ERROR-PRONE REPAIR INDUCED BY MUTANT DNA METHYLTRANSFERASES

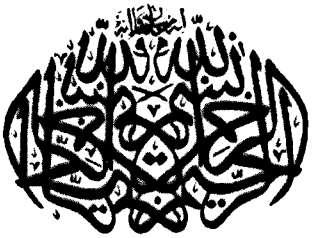
ABDULAZIZ MOHAMMED A. AL-SWAILEM *M.SC.*

Submitted to the University of Sheffield for the degree of Doctor of Philosophy

Department of Molecular Biology and Biotechnology

University of Sheffield

November, 1998



In The Name Of Allah (God), Most Gracious, Most Merciful

To

My parents and my Family

Table of Contents

ACKNOWLEDGEMENT

Firstly, I would like to thank my parents, Mohammed and Latefah, for their continual support throughout my education and their wish that I be happy in whatever I chose to do. I am also grateful to my wife Hayam Al-Amoud and my children Hanan, Faisal, Maryam, Khawlah, and Mohammed for their patience and understanding during the course of this study and expatriation. I am grateful to my brothers, sisters for their support.

I would like to thank Dr. David Hornby, my supervisor for his continuous advice and encouragement with the project, who picked me up when I was in defeat in the middle of my course due to experimental problems. I am indebted to my friends in D29 (not the red D29) who have helped me over this course; Abduaziz Al-Dukhayil, Adel, Qisar, Maryam, Munseng, and Mark. Also I would like to thank Dr. Graham Walker for his advise and support.

Finally, I would like to thanks both King Abdulaziz City for Science and Technology (KACST) for providing financial support for this project and the Molecular biology and Biotechnology department for their support.

Error-Prone Repair Induced by Mutant DNA Methyltransferases

Submitted to the University of Sheffield for the
degree of Doctor of Philosophy

November 1998

ABDULAZIZ MOHAMMED A. AL-SWAILEM *M.SC.*

SUMMARY

Organisms utilise cytosine-5 DNA methylation to expand their repertoire of genetic transactions. Structural studies of DNA cytosine-5 methyltransferase have revealed that DNA methyltransferases incorporate nucleotide flipping into their catalytic cycle in order to access the otherwise buried pyrimidine ring from within duplex DNA. Interestingly, substituting the catalytic nucleophile Cys with Gly can produce cytotoxic forms of the bacterial methyltransferases and cause rearrangements in the DNA. In this study the generality of the cytotoxic effect has been studied on both mono and multi-specific methyltransferases. The effect of dimerisation of methyltransferases on the rearrangement event and the specificity of DNA damage have been defined. The involvement of two DNA repair proteins RecA and UmuDC has been studied. The wild type and mutant multispecific methyltransferase (M.SPRI) has been transcribed and translated *in vitro* and the proteins studied using surface plasmon resonance technique. The experiments described here demonstrate for the first time how a high affinity, catalytically deficient DNA methyltransferase induces error-prone deletions in *E.coli*.

PLASMIDS USED IN THIS STUDY

Abbreviation	Description	Size (bp)
pES1	Gene encoding M.*SPR*I cloned to pET22b	6844
pES1$_{C79G}$K	Gene encoding M.SPRI-glycine mutant inactivated by the kanamycin^r gene cloned into pET22b	8096
pES2	Gene encoding GST-M.*SPR*I cloned into pET22b	7555
pES2$_{C79G}$K	Gene encoding GST-M.SPRI$_{C78G}$ inactivated by the kanamycin^r gene cloned into pET22b	8807
pGM2	Gene encoding M.*Msp*I cloned into pGEX2T	6201
pGM2K	Gene encoding M.*Msp*I inactivated by Kanamycin^r gene cloned into pGEX2T	7301
pGM2d1	Gene encoding M.*Msp*I inactivated by deleting one bp cloned into pGEX2T	6200
pGM2$_{C174G}$K	Gene encoding M.*Msp*I$_{C174G}$ mutant inactivated by kanamycin^r gene cloned into pGEX2T	7371
pGM2$_{C174G}$d1K	Gene encoding M.*Msp*I$_{C174G}$ mutant inactivated by one bp deletion and kanamycin^r gene cloned into pGEX2T	7370
pGS2	Gene encoding M.*SPR*I cloned into pKG	6343
pGS2$_{C79G}$K	Gene encoding M.*SPR*I$_{C78G}$ inactivated by the kanamycin^r gene cloned into pKG	7595
pGS2K	Gene encoding M.*SPR*I inactivated by the kanamycin^r gene cloned into pGK	7595
pLS1$_{C78G}$K	Gene encoding M.*SPR*I$_{C78G}$ inactivated by the kanamycin^r gene cloned into pLitmus28	5461
pLS2$_{C78G}$K	Gene encoding GST-M.*SPR*I$_{C78G}$ inactivated by the kanamycin^r gene cloned into pLitmus 28	6149
pUM1	Gene encoding M.*Msp*I cloned into pUC19	3928
pUM1$_{C174G}$K	Gene encoding M.*Msp*I$_{C174G}$ inactivated by the kanamycin^r gene cloned into pUC19	5098
pUM1$_{C174G}$d1K	Gene encoding M.*Msp*I$_{C174G}$ inactivated by deleting one bp and the kanamycin^r gene cloned into pUC19	5097
pUC19dX	pUC19 with deleted *Xba*I site	2686
pGH2K	Gene encoding M.*Hha*I inactivated by the kanamycin^r cloned into pGEX2T	7223
pGH2$_{C81G}$K	Gene encoding M.*Hha*I$_{C81G}$ inactivated by the kanamycin^r cloned into pGEX2T	7223

Where: **Δ** = Deletion, **1** = monomeric (without glutathione transferase gene), **2** = Dimeric (with glutathione transferase gene), **p** = plasmid, **E** = pET22b, **G** = pGEX2T, **U** = pUC19, **L**= Litmus28, CxG = C → G mutant where x is the amino acid number for C in the Mtase gene, **H** = gene encoding M.*Hha*I, **M** = gene encoding M.*Msp*I, **S** = gene encoding M.*SPR*I, and **K** = kanamycin^r gene

ABBREVIATIONS

AdoMet	S-adenosyl-L-methionine
5-azaC	5-azacytidine
aa	Amino acid
AdoHcy	S-adenosyl-L- homocysteine
amp	Ampicillin
BIA	Biomolecular interaction analysis
bp	Base pair(s)
DNA	Deoxyribonucleic acid
ds	Double strand
DSBs	Double strand breaks
E.C.	Enzyme commission
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetra acetic acid
Enase	Endonuclease
GST	Glutathione-S-transferase (E. C. 2.5.1.18.)
Hr	Hours
IPTG	Isopropyl-1-thio-β-D-galactoside
Kan^r	Kanamycin resistance
Kb	Kilo base pair (1000bp)
LB	Luria Bertani medium
MQ	Milli Q water
Mtase	Methyltransferase
O/N	Over night
OD	Optical density
PCR	Polymerase chain reaction
pfu	Plaque Forming Unit
RM	Restriction and Modification
RU	resonance units
SA	Streptavidin
Sd	Standard deviation
SDS	Sodium dodecyl sulphate
SPR	Surface plasmon resonance
ss	Single strand
TLS	Translesion Synthesis
TRD	Target recognition domain
TRS	Target Recognition Site(s)
UV	Ultraviolet
X-gal	X-galactosidase

All other abbreviations appear in the text or follow the internationally set out rules in "Biochemical Nomenclature and Related Document" (1978).

CHAPTER ONE:
GENERAL INTRODUCTION

Synopsis of Chapter One

- General review of Restriction and Modification Systems.
- General overview of prokaryote DNA Methyltransferases.
- A detailed discussion of C5-Mtase structure and mechanism of action.
- General discussion of DNA Repair.
- Critical review of the SOS system repair pathway.
- Aims of this thesis.

1.1 INTRODUCTION

Following cell division in bacteria, each daughter cell will have received one copy of chromosomal DNA. The enzymatic processes that catalyse and accompany replication include modification, restriction and repair. In the following chapter I will discuss each process briefly dealing with restriction-modification and subsequently DNA repair. The aim of the introduction is to consider how these processes are regulated and how when the genes encoding modification enzymes are mutated, these influence DNA repair processes in limiting the damage inflicted.

1.2 RESTRICTION AND MODIFICATION

The first observations of the phenomenon of restriction and modification (RM) were made between 1952 and 1953 (Luria and Human, 1952; Bertani and Weigle, 1953). Bertani and Weigle observed that stocks of λ phage prepared from *E.coli* strain C grew poorly when propagated on a different strain, K-12. Rare phage that escaped this effect grew with equal efficiency on either host. They coined the term "restriction" to describe this observation. It is now known that specific enzymes, restriction endonucleases, that catalyse the hydrolysis of specific DNA sequences underpin this phenomenon.

Twelve years later Gold and Hurwitz (Gold and Hurwitz, 1964) were the first to propose that a DNA methyltransferase (Mtase) was responsible for the phenomenon of modification in *E.coli*, and later Kuehnlein et al. (1969) isolated an enzyme from *E.coli* which could methylate DNA. Finally this enzyme was purified by Kuehnlein and Albert (1972). Similarly, during this period, reports were published concerning DNA methyltransferases (Mtases) associated with murine chromatin and solubilised rat liver extracts. By 1975 a human DNA Mtase had been purified by Roy and Weissbach see (Weissbach, 1993).

Restriction is the result of the action of an endonuclease that recognises a specific DNA sequence and cleaves as a consequence of this recognition event.

Modification is catalysed by DNA Mtases, which recognise and methylate the same DNA sequence to protect the host chromosome from endonuclease action. The precise cleavage pattern that is associated with the restriction enzyme pathway defines the category into which the enzyme is placed. Phage or any foreign DNA, become resistant to restriction ocurs when the Mtase modifies each phage borne recognition site prior to endonuclease action (Heitman, 1993).

1.2.1 BIOLOGICAL ROLES OF RESTRICTION AND MODIFICATION SYSTEM.

The widespread occurrence of methylated bases in DNA of various organisms, the sequence specificity of the various DNA Mtase and the non-random distribution of the methylated bases along the chromosomes strongly suggest that modified bases in DNA are biologically significant.

Although, in general the biological significance of methylated bases in DNA is still an enigma, a number of recent experiments strongly suggest that DNA methylation maybe involved in major biological processes. Since the methyl groups lie in the major groove of the double helix, they are exposed to the surroundings and could serve as signals influencing protein-DNA interaction. Some of the proposed biological roles of RM are:

1- **A primitive prokaryotic: "Immune system"** The only biological phenomenon, for which it has been shown unambiguously that methylation of DNA plays a role, is that of restriction and modification in bacteria. Modification of specific DNA sequences render those sequences refractory to the cognate endonuclease. Restriction endonucleases hydrolyse the foreign DNA that enters the cell, thereby protecting the cell from genetic subversion. Mtases modify the cell's own DNA thereby protecting it from similar destruction (Smith, 1979). However this role of Mtases has been re-considered after the finding that some RM systems behave as "selfish" genes (see section 1.2.6).

2- **Regulation of gene expression:** The expression of several genes from *E.coli* and its bacteriophages and transposons is affected by DNA methylation (Marinus, 1987). These changes in the rate of gene expression are frequently due

to the presence of *dam* recognition sites within the binding sites of proteins such as RNA polymerase (Russell and Zinder, 1987).

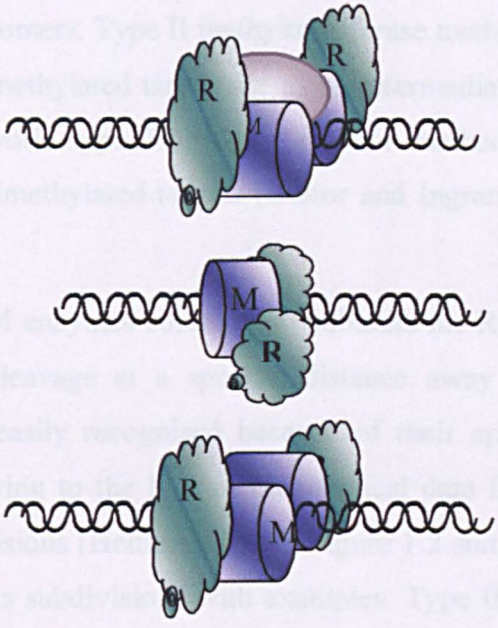
3- **Initiation of replication:** In prokaryotes such as *E.coli* (and genetic elements like plasmid P1) the origin of replication is characterised by the presence of multiple GATC sequences: this is the sequence that is methylated by the Dam Mtase (see below). As a consequence of this, methylation phenomenon, replication synchrony is abolished in the absence of the enzyme (Adams, 1990). DNA replication is dependent upon the presence of a specific transcript, which starts outside the origin region and terminates at a number of sites within the origin where DNA synthesis is initiated. These sites include the multiple GATCs, and it is possible that these sequences are recognised by an initiation protein when sites are fully methylated or that methylation reduces the stability of the duplex, thereby allowing unwinding to occur (Barras and Marinus, 1989; Adams, 1990).

4- **DNA Repair**: When a normal, but incorrect base is introduced into DNA at replication, it is usually removed by a proofreading mechanism before the next nucleotide is added. When this mechanism fails, the product is a duplex DNA molecule with a mismatch. A repair system, which can correct the mismatch, is therefore required to discriminate between two normal bases, one of which is incorrect. A random process would accentuate the error in 50% of cases, but a mismatch correction system found in *E.coli* utilises the lack of methylation on the newly synthesised (incorrect) strand to determine which strand requires repair (Adams, 1990) (see section 1.4 for details).

1.2.2 RESTRICTION AND MODIFICATION SYSTEMS: NOMENCLATURE AND PHYSICAL PROPERTIES

Several kinds of RM enzymes have been discovered, more than 100 RM genes have been cloned and approximately fifty have been sequenced. Roberts, (personal communication) has proposed a new nomenclature for RM enzymes based on the old nomenclature with some modifications. Based on the properties of the purified enzymes, RM enzymes fall into three major classes: type I, type II, and type III. Each division can be classified into subdivisions, reflecting the biochemical and sequence recognition characteristics of the enzyme (see figure 1.1 and 1.2).

Figure 1.1: Diagram shows the three RM classes.
The difference between the three types of RM are illustrated, blue indicates the Modification (M) subunits, green Restriction (R) subunits and gray specificity (S) subunit.



Type I RM enzymes, are the most complex and consist of three subunits (R = restriction, M = modification, and S = specificity) which form pentameric (R_2M_2S) multifunctional enzyme complexes. These complexes cleave at random sites distant from the site of recognition, unwind DNA to form a loop between sites of recognition and cleavage, hydrolyse ATP, and are inactive after a single round of DNA scission. This type contains one subdivision namely IB Figures (1.1 and 1.2). Type I enzymes, require S-adenosyl-L-methionine (AdoMet) for both cleavage and modification. Cleavage occurs at considerable, and variable, distances from the recognition sequence. The recognition sequences are asymmetric and bipartite; they comprise two sub-sequences, three and four bp in length, separated by six to eight non-specific bp. The S subunit determines specificity for both restriction and modification (Daniel et al. 1988).

Type II RM enzymes, consist of two separate enzymes, which operate independently: the endonuclease cleaves at unique sites either within or close to their recognition sequence. This type can be divided into ten subdivisions (figure 1.2). Type II Mtases, catalyse the methylation of sites at adenine or cytosine at the N6-mA or 5-mC and N4-mC respectively (figure 1.3), they require AdoMet for modification. The recognition sequences are essentially symmetric, comprising four to eight specific bp, (sometimes they may include additional bp in the form

of non-specific interruptions). The endonuclease cleaves symmetrically (except IIs) within the sequences and the endonucleases act mainly as homodimers while the Mtases are monomers. Type II methyltransferase methylate their targets in two steps, with a hemimethylated target site as an intermediate reaction product. The activity of prokaryotic type II Mtases are not markedly different with non-methylated or hemimethylated targets (Bestor and Ingram, 1983) (See figure 1.1 and table 1.2).

Type III RM enzymes contain two subunits for R and two for M (R_2M_2), require ATP for cleavage at a specific distance away from their recognition sequence and are easily recognised because of their apparent similarity at the sequence level. Owing to the lack of biochemical data for this type of enzyme, there are no subdivisions (Heitman, 1993). Figure 1.2 summarise the specification for each type and its subdivisions with examples. Type III RM enzymes produce N6-mA (Meisel et al. 1991), requires AdoMet for cleavage and unlike type II enzymes modification is stimulated by AdoMet. The M subunits also methylate DNA in the absence of R. Restriction on the other hand requires the association of the R and M subunits: cleavage requires ATP. The recognition sequences are asymmetric, uninterrupted, five to six bp in length and contain A only in one strand therefore only one strand of any target sequence can be methylated. Cleavage occurs approximately 25 bp to one side of the sequence. Only one strand of the recognition sequence becomes methylated, in apparent violation of the rule that both strands must be methylated to preserve modification during replication. However, cleavage takes place only when two unmethylated sites are present in the DNA, in opposite orientations. Since one site or the other remains modified after passage of the replication fork, modification is preserved during replication (Krüger et al. 1990).

1.2.3 GENETICS OF RESTRICTION AND MODIFICATION SYSTEMS.

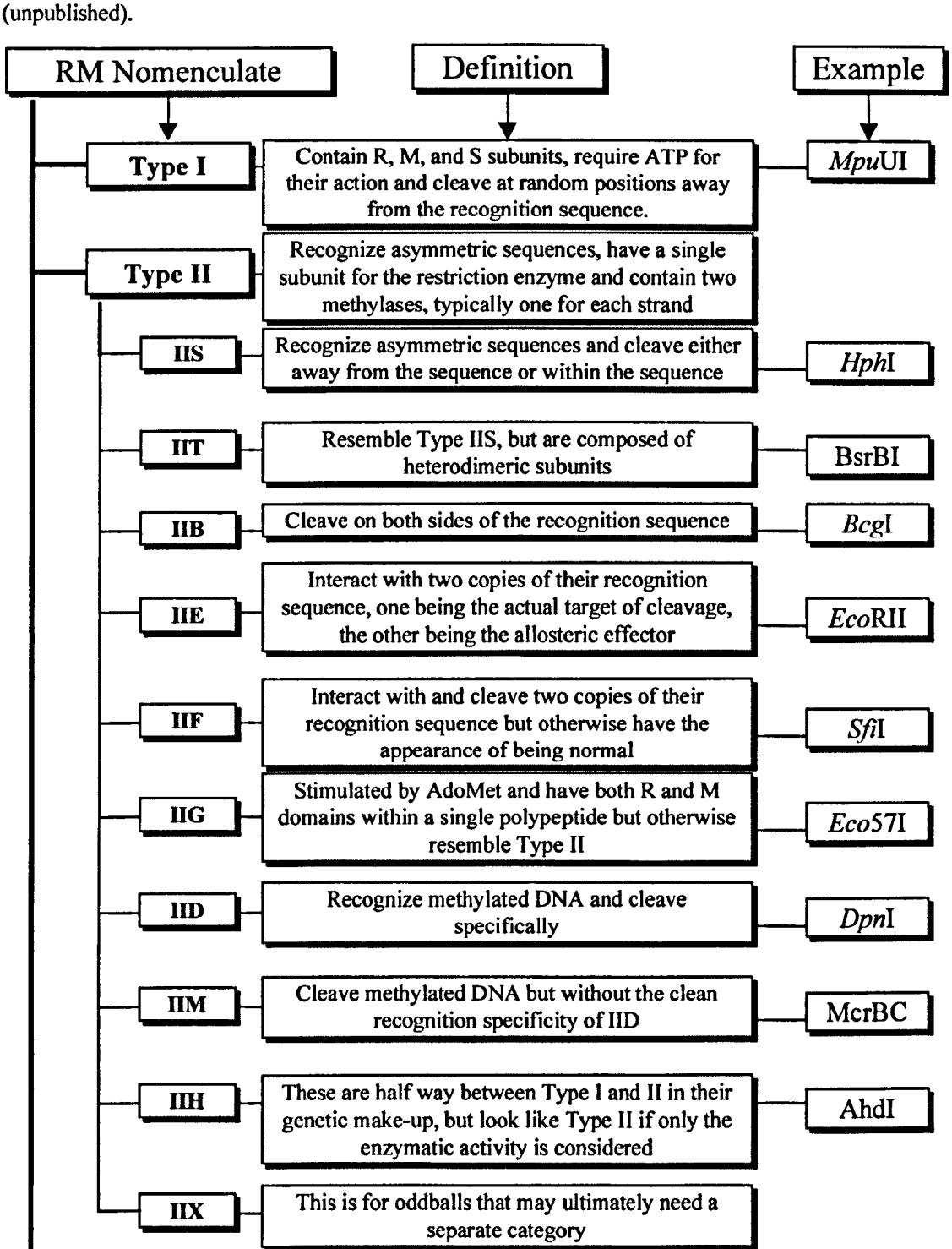
The M and S genes in type I enzymes are transcribed as a single operon, and the R gene is transcribed separately but in the same direction. However, in type II systems the R and M genes are linked. Sometimes the R gene precedes M; at other times the M gene precedes R and in several systems the genes have

Table 1.2: Some physical and genetic properties of RM systems from (Trautner and Noyer-Weidner, 1993; Wilson, 1991) with modification.

Types of R/M system	Structure	Co-factor requirement		Type of methylation observed	Location of target recognising domain(s)	Site of cleavage with respect to recognition sequence	Genetic
		R	M				
I	Enzyme complex consisting of Enase, MTase and specificity determining subunits	ATP SAM Mg^{++}	SAM ATP Mg^{++}	N6-mA	Within specificity determining subunit	Remote from target	M and S proteins transcribed as same operon, R transcribed as another operon in the same direction.
II	Enase and MTase represent separate enzymes	Mg^{++}	SAM	5-mC, N4-mC N6-mA	Within individual enzymes	Within the target or at defined location within 20 bp on one side of the target	R and M genes occur in all linkage configurations
III	Enzyme complex consisting of ENase and MTase subunits	ATP SAM Mg^{++}	SAM ATP Mg^{++}	N6-mA	Within MTase subunit	At a defined distance of 25-27 bp from the 3` site of the target R requires two targets in opposite orientation for scission	R and M transcribed as same operon

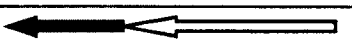
Figure 1.2: The three class of RM system.

Diagram shows the three types of RM system I, II, and III definition of each division and subdivision are illustrated with one example for each. This classification was based on Robert's proposal (unpublished).



opposite orientations; some diverge while others converge. Type III R and M genes are transcribed as a single operon in the same orientation (Wilson, 1991a) (see table 1.1)

Table 1.1:Prototype target sequences and genes structures of RM systems. The specificity of the RM systems shown is indicated, with the methylated bases underlined. The arrows give the direction of gene transcription. Open, solid, and shaded arrows describe genes for Mtases, Enases and the specificity determining subunits. The numbers of aa of the corresponding proteins is given below the arrows (Noyer-Weidner and Trautner, 1993).

System	Genes	Specificity	Gene organisation
Type I *EcoK*	RMS	AACN$_6$GTGC TTGN$_6$C<u>A</u>CG	 1090 bp 529 464
Type II M.*Msp*I	RM	C<u>C</u>GG	 262 418
TypeII *Bsu*BI	RM	CTGC<u>A</u>G	 311 501
Type III *EcoP*I	RM	AG<u>A</u>CC	 970 646

1.2.4 RESTRICTION OF METHYLATED DNA

The observation that Mtase encoding genes could not be cloned in some *E.coli* strains led to the discovery of three *E.coli* K-12 restriction systems directed against methylated DNA (Heitman and Model, 1987). Restriction of methylated DNA by *E.coli* depends on the products of the *mcrA* (modified cytosine restriction) gene, the *mcrBC* operon and the *mrr* (modified adenine recognition and restriction) gene (Noyer-Weidner and Trautner, 1993). Depending on the DNA modification type eliciting restriction, these genes are generally referred to as Mcr, if restriction depends on the presence of modified cytosine in DNA, or Mrr, if restriction depends on adenine modification. Further sub-specification of phenotypic designations (e.g., McrA, McrBC or McrF) reflects differences in the specificities of the *mcrA, mcrBC* and *mrr* encoded restriction activities for DNA substrates carrying an identical type of modified base (e.g. 5-mC).

So far only the McrBC restriction enzyme has been characterised biochemically (Sutherland et al. 1992; Stewart and Raleigh, 1998). DNA cleavage by McrBC depends on the presence of the general sequence 5'$R^mC(N_{40-80})R^mC3'$. In response to two appropriately spaced methylated cytosines, even if present in only one strand of DNA, the McrBC activity catalyses double-strand breaks at multiple positions of the intervening region. ATP, which serves as a cofactor for most complex Enases, inhibits McrBC restriction. One unique feature of McrBC restriction is its absolute dependence on GTP, which is hydrolysed in the reaction (Sutherland et al. 1992)

The biological function of the nonessential *E.coli* restriction activities directed against methylated DNA might be one or more of the following:

1- Protection of *E.coli* strains against infection by phages which escape host restriction by incorporation of modified bases into their DNA (Krüger and Bickle, 1983)

2- DNA repair: by counteracting the mutagenic potential derived from spontaneous deamination of 5-mC to thymine using the capacity of the McrA, McrBC and McrF activities to restrict and remove 5-mC containing DNA (Wilson and Murray, 1991b).

Modification-dependent restriction activities have also been used to facilitate gene cloning. The McrBC activity served to establish an *E.coli* K-12 vector system allowing the positive selection of recombinant plasmid. The isolation of strains encoding temperature-sensitive McrA and McrBC activities has further provided the basis for the construction of strains facilitating the identification of recombinant DNA encoding non *E.coli* specific DNA-Mtases (Noyer-Weidner and Trautner, 1993).

1.2.5 ARE RESTRICTION AND MODIFICATION GENES "SELFISH"?

One of the possible roles of RM systems is to protect bacteria from infection by foreign DNA. This hypothesis has been challenged by Naito et al. (Naito et al. 1995), who question " whether such a cellular defence hypothesis explains the extreme diversity and specificity of sequence recognition". Two

groups have provided convincing support for the idea that type II RM systems can serve to maintain the presence of the plasmid that encodes them (Naito et al. 1995; Kulakauskas et al. 1995). Their data are consistent with the concept of a restriction enzyme as a "toxin" that is normally neutralised by an "antidote": the modification enzyme. When a cell continues to divide following the loss of the plasmid, the level of antidote may become insufficient to protect all the restriction targets from the endonuclease, and consequently, cells lacking the plasmid will die. Type II RM genes that behave in this may be regarded as "selfish" because there would be direct selection for their retention rather than their presence being maintained as a consequence of the advantage they confer on the host bacterium. O'Neill et al. (1997) tested this hypothesis with the type I RM genes encoding *Eco*KI. At both the plasmid and the chromosomal level, they found that type I genes are not "selfish" when plasmid-borne. They also found that functional type I RM genes when present on the chromosome, are readily replaced by mutant alleles and by alleles encoding a type I RM system of different specificity (O'Neil et al. 1997). There are therefore two hypotheses, which may account for the prevalence of RM genes: one is as a "cellular defence" and the other, the "selfish gene" hypothesis (Kobayashi, 1996), clearly more experiments are required to resolve these issues.

1.3 DNA METHYLATION

It is well known that DNA is a very stable molecule; yet with each round of replication, this structure is modified by base methylation in nearly all cells and organisms. In this section methylation will be discussed from both a chemical and biological viewpoint with particular, reference to type II DNA Mtases.

1.3.1 THE CHEMISTRY OF DNA METHYLATION

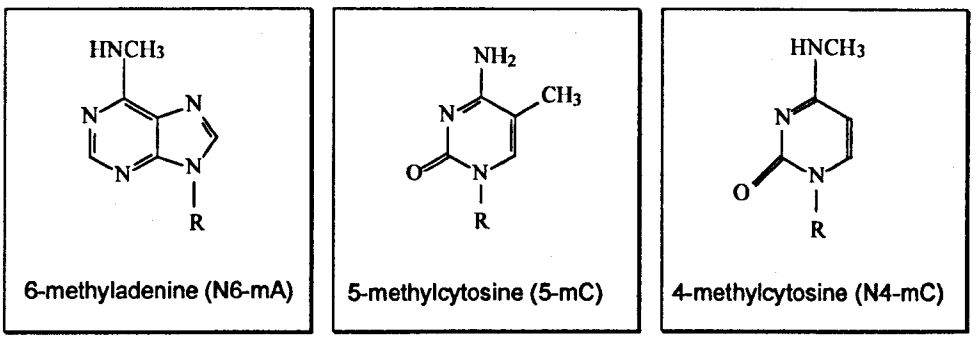
As well as the four major bases, the DNA of most organisms contains one or more minor bases. The commonest of these is 5-methylcytosine (5-mC), but N^6-methyladenine (N6-mA) and N^4-methylcytosine (N4-mC) are also found, particularly in prokaryotes (figure 1.3). The large spectrum of different Mtases identified in the prokaryotes can be attributed to three factors related to the specificity of these enzymes:

1- There are three categories of Mtases: those which generate 5-mC, N4-mC and N6-mA.

2- Mtases of the three categories methylate a particular base only when it is part of a DNA "target" sequence specifically recognised by each enzyme.

3- There are many isomethylomeric Mtases, i.e. enzymes from different sources with the same specificity.

Figure 1.3: Diagram showing the three methylated bases.



In vertebrates 3-6% of DNA cytosine is methylated, but this value decreases in many insects and in single-celled eukaryotes there is no detectable 5-methyl cytosine (Adams, 1990). In contrast, plant may have 30% of their DNA cytosine methylated (Adams and Burdon, 1985)

Methylation by type II Mtases requires AdoMet as a methyl group donor. However, in the case of the *E.coli* Dam Mtase, (and type I restriction enzymes) AdoMet is also an allosteric effector (Bergerat et al. 1991)

1.3.2 PRIMARY STRUCTURE MOTIFS IN C5-MTASES

A comparison of some 36 mono and multi-specific C5-Mtases indicated that amino acid sequences, which were highly conserved amongst all of these enzymes, alternated with less conserved regions. This pattern of enzyme structure was recognised independently in several laboratories (Som et al. 1987; Lauster et al. 1989; Posfai et al. 1989). In the absence of uniform criteria in the definition of conserved and variable sequence motifs, different consensus patterns have been

proposed by these laboratories in the description of conserved and variable elements. In this section the nomenclature advocated by Posfai will be followed (Posfai et al. 1989).

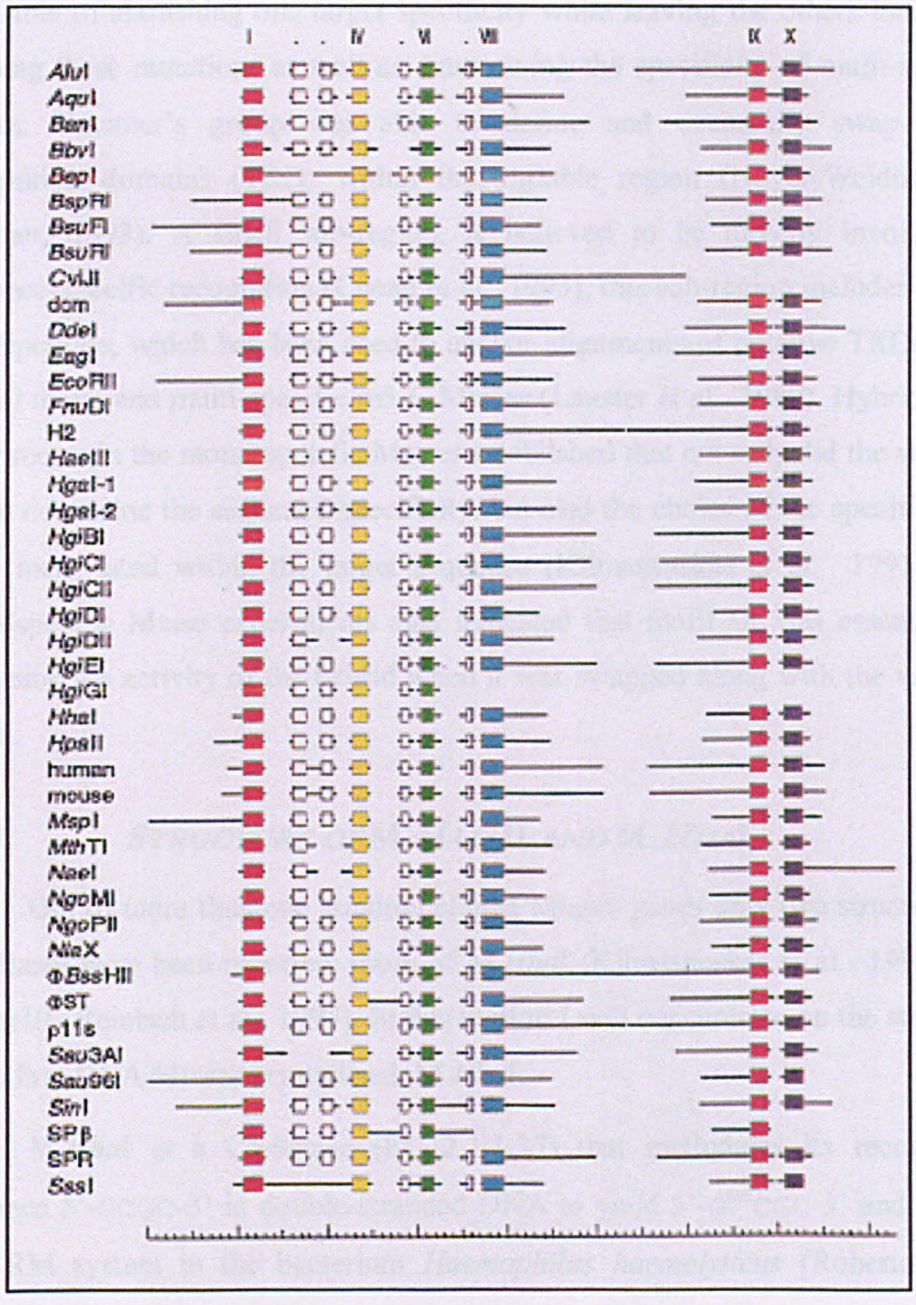
Ten conserved regions (I-X) have been found to be arranged in the same order, within a protein of about 40 kDa (Kumar et al. 1994). The few eukaryotic DNA C5-Mtases for which sequence information is known have at least eight of these conserved regions concentrated in the C-terminal part of the protein. In the original analysis of 13 m5-C Mtases, five motifs were considered highly conserved. Re-analysis of 36 sequences suggested that there are six rather than five well-conserved motifs (Kumar et al. 1994). Figure 1.4 shows a schematic diagram of the ten conserved motifs in a representative subset of 12 m5-C Mtases. This is abstracted from an updated alignment of the published sequences of 36 m5-C Mtases (Kumar et al. 1994). The six most strongly conserved motifs I, IV, VI, VIII, IX, and X show only minor changes in the degree of conservation of residues (Figure 1.4) However, a few sequences have been detected in which unambiguous assignments for some of the motifs are not possible. These motifs II, III, IX, and X are either more tolerant to sequence variation or nonessential for function in some of the proteins.

The distinction within the primary amino acid sequence of conserved and variable regions suggested a modular structure of C5-Mtases, in which a core of conserved regions determining universal steps in the DNA methylation reaction had become associated with different variable regions.

Motif I FXGXG (where X is any amino acid) forms a critical part of the AdoMet binding site, Motif IV contains an invariant PC dipeptide that is known to be part of the catalytic site, which is responsible for covalent binding to the target cytosine in DNA (figure 1.5). Substitution of the Cys in this dipeptide leads to loss of function in M.*Eco*RII, M.*Hha*I, M.*Hae*III, M.*Msp*I, and Dcm. Direct identification of the residue covalently attached to carbon-6 in the trapped intermediates formed with the suicide substrate 5-flouro-2-deoxycytidine (see section 1.3.5.1) (Wyszynski et al. 1991) confirmed that this mechanism proceeds through a covalent enzyme:DNA intermediate. This covalent bond between Cys81 and carbon-6 of cytosine is clearly visible in the M.*Hha*I-DNA structure, since an

Figure 1.4. Schematic shows the alignment of 45 m5C-Mtases.

The six highly conserved motifs are coloured (red-motif I, FGG; yellow-motif IV, PC; green-motif VI, ENV; cyan-motif VIII, QRR; magenta-motif IX, RE; blue-motif X, GN). Where boxes are missing, the motif could not be unambiguously assigned. Breakpoints within unaligned segments were arbitrarily placed in the centre of the segments (Kumar *et al.*, 1994).



oligonucleotide containing 5FC was used for co-crystallisation (Klimasauskas et al. 1994).

The variable region, located between conserved motifs VIII and IX, shows the greatest heterogeneity in size, sequence and composition among C5-Mtases (see figure 1.4). In multi-specific enzymes, a point mutation in the variable region is capable of abolishing one target specificity while leaving the others intact. By mapping these mutations as well as determining the specificity of multi-specific Mtases, Trautner's group was able to define and eventually swap target recognition domains (TRD) within the variable region (Noyer-Weidner and Trautner, 1993). A small sub-region, is believed to be directly involved in sequence-specific recognition (Cheng et al. 1993), this sub-region includes a Thr-Leu dipeptide, which has been used to anchor alignments of putative TRDs from several mono and multi-specific m5-C Mtases (Lauster et al. 1989). Hybrid swap experiments in the mono-specific Mtases established that not only did the variable region determine the sequence-specificity, but also the choice of the specific base to be methylated within the target sequence (Klimasauskas et al. 1991). The mono-specific Mtase experiments also indicated that motif IX was essential for increasing the activity of the hybrid when it was swapped along with the variable region.

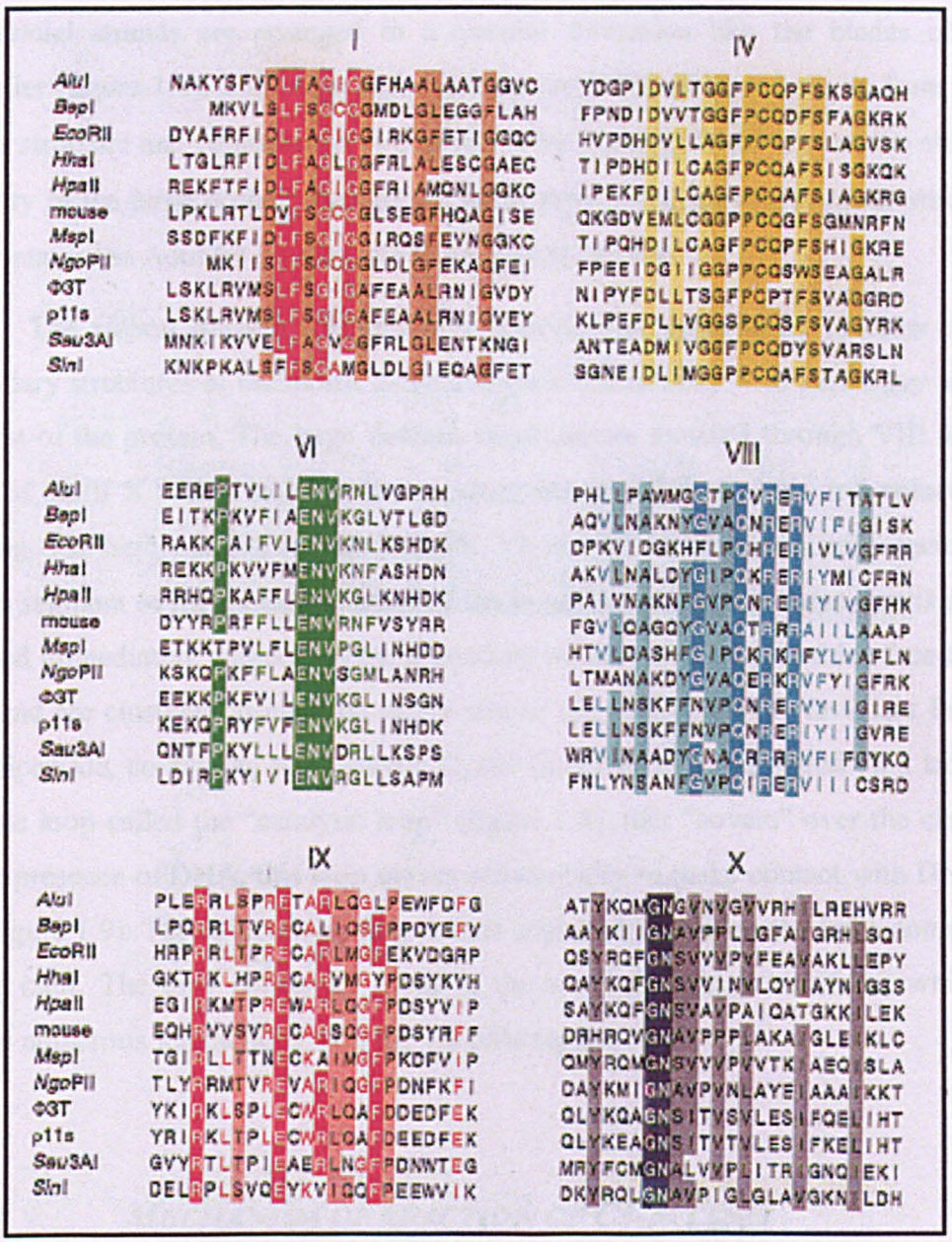
1.3.3 STRUCTURE OF M.*HAE*III *AND* M.*HHA*I

Out of more than one hundred cloned Mtases genes only two structures of C5-Mtases have been reported: those of M.*Hha*I (Klimasauskas et al. 1994) and M.*Hae*III (Reinisch et al. 1995). In this section I will concentrate on the structure of the first DNA Mtases crystallised, M.*Hha*I.

M.*Hha*I is a C5-Mtase (EC 2.1.1.37) that methylates its recognition sequence 5'-GCGC-3' in double-stranded DNA to yield 5'-G^{me}CGC-3' and is part of a RM system in the bacterium *Haemophilus haemolyticus* (Roberts et al. 1976). M.*Hha*I is one of the smallest of the C5- Mtases, containing 327 amino acids having a MW of 37000 kDa. The structure of M.*Hha*I is illustrated schematically in figure 1.6. The molecule, with dimensions of 40 x 50 x 60 Å, is folded into three parts, a large domain, a small domain, and a hinge region,

Figure1.5. Representative sequences of the six highly conserved motifs.

Four degrees of conservation ranging from complete to none are denoted in order of decreasing conservation by: solid (dark) coloured background with white text, stippled coloured background with black text, white background with coloured text, and white background with black text. In the case of motif IV, only black text was used to aid visibility, so only three levels of conservation are indicated. The colour scheme is as in figure. 1.4 (Kumar *et al.,* 1994)



corresponding to the three sides surrounding a cleft, which is wide enough to bind DNA. The large domain (amino acids 1-193 and amino acids 304-327) is a mixed α/β structure consisting of the N-terminal two-thirds of the protein followed by a crossover connection to an additional α helix from the C-terminus. The small domain (amino acids 194-275) consists of seven strands, five up and down antiparallel strands are arranged in a circular formation like the blades of a propeller (figure 1.6). The hinge region (amino acid 276-303), is built up from an α-β-α structure and connects the two domains by forming the bottom of the cleft. Activity in the large domain face of the cleft, next to the carboxyl end of strand β1, contains the AdoMet binding site (Cheng et al. 1993).

The ribbon diagram (Figure 1.7) provides an opportunity to view the secondary structures of the motifs as well as their interactions with each other and the rest of the protein. The large domain encompasses motifs I through VIII and most of motif X which includes the residues essential for catalysis and cofactor binding. The highly conserved motifs I, IV, VI, and VIII form the "core" structure of this segment of the molecule. Most of the invariant residues in these motifs are situated immediately adjacent to the secondary structures in loops, which face the cleft and are clustered around the active site of the molecule. The invariant Pro-Cys dipeptide, containing the catalytic Cys81 (in motif IV), is situated on a large flexible loop called the "catalytic loop" (figure 1.8), that "hovers" over the cleft. In the presence of DNA, this loop moves substantially to make contact with DNA (see figure 1.9). The cofactor-binding pocket is embedded within the large domain in the cleft. The only conserved motif in the small domain is motif IX, which makes numerous interactions with the variable region.

1.3.4 MECHANISM OF REACTION OF C5-MTASES

Cytosine-5 methylation has been termed a "chemically improbable reaction", as it must overcome formidable energetic and stereo-chemical barriers. Santi et al. (1983) solved the energy problem by proposing that DNA cytosine methyltransferases might use a reaction mechanism analogous to that of thymidylate synthetase, and suggested the C5 position of cytosine is activated by covalent addition of an enzyme cysteine thiolate to the neighbouring C6 position

Figure 1.6: Schematic drawing of the secondary structure of M.*Hha*I.

The two structural domains of the molecule are seen clearly. Rectangles and arrows indicate α helices (lettered) and β strands (numbered), respectively. The thick dashed lines indicate hydrophobic interactions, and the thin dashed lines indicate electrostatic interactions or hydrogen bonds. The colour scheme is as in figure 1.4

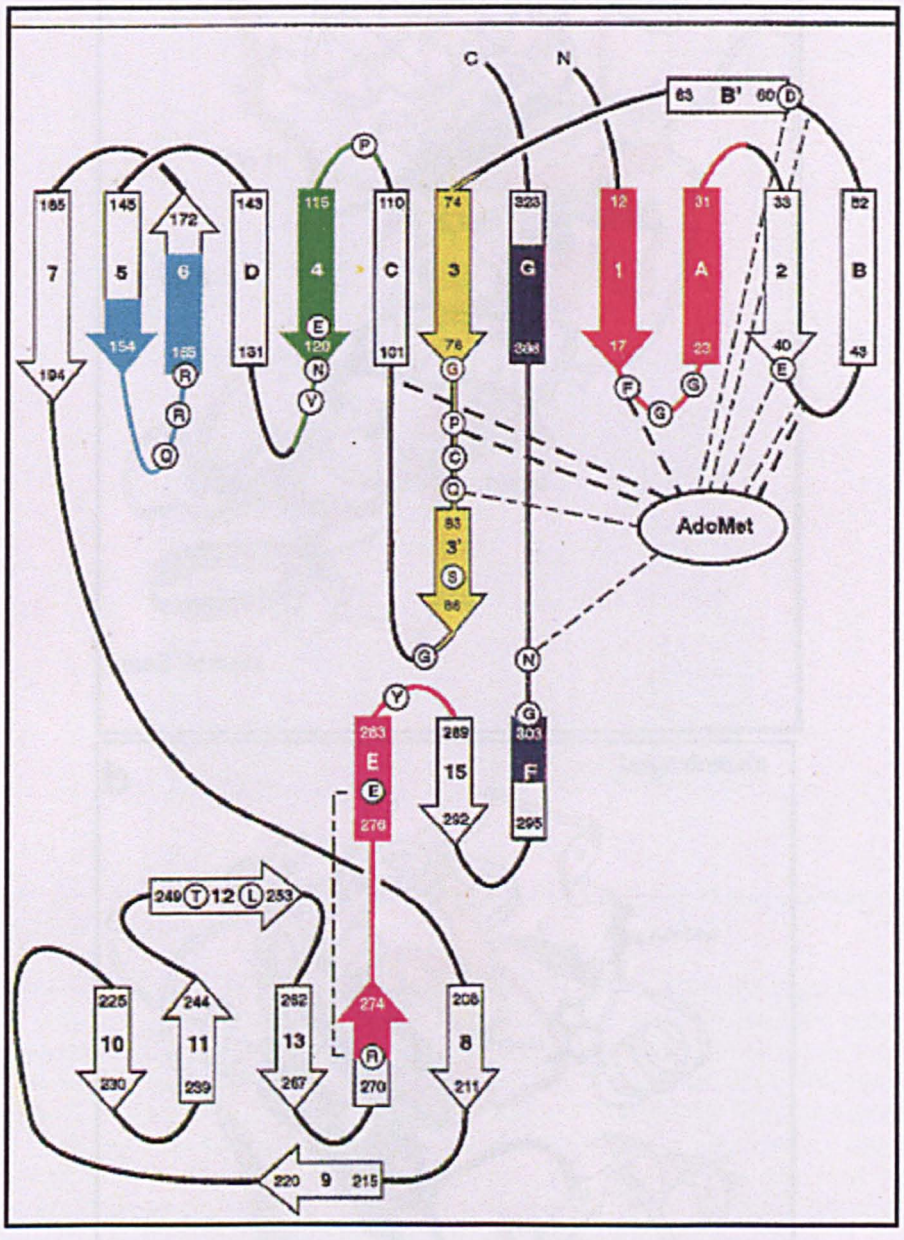
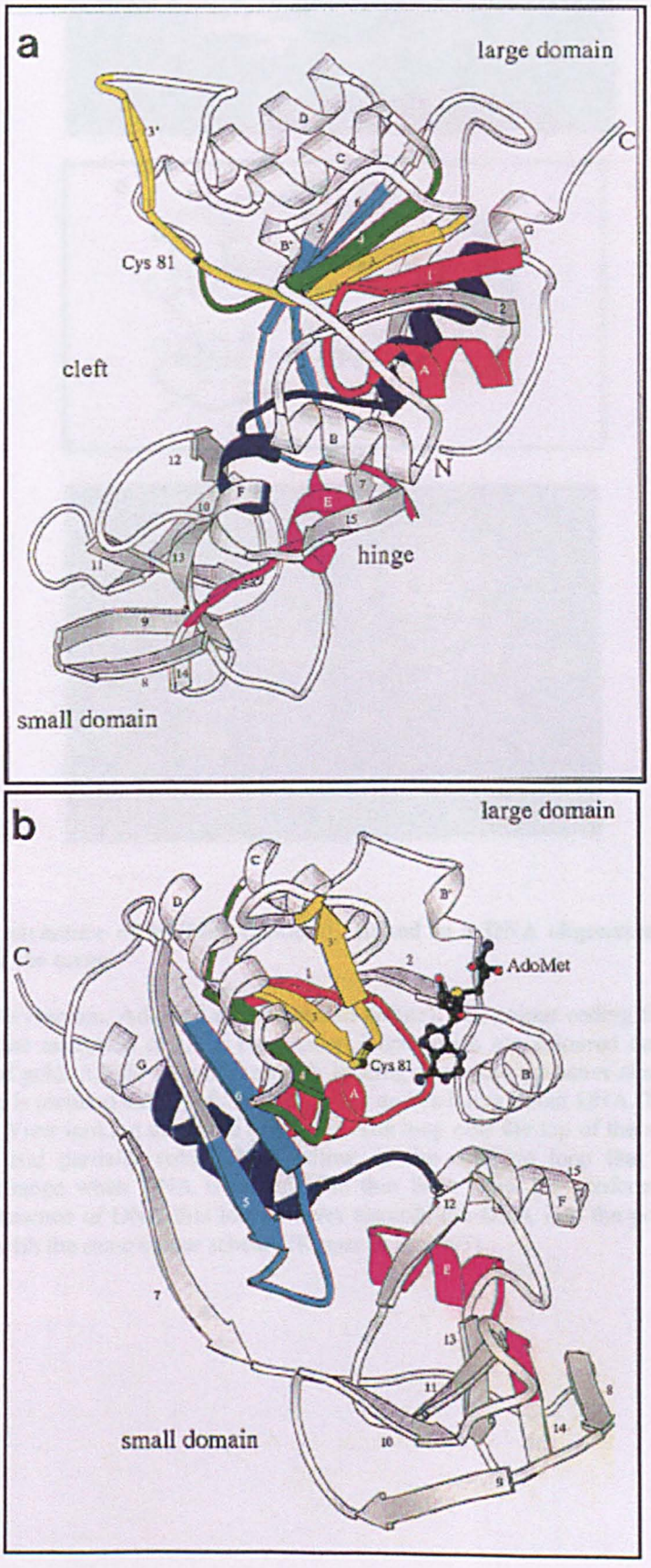


Figure 1.7. Ribbon diagrams showing the domain organisation of M.*Hha*I

The large (amino acids 1-193 and amino acids 304-327) and small (amino acids 194-275) domains are marked. The SAM is omitted from (a). The motifs are coloured as in figure 1.4. The cleft is large enough to accommodate DNA. Side view showing the proposed DNA-binding cleft. Shown in (b) is the view into the cleft.



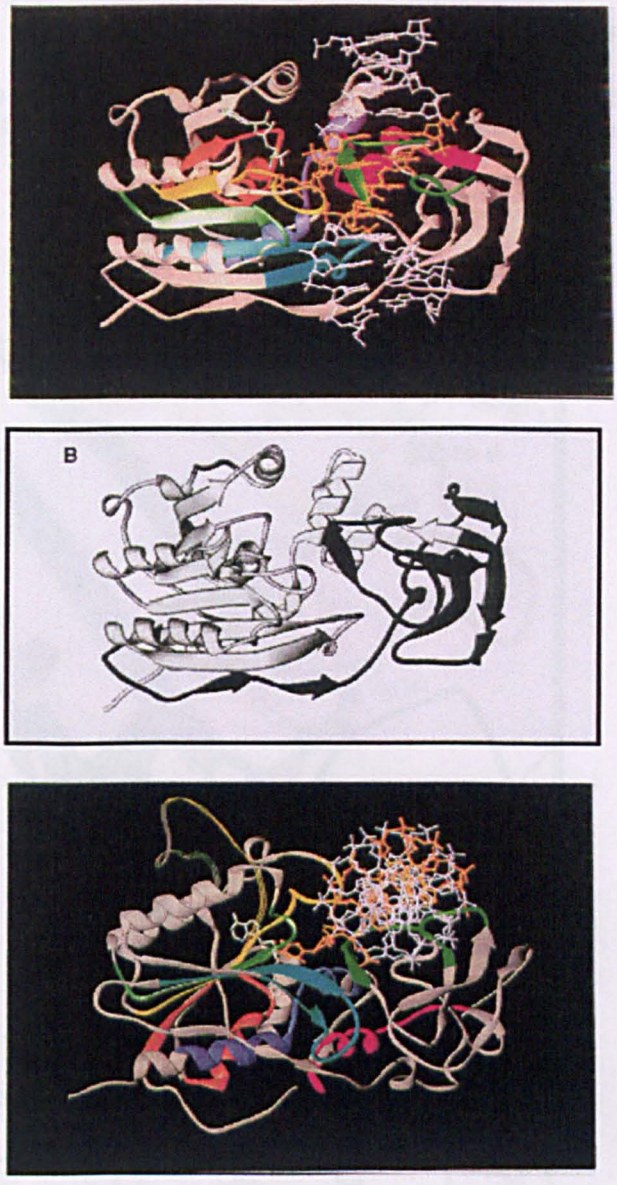
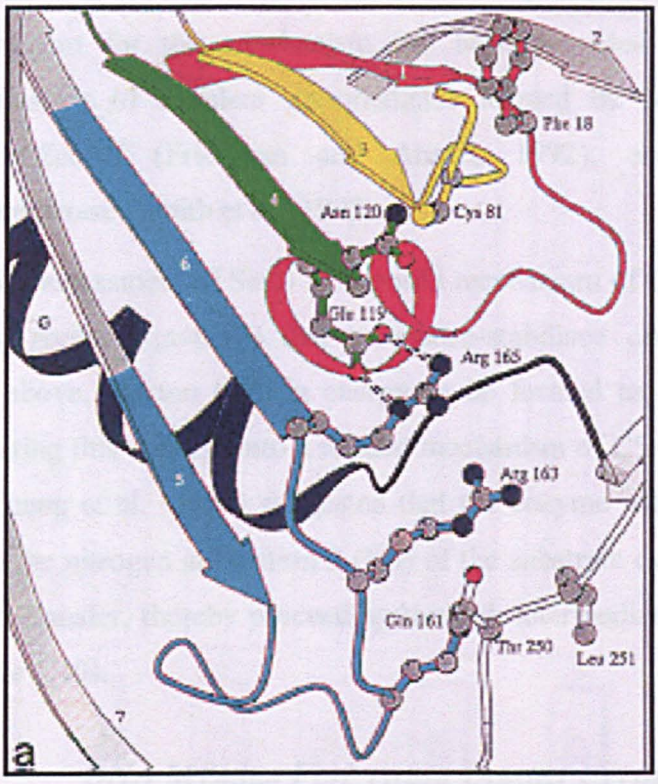


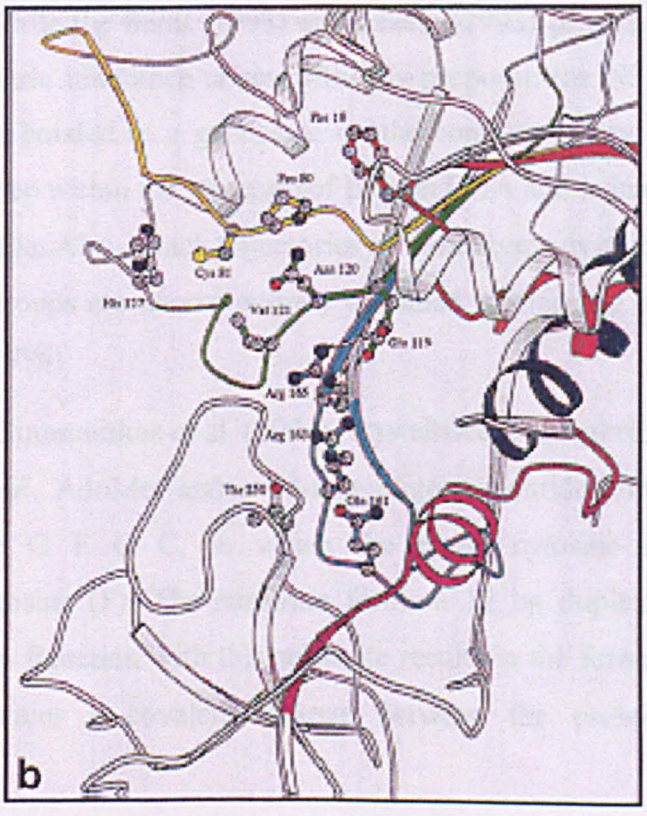
Figure 1.8. The structure of M.*Hha*I covalently linked to a DNA oligonucleotide containing 5-fluorocytosine at the target.

The product of the reaction, AdoHcy, also presented (white). The colour coding for the motifs is as in figure 1.4, with the exception of DNA recognition loops which are coloured dark green. The target sequence coloured gold. A Side view of the DNA, looking down into the active site cleft of the enzyme. The large domain is pictured on the left. B. Same view as in a but with out DNA. The variable region is shaded black. C. View looking down the helix axis. The loop near the top of the structure, drawn with three thin lines and partially coloured in yellow, is the catalytic loop that undergoes a major conformational change when DNA is bound. The thin lines show the conformation without DNA present. In the presence of DNA this loops moves towards the DNA into the position shown by the solid-filled loop with the same colour scheme (Kumar et al. 1993).

Figure 1.9. Structure of the DNA-binding cleft of M.*Hha*I

Key residues and motifs are coloured as in fig 1.4. (a) View into the cleft from the small domain. (b) Close up view from an angle similar to figure1.7 a showing the key conserved residues (Cheng *et al.*, 1993).





to create the C5 carbanion (carbon-centered anion), which would then attack the sulphur-linked methyl group of AdoMet. Following methyl transfer, abstraction of the C5 proton would restore the C5-C6 double bond and allow release of the covalent intermediate as a free enzyme by β elimination (Santi et al. 1983) (figure 1.10). This intermediate can be trapped if the target cytosine base has a fluorine atom substituted for a hydrogen atom at the C5 position (Osterman et al. 1988). Direct support for this mechanism has been provided by the isolation and characterisation of covalent intermediates formed by M.*Hae*III (Chen et al. 1991), *M.EcoR*II (Friedman and Ansari, 1992), and the human DNA methyltransferase (Smith et al. 1992).

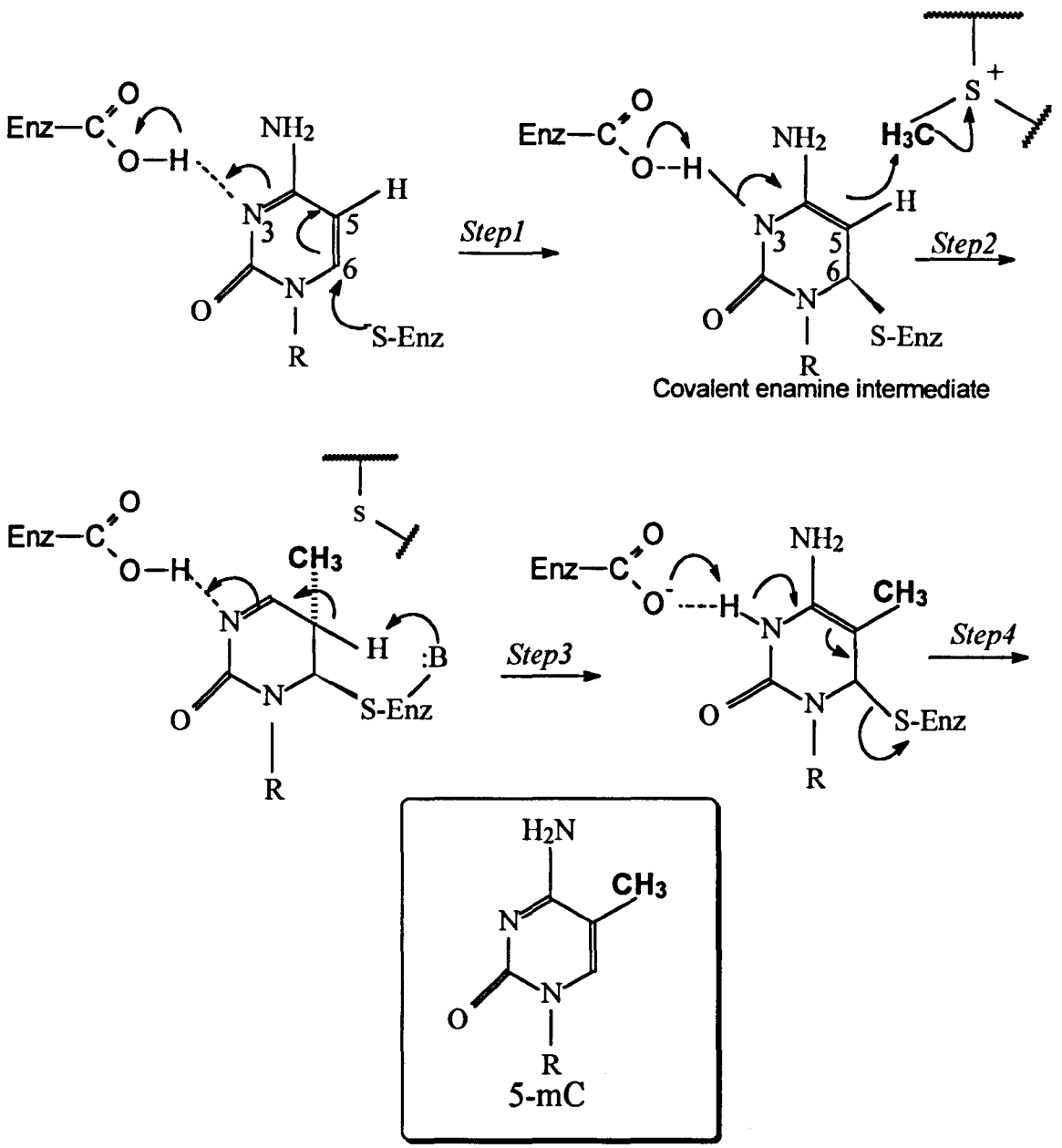
A reassessment of Santi's proposed mechanism of C5-Mtases is warranted owing to recent arguments that resonance-stabilised carbanions, of the kind invoked above, are too high in energy to be formed as discrete intermediates. Incorporating this concept into a revised mechanism of C5-Mtases, Cheng and his group (Cheng et al. 1993) suggested that the enzyme must transfer a proton to and from the nitrogen at position 3 (N3) of the substrate cytosine during catalysis of methyl transfer, thereby proceeding through intermediates known as enamines (see figure 1.10).

1.3.4.1 DNA MTASES FLIP THEIR TARGET CYTOSINE

While the Santi (1993) and Cheng (1993) proposals are widely accepted, severe steric hindrance arises with this proposal: the N3 position of cytosine is hydrogen bonded to a guanosine on the complementary strand, and is therefore buried deep within the structure of B form DNA and is inaccessible to the enzyme nucleophile. Also, attack trajectories of the enzyme cysteinyl thiolate and AdoMet methyl groups are almost entirely occluded by flanking nucleotides (Nelson and Bestor, 1996)

Klimasauskas et al. (1994), crystallised a complex formed upon incubation of M.*Hha*I, AdoMet and a 13-mer oligo-nucleotide containing the recognition sequence G F G C, in which the target cytosine was substituted by 5-flourocytosine (F). The substrate forms a 12 bp duplex with a single base 5' overhang. Reaction with this substrate results in the formation of an intermediate that contains a covalent linkage between the protein and DNA. In this

Figure 1.10: Stepwise mechanism of methyl transfer catalysed by DNA C5-Mtases, Chen *et al.*, (1993) in M.*Hha*I, Enz-COOH is Glu119, Enz-S is Cys 81, and Enz-B: is unknown Klimasauskas et al. (1994).



intermediate, the methyl group has been transferred to the C5 position, but elimination is inhibited because the fluorine on C5 cannot be released as F^+. During the course of this reaction, AdoMet is converted to AdoHcy, which is still present in the complex. From the structure, it was found that M.*Hha*I methyltransferase binds DNA as a monomer and the cleft provides the binding site for DNA (figure 1.11). The 20-residue active site loop (amino acids 80-99) is located in the large domain (see section 1.3.3 and figure 1.6). This loop contains six highly conserved residues, including the catalytic nucleophile Cys-81 and it undergoes a large conformational change as it moves towards the DNA-binding cleft (figure 1.12). The enzyme contacts the DNA minor groove through three residues (Ser-85, Lys-89, and Arg-97), which interact with the sugar-phosphate backbone. The sulfhydryl group of Cys-81 is brought into close proximately with the target cytosine, allowing the formation of a covalent link.

Interestingly, in the structure of the DNA they found that the enzyme bound G-C base pair that contains the target cytosine was completely disrupted, it was flipped completely out of the DNA helix. The exclusion of the cytosine from the helix is accompanied by substantial distortions of the phosphodiester backbone on the same strand. Both phosphates flanking the target cytosine are shifted significantly away from their corresponding positions in the unbound oligonucleotide, while further distortions extend along the phosphodiester backbone in the 5' direction. These changes increase the phosphate (P-P) distance between the two strands and allow the target cytosine to flip through the minor groove and out of the helix.

The target cytosine is held in place by a number of specific interactions with residues that are invariant in all members of the m5C-Mtases family. As shown in figure 1.13, the non-terminal nitrogen in the side chain of Arg-165 forms a direct hydrogen bond to the oxygen of the target cytosine ring at position 2 and its guanidine group makes a salt bridge with the phosphate of Cytosine-2 (GCG<u>C</u>). The side chain of Arg-165 clearly undergoes a conformational change in the presence of DNA, since in the binary complex of M.*Hha*I-AdoMet the side chain of Arg-165 interacts with Glu-119 through ion pairing. The side chain of Glu-119 form hydrogen bonds to N3 and N4 of the cytosine, and the main chain

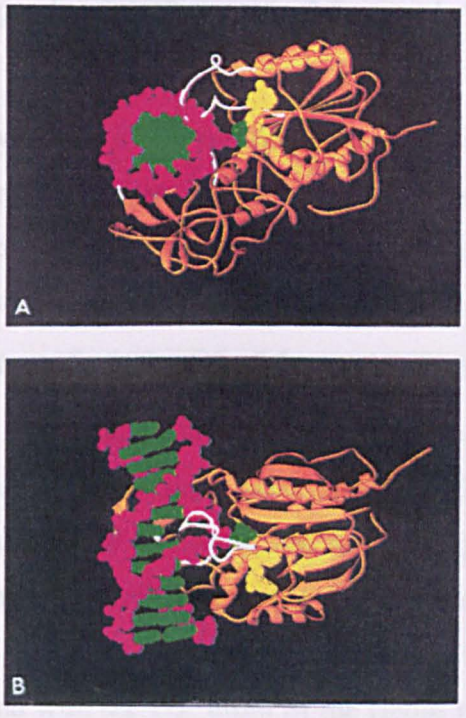


Figure 1.11: Graphic representation of the complex of M.*Hha*I covalently bound to a 13-mer DNA duplex containing its recognition sequence the end product of the reaction, AdoHcy, is also presented (yellow). The protein is in brown, the DNA backbone is in magenta and the DNA bases are in green, and the active site loop and the two recognition loops are in white. Klimasauskas *et al.,* (1994).

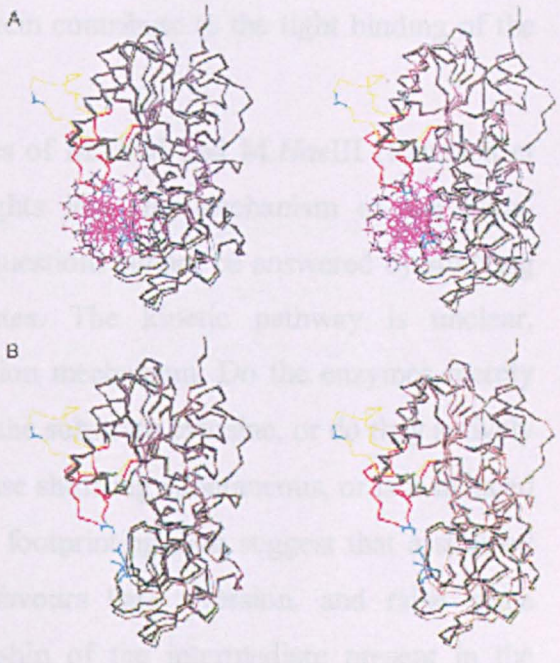


Figure 1.12: Stereo showing the superimposition of the ternary complex of DNA-M.*Hha*I.-AdoHcy and the binary structure of M.*Hha*I-AdoMet.
The Cαchains of M.*Hha*I from the two structures are shown as light (binary complex) or heavy (ternary complex). The 20-residue active-site loop in the large domain (wellow, binary complex; red, ternary complex) undergose a large conformational change upon DNA binding. The extreme movment is about 25 Å toward the cleft by Ser-87 (blue). The small domain is also shiffted toward the cleft with a movement of about 3 Å, shown by Gln-237 (blue). The catalytic nucleophile, Cys-81, in both complexes is in green. The root mean square (rms) deviation (except the 20-residue loop) between all α-carbons was 0.8 Å; between core atoms of the large domain (197 atoms) and between core atoms of the small domain (110 atoms) it was view with DNA. A) Stereo view with DNA. B) Stereo view without DNA Klimasauskas et al. (1994).

carbonyl oxygen of Phe-79 forms a hydrogen bond with the N4 of the extrahelical cytosine.

In the same experiment, AdoMet was used in the reaction mixture initially but AdoHcy was found in the cofactor-binding pocket. This is because under the reaction conditions methylation can proceed through the steps of covalent bond formation and methyl transfer. The sulphur atom, to which the methyl group is attached in AdoMet, is located about 5Å away from the C5 of the target cytosine. The bound cofactor has no direct contact with DNA. The binding pocket for the purine ring of AdoHcy is formed by the hydrophobic side chains of Trp-41, Phe-18, and Pro-80, and negatively charged side chains of Glu-40 and Asp-60 (figure 1.14). The aromatic ring of Phe-18 lies nearly perpendicular to the AdoMet purine ring, while the indole ring of Trp-41 lies parallel to the purine ring and the side chain nitrogen atom makes a hydrogen bond with the O2' atom of the ribose ring. Two charged residues, Glu-40 and Asp-60, interact directly with the O2' and O3' atoms of the ribose ring and the exocyclic amino group of the purine ring, respectively. These and the other extensive contacts, both direct and water-mediated, between AdoHcy and the protein contribute to the tight binding of the cofactor.

Although the co-crystal structures of M.*Hha*I and M.*Hae*III (Reinisch et al. 1995) have given important insights into the mechanism of the trans-methylation reaction, several important questions cannot be answered by studying the trapped transition-state intermediates. The kinetic pathway is unclear, especially with regard to the base eversion mechanism. Do the enzymes merely capture an extreme thermal excursion of the substrate cytosine, or do they actively participate in the eversion reaction? Is base shuffling spontaneous, or is it induced by interaction with the enzyme? Recent footprinting data suggest that a strained intermediate may exert a force that favours base eversion, and raise some interesting questions about the relationship of the intermediate present in the crystal to that in solution (Renbaum and Razin, 1995).

Figure 1.13: The active site of M.*Hha*I.

a) The flipped out target cytosine (Cyt2) is linked to several invariant residues. The nonterminal nitrogen (N_ε) of Arg-165, the main chain carbonyl oxygen of Phe-79, and the side chain carboxyl group ($O_{\varepsilon 1}$ and $O_{\varepsilon 2}$) of Glu-119 from direct hydrogen bonds with the O2, N4, and N4 atoms of the target cytosine ring, respectively.

b) Stereo view showing the sulfhydryl group of Cys-81 and AdoHcy are on the opposite side of the cytosine ring. One water molecule makes hydrogen bonds with $O_{\varepsilon 2}$ of Glu-119, the main chain oxygen of residue 304, and terminal nitrogen atom of the homocysteine moiety of AdoHcy.

Klimasauskas *et al.*, (1994).

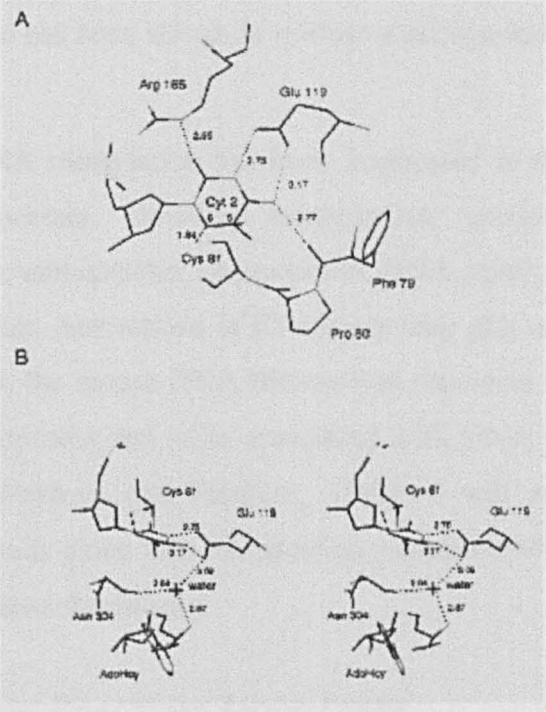
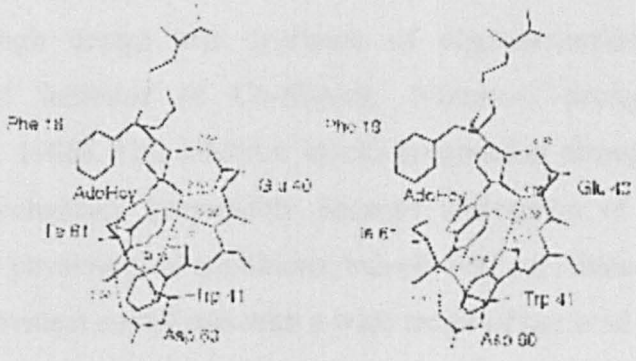


Figure 1.14: The cofactor binding site.

The N6 of the purine ring interacts with $O_{\varepsilon 1}$ of Asp-60, and N3 from hydrogen bonds with the main chain of residues 61 and 41,respectively. Two oxygen atoms (O2' and O3') of the ribose ring interact with the terminal carboxyl group of Glu-40. Note that the terminal amino and carboxyl groups of the homocysteine moiety have both direct and water-modiated contacts (w) to the main chain atoms of residues 18(w), 21, 23, 78(w), 301(w), as well as the side chains of Glu-119(w), Asn-304(w), and Ser-305. These additional interactions are omitted for clarity. Klimasauskas *et al.*, (1994).



1.3.4.2 DNA METHYLATION IN EUKARYOTES

There are just two types of Mtases in eukaryotes, C5-Mtases and N-4-Mtases. The latter is very few while the majority is the former. Eukaryotic DNA Mtases are very large, single subunit enzymes that preferentially methylate hemi-methylated DNA (Adams, 1990). Unlike the prokaryotic enzymes, the eukaryotic DNA Mtases must enter the nucleus to reach chromosomal DNA. The N-terminal part of the mouse enzyme has been shown to contain a nuclear localisation signal (Adams, 1995).

In eukaryotes DNA methylation has been implicated in the control of a number of cellular processes, including transcription, genomic imprinting, developmental regulation, mutagenesis, transposition, DNA repair, X inactivation, and chromatin organisation. Aberrations in C5-Mtases may play a role in human genetic disease. Recently, the mouse DNA Mtases was shown to be essential for normal embryonic development and to be associated with DNA replication foci (Adams, 1995; Noyer-Weidner and Trautner, 1993). I will not discuss the eukaryotic Mtases in details since this introduction concentrated in prokaryotic Mtase (see the above reviews for details).

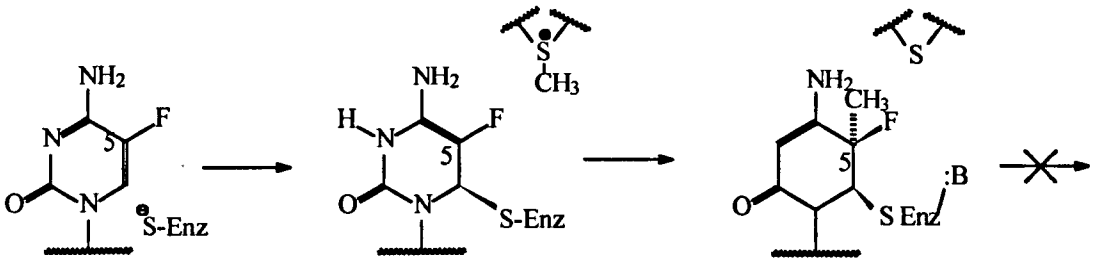
1.3.5 C5-MTASE REACTION INHIBITORS

1.3.5.1 5-FLOURO-2'-DEOXYCYTIDINE (FDC)

Structural studies on complexes of catalytic DNA-binding proteins are complicated by the fact that, under ordinary circumstances, the protein and DNA are associated tightly with one another only for the fleeting instant during which the covalent chemistry takes place. This problem has been overcome in the case of C5-Mtases through design and synthesis of oligonucleotides containing a mechanism-based inhibitor of C5-Mtases, 5-fluoro-2'-deoxycytidine (FdC) (Osterman et al. 1988). This inhibitor blocks progression through step 3 (figure 1.10) of the mechanism, presumably because abstraction of F^+ is virtually impossible under physiological conditions; indeed, FdC substituted DNA has been shown to form covalent complexes with a wide range of bacterial and mammalian m5C Mtases (Osterman et al. 1988; Friedman and Ansari, 1992; Smith et al. 1992). However, H and F bonded to C possess a similar van der Waals radius,

hence the trapped intermediate thus formed (figure 1.15) is expected to resemble closely the native intermediate formed with C.

Figure 1.15: Simplified mechanism of Mtase inactivation by FdC, which block progression through step 3 of the reaction mechanism.

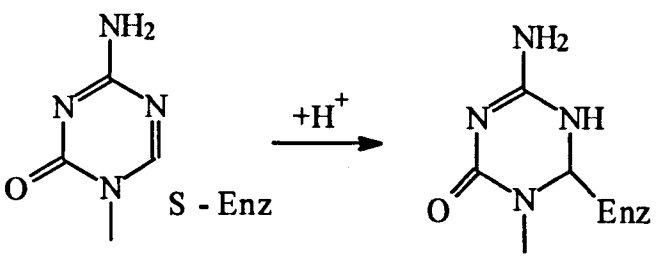


Trapped Covalent Intermediate

1.3.5.2 5-AZA-CYTIDINE (5-AZAC)

5-Azacytidine (5-azaC), an analogue of cytidine in which the C-H group at position 5 is replaced by a nitrogen, has a multitude of biological effects in a wide variety of organisms (Bhagwat and Roberts, 1987). The observation by Friedman (Friedman, 1981) that a number of bacterial DNA m5-C Mtases are inactivated by incubation with DNA containing 5-azaC provided the first clue to the mechanism by which 5-azaC causes hypomethylation. Santi et al. (1983) suggested that 5-azaC acts as a mechanism-based inhibitor of cytosine methylation by forming a covalent bond to the position 6 of cytosine as a normal intermediate in their action. In the normal reaction after the transfer of the methyl group from AdoMet to position 5 of cytosine, the enzyme is released. When 5-azaC incorporated into DNA or RNA at the site of methylation sabotages this reaction by making the transfer of the methyl group to position 5 impossible. Therefore, cytosine methylase should form stable covalent complexes with DNA containing 5-azaC (figure 1.16).

Figure 1.16: Proposed structure of the complex between the methylase and 5-azacytosine containing DNA (Wyszynski et al. 1991)



This prediction has been confirmed for the *Hpa*II, *Eco*RII, M.*Msp*I, and Dcm methylases (Bhagwat and Roberts, 1987). Formation of such complexes has been demonstrated *in vitro* and *in vivo* (Friedman and Som, 1993). The tight complexes formed between DNA C5-Mtases and 5-azaC-DNA, unless repaired, might interfere with cellular processes such as replication, transcription, and recombination, and thereby contribute to the lethality observed with the strains producing a DNA C5-Mtase. Inhibition of DNA strand exchange and blocking of transcription at the elongation step by complex formed with C5-Mtases and 5-azaC-DNA has been reported (Som and Friedman, 1994; Taylor et al. 1993).

1.4 DNA REPAIR

Since living cells require the correct functioning of thousands of proteins, each of which could be damaged by a mutation at many different sites in its gene, it is clear why DNA sequences are usually passed on unchanged if progeny are to have a good chance of survival. The challenge for the cell is twofold. First, the enzymatic machinery that replicates DNA must be inherently accurate. Second, the cell must repair accidental damage to DNA that would destroy its function. Since mutations result when damage changes the coding properties of DNA bases, an organism could not survive the natural rate of damage to its DNA without specific enzymatic mechanisms to repair damaged sites. In fact, DNA repair is so important that a bacterium may devote several percent of its genome to specifying and controlling the enzymes involved (Friedberg et al. 1995).

DNA repair has become a very wide subject and it is impossible to cover it fully. In this section I shall concentrate on DNA repair in prokaryotes in general and the genetics and enzymology of error-prone repair in some detail.

1.4.1 *DNA DAMAGE*

DNA is subject to damage in its chemistry or sequence. Many of these changes arise as a consequence of errors introduced during replication, recombination, and repair itself. Other forms of damage arise from the inherent

instability of specific chemical bonds that constitute the normal chemistry of nucleotides under physiological conditions of temperature and pH. Finally, the DNA of living cells reacts very easily with a variety of chemical compounds and a smaller number of physical agents, many of which are present in the environment. Some of these chemicals are products of the metabolism or decomposition of other living forms with which many organisms coexist. Others, particularly in recent decades, are man-made and contribute to the genetic insult faced by individuals living in highly industrialised communities.

From the above we can group DNA damage into two major classes **spontaneous** and **environmental** (Friedberg et al. 1995). Spontaneous damage of DNA includes mismatches (Loeb and Kunkel, 1981), spontaneous alteration in the chemistry of DNA bases, and loss of bases (depurination and depyrimidination) (Lindahl, 1993). Environmental damage to DNA is caused by physical agents such as UV radiation (Setlow, 1966) and ionising radiation (Goodhead, 1989) or chemical such as alkylating damage agents (Richardson and Richardson, 1990).

Living cells exhibit different responses to these diverse forms of damage. Table 1.3 summarises the spectrum of DNA damage response.

Table 1.3: Cellular responses to DNA damage (Friedberg et al. 1995).

Response	Mechanism
Reversal of DNA damage	Enzymatic photoreactivation
	Repair of O^6alkylguanine, O^4- alkylthymine, and alkylphosphotriesters
	Ligation of DNA strand breaks
Excision of DNA damage	Base excision repair
	Nucleotide excision repair
	Mismatch repair
Tolerance of DNA damage	Replicative bypass of template damage with gap formation and recombination
	Translesion DNA synthesis

1.4.2 REPAIR BY REVERSAL OF DNA DAMAGE

Photoreactivation (PR) of DNA: in this repair pathway pyrimidine dimers are the target of the ubiquitous enzyme photolyase, which binds as a dimer and catalyses a photochemical reaction, using visible light, that changes the cyclobutane ring induced by UV light back to individual pyrimidine bases (Friedberg et al. 1995).

Alkylation: this type of damage is removed by the cellular enzyme O^6-methylguanine methyltransferase, encoded by the gene *ada*, which recognises the O^6-methylguanine in the DNA duplex and removes the methylgroup by transfering it to a cysteine at the active site of the enzyme. The same enzyme also removes possibly disruptive methyl groups from phosphate of the DNA backbone (Teo et al. 1984). One of the features of this reaction is that there is no means of regenerating the unmethylated enzyme: a new enzyme molecule is spent for each methyl group removed (Lindahl, 1982).

Ligation of DNA strand breaks: damage is caused by agents that promote the hydrolysis of phosphodiesters in duplex DNA, like ionising radiation (such as X-ray and γ rays). The repair of these lesions requires various processing reactions that remove such damage before the ends can be rejoined. However, in *E.coli* some single stranded breaks in DNA produced by ionising radiation under anoxic conditions are repaired by simple rejoining of the ends (Jacobs et al. 1972), and such repair may be considered an example of the direct reversal of DNA damage directed by DNA ligase. DNA ligase is a highly specific enzyme which is ubiquitous in its distribution and plays a role in most known biochemical pathways which require the rejoining of strand breaks in DNA (Lehman, 1974).

The reversal of damage in DNA is obviously the most direct mode of DNA repair and suggests a number of distinct advantages to a living cell.

1- Only a single gene product is required which is a highly economical use of genetic information. However, the alkyltransferase mode of repair is clearly energetically expensive, since an entire protein molecule is expended for each reaction.

2- As a corollary of the previous statement, the reversal mode of DNA repair is kinetically advantageous since it presumably occurs more rapidly than multi-step biochemical pathways such as excision repair (see below).

3- In general, these processes would tend to be relatively error free because of their high degree of specificity.

1.4.3 EXCISION REPAIR

The double-stranded nature of DNA allows almost any damage to be repaired, no matter what the chemical change may be. Because a double helix holds two interconvertible copies of the genetic message, the loss of one is not serious: the information necessary to renew a damaged segment is present in the complementary strand.

Excision repair of DNA can be initiated by three distinct biochemical mechanisms. If the genome contains damaged, mispaired or, inappropriate bases (such as uracil), base excision repair will take place. Excision of DNA containing certain mispaired bases can occur by mismatch repair, and excision of DNA containing many different types of base damage can be effected by a more generic damage-specific endonulease (nucleotide excision repair) (Friedberg et al. 1995).

Base excision repair, involves a battery of enzymes called **DNA glycosylases**. Each DNA glycosylase recognises any altered base in DNA and catalyses its hydrolytic removal. There are at least six types of these enzymes, including those that remove deaminated Cs, deaminated As, different types of alkylated or oxidised bases, bases with opened rings, and bases in which a carbon-carbon double bond has been accidentally converted to carbon-carbon single bond (Friedberg et al. 1995). One of the general mechanisms that operates in all cases, the removal of a deaminated C by uracil DNA glycosylase is shown in figure 1.17 (Alberts et al. 1994). Base excision repair starts with a DNA glycosylase (such as uracil DNA glycosylase) which removes an accidentally deaminated cytosine in DNA. After the action of this glycosylase (or another DNA glycosylase that recognises a different kind of damage) the sugar phosphate with the missing base is excised by the sequential action of **AP** endonuclease and a phosphodiesterase. The single nucleotide gap is then filled by the subsequent action of a DNA polymerase and DNA ligase. The net result is that the U is restored to a C. The AP

endonuclease derives its name from the fact that it recognises any site in the DNA helix that contains a deoxyribose sugar with a missing base; such sites can arise either by the loss of a purine (**a**purinic sites) or by the loss of a pyrimidine (**a**pyrimidinic) (Alberts et al. 1994).

Many of the known forms of base damage, particularly those caused by the interaction of DNA with environmental agents, are not recognised by specific DNA glycosylases. There is yet another class of repair-specific enzymes, which generate incisions in DNA at or near sites of base damage and define a second mode of excision repair (Lindahl, 1982).

The **Nucleotide excision repair** pathway is capable of removing almost any type of DNA damage that creates a large change in the DNA double helix, such as the various pyrimidine dimers (T-T, T-C, and C-C) caused by UV light. In these cases a large multi-enzyme complex scans the DNA for a distortion in the double helix rather than for a specific base change. Once a bulky lesion is found, the phosphodiester backbone of the abnormal strand is cleaved on both sides of the distortion, and the portion of the strand containing the lesion (an oligonucleotide 12 bp in length) is "peeled away" from the DNA double helix by a **DNA helicase** enzyme. The gap produced in the DNA helix is then repaired in the usual manner by DNA polymerase and DNA ligase (Hoeijmakers, 1993).

There are more than 25 genes involved in the nucleotide excision repair which recognises and removes DNA damage, among these *uvrA, uvrB,* and *uvrC* which encode separate subunits of a single enzyme (UvrABC) (see figure 1.17) (Sancar and Rupp, 1983).

Some incorrectly paired bases escape even the proofreading activity of the bacterial DNA polymerase. In *E. coli,* the final degree of accuracy is the responsibility of a **Mismatch repair** system comprising mismatch correction enzymes. The enzyme scans newly replicated DNA for mismatched base pairs and removes a single-stranded segment containing the wrong nucleotide, thereby allowing a DNA polymerase to insert the correct base when it fills the resulting gap (Wagner and Meselson, 1976). The obvious problem that this entails is that of distinguishing which base of a mismatched pair is wrong, because both are natural components of DNA. A specific signal controlled by a timing device directs the

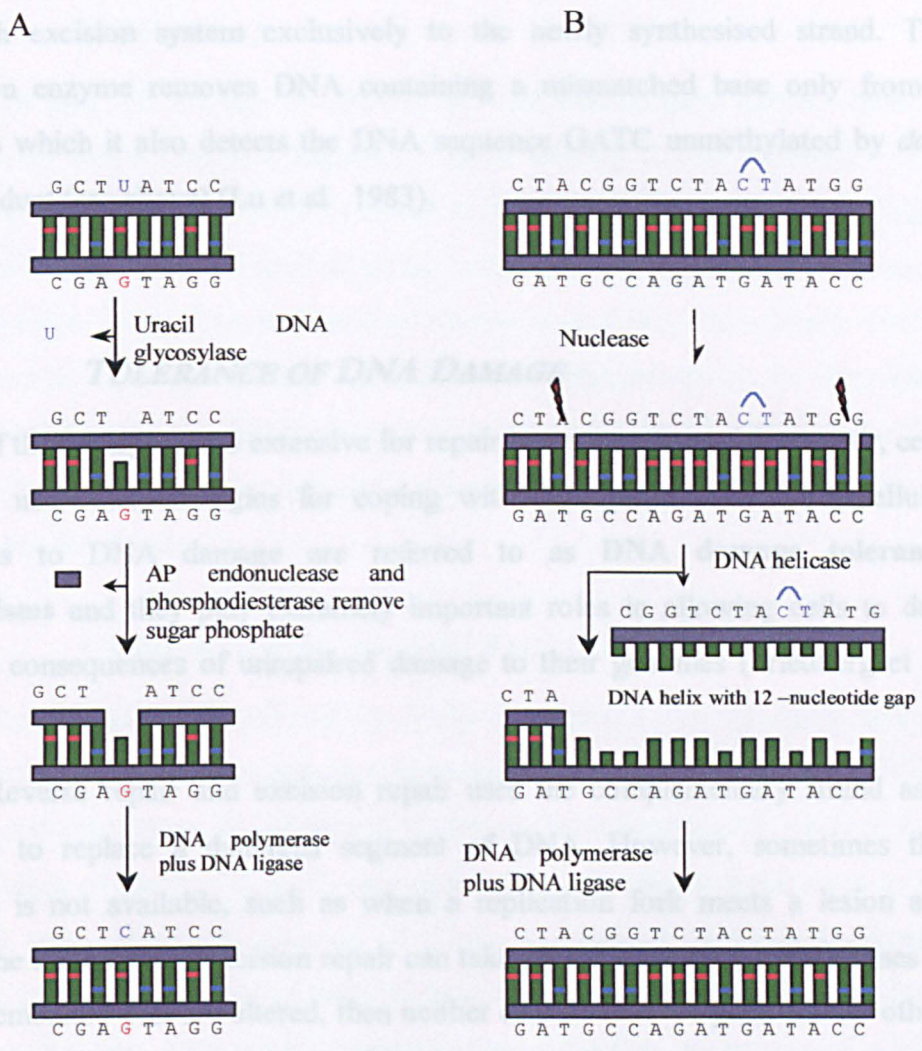


Figure 1.17: model shows two types of excision repair, a) base excision repair and b) nucleotide excision repair.

mismatch excision system exclusively to the newly synthesised strand. The correction enzyme removes DNA containing a mismatched base only from a strand in which it also detects the DNA sequence GATC unmethylated by *dam* gene product (see above) (Lu et al. 1983).

1.4.4 TOLERANCE OF DNA DAMAGE

If the damage is too extensive for repair by reverse or excision repair, cells still can use other strategies for coping with damaged DNA. These cellular responses to DNA damage are referred to as **DNA damage tolerance mechanisms** and they play extremely important roles in allowing cells to deal with the consequences of unrepaired damage to their genomes (Friedberg et al. 1995).

Reverse repair and excision repair uses the complementary strand as a template to replace a damaged segment of DNA. However, sometimes this template is not available, such as when a replication fork meets a lesion and affects the DNA before excision repair can take place. Moreover, if both bases of a complementary pair are altered, then neither can act as a template for the other. Finally, a double helix that is broken straight across, particularly if a duplex segment is missing altogether, cannot be repaired directly with much chance of the connection being correctly remade. In all these cases, all information is lost at the site of damage, and it can be recovered only by taking a corresponding DNA segment from a separate but identical DNA molecule: this is called **recombination repair.** In *E.coli* the RecA protein is essential for carrying out this function (Szostak et al. 1983). In addition, cells have evolved another class of mechanism for processing damaged DNA which, although not yet fully understood at a biochemical level, appears to involve the polymerisation of DNA past a lesion and is often referred to as **translesion DNA synthesis.** Finally, when the damage that is severe enough to stop DNA synthesis such as a protein bound covalently to the DNA, a series of genes may be induced to repair the damage. These genes encode a pathway known as the **SOS repair** pathway (Friedberg et al. 1995). The SOS repair pathway will be discussed in the next section.

1.4.5 SOS REPAIR

The idea that damage to DNA or that the physiological consequences of such damage might initiate a regulatory signal causing the simultaneous *recA*$^+$ *lexA*$^+$ -regulated derepression of a number of genes in *E.coli*, the product of some or all of which enhance the survival of the cell, was first suggested by Radman in 1973 (Radman, 1974). The international distress signal -Save Our Souls- (SOS), the term he appropriated to describe this phenomenon, is frequently misinterpreted to imply a last ditch attempt by the cell to survive the lethal effect of DNA damage after other cellular responses have failed. SOS regulation is discussed in detail below.

1.4.5.1 CURRENT MODEL FOR SOS REGULATION

In an uninduced *E.coli* cell (i.e. in the absence of DNA damaging agents), the product of the *lexA* gene acts as a repressor for more than 20 genes, including the *recA* and *lexA* genes, by binding to similar operator sequences at each gene or operon. Many of these SOS genes are expressed at significant levels even in the repressed state. When the genome of *E.coli* is damaged or DNA replication inhibited, an intracellular signal for SOS induction is generated. There is evidence to suggest that this intracellular induction signal consists of regions of single-stranded DNA that are generated when the cell attempts to replicate a damaged template or when its normal process of DNA replication is interrupted (Friedberg et al. 1995). The binding of RecA protein to these regions of single-stranded DNA in the presence of nucleoside triphosphates reversibly converts it to an active form, often refered to as RecA*. The LexA protein then diffuses to the activated RecA protein, and interacts with this nucleoprotein complex in a way that results in LexA becoming proteolytically cleaved at a specific Ala-Gly bond near the middle of the protein. This proteolytic cleavage of LexA protein results from the ability of activated RecA to facilitate an otherwise latent capacity of LexA protein to autodigest (figure 1.18).

RecA-mediated cleavage of LexA protein inactivates LexA as a repressor. Thus, as LexA cleavage proceeds after an SOS-inducing treatment, the pools of LexA protein begin to decrease so that various SOS genes, including the *recA*$^+$ gene, are expressed at increased levels and SOS responses mediated by these

genes can be observed. Genes with operators that bind LexA relatively weakly are the first to be turned on fully. If the inducing treatment is sufficiently strong, more molecules of RecA are activated and more molecules of LexA are cleaved. As the pools of LexA decline to very low levels, even genes whose operators bind LexA very tightly are expressed at maximal levels.

As the cell begins to recover from the SOS-inducing treatment, the regions of single-stranded DNA disappear as a consequence of various DNA repair processes and RecA molecules return to their nonactivated state. Continued synthesis of LexA protein then leads to an increase in LexA pools, which, in turn, leads to repression of SOS genes and a return to the uninduced state (Friedberg et al. 1995). Figure 1.18 shows schematically the basic regulatory mechanism of the SOS system at a molecular level.

The proteolytic cleavage of LexA when it interacts with activated RecA is central to SOS regulation. At first, it was assumed that the RecA protein was acting as classical protease. However, experiments by Little and his colleagues have led to the view that activated RecA protein acts indirectly by stimulating an otherwise latent capacity of LexA to autodigest (Little, 1984). Working as **co-protease** is one of the activities of the multifunctional RecA.

1.4.5.2 INDUCTION OF THE SOS RESPONSE

The *in vitro* requirements for RecA protein mediated cleavage of LexA are well established: RecA protein, ATP, and single-stranded DNA (Sassanfar and Roberts, 1990). However, it has been considerably more challenging to determine the nature of the *in vivo* inducing signal that leads to RecA activation and subsequent LexA cleavage. Recent work has supported the unifying view, that at least for many inducing treatments, the ultimate signal for SOS induction *in vivo* is the generation of regions of single-stranded DNA within the cell (Masek et al. 1995; Friedberg et al. 1995; Drlica et al. 1980). Most SOS-inducing agents do not directly cause breaks in DNA but, rather, create lesions that alter the chemical structure of the bases thereby interfering with base pairing.

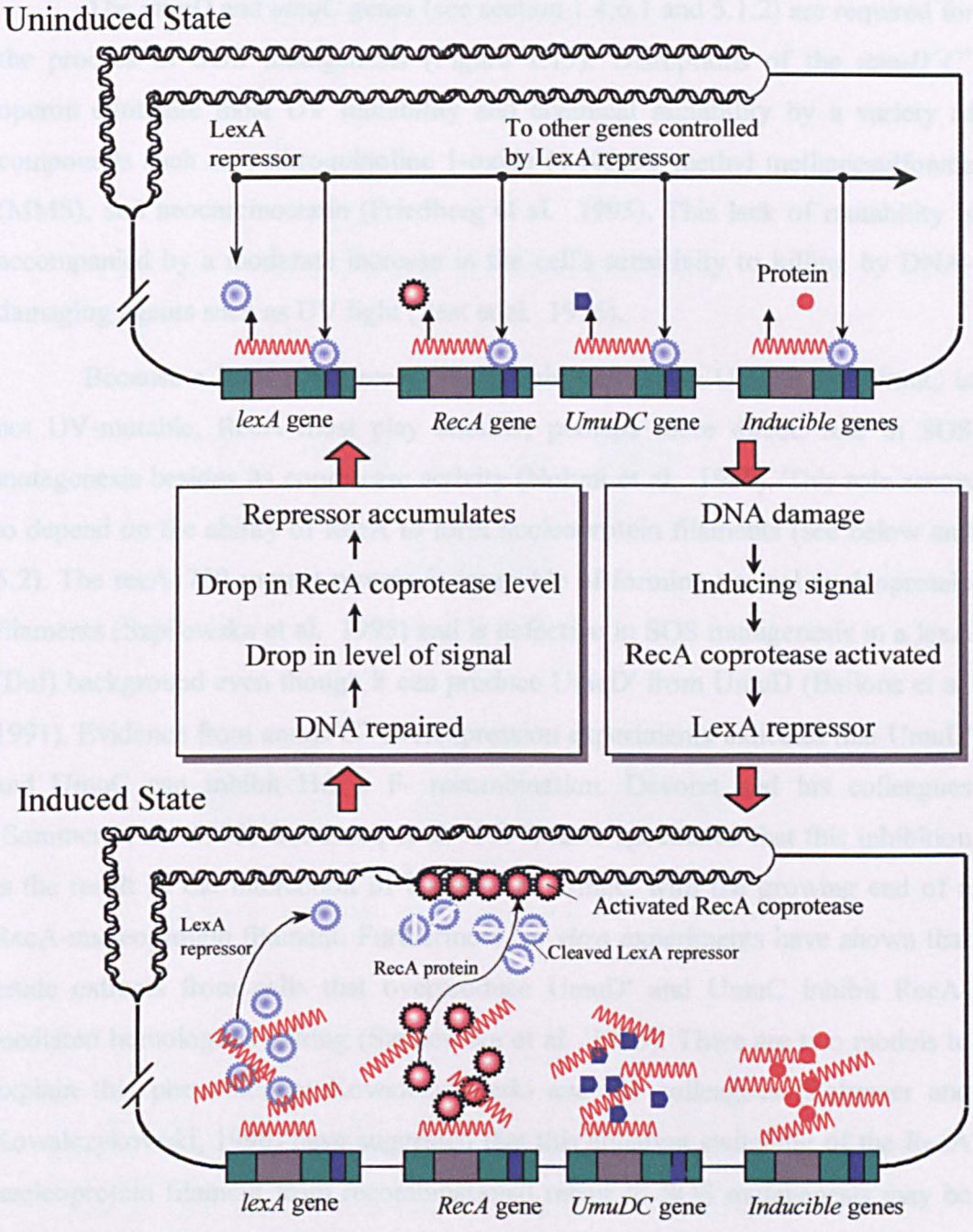


Figure 1.18: Model for SOS regulation.
Diagrammatic representation of the mechanism by which the *lexA-recA* regulon is regulated. In the uninduced state (top), LexA repressor protein constitutively expressed in small amounts is bound to the *lexA* operator and to the operators of the *recA* gene and other genes under LexA control. These genes are still able to express small amounts of the protein they encode; thus, there is some RecA protein constitutively present in uninduced cells. Following DNA damage (e.g., the presence of a pyrimidine dimer near the replication fork after induction by UV radiation), the coprotease activity of existing RecA protein is activated, probably by binding to the single stranded DNA in the gaps created by discontinuous DNA synthesis past the dimers. The interaction between LexA and activated RecA results in the proteolytic cleavage of LexA. In the induced state (bottom), derepression of the *recA*$^+$ gene results in the production of large amounts of RecA protein. Other genes under LexA control are also derepressed, although not necessarily with identical kinetics. When the induction signal disappears (probably by repair the single-strand gap), the level of active coprotease drops, LexA repressor accumulates, and genes under LexA control are once again repressed. Reprinted from (Friedberg *et al.*, 1995 with modification)

1.4.6 SOS MUTAGENESIS

The *umuD* and *umuC* genes (see section 1.4.6.1 and 5.1.2) are required for the process of SOS mutagenesis (Figure 1.19). Disruptions of the $umuD^{+}C^{+}$ operon eliminate most UV mutability and chemical mutability by a variety of compounds such as 4-nitroquinoline 1-oxide (4-NQO), methyl methanesulfonate (MMS), and neocarcinostatin (Friedberg et al. 1995). This lack of mutability is accompanied by a moderate increase in the cell's sensitivity to killing by DNA-damaging agents such as UV light (Peat et al. 1996).

Because a lexA (Def) recA (Def) strain expressing UmuD' and UmuC is not UV-mutable, RecA must play another, perhaps more direct, role in SOS mutagenesis besides its coprotease activity (Nohmi et al. 1988). This role seems to depend on the ability of RecA to form nucleoprotein filaments (see below and 5.2). The recA1730 mutant protein is incapable of forming normal nucleoprotein filaments (Szpilewska et al. 1995) and is defective in SOS mutagenesis in a lexA (Def) background even though it can produce UmuD' from UmuD (Bailone et al. 1991). Evidence from $umuD'C^{+}$ overexpression experiments indicates that UmuD' and UmuC can inhibit Hfr x F- recombination. Devoret and his colleagues (Sommer et al. 1993; Boudsocq et al. 1997) have speculated that this inhibition is the result of the interaction of UmuD' and UmuC with the growing end of a RecA-nucleoprotein filament. Furthermore, *in vitro* experiments have shown that crude extracts from cells that overproduce UmuD' and UmuC inhibit RecA-mediated homologous pairing (Szpilewska et al. 1995). There are two models to explain this phenomenon: Kowalczykowski and his colleagues (Rehrauer and Kowalczykowski, 1996) have suggested that this apparent switching of the RecA nucleoprotein filament from recombinational repair to SOS mutagenesis may be due to the aforementioned competition between coprotease substrate, i.e., UmuD, and the secondary dsDNA molecule for the same binding site on the RecA nucleoprotein filament. The second model of RecA's role in SOS mutagenesis involving a targeting mechanism was proposed by Woodgate and his group (Frank et al. 1993) in which the RecA-nucleoprotein filament serves to target the Umu proteins to damaged ssDNA. In support of both models, affinity chromatography has been used to demonstrate an interaction of UmuD, UmuD', and UmuC with RecA. In light of these hypotheses, it is interesting to note that a

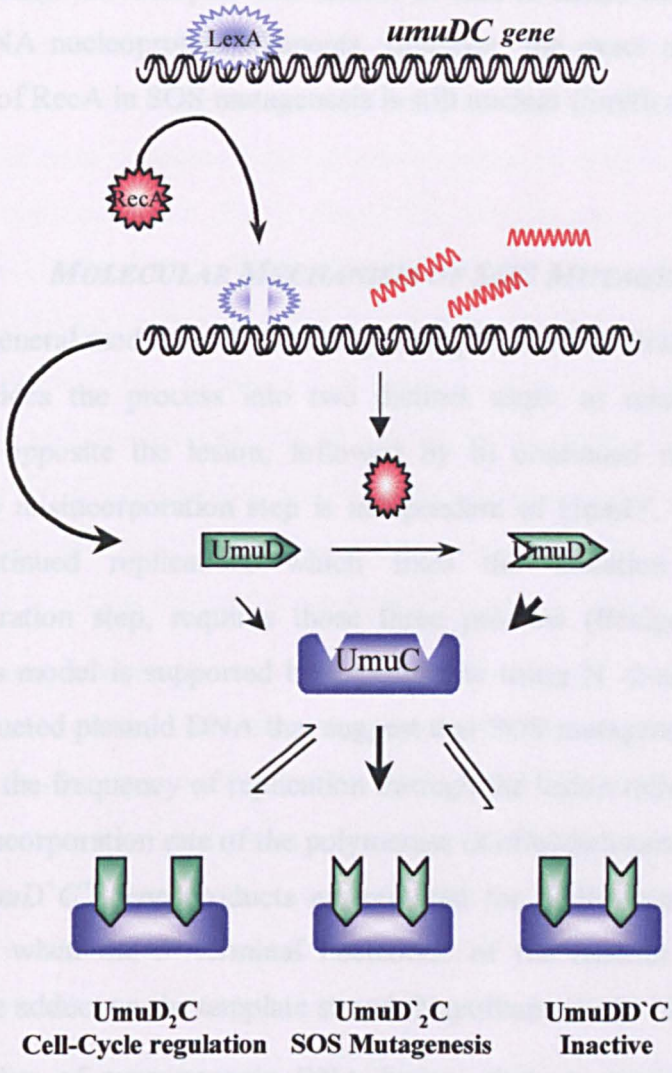


Figure 1. 19: Regulation of the umuDC operon by RecA and LexA.
DNA damage generates a signal that converts RecA to RecA*. RecA* mediates the cleavage of the LexA repressor that results in the induction of the umuDC operon as well as the rest of the SOS response genes. RecA* can also mediate the processing of UmuD to the shortened UmuD' molecule. UmuD and UmuD' can interact with UmuC in a variety of combinations. The Umu(D)2C complex seems to be involved in regulating the E. coli cell cycle after DNA damage. The Umu(D')2C complex is active in SOS mutagenesis (translesion synthesis). The third complex, UmuDD'C, does not appear to have an activity, but it may play a role in shutting off SOS mutagenesis by sequestering UmuD'. Reprinted from (Smith and Walker, 1998)

purified Umu(D')2C complex was shown to bind to naked ssDNA, as well as to RecA-ssDNA nucleoprotein filaments. However, the exact nature of this more direct role of RecA in SOS mutagenesis is still unclear (Smith and Walker, 1998).

1.4.6.1 MOLECULAR MECHANISM OF SOS MUTAGENESIS

A general model for SOS mutagenesis proposed by Bridges and Woodgate (1985) divides the process into two distinct steps: a) misincorporation of a nucleotide opposite the lesion, followed by b) continued replication past the lesion. The misincorporation step is independent of UmuD', UmuC, and RecA, while continued replication, which fixes the mutation induced in the misincorporation step, requires those three proteins (Bridges and Woodgate, 1985). This model is supported by experiments using N -2-acetylaminofluorene (AAF)-adducted plasmid DNA that suggest that SOS mutagenesis results from an increase in the frequency of replication through the lesion rather than an increase in the misincorporation rate of the polymerase (Koffel-Schwartz et al. 1996), and that the *umuD*$^{+}$*C*$^{+}$ gene products are required for replication of AAF-adducted DNA only when the 3' terminal nucleotide of the nascent strand is directly opposite the adduct on the template strand (Napolitano et al. 1997).

Studies of premutagenic DNA lesions that are processed by the SOS mutagenesis system have indicated the following: (1) that lesions such as abasic sites, UV-induced thymine-thymine cyclobutane dimers, and pyrimidine-pyrimidone (6-4) photoproducts present virtually complete blocks to DNA replication; (2) that replication through these lesions and the mutation frequencies at these sites are greatly enhanced in SOS-induced cells; and (3) that the translesion synthesis that occurs in SOS-induced cells processes some premutageneic lesions differently from the basal level that is present in uninduced cells, providing further evidence that a modified DNA polymerase is functioning in SOS-induced cells (Smith and Walker, 1998).

The specific molecular details of *umuD*$^{+}$*C*$^{+}$-mediated translesion synthesis, the underlying cause of SOS mutagenesis, are still not understood. A number of models have been proposed for the molecular mechanism of this process, including a suppression of a DNA polymerase's proofreading ability, an

increase in its processivity, or a relaxation of its requirement for a proper Watson-Crick DNA structure for continued replication (Friedberg et al. 1995). The identity of the DNA polymerase (DNA Pol) that participates in translesion synthesis has been examined, and it has been shown that two of *E. coli*'s three DNA polymerases, DNA Pol I and DNA Pol II are not required for SOS mutagenesis (Iwasaki et al. 1990). DNA Pol III, the major replicative polymerase in *E. coli*, has been implicated in SOS mutagenesis (Hagensee et al. 1987). Both *in vivo* and in vitro, translesion synthesis requires UmuC and UmuD' or homologous mutagenesis proteins as well as RecA, even when a DNA Pol III lacking, the 3'→5' exonuclease proofreading subunit, is used (Rajagopalan et al. 1992). Therefore, it appears that the suppression of the proofreading function of DNA Pol III is not the mechanism by which translesion synthesis occurs.

Overexpression of the $umuD^{+}C^{+}$ operon from a multi-copy plasmid in lexA(Def) cells, which constitutively express the SOS response genes, causes a cold-sensitive growth phenotype at 30°. This $umuD^{+}C^{+}$-mediated cold sensitivity for growth is accompanied by a rapid, reversible, UmuDC-dependent inhibition of DNA synthesis at the restrictive temperature that is dependent on the cell being in lag phase (Smith and Walker, 1998). In addition, overexpression of either ε or β, the dimeric processivity element that encircles the template DNA and tethers the core of DNA Pol III to the DNA, partially suppresses SOS mutagenesis, perhaps by titrating out the Umu proteins or by competing with the Umu proteins for a binding site on the DNA Pol III holoenzyme. These data suggest that an interaction becomes between the Umu proteins and the *E. coli* DNA replication machinery (Smith and Walker, 1998).

Echols and his colleagues (Rajagopalan et al. 1992) were able to reconstitute translesion synthesis *in vitro* using purified UmuC, UmuD', RecA, DNA Pol III, single-stranded DNA binding protein (SSB), and a linear ssDNA template with a synthetic abasic lesion engineered at a defined location. In these studies, ~5% of the primers were extended through the abasic site. In the absence of RecA, UmuC, or UmuD', the background level of translesion synthesis was ~0.5%. These values are similar to those reported in studies of translesion synthesis *in vivo* with either an abasic site-containing M13mp7-based vector or an AAF-adducted plasmid (Koffel-Schwartz et al. 1996), which indicated a 0.1–

0.7% or a 0.42% translesion synthesis frequency in the absence of SOS induction and a 5–7% or a 12.4% frequency after the SOS response is induced, respectively. An alternative *in vitro* system for dissecting translesion synthesis using crude bacterial cell extracts of varying genotypes has shown a requirement of *recA*$^+$ and *umuC*$^+$ for UV mutagenesis (Cohen-Fix and Livneh, 1994). These *in vitro* systems will be useful in dissecting the complex set of molecular interactions between the Umu proteins, RecA, DNA Pol III, and the damaged DNA template that are required for translesion synthesis. Understanding the mechanisms by which a DNA polymerase can replicate through regions of damaged DNA will undoubtedly teach us much about the manner by which DNA Pol III monitors itself and the template DNA strand during normal DNA replication.

1.5 AIMS

To determine how a mutant form of C5-Mtases in which the essential Cys (motif IV) is substituted by Gly induces error-prone deletions. This will include:

1- Definition of the DNA damage (lesion)

2- Establishing a rapid method for analysing the deletion events in populations of colonies.

3- Establishing the generality of the observation with several Mtases (using more than one type of mono-specific and a multi-specific DNA Mtase).

4- Establishing whether deletion occurs with monomeric and dimeric form of Mtases.

5- Identifying any genes (and proteins) that contribute to deletion and whether this deletion process is dependent on the SOS repair pathway.

CHAPTER TWO:
MATERIALS AND METHODS

SYNOPSIS OF CHAPTER TWO

- List of suppliers of general laboratory reagents and equipment used in this study.
- List and composition of the common buffers used in this study.
- List of genotype for strains used in this study
- General laboratory methods used in this study.

2.1 MATERIALS

2.1.1 CHEMICALS, ENZYMES AND GENERAL REAGENTS

All materials used in the experiments were high quality chemicals and enzymes. Restriction endonuclease used were from New England Biolabs (NEB; Beverly, MA, USA) or MBI Fermentas (Vilnius, Lithuania). Reagents were optained from Sigma (Poole, Dorest, UK), Fisher (Loughborough, Leics.,UK) and BDH (Lutterworth, Leics., UK). *E.coli* culture media were supplied by Oxoid and Difco Laboratories, TNT® T7 Quick Coupled Transcription/Translation from Promega. Most of the general materials used in the experiments are listed below:

Ampicillin	Sigma
Bacteriophage CE6	Novagen
Calf intestine alkaline phospatase (CIP)	Boehringer Mannheim
Chicken egg white lysozyme	Sigma
Chloramphenicol	Sigma
Glutathione agarose beads	Sigma
Kanamycin	Sigma
pUC-4K plasmid	Pharmacia Biotech
Ribonuclease A	Worthington Biochemicals
Sephadex G-25 Superfine	Pharmacia
Taq Dyedeoxy™ Teminator cycle sequencing kit	Applied Biosystems Inc.
Vent$_R$®DNA polymerase	N.E.B.

2.1.2 EQUIPMENT

Equipment used to carry out this work were as follows:

Equipment	Manufacturer
Pipettes	(P20, P200 and P1000 µL) Gilson
Balances	Mettler 0-60g model AJ 100 and Nettler 0-800g model K7
Beckman DU® 640	Beckman
BIAcore Chip	Pharmacia

BIAcore2000	Pharmacia
Camera	Polaroid
Centrifuges	Beckman J2, Beckman TL-100 ultracentrifuge, M.S.E. Centaur 2, and M.S.E. Microcentaur
Gel Drier	Model 483 slab drier BIO-RAD
Heating blocks:	Dri-Block DB1 Techne
Horizontal slab gel	mini- and wide mini-subcell BIO-RAD
Incubators	Model S.I. 60, Stuart Scientific.
PC Scanner	Hewlett-Packard Flatbed Scanjet-4C
PH meter	Denver Basic with Tris Electrode Denver Instruments
Power supplies	Model 200/20 BIO-RAD
Spectrophotometer	LKB-Biochrom Ultraspec 4050.
tThermal cycler	GeneE Techne (Programmable)
Transilluminator	Genetic Research Instruments Ltd.
Vacuum Dryers	Hetovac VR-I rotary evaporator, Heto-Intermed.
Vertical slab gel	Model SE600 Hoeffer Scientific Ltd.
Vertical slab gel	Mini Protean II BIO-RAD
Waterbaths	Grant Instruments, Cambridge.

2.1.3 STANDARD BUFFERS AND SOLUTIONS

Buffer	Composition
1 X SDS Sample Buffer	500 mM Tris-HCl (pH 6.8), 10% (v/v) glycerol, 10% (w/v) SDS, 5% (v/v) 2-mercaptoethanol and 0.005% (w/v) bromophenol blue.
1 X STE	10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0) and 100 mM NaCl.
1 X TAE	40 mM Tris-acetate and 1 mM EDTA (pH 8.0).
1 X TBE	100 mM Tris, 80 mM Boric acid and 2.5 mM EDTA (pH 8.3).
1 X TE	10 mM Tris-HCl (pH 8.0) and 1 mM EDTA (pH 8.0)
6 X AGE Loading Buffer	0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF and 40% (w/v) sucrose in MQ water.
LA medium (1L) Luria-Bertani agar	10 g tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar
LB medium (1L) Luria-Bertani	10 g tryptone, 5 g yeast extract and 10 g NaCl
PBS Phosphate Buffered Saline	137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4H_2O$ pH 7.4
RF1	100 mM RbCl, 50 mM $MnCl_2.4H_2O$, 30 mM K acetate (pH 7.5), 10 mM $CaCl_2.H_2O$ and 15% Glycerol (w/v)

	Final pH 5.8 using acetic acid
RF2	10 mM RbCl, 75 mM $CaCl_2.2H_2O$, 10 mM MOPS and 15% Glycerol (w/v) final pH 6.8 using NaOH
TBS pH 7.4 Tris Buffer Saline	20 mM Tris pH 7.4 and 500 mM NaCl.
Tris-Acetate-EDTA (TAE) AGE Buffer pH 8.0	Prepared as a 50X stock concentrate: 242g Tris, 57.1 ml glacial acetic acid and 100 ml 500 mM EDTA pH 8.0

2.1.4 STRAINS AND PLASMIDS

Strains used in the experiments and their genotypes.

Strain	*Genotype*[1]	Source
HB101	*thi*-1, *hsdS*20 (r_B^-,m_B^-), *supE*44, *recA*13, *ara*-14, *leuB*6, *proA*2, *lacY*1, *rpsL*20 (str^r) *xyl*-5 *mtl*-1	NBL Gene Science
RRI	isogenic with HB101 except for $recA^+$	NBL Gene Science
AB1157	*ara*-1 *thr*-1 *leuB*6 *lacY*1 *argE*3 *Δ(gpt proA)*62 *mtl*-1 *xyl*-5 *rpsL*31 *tsx*-33 *supE*44 *galk*2 *glnV*44 *hisG*4(Oc) *rfbD*1 *kdgk*51 *rfb*-1 *mgl*-51 *thi*-1	G. Walker MIT USA
$AB1157^{RecA}$	AB1157 *recA*13 (known as JC2926)	G. Walker and A.J. Clark
$AB1157^{UmuDC}$	AB1157 *UmuDC::cat*(595) (known as GW8017)	G. Walker
GW1000	AB1157 *recA*441 *sulA* ilv^{ts} Δ*lacU*169 pro^+	G. Walker
$GW1000^{RecA}$	GW1000 *recA::Tet*	G. Walker

The Multispecific Mtase M.SPRI gene was a gift from Prof. Trautner T. A. (Berlin, Germany).

2.1.4.1 TESTING RECA⁻ STRAINS BY UV SENSITIVITY.

RecA⁻ strain was tested by making a horizontal strip of cells across an LA plate, $RecA^+$ cells were used as a control. Part of the plate was covered with a piece of cardboard, the rest of it was exposed to UV light for different periods of time. $RecA^-$ cells show increased sensitivity to UV light: those in the unshielded part of the plate should be killed by certain level of irradiation (see figure 2.1).

[1] Major genetic markers are detailed in appendix two at the end of this thesis.

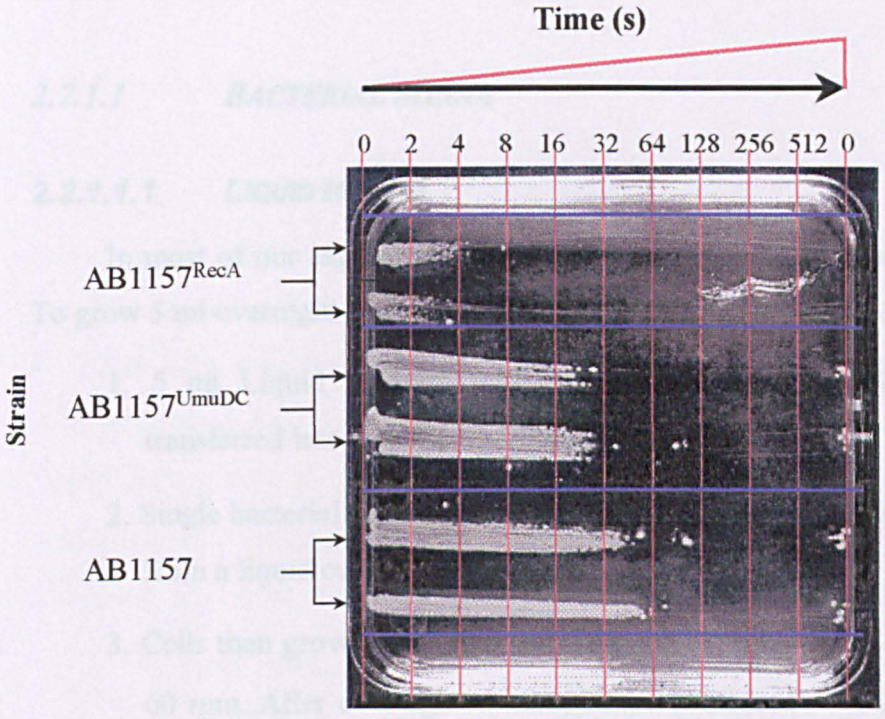


Figure 2.1: Testing the RecA⁻ genotype.
Using a toothpick, a horizontal stripe of cells was spread across an LA plate.The plate was covered with cardbord, and irradiated with 300 ergs/cm^2 of 254 nm UV light. First two lanes are two colonies from AB1157^{RecA}, second two lanes AB1157^{UmuDC}, and last two lanes are AB1157 wild type. Plates were incubated at 37°C O/N. RecA⁻ cells are killed within the first 4 seconds, UmuDC⁻ cells killed within the 30 seconds while the wild type cells were killed after 60 seconds.

2.2 METHODS

2.2.1 GROWTH AND STORAGE OF BACTERIAL STRAINS AND BACTERIOPHAGE

E. coli were grow in solid media to isolate single colonies or in liquid media for suspension culture. To grow either liquid or solid cultures fresh cells from a -70°C stock were used.

2.2.1.1 BACTERIAL MEDIA

2.2.1.1.1 LIQUID MEDIUM

In most of our experiments ***LB medium*** was used (sambrook et al. 1989). To grow 5 ml overnight culture:

1. 5 ml Liquid medium with appropriate antibiotic (if required) was transferred into a sterile 15-20 ml tube.
2. Single bacterial colony inoculated from a fresh agar plates or 10µl of cells from a liquid culture.
3. Cells then grown at 37°C to saturation in a shaker or on a roller drum at 60 rpm. After overnight culture, strains produced ~4 x 10^9 bacterial /ml depending upon the medium, degree of aeration, the strain, and the temperature.

For large volume cultures, an overnight culture was diluted 1:100 in an Erlenmeyer or baffle flask (≥5 times the volume of the culture) and was grown at 37°C overnight with vigorous agitation at 300 rpm.

2.2.1.1.2 SOLID MEDIUM

Solid media was used for obtaining pure cultures, by streaking colonies with an inoculating loop onto an agar plate with appropriate antibiotic (if required). This plate could be stored for several weeks at 4°C after sealing the edge with Parafilm.

2.2.1.2 BACTERIAL STOCKS

Liquid cultures of bacteria grown in LB medium were used for daily experiments, the liquid cultures were stored at 4°C and used within 3 days. For longer storage agar plates were used and were stored at 4°C for 4-6 weeks. To store bacteria for years or to keep certain strains, bacteria in the stationary phase (liquid culture at 37°C for 6-8 hr) were stored in Eppendorf tubes at -70°C in solutions of 15-50 % glycerol.

2.2.2 DNA PURIFICATION

2.2.2.1 PLASMID DNA PURIFICATION

Small scale plasmid isolation (referred to as the 'miniprep' method) from *E.coli* (5-30 ml culture) was performed by the alkaline lysis method (Sambrook et al. 1989) with the following modification. Following the addition of an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1), the plasmid mixture was mixed very gently in a shaker for 20 minutes. For high purity DNA used in nucleotide sequencing experiments and *in vitro* transcription/translation, Wizard® 'miniprep' was used according to the manufactures instructions

For large scale plasmid DNA purification, cells harbouring the plasmid were grown in 50 ml of LB with the appropriate antibiotic overnight, and plasmid DNA was then purified using QIAGEN® columns according to the manufacturers instructions with the following minor modification: after propan-2-ol precipitation and 70% ethanol wash, samples were recentrifuged for 10 miniutes and then dried under vacuum and redissolved in an appropriate volume of 1X TE buffer or MQ water.

2.2.2.2 DNA FRAGMENT PURIFICATION

DNA fragments were purified from agarose gel slices or from solution using a **Jetsorb** DNA extraction from GENOMED or a Gene Clean kit (BIO-101 Inc.). The manufacturer's protocols were followed in both cases. DNA was eluted by incubation in 1X TE buffer or MQ water.

2.2.2.3 DEPROTEINATION OF DNA SAMPLES

Proteins were removed from DNA samples by phenol/chloroform extraction. Phenol, pre-equilibrated at pH 8.0 in 100 mM Tris-HCl pH 8.0, was mixed with chloroform and isoamyl alcohol (25:24:1 respectively) as described by Sambrook et al. (1989). The mixture was stored at 4°C in a light-tight bottle. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was mixed with the DNA solution by vortexing and was then centrifuged in a micro-centaur bench-top centrifuge at 13,000 rpm for 5 minutes at room temperature. The upper aqueous layer was carefully decanted into a fresh microfuge tube so as not to disturb the aqueous-organic layer interface where the protein was partitioned. Residual traces of phenol/chloroform/isoamyl alcohol were removed by ether extraction if required. An equal volume of ether was mixed with the DNA solution and centrifuged in a micro-centaur bench-top centrifuge at 13,000 rpm for 1 minute at room temperature. The upper ether layer was discarded and traces of ether were removed by blowing air on to the surface of the solution with a Pasteur pipette.

2.2.2.4 ETHANOL PRECIPITATION OF DNA SAMPLES

DNA was precipitated by the addition of 1/10 volume of a potassium acetate solution pH 4.8 (3M with respect to potassium and 5M with respect to acetate) and 3 volumes of -20°C absolute ethanol. Plasmid DNA isolated by the miniprep method contained previously added potassium acetate solution so only ethanol addition was required. The solution was vortexed, incubated at -20°C for 20-30 minutes and then spun in a micro-centaur bench-top centrifuge at 13,000 rpm for 15 minutes at 4°C. the pellet was washed in 70% (v/v) ethanol and spun again for 5 minutes. The pellet was dried under vacuum (Hetovac VR-I rotary evaporator) for 10 minutes and then dissolved in an appropriate volume of MQ water (usually 50 µl for small scale, miniprep-purified plasmid DNA and 200 µl for large scale).

2.2.3 DNA QUANTIFICATION

2.2.3.1 ABSORBANCE QUANTIFICATION

DNA samples were quantified using the absorbance of the sample at 260 nm. An OD_{260} reading of 1 corresponding to a concentration of 50 μg/ml for double stranded DNA, and 40μg/ml for single stranded DNA (Ausubel, 1992). The ratio of OD_{260} : OD_{280} indicates the level of protein or phenol contamination - pure preparations of DNA have OD_{260}:OD_{280} of 1.8- a ratio significantly less than this value indicates the presence of either contaminant.

The concentration of oligodeoxynucleotides was determined using the equation:

Concentration (mM) $= A_{260} / E$

where: A_{260} = absorbance at 260 nm

E = the sum of all the mM extinction coefficients of the nucleotides present: nεdATP, nεdTTP, nεdCTP, and nεdGTP, where n = number of bases present

The path length was 1 cm in all cases and the mM extinction coefficients for the nucleotides are εdATP = 15.2, εdTTP = 8.4, εdCTP = 7.05, and εdGTP = 12.01. The absorbance at 260 nm was taken using a LKB spectrophotometer with 1000 fold dilutions of oligodeoxynucleotide samples in 1ml quartz cuvettes.

2.2.3.2 FLUORESCENCE QUANTIFICATION

For quick and roughly ds DNA quantification fluorescence comparison was used by running appropriate amount of DNA in 1.0% agarose gel along with a 0.5μg ladder DNA marker (of known concentration) followed by using ethidium bromide staining (see figure 2.2).

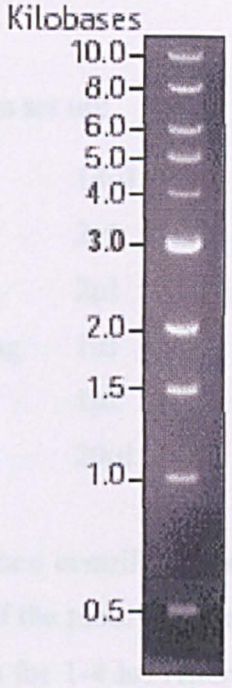
2.2.4 RECOMBINANT DNA METHODS

2.2.4.1 DNA DIGESTION

The DNA was digested with appropriate restriction enzymes to generate compatible ends for cloning or to ensure that the correct fragment had been

A) 1 Kb Ladder Marked

Fragment	Base Pairs	DNA amount (ng)
1	10,002	41
2	8,001	41
3	6,001	50
4	5,001	41
5	4,001	32
6	3,001	124
7	2,000	63
8	1,500	46
9	1,000	31
10	500	31



B) 100 bps Ladder Marker

Fragment	bp	DNA amount (ng)
1	1,500	46
2	1,200	37
3	1,000	91
4	900	27
5	800	24
6	700	21
7	600	18
8	500	91
9	400	37
10	300	27
11	200	24
12	100	55

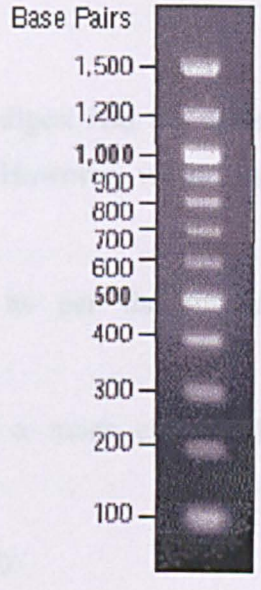


Figure 2.2: The 1Kb and 100 bp DNA ladder can be used for approximating the mass of DNA in comparably intense samples of similar size. The approximate mass of DNA in each of the bands in the ladders is as illustrated in the tabels opposit to the agarose photographs in a and b (assuming a 0.5 ug loading). A) 1% agarose gel for the 1Kb ladder marker and b) 100 bp ladder marker both markers were run for 1 hour at 100 volts.

inserted to a given vector. Plasmid DNA, compared with DNA fragments often required a higher quantity of restriction enzyme in order to complete digestion.

2.2.4.1.1 SINGLE DIGESTION

For single digestions the following reaction was set up:

MQ water	14µl
Restriction enzyme 10X buffer	2µl
BSA 1 mg/ml (if required)	2µl
DNA sample (in water or TE buffer) 0.2-1.0 µg	1µl
Restriction enzyme, 2-10U	1µl
Total volume	20µl

The reaction was mixed gently by pipetting then centrifuged briefly in a microcentrifuge to collect the contents at the bottom of the tube. The reaction was incubated at the optimum temperature (usually 37°C) for 1-4 hr. Afterwards 1µl of blue dye loading buffer was added to the reaction prior to electrophoresis.

2.2.4.1.2 MULTIPLE DIGESTION

If all the restriction enzymes in a multiple digest had the same optimal buffer, setting up the reaction was straightforward. However, when this was not the case, one of three options was followed:

1- A compatible buffer was chosen as per the manufacturer's instructions.

2- An isoschizomer was chosen with a more compatible buffer requirement.

3- Each digest was performed sequentially.

2.2.4.2 *ANALYSIS OF RESTRICTED PLASMID DNA BY AGAROSE GEL ELECTROPHORESIS*

Restriction plasmid DNA fragments were separated on 1% agarose gels. These gels were prepared as follows: 0.5g agarose was added to 50 ml 1X TAE buffer. The agarose was dissolved by heating the solution in a microwave on high

power (~2 min). When all the agarose had dissolved, the liquid gel was cooled to around 55°C and then poured into a pre-prepared gel former and left to set at room temperature.

The gels were run for one hour at 100V (~0.15mA). After this time the gels were stained in ethidium bromide (10mg/ml stock) for 15 minutes at room temperature and the separated DNA fragments were visualised using UV light (wavelength 310 nm). To obtain a permanent record of gels, photographs were taken using a Polaroid DS-34 direct screen instant camera with an orange filter. Photographs were then scanned into a PC using a flatbed scanner (Hewlett-Packard Flatbed Scanjet - 4C) and stored on disc. Depending on the expected fragment size, 1kb or 100bp ladder marker used (see figure 2.2).

2.2.4.3 LIGATION

After the vector and insert DNA had been prepared for ligation, various vector:insert DNA ratios were tested in order to find the optimum ratio for a particular vector and insert. In most cases a 1:3 molar ratio worked well. The following equation was used to calculate the DNA concentration used for each ligation mixture:

$$\frac{\text{ng of vector X kb size of insert}}{\text{Kb size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}} = \text{ng of insert}$$

Then using the appropriate vector : insert ratios samples were set up with 1U T4 DNA ligase, 2µl ligase 10X buffer and MQ to final volume 20µl. For blunt ends ligation the reaction was incubated at 16°C overnight while 1-3 hr at room temperature was considered sufficient cohesive ends. Following the ligation reaction, the ligated DNA transformed then into competent cells of an appropriate host cells.

2.2.4.4 TRANSFORMATION

Transformation of *E.coli* cells with DNA plasmid was performed by the method using rubidium chloride adapted from Hanahan (1985). This method gave transformation efficiencies of up to 2 x 10^9 cfu/µg plasmid DNA (Hanahan, 1985).

2.2.4.4.1 PREPARATION OF COMPETENT CELLS

1. Cells were streaked from a frozen stock and grown on LA plates O/N then a single colony was inoculated into 3 ml LB with a suitable antibiotic at 37°C for about 16-20 h.
2. 0.5 ml of the growing cells were inoculated into 50 ml (1%) in an Erlenmeyer flask containing LB and the same antibiotic. The cells were then incubated at 37°C with moderate agitation until the cell density was 4-7 x 10^7 viable cells/ml.
3. The culture was collected into 50 ml centrifuge tubes and chilled on ice for 10 - 15 min.
4. The pellet was collected by centrifugation at 2000 - 3000 rpm for 12 -15 min at 4°C.
5. The pellet was resuspended by moderate vortexing in 17 ml of RF1 (1/3 of the volume collected), and was incubated on ice for 15 min.
6. The pellet was collected again, as in step 4, and resuspended in 4 ml RF2 (1/12.5 of the original volume), and incubated on ice for 15 min.
7. The cells were distributed in 0.2 ml aliquots into chilled 1.5 ml micro-centrifuge tubes, and immediately frozen using liquid nitrogen, and stored at -70°C.

2.2.4.4.2 USE OF FROZEN COMPETENT CELLS

1. Tubes were removed from the freezer (-70°C) and thawed at room temperature until the cell suspension was just liquid, and was then placed on ice.
2. The DNA solution (in a volume of <20 µl) at a concentration of ≈ 50 ng was added to the competent cells, the DNA was evenly mixed with the cells and were incubated on ice for 30 min.
3. Cells were heat shocked at 42°C for 90 seconds, and were then chilled immediately by returning them to ice.
4. 0.8 ml (4X the original volume of competent cells) of LB was added and the cells were incubated at 37°C with moderate agitation for 45 min.

5. 200 µl competent cells were then spread on to preheated LA plates with the appropriate antibioticat 37°C for 18 h.

2.2.5 POLYMERASE CHAIN REACTION (PCR)

The polymerase chain reaction (PCR) (Ausubel, 1992) was used to amplify ds DNA. Vent$_R$® DNA polymerase was used to catalyse the reaction with the buffer (10X PCR buffer) provided by the supplier. Ultra-pure dNTPs were stored as a mixture of all four at a concentration of 5mM at -20°C and were used at a final concentration of 0.2mM. Primers were stored at -20°C as 5µM stocks and were used at a concentration of 0.2µM. A DNA template concentration of 1ng/µl was generally employed. The reaction was set up as follow:

1- The reagents were mixed at the concentration described in a thin-walled 0.5ml microcentrifuge tube prior to addition of the polymerase in a final volume of 49µl.

2- - The reaction was heated at 95°C for 5 minutes and was centrifuged briefly to collect condensed water, and was then chilled on ice for 5 minutes.

3- One µl (2U) of the polymerase enzyme was added and the contents were mixed gently.

4- The thermal cycler programme used consisted of a denaturation step at 94°C for 45 seconds, a high thermal ramp (40°C/minute) to 46-50°C (depending on the primer length), 1 minute annealing, a rapid thermal ramp to 72°C, extension at 72°C, for 1 minute and then a rapid thermal ramp to 94°C. The reaction was allowed to proceed for 30 cycles.

5- Amplified DNA was purified by agarose gel electrophoresis and Gene Clean (see sections 2.2.5.2 and 2.2.2.2).

2.2.5.1 SITE DIRECTED MUTAGENESIS

Site directed mutagenesis (Ho et al. 1989) was employed to mutate the Cys codon to a Gly codon in motif IV of Mtases. Such procedures require one mutagenic primer and another universal one, which contain suitable restriction site to facilitate subsequent cloning steps. Essentially, this method makes use of

two steps, a PCR step using a mutagenic and universal primer. In the second stage, the resultant PCR product is treated with an appropriate pair of restriction enzymes and the fragment ligated into a suitable vector.

2.2.5.2 COLONY PCR

Single colonies were picked from an overnight agar plate using a 200µl pipette tip or a sterile toothpick. Colonies that were at least 1mm in diameter were chosen and as many cells as possible were taken. Bacteria were then transferred to a 1.5ml tube containing 50µl of MQ water and vortexed to disperse the pellet. Tubes were placed in boiling water for ten minutes to lyse the cells and denature any endogenous Dnase, followed by centrifugation at 12000g for 1 min to remove cell debris. After cooling to room temperature, 35µl of cell extracts were then added to the premixed primers, selected polymerase buffer, dNTPs, and the final volume was made up to 50µl with MQ water.

2.2.6 DNA SYNTHESIS

Oligodeoxynucleotides were used as primers for the amplification of DNA by PCR, sequencing, or in protein-DNA binding assays, and were synthesised by the Biomolecular Synthesis Service of the Krebs Institute, University of Sheffield on an Applied biosystems DNA/RNA synthesiser, model 381A, by standard phosphoramidite method (Brown and Brown, 1991). The 3'[(2-cyanoethyl)-(N,N-diisopropyl)] (CE) –phosphoramidities: 5'-dimethoxytrityl-2'-deoxyadenosine (dA), 5'-dimethoxytrityl-2'-deoxycytidine (dC), 5'-dimethoxytrityl-2'-deoxyguanosine (dG), and 5'-dimethoxytrityl-2'-deoxythymidine (dT) were all purchased from Glen Research.

Oligodeoxynucleotides, provided in solution after deprotection overnight in 30% ammonia were precipitated using ethanol and potassium acetate solution (see section 2.2.2.4) and were resuspended in an appropriate volume of MQ water. All the oligodeoxynucleotides used in this work are listed in appendix 3.

2.2.7 DNA SEQUENCING

DNA sequencing was performed on an Applied Biosystems Model 373A DNA sequencing system by the Biomolecular Synthesis Service off the Krebs

Institute, University of Sheffield. Template DNA was isolated from *E.coli* and purified using a Wizard® column as described in section 2.2.2.1 . Cycle sequencing reactions were performed using the *Taq* Dyedeoxy Terminator Cycle Sequencing kit as described in the manufacturer's protocols using 1 μg template DNA and 3.2 pmol of oligodeoxynucleotide primer. A modification was made to the recommended thermal cycling programme:

*Rapid thermal ramp to 96°C at 96°C for 30 seconds

*Rapid thermal ramp to 50°C at 50°C for 30 seconds

(That recommended by the manufacturers was 15 seconds)

*Rapid thermal ramp to 60°C at 60°C for 4 minutes

25 cycles

Extension products were concentrated using sodium acetate and ethanol precipitation as described in the manufacturer protocols. Samples were dried under vacum and were provided to the Biomolecular Synthesis Service at Krebs Institute University of Sheffield. Nucleotide sequencing data were provided as a computer file in floppy disc.

2.2.8 PROTEIN EXPRESSION

2.2.8.1 BACTERIOPHAGE CE6 INDUCTION OF T7 REGULATED GENES

Bacteriophage CE6 is a recombinant lambda phage containing the T7 polymerase gene. It used to provide a source of T7 RNA polymerase to the cells carrying pET recombinant plasmids (Studier and Moffatt, 1986).

The following protocol has been used for expression of target genes cloned in pET22b expression vectors.

1- Cells harbouring a Mtase gene cloned into pET22b, were grown in LB supplemented with 0.2% maltose and the appropriate antibiotic.

2- When the OD_{600} was between 0.6 and 1.0, $MgSO_4$ was added to a final concentration of 10 mM and λCE6 stock to final concentration of 2-4 X 10^9 pfu/ml.

3- The infected cells were left to grow for 3 hr

2.2.8.2 IN VITRO EXPRESSION OF PROTEINS

Transcription and translation *in vitro* has been designed for the efficient synthesis of protein from a variety of DNA templates containing T7 promoters upstream from coding sequences, including supercoiled plasmids, linearised plasmids, and PCR products. The system is based on transcription with T7 RNA polymerase followed by translation in an optimised rabbit reticulocyte lysate (Perara and Lingappa, 1985).

In addition to circular plasmid DNA, linear DNA templates, such as those generated by PCR or by restriction enzyme digestion also can be transcribed/translated using the T7 System. For maximal expression from such linear templates, it is recommended that approximately 20bp is present upstream of the T7 RNA Polymerase promoter for efficient promoter binding. It is also important to add a stop codon (usually UAA) to truncated gene products in order to prevent ribosomes from stalling at the ends of linear DNA. Smaller amounts of PCR DNA templates were used in these experiments compared with plasmid DNA (e.g., for a 1 kb PCR product, 0.1- 0.2 µg for each 50µl TNT® Quick reaction). In some cases the PCR-generated DNA was titrated to determine the optimal level for both expression and fidelity.

To reduce the risk of RNase contamination, gloves were worn during the experiment, and microcentrifuge tubes and pipette tips were RNase-free. The following steps were followed:

1. The TNT® Quick Master Mix reagents was removed from storage at -70°C then rapidly thawed by hand warming and placed on ice.
2. Following the example below the reaction components were assembled in a 0.5 ml micro-centrifuge tube.

Component	Standard Reaction
Quick Master Mix	40µl
Methionine, 1mM	1µl
DNA template(s)	2µl (0.5µg/µl)
Nuclease-Free Water	to a final volume of 50µl

3. The reaction was incubated at 30°C for 60-90 minutes.

4. The results and proteins were analysed by biomolecular interaction analysis (BIA).

2.2.9 MTASES ACTIVITY ASSAY

When a plasmid carries a DNA Mtase gene, that plasmid is fully methylated at the sites on the plasmid that are recognised by the enzyme. This phenomenon facilitates the analysis of active and inactive mutant enzymes e.g. pUC18 carrying the M.*Msp*I gene has approximately 20 sites. All sites are protected from R.*Msp*I digestion when the plasmid is recovered from the cells by conventional "mini-prep" analysis. If the active site cysteine is replaced e.g. by alanine, the mutant plasmid is no longer resistant to R.*Msp*I cleavage.

CHAPTER THREE:

THE EFFECT OF MOTIF IV CYS TO GLY MUTATION IN MONOSPECIFIC MTASES ON E.COLI

SYNOPSIS OF CHAPTER THREE

- A brief overview of the motif IV mutation: PC to PG.
- Details of the construction and features of the vectors used in this study.
- Constructing M.*MspI*$_{C174G}$ gene expressed in dimeric and monomeric forms.
- Testing the generality of the effect a monospecific Gly mutants on *E.coli*.
- Identifying Cis and Trans deletions.
- Nucleotide sequencing of deleted plasmids.
- Discussion of chapter three results.

3.1 INTRODUCTION

Several studies have shown that the replacement of the conserved cysteine in motif IV with serine, valine, or tryptophan for example abolished catalysis in C5 Mtases (Gabbara et al. 1995). It was first observed for the M.*SPR*I multispecific Mtase (see section 4.1) that replacement of the cysteine residue in the motif IV (PC) with serine destroyed catalytic activity (Wilke et al. 1988). A later more extensive study of the *Eco*RII Mtase (M.*Eco*RII recognition sequence CCWGG), showed that replacement of the conserved cysteine with serine, valine or tryptophan abolished catalysis (Wyszynski et al. 1991). Interestingly, glycine substitution of the catalytic thiol both abolished activity and proved cytotoxic to *E. coli* (Wyszynski et al. 1991). Mi and Roberts (1993) found that M.*Hha*I exhibits similar properties when the conserved Cys81 in the PC motif is altered. Mutants in which Cys81 was replaced with Arginine, Histidine or Serine lost methyltransferase activity. In contrast, the Gly81 mutant lost activity and was cytotoxic.

Other studies have shown that cytosine Mtases interact with DNA containing 5-azacytidine (5-azaC), an analogue of cytidine in which carbon-5 is replaced by a nitrogen. This analogue acts as a specific inhibitor of C-5Mtases through the formation of irreversible covalent complexes with the enzyme (Bhagwat and Roberts, 1987).

Interestingly when the cysteine is replaced by glycine in motif IV, the mutant enzyme becomes cytotoxic and plasmids are difficult to recover. However if the mutant plasmid (containing C81G) is used to transform a $RecA^+$ host, the plasmids are recovered but they contain deletions between *Msp*I recognition sites which fall within the M.*Msp*I gene (Hurd and Hornby unpublished data). A similar effect would be anticipated when cells, which contain wild type M.*MspI*, are grown in the presence of 5-azaC. This analogue promotes the formation of covalent protein:DNA complexes (*Msp*I : CCGG complex) at a level which is dependent on the extent of 5-azaC incorporation into newly replicated DNA (Som and Friedman, 1994; Taylor et al. 1993).

3.2 MUTAGENESIS OF THE CATALYTIC CYS OF M.MSPI AND ITS CONSEQUENCES FOR E.COLI SURVIVAL

3.2.1 MSPI C-5 DNA MTASE

M.*Msp*I C-5 Mtase is a type II, monomeric enzyme that catalyses the transfer of a methyl group from AdoMet to the C-5 position of the outer 5` deoxycytosine base within the tetra-nucleotide sequence (5`-**CCGG**-3`) of DNA (Lin et al. 1989). R.*Msp*I was originally isolated from an organism present as a contaminant in a culture, it was characterised as *Moraxella* species, and that formed the basis for the name *Msp*I. However, subsequent tests of this organism suggest that the original identification was incorrect and it has since been variously characterised as an *Acinetobacter* species, or *Flavobacterium* species. For historical reason, it is still referred to as *Moraxella* species (Lin et al. 1989).

The gene encoding the *Msp*I restriction and modification system was first cloned by Walder et al. (1983) from a genomic library of *Moraxella* DNA. Recombinants in pBR322, carrying the gene for the modification enzyme were enriched from a random library followed by restriction with *Msp*I. M.*Msp*I was then sub-cloned into ps3 (a derivative of pUC19) by Dubey et al. (1992) to yield M.*Msp*I which overexpresses at least 10% of total cellular protein (Dubey et al. 1992). However, the construct used in our experiments was the GST-M.*Msp*I fusion described by Taylor et al. (1993) which expressed M.*Msp*I at high levels following induction with IPTG (Taylor et al. 1993). The purified fusion enzyme behaved in a manner identical to that of the WT enzyme (Dubey et al. 1992) and could be over-expressed at a high levels in *E.coli* grown at 30°C and readily purified in one step by glutathione affinity chromatography (Taylor et al. 1993).

M.*Msp*I comprises 418 amino acid (Lin et al. 1989) (see Figure 3.1) and the primary structure of the protein shares the same organisation as the 45 or so other bacterial C5-Mtases whose primary structure are known (Kumar et al. 1994; Lauster et al. 1989; Posfai et al. 1989; Wilson and Murray, 1991b) (see section 1.3.2 and figure 1.4 and 1.5). All related Mtases consist of ten conserved motifs, of which six are highly conserved (illustrated in figure 3.1)

Figure 3.1: DNA sequence and deduced amino acids sequence of M.*Msp*I$_{cass}$ gene

The red arrows indicate start and stop codons. The invariably conserved amino acids present in the six most highly conserved motifs are coloured the same as in figure (1.4), the highly conserved amino acids among those motifs are in bold (see figure 1.5). Numbering is on the left, black for DNA sequence and red between brackets for amino acids. The targeted C174 in motif IV is in bold and underlined. The unique restriction sites used in the experiments *Bsp*EI (TCCGGA), *Nco*I (CCATGG), and *Xba*I (TCTAGA) are underlined.

```
   1  → ATG AAA CCT GAA ATA TTG AAA TTG ATT CGT AGT AAA TTG GAT CTA
(1)      M   K   P   E   I   L   K   L   I   R   S   K   L   D   L
  46  ACT CAA AAG CAA GCA TCT GAA ATT ATT GAA GTT AGT GAT AAA ACA TGG
(16)   T   Q   K   Q   A   S   E   I   I   E   V   S   D   K   T   W
  94  CAA CAG TGG GAA TCA GGT AAA ACA GAA ATG CAT CCT GCT TAT TAC TCT
(32)   Q   Q   W   E   S   G   K   T   E   M   H   P   A   Y   Y   S
 142  TTT CTC CAA GAA AAA CTT AAA GAT AAG ATT AAC TTT GAG GAA CTG AGC
(48)   F   L   Q   E   K   L   K   D   K   I   N   F   E   E   L   S
 190  GCT CAA AAA ACA CTC CAA AAG AAG ATA TTC GAT AAA TAC AAT CAG AAT
(64)   A   Q   K   T   L   Q   K   K   I   F   D   K   Y   N   Q   N
 238  CAA ATT ACT AAA AAT GCG GAA GAA TTA GCT GAA ATT ACT CAT ATT GAA
(80)   Q   I   T   K   N   A   E   E   L   A   E   I   T   H   I   E
 286  GAA AGA AAA GAT GCT TAT AGC AGT GAT TTC AAA TTT ATT GAT TTA TTT
(96)   E   R   K   D   A   Y   S   S   D   F   K   F   I   D   L   F
 334  TCT GGT ATC GGT GGA ATA AGG CAA TCA TTC GAA GTG AAT GGC GGA AAA
(112)  S   G   I   G   G   I   R   Q   S   F   E   V   N   G   G   K
 383  TGC GTC TTT TCG TCT GAA ATA GAT CCC TTT GCA AAA TTT ACA TAT TAT
(128)  C   V   F   S   S   E   I   D   P   F   A   K   F   T   Y   Y
 430  ACA AAT TTT GGT GTA GTC CCT TTT GGA GAT ATT ACA AAA GTT GAG GCT
(144)  T   N   F   G   V   V   P   F   G   D   I   T   K   V   E   A
 478  ACA ACT ATT CCG GAG CAT GAC ATT TTA TGT GCA GGA TTT CCG TGT CAG
(160)  T   T   I   P   E   H   D   I   L   C   A   G   F   P   C   Q
 526  CCA TGG AGC CAT ATG GGT AAA AGA GAA GGC TTT GAA CAT CCA ACT CAA
(176)  P   W   S   H   M   G   K   R   E   G   F   E   H   P   T   Q
 574  GGA ACA ATG TTC CAT GAA ATT GTT CGT ATC ATT GAA ACA AAA AAA ACC
(192)  G   T   M   F   H   E   I   V   R   I   I   E   T   K   K   T
 622  CCA GTT TTA TTT CTA GAA AAT GTT CCT GGT CTC ATT AAT CAT GAT GAC
(208)  P   V   L   F   L   E   N   V   P   G   L   I   N   H   D   D
 670  GGA AAT ACA TTA AAA GTC ATC ATT GAA ACA CTA GAA GAT ATG GGC TAC
(224)  G   N   T   L   K   V   I   I   E   T   L   E   D   M   G   Y
 718  AAA GTT CAT CAT ACG GTA CTA GAC GCT AGT CAT TTT GGT ATC CCA CAA
(240)  K   V   H   H   T   V   L   D   A   S   H   F   G   I   P   Q
 766  AAA CGT AAG CGT TTT TAC TTA GTA GCT TTC CTA AAT CAA AAT ATT CAC
(256)  K   R   K   R   F   Y   L   V   A   F   L   N   Q   N   I   H
 814  TTT GAG TTT CCT AAA CCT CCA ATG ATT TCT AAA GAT ATC GGT GAA GTT
(272)  F   E   F   P   K   P   P   M   I   S   K   D   I   G   E   V
 862  TTA GAA AGC GAT GTA ACT GGA TAT AGC ATT TCA GAG CAT TTA CAA AAA
(288)  L   E   S   D   V   T   G   Y   S   I   S   E   H   L   Q   K
 910  AGT TAT CTC TTT AAA AAA GAT GAT GGT AAA CCT TCT TTA ATT GAC AAA
(304)  S   Y   L   F   K   K   D   D   G   K   P   S   L   I   D   K
 958  AAT ACG ACT GGG GCA GTT AAA ACA TTA GTT TCT ACC TAT CAC AAA ATT
(320)  N   T   T   G   A   V   K   T   L   V   S   T   Y   H   K   I
1006  CAA CGA TTA ACA GGT ACT TTT GTT AAG GAT GGA GAA ACA GCT ATC CGT
(336)  Q   R   L   T   G   T   F   V   K   D   G   E   T   A   I   R
1054  CTT TTA ACA ACG AAT GAA TGC AAA GCT ATT ATG GGT TTT CCA AAA GAT
(352)  L   L   T   T   N   E   C   K   A   I   M   G   F   P   K   D
1102  TTT GTT ATT CCT GTA TCA AGA ACT CAG ATG TAC CGT CAA ATG GGT AAC
(368)  F   V   I   P   V   S   R   T   Q   M   Y   R   Q   M   G   N
1150  TCT GTC GTA GTT CCG GTG GTT ACA AAA ATT GCA GAA CAG ATT AGT TTG
(384)  S   V   V   V   P   V   V   T   K   I   A   E   Q   I   S   L
1198  GCA CTA AAA ACT GTT AAC CAA CAA TCC CCG CAA GAA AAC TTT GAA TTA
(400)  A   L   K   T   V   N   Q   Q   S   P   Q   E   N   F   E   L

1246  GAA CTC GTT TAA ←
(416)  E   L   V   STOP
```

In this study M.*Msp*I will be mutated in the region encoding the highly conserved amino acid sequence in motif IV Cys (C174) to Glycine (this mutation will be referred to as (C174G) to study the effect on the *E.coli* cells survival.

3.2.2 THE USE OF PGEX2T FOR MUTANT CONSTRUCTION

The pGEX series of vectors (Smith and Johnson, 1988; Guan and Dixon, 1991) express an active 26kD glutathione-S-transferase (GST) from *Schistosoma japonicum* under the control of an IPTG inducible *tac* promotor (de Boer et al. 1983). This promoter, derived from the *trp* and the *lac*UV5 promoters, is repressed in the absence of IPTG due to the plasmid encoded *lac* repressor. At the 3' end of the GST gene a polylinker region containing unique restriction sites is followed by translational stop codons in all three reading frames. Foreign DNA sequences ligated into these unique sites can be expressed to high levels in *E.coli* as fusion products with GST. GST fusion proteins can be purified by glutathione affinity chromatography (Smith and Johnson, 1988).

Plasmid pGEX2TMspI (pGM2), is a construct (Taylor et al. 1993) which contains genes coding for the lac repressor protein, β lactamase which confers ampicillin resistance and a glutathione-S-transferase-M.*Msp*I (GST-M.*Msp*I) fusion protein under the transcriptional control of the P_{tac} promoter. The ampicillin resistance gene (β-lactamase 928 bp) provides the means of selection during the experiments when ampicillin is present in the growth medium.

The GST-M.*Msp*I fusion protein has full methyltransferase activity, and the GST component provides a means of purifying the enzyme by affinity chromatography. Moreover, since GST is a homo-dimer, M.*Msp*I will be an artificial dimer, (see figure 3.2). The P_{tac} promoter sequence (61 bp) which precedes the GST-M.*Msp*I gene facilitates control of GST-M.*Msp*I expression. Expression of the *lacI*^{q} gene (1080 bp) produces the Lac repressor which upon binding to the P_{tac} operator represses expression of GST-M.*Msp*I. However, the P_{tac} promoter sequence is 'leaky', in that LacI repression is not complete and therefore some GST-M.*Msp*I expression does occur in the absence of IPTG.

Figure 3.2: M.*Msp*I-GST fusion protein

Diagram illustrating the dimeric form of the M.*Msp*I –GST fusion protein.

3.2.3 CONSTRUCTION OF A PLASMID ENCODING THE C174G MUTATION OF DIMERISED M.MSPI

pGEX-M.*Msp*I$_{cass}$ was a gift from Dr. P. Hurd. It was made by introducing three unique sites into the region of the gene that encodes motif IV namely *Bsp*EI, *Nco*I, and *Nde*I. Although these restriction sites have not been introduced "silently" three amino acids were changed, namely; Q164 to E, F177 to W, and I180 to M, however these changes do not affect the enzyme activity (Hurd, 1996). The general features of the plasmid are shown in Figure 3.1.

One of the advantages of this "cassette" vector is that it facilitates subsequent mutagenesis experiments in motif IV. This feature was used to mutate C174 to G (see next section). A second feature is the presence of a new *Bsp*EI site (5'-TCCGGA-3') which contains a *Msp*I site. Methylation of this site was used as an indicator of the activity of the encoded enzyme.

3.2.3.1 INACTIVATION OF THE M.MSPI GENE BY INSERTION OF A KANAMYCIN^R GENE CASSETTE.

Since M.*Msp*I$_{C174G}$ gene is potentially lethal, in order to generate this mutant and to study its effects on *E.coli* it is necessary to reversibly inactivate the gene. In addition a selective method is required to identify the inactivate the M.*Msp*I gene. The Kanamycin resistance gene (Kanamycin^r) was selected for this purpose and was inserted into the unique *Xba*I site in the M.*Msp*I gene.

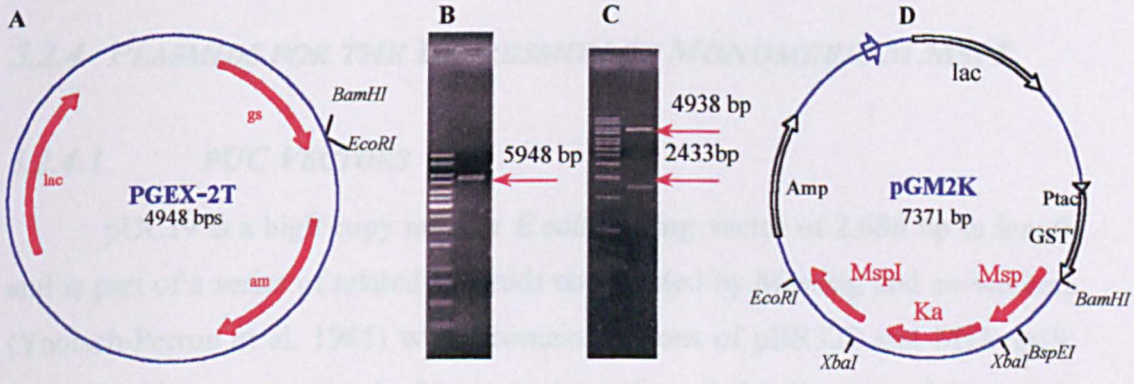
The Kanamycin^r gene from pUC-4K (Pharmacia) was used in the experiment. However, there was no suitable unique site in the M.*Msp*I gene flanking the kanamycin^r gene cassette and so an *Xba*I restriction site was introduced in the kanamycin^r gene using two oligodeoxynucleotide primers KANX3 and KANX5. The sequence of these primers are given in appendix 3.

The PCR reaction conditions were as described in section 2.2.5 with KANX5 as the downstream primer and KANX3 as the upstream primer with pUC-4K as template. The thermal cycler programme used consisted of: denaturation step at 94°C for 1 minute, high thermal ramp (40°C/minute) to 50°C, annealing for 1 minute at 50°C, rapid thermal ramp to 72°C, extension at 72°C, for 1 minute and then rapid thermal ramp to 94°C. The reaction was allowed to proceed for 30 cycles. The amplified 1180 bp PCR product was purified from a 1% agarose gel by the Gene Clean method (section 2.2.2.2) and cleaved with *Xba*I: pGM2 was also cleaved with *Xba*I. Following deproteinisation and precipitation (section 2.2.2), the Kanamycin^r gene (PCR product) and pGM2 were ligated as described in section (2.2.5.3).

The ligation mixture was used to transform *E.coli* RRI competent cells as described in section (2.2.6.4). The cells were plated on kanamycin and ampicillin containing LA plates. Possible recombinants were picked, grown in LB media (with ampicillin and kanamycin) and plasmid DNA was isolated using the 'mini-prep' procedure (see section 2.2.2.1). Samples of plasmid DNA prepared in this way were digested with *BamH*I and *EcoR*I, followed by analytical agarose gel electrophoresis. In this way it was possible to identify recombinant plasmids by the presence of a 2433bp fragment corresponding to the M.*Msp*I and kanamycin^r gene (see figure 3.3). The final construct was named **pGM2K**.

Figure 3.3: The pGEX2T and pGM2K vectors.

The pGEX2T vector has been used to clone the gene encoding M.*Msp*I in the dimerised form. A) Map of pGEX2T showing the *BamH*I and *EcoR*I sites. b) Photograph of an 1% agarose gel showing the product of digestion of pGEX2T which the *BamH*I/*EcoR*I band (4938 bp) has been recovered and used for cloning the M.*Msp*I gene. C) Photograph of a 1% agarose gel given the product of digestion of pGM2K with *BamH*I and *EcoR*I showing the M.*Msp*I and Kanamycin gene (2433bp) fragments. D) Map of pGM2K showing the *Xba*I site used for cloning the kanamycin gene, the *BamH*I and *EcoR*I sites used to clone the M.*Msp*I gene.



3.2.3.2 GENERATION OF PLASMID PGM2$_{C174G}$

As described above C174 has been targeted for mutation in this study and pGM2 was designed to facilitate rapid substitution of the Cys codon by oligodeoxynucleotide exchange. Two oligodeoxynucleotides were designed to mutate C174 to G: ABF22 and ABF23 (see figure 3.4). These two oligodeoxynucleotides encode the mutation C174 to G and introduce a new *Msp*I site in the gene that facilitate identification of mutants can be used as a marker for the mutated gene.

pGM2K was digested with two enzymes: *Bsp*EI and *Nco*I sequentially followed by 1% agarose gel electrophoresis to isolate the *Bsp*EI/*Nco*I fragment from the rest of the vector (7331bp) which was purified as described in section 2.2.2.2 and subsequently ligated to the annealed oligonucleotides. Finally the pGS2K encoding the C174G mutant was used to transform *E.coli* RRI cells using kanamycin and ampicillin selection (see figure 3.5). The constructed plasmid DNA was then verified by appropriate restriction enzymes and named **pGM2$_{C174G}$K.**

Another construct was made using the same strategy to be used as a control, which encodes M.*Msp*I$_{C174G}$ with one base pair deleted to facilitate comparison between active and inactive mutant Mtase coding region. This construct was made by deleting one base from that region of the M.*Msp*I gene encoding motif IV (see figures 3.4 and 3.5). This frame shift mutant should transform *E.coli* efficiently without any negative effect and should act as a useful control for transformation efficiency. Two oligodeoxynucleotides were used to construct this vector ABF22d and ABF23d (see appendix 3 for sequences) and it is named **pGM2$_{C174G}$d1K.**

*3.2.4 PLASMIDS FOR THE EXPRESSION OF MONOMERIC M.*MSP*I*

3.2.4.1 PUC VECTORS

pUC19 is a high copy number *E.coli* cloning vector of 2,686 bp in length and is part of a series of related plasmids constructed by Messing and co-workers (Yanisch-Perron et al. 1985) which contain portions of pBR322 and M13mp19. The pUC vectors encode the N-terminal portion of the *E.coli lacZ* gene that

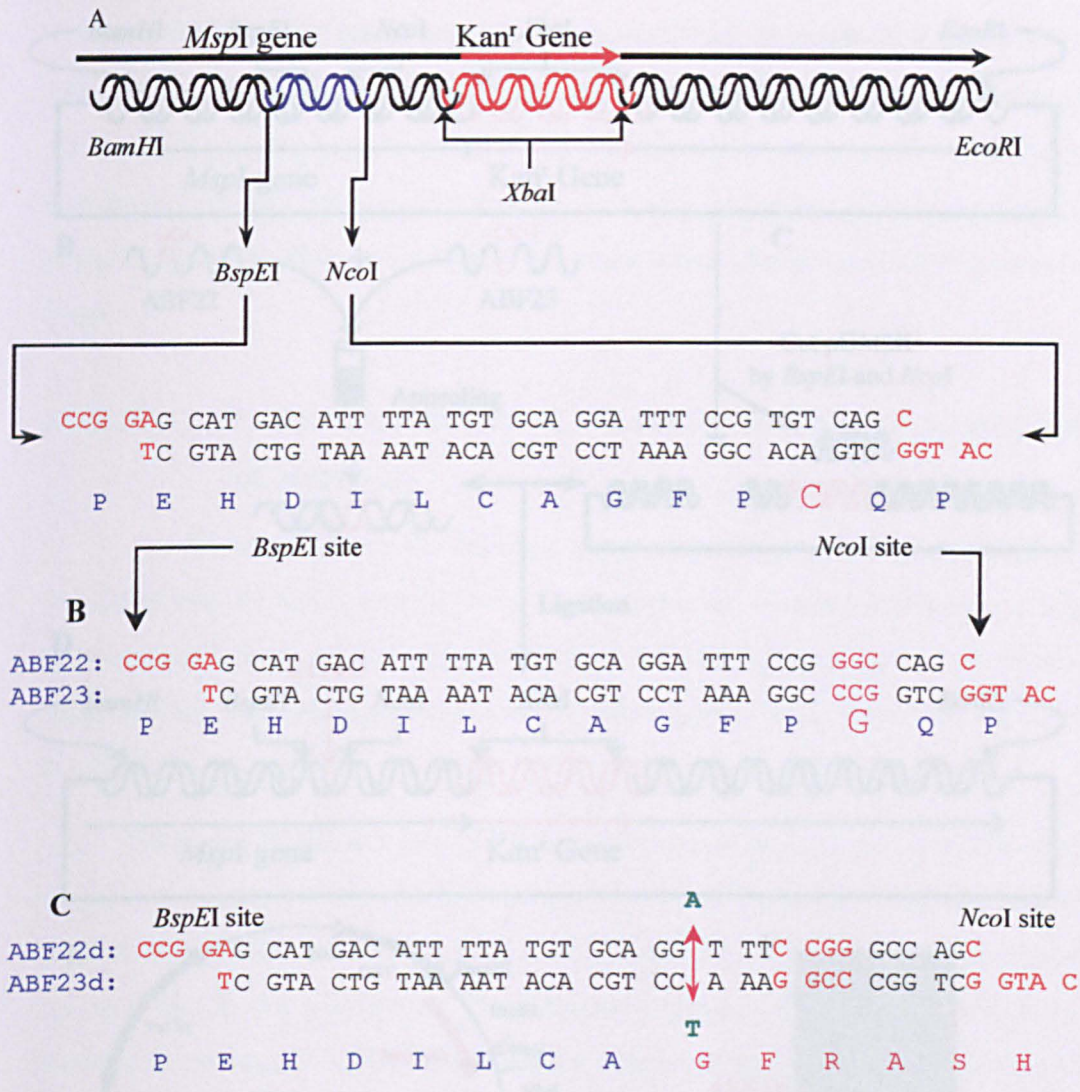


Figure 3.4: Diagram illustrating the sequences and positions of the oligodeoxynucleotides used to mutate M.*Msp*I gene at C_{174} to G. A) Diagram of the M.*Msp*I kanamycin gene with the unique restriction sites used followed by the wild type sequence of the 30 bp between *Bsp*EI and *Nco*I sites. B) The sequence of ABF22 and ABF23 the oligodeoxynucleotides used to mutate C174 to G. C) the same sequence used to mutagenise the sequence encoding C174G however one bp was deleted as a control for transformation efficiency.

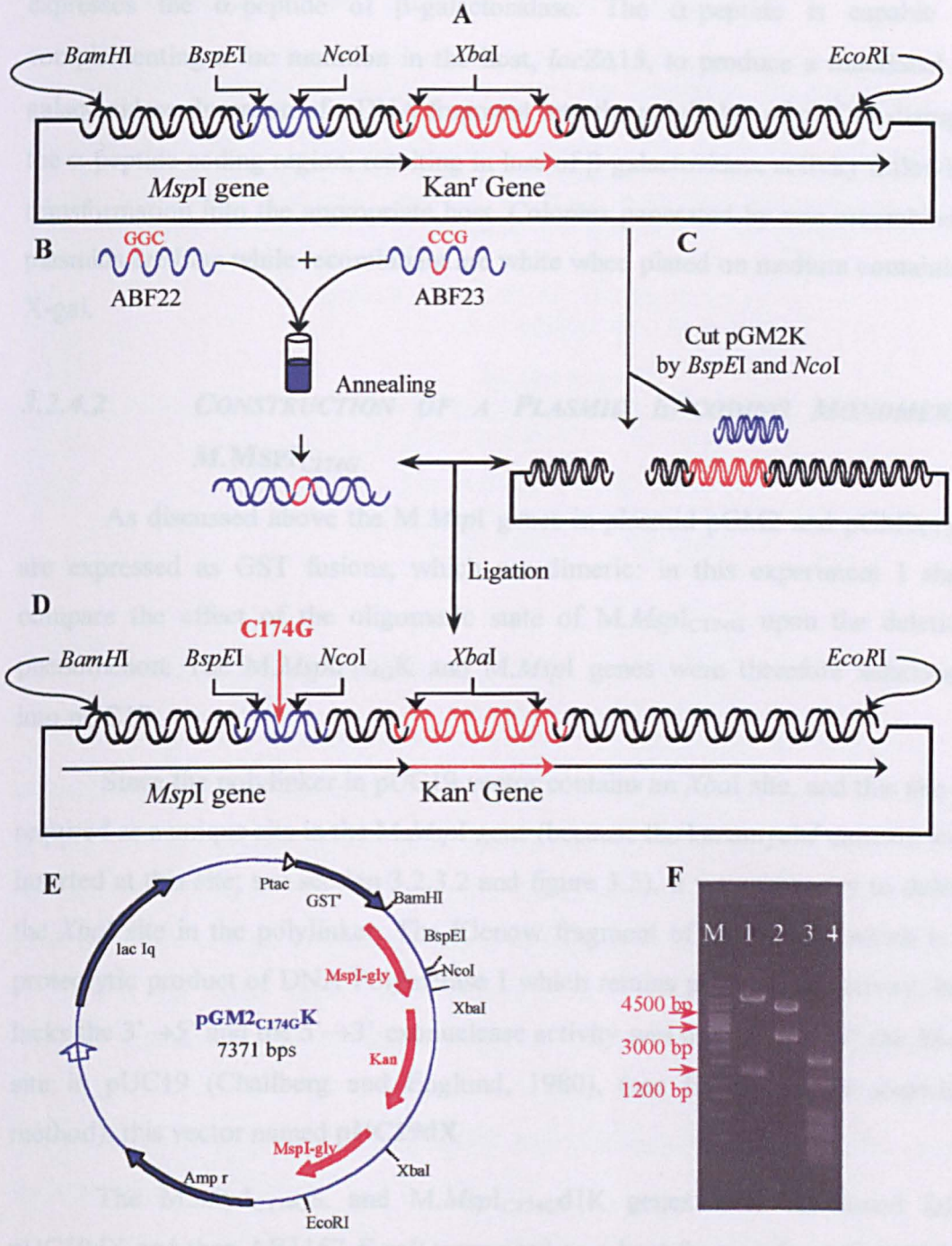


Figure 3.5: Diagram illustrating the methodology used to construct the pGM2$_{C174G}$K.

The unique restriction sites in pGM2K vector was used to mutagenesis C174G in the inactivated *MspI* gene are presented in (a), the two oligodeoxynucleotides encoding the C174G were annealed using thermal cycler (b) and see text for details. The pGM2K was digested then with *Bsp*EI and *Nco*I sequentially (c) the cut pGM2K vector was run on a 1% agarose gel electrophoresis to remove the *Bsp*EI/*Nco*I fragment (40bp) and the band corresponding to the rest of the vector (7331bp) was recovered and subjected to Gene Clean purification. The purified vector (pGM2K) was then ligated to the annealed insert (d). The pGS2K encoding $C_{174}G$ plasmid DNA was used to transform *E.coli* RRI cells using kanamycin and ampicillin selection. Possible recombinants were picked, grown in 5ml of LB media (with ampicillin and kanamycin) and plasmid DNA was isolated using the 'mini-prep' procedure. The constructed plasmid DNA was named **pGM2$_{C174G}$K**. In (e) map of the final construct pGM2$_{C174G}$K. And in (f) Photograph of a 1% agarose gel showing the digested pGM2K with *Xba*I (lane 1) *BamH*I and *EcoR*I (lane 2), *Msp*I (lane 3). 1Kb marker used.

expresses the α-peptide of β-galactosidase. The α-peptide is capable of complementing a *lac* mutation in the host, *lacZ*Δ15, to produce a functional β-galactosidase. Insertion of a DNA fragment into the polylinker of pUC19 disrupts the α-peptide coding region, resulting in loss of β-galactosidase activity following transformation into the appropriate host. Colonies generated by non-recombinant plasmids are blue while recombinant are white when plated on medium containing X-gal.

3.2.4.2 *CONSTRUCTION OF A PLASMID ENCODING MONOMERIC M.*MSP*I*$_{C174G}$

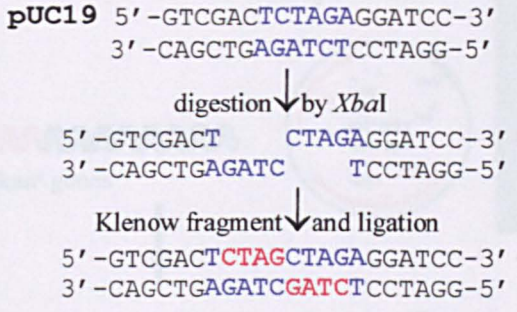
As discussed above the M.*Msp*I genes in plasmid pGM2 and pGM2$_{C174G}$ are expressed as GST fusions, which are dimeric: in this experiment I shall compare the effect of the oligomeric state of M.*Msp*I$_{C174G}$ upon the deletion phenomenon. The M.*Msp*I$_{C174G}$K and M.*Msp*I genes were therefore subcloned into pUC19.

Since the polylinker in pUC19 vector contains an *Xba*I site, and this site is required as a unique site in the M.*Msp*I gene (because the kanamycin^r cassette was inserted at this site; see section 3.2.3.2 and figure 3.5), it was necessary to delete the *Xba*I site in the polylinker. The Klenow fragment of DNA PolI, which is a proteolytic product of DNA Polymerase I which retains polymerase activity, but lacks the 3`→5` and the 5`→3` exonuclease activity was used to "fill in" the *Xba*I site in pUC19 (Challberg and Englund, 1980), (see figure 3.6 for stepwise method): this vector named **pUC19dX**

The M.*Msp*I$_{C174G}$K and M.*Msp*I$_{C174G}$d1K genes were subcloned into pUC19dX and then AB1157 *E.coli* were used as a host for transformation of the plasmid. Possible recombinants were picked, grown in LB media with ampicillin and kanamycin for selection and plasmid DNA was isolated using the 'mini-prep' procedure. The constructed plasmid DNA was checked by digestion with *BamH*I and *EcoR*I (see figure 3.7) the constructs were named **pUM1$_{C174G}$K** and **pUM1$_{C174G}$d1K.**

Figure 3.6: deletion of *Xba*I site in pUC19 using Klenow fragment enzyme.

The pUC19 vector was digested with *Xba*I and after 3 hrs incubation at 37°C the DNA was recovered from an agarose gel and subjected to Gene Clean purification. The *Xba*I digested site was then filled using Klenow fragment enzyme, 2µl buffer, and 2µl dNTP (5µM). The reaction was incubated at 25°C for 15 minutes followed by immediate DNA precipitation, deprotenisation, and then ligation. The ligated DNA was used to transform *E.coli* AB1157 competent cells. Possible recombinants were picked, grown in 5ml of LB media (with ampicillin) and plasmid DNA was isolated using the 'mini-prep' procedure. The plasmid DNA was checked by digestion with *Xba*I. This vector named **pUC19dX**



3.3 HHA*I C5 DNA MTASE*

M.*Hha*I is a C5-Mtase (EC 2.1.1.37) that methylates the sequence 5'-GCGC-3' in double-stranded DNA to yield 5'-G^{me}CGC-3' and is part of an RM system in *Haemophilus haemolyticus* (Roberts et al. 1976). M.*Hha*I is one of the smallest of the C5-Mtases, containing 327 amino acid having a MW of 37 kDa. Its gene has been cloned, sequenced and over-expressed in *E.coli* (Wu and Santi, 1987; Wu and Santi, 1988; Klimasauskas et al. 1991) (see section 1.3.3 for M.*Hha*I structure and mechanism of reaction). The full DNA sequence and amino acid translation of M.*Hha*I is illustrated in figure 3.8. in this experiment two constructs of M.*Hha*I were used pGH2K and pGH2$_{C81G}$K (these construct were made previously in the lab).

3.4 EXPERIMENTAL DESIGN

3.4.1 TRANSFORMATION

The Kanamycin^r gene was used to inactivate the Mtase gene and therefore in order to restore the activity the inserted kanamycin^r gene was removed by digestion with a suitable nuclease (*Xba*I for M.*Msp*I gene and *Sal*I for M.*Hha*I). After gel purification, the Mtase encoding DNA was subjected to ligation and

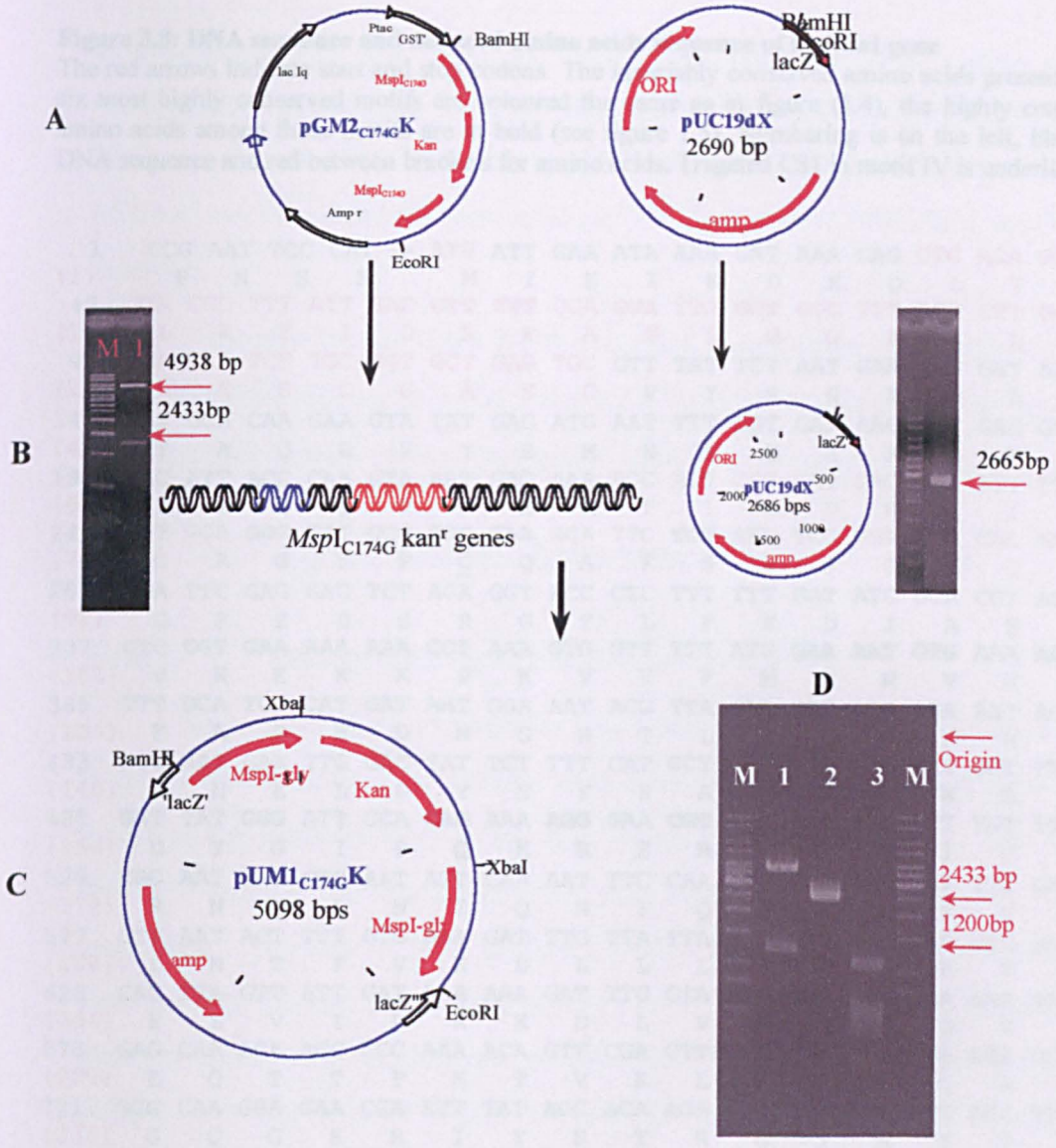


Figure 3.7: Stepwise constructing of pUM1$_{C174G}$K and pUM1$_{C174G}$d1K.
The pGM2$_{C174G}$K, pGM2$_{C174G}$Kd1, and pUC19dX were digested by *BamH*I and *EcoR*I (a). The band corresponding to the *Msp*I-Kanamycin genes (2433 bp) and pUC19dX (2665 bp) were purified from agarose gel (b). The *Msp*I$_{C174G}$K and *Msp*I$_{C174G}$d1K genes were cloned to pUC19dX (b). And then subjected to transform *E.coli* AB1157. Possible recombinants were picked, grown in LB media with ampicillin and kanamycin and plasmid DNA was isolated using the 'mini-prep' procedure. The constructed plasmid (pUM1$_{C174G}$K illustrated in C) DNA was checked by digestion with *Xba*I in one reaction (lane 1) and *BamH*I and *EcoR*I (lane 2) and *Msp*I (lane 3).the one Kb ladder marker used.

Figure 3.8: DNA sequence and deduced amino acids sequence of M.*Hha*I gene
The red arrows indicate start and stop codons. The invariably conserved amino acids present in the six most highly conserved motifs are coloured the same as in figure (1.4), the highly conserved amino acids among those motifs are in bold (see figure 1.5). Numbering is on the left, black for DNA sequence and red between brackets for amino acids. Trageted C81 in motif IV is underlined.

```
  1   CCG AAT TCC CAT → ATG ATT GAA ATA AAA GAT AAA CAG CTC ACA GGA
(1)     P   N   S   H     M   I   E   I   K   D   K   Q   L   T   G
 49  TTA CGC TTT ATT GAC CTT TTT GCA GGA TTA GGT GGC TTT AGA CTT GCT
(12)   L   R   F   I   D   L   F   A   G   L   G   G   F   R   L   A
 97  TTA GAA TCT TGC GGT GCT GAG TGC GTT TAT TCT AAT GAA TGG GAT AAA
(28)   L   E   S   C   G   A   E   C   V   Y   S   N   E   W   D   K
145  TAT GCA CAA GAA GTA TAT GAG ATG AAT TTT GGT GAA AAG CCT GAG GGC
(44)   Y   A   Q   E   V   Y   E   M   N   F   G   E   K   P   E   G
193  GAC ATT ACC CAA GTA AAT GAG AAA ACC ATT CCT GAT CAC GAC ATT TTA
(60)   D   I   T   Q   V   N   E   K   T   I   P   D   H   D   I   L
241  TGT GCA GGG TTT CCG TGC CAA GCA TTC TCT ATT TCC GGA AAA CAA AAA
(76)   C   A   G   F   P   C   Q   A   F   S   I   S   G   K   Q   K
289  GGA TTC GAG GAC TCT AGA GGT ACC CTC TTT TTT GAT ATC GCA CGT ATT
(92)   G   F   E   D   S   R   G   T   L   F   F   D   I   A   R   I
337  GTC CGT GAA AAA AAA CCT AAA GTG GTT TTT ATG GAA AAT GTG AAA AAT
(108)  V   R   E   K   K   P   K   V   V   F   M   E   N   V   K   N
385  TTT GCA TCG CAT GAT AAT GGA AAT ACG TTA GAA GTT GTA AAA AAT ACA
(124)  F   A   S   H   D   N   G   N   T   L   E   V   V   K   N   T
433  ATG AAT GAA TTG GAC TAT TCT TTT CAT GCT AAA GTA TTA AAT GCT TTA
(140)  M   N   E   L   D   Y   S   F   H   A   K   V   L   N   A   L
481  GAT TAT GGG ATT CCA CAG AAA AGG GAA CGT ATC TAT ATG ATT TGT TTT
(156)  D   Y   G   I   P   Q   K   R   E   R   I   Y   M   I   C   F
529  CGC AAT GAT CTC AAT ATT CAA AAT TTC CAA TTT CCA AAA CCT TTT GAG
(172)  R   N   D   L   N   I   Q   N   F   Q   F   P   K   P   F   E
577  CTT AAT ACT TTT GTG AAA GAT TTG TTA TTA CCT GAT AGC GAG GTG GAA
(188)  L   N   T   F   V   K   D   L   L   L   P   D   S   E   V   E
625  CAC TTA GTT ATT GAT AGA AAA GAT TTG GTA ATG ACA AAC CAA GAA ATT
(204)  H   L   V   I   D   R   K   D   L   V   M   T   N   Q   E   I
673  GAG CAA ACA ACC CCC AAA ACA GTT CGA CTT GGT ATT GTA GGA AAA GGT
(220)  E   Q   T   T   P   K   T   V   R   L   G   I   V   G   K   G
721  GGG CAA GGA GAA CGA ATT TAT AGC ACA AGA GGC ATT GCA ATT ACC TTA
(236)  G   Q   G   E   R   I   Y   S   T   R   G   I   A   I   T   L
769  TCT GCT TAT GGT GGC GGC ATT TTC GCT AAG ACA GGG GGA TAT TTA GTA
(252)  S   A   Y   G   G   G   I   F   A   K   T   G   G   Y   L   V
817  AAC GGG AAG ACA CGG AAA TTA CAC CCT AGA GAG TGT GCT AGA GTA ATG
(268)  N   G   K   T   R   K   L   H   P   R   E   C   A   R   V   M
865  GGC TAC CCA GAT AGT TAT AAA GTC CAC CCG TCA ACC AGC CAA GCA TAT
(284)  G   Y   P   D   S   Y   K   V   H   P   S   T   S   Q   A   Y
913  AAA CAA TTT GGT AAC TCA GTT GTT ATC AAT GTA CTT CAA TAT ATT GCT
(300)  K   Q   F   G   N   S   V   V   I   N   V   L   Q   Y   I   A

961  TAT AAC ATT GGT TCA TCA TTA AAT TTC AAA CCA TAT TAA ←
(316)  Y   N   I   G   S   S   L   N   F   K   P   Y  STOP
```

transformation. In the M.*Msp*I case four plasmids were used namely: pGM2$_{C174G}$K, pGM2$_{C174G}$ d1K, pUM1$_{C174G}$K, and pUM1$_{C174G}$ d1K the purified DNA length was 6201 bp, 6200 bp, 3928 bp, and 3927 bp respectively. In the M.*Hha*I case two plasmids were used (pGH2$_{C81G}$K and pGH2K) both are 5949 bp in length.

Typically, 30 ng of the DNA was used to transform pre-prepared (200µl) competent cells (see sections 3.4.2 and 2.2.4.4). In some cases however, due to the toxicity of the mutant Mtases, large-scale transformations were required.

3.4.2 DETECTION OF DELETIONS

3.4.2.1 DETECTING DELETIONS USING PCR

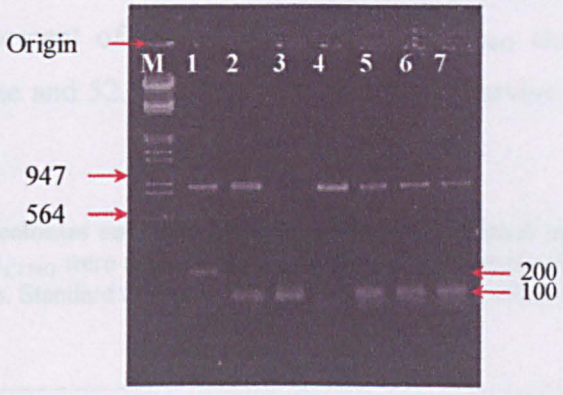
For rapid analysis of plasmids recovered from cells, colony PCR was used to check for deletions. Four primers were designed, two for amplifying the region encoding the catalytic loop in the M.*Msp*I gene the other two as controls since they will amplify the ampicillin gene, (see appendix 3 for the oligonucleotides sequences,). The two M.*Msp*I primers (ABF13 and ABF14) flank the two *Msp*I sites found in the region encoding motif IV in the M.*Msp*I gene (see figure 3.1). If there are deletions between these two sites, the PCR product will be 170 bp while if there is no deletion, the fragment will be 200 bp. Another way of checking the deletion was exposing the PCR product to *Bsp*EI, this should cut into the middle of the PCR product. If there are deletions it will give one band 170 bp in length: However, if there are no deletions digestion of the PCR product will yield two 100 bp fragments (see figure 3.9).

The previous two primers ABF13 and ABF14 were designed to detect deletion from the middle of the M.*Msp*I gene. In addition another two primers were used to detect deletion in the M.*Msp*I gene, they were ABF17 and ABF18 (see appendix 3). ABF17 will prime at the 5' end of the M.*Msp*I gene while ABF18 will primes in the 3' end of the gene. In addition a control pair of primers that amplified the β-lactamase gene were used. ABF15 and ABF16 prime at the 5' and 3' ends of the ampicillin^r gene respectively to yield a 928 bp β-lactamase gene fragment. The sequences of ABF13, ABF14, ABF15 and ABF16 are given in appendix 3. The colony PCR was carried out as described in section 2.2.5.2

then the results analysed by running 5 µl of the product on a 1.7 % agarose gel as shown in figure 3.9 for an example of the data.

Figure 3.9: Deletion detection by the PCR method.

Photograph of a 1.7 % agarose gel of 5 µl PCR products from pGM2 using four primers ABF13, 14, 15, and 16.to amplify the ampicillin gene and 200 bp of the M.*Msp*I gene. The first lane (M) is marker λDNA *BamHI*/*EcoR*I; lane 1 pGM2 using the four primers; lane 2 the products in lane 1 digested by *Bsp*EI; lane 3 is the control using the same template but two primers ABF 13, and 14, lane 4 another control using the other two primers ABF15, and 16. The remaining three lanes 5-7 are products of $pUM1_{C174G}$ colony PCRs. There were no deletion in these three samples. Samples loaded in lanes 2-7 were digested with *Bsp*EI.



3.4.2.2 RESTRICTION MAPPING OF DELETION MUTANTS.

The other method used to analyse the deletions was restriction mapping using suitable restriction enzymes. In the case of the M.*Msp*I_{C174G} experiments, three enzymes were used *BamH*I, *Bsp*EI, and *EcoR*I. Digestion of the M.*Msp*I gene by these three enzymes will yield two fragments 491 bp (*BamHI*/*Bsp*EI) and 772 bp (*Bsp*EI/*EcoR*I) plus the rest of the plasmid (4938bp in the pGM2 plasmid) and (2665bp in the pUM1 plasmids). If there are deletions one or more of these three enzymes will not cut or they will cut, but the fragment will be different in size than expected. As described above, there are two advantages of using *Bsp*EI restriction enzyme first to check the activity of the M.*Msp*I (*Bsp*EI will not cut if the outer C of the TCCGGA is methylated by M.*Msp*I), the second, since it is nearly in the middle of the gene and between two *Msp*I sites is ideal for detecting the deletions in the gene.

Two restriction enzymes were used to detect deletions in M.*Hha*I gene, *EcoR*I and *Sal*I, these two enzymes are expected to yield a 991bp fragment (and the plasmid). Since there was just one *Hha*I site in the M.*Hha*I gene it was easy to detect a deletion within the gene itself.

3.5 DELETION IN THE GENE ENCODING M.MSPI

3.5.1 DELETION IN THE GENE ENCODING MONOMERIC M.MSPI

pUM1$_{C174G}$ and pUM1d1$_{C174G}$ were used to transform AB1157 competent cells. Ninety nine percent of the cells in the pUM1$_{C174G}$ carried the plasmid encoding active Mtase and 52.9% of them were able to survive. See table 3.1 for details.

Table 3.1: Number of colonies survived after transforming mutated monospecific Mtases. pUM1$_{C174G}$, and pUM1d1$_{C174G}$ were used to transform AB1157 competent cells. Results presented are the mean of triplicates. Standard deviation (sd) and transformation efficiency are calculated.

	pUM1d1$_{C174G}$	pUM1$_{C174G}$
Number of colonies	**1950**	**1033**
Sd	**86.6**	**57.7**
Transformation efficiency	**3.9 X 10^5**	**2 X 10^5**
Percentage of survived cells (%)	**52.9**	

Twelve colonies were selected from the pUM1$_{C174G}$ plate and were subjected to colony PCR (see section 2.2.5.2) using four primers ABF13, ABF14, ABF15, and ABF16 as described in section 3.4.4.1 and figure 3.10 shows an agarose gel of the PCR products after digestion of the DNA with *Bsp*EI. It is clear from these bands that there was no deletion in any plasmid recovered within the 200bp amplified.

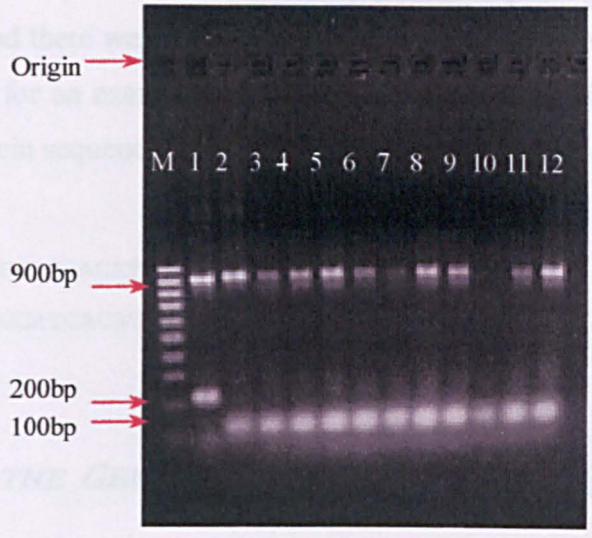


Figure 3.10: Colony PCR products after digestion by *Bsp*EI.

1.7 % agarose gel illustrates eleven products of colony PCR after digestion with *Bsp*EI to find out if there are deletion in motif IV. The control ampicillin products (928bp) are shown. The marker was 1µl of 100bp ladder, lane 1 pUM1 uncut as control, and lanes 2-12 are the colony PCR products 5µl of the PCR products digested and loaded in the gel from each sample.

Plasmid DNA was prepared from the twelve colonies to see whether deletions were present elsewhere in the M.*Msp*I gene. DNA was digested by *BamH*I, *EcoR*I, and *Bsp*EI as described in section 3.4.2.2. All the plasmids were not methylated by M.*Msp*I and they all gave the expected restriction map (see figure 3.11).

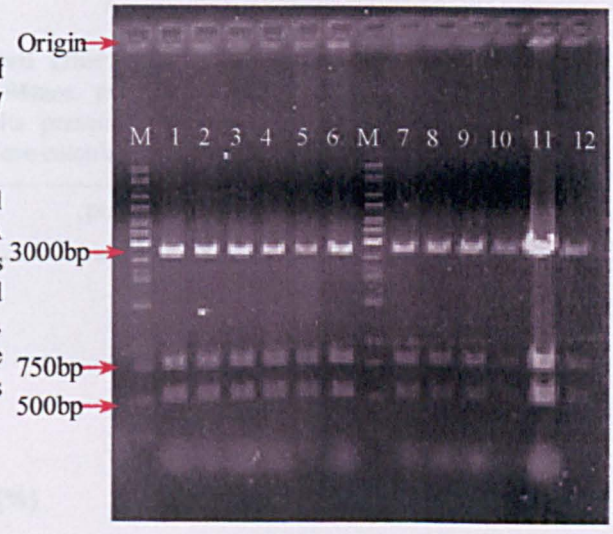


Figure 3.11: Restriction mapping of DNA isolated from AB1157 transformed with pUM1$_{C174G}$.

Photograph of a 1 % agarose gel illustrated the digestion of 11 DNA samples isolated from AB1157 cells transformed with monomeric mutated M.*Msp*I with *BamH*I, *EcoR*I, and *Bsp*EI. 1kb ladder marker was used and lane one pUM1d1 digested by the same enzymes as control.

3.5.1.1 Nucleotide Sequences of Plasmids Recovered from Cells Transformed with pUM1$_{C174G}$.

The nucleotide sequences of three plasmids isolated from cells transformed with pUM1$_{C174G}$ were obtained using two primers; ABF17 and ABF18 (see appendix 3 for table of oligonucleotides, section 2.2.7 for sequencing

methods). Aligning the nucleotide sequences obtained from the three plasmids showed that they were identical and there were no deletions. However, there were some point mutations (see below for an example). However this particular point mutation does not change the protein sequence, Since ACT and ACA both encode Threonine.

GTTGAGGCTACAACTATTCCGGAGCATGACATTTTATGTGCAGGATTTCCGGGCCAG

GTTGAGGCTACAACaATTCCGGAGCATGACATTnTATGTGCAGGATTTCCGGGCCAG

3.5.2 *DELETIONS IN THE GENE ENCODING DIMERIC GST-M.MSPI*

$pGM2_{C174G}$ and $pGM2d1_{C174G}$ were used to transform AB1157 competent cells in order to test the effect of the glycine mutant on cell survival. There were no colonies obtained in a typical transformation experiment. However, large-scale transformations did produce a significant number of transformants. Only 3.8% of the colonies were able to survive compared with the control transformation (see table 3.2).

Table 3.2: Analysis of colonies recovered after *E.coli* was transformed with plasmids encoding mutant, dimeric monospecific Mtase. $pGM2_{C174G}$ and $pGM2d1_{C174G}$ were used to transform AB1157 competent cells. Results presented are the mean of triplicates. Standard deviation (sd) and transformation efficiency are calculated.

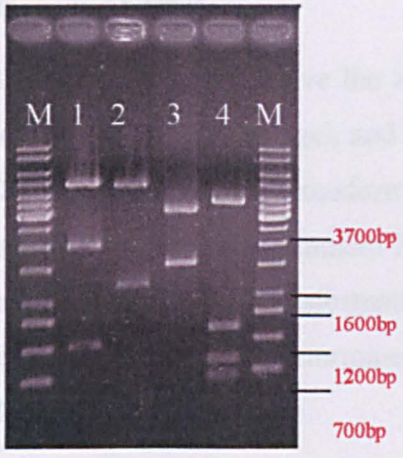
	$pGM2d1_{C174G}$	$pGM2_{C174G}$
Number of colonies	1666.6	63.3
sd	577.3	15.2
Transformation efficiency	3.3×10^5	1.2×10^4
Percentage of repaired cells (%)	**3.8**	

Plasmid DNA was prepared from ten colonies to see whether deletions were present in the M.*Msp*I gene. DNA was digested by *BamH*I, *EcoR*I, and *Bsp*EI as described in section 3.4.2.2. All the plasmids were not methylated by M.*Msp*I and they gave two patterns of restriction map (see figure 3.12). These two

patterns of bands indicate that there was rearrangement and deletion in the plasmid DNA and each one forms 50% of the total population (see figure 3.12).

In the first pattern when plasmid DNA was digested by *BamH*I, *Bsp*EI, and *EcoR*I two bands results in the agarose gel 3500 and 1600 bp which shows that the plasmid has lost 1100 bp. Strangely this pattern suggest that the M.*Msp*I gene is 1600 bp instead of 1200 bp, whilst the total size of the plasmid is reduced by 1100 bp. The second plasmid pattern when DNA was digested by the same three enzymes, yields four bands instead of three 3700, 700, 400, and 300 bp: the bands show that the M.*Msp*I gene is now 1400 bp in size (700 + 400 + 300) and the total plasmid DNA size is 5100 bp. Interestingly new restriction sites have been introduced into the M.*Msp*I gene (see figure 3.12).

Figure 3.12: Photograph of 1% agarose gel of pGM2$_{C174G}$ DNA isolated from survived cells. The pGM2$_{C174G}$ DNA was digested by *BamH*I, *EcoR*I, and *Bsp*EI. Two patterns arise from the digested DNA. Lanes 1 and 2 are controls pGM2K and pGM2 respectively. Digested pGM2 gave two bands due to the methylation by M.*Msp*I (*Bsp*EI does not cut). Lane 3 and 4 the two patterns of recovered pGM2$_{C174G}$ DNA which shows the rearrangement in the vector.



The nucleotide sequences of recovered plasmids showing the different restriction maps, were obtained using three primers; ABF17, ABF15 and ABF18 (see appendix 3 for table of oligonucleotides and section 2.2.7 for sequencing method). ABF15 was used to sequence the β-lactamase gene as control for the sequencing. Nucleotide sequencing of the "rearrangement" plasmid failed while β-lactamase control sequencing was successful.

For a more detailed understanding of the rearrangement, restriction mapping was carried out using four restriction enzymes *BamH*I, *Xba*I, *EcoR*I, and

*Pst*I. Since all the cells grew on ampicillin plates, the β-lactamase gene must be intact and because *Pst*I is a unique site in the β-lactamase gene, it was used as an indicator for the size of rearranged DNA (see figure 3.13). From the restriction mapping data presented in figure 3.13, it can be concluded that in the first type of rearrangement there was a deletion of 1400bp between *BamH*I and *Pst*I, 300bp of that deletion between *Xba*I and *EcoR*I. Furthermore, there was an insertion of 300bp between *Pst*I and *BamH*I. The net result is a deletion of 1100 bp between *BamH*I and *Xba*I. In the other type of rearrangement, there was an insertion and a deletion, the insertion was 1540bp between *EcoR*I and *Pst*I and the deletion was 2240bp between *Pst*I and *EcoR*I (see figure 3.13)

3.6 DELETION IN THE GENE ENCODING M.HHAI

So far DNA damage has been found in plasmids encoding the M.*Msp*I$_{C174G}$ gene. Does the same error-prone deletion occur with other monospecific Mtases? In order to test the generality of this phenomenon, a similar set of experiments was repeated using M.*Hha*I$_{C81G}$.

pGH2$_{C81G}$K was digested with *Xba*I, to remove the kanamycin^r gene and the rest of the DNA was purified from a 1% agarose gel, and ligated to regenerate the M.*Hha*I$_{C81G}$ gene. Three plasmids were used to transform AB1157 competent cells: a) pGH2$_{C81G}$ to study the effect of the glycine mutant in the cells, b) pGH2 as positive control and to compare between the transformation efficiency in the wild type and the mutant, and c) pLitmus to determine the transformation efficiency of the competent cells.

The transformation efficiency of AB1157 competent cells was 1X10^6: with the wild type M.*Hha*I there were 90 colonies while there was 15 colonies recovered after transforming the M.*Hha*I$_{C81G}$ mutant plasmid. There were no colonies in the control kanamycin plate therefore all of the colonies from the M.*Hha*I$_{C81G}$ plate were carrying the active gene and no "uncut" background plasmid. The percentage of repair was 16.6%. However, in order to corroborate these findings, 8 colonies were grown and plasmid DNA was restriction mapped with *EcoR*I and *Sal*I.

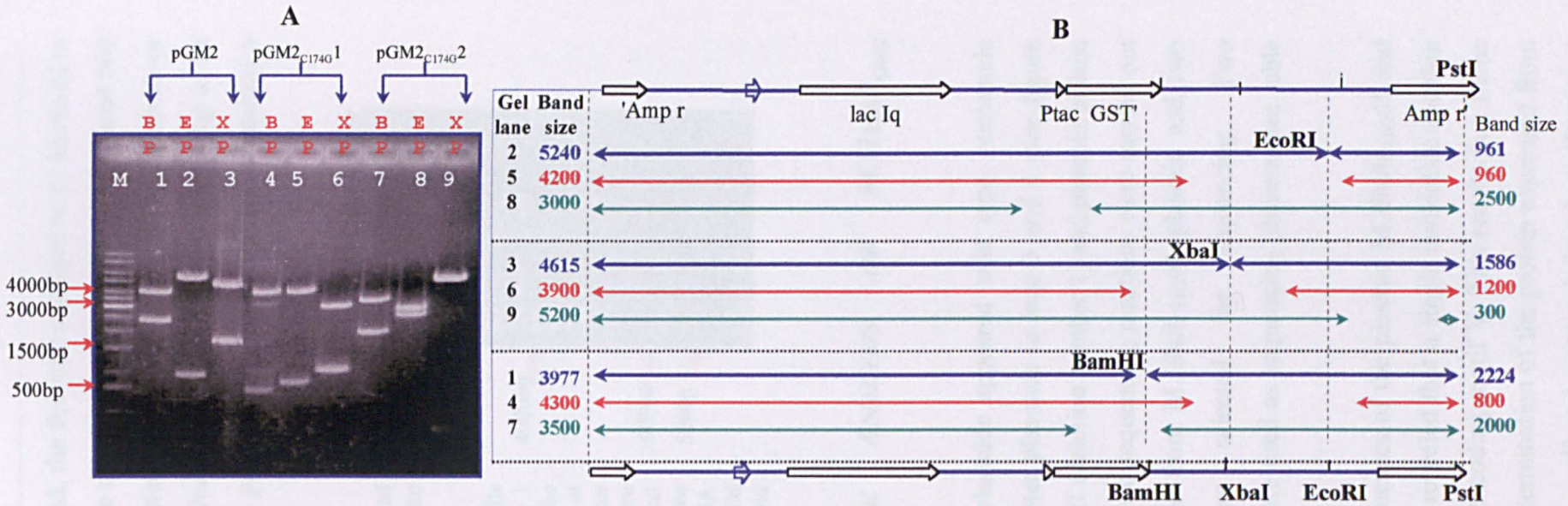
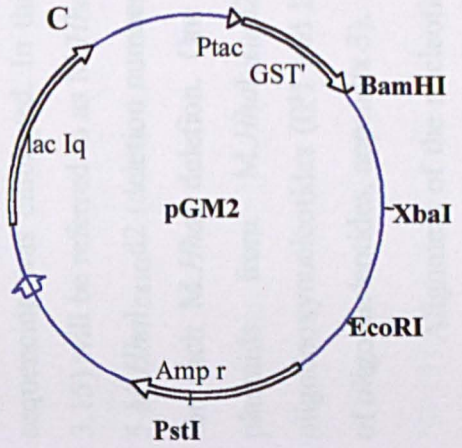


Figure 3.13: Restriction analysis of the two rearrangements in $pGM2_{C174G}$
pGM2 and the two plasmids rearranged were digested with *Pst*I and one of the three enzymes BamHI, XbaI, and EcoRI. In (A) photograph of an agarose gel; lanes 1, 2, and 3 pGM2 was digested with *Pst*I and *BamH*I (PB), *Pst*I and *EcoR*I (PE), and *Pst*I and *Xba*I (PX) respectively; lanes 4, 5, and 6 the rearranged DNA 1 were digested with the same enzymes in the same order; lanes 7, 8, and 9 the rearranged DNA 2 was digested with the same enzymes in the same order. In (B) a diagram showing the fragment sizes from each lane, to simplify comparison of the results the fragment sizes produced from each reaction are presented together. The smaller fragment sizes are presented on the right and the larger on the left: the fragment sizes are presented opposite each line. The pGM2 bands are shown in blue, rearranged $pGM2_{C174G}1$ in red and rearranged $pGM2_{C174G}2$ in green. (C) A map of pGM2 the enzymes sites in the plasmid are shown.



Two patterns of bands were found, the first one appeared to be identical to the wild type M.*Hha*I gene in length, the other contained a deletion of around 560 bp, and since both plasmids were digested with *EcoR*I and *Sal*I the deletions or point mutation(s) were within the M.*Hha*I gene itself. Plasmids appearing to be wild type represented 75% of the total plasmid population, the rest contained a visible deletion (see figure 3.14).

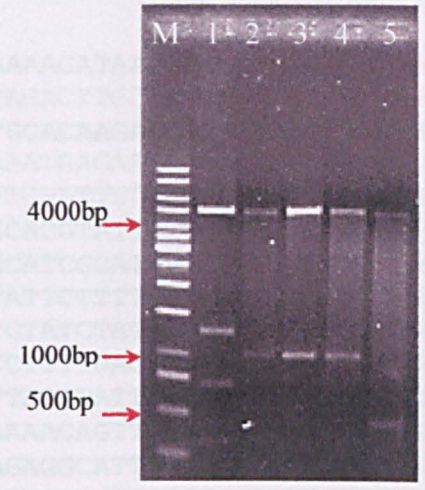


Figure 3.14: Restriction mapping of DNA with *EcoR*I and *Xho*I illustrate the deletion in M.*Hha*I gene.

Plasmid DNA prepared from surviving cells were digested with *EcoR*I and *Xho*I. The 1 % agarose gel shows the deletions in the M.*Hha*I$_{C81G}$ gene after transformation of *E.coli* AB1157. 1kb DNA ladder marker was used. Lane 1 pGH2K, four bands were produced owing to the presence of two *Sal*I sites. Lane 2 pGH2 and lanes 3-5 three examples of pGH2$_{C81G}$. It is clear that DNA in lanes 3 and 4 plasmids appear as wild type (~991 bp) while in lane 5 the plasmid differs by ~550 bp

3.6.1 *Nucleotide Sequence Analysis of M.*HhaI$_{C81G}$ *Encoding Plasmids*

While 75% of the recovered plasmids appeared wild type, nucleotide sequencing was employed. In this chapter plasmids in lane 3 and 4 (see figure 3.15) will be referred to as M.*Hha*I$_{C81G}$d1 (deletion number 1) and plasmid in lane 5 M.*Hha*I$_{C81G}$d2 (deletion number 2). Two sequencing reactions were carried out for each M.*Hha*I deletion. One sample from M.*Hha*I$_{C81G}$d1 plasmid and two plasmids from M.*Hha*I$_{C81G}$d2 were selected for sequencing. Two oligodeoxynucleotides (IP1 and IP4) were used as sequencing primers (see table of oligonucleotides, appendix 3).

Alignment of the nucleotide sequences of the plasmid M.*Hha*I$_{C81G}$d1 and M.*Hha*I$_{C81G}$d2 illustrated in figure 3.15 revealed that a single nucleotide deletion had occurred in the region of the gene encoding motif V. This results in a frame shift after base 307 (A) and subsequent termination of the protein expressed from this gene. The full sequences of M.*Hha*I$_{C81G}$d1 and protein deduced are illustrated in figure 3.15.

Figure 3.15: Nucleotides sequence and amino acid sequence deduced from M.*Hha*I$_{C81G}$d1

(a) Nucleotide sequence of M.*Hha*I$_{C81G}$d1 colours are the same as those for the primary structure motifs figure 1.4). The unique *Hha*I site (GCGC), and sequences encoding Gly (GGC) in motif IV are in bold and underlined. The last base before the deletion and first base after ligation is in bold green. The *Eco*RI and *Xho*I sites flanking the gene are in blue bold. Small letters to show the difference between the mutant and the wild type M.*Hha*I gene.

(b) Protein deduced from sequence in (A) aligned with the wild type M.*Hha*I protein coloured as in figure 1.4. Unmatched amino acids are coloured by green starting from amino acid G98 and downstream to the S121, which replaced with stop codon.

A

5′ **GGATCC**CCGAATTCCCAT**ATG**ATTGAAATAAAAGATAAACAGCTCACAGGATTACGCT
TTATTGACCTTTTTGCAGGATTAGGTGGCTTTAGACTTGCTTTAGAATCTTGCGGTGCTG
AGTGCGTTTATTCTAATGAATGGGATAAATATGCACAAGAAGTATATGAGATGAATTTTG
GTGAAAAGCCTGAGGGCGACATTACCCAAGTAAATGAGAAAACCATTCCTGATCACGACA
TTTTAT**GcGC**AGGGTTTCCG**gGC**CAAGCATTCTCTATTTCCGGAAAACAAAAAGGATTCG
AGGACTCTA**GG**GTACCCTCTTTTTTGATATCGCACGTATTGTCCGTGAAAAAAAACCTAA
AGTGGTTTTTATGGAAAATGTGAAAAATTTTGCATCGCATGATAATGGAAATACGTTAGA
AGTTGTAAAAAATACAATGAATGAATTGGACTATTCTTTTCATGCTAAAGTATTAAATGC
TTTAGATTATGGGATTCCACAGAAAAGGGAACGTATCTATATGATTTGTTTTCGCAATGA
TCTCAATATTCAAAATTTCCAATTTCCAAAACCTTTTGAGCTTAATACTTTTGTGAAAGA
TTTGTTATTACCTGATAGCGAGGTGGAACACTTAGTTATTGATAGAAAAGATTTGGTAAT
GACAAACCAAGAAATTGAGCAAACAACCCCCAAAACAGTTCGACTTGGTATTGTAGGAAA
AGGTGGGCAAGGAGAACGAATTTATAGCACAAGAGGCATTGCAATTACCTTATCTGCTTA
TGGTGGCGGCATTTTCGCTAAGACAGGGGGATATTTAGTAAACGGGAAGACACGGAAATT
ACACCCTAGAGAGTGTGCTAGAGTAATGGGCTACCCAGATAGTTATAAAGTCCACCCGTC
AACCAGCCAAGCATATAAACAATTTGGTAACTCAGTTGTTATCAATGTACTTCAATATAT
TGCTTATAACATTGGTTCATCATTAAATTTCAAACCATAT**GTCGAC** -3′

B

(wt) SPNSHMIEIKDKQLTGLRFID**LFAGLGG**FRLALESCGAECVYSNEW
(mutant) SPNSHMIEIKDKQLTGLRFIDLFAGLGGFRLALESCGAECVYSNEW

(wt) DKYAQEVYEMNFGEKPEGDITQVNEKTIPDHDILCAGFPCQAFSIS
(mutant) DKYAQEVYEMNFGEKPEGDITQVNEKTIPDHDILCAGFP**g**QAFSIS

(wt) GKQKGFEDSRGTLFFDIARIVREKKPKVVFMENS
(mutant) GKQKGFEDSRvpSflishvLsVkKNlkwfLwkm-

The other type of deletion was much larger and easily recognised by restriction mapping, it was 560 bp shorter than the wild type gene. Two samples were sequenced both gave exactly the same sequence with the same deletion in the protein starting from T302 in the sequence encoding motif V downstream of G865 in the sequence encoding motif IX. The full nucleotide sequence of the M.*Hha*I$_{C81G}$d2 and the deduced amino acid sequence are illustrated in figure 3.16.

Figure 3.16: Nucleotide sequence and amino acid sequence deduced from M.*Hha*I$_{C81G}$d2

(a) Nucleotide sequence of M.*Hha*I$_{C81G}$d2. Motifs are coloured same as in figure 1.4, the unique *Hha*I site (GCGC), and sequences encoding Gly (GGC) in motif IV are in bold and underlined. The last base before deletion and first base after ligation are in bold green. The *Eco*RI and *Xho*I sites flanking the gene are in blue bold. Small letters show that there are differences between the wild type M.*Hha*I gene and this type of deletion. 563 bp are deleted from 303 to 865 (see figure 3.8 for sequence numbering).

(b) Amino acid sequence deduced from the sequence in (A) aligned with the wild type M.*Hha*I protein coloured the same as in figure 1.4. Unmatched amino acids are coloured by green, changes in amino acids started from amino acid S96 and down stream to the stop codon replacing L100.

A

5′ **GGATCC**CCGAATTCCCAT**ATG**ATTGAAATAAAAGATAAACAGCTCACAGGA
TTACGCTTTATTGACCTTTTTGCAGGATTAGGTGGCTTTAGACTTGCTTTAGA
ATCTTGCGGTGCTGAGTGCGTTTATTCTAATGAATGGGATAAATATGCACAAG
AAGTATATGAGATGAATTTTGGTGAAAAGCCTGAGGGCGACATTACCCAAGTA
AATGAGAAAACCATTCCTGATCACGACATTTTAT**GcGC**AGGGTTTCCGgGCCA
AGCATTCTCTATTTCCGGAAAACAAAAAGGATTCGAGGAC**TG**GCTACCCAGAT
AGTTATAAAGTCCACCCGTCAACCAGCCAAGCATATAAACAATTTGGTAACTC
AGTTGTTATCAATGTACTTCAATATATTGCTTATAACATTGGTTCATCATTAA
ATTTCAAACCATAT**GTCGAC** 3′

B

(wt) SPNSHMIEIKDKQLTGLRFID**LFAGLG**GFRLALESCGAECVYSNEW
(mutant) SPNSHMIEIKDKQLTGLRFIDLFAGLGGFRLALESCGAECVYSNEW

(wt) DKYAQEVYEMNFGEKPEGDITQVNEKTIPDHDILCAGF**PCQ**AF**S**IS
(mutant) DKYAQEVYEMNFGEKPEGDITQVNEKTIPDHDILCAGFPgQAFSIS

(wt) GKQKGFEDSRGTL
(mutant) GKQKGFEDwlpr-

3.7 SUMMARY OF CHAPTER 3 RESULTS

1- Plasmids carrying a mutation Motif IV Cys to Gly in monospecific Mtases show reduced transformation efficiency and are weakly cytotoxic to *E.coli.*

2- All recovered plasmids contained mutations suggesting that an efficient form of error-prone repair had take place.

3- Nucleotide sequence analysis revealed that the repair processes leads to DNA damage.

4- The dimerised form of mutated Mtases is required to induce DNA damage.

3.8 DISCUSSION

In this chapter it has been found that plasmids encoding monospecific Mtases with Cys substituted by Gly, in the catalytic motif reduce transformation efficiency by approximately 100 times. This reduction in transformation efficiency is however significant by less than that found when a plasmid encoding an active Mtase is transformed into an *mcrBC*$^{+}$ host where a 10^{6} fold reduction in recovered transformants would be expected. This cytotoxic effect was tested using two monospecific Mtases in this chapter: M.*Hha*I and M.*Msp*I: both gave essentially the same result. Therefore, the Cys to Gly in motif IV conveys a weak cytotoxic effect to host cells.

Further analysis of the plasmids isolated from the transformed colonies showed that all contained mutations in the Mtase gene. Therefore cells respond to mutant Mtase expression by inducing error-prone repair. The error-prone repair is induced at relatively high frequency since transformation efficiency is only reduced by 25 fold (M.*Msp*I$_{C174G}$) and 6 fold (M.*Hha*I$_{C81G}$) as shown in table 3.3.

In this study the deletion phenomena has been extended to include more than one type of Mtase namely M.*Msp*I and M.*Hha*I. Nucleotide sequence

analysis provides the most definitive means of examing the molecular products of this error-prone repair. It has been shown that the repair process can lead to deletions of between 1 and 563 bp and can also give rise to combinations of deletion and insertion which may originate from some form of recombination events (see figure 3.13, 3.15 and 3.16). However the deletions analysed were not precisely linked to Mtases recognition sites as had been observed earlier (Hurd and Hornby, unpublished data). Instead deletions appear to start from any sequence in the DNA (see figure 3.17 and appendix 4).

Table 3.3: Transformation efficiencies of plasmids encoding the mutant and control monospecific Mtases and the percentage of mutations observed.

Plasmid	**Transformation efficiency ($X10^6$)**	**Percentage of colonies mutated**
$pGH2_{C81G}K$ (control)	15	0
$pGH2_{C81G}$ (mutant)	2.5	100
pGM2d1 (control)	33	0
$pGM2_{C174G}$ (mutant)	1.2	100

Since all experiments were initially carried out with GST fusion vectors, it was important to establish whether dimerisation of the mutant Mtases had any effect on the repair processes. It has been shown clearly that deletion and rearrangement repair required the Mtases to be dimeric. When parallel experiments were carried out with monomeric M.*Msp*I only point mutations and no evidence of deletions were obtained.

All the plasmid isolated from cells transformed with GST-M.*Msp*I_{C174G} were rearranged either with deletion or insertion (see section 3.5). on the other hand, plasmids isolated from cells transformed with GST-M.*Hha*I_{C81G} 25% of the recovered colonies carried large deletions, with single base deletions in the remainder. Moreover the recovery percentage in the GST-M.*Msp*I_{C174G} experiment was 3.8% while in the GST-M.*Hha*I_{C81G} it was more than four times higher (16.6%). This is perhaps related to the presence of target recognition sites in the gene. In the GST-M.*Msp*I gene there are three sites whilst just one site present in the GST-M.*Hha*I gene. The percentage of cells recovered from in the monomer-M.*Msp*I_{C174G} encoding plasmids was 52% which is 13.6 and 3.1 fold

higher than plasmids expressing dimerised M.*Msp*I$_{C174G}$ and M.*Hha*I$_{C81G}$ respectively.

These data may be compared with experiments by the laboratories of Bhagwat (Wyszynski et al. 1991) and Roberts (Mi and Roberts, 1993). Replacement of the conserved Cys with Ser, Val or Trp abolished catalysis M.*EcoR*II (Wyszynski et al. 1991). These authors also they reported that a Gly substitution of the catalytic motif both abolished catalysis and proved cytotoxic to *E.coli*: the mutant allele could only be maintained in the cells when it is poorly expressed (Wyszynski et al. 1991). Mi and Roberts (1993) extended this observation to M.*Hha*I and concluded that the cytotoxic effect was due to high affinity recognition of target sequences in the genomic of the host by C81G mutated M.*Hha*I (see section 6.4). Recently in this laboratory it was found that this observation is also true for M.*Msp*I, but colonies recovered following transformation of *E.coli* RRI all contained precise deletion of 30 bp between the two *Msp*I sites within the M.*Msp*I gene.

One of the most interesting results to emerge from the M.*Hha*I$_{C81G}$ experiment here is that both the two types of deletions found start from the same region in the M.*Hha*I gene. The deletion started 60 bp or 56 bp downstream of the unique *Hha*I site in the gene for M.*Hha*I$_{C81G}$d1 and M.*Hha*I$_{C81G}$d2 respectively (see section 3.3.6.1). As shown in figure 3.17, M.*Hha*I$_{C81G}$d1 started from the G98 codon and M.*Hha*I$_{C81G}$d2 started from the S96 codon. This result suggests that deletions have arisen by a similar mechanism. The proposed mechanism of deletion will be discussed in chapter 7.

There are at least two factors that could determine the toxicity level of a mutant Mtase gene: the size of encoded protein, and the number of target recognition sites present in the Mtase gene. Using dimerised forms of mutated Mtases will double the molecule size of the mutated protein and nucleoprotein complex. On the other hand increasing the number of recognition sites will increase the density of protein bound to the DNA and subsequently increase the lethal effect of the protein.

As expected, the standard deviation (sd) was high when measuring the number of colonies able to survive (and the corresponding controls). This is due to

the wide variety of factors that influence transformation, including the number of viable cells in the competent cell preparation, the small differences in the concentration of DNA used to transform the cells, and handling during the experiments. DNA was prepared for each set of experiments and the same concentration was used to transform the *E.coli* and the one-bp shift mutation provided the best control for studying the effect of M.*Msp*I$_{C174G}$. In spite of attempts to minimise these factors e.g. by preparing large batches of competent cells the variety of transformation results could not be reduced.

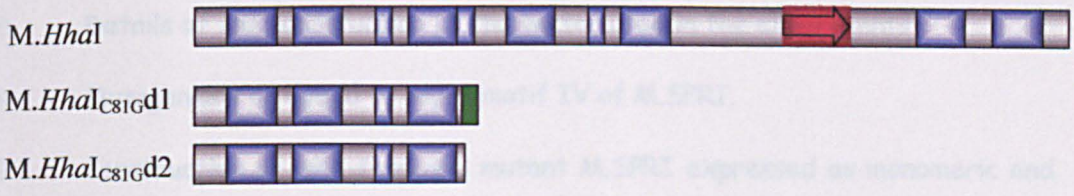
Figure 3.17: protein deduced from deletions in M.*Hha*I$_{C81G}$.

a). Protein sequences deduced from M.*Hha*I$_{C81G}$d1 and M.*Hha*I$_{C81G}$d2 the upper lane is the wild type M.*Hha*I and the end of motif IV is shown in yellow and the sequence of protein after the deletion in green. B) diagram shows the motifs in the two deletions compared with the wild type.

a)

M.*Hha*I	GKQKGFEDSRGTLFFDIARIVREKKPKVVFMENS
M.*Hha*I$_{C81G}$d1	GKQKGFEDSRVPSFLISHVLSVKKNLKWFLWKMsotp
M.*Hha*I$_{C81G}$d2	GKQKGFEDWLPRstop

b)



CHAPTER 4:

THE EFFECT OF THE MOTIF IV CYS TO GLY MUTATION IN THE MULTISPECIFIC MTASE M.SPRI, ON E.COLI

SYNOPSIS OF CHAPTER FOUR

- A brief review of Multispecific Mtases.
- A brief review of M.SPRI Mtase.
- Details of the construction of the vector used in the experiments
- Mutagenesis of Cys78 to Gly in motif IV of M.SPRI.
- Construction of wild type and mutant M.SPRI expressed as monomeric and dimeric forms.
- Methodology of deletion detection.
- Classification of deletions found in the M.SPRI$_{C78G}$ gene.
- Nucleotide sequencing of the deletions found in the M.SPRI$_{C78G}$ gene.
- Discussion of results obtained from Chapter Four

4.1 INTRODUCTION

So far it has been found that dimeric forms of M.*Msp*I$_{C174G}$ and M.*Hha*I$_{C81G}$, provoke deletions *in vivo* in *E.coli.* It is therefore of interest to establish how a multispecific Mtase such as M.SPRI, which should increase the frequency of protein DNA encounter, affects *E.coli.* Furthermore, can the multispecific recognition property of this type of Mtase be used to study the specificity of the deletion phenomenon?

4.2 MULTISPECIFIC DNA MTASES

The most commonly observed function of restriction and modification systems is to protect a host against foreign DNA (see section 1.2.1). The concept that the primary role of restriction is to prevent phage infection is underlined by the finding that many bacteriophages have gained the potential to infect cells with R/M systems by sophisticated mechanisms overcoming host restriction(Krüger and Bickle, 1983).

'Self methylation' by phage-encoded Mtase is one of the mechanisms by which bacteriophage protect themselves against restriction. A number of virulent and temperate bacteriophage of a variety of phylogenetically different hosts encode solitary type II Mtases which recognise the same target as that of the RM system of the host which they infect. SPR is a *Bacillus subtilis* phage, which encodes a "multispecific" Mtase, having the capacity to methylate several different DNA target sequences (Buhk et al. 1984). One of these multiple specificities may be identical to the RM system present in the bacterial host lysogenised by the bacteriophage.

The DNAs of *B. subtilis* bacteriophages M.SPR, SPβ, ϕ3T, and ρ11 are resistant to degradation by certain restriction enzymes (Buhk et al. 1984). This resistance was shown to be due to a specific modification of the DNA during phage growth which is mediated through the action of phage-encoded DNA Mtases. The genes coding for such enzymes have been identified genetically in all phages and have subsequently been cloned in *E.coli* and *B. subtilis* vectors (Noyer-Weidner et al. 1985). When expressed in both organisms, they lead to

specific methylation of DNA. In this study we shall focus on M.SPRI as an example of a multispecific Mtase.

4.3 THE MULTISPECIFIC MTASE M.SPRI

M.SPRI is a type II multispecific C5-Mtase enzyme that catalyses the transfer of a methyl group from SAM to the C-5 position of the underlined cytosine within the nucleotide sequences GGCC (*Hae*III), CCGG (*Msp*I) and CCA/TGG (*EcoR*II) (Behrens et al. 1983; Tran-Betcke et al. 1986; Trautner et al. 1980). The gene encoding the M.SPRI gene was first cloned by Behrens et al. (1983) from a genomic library of *Bacillus subtilis* phage DNA.

The M.SPRI gene encodes a 50 kDa protein (Buhk et al. 1984) (see figure 4.2), and shows the typical primary structure of a C-5 Mtase (Kumar et al. 1994; Lauster et al. 1989; Posfai et al. 1989; Wilson and Murray, 1991b) (see section 1.3.2 and figure 1.5). M.SPRI like other multispecific Mtases differs from the monospecific C-5 Mtases in that it has more than one TRD: the TRDs of multispecific Mtases are structurally and functionally independent (Walter et al. 1992).

4.3.1 SUBCLONING THE M.SPRI GENE

M.*SPR*I was subcloned into pGEX-KG vector by introducing new cloning sites (*BamH*I and *Xho*I) via PCR. Two primers used were AWSPR1 and AWSPR2 (see appendix 3 for sequence). The PCR product was cloned in the pGEX-KG polylinker between *BamH*I and *Xho*I to produce the 6277 bps construct (see figure 4.1). In this thesis pGEX-KG M.SPRI expressing a wild type version of M.SPRI is named **pGS2.**

4.3.2 THE NUCLEOTIDE SEQUENCE OF THE M.SPRI GENE

Before initiating experiments, the complete nucleotide sequence of M.SPRI in pGEX-KG was determined (see figure 4.2). Three primers were used ABF 20, 25, and 27 (see section 4.4.2 and table 4.1 for the sequences of these oligonucleotides).

Figure 4.1: pGS2 map.
Restriction map of pGS2, this vector contains the multispecific Mtase M.*SPR*I cloned into pGEX-KG downstream of the GST. All the unique resteriction sites used in the cloning and analysing the gene are in blue, the three sites recognised by M.*SPR*I are illustrated, H for *Hae*III, M for *Msp*I, and E for *Eco*RII.

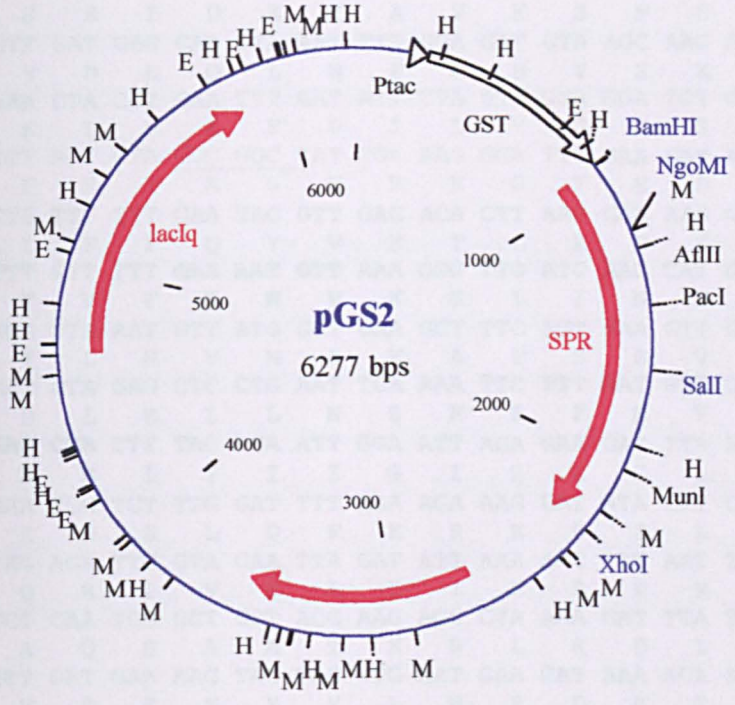


Figure 4.2: DNA sequence and deduced amino acids sequence of M.*SPR*I gene

The invariably conserved amino acids present in the six most highly conserved motifs are coloured the same as in figure (1.4), the highly conserved amino acids among those motifs are in bold (see figure 1.5). Numbering is on the left, black for DNA sequence and red between brackets for amino acids. The targeted C78 in motif IV is in bold and underlined. The unique restriction sites used in the experiments *Ngo*MI, *Sal*I, *Sph*I, and *Mun*I are underlined.

```
   1        TTG GGT AAA CTA CGT GTA ATG AGT CTT TTT AGT GGG ATC GGT GGA
 (1)          L   G   K   L   R   V   M   S   L   F   S   G   I   G   G
  46  TTT GAA GCT GCA CTA AGA AAC ATT GGG GTT GGT TAT GAG CTG GTT GGT
(16)   F   E   A   A   L   R   N   I   G   V   G   Y   E   L   V   G
  94  TTT AGT GAG ATT GAT AAA TAT GCN GTC AAA TCT TTT TGT GCA ATT CAC
(32)   F   S   E   I   D   K   Y   A   V   K   S   F   C   A   I   H
 142  AAC GTT GAT GAG CAA TTA AAT TTT GGA GAT GTA AGC AAG ATT GAT AAG
(48)   N   V   D   E   Q   L   N   F   G   D   V   S   K   I   D   K
 190  AAA AAA CTA CCT GAA TTT GAT ATT TTA GTT GGA GGA TCT CCT TGT CAA
(64)   K   K   L   P   E   F   D   I   L   V   G   G   S   P   C   Q
 238  AGC TTT AGT GTA GCC GGC CAT CGA AAG GGA TTT GAA GAT ACA AGA GGT
(80)   S   F   S   V   A   G   H   R   K   G   F   E   D   T   R   G
 286  ACC TTG TTT TTT CAA TAC GTT GAG ACA CTT AAG GAA AAG CAA CCA AAG
(96)   T   L   F   F   Q   Y   V   E   T   L   K   E   K   Q   P   K
 334  TTT TTT GTT TTT GAA AAT GTT AAA GGG TTG ATC AAC CAT GAT AAA GGA
(112)  F   F   V   F   E   N   V   K   G   L   I   N   H   D   K   G
 382  AAT ACA TTA AAT GTT ATG GCT GAA GCT TTC AGT GAA GTT GGG TAC AGA
(128)  N   T   L   N   V   M   A   E   A   F   S   E   V   G   Y   R
 430  ATT GAT CTA GAG CTC CTG AAT TCA AAA TTC TTT AAT GTT CCA CAA AAT
(144)  I   D   L   E   L   L   N   S   K   F   F   N   V   P   Q   N
 478  AGG GAG CGA CTT TAC ATA ATT GGA ATT AGA GAA GAT TTA ATT AAA AAT
(160)  R   E   R   L   Y   I   I   G   I   R   E   D   L   I   K   N
 526  GAA GAA TGG TCT TTG GAT TTT AAA AGA AAG GAT ATA CTT CAA AAA GGG
(176)  E   E   W   S   L   D   F   K   R   K   D   I   L   Q   K   G
 574  AAA CAG AGA TTG GTA GAA TTA GAT ATT AAA AGC TTT AAT TTT AGA TGG
(192)  K   Q   R   L   V   E   L   D   I   K   S   F   N   F   R   W
 622  ACA GCT CAA TCG GCT GCT ACG AAG AGG CTA AAA GAT TTA TTA GAA GAA
(208)  T   A   Q   S   A   A   T   K   R   L   K   D   L   L   E   E
 670  TAC GTT GAT GAA AAG TAC TAC TTG AAT GAA GAT AAA ACA AAC AGT TTG
(224)  Y   V   D   E   K   Y   Y   L   N   E   D   K   T   N   S   L
 718  ATC AAA GAG TTG TCT ACA AGT CGA CTT AAT GAA AAT CTT ACT GTT GAG
(240)  I   K   E   L   S   T   S   R   L   N   E   N   L   T   V   E
 766  CAA GTA GGT AAC ATT AAT CCC TCT GGT AAT GGA ATG AAT GGA AAT GTT
(256)  Q   V   G   N   I   N   P   S   G   N   G   M   N   G   N   V
 814  TAT AAT TCA TCT GGA TTA AGC CCC ACA ATT ACC ACT AAT AAA GGA GAG
(272)  Y   N   S   S   G   L   S   P   T   I   T   T   N   K   G   E
 862  GGA CTG AAA ATT GCA GTT GAG TAC TCC AGA AAA AGC GGG CTT GGA CGA
(288)  G   L   K   I   A   V   E   Y   S   R   K   S   G   L   G   R
 910  GAA CTT GCT GTA TCT CAT ACG CTT TCT GCT TCT GAC TGG AGA GGA TTG
(304)  E   L   A   V   S   H   T   L   S   A   S   D   W   R   G   L
 958  AAT AGG AAC CAA AAA CAA AAT GCA GTT GTT GAG GTA AGG CCA GTA TTA
(320)  N   R   N   Q   K   Q   N   A   V   V   E   V   R   P   V   L
1006  ACC CCA GAA AGG GGG GAG AAG CGA CAA AAT GGA AGA AGA TTT AAA GAT
(336)  T   P   E   R   G   E   K   R   Q   N   G   R   R   F   K   D
1054  GAC GGT GAA CCA GCA TTT ACA GTA AAC ACA ATT GAC AGA CAC GGG GTA
(352)  D   G   E   P   A   F   T   V   N   T   I   D   R   H   G   V
1102  GCG GTT GGA GAG TAT CCA AAA TAC AGA ATT AGA AGA TTA ACA CCG TTA
(368)  A   V   G   E   Y   P   K   Y   R   I   R   R   L   T   P   L
1150  GAG TGC TTT AGG CTA CAG GCT TTT GAT GAC GAA GAT TTT GAA AAA GCT
(384)  E   C   F   R   L   Q   A   F   D   D   E   D   F   E   K   A
1198  TTT GCT GCG GGA ATA AGT AAC TCA CAA TTA TAT AAG CAA GCC GGT AAT
(400)  F   A   A   G   I   S   N   S   Q   L   Y   K   Q   A   G   N
1246  TCA ATT ACT GTA ACT GTG CTT GAG TCA ATA TTC AAG GAA TTA ATA CAT
(416)  S   I   T   V   T   V   L   E   S   I   F   K   E   L   I   H
1294  ACA TAC GTT AAT AAA GAA TCT GAA TAA CTC GAC GCG TCT GCA GAA GCT
(432)  T   Y   V   N   K   E   S   E STOP L   D   A   S   A   E   A
1342  TCT AGA ATT CGA GCT CCC GGG TAC CAT GGC ATG CAT CGA ATC TCG AGC
(448)  S   R   I   R   A   P   G   Y   H   G   M   H   R   I   S   S
```

4.4 EXPERIMENTAL DESIGN

In order to establish whether M.SPRI$_{C78G}$ provoked deletions in *E.coli,* the experimental design was essentially the same as the previous experiments (section 3.4), using the monospecific enzymes. First the M.SPRI gene was inactivated using a kanamycin resistance gene, then the highly conserved sequence in motif IV PCQ was mutated to PGQ (see figure 4.2). Reactivation was achieved by removing the kanamycin resistance gene, after that *E.coli* cells of various genotype were transformed. The experiments were carried out using monomeric and dimerised-M.SPRI$_{C78G}$.

4.4.1 INACTIVATION OF M.SPRI BY THE KANAMYCIN^R GENE

The first step in this experiment was inactivation of M.SPRI by insertion of the kanamycin resistance gene. There are several advantages to be gained by using this method. First selection of the preferred colonies using media with ampicillin and kanamycin antibiotics is facilitated, second inactivation of the M.SPRI gene is effected, and thirdly there is no conflict with the unique restriction site in the M.SPRI gene used to insert and excise the kanamycin^r gene cassette.

There are four unique restriction sites in the M.SPRI gene suitable to be used as kanamycin gene cloning sites *Afl*II (CTTAAG), *Pac*I (TTAATTAA), *Sal*I (GTCGAC), and *Mun*I (CAATTG). The *Sal*I site (in the region encoding *EcoR*II TRD) was chosen because it is also present in the polylinker flanking the kanamycin^r gene in pUC4K.

The kanamycin^r gene was removed from pUC4K (Pharmacia) by digestion with *Sal*I. After purification of the 1252 bp DNA fragment from an agarose gel it was ligated into pGS2 predigested with *Sal*I. Ligated DNA was used to transform *E.coli* AB1157 cells, which were then plated onto LA plates containing kanamycin and ampicillin. The advantage of using *E.coli* AB1157 cells in the transformations is that they afford both kanamycin selection and the *mcr*$^+$ genotype of the cells will lead to restriction of DNA which is C5 methylated. Possible recombinants were picked, grown in LB (supplemented with ampicillin and kanamycin) and plasmid DNA was isolated using the 'mini-prep' procedure

(see section 2.2.2.1). The plasmid DNA was analysed by digestion with *Sal*I to check for the presence of the kanamycin^r gene, and *Msp*I to check for the activity of the plasmid encoding M.SPRI (see figure 4.3) The final plasmid DNA was named **pGS2K**.

4.4.2 MUTAGENESIS OF C78 TO G IN M.SPRI

Site directed mutagenesis was employed to mutate C78G in motif IV, using two primers ABF19 and ABF20 (the sequence and restriction sites are shown below). The ABF19 oligodeoxynucleotide contains a *BamH*I site and a new *Msp*I site by changing CTA CGT to CTC CGG (see Figure 4.2), both are silent mutations. The second oligodeoxynucleotide is ABF20, the upstream primer, contains an *NgoM*I site (5'-GCCGGC-3') and mutates the highly conserved amino acid in motif IV PCQ to PGQ by changing TGT (Cys) to GGC (Gly).

ABF 19 (forward primer 31 bp):

*BamH*I *Msp*I

5'-GCGC GGA TCC TTG GGT AAA CTC CGG GTA ATG-3'

ABF 20 (reverse primer 39 bp):

*NgoM*I C79G

5'-GCGC GCC GGC TAC TCT AAA GCT TTG GCC AGG AGA TCC TC-3'

In the PCR reaction pGS2 was used as the template for amplification of the mutated PCR fragment, ABF19 and ABF20 were used as downstream and upstream primers respectively and the PCR reaction conditions were as described in section 2.2.5. The PCR fragment was analysed by agarose gel electrophoresis as shown in figure 4.4b. pGS2K and the PCR fragment were digested with *BamH*I and *NgoM*I sequentially as described in section 2.2.4.1.2 and the required bands (7340bp for the vector 255 bp for the PCR fragment) were purified from an agarose gel (see figure 4.4 b and c).

The digested PCR product and the vector (digested pGS2K) were ligated and used to transform *E.coli* AB1157 cells. Possible recombinants were picked, grown in LB media (with ampicillin and kanamycin) and plasmid DNA was

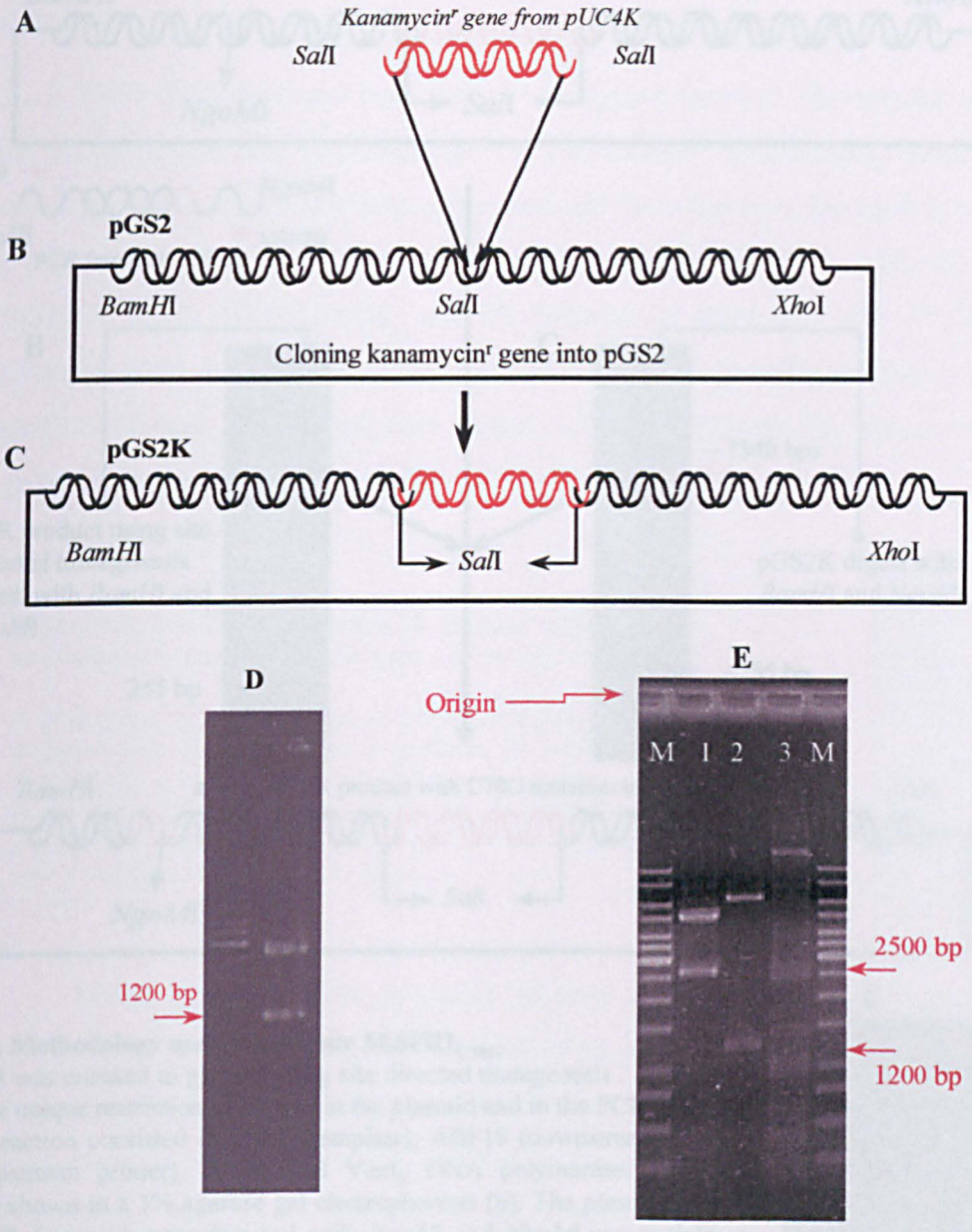
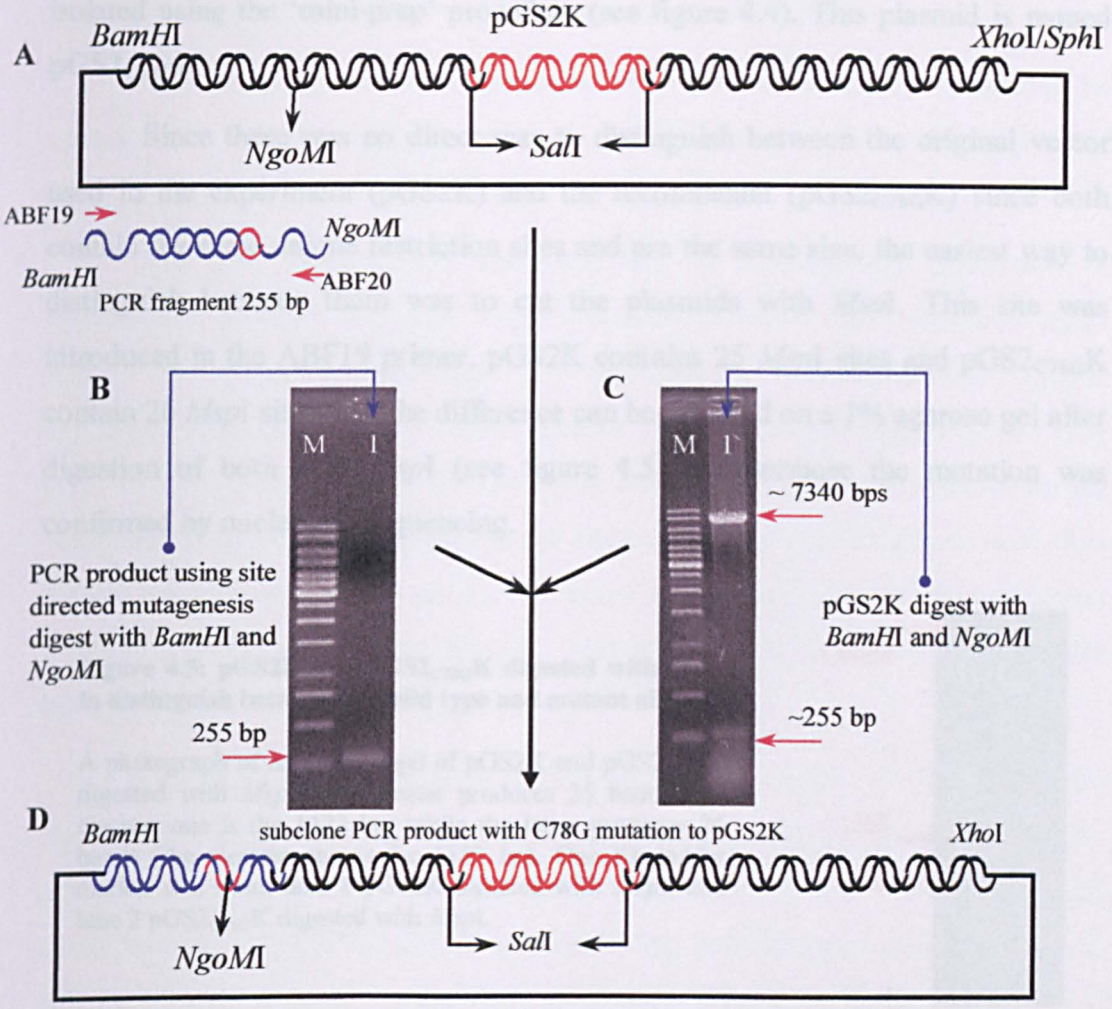


Figure 4.3: inactivation of M.SPRI by kanamycin resistance gene.
The kanamycin^r gene was removed from pUC4K plasmid (Pharmacia) by digestion with *SalI* followed by purification of the 1252 bp DNA fragment from an agarose gel (a and d). pGS2 was also digested with *SalI* followed by deproteinisation and precipitation of the 6277 bp DNA. The purified kanamycin^r gene was ligated with the digested pGS2 (b). The ligated DNA was used to transform *E.coli* AB1157 cells (c). In (e) the recombinant DNA was digested with *BamHI* and *XhoI* in lane 1, *SalI* in lane 2, and *MspI* in lane 3 to check the activity.



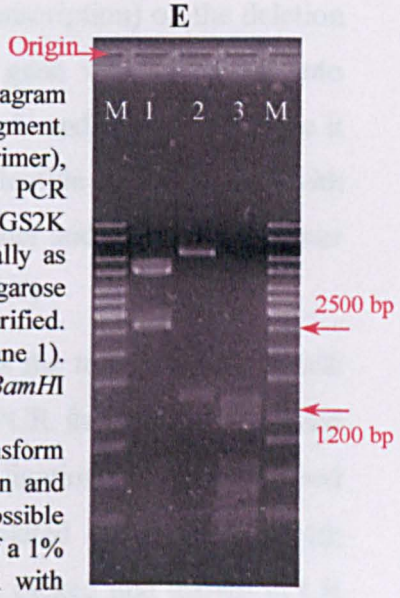


Figure 4.4: Methodology used to generate M.SPRI$_{C78G}$.
Cysteine 78 was mutated to glycine using site directed mutagenesis . A) diagram showing the unique restriction sites used in the plasmid and in the PCR fragment. The PCR reaction consisted of pGS2 (template), ABF19 (downstream primer), ABF20 (upstream primer), dNTP, and Vent$_R$ DNA polymerase. The PCR fragment is shown in a 1% agarose gel electrophoresis (b). The plasmid pGS2K and the PCR fragment were digested with *BamH*I and *NgoM*I sequentially as described in section 2.2.5.1.2. DNA was then electrophoresed on a 1% agarose gel and the vector band (7340 bp) and the PCR fragment (255 bp) were purified. B) photograph of a 1% agarose gel showing the 255 bp PCR fragment (lane 1). C) photograph of a 1% agarose gel showing the digested pGS2K with *BamH*I and *NgoM*I the two fragments are illustrated 7340 and 255 bp.
The digested pGS2K and PCR product were ligated and then used to transform *E.coli* AB1157 competent cells and plated on LA plates with ampicillin and kanamycin. D) a diagram illustrating the pGS2$_{C78G}$K plasmid. Possible recombinants were picked and were grown on LB media. E) photograph of a 1% agarose gel illustrating the digestion of the final construct pGS2$_{C78G}$K with *BamH*I and *Sph*I (lane 1) to show the size of the M.SPRI and kanamycin^r genes (2631), *Sal*I (lane 2) to show the presence of kanamycin^r gene, and *Msp*I (lane 3) to check the activity of M.SPRI gene. The one kb ladder marker was used in all the agarose gels see figure 2.2 for band size.

isolated using the 'mini-prep' procedure (see figure 4.4). This plasmid is named **pGS2$_{C78G}$K**

Since there was no direct way to distinguish between the original vector used in the experiment (pGS2K) and the recombinant (pGS2$_{C78G}$K) since both contain the same unique restriction sites and are the same size, the easiest way to distinguish between them was to cut the plasmids with *Msp*I. This site was introduced in the ABF19 primer. pGS2K contains 25 *Msp*I sites and pGS2$_{C78G}$K contain 26 *Msp*I sites, and the difference can be detected on a 1% agarose gel after digestion of both with *Msp*I (see figure 4.5). Furthermore the mutation was confirmed by nucleotide sequencing.

Figure 4.5: pGS2K and pGS2$_{C78G}$K digested with *Msp*I to distinguish between the wild type and mutant alleles.

A photograph of an agarose gel of pGS2K and pGS2$_{C78G}$K digested with *Msp*I. The former produces 25 bands (the clearest one is the 1422 bp) while the latter produces 26 bands (the clearest one is the 1183 bp). One Kb ladder marker was used, lane 1 pGS2K digested with *Msp*I, and lane 2 pGS2$_{C78G}$K digested with *Msp*I.

To test the effect of the promoter (or level of transcription) on the deletion activity of mutated M.SPRI, the GST-M.SPRI$_{C78G}$K gene was subcloned into pLitmus 28. As there is no unique restriction site at the 5` end of the GST gene it was necessarily to introduce one. ABF28 an oligonucleotide was designed with two restriction sites *EcoR*I and *Sac*I in the forward primer and the reverse primer was ABF27 (see appendix 3 for primers list and figure 4.7)

A PCR reaction was setup, using pGS2$_{C78G}$K as the template and Vent® DNA polymerase as described in section 2.2.5. The PCR fragment was cloned into pLitmus 28 using *EcoR*V as a unique site, and the ligation reaction was used to transform *E.coli* AB1157 cells which were then plated on LA plates with ampicillin and kanamycin. Possible recombinants were picked and grown in LB with ampicillin and kanamycin and DNA was isolated and analysed with appropriate restriction enzymes (see figure 4.7). This plasmid was named **pLS2$_{C78G}$K**

4.4.3 GENERATION OF MONOMERIC M.SPRI

The effect of oligomeric status on the deletion phenomenon induced by multispecific Mtases was investigated by cloning M.SPRI wild type and M.SPR$_{C78G}$ genes into pET22b vector.

The pET System was developed for the cloning and expression of recombinant proteins in *E.coli*. Target genes are cloned in pET plasmids under control of strong bacteriophage T7 transcription and translation signals; expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so active that almost all of the cell's resources are converted to target gene expression; the desired product can constitute more than 50% of the total cell protein after a few hours of induction. Studier and colleagues (1986) originally constructed the pET vectors, and the newer pET derivatives have been developed at Novagen.

The M.SPRI gene from pGS2 and pGS2$_{C78G}$K were subcloned into the *BamH*I and *Xho*I sites in the pET22b polylinker (see figure 4.1 and 4.6). pGS2$_{C78G}$K and pET22b were digested with *BamH*I and *Xho*I, the band corresponding to M.*SPR*I$_{C78G}$K genes (2643 bp) and the band corresponding to pET22b (5453 bp) were excised from an agarose gel and purified as described in section 2.2.2.2 and then used in ligations. The ligated DNA was used to transform *E.coli* AB1157 competent cells. Possible recombinants were picked, grown in LB media (with ampicillin and kanamycin) and plasmid DNA was isolated using the 'mini-prep' procedure and checked by digestion with *BamH*I and *Xho*I (see Figure 4.6). This plasmid was named **pES1$_{C78G}$K.**

In order to clone the monomeric wild type M.SPRI into pET22b, pES1$_{C78G}$K and pGS2 were digested with *BamH*I and *Xho*I (see Figure 4.1) and the pET22b vector and M.SPRI gene were cut and purified from an agarose gel. DNA was ligated as described in section 2.2.4.3 and then the ligation mixture was used to transform *E.coli* RRI. Possible recombinant colonies were picked, grown in LB media (supplied with ampicillin) and plasmid DNA was isolated using the 'mini-prep' procedure (see Figure 4.6 g and h) and checked by digestion with *BamH*I and *Xho*I. This plasmid was named **pES1.**

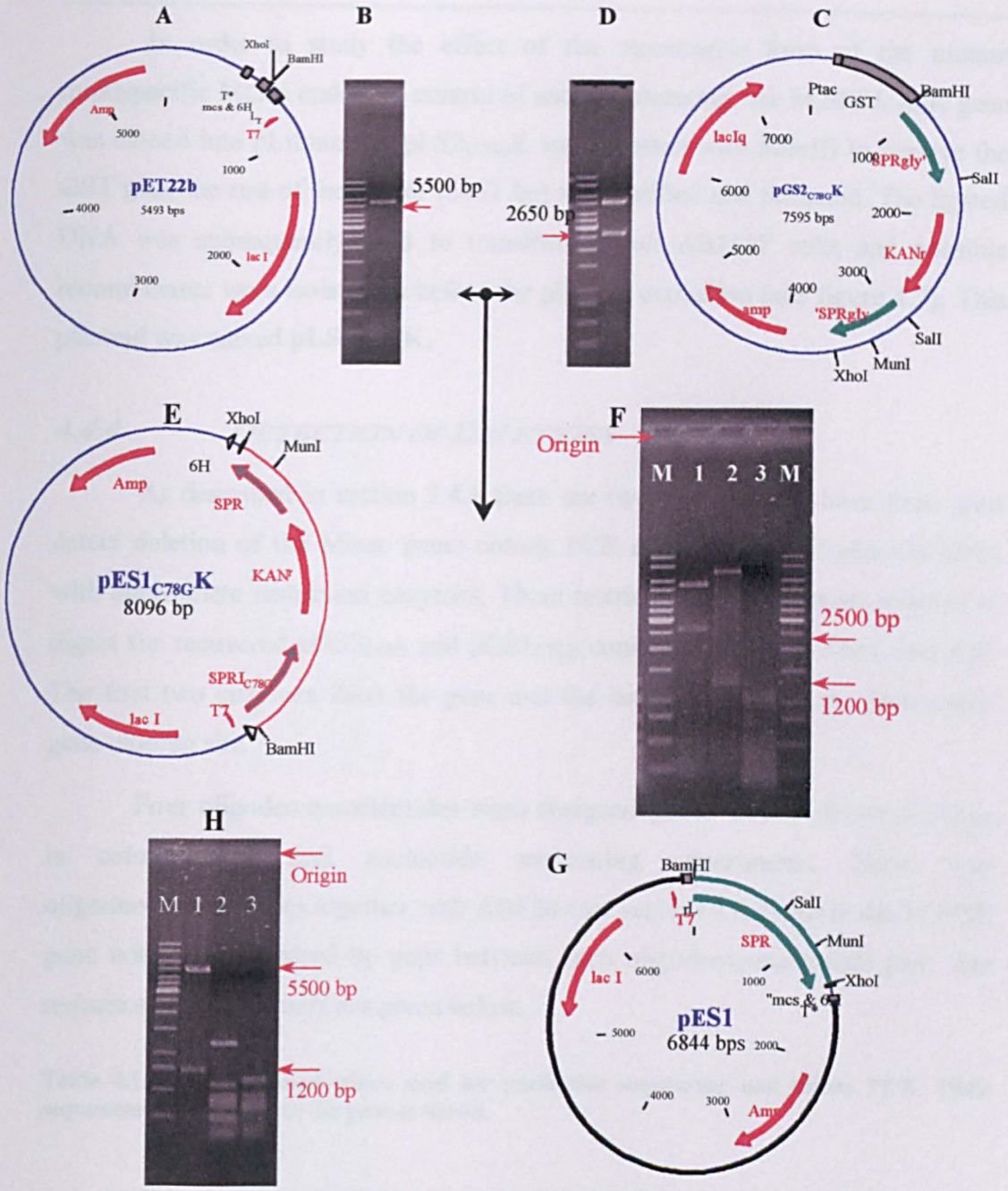


Figure 4.6: Cloning of M.SPRI$_{C78G}$K and M.SPRI into pET22b.
pES1$_{C78G}$K was cloned by digestion of pGS2$_{C78G}$K and pET22b with *BamH*I and *Xho*I. After running a 1% agarose gel, the bands corresponding to the M.SPRI$_{C78G}$K genes (2643 bp) and the band corresponding to pET22b (5453 bp) were cut and purified. A) pET22b, b) photograph of the agarose gel showing digested pET22b, C) pGS2$_{C78G}$K and D) photograph of the agarose gel showing digested pGS2$_{C78G}$K. The purified DNA was ligated and used to transform *E.coli* AB1157. Recombinant colonies were picked, grown in LB media (with ampicillin and kanamycin) and plasmid DNA was isolated using the 'miniprep' procedure and checked by digestion with appropriate restriction enzymes. E) pES1$_{C78G}$K, and F) photograph of an agarose gel showing pES1$_{C78G}$K digested with *BamH*I and *Xho*I (lane 1) to check for the presence of the right gene, *Sal*I (lane 2) to check the kanamycin^r gene , and *Msp*I (lane 3) to check the activity of M.SPRI.

In order to clone wild type M.SPRI in pET22b, pES1$_{C78G}$K and pGS2 were digested by *BamH*I and *Xho*I after a 1% agarose gel and DNA purification the M.SPRI and pET22b fragments were ligated and used to transform RRI. G) pES1, and H) photograph of an agarose gel showing the digestion of the final plasmid with *BamH*I and *Xho*I (lane 1), *Msp*I (lane 2), and *Hae*III (lane 3). The one Kb ladder was used in all the gels as a size marker (see figure 2.2).

In order to study the effect of the monomeric form of the mutant multispecific Mtase under the control of another promoter, the M.SPRI$_{C78G}$K gene was cloned into pLitmus 28. pLS2$_{C78G}$K was digested with *BamH*I to remove the GST gene the rest of the vector (5461 bp) was purified and religated. The ligated DNA was subsequently used to transform *E.coli* AB1157 cells and possible recombinants were isolated as before for plasmid extraction (see figure 4.7). This plasmid was named **pLS1$_{C78G}$K.**

4.4.4 DETECTION OF DELETIONS

As described in section 3.4.3 there are two methods that have been used detect deletion of the Mtase gene: colony PCR and digestion of plasmid DNA with appropriate restriction enzymes. Three restriction enzymes were selected to digest the recovered pGS2$_{C78G}$ and pGS1$_{C78G}$ constructs: *BamH*I, *Xho*I, and *Sal*I. The first two enzymes flank the gene and the last one provides the kanamycin^r gene cloning site.

Four oligodeoxynucleotides were designed (table 4.1) to detect deletions in colony PCR and nucleotide sequencing experiments. These four oligodeoxynucleotides together with ABF20 (see section 4.4.2) cover the M.SPRI gene with three hundred bp gaps between each oligodeoxynucleotide pair. The sequences of the primers are given below.

Table 4.1: Oligodeoxynucleotides used for nucleotide sequencing and colony PCR. Their sequences and position in the gene is shown.

Name	Sequence 5' → 3'	Position in the gene
ABF24	GCGC TTTTTGAAGTATATCCTTTC	567
ABF25	GCGC TTTCAGTCCCTCTCCTTTAT	866
ABF26	GCGC AGCCTGTAGCCTAAAGCACT	1169
ABF27	GCGC GCCCCGACACCCGCCAACAC	170 bp downstream the gene

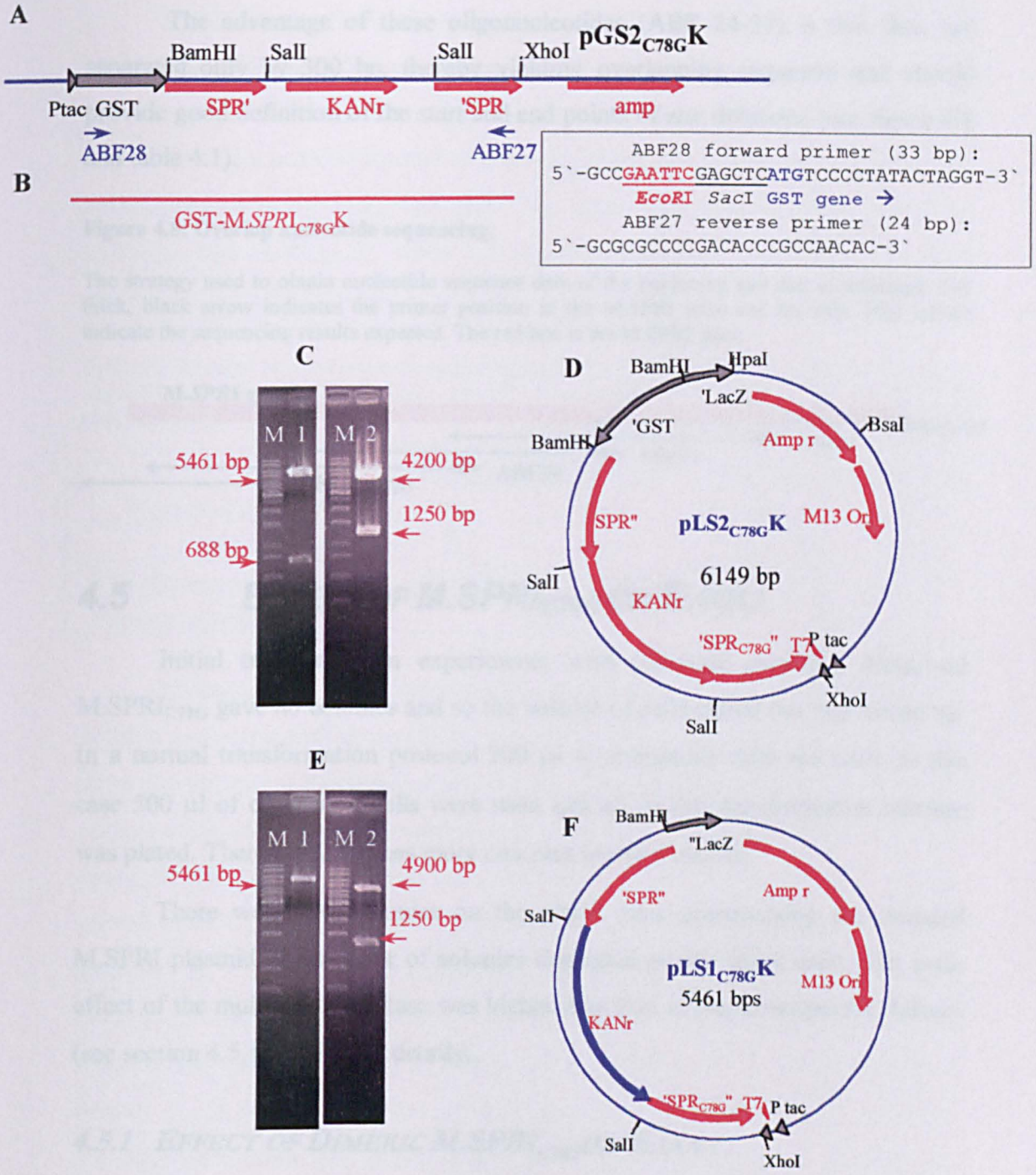


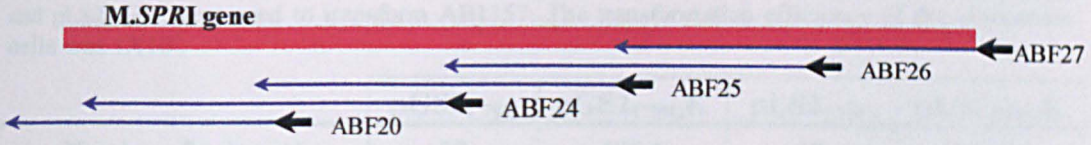
Figure 4.7: Constructing of $pLS2_{C78G}K$ and $pLS1_{C78}K$.

$pGS2_{C78G}K$ was used to amplify the GST and $M.SPRI_{C78G}K$ genes using ABF28 and ABF27 as forward and reverse primers respectively (b) sequences of the primers are shown on the box opposite (b). The PCR product was cloned into predigested pLitmus 28 cut with *EcoR*V and was subsequently used to transform AB1157s. C) DNA was isolated using "miniprep" procedure and digested with *BamH*I (lane 1) to remove GST gene and *Sal*I (lane 2) to check the presence of kanamycin^r gene. D) $pLS2_{C78G}K$. The 5461 bp band in (C) lane 1 was cut, purified, and religated and used to transform AB1157. DNA was isolated using "miniprep" procedure. E) $pLS1_{C78G}K$ was digested with *BamH*I (lane 1) and *Sal*I (lane 2) to check the presence of kanamycin gene. F) $pLS1_{C78G}K$.

The advantage of these oligonucleotides (ABF 24-27) is that they are separated only by 300 bp, thereby yielding overlapping sequence and should provide good definition of the start and end points of any deletions (see figure 4.8 and table 4.1).

Figure 4.8: Overlap nucleotide sequencing.

The strategy used to obtain nucleotide sequence data of the beginning and end of deletions. The thick, black arrow indicates the primer position in the M.SPRI gene and the thin, blue arrows indicate the sequencing results expected. The red box is the M.SPRI gene.



4.5 EFFECT OF M.SPRI$_{C78G}$ ON E.COLI

Initial transformation experiments with plasmids encoding dimerised M.SPRI$_{C78G}$ gave no colonies and so the volume of cells plated out was scaled up. In a normal transformation protocol 200 µl of competent cells are used, in this case 500 µl of competent cells were used and all of the transformation mixture was plated. Therefore 2.5 times more colonies would expected.

There were few colonies on the plates after transforming the mutated M.SPRI plasmid, the number of colonies depended on the strain used. The toxic effect of the multispecific Mtase was higher than that of the monospecific Mtases (see section 4.5.1 and 7.1 for details).

4.5.1 EFFECT OF DIMERIC M.SPRI$_{C78G}$ON E.COLI

pGS2$_{C78G}$K and pGS2K were digested with *Sal*I, then the pGS2$_{C78G}$ and pGS2 vectors were purified from an agarose gel and subsequently used for ligation. pGS2$_{C78G}$ and pGS2 DNA were used to transform *E.coli* AB1157 cells. The transformation mixtures were plated onto square plates (12 cm). A small volume of the transformation mixture (2.5 µl) was plated on plates prepared with both ampicillin and kanamycin to control for undigested (or religated) vectors containing the kanamycin^r gene.

AB1157 strain is *mcr*$^+$ and therefore pGS2 can not be used as a positive transformation control, however it can be used as a negative control because all of the colonies transformed with pGS2 should not grow. On the other hand pGS2K can be used as a positive control by removing the kanamycin^r gene, religating and transforming it to the competent cells. The same experiment was repeated using $pLS2_{C78G}$ and $pLS2_{C78G}K$. Table 4.2 shows the number of colonies that were obtained in each experiment.

Table 4.2: Number of colonies obtained in the experiment in which $pGS2_{C79G}$, $pGS2_{C79G}K$, $pLS2_{C78G}$, and $pLS2_{C78G}K$ were used to transform AB1157. The transformation efficiency of the competent cells was $1X10^6$.

	$pGS2_{C79G}$	$pGS2_{C79G}K$	$pLS2_{C78G}$	$pLS2_{C78G}K$
Number of colonies	10	456.6	67	3466.6
Sd	2	30.1	12	550.7
Transformation efficiency	$2 X 10^2$	$1 X 10^4$	$1.3 X 10^3$	$7 X 10^4$
Percentage cells surviving	2.19		1.9	

Colonies were picked after transforming the $pGS2_{C78G}$ and plasmids were isolated and subjected to restriction mapping, PCR, and nucleotide sequencing. Since the pattern of deletions obtained with both monospecific and multispecific Mtases apparently need not be directly related to their recognition specificity, two strains of AB1157 were used as host. The first is a *recA* mutant (JC2926) and the second *umuDC* mutant (GW8017). In the following experiments, the analysis of deletions is restricted to a comparison of the pattern of deletion that arise following transformation of these two mutant strains with $pGS2_{C78G}$. A more detailed analysis of the role of repair genes in deletion formation appears in chapter 5. In total, five types of deletions were isolated together with one rearrangement when the above experiment was repeated using JC2926 and GW8017. These can be classified into three types: a small deletion (1-50 bp), a large deletion (51-1000 bp), and a rearrangement. The previous experiment was repeated using isogenic strains of AB1157 to determine whether any DNA repair genes are involved in this phenomenon (see chapter 5).

4.5.1.1 SMALL DELETIONS

When plasmid DNA from the above experiment was characterised by restriction analysis with *BamH*I, *Sal*I, and *Xho*I, two patterns of small deletion

were found. At first one of the small deletions looked like a wild type M.SPRI on agarose gels since all the restriction enzymes used gave fragments of the expected size. Ten percent of plasmids conformed to this pattern and the plasmid was named **pGS2$_{C78G}$Δ6**. Figure 4.9 illustrates the restriction map of this type of deletion.

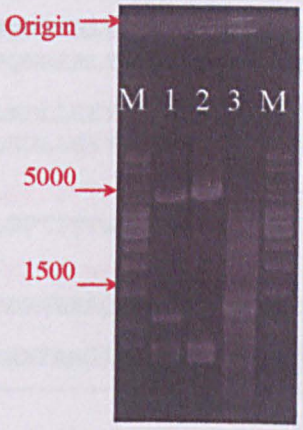


Figure 4.9: Restriction mapping of pGS2$_{C78G}$Δ6.

Plasmid DNA isolated from a colony was subjected to restriction mapping with *BamH*I, *Sal*I, and *Xho*I. Photograph of a 1 % agarose gel electrophoresis shows pGS2K (lane 1), and pGS2gΔ6 (lane 2). pGS2$_{C78G}$Δ6 plasmid was analysed by *Msp*I digestion (lane 3). One Kb ladder marker used in the agarose gel.

The restriction analysis method showed that the pGS2$_{C78G}$Δ6 plasmid was similar to the wild type in size. To find out whether a small deletion had occurred, the M.*SPR*$_{C78G}$Δ6 gene was sequenced (see sections 2.2.7 and 4.4.4). Alignment of the wild type and experimental gene, showed that six base pairs have been deleted from the M.SPRI gene (see Figure 4.10).

The nucleotide sequencing results indicate that the deletion started from GAG T and ended at AA TAC, therefore religation had occurred at GAG TAA TAC (E-Stop-Y). The predicted protein will terminate just at the end of the TRDs and the protein deduced will lack motifs IX and X. Figure 4.11 shows the protein sequence deduced from pGS2$_{C78G}$Δ6 compared with the wild type.

Figure 4.10: Nucleotide sequence results for pGS2$_{C78G}$Δ6.

The plasmid pGS2$_{C78G}$Δ6 was sequenced as described in section 4.4.5. And the results show a deletion of 6 base pairs. a) the M.SPRI wild type nucleotide sequence is given in the upper line, and amino acids deduced from the sequence are in the lower line. The deleted part of the gene is coloured in red and the start and end of the deletion is coloured in blue. B) Shows the mutated plasmid sequence and the deduced amino acids sequence.

(A) Wild type	GTT	GGA	GAG	TAT	CCA	AAA	TAC	AGA	ATT
	V	G	E	Y	P	K	Y	R	I
(B) pGS2$_{C78G}$Δ6	GTT	GGA	GAG	**T**--	---	**-AA**	TAC	AGA	ATT
	V	G	E	---	---	**-Stop**	---	---	

Figure 4.11: Comparison between the primary structures of M.MsprI and M.*SPR*$_{C78G}$Δ6.

Comparison between the amino acid sequence deduced from M.SPRI and M.SPRI$_{C78G}$Δ6, the upper line is the wild type, conserved motifs are coloured in blue and TRD in red, the lower line is M.SPRI$_{C78G}$Δ6.

```
(wt)   RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIHNVDEQLNFGDV
(mut)  RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIHNVDEQLNFGDV

(wt)   SKIDKKKLPEFDILVGGSPCQSFSVARHRKGFEDTRGTLFFQYVETLKEKQPKFFVFENV
(mut)  SKIDKKKLPEFDILVGGSPGQSFSVPRHRKGFEDTRGTLFFQYVE?LKEKQPKFFVFENV

(wt)   KGLINHDKGNTLNVMAEAFSEVGYRIDLELLNSKFFNVPQNRERLYIIGIREDLIKNEEW
(mut)  KGLINHDKGNTLNVMAEAFSEVGYRIDLELLNSKFFNVPQNRERLYIIGIREDLIKNEEW

(wt)   SLDFKRKDILQKGKQRLVELDIKSFNFRWTAQSAATKRLKDLLEEYVDEKYYLNEDKTNS
(mut)  SLDFKRKDILQKGKQRLVELDIKSFNFRWTAQSAATKRLKDLLEEYVDEKYYLNEDKTNS

(wt)   LIKELSTSRLNENLTVEQVGNINPSGNGMNGNVYNSSGLSPTITTNKGEGLKIAVEYSRK
(mut)  LIKELSTSRLNENLTVEQVGNINPSGNGMNGNVYNSSGLSPTITTNKGEGLKIAVEYSRK

(wt)   SGLGRELAVSHTLSASDWRGLNRNQKQNAVVEVRPVLTPERGEKRQNGRRFKDDGEPAFT
(mut)  SGLGRELAVSHTLSASDWRGLNRNQKQNAVVEVRPVLTPERGEKRQNGRRFKDDGEPAFT

(wt)   VNTIDRHGVAVGEYPKYRIRRLTPLECFRLQAFDDEDFEKAFAAGISNSQLYKQAGNSIT
(mut)  VNTIDRHGVAVGE----------------------------------------------
```

A second type of small deletion was found in 30% of colonies. When restriction analysis was used to classify the deletion mutations, the data indicate that there is a deletion of the *Sal*I site (see figure 4.12). To corroborate this observation, the appropriate plasmid DNA was digested with *Sal*I; and it remained undigested. This type of deletion will be referred as **pGS2$_{C78G}$Δ32.**

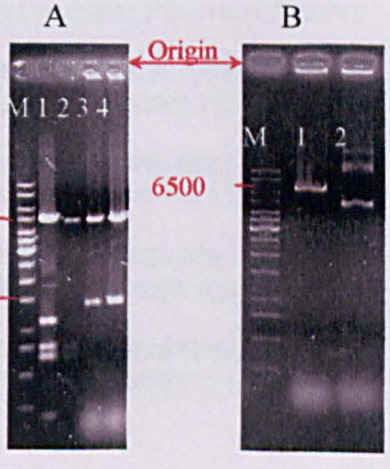


Figure 4.12: Restriction mapping of pGS2$_{C78G}$Δ32. Digested DNA gave just two bands one band equivalent to the M.SPRI gene (≈1400 bp) and the other one the rest of the plasmid (≈4950 bp) A) The DNA was digested with *BamH*I, *Sal*I, and *Xho*I using red buffer, lane 1 pGS2K, lane 2 pGS2, lane 3 and 4 two samples of pGS2$_{C78G}$Δ32. B) To double-check the presence of *Sal*I site the pGS2$_{C78G}$Δ32 was digested by *BamH*I and *Sal*I in lane 1 and by *Sal*I in lane two. One Kb ladder marker used.

Nucleotide sequence analysis was used to define the molecular details of the deletions. The five sequencing reaction were carried out using ABF20, 24, 25, 26, and 27 oligonucleotides (see sections 2.2.7 and 4.4.4), aligning the sequencing

results with the wild type M.SPRI gene shows that there was a deletion of 32 bp which started from the *Sal*I site (see figure 4.13).

Figure 4.13: Nucleotide sequence results for pGS2$_{C78G}$Δ32.

The plasmid pGS2$_{C78G}$Δ32 was sequenced as described in section 4.4.4, sequencing results show a deletion of 32 bp. a) the M.SPRI wild type sequence in the upper line, and the deduced amino acids sequence are in the lower line. The deleted part of the gene is coloured in red and the start and end of the deletion is coloured in blue. B) Shows the mutated plasmid sequence and amino acid sequence deduced from it.

(A)	ACA	AGT	CGA	CTT	AAT	GAA	AAT	CTT	ACT	GTT	GAG	CAA	GTA	GGT	AAC	ATT
	T	S	R	L	N	E	N	L	T	V	E	Q	V	G	N	I
Wild type																
(B)	ACA	AGT	CGA	---	---	---	32 bps deleted					---	---	---	--T	AAC ATT
	T	S	R	---	---	---	---	---	---	---	---	---	---	---	---	-Stop
pGS2$_{C78G}$Δ32																

The nucleotide sequencing data indicate that the deletion started from CGA and ended at TAA, religating as a stop codon, and the M.SPRI$_{C78G}$Δ32 protein terminates at amino acid 249 two amino acids before the TRDs. Figure 4.14 shows the comparison between the M.SPRI wild type and M.SPRI$_{C78G}$Δ32.

Figure 4.14: Comparison between the primary structure of M.SPRI and M.SPRI$_{C78G}$Δ32.

Sequencing data were translated using clone manager (software) and a comparison was made using align (software). The upper line is the wild type with conserved motifs coloured in blue and TRD in red, the sequence of M.SPRI$_{C78G}$Δ32 is given in the lower line.

(wt) RSLGKLR**VMSLFSGIGGFEAALRNIGV**GYE**LVGFSEIDKYAVKSFCAIH**NVD
(mut) RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIHNVD

(wt) EQLNF**GDVSK**IDKKKLPE**FDILVGGSPCQSFSVARHRKGF**E**DTRGTLFF**QYV
(mut) EQLNFGDVSKIDKKKLPEFDILVGGSPGQSFSVARARKGFEDTRGTLFFQYV

(wt) ETLKEK**QPKFFVFENVKGLINHDKGN**TLNVMAEAFSE**VGYRID**LEL**LNSKFF**
(mut) ETLKEKQPKFFVFENVKGLINHDKGNTLNVMAEAFSEVGYRIDLELLNSKFF

(wt) **NVPQNRERLYIIGI**REDLIKNEEWSLDFKRKDILQKGKQRLVELDIKSFNFR
(mut) NVPQNRERLYIIGIREDLIKNEEWSLDFKRKDILQKGKQRLVELDIKSFNFR

(wt) WTAQSAATKRLKDLLEEYVDEKYYLNEDKTNSLIKELSTSRLNENLT**VEQ**
(mut) WTAQSAATKRLKDLLEEYVDEKYYLNEDKTNSLIKELSTSR---------

4.5.1.2 LARGE DELETIONS

The other type of deletion found was a large deletion of more than 0.1Kb. In the M.SPRI$_{C78G}$ experiment three large deletions were found in which: 645, 718, and 856 base pairs were deleted; these will be named according to the number of base pairs deleted, M.SPRI$_{C78G}$Δ645, M.SPRI$_{C78G}$Δ718, and M.SPRI$_{C78G}$Δ856 respectively

Plasmid DNA encoding M.SPRI$_{C78G}$Δ645 was digested with *BamH*I, *Sal*I, and *Xho*I as described in section 4.4.4. The 1% agarose gel in figure 4.15 shows two bands: one was ≈650 bp and the other corresponds to the rest of the vector. Since one band (≈650 bp) corresponded to the M.SPRI gene (two bands of 742 and 649 bp were expected) this means that there are either small deletions at one end of the gene which has led to the deletion of one of the sites flanking the gene (*BamH*I and *Xho*I), or there is a large deletion which has deleted half of the gene including the *Sal*I site.

Figure 4.15: pGS2$_{C78G}$Δ645 digested with *BamH*I, *Sal*I, and *Xho*I.

An agarose gel shows digestion of pGS2$_{C78G}$Δ645 with *BamH*I, *Sal*I, and *Xho*I. One kb ladder marker was used. Lane 1 pGS2, lane 2 pGS2K, lane 3 and 4 two samples of pGS2$_{C78G}$Δ654. One band corresponding to 750 bp appears in the gel (see lower arrow in lane 2)

Nucleotide sequence of the M.SPRI gene for the plasmid pGS2$_{C78G}$ using the same method described in section 4.4.4 shows that two of the sequencing reactions did not work ABF24 and ABF25 presumably owing to deletion of the complementary sequences in the M.SPRI$_{C78G}$Δ645 gene. Aligning the other three sequences together with the wild type shows that a deletion of 645 bp has taken place from the plasmid (see figure 4.16).

Figure 4.16: Sequencing results of pGS2$_{C78G}$Δ645.

Three out of five nucleotide sequencing reactions worked for this type of deletion. A) The M.SPRI wild type sequence is given in the upper line, and the deduced amino acids from the sequence are in the lower line. The deleted part of the gene is coloured in red and the start and end of the deletion is coloured in blue. B) The mutated plasmid sequence, and amino acids deduced from the sequence are in the lower line.

(A)	TTT	AGT	GTA	GCC	GGC	≈	≈	≈	≈	≈	≈	≈	≈	≈	≈	AAA	AGC	GGG	CTT
Wild type	F	S	V	A	R											K	S	G	L
(B)	TTT	AGT	GTA	**GC**-	-	645	bps	deleted	-	--**C**	GGG	CTT
pGS2$_{C78G}$Δ645	F	S	V			215	aa	deleted		A	G	L

Unfortunately, it was not clear whether the deletion started from AGC and ligated at CGG or it start from GCC and ligated at CGG. However, ligation

produced was in frame protein of 215 amino acid with internal deletion between the end of motif IV to the middle of the TRDs (all of the *EcoR*II TRD and part of the *Msp*I TRD was deleted) as shown in figure 4.17.

Figure 4.17: Comparison between the primary structure of M.SPRI and M.SPRI$_{C78G}$Δ645.

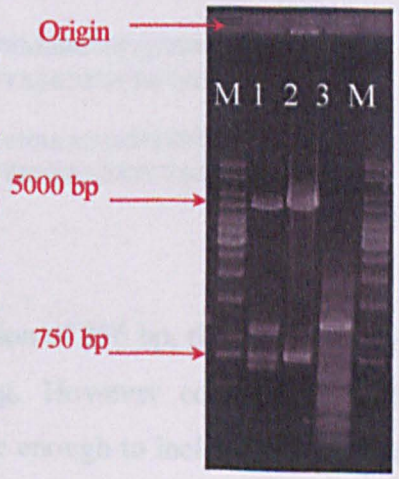
Sequencing data were translated using clone manager and a comparison was made using align (software). The upper line is the wild type with conserved motifs is coloured in blue, *EcoR*II TRD in green, *Msp*I TRD in red, and *Hae*III in light green. The sequence of M.*SPR*$_{C78G}$Δ645 is given in the lower line.

```
(wt)   RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIHNVDEQLNF
(mut)  RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIHNVDEQLNF

(wt)   GDFSKIDKKKLPEFDILVGGSPCQSFSVARHRKGFEDTRGTLFFQYVETLKEKQPKF
(mut)  GDFSKIDKKKLPEFDILVGGSPGQSFSVA-------------------------

(wt)   FVFENVKGLINHDKGNTLNVMAEAFSEVGYRIDLELLNSKFFNVPQNRERLYIIGIR
(mut)  ------------------------------------------------------

(wt)   EDLIKNEEWSLDFKRKDILQKGKQRLVELDIKSFNFRWTAQSAATKRLKDLLEEYVD
(mut)  ------------------------------------------------------

(wt)   EKYYLNEDKTNSLIKELSTSRLNENLTVEQVGNINPSGNGMNGNVYNSSGLSPTITT
(mut)  ------------------------------------------------------

(wt)   NKGEGLKIAVEYSRKSGLGRELAVSHTLSASDWRGLNRNQKQNAVVEVRPVLTPERG
(mut)  ----------------GLGRELAVSHTLSASDWRGLNRNQKQNAVVEVRPVLTPERG

(wt)   EKRQNGRRFKDDGEPAFTVNTIDRHGVAVGEYPKYRIRRLTPLECFRLQAFDDEDFEK
(mut)  EKRQNGRRFKDDGEPAFTVNTIDRHGVAVGEYPKYRIRRLTPLECFRLQAFDDEDFEK

(wt)   AFAAGISNSQLYKQAGNSITVTVLESIFKELIHTYVNKESE
(mut)  AFAAGISNSQLYKQAGNSITVTVLESIFKELIHTYVNKESE
```

The second type of large deletion leads to the loss of 718 bp. Restriction mapping gave two bands instead of the expected three bands, as with the previous deletion. However, in this case the band corresponding to the M.SPRI gene was ≈750 bp which means that deletion of one of the three sites used to digest the gene had occurred (see figure 4.18).

Figure 4.18: pGS2$_{C78G}$Δ718 digested with *BamH*I, *Sal*I, and *Xho*I

Photograph of a 1% agarose gel shows the digestion of pGS2$_{C78G}$Δ718. The digestion yields two bands, which indicates that there is a deletion of one of the sites used for digestion or a large deletion including the *Sal*I site. Lane 1 is pGS2K, lane 2 pGS2$_{C78G}$Δ718 digested with *BamH*I, *Sal*I, and *Xho*I, and lane 3 pGS2$_{C78G}$Δ718 digested with *Msp*I. One Kb ladder marker used



Nucleotide sequence analysis of this of pGS2$_{C78G}$Δ718 shows that there is a deletion of 718 base pairs at the middle of the gene. Since there are similarities between the start and end of the deletion sites the deletion may start from CAC and end at AAA, or start from AAA and end at GCT (see figure 4.19).

Figure 4.19: Sequencing results of pGS2$_{C78G}$Δ718.

Two out of five sequencing reactions worked for this type of deletion. A) the M.SPRI wild type sequence is given in the upper line, and the amino acids deduced from the sequence are in the lower line. The deleted part of the gene is coloured in red and the start and end of the deletion is coloured in blue. B) The mutated plasmid sequence, amino acids deduced from the sequence are in the lower line.

(A)	TTT	AAT	GTT	CCA	CAA	AAT	≈≈≈≈≈≈≈≈≈≈≈	TTT	GAA	AAA	GCT	TTT
Wild type	F	N	V	P	Q	N			F	E	K	A	F
(B)	TTT	AAT	GTT	CCA	**C-**	**--718 bps deleted-**			**--A**	**A**AA	GCT	TTT
pGS2$_{C78G}$Δ718	F	N	V	P		239 aa deleted				Q	K	L	

The nucleotide sequencing results indicate that the protein is altered in the middle of motif VIII and is then terminated prematurely after six mutant amino acids (see figure 4.20).

Figure 4.20: Comparison between the primary structure of M.SPRI and M.SPRI$_{C78G}$Δ718.

Sequencing data were translated using clone manager and a comparison was made using align (software). The upper line is the wild type conserved motifs are coloured in blue. M.SPRI$_{C78G}$Δ718 is in the lower line the unmatched amino acids are coloured in red. The deletion starts from middle of motif VIII and 239 amino acid are deleted.

```
(wt)   RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIH
(mut)  RSLGKLRVMSLFSGIGGFEAALRNIGVGYELVGFSEIDKYAVKSFCAIH

(wt)   NVDEQLNFGDVSKIDKKKLPEFDILVGGSPCQSFSVARHRKGFEDTRGT
(mut)  NVDEQLNFGDVSKIDKKKLPEFDLLVGGSPGQSFSVAGHRKGFEDTRGT

(wt)   LFFQYVETLKEKQPKFFVFENVKGLINHDKGNTLNVMAEAFSEVGYRID
(mut)  LFFQYVETLKEKQPKFFVFENVKGLINHDKGNTLNVMAEAFSEVGYRID

(wt)   LELLNSKFFNVPQNRERLY
(mut)  LELLNSKFFNVPQKLLLRE
```

The last type of deletion observed is a deletion of 856 bp, the nature of this deletion was not clear from restriction mapping. However comparison with control plasmids suggest that the deletion was large enough to include most of the M.SPRI gene (see figure 4.21).

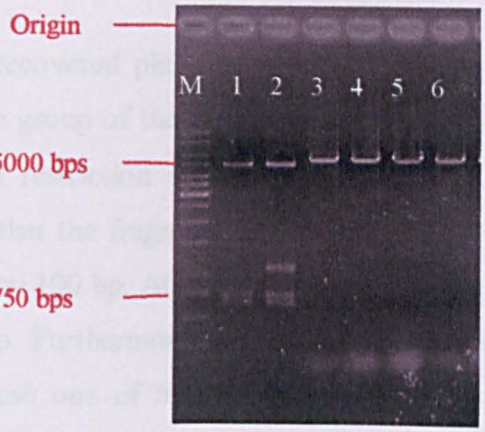


Figure 4.21:pGS2$_{C78G}$Δ856 digested with *BamH*I, and *Xho*I.

Photograph of a 1% agarose gel shows the digestion of pGS2$_{C78G}$Δ856. The digestion yields one band, which indicates that there is a deletion of two of the sites used for digestion or a large deletion. Lane 1 is pGS2, lane 2 pGS2K, lane 3, 4, and 5 are three samples of the pGS2$_{C78G}$Δ856. The one Kb ladder used as a marker

Nucleotide sequence analysis of the recovered plasmids (section 4.4.4) shows that the large deletion starts at the 3' end of the GST gene and terminates downstream of the TRDs in the M.SPRI gene. Interestingly there are again similarities between the deletion start and end sequences (see figure 4.22). The deletion starts from the AGCAAGTA sequence and ends at the same sequence 856 base pairs down stream in the DNA. In this type of deletion no M.SPRI protein will be expressed since the gene is ligated at a stop codon 25 codons before the start of the M.SPRI gene (see figure 4.22).

Figure 4.22: Nucleotide sequence results of pGS2$_{C78G}$Δ856.

Three out of five sequencing reactions worked for this type of deletion. A) the M.SPRI wild type sequence is given in the upper line, and the amino acids deduced from the sequence are in the lower line. The deleted part of the gene is coloured in red and the start and end of the deletion is coloured in blue. B) The mutated plasmid sequence and amino acids deduced from the sequence are in the lower line. C) Sequencing results were translated using clone manager and comparison was made using align software. The upper line is the wild type and M.SPRI$_{C78G}$Δ856 is in the lower line.

(A)	TTG	AAA	TCC	AGC	AAG	TAT	≈≈≈≈≈≈≈≈≈≈≈			GTT	GAG	CAA	GTA	GGT
Wild type	L	K	S	S	K	Y				V	E	Q	V	G
(B)	TTG	AAA	TCC	---	856	bps	deleted	-	---	-AG	CAA	G**TA**	**G**GT	
pGS2$_{C78G}$Δ856	L	K	S	---	285	aa	deleted	---	-	-	S	K	**Stop**	

C)

(wt) **DKY**LKSSKYIAWPLQGWQATFGGGDHPPKSDLVP**RGSLGK**LRVMS
(mut) DKYLKSSK- M.*SPRI* gene start

4.5.1.3 REARRANGEMENTS PRODUCED IN PLASMIDS ENCODING M.SPRI$_{C78G}$

Unfortunately not all of the recovered plasmid restriction data could be reconciled with a deletion event. One group of the recovered plasmids gave two species of 1500 and 2600 bp when restriction analysis was carried out with *BamH*I, *Sal*I, and *Xho*I. This means that the fragment encoding the M.SPRI$_{C78G}$ gene had suffered an increase in size by 100 bp. At the same time the total size of the plasmid was reduced by 2200 bp. Furthermore in addition to the insertion, there appears to be a deletion because one of the three enzymes used in the restriction mapping did not cut. Nucleotide sequencing of the rearranged M.SPRI$_{C78G}$ gene using five different primers (see section 4.4.4) failed to solve the problem.

4.5.2 DELETIONS IN THE PLASMID ENCODING MONOMERIC M.SPRI$_{C78G}$

Since M.SPRI and M.SPRI$_{C78G}$ were cloned into pET vectors and were used to transform AB1157 *E.coli* which does not carry T7 RNA polymerase gene, the Mtases genes are transcriptionally inactive and should not cause any problem to the host cells. In order to express a target gene in pET two options exist: a host bearing the T7 RNA polymerase gene (λDE3 lysogen) can be employed, but because AB1157 is preferred host in this work the cells were supplied with DE3 by infection with λCE6 phage.

pES1$_{C78G}$ was used to transform AB1157 competent cells and expression of the Mtase was induced by CE6 phage as described in section 2.2.8.1. Cells were then plated on LA containing ampicillin. An equal amount of uninduced *E.coli* AB1157 cells were plated in the same way. The numbers of the colonies obtained from the experiments are shown in table 4.3.

In another experiment, the influence of expression levels of pLS1$_{C78G}$ and pLS1$_{C78G}$K was examined using a Tac promoter. Numbers of colonies recovered in this experiment were higher than with the dimeric M.SPRI$_{C78G}$. Table 4.3 summarises the data.

Table 4.3: Number of colonies recovered from the experiments after transformation with, $pLS1_{C78G}$, and $pLS1_{C78G}K$ and induction of $pES1_{C79G}$, in AB1157. The transformation efficiency of the competent cells was $1X10^6$.

	$pES1_{C79G}$ induced	$pES1_{C79G}$ uninduced	$pLS1_{C78G}$	$pLS1_{C78G}K$
Number of colonies	**22.6**	**241.6**	**250**	**3160**
Sd	**7**	**57.9**	**24**	**557.5**
Transformation efficiency	-	-	**$5 X 10^3$**	**$6.3 X 10^4$**
Percentage of survived cells	**9.3**		**7.9**	

Ten colonies were recovered and were used to isolate plasmid DNA which was digested with *BamHI* and *XhoI* to analyse any deletions. Figure 4.23 shows two examples from each plate, which indicates that no deletions had occurred.

Figure 4.23: Monomer $M.SPRI_{C78G}$ digested with *BamHI* and *XhoI*.

Photograph of 1% agarose gel shows sample of the digestion of plasmid isolated from cells carrying $pES1_{C78G}$ survived after induction by λCE6 (lanes 2-5) control pES1 in lane 1.The one Kb ladder marker used.

As presented above three samples in the $pUM1_{C174G}$ experiment were sequenced (see section 3.5.1) and there were no deletions. Furthermore the gel shown in figure 4.23 shows no difference between the wild type and mutant M.SPRI.

4.6 SUMMARY OF RESULTS

1- Dimerised M.SPRI$_{C78G}$ convey deletions to plasmid DNA in a manner similar to the monospecific enzymes.

2- The level of repair of plasmids encoding M.SPRI$_{C78G}$ DNA damage is two folds lower than observed with monospecific enzymes.

3- Nucleotide sequence analysis of deletions revealed that in AB1157 a single large (718 bp) deletion predominates.

4- Parallel experiments in which pGS2$_{C78G}$ was used to produce deletions in two repair deficient strains of *E.coli* revealed that strain-specific deletion patterns had arisen.

5- Nucleotide sequence analysis of M.SPRI$_{C78G}$ induced deletions revealed that there is no correlation between the recognition sequence of the Mtase and the sequence at the junction of the repaired lesion.

6- Nucleotide sequences analysis of M.SPRI$_{C78G}$ induced large deletions revealed that there are homologous sequence at the junction of the repaired lesion.

4.7 DISCUSSION

In this chapter it has been found that the mutation of Cys to Gly in motif IV in the multispecific Mtases is also cytotoxic to cells. Results presented in this chapter support the finding in chapter three, that the lethal effect on cells is dependant on the oligomeric nature of the mutated Mtases: using dimerised-M.SPRI$_{C78G}$, 2% of the cells were able to survive while in the monomer-M.SPRI$_{C78G}$ the survival level was approximately 8% (see table 4.3). More interestingly, all of the recovered cells following transformation with plasmid encoding dimerised M.SPRI$_{C78G}$ carried deletions, while those cells transformed with plasmid encoding monomeric M.SPRI$_{C78G}$ were without deletions. Based on these observations and on the monospecific Mtases results in chapter 3, the deletion event appears to be dependent on the dimeric state of the mutated Mtase and the presumed protein: DNA complex formation *in vivo*.

I have shown that only 2% of cells were able to recover from the lethal effect of the presumed protein:DNA complexes formed by the high affinity binding of the M.SPRI$_{C78G}$ at its TRSs. In contrast the percentage of repair in those cells transformed with mutated, dimerised monospecific Mtases was approximately 4%. Percentage recovery of colonies following transformation with a plasmid encoding M.SPRI$_{C78G}$ was found to be dependent on the strain used.

Nucleotide sequence analysis of deletions revealed that in AB1157 a single large (718 bp) deletion predominates. Parallel experiments in which pGS2$_{C78G}$ was used to produce deletions in two repair deficient strains of *E.coli* revealed that strains-specific deletion patterns can arise. This variation in survival frequency will be addressed in chapter 5.

All the experiments with the mutated Mtases were initially carried out with the pGEX vectors. The influence of these plasmids was determined by repeating the experiments with two different cloning vectors. pLitmus28 gave the same results as pGEX; the percentage of recovery for pGS2$_{C78G}$ was 2.19% and for pLS2$_{C78G}$.it was 1.9%. The toxic effect of the mutated Mtase requires the protein to be dimeric but is not influence by the choice of vector.

One of the issues addressed in this chapter was the specificity of the deletion process. As mentioned before, multispecific M.SPRI recognises three sites CCGG, GGCC, and CCWGG. It was hoped that this multi-specificity would help to answer the above question. One of the possibilities based on preliminary data from Hurd and Hornby (unpublished) was that the protein would bind to any two recognition sequence as a dimer and subsequently deletion would arise between the two recognition sites. (see figure 4.24).

Figure 4.24: The M.SPRI deletion specificity.

Assuming that deletion will occur between two recognition sites, by using a mutated multispecific Mtase the deletion specificity can be determined. The ligated sequence would arise between the red, green, or blue arrows if deletion occurred presisely between recognition sites.

According to our finding in this chapter the sequencing of the recovered plasmids indicates that the deletion start-end points are far from the TRS (see table 4.4 and figure 4.25) and that earlier observations of precise deletions were merely fortuitous. In one case ($pGS2_{C78G}\Delta645$), the deletion started from CCGG (*Msp*I site) and ended 98 bp before the next TRS. There was no significant correlation between the number of bases before ($F = 1.32$, P-value = 0.316) and after the deletions ($F = 1.306$, P-value = 0.322).

Table 4.4: Number of bases to the nearest target recognition sequence in the observed deletions.

The three target recognition sequences CCGG (*Msp*I), GGCC (*Hae*III), and CCWGG (*Eco*RII) are shown, also the any nearest recognition sequence calculated for all the deletions founds. Numbers in the table indicate the number of bases. (≺) Nearest recognition site before the deletion and (≻) nearest recognition site after the deletion

Gene	*Nearest target DNA sequence*							
	Any site		CCGG		GGCC		CCWGG	
	≺	≻	≺	≻	≺	≻	≺	≻
$pGS2_{C78G}\Delta6$	116	119	860	119	116	505	1281	2155
$pGS2_{C78G}\Delta32$	487	221	487	466	506	221	908	2502
$pGS2_{C78G}\Delta645$	16	98	236	341	16	98	418	2449
$pGS2_{C78G}\Delta718$	218	47	218	47	237	363	638	2223
$pGS2_{C78G}\Delta856$	74	231	1075	476	375	231	74	2582

One other possibility considered was that the deletion may start or end at a specific motif in the M.SPRI gene, or at specific sequence even if that sequence is not one of the target sequences recognised by M.SPRI enzyme. Deletions appeared to start at any sequence in the gene; just before the gene ($pGS2_{C78G}\Delta856$), at the middle of motif IV ($pGS2_{C78G}\Delta456$), at motif VIII ($pGS2_{C78G}\Delta718$), and in the variable region before and after the TRDs ($pGS2_{C78G}\Delta32$ and $pGS2_{C78G}\Delta6$ respectively), which suggests that there is no direction relation between the deletion event and the position in the gene (see figure 4.26). Furthermore, nucleotide sequencing results shows that there are no similarities between the different types of deletion start end points (see table 4.5).

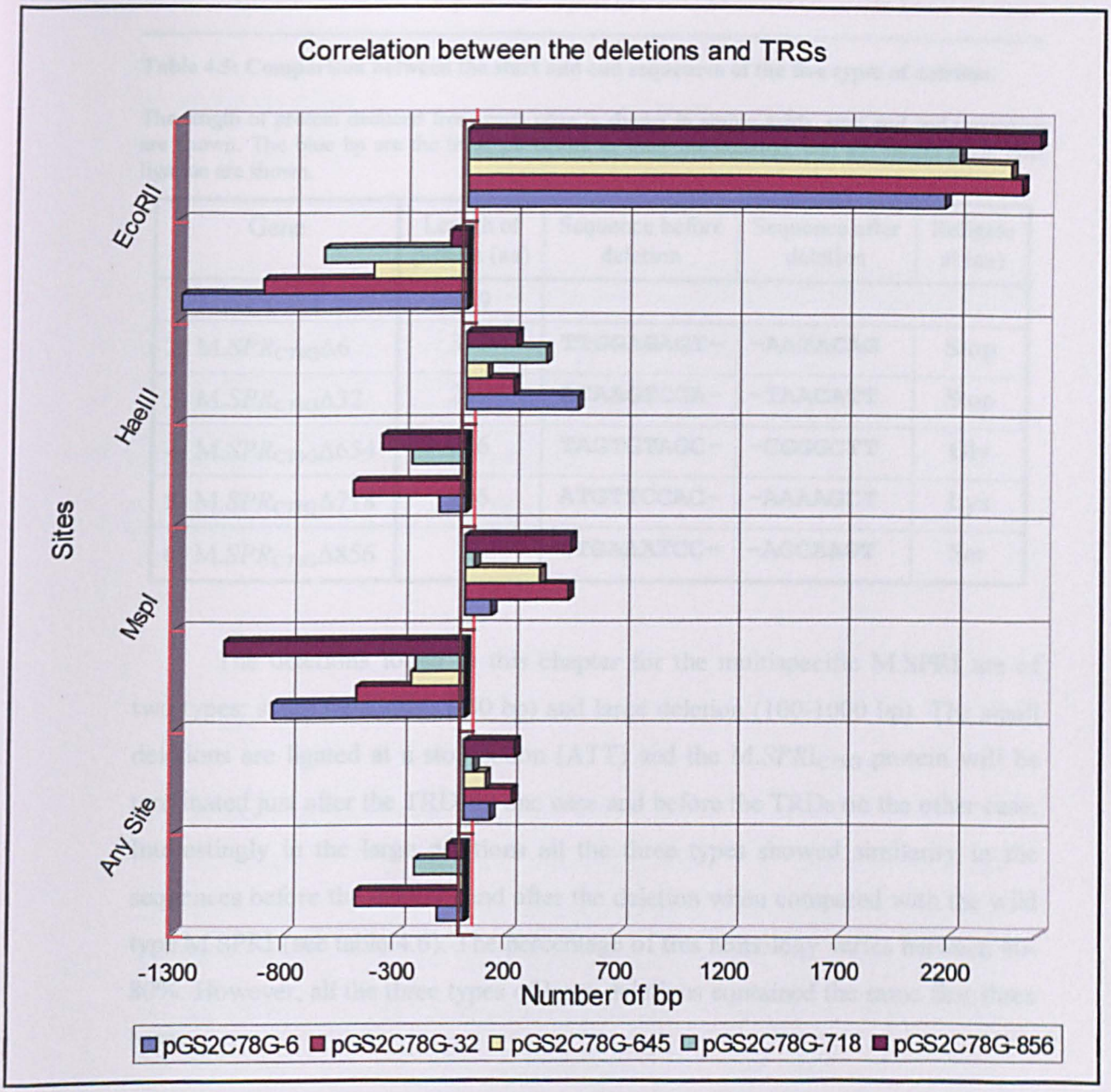


Figure 4.25: Correlation between the deletion start-end points and the target recognition sites (TRSs).

The number of bp to the nearest TRS before and after each deletion were calculated and plotted in this diagram. Also the number of bp to the nearest TRS were plotted. The red line in the middle of the chart indicates the deletion start-end point. Numbers in negative values are before the deletion and numbers in the positive values are after the deletions. Data are presented in table 4.4.

Table 4.5: Comparison between the start and end sequences of the five types of deletion.

The length of protein deduced from each gene is shown in amino acids, start and end sequences are shown. The blue bp are the three pb before or after the deletion, and the amino acids after ligation are shown.

Gene	Length of protein (aa)	Sequence before deletion	Sequence after deletion	Religate at (aa)
1) M.*SPR* WT	439			
2) M.*SPR*$_{C78G}$Δ6	373	TTGGAGAGT-	-AATACAG	Stop
3) M.*SPR*$_{C78G}$Δ32	249	ACAAGTCGA-	-TAACATT	Stop
4) M.*SPR*$_{C78G}$Δ654	226	TAGTGTAGC-	-CGGGCTT	Gly
5) M.*SPR*$_{C78G}$Δ718	186	ATGTTCCAC-	-AAAAGCT	Lys
6) M.*SPR*$_{C78G}$Δ856	0	TTGAAATCC-	-AGCAAGT	Ser

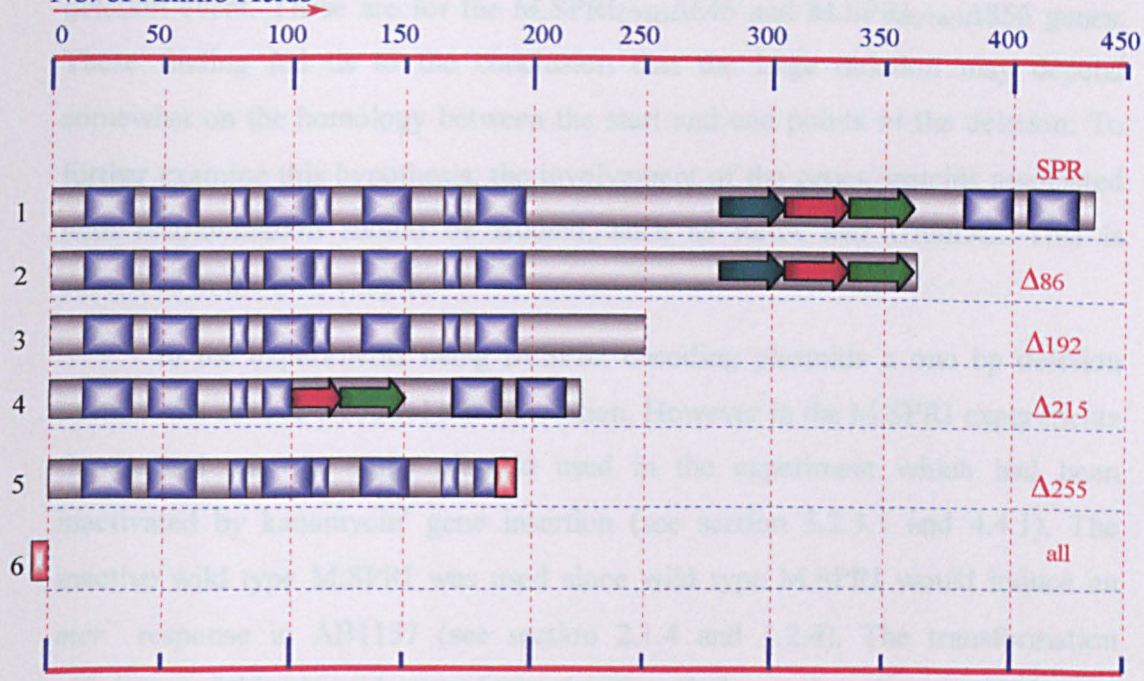
The deletions found in this chapter for the multispecific M.SPRI are of two types: small deletions (1-50 bp) and large deletion (100-1000 bp). The small deletions are ligated at a stop codon (ATT) and the M.*SPRI*$_{C78G}$ protein will be terminated just after the TRDs in one case and before the TRDs on the other case. Interestingly in the large deletions all the three types showed similarity in the sequences before the deletion and after the deletion when compared with the wild type M.SPRI (see table 4.6). The percentage of this homology varies between 40-80%. However, all the three types of large deletions contained the same first three bases.

Table 4.6: Comparison between the sequences before and after the deletion in large deletion.

The sequence before and after deletion for each deletion type has been aligned with the wild type individually. Red bases illustrate the three bases before deletion, and blue bases the homologous bases after deletion. The percentages of homology are given in the last column.

Gene	Sequences before deletion	Sequences after deletion	Homology of the first 10 bp (%)
M.*SPR* WT	GTGTAGC	CGGCCATCGA	
M.SPRI$_{C78G}$Δ645	GTGTAGC	CGGGCTTGGA	70
M.*SPR* WT	GTTCCAC	AAAATAGGGA	
M.SPRI$_{C78G}$Δ718	GTTCCAC	AAAAGCTTTT	40
M.*SPR* WT	GAAATCC	AGCAAGTATA	
M.SPRI$_{C78G}$Δ856	GAAATCC	AGCAAGTAGG	80

A



B

1) M.*SPR* gene	(439 aa)			
2) M.*SPR*gΔ6	(373 aa)	TTGGAGAGT--	6 bps	--AATACAG
3) M.*SPR*gΔ32	(249 aa)	ACAAGTCGA--	32 bps	--TAACATT
4) M.*SPR*gΔ654	(226 aa)	TAGTGTAGC--	654 bps	--CGGGCTT
5) M.*SPR*gΔ718	(186 aa)	ATGTTCCAC--	718 bps	--AAAAGCT
6) M.*SPR*gΔ856	(0 aa)	TTGAAATCC--	856 bps	--AGCAAGT

Figure 4.26: Summaries of the five type of deletions found using M.SPRI with Gly mutation in motif IV.

A) Comparison between the deletion mutants with M.SPRI, number of base pairs deleted in each type is indicated in red. Blue boxes indicate the conserved motifs in the protein to clarify which motif is deleted in the protein. Arrows in the diagram indicate the target recognition domains for *Eco*RI, M.*Msp*I, and M.*Hha*I. Red boxes indicate the amino acids translated which do not match the reading frame of the original protein.

B) Sequences at the sites where the deletion start and end points.

Interestingly there are just two sequences identical in sequence after the deletion event. These are for the M.SPRI$_{C78G}$Δ645 and M.SPRI$_{C78G}$Δ856 genes. These finding led us to the conclusion that the large deletion may depend somewhat on the homology between the start and end points of the deletion. To further examine this hypothesis, the involvement of the genes/proteins associated with recombination should be studied, such as RecA and UmuDC. This is described in the next chapter.

In the experiments using M.SPRI encoding plasmids a one bp deletion mutant was used as a control transformation. However in the M.SPRI experiments the control was the same plasmid used in the experiment which had been inactivated by kanamycin^r gene insertion (see section 3.2.3.1 and 4.4.1). The inactive wild type M.SPRI was used since wild type M.SPRI would induce an *mcr*$^+$ response in AB1157 (see section 2.1.4 and 1.2.4). The transformation efficiency of this plasmid control was significantly lower than the control plasmid pLitmus28 (7X10^4 cfu/μg for pLS2$_{C78G}$K and 6.3X10^4 cfu/μg for pLS1$_{C78G}$K compared with 1X10^6 cfu/μg for pLitmus28). Furthermore using the same plasmids to transform RRI *E.coli* cells (*mcr*$^-$) gave good transformation efficiencies 1X10^6 cfu/μg. Those observations lead us to suggest that the M.SPRI inactivated by kanamycin^r gene is "leaky" and protein is expressed in an active form even with the presence of kanamycin^r gene cassette. Trautner and colleges (Walter et al. 1992) found that multispecific Mtases can accommodate inert material of non-Mtase origin within their variable region without losing their activity. This observation now has been extended by work in Hornby's lab (unpublished) in which insertion of long fragments of DNA into the DNA encoding the variable region in M.SPRI did not lead to loss of activity. As presented in figure 4.1 and 4.2 the *Sal*I site used to insert the kanamycin^r gene is located in the variable coding region. However, since only trace amounts of the active M.SPRI will be expressed from the M.SPRI-Kanamycin^r gene (figure 3.3e and 3.4e) only a relatively weak *mcr*$^-$ response will be induced.

The low amount of the M.SPRI protein expressed even with the kanamycin^r gene cassette inserted into the *Sal*I site presumably explains the low transformation efficiency obtained with the control plasmid. Since in each experiment the active M.SPRI$_{C78G}$ and inactive M.SPRI$_{C78G}$K were used to

transform *E.coli* (see section 4.4) this will subsequently suppress the percentage of recovered cells in the previous experiments. This expectation is in agreement with our conclusion in chapter three that the lethal effect of the C to G mutation increases with the number of sites recognised by the Mtase. There are 11 TRSs in the M.SPRI gene and the percentage of recovery was less than 2% which is significantly lower than the effect of M.*Msp*I$_{C174G}$ (3.8% and which contains three TRS in the gene) but may in fact be an over-estimation of repair efficiency.

CHAPTER FIVE:
DNA REPAIR GENES INVOLVED IN THE DELETION EVENT

SYNOPSIS OF CHAPTER FIVE

- A brief review of RecA.
- A brief review of UmuDC.
- Description of experimental design.
- Demonstration of the involvement of RecA in the deletion event.
- Demonstration of the involvement of UmuDC in the deletion event.
- Discussion of results obtained from chapter five.

5.1 INTRODUCTION

It is proposed that the deletion phenomenon found with Mtase mutants in which the conserved Cys in motif IV had been changed to Gly, is due to the high affinity binding between the mutant protein and DNA. This unique type of lesion (macro-lesion) is clearly repaired by certain strains through some form of error-prone repair, which leads to mutations in the target gene. As described in section 1.4, there are two types of repair dealing with the lesions affecting the DNA replication, SOS repair and SOS mutagenesis. The key player in SOS repair is the RecA protein, while for SOS mutagenesis it is the UmuDC proteins. In this chapter I shall discuss the involvement of RecA and UmuDC in the deletion event using three isogenic strains carrying different mutations in these repair genes.

5.1.1 *THE RECA PROTEIN*

The gene encoding *recA* was first shown in 1965 to be essential for genetic recombination and resistance to UV irradiation (Claek and Margulies, 1965). The RecA protein has been found in all bacteria, including *Mycoplasma* within its minimal genome. The *recA* gene has been sequenced in over 60 bacterial species (Roca and Cox, 1997). Structural homologues of the RecA protein have been found in eukaryotes from yeast to human (Story et al. 1993). The structure of *E.coli* RecA has been determined as both a monomer and polymer (Story et al. 1992). These studies revealed there is a major central domain flanked by two smaller subdomains at the N and C termini. Monomers in the crystal are packed to form a continuous spiral filament, with six monomers per right-handed helical turn. The filament exhibits a deep helical groove, which can accommodate up to three DNA strands (Story et al. 1992). The alignment of 64 full-length bacterial RecA sequences indicates that RecA protein is highly conserved (Roca and Cox, 1997). The RecA protein is a 352-aa polypeptide of M_r 37,842. This modestly sized polypeptide has separate binding sites for at least three stands of DNA, ATP, the LexA and λ repressors, and other RecA monomers. The active species of RecA protein implicated in most of its activities has a structure in which RecA protein monomers are assembled into a right-handed helical filament on DNA.

Functions of RecA protein in vivo:

1- **Coprotease function:** as discussed in section 1.4.

2- **DNA recombination:** RecA is able to pair two homologous DNA molecules if one of them is single stranded or partially single stranded (Kowalczykowski and Eggleston, 1994). The pairing of single stranded or partially single-stranded DNA with duplex DNA that has suitable ends leads to a further reaction: an exchange of strands that can proceed for several kilo-bases. When a partially single-stranded molecule reacts with a fully duplex molecule that has suitable ends, strand exchange is reciprocal and produces the classical recombination intermediate known as a Holliday junction (Holliday, 1964). One of the roles for DNA recombination is to repair damaged DNA, in this case the correct "information" is supplied by a different but homologous DNA molecule. (Friedberg et al. 1995). Figure 5.1 summarises some of the current models for recombinational DNA repair.

3- **Chromosome Partitioning:** the abnormal chromosome complement in cells lacking *recA* function has led to a proposal that the RecA protein has a role in the proper partitioning of chromosomes at cell division. Part of the role of RecA in partitioning may be indirect, mediated by the protection of the chromosome from the nuclease digestion that occurs in the absence of RecA. RecA may also be required to produce the tension required for chromosomal partitioning (Skarstad and Boye, 1993).

4- **Induced stable replication:** DNA damage brings about elevated levels of DNA replication that are not dependent on initiation at the bacterial replication origin, *oriC*. This phenomenon is called induced stable DNA replication. Replication is closely linked to recombination and recombinational DNA repair (Asai and Kogoma, 1994). It has been suggested that single-stranded 3' ends involved in an early stage of the recombinational processes might, after being paired with a new complementary strand, act as primers for this DNA synthesis (Roca and Cox, 1997).

5- **SOS mutagenesis:** as described above (1.4.5). It should be noted that the RecA protein has at least three roles in SOS mutagenesis. First, some of the required activities are induced as part of the SOS response. Second, at least

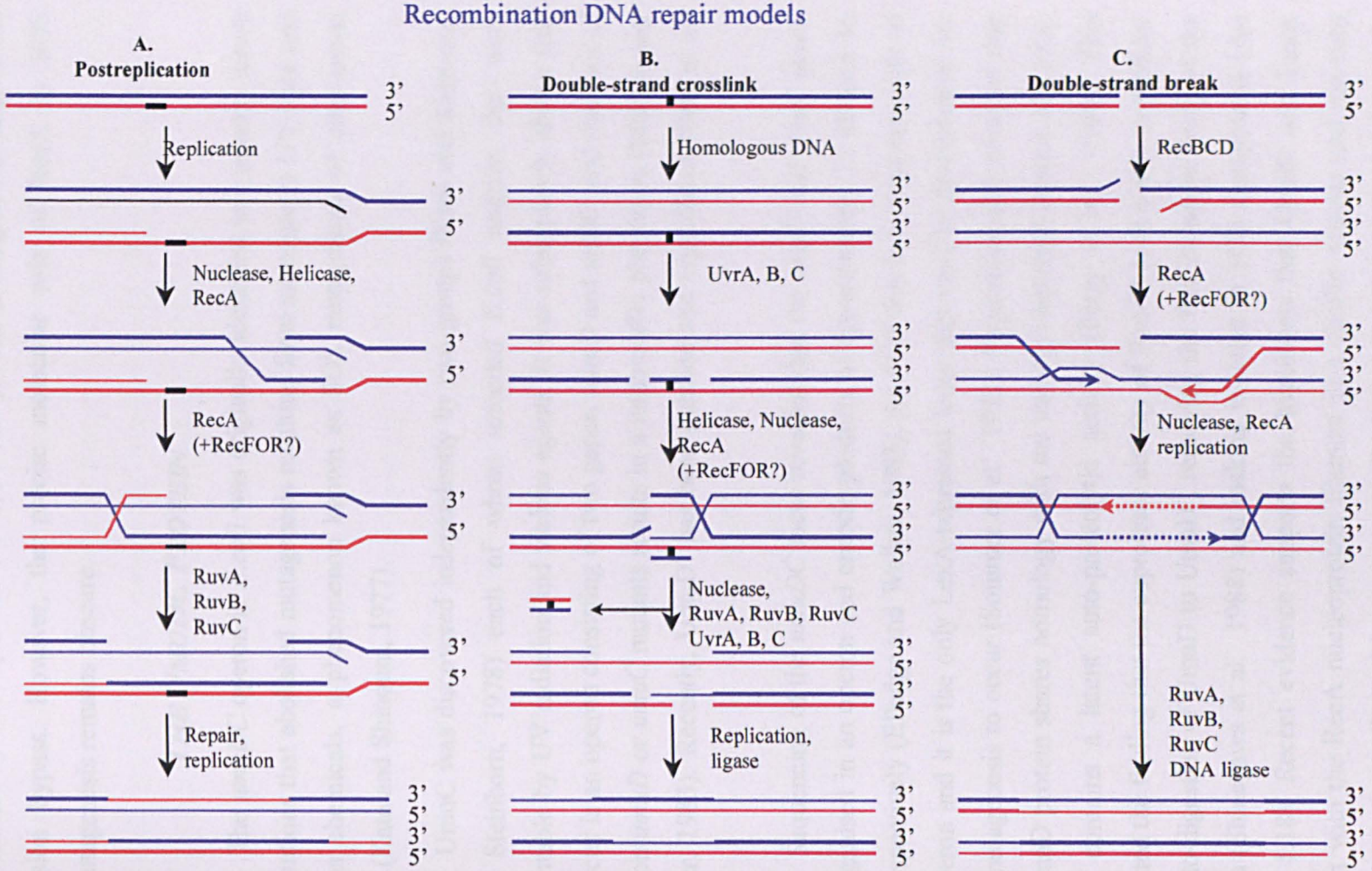


Figure 5.1: Models for recombination DNA repair. Adabted from (Roca and Cox, 1997)

one of the required proteins is proteolytically processed in a reaction facilitated by RecA protein. Finally, RecA protein participates directly in the lesion bypass. However, the precise molecular role of RecA in SOS mutagenesis remains obscure.

5.1.2 THE UMUDC PROTEIN

The *umuDC* operon in *E.coli* was originally identified in a genetic screen of mutations that abolished mutagenesis resulting from exposure to UV light and various chemicals, a phenomenon known as SOS mutagenesis or error-prone repair (Kato and Shinoura, 1977).

UmuC was discovered independently by two groups (Kato and Shinoura, 1977; Steinborn, 1978) each of whom screened *E.coli* mutants that were nonmutable by UV radiation and various agents. It was subsequently shown that this locus is an operon consisting of two genes, *umuD* and *umuC*, and that loss of function *umuD* or *umuC* mutants results in a nonmutable phenotype (Elledge and Walker, 1983). Recently, UmuD` has been purified and crystallised (Peat et al. 1996)

Sequencing of the *umuDC* locus revealed that the *umuD* and *umuC* genes are organised in an operon and encode products of approximately 15 kD and 45 kD, respectively (Elledge and Walker, 1983). An SOS box is located upstream of the operon and it is the only LexA-repressed locus that must be de-repressed for SOS mutagenesis to occur (Sommer et al. 1993). Sequencing also revealed that the UmuD protein shares homology with the carboxyl-terminal domain of LexA, which contains a latent auto-proteolytic activity (Perry et al. 1985). This prompted the finding that the coprotease activity of RecA* is also able to mediate the auto-digestion of UmuD to UmuD`, removing the first 24 amino acids of the protein (Shinagawa et al. 1988) and making it active in SOS mutagenesis (see figure 1.18). Recent evidence supports the hypothesis that UmuD and LexA interact with the RecA nucleoprotein filament in a similar manner and are then stimulated to autodigest (Nastri et al. 1997). It is possible that the differences in the interactions of UmuD and LexA with the RecA nucleoprotein filament are responsible, in part, for the fact that the rate of RecA*-mediated cleavage of UmuD is slower than that of LexA (Woodgate and Ennis, 1991). *In vivo*, this

delayed cleavage is manifested as a delay of maximal SOS mutagenic activity by approximately 30 min after UV irradiation. Recently, evidence of further post-translational regulation of SOS mutagenesis has been reported. Lon, an ATP-dependent protease, has been shown to play a role in the degradation of UmuC and UmuD, and the ClpXP protease is involved in the degradation of UmuD' when it is in a heterodimer with UmuD (Frank et al. 1996).

The nature of the UmuC protein has remained much more elusive than that of its partners UmuD and UmuD`. The initial purification of an active protein was challenging and required denaturation followed by a gradual refolding in the presence of equimolar amounts of ribosomal S9 protein (Woodgate and Ennis, 1991). This purified UmuC was shown to have a ssDNA binding activity *in vitro* (Cayrol et al. 1995), which was also seen with a purified, soluble Umu(D`)2C complex. It has been estimated that one Umu(D`)2C complex is bound per 20 nucleotides of ssDNA. However, activity of this complex in translesion synthesis has not been reported. The biological consequences of the ssDNA binding activity are not yet clear, but the ability of UmuC and the Umu(D`)2C complex to bind ssDNA could be involved in interactions with the DNA polymerase at the replication fork or in the direction of the Umu proteins to damaged DNA (Smith and Walker, 1998).

The regulation of the *umuDC* operon by RecA and LexA can be summarised as follows: DNA damage generates a signal that converts RecA to RecA*. RecA* mediates the cleavage of the LexA repressor that results in the induction of the *umuDC* operon as well as the rest of the SOS response genes. RecA* can also mediate the processing of UmuD to the shortened UmuD` molecule. UmuD and UmuD` can interact with UmuC in a variety of combinations. The Umu $(D)_2C$ complex seems to be involved in regulating the *E. coli* cell cycle after DNA damage (Smith and Walker, 1998). The Umu $(D`)_2C$ complex is active in SOS mutagenesis (translesion synthesis). The third complex, UmuDD`C, does not appear to have an activity, but it may play a role in shutting off SOS mutagenesis by sequestering UmuD` (Smith and Walker, 1998) (see figure 1.19).

5.2 EXPERIMENTAL DESIGN

5.2.1 ISOGENIC STRAINS USED IN THE EXPERIMENT

At the beginning of the experiment RRI and HB101 were used as wild type and *recA*⁻ strains respectively, RRI (the mutated form of HB101) was isolated by plating HB101 on methyl methanesulfonate (MMS) and viable cells were picked (Boyer and Dussoix, 1969). Thus RR1 is not a wild type parent of HB101 but rather an MMS-resistant revertant of HB101. The *recA* phenotype of RRI is ill-defined (Walker, personal communication). In addition RR1 may carry other mutations as a consequence of the MMS exposure. These possible mutations, if there are any, may well affect the interpretation of data. Finally, as described in the introduction, the role of UmuDC proteins in the deletion event will be studied and there are no isogenic strains of RRI *umuDC* carrying *umuDC* mutants. For these reasons a set of strains from Graham Walker's laboratory were used in all of the remaining experiments.

The relative genotype of AB1157 *E.coli* along with two isogenic mutants; first AB1157^{recA} (known as JC2926) and second AB1157^{umuDC} (known as GW8017) are given in table 5.1. Since a wild type Mtase can not be used with these strains (owing to an *mcr*$^+$ conflict) the Kanamycin^r inactivated Mtase gene was used as a control throughout.

Table 5.1: The genotypes of the strains used in these experiments.

Strain	Genotype
HB101	*hsdS*20 (r_B^-,m_B^-), *recA*13, *rpsL*20 (str^r) *xyl*-5 *mtl*-1
RRI	Isogenic with HB101 except for carrying a *recA*$^+$
AB1157	r_B^+,m_B^+, *recA*$^+$
AB1157^{recA}	AB1157 *recA*13 (known as JC2926)
AB1157^{umuDC}	AB1157 *umuDC*::*cat*(595) (known as GW8017)

For full genotype see section 2.1.4

5.2.2 MUTATED MTASES USED TO TEST THE ROLE OF RECA AND UMUDC IN THE DELETION EVENT

The multispecific Mtase M.SPRI was used to study the role of RecA and UmuDC proteins in the deletion event provoked by $pGS2_{C78G}$. The advantage of employing M.SPRI is that it recognises three binding sites (CCGG, GGCC, and

CCWGG) and so the toxic effects are higher than observed with monospecific Mtase (see section 4.6). To examine the role of RecA and UmuDC in the deletion event, the experiments were repeated with dimerised monospecific Mtase and also with monomeric mono and multi-specific Mtases subcloned to pLitmus28.

5.2.3 DETERMINATION OF DELETION FREQUENCIES

In all of the experiments described here the mutated Mtase was re-activated by removing the kanamycin^r gene as described earlier (see section 4.4.1) followed by re-ligation of the Mtase gene. When using the mutant Mtases; 10 µl of the transformation mixture was plated on LA plates containing kanamycin (to check that all the transformants are not kanamycin resistant) the rest of the transformation mixture (990µl) was plated on LA plates containing ampicillin. Colonies that grow on ampicillin plates were counted and subsequently tested by re-plating them on kanamycin plates. For the control plasmids (pUMΔ1$_{C174G}$, pGM2Δ1$_{C174G}$, pES1K, and pGS2K) only 20% of the transformation mixture was plated out. Since the transformation efficiency varied between the three strains used, all the colony counts numbers were then adjusted according to the transformation efficiency obtained with pLitmus28. The experiments were repeated five times and the results shown are the mean of these experiments (figure 5.2).

5.3 FREQUENCY OF SURVIVAL ON E.COLI AB1157 TRANSFORMED WITH PLASMIDS ENCODING MUTATED MTASES IN MOTIF IV CYS TO GLY

pGS2$_{C78G}$ and pGS2K were used to transform AB1157 cells as described in section 2.2.6.4 and 5.2.3. After testing the recovered colonies 10 were found to be able to grow on ampicillin but not on kanamycin. The transformation efficiency of the AB1157 competent cells was $2X10^6$ and 125 colonies grew on the pGS2K plate. All the pGS2$_{C78G}$ derived transformants were picked and grown in 5 ml LB with ampicillin overnight. DNA was subsequently isolated and was digested with *BamHI*, *SalI*, and *XhoI*. After 1% agarose gel electrophoresis one

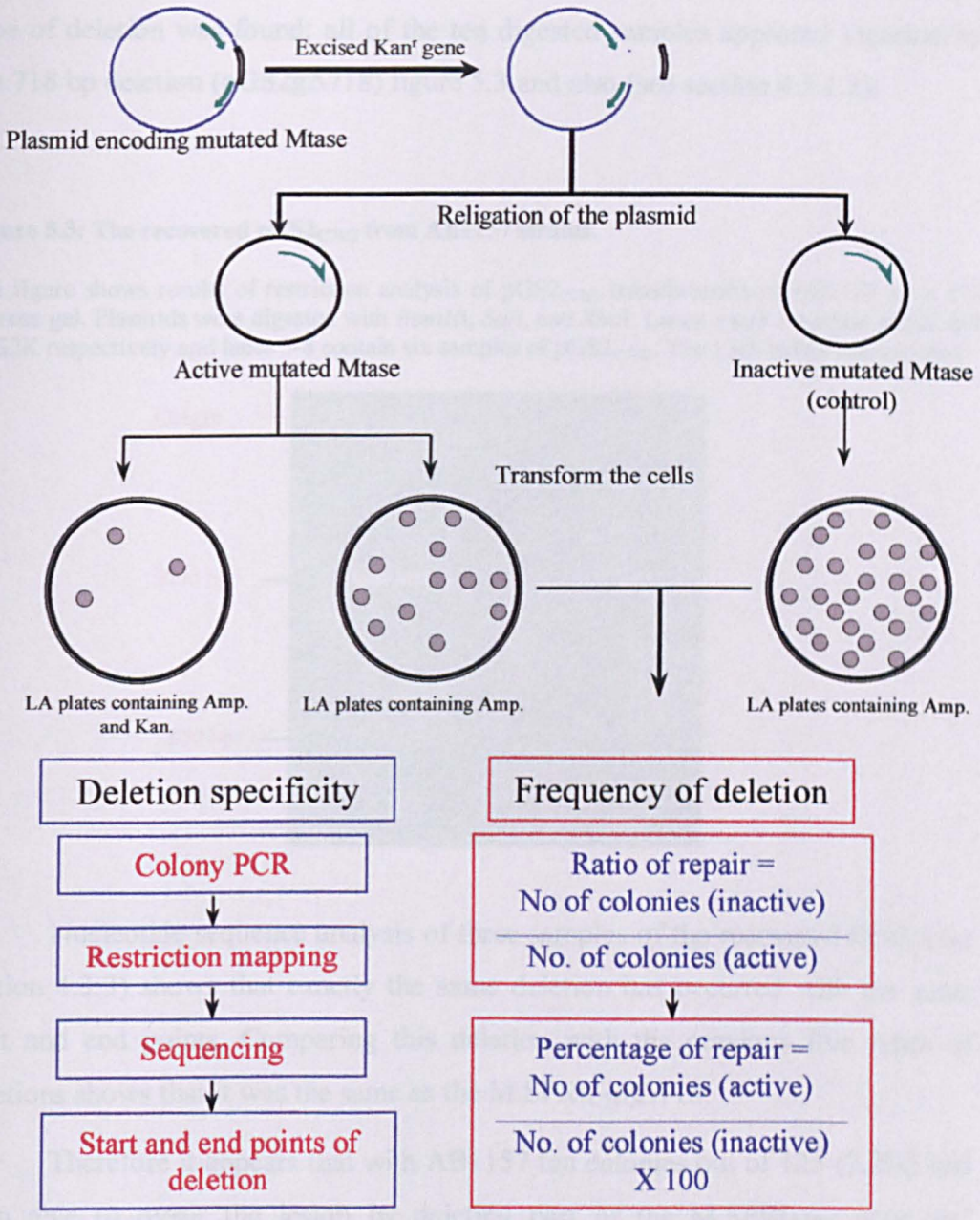
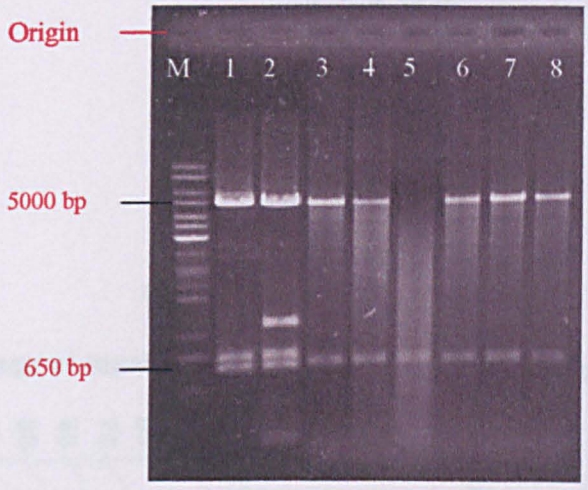


Figure 5.2 Experimental design to determine the frequency and nature of deletions. The mutated Mtases are reactivated by removing the kanamycin^r gene and are used to transform the required strain. Comparison between the transformation efficiencies with wild type and mutant Mtases provides measure of the frequency of deletion. Deletion specificity is determined by colony PCR, restriction mapping and nucleotide sequencing.

type of deletion was found: all of the ten digested samples appeared identical to the 718 bp deletion (pGS2gΔ718) figure 5.3 and also (see section 4.5.1.2).

Figure 5.3: The recovered $pGS2_{C78G}$ from AB1157 strains.

The figure shows results of restriction analysis of $pGS2_{C78G}$ transformants of AB1157 on a 1% agarose gel. Plasmids were digested with *BamH*I, *Sal*I, and *Xho*I. Lanes 1 and 2 contain pGS2 and pGS2K respectively and lanes 3-8 contain six samples of $pGS2_{C78G}$. The 1 Kb ladder marker used.



Nucleotide sequence analysis of three samples of the recovered DNA (see section 4.3.3) shows that exactly the same deletion has occurred with the same start and end points. Comparing this deletion with the previous five types of deletions shows that it was the same as the $M.SPRI_{C78G}$Δ718.

Therefore it appears that with AB1157 ten colonies out of 125 (2.2%) had been able to repair the lesion by deleting part of the $M.SPRI_{C78G}$ gene and subsequently inactivating the toxic protein.

The experiment was repeated using $pLS2_{C78G}$, $pLS1_{C78G}$ and a mono-specific Mtase (*Msp*I) using $pUM1_{C174G}$ for the monomer Mtase and $pGM2_{C174G}$ for the dimerised Mtase. Inactivated M.*Msp*I produced by introducing a single base deletion (see chapter 3) was used as a transformation control. The same methods described above were used and the numbers of colonies obtained are given in table 5.2 and plotted in figure 5.4.

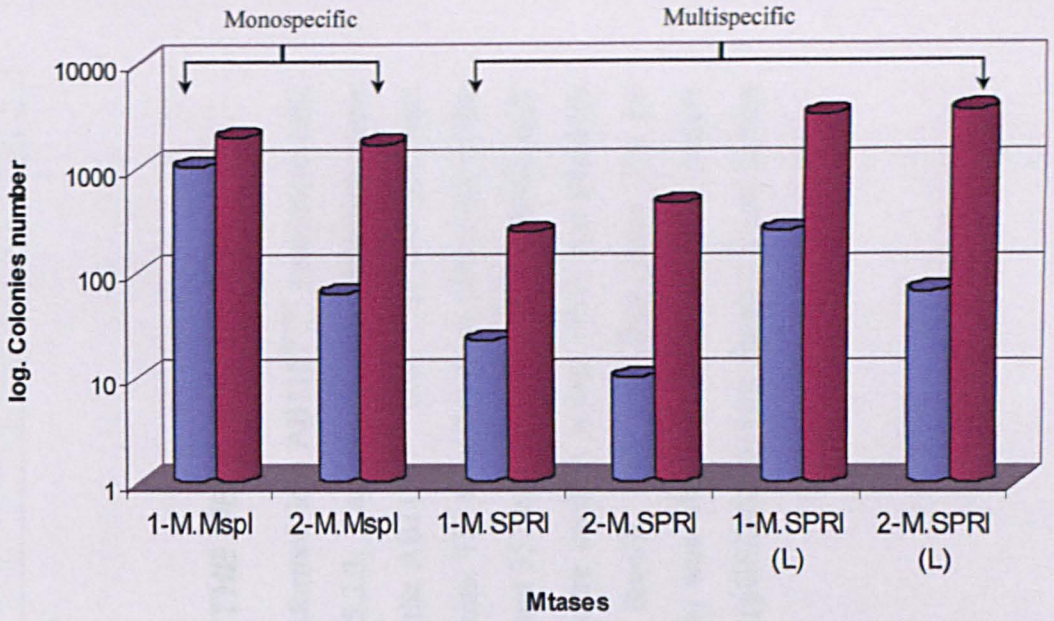
Figure 5.4: Number of colonies obtained after transformation AB1157 with different Mtases wild type and mutant plasmid.

Six types of mutant Mtases were used to check the ability of AB1157 *E.coli* cells ***(recA*** $^{+}$ and ***umuDC*** $^{+}$***)*** to repair the damage caused due to the high affinity binding between a Mtase and DNA. The inactive Mtases were $pUM1_{C174G}d1$, $pGM2_{C174G}d1$, $pES1_{C78G}K$, $pGS2_{C78G}K$, $pLS1_{C78G}K$, and $pLS2_{C78G}K$ respectively.

A) Logarithmic plot of the number of cells obtained versus the Mtases species.

B) Logarithmic plot of the ratio of repair for each Mtase. Repair ratio were calculated by dividing the number of colonies (see section 5.2.3) obtained with the control Mtase (inactive) by the number of colonies from an "active" mutant Mtase. Where (1) prior to the Mtase mean monomeric and (2) mean dimeric statues.

A



B

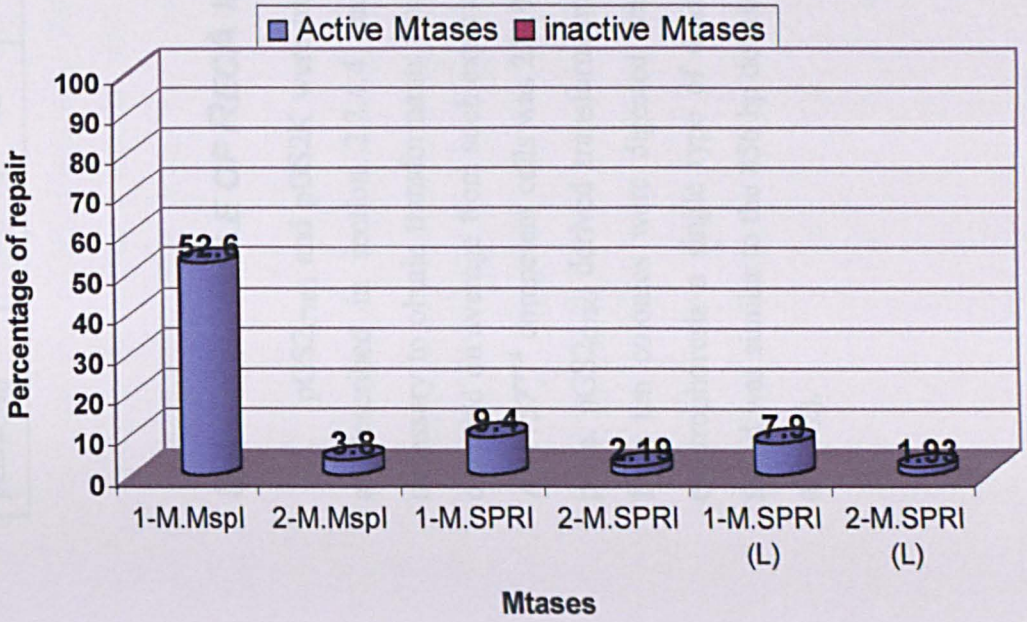


Table 5.2: Repair ratio of plasmids used to transform AB1157.

Six types of mutant Mtases were used to check the ability of *E.coli* with *recA* and *umuDC* alleles to repair the damage caused due to the high affinity binding between Mtase and DNA. The inactive Mtases were encoded in pUM1$_{C174G}$Δ1, pGM2$_{C174G}$Δ1, pES1$_{C78G}$K, pGS2$_{C78G}$K, pLS1$_{C78G}$K, and pLS2$_{C78G}$K respectively. The repair ratio was calculated by dividing the number of colonies (see section 5.2.3) obtained with the control Mtase (inactive) by the number of colonies obtained with the various active mutant Mtases. The percentage of repair was calculated by dividing the number of colonies from the active Mtase by the transformation number of colonies from the inactive Mtase transformation. The number of colonies, **No**, is the mean of five experiments and **sd** standard deviation.

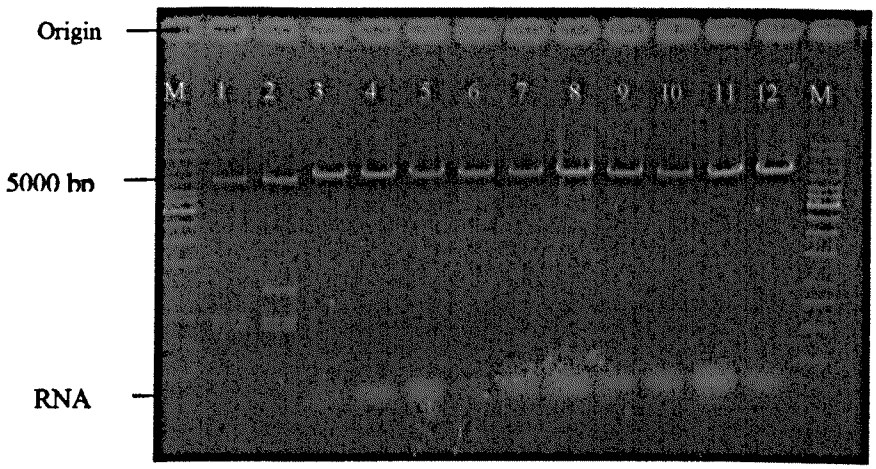
Plasmid	Active Mtases		Inactive Mtases		Ratio of repair	% of repair
	No	*sd*	*No*	*sd*		
pUM1$_{C174G}$	1033.3	57.7	1950	86.6	1 : 1.89	***52.6***
pGM2$_{C174G}$	63.3	15.2	1666.6	577.3	1 : 26.3	***3.8***
pES1$_{C78G}$	22.6	7	241.6	57.9	1 : 10.6	***9.4***
pGS2$_{C78G}$	10	2	456.6	30.1	1 : 45.6	***2.19***
pLS1$_{C78G}$	250	24	3160	557.5	1 : 12.6	***7.9***
pLS2$_{C78G}$	67	12	3467	550.8	1 : 51.7	***1.93***

5.4 ROLE OF RECA IN THE DELETION EVENT

pGS2$_{C78G}$ and pGS2K were transformed into AB1157^{recA} competent cells as described in section 2.2.4.4 and 5.2.3. Large-scale transformation were necessary to obtain transformants from the AB1157^{recA} host. Five colonies were obtained on average from such experiments. The transformation efficiency of the AB1157^{recA} competent cells was 2X10^5 and 357 colonies grew on the pGS2$_{C78G}$K plate. pGS2$_{C78G}$ derived transformants were used to isolate DNA and plasmids from ten colonies were digested with *BamHI*, *SalI*, and *XhoI*. After 1% gel electrophoresis a single type of deletion was obtained. The restriction pattern found was similar to the 856 bp deletion (pGS2gΔ856) (see figure 5.5 and section 4.5.1.2).

Figure 5.5: Analysis of the recovered pGS2$_{C78G}$ from AB1157^{recA} transformations.

The figure shows the results of restriction analysis of pGS2$_{C78G}$ transformants of AB1157^{recA} on a 1% agarose gel. Plasmids were digested with *BamHI*, *SalI*, and *XhoI*. Lanes 1 and 2 contain pGS2 and pGS2K respectively and lanes 3-12 contain ten samples of pGS2$_{C78G}$. The 1 Kb ladder was used as a marker.



Nucleotide sequence analysis of 3 samples of the recovered DNA (see section 4.3.3) showed that all contained exactly the same deletion with the same start and end points. Comparing this deletion with the previous five types of deletions shows that it was the same as M.SPRI$_{C78G}$Δ856. The experiment was repeated using other Mtases (see section 5.3) using the same methods described above and the numbers of colonies obtained are given in table 5.3 and plotted in figure 5.6.

Table 5.3: Repair ratio and percentage of repair of plasmids used to transform AB1157^{RecA}.

Six types of mutant Mtases were used to check the ability of *E.coli* with ***recA***$^-$ and ***umuDC*** alleles to repair the damage caused due to the high affinity binding between Mtase and DNA. The inactive Mtases were encoded in pUM1$_{C174G}$Δ1, pGM2$_{C174G}$Δ1, pES1$_{C78G}$K, pGS2$_{C78G}$K, pLS1$_{C78G}$K, and pLS2$_{C78G}$K respectively. The repair ratio was calculated by dividing the number of colonies (see section 5.2.3) obtained with the control Mtase (inactive) by the number of colonies obtained with the various active mutant Mtases. The percentage of repair was calculated by dividing the number of colonies from the active Mtase by the transformation number of colonies from the inactive Mtase transformation. The number of colonies, **No**, is the mean of five experiments and **sd** standard deviation.

Plasmid	Active Mtases		Inactive Mtases		Ratio of repair	Percentage of repair
	No	*sd*	*No*	*sd*		
pUM1$_{C174G}$	166.6	117.1	1866.6	230.9	1 : 11.2	***8.9***
pGM2$_{C174G}$	3.6	2.6	666.6	305.5	1 : 185	***0.54***
pES1$_{C78G}$	1	0	193.6	73.1	1 : 193.6	***0.51***
pGS2$_{C78G}$	1	0	357.3	39.6	1 : 357.3	***0.27***
pLS1$_{C78G}$	16	6	1853	362.3	1 : 115.8	***0.86***
pLS2$_{C78G}$	9	1	3245	817.8	1 : 360	***0.27***

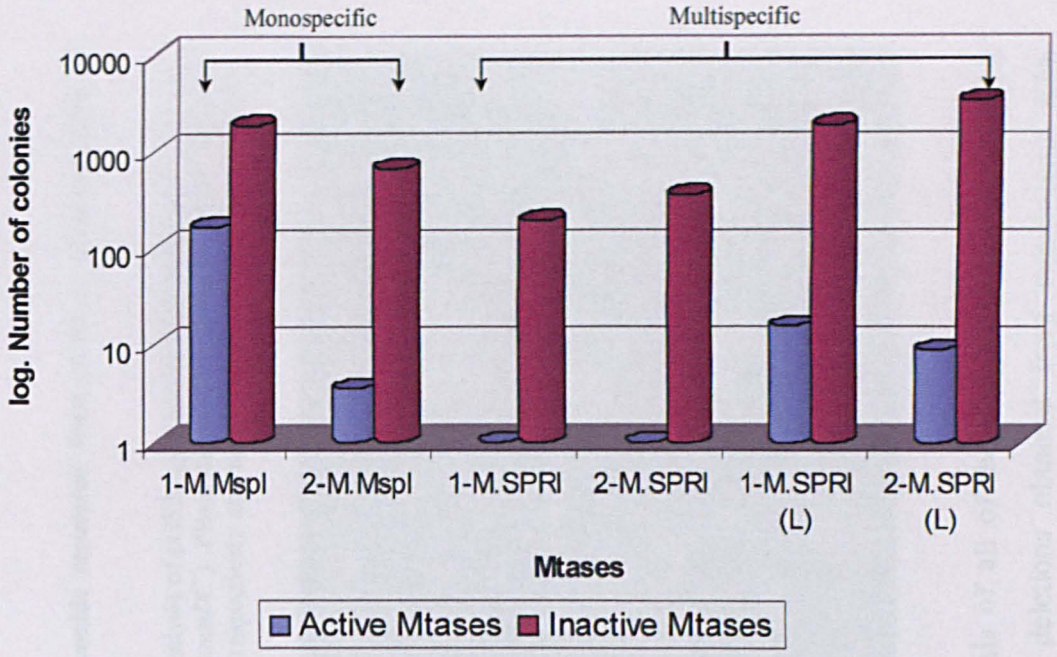
Figure 5.6: Number of colonies obtained after transformation AB1157^{recA} with different Mtases wild type and mutant plasmid.

Six types of mutant Mtases were used to check the ability of AB1157 *E.coli* cells ***(recA$^-$*** and ***umuDC$^+$)*** to repair the damage caused due to the high affinity binding between a Mtase and DNA. The inactive Mtases were pUM1$_{C174G}$d1, pGM2$_{C174G}$d1, pES1$_{C78G}$K, pGS2$_{C78G}$K, pLS1$_{C78G}$K, and pLS2$_{C78G}$K respectively.

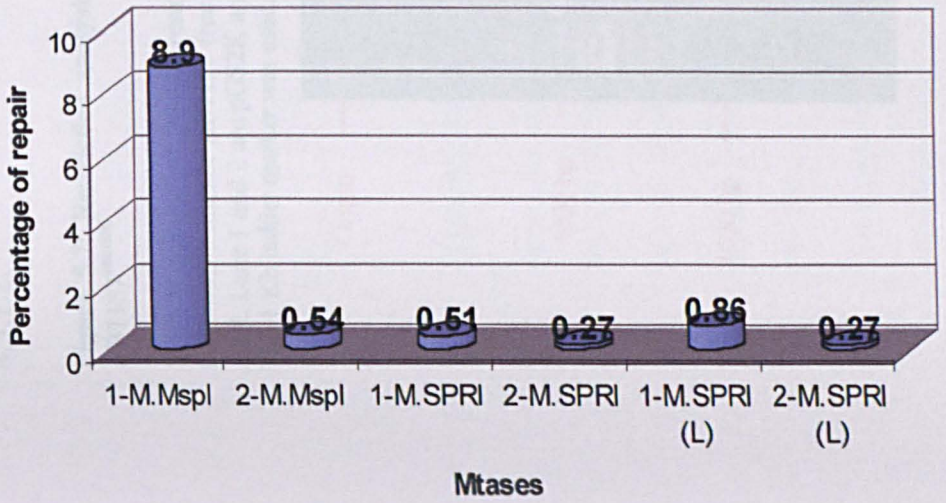
A) Logarithmic plot of the number of cells obtained versus the Mtases species.

B) Logarithmic plot of the ratio of repair for each Mtase. Repair ratio were calculated by dividing the number of colonies (see section 5.2.3) obtained with the control Mtase (inactive) by the number of colonies from an "active" mutant Mtase. Where (1) prior to the Mtase mean monomeric and (2) mean dimeric statues.

A



B



$pGS2_{C78G}\Delta 6$, $pGS2_{C78G}\Delta 32$, and $pGS2_{C78G}\Delta 645$ and the percentage of each type of the deletions was 11.1, 33.3, 22.2, respectively and 22.2 for the rearrangement.

The experiment was repeated using other Mtases (see section 5.3) using the same methods described above and the numbers of colonies obtained are given in table 5.4 and plotted in figure 5.8.

Table 5.4: Repair ratio and percentage of repair of plasmids used to transform AB1157^{umuDC}.

Six types of mutant Mtases were used to check the ability of *E.coli* with ***recA***$^+$ and ***umuDC***$^-$ alleles to repair the damage caused due to the high affinity binding between Mtase and DNA. The inactive Mtases were encoded in $pUM1_{C174G}\Delta 1$, $pGM2_{C174G}\Delta 1$, $pES1_{C78G}K$, $pGS2_{C78G}K$, $pLS1_{C78G}K$, and $pLS2_{C78G}K$ respectively. The repair ratio was calculated by dividing the number of colonies (see section 5.2.3) obtained with the control Mtase (inactive) by the number of colonies obtained with the various active mutant Mtases. The percentage of repair was calculated by dividing the number of colonies from the active Mtase by the transformation number of colonies from the inactive Mtase transformation. The number of colonies, **No**, is the mean of five experiments and **sd** standard deviation.

Plasmid	Active Mtases		Inactive Mtases		Ratio of repair	% of repair
	No	*sd*	*No*	*sd*		
$pUM1_{C174G}$	533.3	152.7	1700	264.5	1 : 3.2	***31.4***
$pGM2_{C174G}$	26.6	11.1	966.6	251.6	1 : 36.3	***2.75***
$pES1_{C78G}$	15.3	4.1	238.3	61.7	1 : 15.6	***6.4***
$pGS2_{C78G}$	5	2	452.3	33.8	1 : 90.5	***1.1***
$pLS1_{C78G}$	85	25	2807	275.9	1 : 33	***3***
$pLS2_{C78G}$	29	20.1	3214	585.4	1 : 110.8	***0.9***

5.6 SUMMARY OF RESULTS

1- *recA* and *umuDC* gene products affect the quality and quantity of repair but are not essential for error-prone repair.

2- *umuDC* products provoke single deletion events.

3- Mutant Mtase dimers induced lower repair efficiencies than monomeric mutants.

4- The frequency of repair is inversely proportional to frequency of TRSs in the plasmid encoding the mutant Mtase.

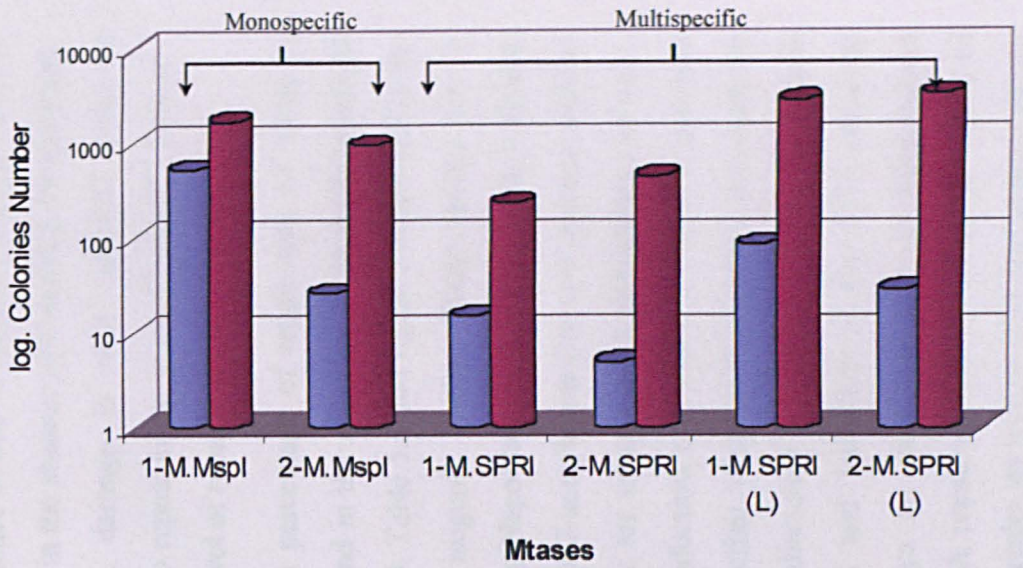
Figure 5.8: Number of colonies obtained after transformation AB1157^{umuDC} with different Mtases wild type and mutant plasmid.

Six types of mutant Mtases were used to check the ability of AB1157 *E.coli* cells ***(recA$^+$*** and ***umuDC$^-$)*** to repair the damage caused due to the high affinity binding between a Mtase and DNA. The inactive Mtases were pUM1$_{C174G}$d1, pGM2$_{C174G}$d1, pES1$_{C78G}$K, pGS2$_{C78G}$K, pLS1$_{C78G}$K, and pLS2$_{C78G}$K respectively.

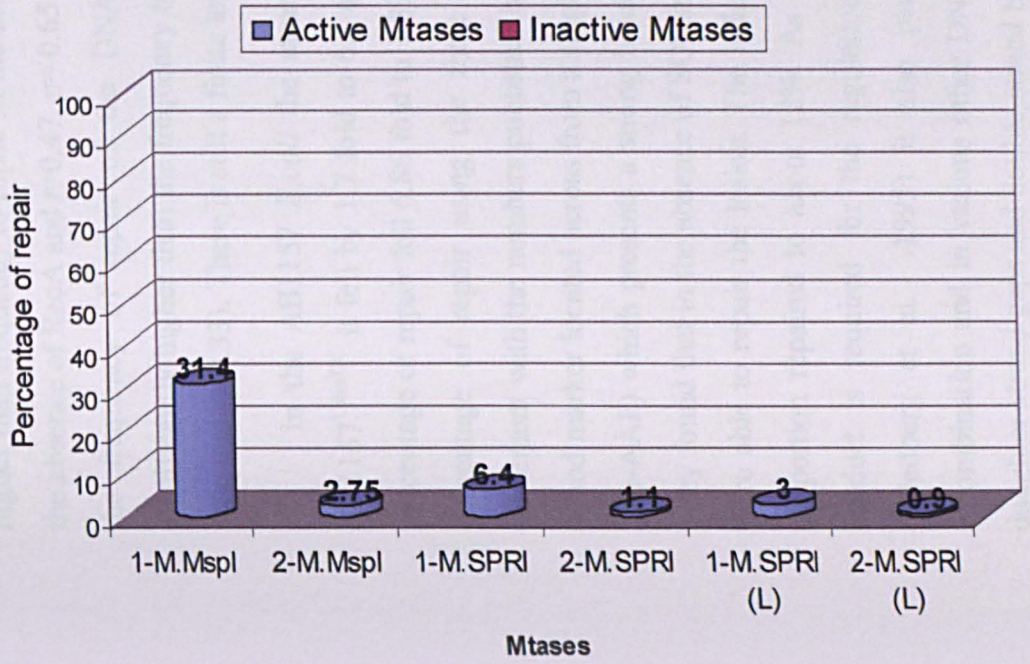
A) Logarithmic plot of the number of cells obtained versus the Mtases species.

B) Logarithmic plot of the ratio of repair for each Mtase. Repair ratio were calculated by dividing the number of colonies (see section 5.2.3) obtained with the control Mtase (inactive) by the number of colonies from an "active" mutant Mtase. Where (1) prior to the Mtase mean monomeric and (2) mean dimeric statues.

A



B



5.7 DISCUSSION

Most of the molecular mechanisms of the repair enzymes in *E.coli* are still obscure especially those related to the SOS response and SOS mutagenesis. In this chapter I set out to explain the involvement of *umuDC* and *recA* gene products in the deletion event. The experiments carried out in this chapter provide evidence that the deletion phenomenon is related to known repair genes since the viability of the cells and the deletion types varies according to the presence of certain DNA repair genes. It would also seem that the presumed Mtase:DNA complex formed through the use of a mutated Mtases interferes with replication leading to error-prone repair.

It is possible to deduce from the results listed in this chapter that the deletion event can occur in the presence and absence of RecA or UmuDC proteins. The frequency of DNA repair in $recA^{+}$ $umuDC^{+}$ hosts are significantly higher than frequency to repair in the absence of one of them (t= 1.19, p= 0.28 in the absence of RecA and t=0.47, p= 0.65 in the absence of UmuDC). Furthermore, the frequency of repair of the DNA damage in $recA^{+}$ $umuDC^{-}$ hosts is significantly higher than the frequency to repair in the in $recA^{-}$ $umuDC^{+}$ host (t= 1.06 and p=0.33). There is still a finite level of repair even without RecA.

In the AB1157 *E.coli* the mean percentage of repair was 13 while in $AB1157^{UmuDC}$ it fell by 1.7 fold to 6.7 and in the $AB1157^{RecA}$ strain the mean of percentage of repair fell 6.86 fold to 1.89. Table 5.5 and figure 5.8 summarise the percentage of repair using the three isogenic strains. These results are in agreement with the numbers published by Napolitano et al (1997) using a genetic strand marker located across from a single 2-acetylaminofluorene guanine adducts (dG-AAF) which presents a strong block for various DNA polymerases *in vitro*. They found that in the absence of SOS induction, only about 0.5% of the colonies were able to repair the lesion. The induction of the SOS response increased the proportion repaired to about 12%. As discussed in section 5.1 the *recA* gene product is required for the regulation and induction of the SOS response (Friedberg et al. 1995) it also plays an important role in homologous recombination and in various other DNA repair events, including the repair of daughter-strand gaps and double-strand breaks, as well as SOS mutagenesis (Roca and Cox, 1997). This perhaps explains the very low percentage of repair in those

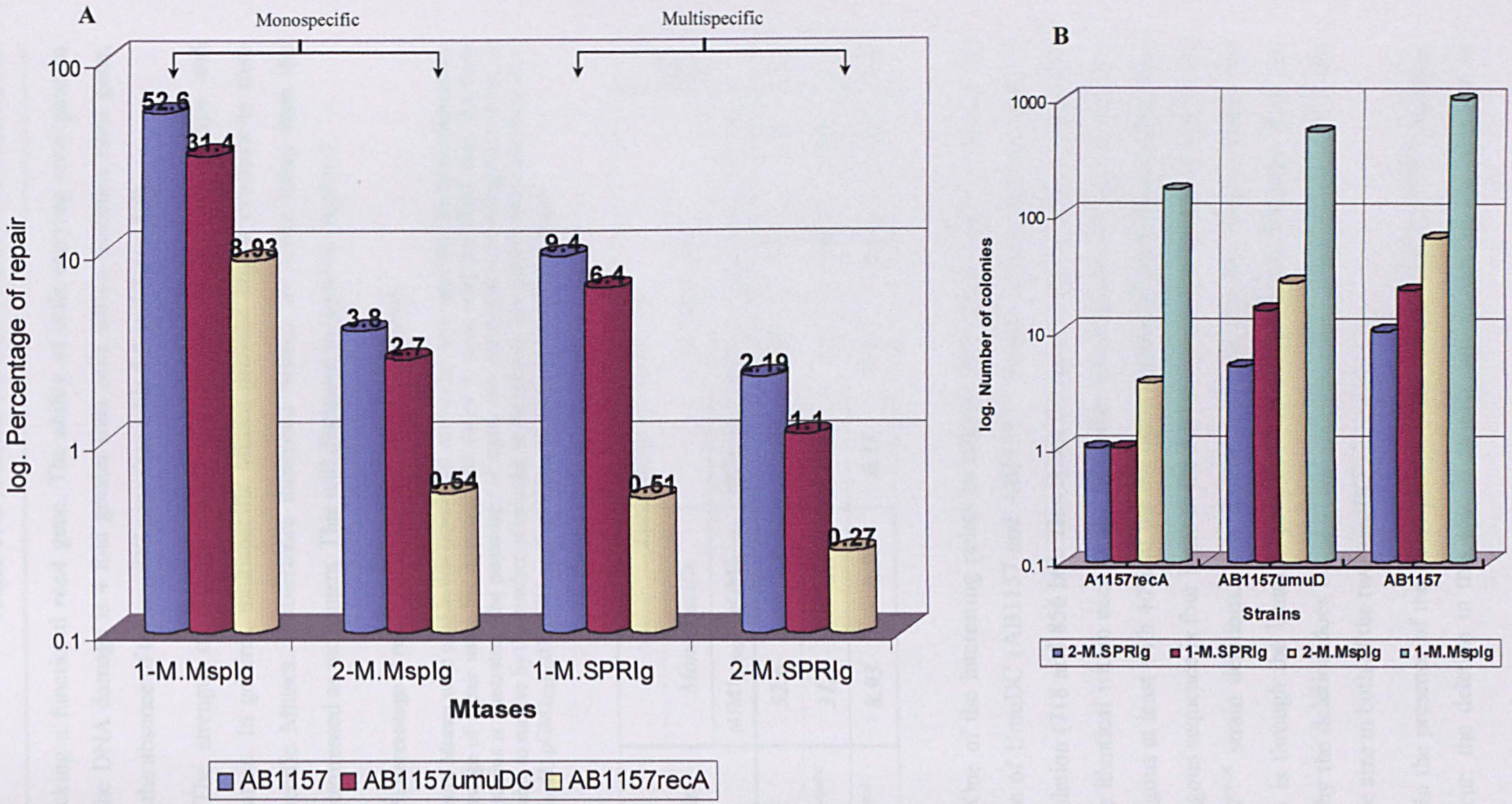


Figure 5.9: Percentage of repair using three strains and different Mtases.

Percentage of repair using multi and mono-specific Mtases in the monomeric and dimeric form. A) logarithmic plot verse the Mtases type the percentage of repair was calculated by dividing the number o colonies able to grow in the plate transformed with mutated Mtase by the control one. B) logarithmic plot of the number of colonies versus the strains. Where (1) prior to the Mtase mean monomeric and (2) mean dimeric statues.

cells lacking a functional *recA* gene. The ability of cells carrying *recA* gene to repair the DNA damage is 4 fold greater than cells with a defective *recA* gene, even in the absence of the *umuDC* gene (see table 5.5 and figure 5.9).

The strength of damage varies with respect to the Mtase type and dimerisation. In general, multispecific Mtases provoke higher cytotoxicity than monospecific Mtases. Furthermore dimerised Mtases are more lethal than the Mtases expressed as monomers. This will discussed in detail in chapter 7.

Table 5.5: Percentage of repair using different strains and Mtases.

Two types of Mtases were used in this experiment, mono and multi-specific. The table summarises the percentage of repair using the monomer and dimer in each case; the effect upon the three isogenic strains is illustrated. The percentage of repair was calculated by dividing the number of colonies able to survive by the number of colonies in the control. The control Mtase plasmids were pUM1$_{C174G}$d1, pGM2$_{C174G}$d1, pES1$_{C78G}$K, pGS2$_{C78G}$K, pLS1$_{C78G}$K, and pLS2$_{C78G}$K.

Strain	*Percentage of repair (%)*						Mean
	Monospecific		*Multispecific*				
	pUM1$_{C174G}$	pGM2$_{C174G}$	pES1$_{C78G}$	pGS2$_{C78G}$	pLS2$_{C78G}$	pLS1$_{C78G}$	
AB1157	***52.6***	***3.8***	***9.4***	***2.19***	***7.9***	***1.93***	***13***
AB1157^{umuDC}	***31.4***	***2.7***	***6.4***	***1.1***	***3.03***	***0.9***	***7.6***
AB1157^{recA}	***8.93***	***0.54***	***0.51***	***0.27***	***0.86***	***0.27***	***1.9***

One of the interesting points to emerge from these results is that in the presence of UmuDC (AB1157 and AB1157^{recA} strains) there is a single type of large deletion (718 and 856 bp). This is not the same in the two strains, however they are identical within each strain. Moreover the deletions start-end points are homologous at least with 40% of the ten bases sequenced after the deletion. This homologous sequence is four bases in the AB1157 strain deletions and 8 bp in the AB1157^{recA} strain deletions. This finding indicates that the main pathway for deletion is through the UmuDC protein (SOS mutagenesis). However it is not clear why the deletion does not start and end from the same sequence and is not the same size in both of the two *umuDC*$^{+}$.

In the presence of the RecA protein (AB1157 and AB1157^{umuDC}) deletion still occur: the deletion in this case is precise and unique in the presence of

umuDC (AB1157 strain). However if there is no *umuDC* gene products, the deletions vary from 6 bp to 645 bp and rearrangements appear. Table 5.6 summarises the deletion types, start end points and their sequences.

Table 5.6: Start and end points of deletions obtained following transformation of repair-deficient strains of *E.coli*.

The five type of deletion and there start-end points sequences are presented, the blue coloured bp are homologous with the sequence just after the deletion in the wild type M.SPRI gene. In the cells expressing wild type *umuDC*, a single type of deletion appears. In the cells proficient in *recA* but deficient in *umuDC*, there were three types of deletions and one type of rearrangement.

Gene	Relative of genotype the host cells	Sequences before deletion	Sequences after deletion	Homology of the first 10 bp (%)	% of the colonies harbouring the deletion
pGS2$_{C78G}$Δ718	***recA*$^+$ *umuDC*$^+$**	TCCAC	AAAAGCTTTT	40	100
pGS2$_{C78G}$Δ856	***RecA13 umuDC*$^+$**	AATCC	AGCAAGTAGG	80	100
pGS2$_{C78G}$Δ6	***recA*$^+$ *umuDC*$^-$**	AGAGT	AATACAGAA	30	11.1
pGS2$_{C78G}$Δ32	***recA*$^+$ *umuDC*$^-$**	GTCGA	TAACATTAAT	40	33.3
pGS2$_{C78G}$Δ645	***recA*$^+$ *umuDC*$^-$**	GTAGC	CGGGCTTGGA	70	22.2

AB1157^{RecA} used in this experiment carried *recA*13 mutant allele, which is mutated in the highly conserved MAW motif (L51F). MAW is a region encompassing residues 42 to 65 and is defined as a unique RecA structural motif with the acronym MAW, for Makes ATP Work (Roca and Cox, 1997).*E.coli* harbouring *recA13* is sensitive to UV light, is recombination defective, cannot induce λ prophage or participate in the autocatalytic cleavage of the LexA repressor (Roca and Cox, 1997) and lack SOS error-prone DNA repair (Howard et al. 1993). In the AB1157 both SOS mutagenesis and recombination repair are active, and in the AB1157^{UmuDC} strain SOS mutagenesis is defective. AB1157^{RecA} lacks SOS mutagenesis and recombination repair activities. As described above (section 1.4.5 and 5.1.1) RecA protein has at least three roles in SOS mutagenesis. First, some of the required proteins are induced as part of the SOS response. Second, at least one of the required proteins is proteolytically processed in a reaction facilitated by RecA. Finally, RecA participates directly in lesion bypass (Roca and Cox, 1997). At least the first and second roles are defective in AB1157^{RecA}. Therefore, accordingly the very low levels of repair observed in the

AB1157^{RecA} was not brought about the standard SOS mutagenesis pathway even in the presence of *umuDC* gene products. However, since the precise molecular role of RecA in SOS mutagenesis remains obscure it is not clear what system was used to repair the macro-lesion in the AB1157^{RecA} (see figure 5.10). Furthermore, the similarity of results from transformation of AB1157 and AB1157^{RecA} (see figure 5.9 and table 5.6) indicate that the main pathway of repair is mediated through *umuDC* gene products, and this repair is more efficient in the presence of the *recA* gene product (13%) than in its absence (1.89%) (see figure 5.10).

The results given above, indicate that the deletion event can proceed through the SOS mutagenesis pathway and this is the main deletion pathway. Furthermore, if this pathway is blocked, the repair proceeds through a recombination repair system (figure 5.9). In their study (Barbe et al. 1986); using 5-azaC to test the survival of *E.coli* and the role of the SOS response in *E.coli* K-12 they report that a *recA13* mutant was more sensitive to 5-azaC than cells proficient in *recA*. On the other hand, 5-azaC was unable to trigger the induction of the *umuDC* operon and no amplification of RecA protein synthesis in either *dcm*$^+$ or *dcm*$^-$ strains was observed. These authors concluded that there was discrimination in the expression of the various SOS genes and that some SOS genes may be induced without amplification of RecA protein synthesis (Barbe et al. 1986). These results may explain the presence of very low level of repair in AB1157^{RecA} however, it is still unclear which pathway is used to repair this damage in the experiments reported here.

One further aspect of the major deletion repair pathway that the homologies in the start and end points of the deletion indicate that this process may take place through a DNA recombination pathway. However, the highest homology of ends (80%) was observed in the AB1157^{RecA} experiments (table 5.6) and as mentioned above, this strain lacks the major recombination protein, RecA, which implies that deletion depends on another pathway. On the other hand the deletions observed in the AB1157^{UmuDC} transformations were without homologies: the presence of rearrangements among the deletions on the AB1157^{UmuDC} however supports the notion that some form of recombination repair has taken place.

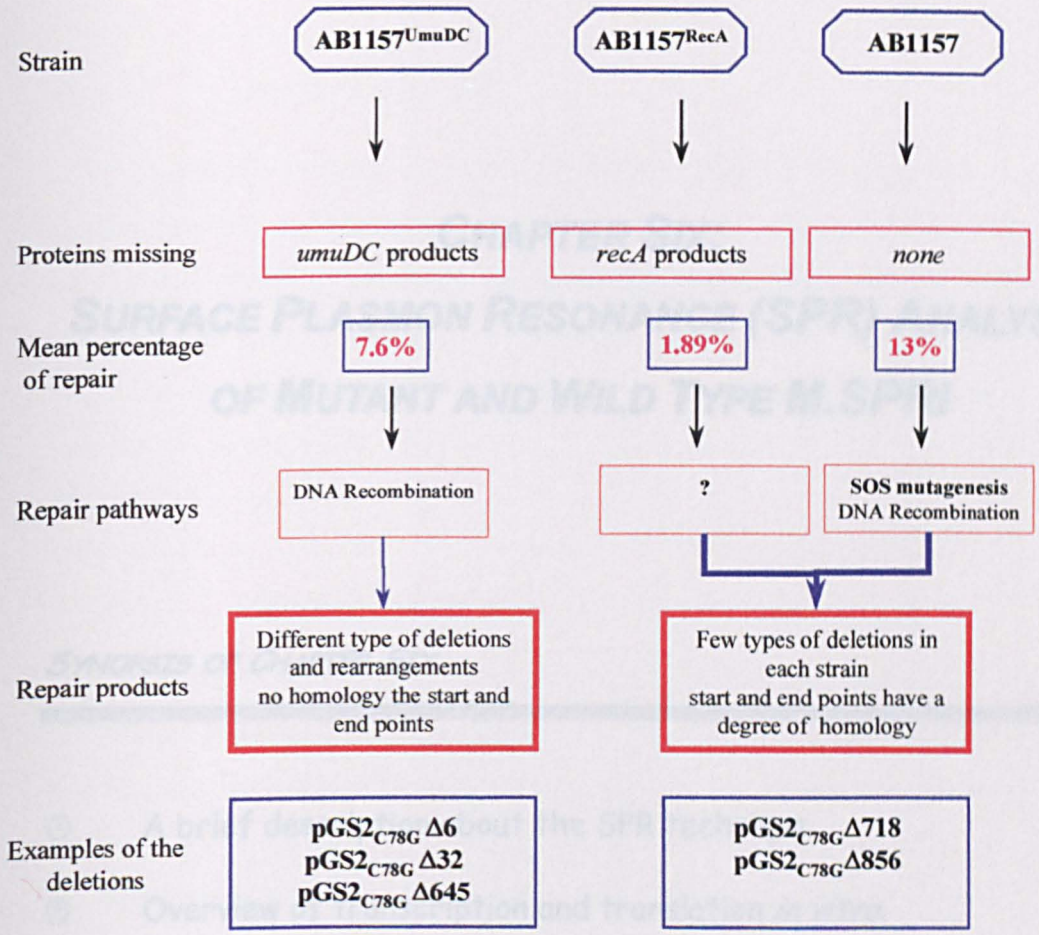


Figure 5.10: Summary of the data provided in chapter 5.

The flow diagrams indicate the types of deletions recovered following transformation of three related *E.coli* strains which are proficient in the DNA recombination and error-prone repair pathway.

CHAPTER SIX:
SURFACE PLASMON RESONANCE (SPR) ANALYSIS OF MUTANT AND WILD TYPE M.SPRI

SYNOPSIS OF CHAPTER SIX

- A brief description about the SPR technique.
- Overview of transcription and translation *in vitro*.
- Calculation of M.SPRI wild type k_{off}.
- Calculation of mutant M.SPRI k_{off}.
- Conclusion of chapter Six.

6.1 INTRODUCTION

Cytotoxic proteins are often difficult to isolate in the quantities necessary for biochemical and biological analysis. In order to overcome this problem, the cytotoxic M.SPRI$_{C78G}$ has been synthesised *in vitro* using a commercial protein synthesis kit from Promega. The synthetic polypeptide can be produced in sufficient quantities for SPR analysis. Here, experiments are described which allow a direct comparison to be made between the binding affinity for DNA of wild type and mutant Mtases.

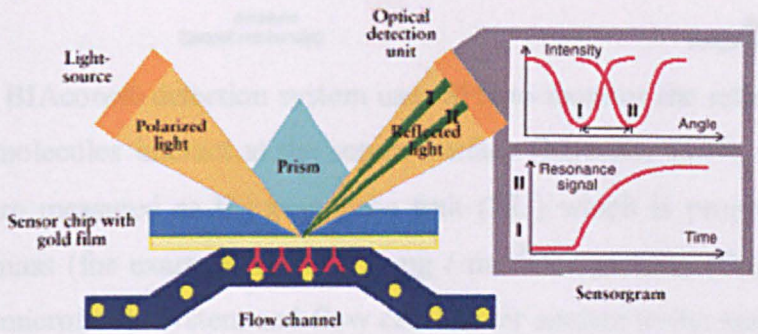
6.1.1 *SURFACE PLASMON RESONANCE (SPR)*

Specific interactions of macromolecules such as proteins and DNA provide a chemical foundation for all cellular processes. Studying dynamics of macromolecular interactions is required to define the mechanism of binding. This kinetic can now be obtained using surface plasmon resonance (**SPR**) biosensors. In 1990 BIAcore® Inc. produced the first commercial biosensor. These instruments allowed the real time analysis of reactions without the need for radioisotopes, moreover the methods requires small amounts of sample and is fully automated (see figure 6.1).

In biological SPR assays, a thin film of metal is applied to a grating which has been moulded into a substrate (typically glass or plastic). The metal film is between 35 and 200 nm thick. In one example, antibodies specific to a particular analyte (e.g. a hormone, drug, protein, nucleotides etc.) are chemically attached to the metal film. When the sensor is exposed to a sample containing that analyte, the binding of the antibody and the analyte causes a change at the metal surface, within the plasmon field, and the shift in the resonant wavelength of the incident light is measured. The size of the shift is proportional to the quantity of the analyte in the sample. Because of the very specific relationship between the antibody and analyte no other molecules in the sample can be mistakenly measured by the sensor (see figure 6.1).

Figure 6.1: Following biomolecular binding in real time using the BIAcore SPR instrument.

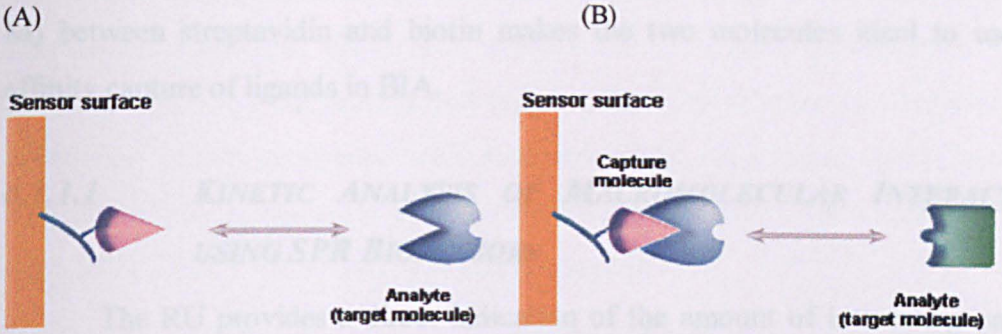
During a binding experiment changes occur as a solution is passed over the surface of a sensor chip. To perform an analysis, one reactant is captured on a sensor surface. The sensor surface forms one wall of a flow cell. Sample containing the other reactant (s) is injected over this surface in a precisely controlled flow. Fixed wavelength light, in a fan-shaped form, is directed at the sensor surface and biomolecular-binding events are detected as changes in the particular angle where SPR creates extinction of light. This change is measured continuously to form a sensorgram, which provides a complete record of the progress of association and dissociation of the reactants.



Biomolecular interaction analysis (BIA) is a label-free technology that monitors molecular binding processes on the surface of a specially prepared biosensor chip. One of the interacting partners (ligand) is attached to the chip and a solution containing the other partner (analyte) is passed over the surface. When the analyte (A) and the ligand (B) interact to form a complex (AB), this biomolecular binding events cause changes in the refractive index at the surface layer, which are detected as changes in the SPR signal. In general, the refractive index change for a given change of mass concentration at the surface layer, is practically the same for all proteins and peptides, and is similar for glycoproteins, lipids and nucleic acids. An analysis can only be as good as the surface on which it is performed, so sensor surfaces must meet the highest demands for precision, stability and reproducibility. Sensor surfaces can be functionalised to detect a wide selection of analytes, and used in direct capture or affinity capture modes. Because binding responses are recorded in real time it is possible to derive kinetic information from the interaction (see figure 6.2).

Figure 6.2: The ligand and analyte in BIAcore SPR.

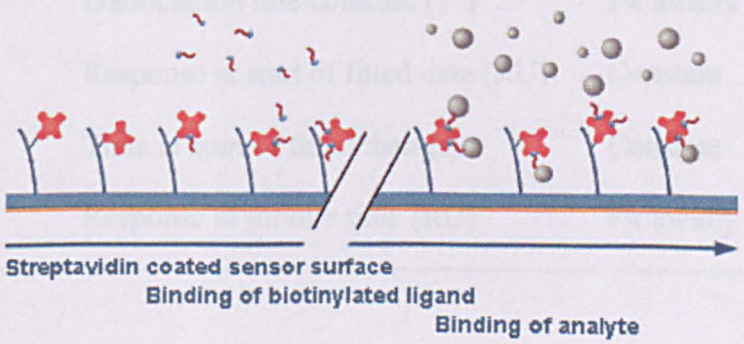
a) Covalent attachment of ligand. b) Reversible attachment of ligand through the streptavidin binding.



The BIAcore® detection system uses SPR to monitor the refractive index change as molecules interact at the sensor surface (Myszka, 1997). The signals generated are measured as the resonance unit (RU) which is proportional to a change in mass (for example 1 RU = 1 pg / mm^2 for protein) (Stenberg et al. 1991). The microfluids system and flow cell deliver analyte to the sensor surface, which is typically coated with a carboxymethyldextran matrix. This dextran matrix provides a flexible anchor for ligand immobilisation, allowing interactions to occur as in solution.

There are many different chips that can be used in BIA one of which is **Sensor Chip SA** with pre-immobilised streptavidin which is designed as a general purpose sensor chip for the capture of biotinylated ligand (figure 6.3). Sensor chip SA has a streptavidin-coated surface ready to bind biotinylated ligands. The streptavidin is covalently linked to a carboxymethylated dextran matrix, optimised for high binding capacity for long nucleic acid fragments (200 bp) and other molecules.

Figure 6.3: Streptavidin covalently linked to the dextran matrix allows convenient capture of biotinylated ligands.



The sensor chip is designed to give high reproducibility. It is also highly resistance to a range of harsh conditions including the pH range 2-13, providing flexibility in the choice of regeneration agents. The remarkable affinity ($K_D \approx 10^{-15}$ M) between streptavidin and biotin makes the two molecules ideal to use for affinity capture of ligands in BIA.

6.1.1.1 KINETIC ANALYSIS OF MACROMOLECULAR INTERACTION USING SPR BIOSENSORS

The RU provides a direct indication of the amount of interacting partner that binds to an immobilised molecule from solution. A response indicates binding, so the technique is used to detect the presence of binders and to investigate the specificity of binding. The shape of the binding curve reflects the kinetics of the interaction-association phase while sample is present and dissociation after the sample has been removed. The level of binding reached at steady state depends on the affinity of the interaction (figure 6.4).

In this experimental section sensorgrams are used to calculate the dissociation rate using BIAevaluation 3.0 programme provided by the instrument supplier. The programme used one to one (Langmuir) dissociation, which, fits the dissociation phase of a 1:1 interaction to the integrated rate equation:

$$\mathbf{R = R_0 e^{-k_{off}(t-to)} + Offset}$$

The Offset term allows the dissociation phase to approach a non-zero baseline. The equation parameters are described in table 6.1.

Table 6.1: Dissociation rate constant parameters.

Parameter	Description	Fitting status
kd	Dissociation rate constant (s^{-1})	Fit locally
R0	Response at start of fitted data (RU)	Constant
t0	Time at start of fitted data (s)	Constant
Offset	Response at infinite time (RU)	Fit locally

6.1.2 TRANSCRIPTION AND TRANSLATION IN VITRO

The TNT® Quick Coupled Transcription/Translation System is a single-tube coupled transcription/translation reactions for eukaryotic *in vitro* translation (Perara and Lingappa, 1985). The original TNT® Coupled Reticulocyte Lysate Systems simplified the process and reduced the time required to obtain *in vitro* translation results compared with standard rabbit reticulocyte lysate systems. Standard rabbit reticulocyte systems commonly use RNA synthesised *in vitro* from SP6, T3 or T7 RNA polymerases (Zubay, 1973).

The TNT® Quick Coupled Transcription/Translation System required the genes to be cloned downstream from the T7 or SP6 RNA polymerase promoters, it also required 0.2-2.0μg of DNA (either circular plasmid DNA or linear DNA such as a PCR fragment). A luciferase-encoding control plasmid and Luciferase Assay Reagent supplied with the TNT® Quick Coupled Transcription/Translation system, which can be used in a non-radioactive assay for rapid (<30 seconds) detection of functionally active luciferase protein.

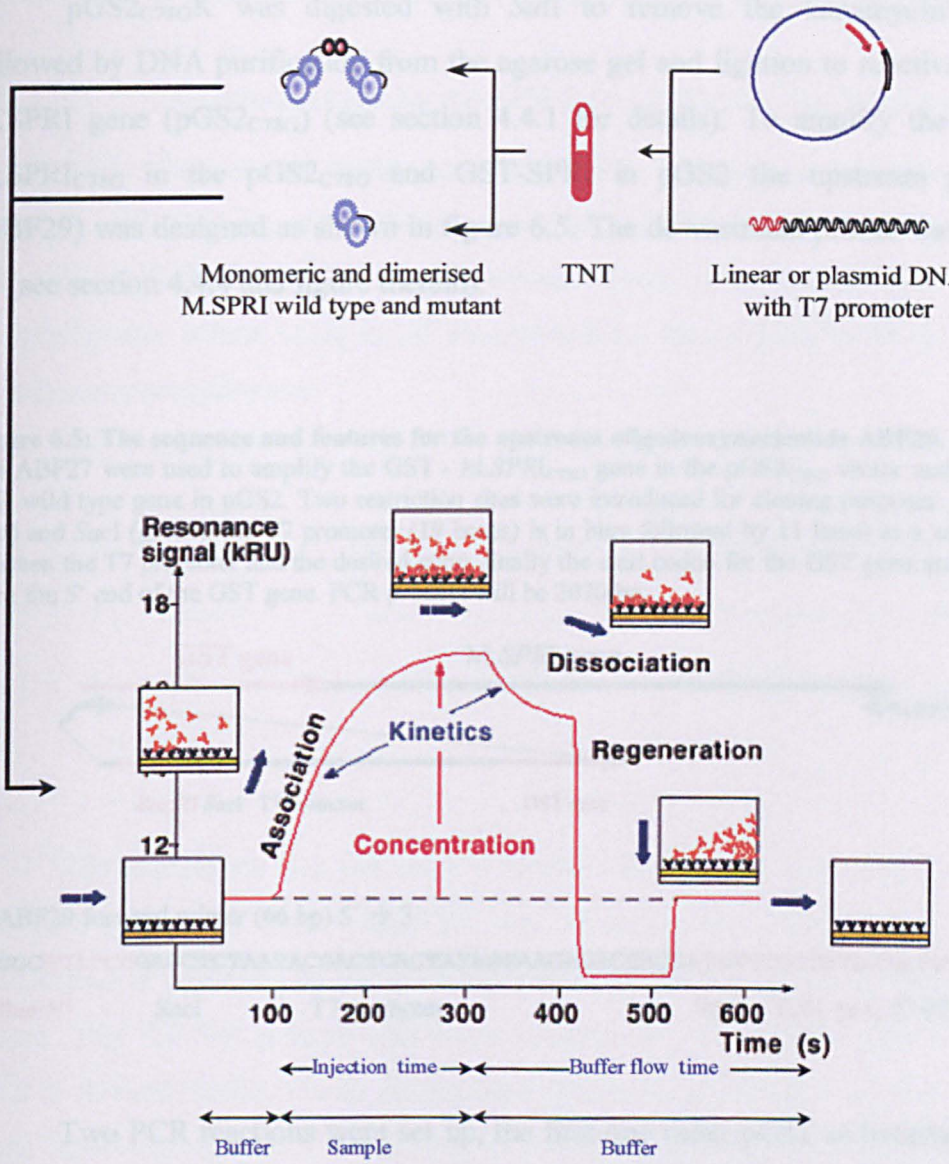
6.2 EXPERIMENTAL DESIGN

6.2.1 THE DNA TEMPLATE FOR THE PRODUCTION OF PROTEIN IN VITRO

Since the TNT® Quick Coupled Transcription/Translation System requires the genes to be cloned downstream from a T7 promoter, and the pGEX2T vectors contains P_{tac} promoter rather than T7, M.SPRI wild type and mutated genes were ligated into the pET22b vector (pET vectors are recommended by the TNT® Quick Coupled System manufacturers). Alternatively a PCR fragment of the required gene was generated with primers designed to introduce the T7 promoter upstream of the gene.

As described in the introduction above, this experiment is designed to study the difference in binding affinity between the wild type and mutant multispecific Mtase in the monomeric and dimerised forms. In the monomeric form pES1$_{C78G}$K and pES1 were used (see section 4.4.3) since both the wild type

Figure 6.4: Design for SPR experiment and Kinetic in formation from sensorgram plot.



Legend

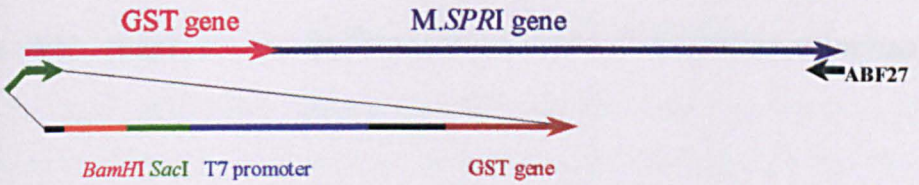
Linear or plasmid DNA encoding M.SPRI wild type and mutant in the monomeric and dimerised form used to transcribe and translated *in vitro* and the produced protein used in SPR.

The progress of an interaction is monitored as a sensorgram. Analyte binds to the surface-attached ligand during sample injection, resulting in an increase in signal. At the end of the injection, the sample is replaced by a continuous flow of buffer and the decrease in signal now reflects dissociation of interactant from the surface-bound complex.

and the mutant are cloned in pET22b vector with out the GST gene, for the dimerised form a PCR fragment was used.

pGS2$_{C78G}$K was digested with *Sal*I to remove the kanamycin^r gene followed by DNA purification from the agarose gel and ligation to reactivate the M.SPRI gene (pGS2$_{C78G}$) (see section 4.4.1 for details). To amplify the GST-M.SPRI$_{C78G}$ in the pGS2$_{C78G}$ and GST-SPRI in pGS2 the upstream primer (ABF29) was designed as shown in figure 6.5. The downstream primer was ABF 27 (see section 4.4.4 and figure therein).

Figure 6.5: The sequence and features for the upstream oligodeoxynucleotide ABF29. ABF29 and ABF27 were used to amplify the GST - M.*SPRI*$_{C78G}$ gene in the pGS2$_{C78G}$ vector and GST - SPR wild type gene in pGS2. Two restriction sites were introduced for cloning purposes: *BamH*I (red) and *Sac*I (green), the T7 promoter (19 bases) is in blue followed by 11 bases as a separator between the T7 promoter and the desired gene, finally the start codon for the GST gene and 18 bp from the 5' end of the GST gene. PCR product will be 2070 bp.



ABF29 forward primer (66 bp) 5` → 3`:

GGCGGATCCGAGCTCTAATACGACTCACTATAGGAACAGACCACCATGTCCCCTATACTAGGTTAT

*BamH*I *Sac*I T7 promoter Start GST gene 5'→3'

Two PCR reactions were set up, the first one using pGS2 as template and the second one using pGS2$_{C78G}$ as template. Vent DNA polymerase was used in this PCR. The thermocycler was run as described in section 2.2.5. The PCR fragments were then analysed by 1% agarose gel electrophoresis and purified using Gene Clean.

6.2.2 *TRANSCRIPTION AND TRANSLATION* IN VITRO

The TNT® Quick Coupled Transcription/Translation System was removed from -70°C and rapidly thawed and divided into five 0.5 ml tubes. DNA was added to each tube as following 1) pES1, 2) pES1$_{C78G}$, 3) GST-M.SPRI (PCR

product), 4) GST-MSPRI$_{C78G}$., and 5) luciferase plasmid as control. Amino acids were added to the reaction which was then incubated at 30°C using the thermocycler. The reactions were kept on ice prior to each BIA experiment.

6.2.3 IMMOBILISATION OF DNA ON STREPTAVIDIN CHIPS

Three biotinylated double stranded 38 mer oligodeoxynucleotides were designed, each one with one of the three sites for M.SPRI, namely *Hae*III, *Eco*RII, and *Msp*I. Single stranded oligodeoxynucleotides were annealed to their complementary strand using equal concentrations. The sequences of the three oligodeoxynucleotides are:

BHA1 5`→3` (*Hae*III): •B-GAATGCTACAGTATCGTGGFCTCACGTACAACATCCAG

BEC1 5`→3` (*Eco*RII): •B-GAATGCTACAGTATCGTCFAGGTCACGTACAACATCCAG

BMS1 5`→3` (*Msp*I): •B-GAATGCTACAGTATCGTCFGGTCACGTACAACATCCAG

The experiment was run using a Sensor Chip SA (see section 6.1.1) with four flow cells (FC). Sensor chip SA was docked in the BIA instrument and equilibrated as described by the manufacturer using methyltransferase buffer. The Sensor Chip SA was primed with three consecutive (1-min) injections of 1 M NaCl in 50 mM NaOH. The ds biotinylated ligands were immobilised on the FCs in the following order FC1 (*Msp*I site), FC2 (*Hae*III site), FC3 (*Eco*RII site), and FC4 (blank). Since the experiment would be run using protein produced *in vitro*, it was necessary to first determine the non-specific binding by the extract and then subtract this binding from the total binding.

In all the experiments the flow buffer used was Mtase buffer (see section 2.1.1) and the dilutions were made using the same buffer to minimise bulk refraction index changes. The BIAcore setup programme is shown in table 6.2.

Table 6.2: BIAcore setup parameters used in the experiments.

Temperature	25.00 °C
Detection mode	Multi
Flow path	1-2-3-4
Flow rate	5µl/min
Data collection	1 Hz
Volume of sample injected	20 µl
Washing to regenerate the chip	5 µl 1M NaCl and 0.01% Triton
Association time	240 Seconds
Dissociation time	600 Seconds
Buffer	Mtase buffer

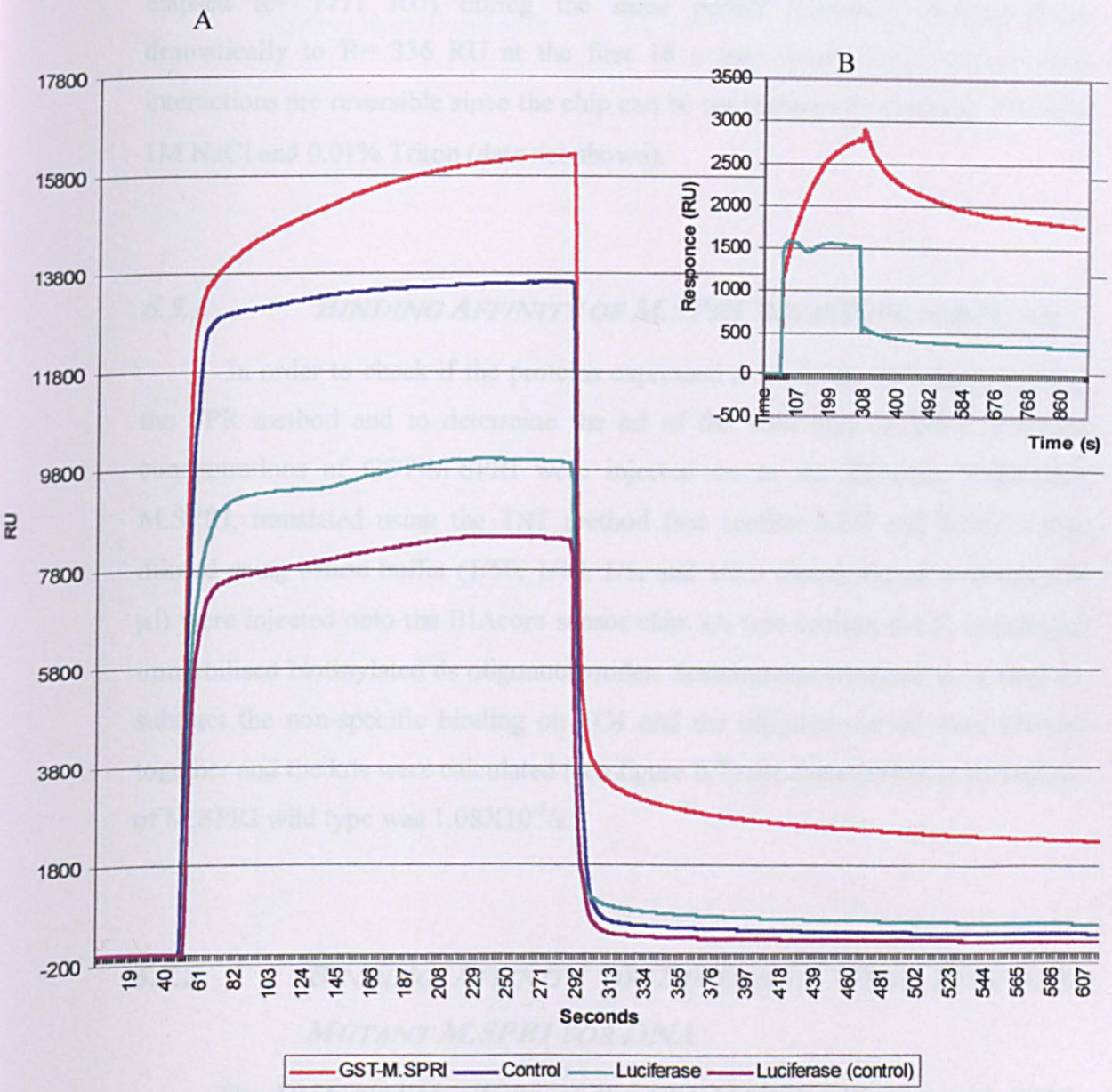
6.3 RESULTS OF M.SPRI:DNA BINDING

In the M.SPRI:DNA binding analysis using BIAcore, two types of controls must be considered, the first one is the non-specific binding between the TNT proteins and the ligand. This can be determined by using a non-Mtase protein, translated using TNT, such as luciferase. Luciferase is used to check whether the TNT has worked efficiently and to determine the level of non-specific binding of the extract to the ligand. The second control required is an estimation of the non-specific binding of proteins to the SA chip. This was determined using the blank flow cell (FC4) which contains streptavidin without DNA.

In the first experiment the non-specific binding between the TNT proteins and the SA was investigated. Equal volumes of luciferase and M.SPRI proteins transcribed using TNT were injected in to the BIAcore and the data obtained are shown in figure 6.6. In a parallel experiment the TNT mixture also was injected to a blank chip and no binding (Rmax) was obtained.

The ejection of luciferase and M.SPRI experiment showed that there was translation of both luciferase and dimerised M.SPRI, (luciferase Rmax = 1557 RU, and M.SPRI Rmax = 2901 RU) since both gave binding comparable to the with TNT reaction protein. The amount of M.SPRI protein bound was higher than

Figure 6.6: Real time binding interactions between ds oligonucleotide and M.SPRI and luciferase proteins.



Legend

Sensorgram from an SPR experiment to determine the non-specific binding between a non-Mtase protein (luciferase) and the ligand and the non-specific binding between the M.SPRI and the blank FC. The injected volumes of M.SPRI and luciferase were equal (20 µl), the ds oligonucleotide with an *Eco*RII site was biotinylated at the 3` end and immobilized on one flow cell of streptavidin coated sensor chip as described in section 6.2.3, and the other FC was blank. A) The red plot is the M.SPRI binding to *Eco*RII site, blue binding to the blank FC as control for non-specific binding, aqua plot binding of luciferase protein to the *Eco*RII site to determine whither there is binding between the TNT reaction mixture and chip, and the pink is the binding to blank FC. All experiments were carried out at 25°C. B) Raw data from sensorgram were used to balanced the plots according to the controls.

the luciferase control. However there was binding for M.SPRI (after 600s, had elapsed R= 1771 RU) during the same period luciferase binding drops dramatically to R= 336 RU at the first 18 s (see figure 6.6b). All of these interactions are reversible since the chip can be regenerated by flushing with 5 µl 1M NaCl and 0.01% Triton (data not shown).

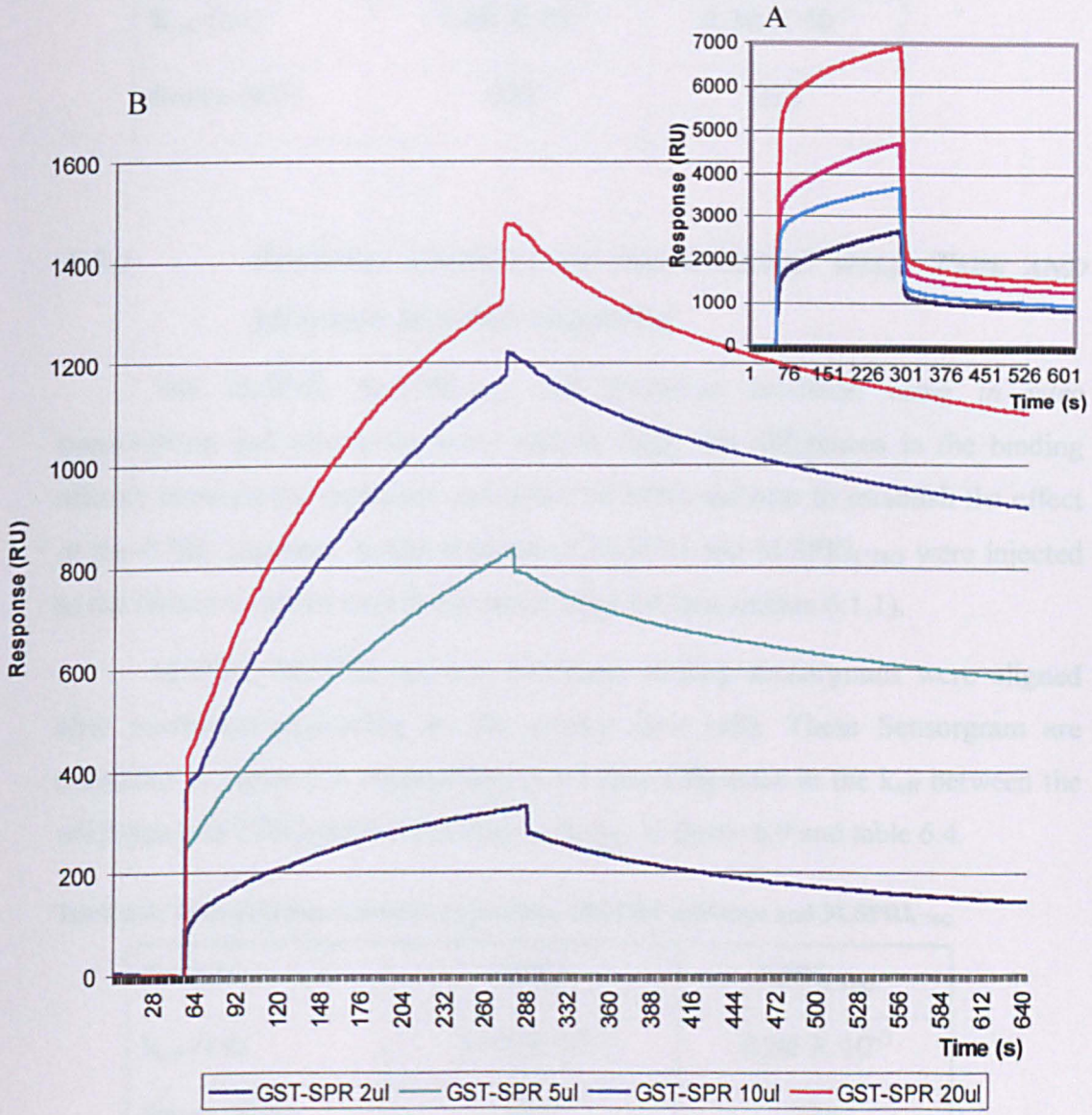
6.3.1 BINDING AFFINITY OF M.SPRI WILD TYPE FOR DNA.

In order to check if the proteins expressed *in vitro* can be analysed using the SPR method and to determine the kd of the wild type M.SPRI, different concentrations of GST-M.SPRI were injected on to the BIAcore. Dimerised M.SPRI, translated using the TNT method (see section 6.2.2 and 2.2.9.2) was diluted using Mtase buffer (1/50, 1/10, 1/5, and 1/2.5 times). Equal volumes (20 µl) were injected onto the BIAcore sensor chip SA (see section 6.1.1) containing immobilised biotinylated ds oligonucleotides. Sensorgrams obtained were used to subtract the non-specific binding on FC4 and the adjusted curves were plotted together and the kds were calculated (see figure 6.7) the dissociation rate constant of M.SPRI wild type was $1.08X10^{-3}/s$

6.3.2 BINDING AFFINITY OF DIMERISED WILD TYPE AND MUTANT M.SPRI FOR DNA

The DNA binding affinity of the GST-M.SPRI wild type and C78G mutant were compared by SPR. Equal volumes of the GST-M.SPRI and GST-M.SPRI$_{C78G}$ proteins were injected onto the BIAcore as described in section 6.2.3. The dissociation was significantly higher for wild type M.SPRI (see figure 6.8). The sensorgram plots were used to analyse the results and the kd was determined for both the wild type and mutant GST-M.SPRI. The results are presented in table 6.3 and figure 6.8. Interestingly the association phase (figure 6.8) is almost the same for both wild type and C78G mutant as expected, furthermore the k_{off}s are significantly different. The C78G mutant has a k_{off} which is about 3 fold lower than the wild type protein.

Figure 6.7: Real time binding interactions between ds oligonucleotide and different concentrations of M.SPRI.



Legend

Sensorgram from an SPR experiment to determine the dissociation constant of dimerised M.SPRI and the ligand. Different dilutions made in Mtase buffer of M.SPRI protein (1, 5, 10, and 20) were injected using the same volume (20 µl). Three ds oligonucleotide with *Msp*I, *Hae*III, and *Eco*RII, sites were biotinylated at the 3` end and immobilized on three flow cells of streptavidin coated sensor chip as described in section 6.2.3, and the fourth FC was blank. The *Eco*RII plots are presented here.

A) Sensorgram of the binding of the four dilutions of dimerised M.SPRI protein. This sensorgram plot used to subtract the non-specific binding and to plot the real response in (B). The red plot is the 20 : l, blue 10 : l, aqua 5 : l, and the pink 2 : l. All experiments were carried out at 25°C.

Table 6.3: Disassociation constant of dimerised M.*SPRI* wild type and M.*SPRI*$_{C78G}$

Protein	GST-*SPRI*	GST-*SPRI*$_{C78G}$
K_{off} (1/s)	1.08×10^{-3}	0.38×10^{-3}
Rmax (RU)	639	579

6.3.3 *BINDING AFFINITY OF MONOMERIC WILD TYPE AND MUTANT M.SPRI FOR DNA*

The M.*SPRI*, M.*SPRI*$_{C78G}$, and luciferase produced using *in vitro* transcription and translation were used to study the differences in the binding affinity between the monomer and dimer M.SPRI and also to establish the effect of the C78G mutation. Equal volumes of M.SPRI and M.SPRI$_{C78G}$ were injected to the BIAcore instruments using sensor chip SA (see section 6.1.1).

M.SPRI, M.SPRI$_{C78G}$, and luciferase binding sensorgrams were aligned after correction (according to the control flow cell). These Sensorgram are presented in figure 6.9. Again there is a 3-fold difference in the k_{off} between the wild type and C78G mutant (30.6%) as shown in figure 6.9 and table 6.4.

Table 6.4: Dissosiciation constant of monomer M.SPRI wild type and M.SPRI$_{C78G}$

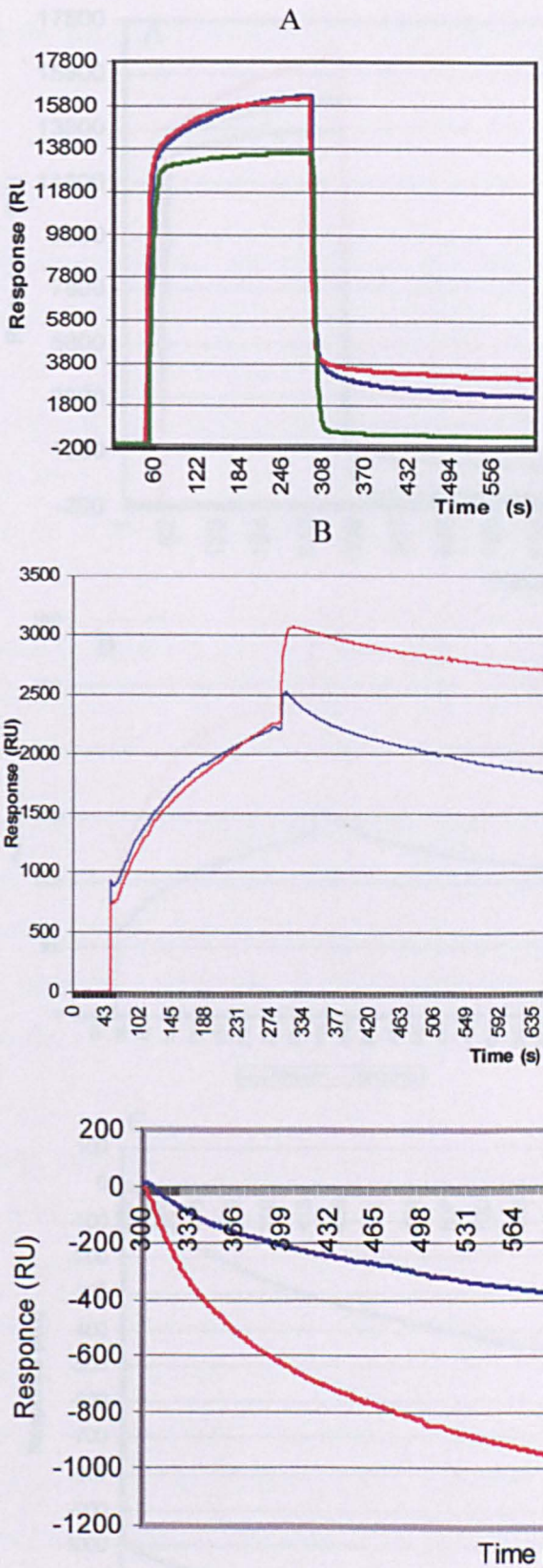
Protein	*SPRI*	*SPRI*$_{C78G}$
k_{off} (1/s)	1.11×10^{-3}	0.34×10^{-3}
Rmax (RU)	777	784

6.4 SUMMARY OF CHAPTER 6 RESULTS

1- Wild type and mutant forms of M.SPRI can be produced *in vitro* both in dimeric and monomeric forms.

2- Mtases produced *in vitro* can be analysed using SPR.

Figure 6.8: Real time binding interactions between ds oligonucleotide and dimerised M.SPRI wild type and C78G mutant.

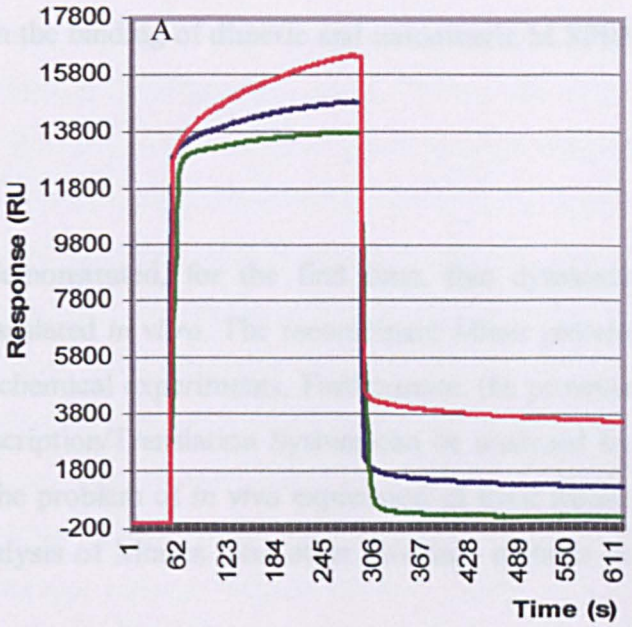


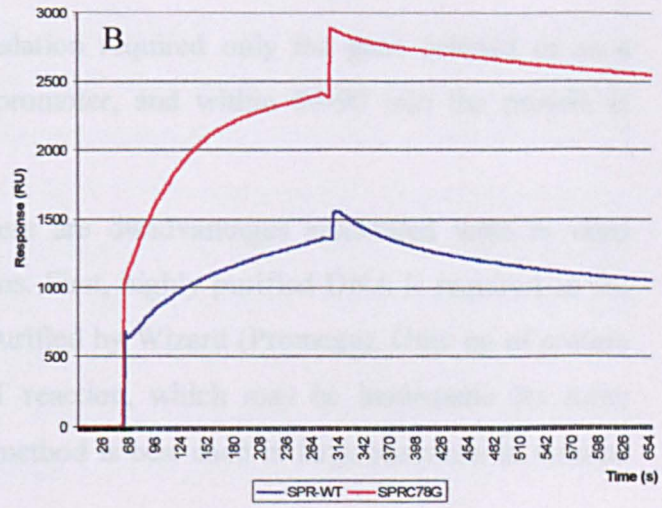
Legend

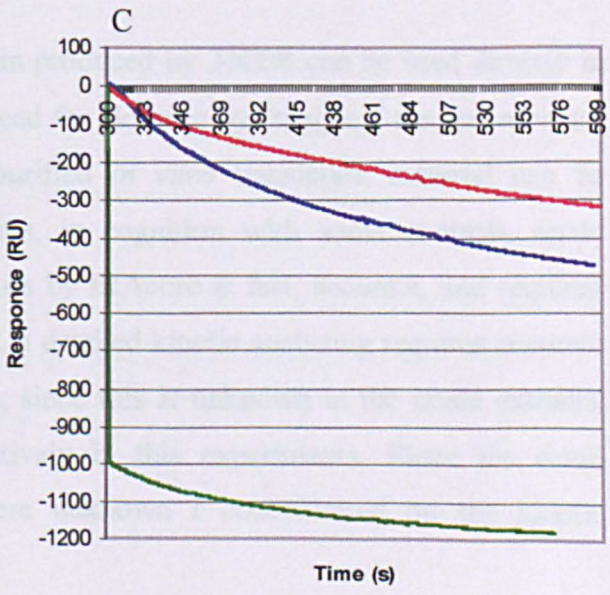
Sensorgram from a SPR experiment to determine the dissociation constant of dimerised M.SPRI and the ligand. Dimerised M.SPRI (blue) and M.SPRI$_{C78G}$ mutant (red) proteins were injected using the same volume (20 : l). Three ds oligonucleotide with *Msp*I, *Hae*III, and *Eco*RII, sites were biotinylated at the 3` end and immobilized on three flow cells of streptavidin coated sensor chip as described in section 6.2.3, and the fourth FC was blank. The *Eco*RII plots are presented here.

A) Sensorgram of the binding of the wild type and mutant dimerised M.SPRI protein and the non-specific binding in FC4 (green). This sensorgram plot was used to subtracted the non-specific binding and to plot the real response in (B). In (C) the dissociation rate is presented after aligning the dimerised M.SPRI (red) and M.SPRI$_{C78G}$ (blue). All experiments were carried out at 25°C.

Figure 6.9: Real time binding interactions between ds oligonucleotide and monomeric M.SPRI wild type and C78G mutant.







Legend

Sensorgram from a SPR experiment to determine the dissociation constant of monomer M.*SPR*I and the ligand. M.*SPR*I (blue) and M.*SPR*I$_{C78G}$ mutant (red) proteins were injected using the same volume (20 : l). Three ds oligonucleotide with *Msp*I, *Hae*III, and *Eco*RII, sites were biotinylated at the 3` end and immobilized on three flow cells of streptavidin coated sensor chip as described in section 6.xx, and the fourth FC was blank. The *Eco*RII plots are presented here.

A) Sensorgram of the binding of the wild type and mutant M.*SPR*I protein and the non-specific binding in FC4 (green). This sensorgram plot was used to subtract the non-specific binding and to plot the real response in (B). In (C) the dissociation rate is presented after aligning the monomeric M.SPRI (red) and M.SPRI$_{C78G}$ (blue). All experiments were carried out at 25°C.

3- The Cys78Gly mutant M.SPRI binds to its TRS with a three fold higher affinity than wild type enzyme.

4- There was no difference in the binding of dimeric and monomeric M.SPRI to DNA.

6.5 DISCUSSION

In this chapter I have demonstrated, for the first time, that cytotoxic Mtases can be transcribed and translated *in vitro*. The recombinant Mtase protein can therefore can be used for biochemical experiments. Furthermore, the proteins produced using the TNT® Transcription/Translation System can be analysed by SPR. This approach overcomes the problem of *in vivo* expression of toxic Mtase mutants, and simplifies rapid analysis of Mtases (and other proteins) mutants *in vitro*.

In vitro transcription/translation required only the gene (cloned or as a PCR product) containing a T7 promoter, and within 60-90 min the protein is available for characterisation.

As with any method, there are disadvantages associated with *in vitro* production of recombinant proteins. First, highly purified DNA is required as the template, usually plasmid DNA purified by Wizard (Promega). Only ng of protein are produced from a single TNT reaction, which may be inadequate for some experiments. On the whole, this method is best used in large reactions in volume with a high sensitivity assay.

As mention above the protein produced by TNT® can be used directly in BIAcore experiments without the need for cleaning by "tagging" the recombinant protein with GST for example, purified *in vitro* translation material can be observed, but for some experiments, in cognition with some controls, crude fraction can be used. Kinetic analysis by BIAcore is fast, accurate, and required small amounts of analyte. However, a detailed kinetic analysing requires accurate estimating of protein concentration, since this is unknown in the crude extracts, the data were evaluated comparatively in this experiments. Since the exact concentrations of the proteins were unknown I concentrated on the kinetic

parameter relating to the dissociation phase this does not require knowledge of the concentration of the analyte.

The toxic effect of a Mtase in which the conserved motif IV Cys is substituted by Gly was firstly reported by Bhagwat's group (Wyszynski et al. 1991). These authors concluded "We have not investigated other possible reasons behind this unusual phenotype in a systematic fashion and it remains a mystery". Two years later Mi and Roberts (1993) extended this finding to M.*Hha*I reciting a different conclusion. They suspected that expression of the M.*Hha*I$_{C81G}$ mutant was cytotoxic because of its enhanced DNA binding to its TRS, which might be expected to interfere with replication and transcription. They supported their idea by a) demonstrating that M.*Hha*I$_{C81G}$ purified from *E.coli* was isolated bound to DNA. b) *in vitro* binding studies of the M.*Hha*I$_{C81G}$ purified free of DNA, showed stronger DNA binding than the wild type enzyme (Mi and Roberts, 1993).

In this study the data are in agreement with Mi and Roberts (1993). Furthermore, the observation has been extended to the multispecific Mtases. Using M.SPRI, the k_{off} for the wild type enzyme was three times higher than that of M.SPRI$_{C78G}$ (1.11 X 10^{-3} /s and 0.34 X 10^{-3} /s respectively). These values of k_{off} are similar to that found by Mi and Roberts for the monospecific Mtase i.e. 1.08 X 10^{-3} /s for wild type and 0.02 X 10^{-3} /s for Gly mutant. The difference in the k_{off} for the mutant M.SPRI is probably due to the difference on methodology used for the determination of k_{off} or may reflect the difference between the mono and multi-specific Mtases. Mi and Roberts (1993) mentioned in their paper "because of the inherent difficulties in measuring accurate Kd values using gel shift assays the absolute values reported here should be treated with caution".

Lastly it is reported here that there is no significant difference between the k_{off} for the M.SPRI in the dimerised and monomeric forms (1.08 X 10^{-3} and 1.11 X 10^{-3} respectively) M.SPRI$_{C78G}$ (0.38 X 10^{-3} for the dimerised form and 0.34 X 10^{-3} for the monomeric). These results support that contention that the toxic effects of these mutant Mtases is due to high affinity binding to their TRS which leads to a block to genetic transactions.

CHAPTER SEVEN:
GENERAL DISCUSSION AND CONCLUSIONS

Synopsis of chapter Seven

1. Definition of the DNA damage caused by mutated Mtases.
2. Discussion of the generality of the deletion with several Mtases.
3. Discussion of the effect of dimerisation on the deletion event.
4. Discussion the involvement of *recA* and *umuDC* gene products in repairing the lesion.
5. Proposal for the deletion mechanism.
6. Future Work

Maintaining the integrity of the genetic information is a major objective of all organisms. In 1953 Watson and Crick proposed that mutations might arise in the DNA (Watson and Crick, 1953). Until recently molecular geneticists tacitly assumed that DNA would be an extremely stable molecule protected from damage by multiple cellular interactions. Subsequently many types of mutations and mutagenesis agents have been discovered, these mutations may be a consequence of environmental damage or may arise spontaneously (see section 1.4.1). Recently a new kind of genetic repair has been found associated with replication arrest or loss of template, which is referred to as error prone-repair or SOS mutagenesis (Friedberg et al. 1995). Given a choice between mutation and death, perhaps the SOS mutagenesis response is a cell's last ditch attempt to survive when DNA damage outstrips its repair capacity (Goodman, 1998). In this study I have found that mono or multi-specific C5-Mtase mutated in motif IV (changing the essential Cys to Gly) are toxic to cells and those that survive harboured deletions in the DNA.

7.1 GENERAL AND SPECIFIC ANALYSIS OF DELETION FORMATION

Generally, the cytotoxicity of mutated (motif IV Cys to Gly) multispecific Mtases is greater than that induced by mutated monospecific Mtases. The percentage of repair using dimerised-M.SPRI$_{C78G}$ was 2.19 and for M.*Msp*I$_{C174G}$ it was 3.8 (see figure 7.1). Moreover, the cytotoxicity of the dimerised Mtases was greater than monomeric Mtases: the percentage of repair using dimerised M.SPRI$_{C78G}$ was 4 times lower than with monomeric M.SPRI$_{C78G}$, and using dimerised M.*Msp*I$_{C174G}$ it was 13.8 times lower than the monomeric M.*Msp*I$_{C174G}$. Furthermore the lethal effect of transformation of *E.coli* with the motif IV C to G mutated Mtases is directly proportional to the number of sites recognised by the Mtases. In the case of M.SPRI$_{C78G}$ (8 TRS) the percentage of repair was 2.19 and 1.98 for pGS2$_{C78G}$ and pLS2$_{C78G}$ respectively, while for M.*Msp*I$_{C174G}$ (3 TRS) it was 3.8, and for M.*Hha*I$_{C81G}$ (1 TRS) it was 16.6 (figure 7.1). These results explain the variation in survival of *E.coli* with different Mtase mutants. Assuming that cytotoxicity is a results of abnormally high affinity protein:DNA interaction,

Figure 7.1: Percentage of repair using different Mtases

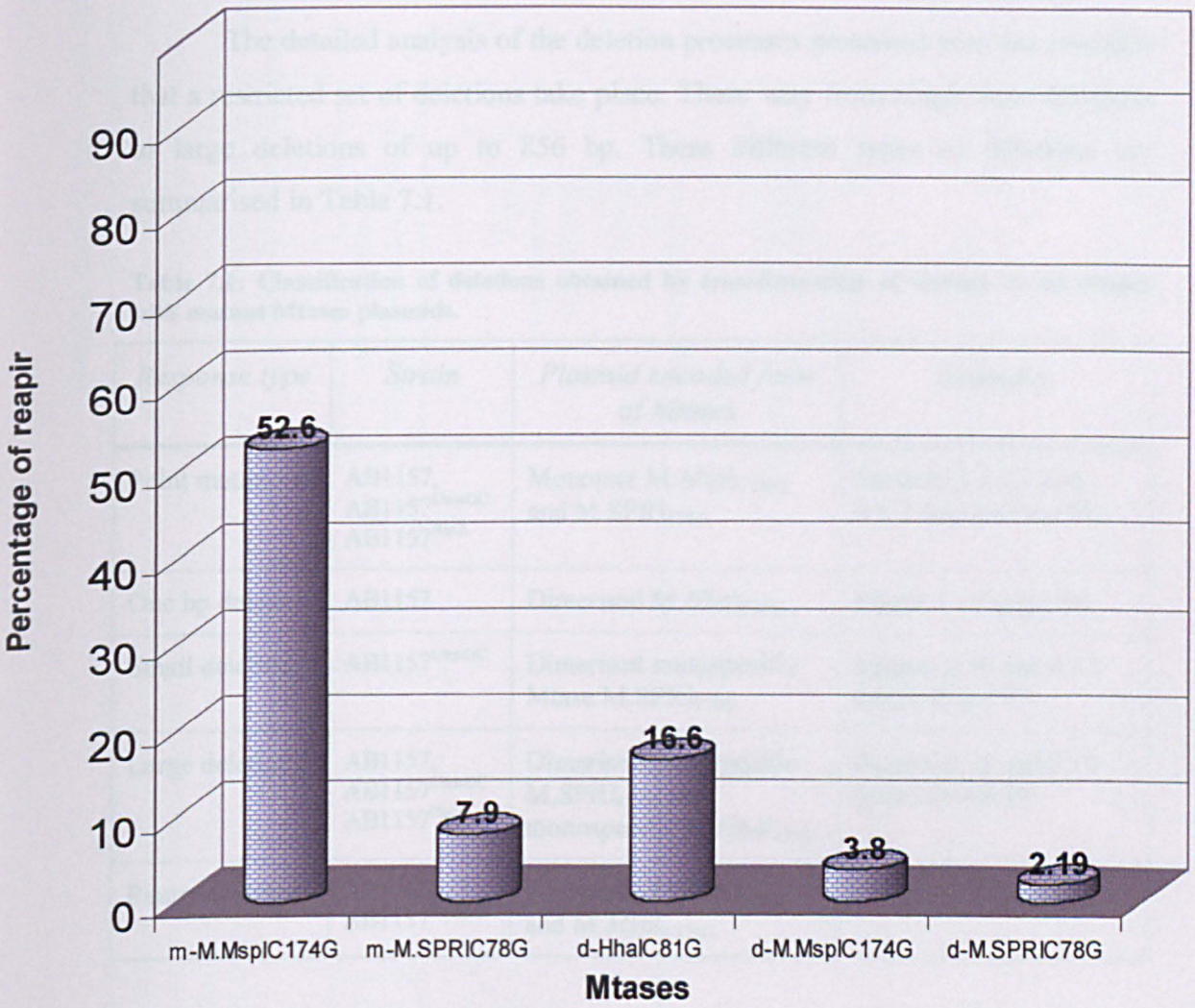


Figure 7.1: Percentage of repair wild type *E.coli* and different Mtases.
Percentage of repair using multi and mono-specific Mtases in the monomeric and dimeric form. logarithmic plot of percentage of repair verse the Mtases type the perecentage of repair was calculated by dividing the number of colonies able to grow in the plate transformed with mutated Mtase by the control one. Where:

m-M.*Msp*I$_{C174G}$	Monomeric mutated M.*Msp*I
m-M.SPRI$_{C78G}$	Monomeric mutated M.SPRI
d-M.*Hha*I$_{C81G}$	Dimeric mutated M.*Hha*I
d-M.*Msp*I$_{C174G}$	Dimeric mutated M.*Msp*I
d-M.SPRI$_{C78G}$	Dimeric mutated M.SPRI

increasing the number of TRSs will increase the number of protein DNA complexes and subsequently increase the demands on cellular repair processes.

The detailed analysis of the deletion processes presented here has revealed that a restricted set of deletions take place. These vary from single base deletions to large deletions of up to 856 bp. These different types of deletions are summarised in Table 7.1.

Table 7.1: Classification of deletions obtained by transformation of various *E.coli* strains with mutant Mtases plasmids.

Response type	*Strain*	*Plasmid encoded form of Mtases*	*Examples*
Point mutations	AB1157, AB1157^{UmuDC} AB1157^{RecA}	Monomer M.*Msp*I$_{C174G}$ and M.SPRI$_{C78G}$	Sections 3.5.1.1 and 4.5.2 (pages 65 and 93)
One bp deletion	AB1157	Dimerised M.*Hha*I$_{C81G}$	Figure 3.15 (page 69)
Small deletions	AB1157^{UmuDC}	Dimerised multispecific Mtase M.SPRI$_{C78G}$	Figures 4.10 and 4.13 (pages 86 and 87)
Large deletions	AB1157, AB1157^{UmuDC} AB1157^{RecA}	Dimerised multispecific M.SPRI$_{C78G}$ and monospecific M.*Hha*I$_{C81G}$	Figures 4.16 and 4.19 (pages 88 and 89)
Rearrangements	AB1157, AB1157^{UmuDC}	Dimerised M.SPRI$_{C78G}$ and M.*Msp*I$_{C174G}$	Figure 3.13 (page 66)

As described in section 3.2.1 recombinant GST when expressed in *E.coli* is dimeric (Ji et al. 1992; Smith, 1993) and therefore Mtase fusions will be artificially dimerised. A dimeric Mtase will presumably bind at two separate TRS adopting one of four possible orientations. A) It may bind to two sites forming a small loop which may led to small deletions such as that observed with M.SPRI$_{C78G}$Δ32 and M.SPRI$_{C78G}$Δ6 (figure 7.2a). B) The dimeric Mtase:DNA complex may bridge two sites forming a large loop in a *cis* configuration which may lead to large deletions such as those observed with M.SPRI$_{C78G}$Δ718 and M.*Hha*I$_{C81G}$d2, or in *trans* (C) by bridging two TRSs from different plasmid molecules which leads to either a one bp deletion such as M.*Hha*I$_{C81G}$d1 or a rearrangement. (D) The monomer Mtase:DNA complex, which binds just at one site, generally appears to produce point mutations (figure 7.2b).

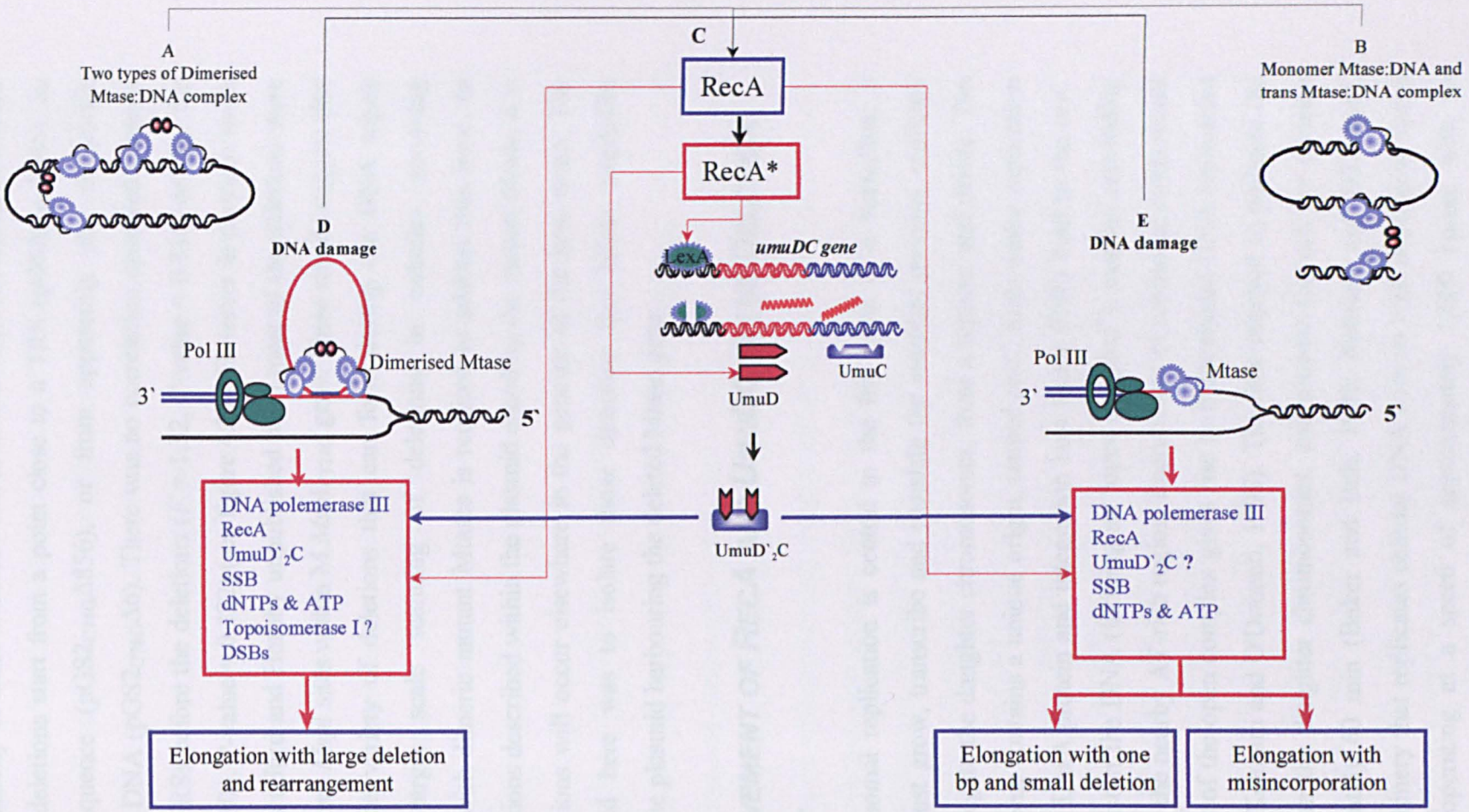


Figure 7.2: Macro-lesion and its effect on *E. coli.*

Mutated Mtase flip the target cytosine out from the DNA helix and bind to carbon 5 with high affinity binding. There are four models of binding the dimerised Mtase bound two sites in the DNA forming large and small loops (a) or the dimerised Mtase bound to two sites in different DNAs (b). The protein:DNA complex lead to blockage of the replication and breakage of DNA which activate RecA protein to induce the SOS response (c). UmuD`$_2$C, pol III, RecA (and may topoisomerase I) work together to repair the lesion by two pathways depends on the protein:DNA complex. Illegitimate recombination which lead to large deletion between short repeat homology (d) and TLS which lead to small deletion, point mutation (depends on UmuD`$_2$C) (e). Large deletions are depends on the homologous and the loop formed by the Mtase.

The specificity of these different deletions still requires further statistical analysis, some deletions start from a point close to a TRS ($pGS2_{C78G}\Delta645$), an homologous sequence ($pGS2_{C78G}\Delta856$), or from apparently any nucleotide sequence in the DNA ($pGS2_{C78G}\Delta6$). There was no correlation observed between the number of TRSs before the deletions ($F = 1.32$, P-value = 0.316) or after the deletion ($F = 1.306$, P-value = 0.322) (see figure 4.27). However in previous work in the laboratory (Hurd and Hornby unpublished data) most of the deletions were found between two *Msp*I sites within M.*Msp*I_{C174G} gene. These results suggest that there are a wide variety of deletions that can form during the DNA repair processes. A larger scale scanning of deletions in colonies surviving transformation with dimeric mutant Mtases is required to address this issue. In addition to deletions described within the plasmid encoding the mutant Mtase, it is likely that deletions will occur elsewhere in the genome of the host strain. The strategy adopted here was to isolate those deletions that retain ampicillin resistance and the plasmid harbouring the deleted Mtase gene.

7.2 INVOLVEMENT OF RECA AND UMUDC IN THE DELETION EVENT

Chromosomal replication is central to the life cycle of a bacterium: a newborn cell must grow, transcribe and translate the essential proteins, replicate its genome segregate the daughter chromosomes, form a septum, and divide. The *E.coli* chromosome contains a unique origin, termed *oriC*, from where replication is initiated. The DnaA protein and integration host factor (IHF) bind to the *oriC* sequence and distort the DNA (forming an "open complex"), causing unwinding of the strands in the nearby AT-rich region. Recruitment of replisome components to the two forks of the open complex gives rise to bidirectional DNA replication (Crooke, 1995; Kelmam and O'Donnell, 1995). The time required to replicate the genome, partition the daughter chromosomes, and prepare the cell for division totals approximately 60 min (Baker and Bell, 1998; Kelmam and O'Donnell, 1995). The machinery that replicates cellular DNA does so with impressive speed and accuracy, operating at a speed of approximately 1,000 bp/sec with an accuracy of one error per 10^7 bp replaced (Kornberg and Baker, 1992). Studies on DNA structural and functional parameters which stimulate deletion formation

indicate a correlation between deletion hotspots and replication pause sites. This correlation was noticed for recombination events between short homologous as well as strictly non-homologous sequences (Bierne et al. 1991) Recent studies show that double strand breaks (DSBs) can arise from the breakage of arrested replication forks (Bierne et al. 1991). Progression of the replication fork may be impeded by DNA secondary structure, DNA damage or DNA-bound proteins, which inhibit the polymerase or the helicase. These replication arrests could lead to DNA breakage (Bierne et al. 1991; Bierne et al. 1997). RecA protein was one among the protein responsible for recombining the double stranded ends with other free ends (Bierne et al. 1997). Based on these data and the data reported in this thesis, the proposed involvement of *recA* and *umuDC* gene products and mechanisms of deletion will be discussed below.

7.2.1 SMALL DELETIONS AND POINT MUTATIONS

It is assumed that *in vivo* Mtase:DNA complexes cause a cessation of normal replication and transcription and may lead to double strand breaks. Organisms have developed various strategies to protect their genomes from the damaging effects of endogenous and exogenous agents. Despite having robust excision repair systems that removes DNA lesions before replication, cells also possess efficient strategies for tolerating lesions persisting at the replication fork during DNA synthesis (Livneh et al. 1993) (see section 1.4). Two basic strategies of lesion tolerance can be distinguished: **(i)** translesion synthesis (TLS) which is a process during which the replication machinery reads through the lesion with an associated risk of fixing a mutation. If the lesion does not impede DNA synthesis, an unmodified replication complex can achieve TLS. However, a replication complex modified by accessory proteins encoded by genes associated with the SOS regulon is required for TLS past lesions that hinder the progression of replication **(ii)** DNA damage avoidance is a general strategy that facilitates replication of damaged DNA templates without the need for the polymerase to read through the lesion. This strategy takes advantage of the information contained in the complementary strand (Napolitano et al. 1997). Two models of damage avoidance have been proposed (Livneh et al. 1993): **(a)** postreplication recombinational repair, in which the DNA polymerase, blocked at a lesion site, dissociates from the DNA and reinitiates replication downstream from the lesion,

leaving a gap that is repaired by a recombination mechanism involving the sister chromatid. (b) Polymerase strand switching, in which the DNA polymerase switches temporarily from the damaged parental template to the undamaged newly synthesised strand of the sister chromatid before returning to the parental template downstream from the lesion. Both damage avoidance models are believed to be efficient and error-free. Our current level of understanding of the molecular mechanisms underlying these processes is rudimentary (Livneh et al. 1993).

The first direct evidence that the *umuD* and *umuC* gene products might be acting by modifying a replication fork so that translesion synthesis could occur on a damaged template was provided by Echols and his colleagues (Rajagopalan et al. 1992). They reported that translesion synthesis over a synthetic abasic site did not occur with DNA polymerase III holoenzyme unless they also added UmuD', UmuC, and RecA (Rajagopalan et al. 1992). In another experiment it was shown that posttranslational processing of UmuD into UmuD` is a critical step in SOS mutagenesis, enabling only the latter protein to interact with UmuC (Jonczyk and Nowicka, 1996). As discussed in section 1.4.6, RecA* is required to stimulate UmuD autodigestion. In other experiments it has been found that transforming AB1157 cells with pGW2123, a *umuD`C*-bearing plasmid, recovers and enhances the frequency of back reversions of the translesion repair (Fabisiewicz and Celina, 1998). The *umuD* and *umuC* genes are subject to both transcriptional and posttranslational regulation as part of *E. coli*'s SOS response to DNA damage (see section 1.4.6 and reviewed in (Friedberg et al. 1995)). Although limited genetic evidence suggests that DNA polymerase III participates in SOS mutagenesis *in vivo*, the relationship of the holoenzyme and its various subunits to this process still remains unclear (Friedberg et al. 1995; Walker, 1998).

Working with *umuDC* is difficult due to two reasons. One is the complex regulatory circuit governing *umuDC* expression and activity. The second has been the decidedly "non-user-friendly" nature of the UmuC protein. Because of the insolubility of UmuC protein, Echols group (Rajagopalan et al. 1992) purified UmuC in its denatured state and obtained the small amounts used in their experiments by renaturation. Very recently two teams manage to over come these

difficulties and published their results essentially contemporaneously (Tang et al. 1998; Reuven et al. 1998).

In both of the two above papers, they used an abasic site as a lesion in different DNA molecules. However Tang et al. (1998) used a UmuD'$_2$C complex while Reuven et al., (1998) taking a different approach to the difficulties of working with UmuC, purified and characterised a fusion protein consisting of the maltose-binding protein joined to the N terminus of UmuC. Reuven et al. (1998) found that the addition of MBP-UmuC and UmuD' to a mixture of DNA polymerase III holoenzyme, RecA, and SSB resulted in robust translesion synthesis with up to 70% of the substrates elongated and past the lesion in a short reaction. In contrast to the Tang et al. (1998) experiments, DNA synthesis past the lesion was quite processive. Some RecA- and SSB-stimulated bypass was observed in the absence of UmuD' and MBP-UmuC but, by analyzing the translesion synthesis products, Reuven et al. (1998) were able to show that the lower amounts of bypass synthesis that occurred in the absence of the Umu protein had resulted from some type of slippage mechanism that most commonly generates a 1 bp deletion (figure7.2). In contrast, when the Umu proteins were present, an additional product of the expected size was present that had evidently resulted from the insertion of a nucleotide, most commonly an A, opposite the abasic site. Based on these observations, Reuven et al. (1998) make the interesting proposal that UmuD' and UmuC function to suppress small deletions, a highly disruptive type of mutation within a coding region, and to promote base substitution, a more mild type of mutation within the same context. *In vivo* experiments provided additional support for their hypothesis (Reuven et al. 1998).

These results are very interesting and germane to the data in this thesis. In the above experiments they used an abasic site as a lesion while in our experiment the lesion was a protein:DNA complex. Considering the above findings it can be concluded that the mechanism of the point mutation, one bp deletion and the small deletion (table 7.1) is consistent with the Reuven et al. and Tang et al. models (figure 7.2e). In contrast with Reuven et al. (1998) the results here show that there is a single base pair deletion even in the presence of UmuD'$_2$C (figure 3.15 and table 7.1). However the one-bp deletion was found only when the smallest Mtase

M.*Hha*I$_{C81G}$ (Cheng et al. 1993) was used. The base substitutions found by Reuven et al. (1998) and Tang et al. (1998) occur opposite the abasic site: in contrast our results shows that it was before the TRS and this maybe due to the physical size of the nucleoprotein complex. Base substitution when observed (section 3.5.1.1) was from T to A in agreement with Reuven et al. (1998). Elswhere Jackson-Grusby et al. (1997) found using the C5-Mtase inhibitor 5-aza2'deoxycytidine (which promotes covalent nucleoprotein complex formation) induced a spectrum of mutations, these mutations are predominately (73%) C:G → G:C but also include C:G→T:A and C:G A:T.

In the experiments by Tang et al. (1998) it was found that the deletions were up to two bp long. These results indicate that the length of deletions may vary dependant upon the lesion. This may explain why small deletions were found when M.SPRI$_{C78G}$ encoding plasmids were used to transform the AB1157^{UmuDC} cells (table 7.1 and section 4.5.1.1).

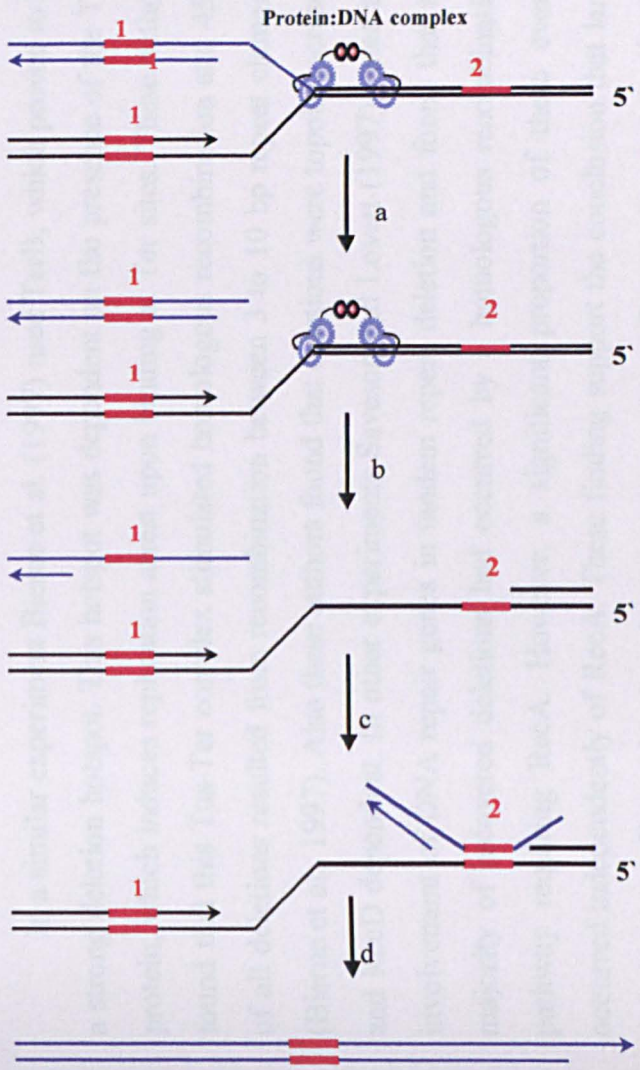
Finally it should be mentioned that the fusion of non-homologous sequences can also be mediated by enzymes whose physiological function involves the generation and joining of DNA ends (Ehrlich, 1989). Topoisomerases, which control the topological state of chromosomes through transient breakage and rejoining of DNA strands, are the main enzymes involved in non-homologous rearrangements (Bierne et al. 1997). Other nicking-closing enzymes, such as site-specific recombinases, also accidentally mediate illegitimate recombination (Ehrlich, 1989).

7.2.2 Models for Generation of Large Deletions (0.1-1 Kb)

While the above mechanism may help to explains the small deletion and base substitution, it is not valid for the larger deletions. Since some of the large deletion occur in the same strain by using the same Mtase (see section 3.6 for example) in different repair products by starting the deletion from the same position in the gene and end at different length (see figure 3.17), which may depends on the loop formed by the dimerised Mtase and the slipped misalignment mechanism directed the deletion. However some large deletions were observed between two repeated sequence in all strains used in these experiment (see table 5.6) which indicate that there is an homology based element to error-prone repair.

Figure 7.3: Models for deletion formation between short direct repeats.

When the replicative polymerase encounters the Mtase lesion, it may dissociate. What may follow is described in the figure for encounter at a single site, although the replicative fork will also approach the other face of the complex. Short repeated sequences (1 and 2) are shown as bold lines. (a) Breakage of the lagging strand template (blue) in the vicinity of the lesion, generating a DSB. (b) Nucleolytic degradation of the exposed 5' ends, generating a 3'-tailed single-strand region. (c) Annealing of complementary sequences 1 and 2. (d) The intermediate is repaired by removal of the 3' tail, gap filling and ligation. A round of replication produces the deletant plasmid molecule. Alternatively, the leading strand template is also broken and repair of both strands leads directly to the deletant plasmid.



The most probable mechanism by which large deletions are generated is through **illegitimate recombination,** i.e. recombination between sequences of little or no homology. This is a multi-pathway phenomenon which is dependent on the presence of short-homologous sequences, at least 3 bp, at the new junction (Bierne et al. 1997; Albertini et al. 1982). Two major mechanisms were proposed for rearrangements joining short-homologous sequences. The slipped mispairing, or copy-choice mechanism, results from template switching by DNA polymerases. It was proposed by Albertini et al. (1982) and is supported by *in vivo* and *in vitro* experiments in *E.coli* (Canceill and Ehrlich, 1996). A second mechanism is conceptually related to the single-strand annealing model. The key event is the occurrence of a DNA double-strand break (DSB) between the repeats, followed by exonucleolytic erosion and pairing of exposed complementary sequences. Such rearrangements using short direct repeats for the pairing step have been described in *E.coli* (Kong and Masker, 1994).

In a similar experiment Bierne et al. (1997) used TerB, which proved to be a strong deletion hotspot. This hotspot was dependent on the presence of the Tus protein, which induces replication arrest upon binding at Ter sites. These authors found that this Tus-Ter complex stimulated homologous recombination and 45% of all deletions resulted from recombination between 3 to 10 bp repeat elements (Bierne et al. 1997). Also these authors found that deletions were topoisomerase I and RecD dependent. In other experiments Saveson and Lovett (1997) tested the involvement of DNA repair genes in tandem repeat deletion and found that the majority of recovered deletions had occurred by a homologous recombination pathway requiring RecA. However, a significant proportion of these events occurred independently of RecA. These finding support the conclusion that large deletions are effected by short homologous sequence. Bearing in mind that in this work large deletions are found in all of the isogenic trains (AB1157, AB1157^{RecA}, and AB1157^{UmuDC}) this indicates that there are other factor(s) which lead to the recombination between two homologous sequence. Large deletions (between short-homologous sequences) result from the repair of the DSBs caused by replication arrest. This repair may proceed by the single-stranded annealing pathway (figure 7.3).

In general, it can be concluded that the mutated Mtases flip the target cytosine from the DNA helix and bind to it with high affinity. This protein:DNA complex leads to a blockage of replication and breakage of DNA which activates RecA protein to induce the SOS response. UmuD`$_2$C, pol III, RecA (and possibly topoisomerase I) work together to repair the lesion by two pathways dependant on the protein:DNA complex. TLS which leads to small deletions and point mutations depends on UmuD`$_2$C whilst illegitimate recombination, which leads to large deletions between short homologous repeats requires homology and free ds DNA.

7.3 FUTURE WORK

In this thesis I have established that dimeric forms of Mtases can provoke DNA damage. There are many unresolved issues that require further investigation, including whether the damaging effects of the mutant methyltransferase can be enhanced by increasing the oligomeric state of the protein e.g. by constructing artificial trimers, tetramers etc. However, perhaps the two most important areas for the future are establishing which gene(s) / proteins are involved in deletion repair and what mechanism(s) exist for this repair.

Towards the end of my thesis, the complete genome of *E.coli* was published. In conjunction with new gene chip technologies, it is now possible to identify groups of genes that are activated in response to a mutagen (for example) in a single experiment (Service, 1998; Chu et al. 1998). This approach would be ideal for determining the number and sequences of those genes involved in this deletion repair process. For example it would be easy to find if the Topo I gene is involved. However, future experiment will be required to establish the molecular basis of deletion and the role(s) that the gene products play in these processes. One technique that offers great promise for the analysis of nucleoprotein complexes is atomic force microscopy (García et al. 1997; Rippe et al. 1997).

Finally a series of large scale experiments are required to produce a sound statistical/population analysis of the frequencies and specificities of the repair process in a range of genetically defined backgrounds.

References

And

Appendixes

REFERENCE LIST

Adams, R.L.P. (1990) DNA methylation. *Biochem.J.* **265**, 309-320.

Adams, R.L.P. (1995) Eukaryotic DNA methyltransferases-structure and function. *BioEssays* **17**, 139-145.

Adams, R.L.P. and Burdon, R.H. (1985) *Molecular Biology of DNA Methylation*, 1 edn. New York: Springer Verlag.

Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982) On the formation of spontaneous deletions: the importance of short sequences of homologies in the generation of large deletions. *Cell* **29**, 319-328.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994) *Molecular Biology of The Cell.*, 3 edn. London: Garland Publishing.

Asai, T. and Kogoma, T. (1994) Homologous recombination-dependent initiation of DNA replication from DNA damage-inducible origins in *Escherichia coli. EMBO J* **176**, 7113

Ausubel, F.M. (1992) *Short Protocols in Molecular Biology*, Second Edition edn. New York: John Wiley & Sons.

Bailone, A., Sommer, S., Knezevic, J., Dutreix, M. and Devoret, R. (1991) A RecA protein deficient in its interaction with the UmuDC complex. *Biochemistry* **73**, 479-484.

Baker, T.A. and Bell, S.P. (1998) Polymerases and the Replisome: Machines within Machines. *Cell* **92**, 295-305.

Barbe, J., Gibert, I. and Guerrero, R. (1986) 5-Azacytidine: survival and induction of the SOS responce in *E.coli. Mutat.Res.* **166**, 9-16.

Barras, F. and Marinus, M.G. (1989) The great GATC:DNA methylation in *E.coli. TIG* **5**, 139-143.

Behrens, B., Pawlek, B.P., Morelli, G. and Trautner, T.A. (1983) Restriction and modification in *Bacillus subtilis:* Construction of hybrid λ and SPP1 phagescontaining a DNA methyltransferase genes from *B. subtilis* SPR. *Molecular and General Genetics* **189**, 10-16.

Bergerat, A., Guschlbauer, W. and Fazakerley, V. (1991) Allosteric and catalytic binding of S-adenosylmethionine to *E.coli* DNA adenine methyltransferase monitored by ^{3}H NMR. *Proc.Natl.Acad.Sci.USA* **88**, 6394-6397.

Bertani, G. and Weigle, J.J. (1953) Host controlled variations bacterial viruses. *Journal of Bacteriology* **65**, 113-121.

Bestor, T.H. and Ingram, V.M. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc.Natl.Acad.Sci.USA* **80**, 5559-5563.

Bhagwat, A.S. and Roberts, R. (1987) Genetic Analysis of the 5-Azacytidine Sensitivity of *Escherichia coli* K-12. *Journal of Bacteriology* **169**, 1537-1546.

Bierne, H., Ehrlich, S.D. and Michel, B. (1991) The replication termination signal *TerB* of the *E. coli* chromosome is a deletion hotspot. *EMBO J* **10**, 2699-2705.

Bierne, H., Ehrlich, S.D. and Michel, B. (1997) Deletions at stalled replication forks occur by two different pathways. *EMBO J* **16**, 3332-3340.

Boudsocq, F., Campbell, M., Devoret, R. and Bailone, A. (1997) Quantitation of the Inhibition of Hfr x F⁻ Recombination by the Mutagenesis Complex UmuD'C. *JMB* **270**, 201-211.

Boyer, H.W. and Dussoix, D.R. (1969) A Complementation Analysis of the Restriction and Modification of DNA in *Escherichia coli* . *JMB* **41**, 459-472.

Bridges, B.A. and Woodgate, R. (1985) Mutagenic repair in *E. coli*: products of the *recA* gene and *umuD* and *umuC* genes act at different steps in UV-induced mutagenesis. *Proc.Natl.Acad.Sci.USA* **82**, 4193-4197.

Brown, T. and Brown, D.J.S. (1991) *Oligonucleotides and Analogues.*, 1 edn. Oxford University Press.

Buhk, H.-J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Günthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) Restriction and modification in *Bacillus subtillus*: nucleotide sequence, functional organization and product of the DNA methytransferase gene of bacterophage SPR. *Gene* **29**, 51-61.

Canceill, D. and Ehrlich, S.D. (1996) Copy-choice recombination mediated by DNA polymerase III holoenzyme from *E.coli*. *Proc.Natl.Acad.Sci.USA* **93**, 6647-6652.

Cayrol, C., Petit, C., Raynaud, B., Capdevielle, J., Guillemot, J.C., Defais, M. and Grzesiuk, E. (1995) Recovery of respiration following the SOS response of Escherichia coli requires RecA-mediated induction of 2-keto-4-hydroxyglutarate aldolase. *Mutation Research - DNA Repair* **92**, 11806-11809.

Challberg, M.D. and Englund, P.T. (1980) Specific labeling of 3' termini with T4 DNA polymerase. *Meth.Enzymol.* **65**, 39

Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* **30**, 11018-11025.

Cheng, X., Kumar, S., Posfai, J., James, W., Pflugrath, J.W. and Roberts, R. (1993) Crystal Structure of the HhaI DNA Methyltransferase Complexed with S-Adenosyl-L-Methionine. *Cell* **74**, 299-307.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I. (1998) The Transcriptional Program of Sporulation in Budding Yeast. *Science* **282**, 699-705.

Clark, A.J. and Margulies, A.D. (1965) Isolation and characterisation of recombinant deficint mutants of *E.coli* K-12. *Proc.Natl.Acad.Sci.USA* **53**, 451-459.

Cohen-Fix, O. and Livneh, Z. (1994) *In vitro* UV mutagenesis associated with nucleotide excision repair gaps in *E. coli*. *Journal of Biological Chemistry* **269**, 4953-4958.

Crooke, E. (1995) Regulation of ChromosomalReplication in E. coli: Sequestration and Beyond. *Cell* **82** , 877-880.

Daniel, A.S., Fuller-Pace, F.V., Legge, D.M. and Murray, N.E. (1988) Distribution and diversity of hsd genes in Escherichia coli and other enteric bacteria. *Journal of Bacteriology* **170**, 1775-1782.

Drlica, K., Engle, E.C. and Manes, S.H. (1980) DNA gyrase on the bacterial chromosome: possibilty of two levels of action. *Proc.Natl.Acad.Sci.USA* 77, 6879-6883.

Dubey, A.K., Mollet, B. and Roberts, R. (1992) Purification and characterization of the *Msp*I DNA methyltransferase cloned and overexpressed in *E.coli*. *Nuclic Acids Research* **20**, 1579-1585.

Ehrlich, S.D. (1989) Illegitimate recombination in bacteria. In: Berg, D.E. and Howe, M.M., (Eds.) *Mobile DNA*, pp. 799-832. Washington, DC: American Society for Microbiology]

Elledge, S.J. and Walker, G.C. (1983) Proteins requred for UV light and chemical mutagenesis: identification of the products of the *umuC* locus of *E.coli*. *JMB* **164**, 175-192.

Fabisiewicz, A. and Celina, J. (1998) DNA mutagenesis and repair in UV-irradiated *E. coli* K-12 under condition of mutation frequency decline. *Mutat.Res.* **402**, 59-66.

Frank, E.G., Gonzalez, M., Ennis, D.G., Levine, A.S. and Woodgate, R. (1996) In Vivo Stability of the Umu Mutagenesis Proteins: a Major Role for RecA. *Journal of Bacteriology* **178**, 3550-3556.

Frank, E.G., Hauser, J., Levine, A.S. and Woodgate, R. (1993) Targeting of the UmuD, UmuD', and MucA' mutagenesis proteins to DNA by RecA protein. *Proc.Natl.Acad.Sci.USA* **90**, 8169-8173.

Friedberg, E.C., Walker, G.C. and Siede, W. (1995) *DNA Repair and Mutagenesis*, 1 edn. Washington, DC: American Society for Microbiology.

Friedman, S. (1981) The inhibition of DNA (cytosine-5) methylases by 5-azacytidine: the effect of azacytosine containing DNA. *Mol.Pharmacol* **19**, 314-320.

Friedman, S. and Ansari, N. (1992) Binding of the EcoRII methyltransferase to 5-flourocytosine-containing DNA: isolation of a boundprotein. *Nucleic Acids Research* **20**, 3241-3248.

Friedman, S. and Som, S. (1993) Induction of *Eco*RII methyltransferase: evidence for autogenous control. *Journal of Bacteriology* **175**, 6293-6298.

Gabbara, S., Sheluho, D. and Bhagwat, A.S. (1995) Cytosine Methyltransferase from *Escherichia coli* in which Active Site Cysteine is Replaced with Serine Is Partially Active. *Biochem.J.* **34**, 8914-8923.

García, R.A., Bustamante, C. and Reich, N.O. (1997) Sequence-specific recognition by cytosine C^5 and adenine N^6 DNA methyltransferases requires different deformations of DNA. *Proc.Natl.Acad.Sci.USA* **93**, 7618-7622.

Gold, M. and Hurwitz, J. (1964) The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. *Journal of Biological Chemistry* **239**, 3858-3865.

Goodhead, D.T. (1989) The initial damage produced by ionizing radiations. *Int.J.Radiat.biol.* **56**, 623-634.

Goodman, M.F. (1998) Purposeful mutations. *Nature* **395**, 221-223.

Guan, K.L. and Dixon, J.E. (1991) Eukaryotic proteins expressed in *E.coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal Biochem* **192**, 262-267.

Hagensee, M.E., Timme, T.L., Bryan, S. and Moses, R.E. (1987) DNA polymerase III of *E. coli* is required for UV and ethyl methanesulfonate mutagenesis. *Proc.Natl.Acad.Sci.USA* **84**, 4195-4199.

Hanahan, D. (1985) Transformation. In: Glover, D.M., (Ed.) *DNA Cloning*, 1 edn. pp. 109-135. Oxford: IRL Press]

Heitman, J. (1993) On the origins, Structures and functions of Restriction-modification enzymes. In: Sotlow, J.K., (Ed.) *Genetic Engineering*, pp. 57-108. New York: Plenum Pree]

Heitman, J. and Model, P. (1987) Site-specific methylases induce the SOS DNA repair response in *E.coli*. *Journal of Bacteriology* **169**, 3243-3250.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.

Hoeijmakers, J.H. (1993) Nucleotide excision repair I: from *E.coli* to yeast. *TIG* **9**, 173-177.

Holliday, R. (1964) A mechanism of gene conversion in fungi. *Genet.Res.* **5**, 282-304.

Howard, B.M., Pinney, R.J. and Smith, J.T. (1993) Function of the SOS process in repair of DNA damage induced by modern 4-quinolones. *J.Pharm.Pharmacol* **45**, 658-662.

Hurd, P.J. (1996) Astudy of the catalytic properties of M.*Msp*I DNA methyltransferase. University of Sheffield. PhD Thesis.

Iwasaki, H., Nakata, A., Walker, G.C. and Shinagawa, H. (1990) The *E. coli polB* gene, which encodes DNA polymeraseII, is regulated by the SOS system. *Journal of Bacteriology* **172**, 6268-6273.

Jackson-Grusby, L., Laird, P.W., Magge, S.N., Moeller, B.J. and Jaenisch, R. (1997) Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. *Proc.Natl.Acad.Sci.USA* **94**, 4681-4685.

Jacobs, A., Bopp, A. and Hagen, U. (1972) In vitro repair of single-stranded breaks in γ-irradiated DNA by polynucleotide ligase. *Int.J.Radiat.biol.* **22**, 431-435.

Ji, X., Zhang, P., Armstrong, R.N. and Gilliland, G.L. (1992) The three dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2Å resolution. *Biochemistry* **31**, 10169-10184.

Jonczyk, P. and Nowicka, A. (1996) Specific in Vivo Protein-Protein Interactions between *Escherichia coli* SOS Mutagenesis Proteins. *Journal of Bacteriology* **178**, 2580-2585.

Kato, T. and Shinoura, Y. (1977) Isolation and characterization of mutants of *E.coli* deficient in induction of mutations by ultraviolet light. *Molecular and General Genetics* **156**, 121-131.

Kelmam, Z. and O'Donnell, M. (1995) DNA polymerase III Holoenzyme: structure and Function of a Chromosomal Replicating Machine. *Annual review of Biochemistry* **64**, 171-200.

Klimasauskas, S., Kumar, S., Roberts, R. and Cheng, X. (1994) HhaI Methyltransferase Flips Its Target Base Out of the DNA Helix. *Cell* **76**, 357-369.

Klimasauskas, S., Nelson, J.L. and Roberts, R. (1991) The sequence specificity domain of cytosine-C5 methylases. *Nucleic Acids Research* **19**, 6183-6190.

Kobayashi, I. (1996) DNA Modification and Restriction: Selfish Behaviour of an Epigenetic System. In: Anonymous *Epigenetic Mechanisms of gene Regulation*, 1 edn. pp. 155-172. Cold Spring Harbor Laboratory Press]

Koffel-Schwartz, N., Coin, F., Veaute, X. and Fuchs, R.P.P. (1996) Cellular strategies for accommodating replication-hindering addcts in DNA: control by the SOS response in *E. coli*. *Proc.Natl.Acad.Sci.USA* **93**, 7805-7810.

Kong, D. and Masker, W. (1994) Deletion between direct repeats in T7 DNA stimulated by double-strand breaks. *Journal of Bacteriology* **176**, 5904-5911.

Kornberg, A. and Baker, T.A. (1992) *DNA replication*, 1 edn. New York: Freeman.

Kowalczykowski, S.C. and Eggleston, A.K. (1994) HOMOLOGOUS PAIRING AND DNA STRAND-EXCHANGE PROTEINS. *Annual review of Biochemistry* **63**, 991-1043.

Krüger, D.H. and Bickle, T.A. (1983) Bacteriophage survival: Multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol.Rev.* **47**, 345-360.

Krüger, D.H., Bickle, T.A., Reuter, M., Pein, C.D. and Schrooeder, C. (1990) *Nuclic Acid Methylation*, 1 edn. New York: Alan R. Liss.

Kuehnlein, U. and Arber, W. (1972) Host specificity of DNA produced by *E.coli* XV. The role of nucleotide methylation in *in vitro* B-specific modification. *JMB* **63**, 9-19.

Kuehnlein, U., Linn, S. and Arber, W. (1969) Host specificity of DNA produced by *E.coli*. *Proc.Natl.Acad.Sci.USA* **63**, 556-562.

Kulakauskas, S., Lubys, A. and Ehrlich, S.D. (1995) DNA restriction-modification system mediate plasmid maintenance. *Journal of Bacteriology* **177**, 3451-3453.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. and Wilson, G.G. (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Research* **22**, 1-10.

Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) Cytosine-specific Type II DNA Methyltransferases A Conserved Enzyme Core with Variable Target-recognizing Domains. *JMB* **206**, 305-312.

Lehman, L.R. (1974) DNA ligase: structure, mechanism, and function. *Science* **186**, 790-797.

Lin, P.M., Lee, C.H. and Roberts, R. (1989) Cloning and characterization of the genes encoding the *Msp*I restriction modification system. *Nucleic Acids Research* **17**, 3001-3011.

Lindahl, T. (1982) DNA repair Enzymes. *Annual review of Biochemistry* **51**, 61-87.

Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.

Little, J.W. (1984) Auotdigestion of *lexA* and phage λ repressors. *Proc.Natl.Acad.Sci.USA* **81**, 1375-1379.

Livneh, Z., Cohen-Fix, O., Skaliter, R. and Elizur, T. (1993) Replication of damaged DNA and the molecular mechanism of ultraviolet light mutagenesis. *Crit.Rev.Biochem.Mol.Biol.* **28**, 465-513.

Loeb, L.A. and Kunkel, T.A. (1981) Fidelity of DNA synthesis. *Annual review of Biochemistry* **52**, 457

Lu, A., Clark, S. and Modrich, P. (1983) Methyl-directed Repair of DNA Base-Pair Mismatch *in vitro*. *Proc.Natl.Acad.Sci.USA* **80**, 4639-4643.

Luria, S.E. and Human, M.G. (1952) A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* **64**, 557-569.

Marinus, M.G. (1987) DNA Methylation in *Escherichia coli*. *Annual review of Genetics* **21**, 113-131.

Masek, F., Fridrichova, I., Pirsel, M. and Sedliakova, M. (1995) DNA repair in Escherichia coli: The dual function of uvr genes. *Neoplasma* **42**, 317-323.

Meisel, A., Krüger, D.H. and Bickle, T.A. (1991) M.*Eco*P15 methylates the second adenine in its target recognition sequence. *Nucleic Acids Research* **19**, 3997-3998.

Mi, S. and Roberts, R. (1993) The DNA binding affinity of *Hha*I methylation is increased by a single amino acid substitution in the catalytic center. *Nucleic Acids Research* **21**, 2459-2464.

Myszka, D.G. (1997) Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Current Opinion In Biotechnology* **8**, 50-57.

Naito, T., Kusano, K. and Kobayashi, I. (1995) Selfish Behaviour of Restriction-Modification Systems. *Science* **267**, 897-899.

Napolitano, R.L., Lambert, I.B. and Fuchs, R.P.P. (1997) SOS Factors involved in translesion synthesis. *Proc.Natl.Acad.Sci.USA* **94**, 5733-5738.

Nastri, H.G., Gusso, A., Lange, C., Walker, G.C. and Knight, K.L. (1997) Mutational analysis of the RecA protein L1 region identifies this area as a probable part of the co-protease substrate binding site. *Molecular Microbiology* **25**, 967-978.

Nelson, H.C. and Bestor, T.H. (1996) Base eversion and shuffling by DNA methyltransferases. *Chemistry and Biology* **3**, 419-423.

Nohmi, T., Battista, J.R., Dodson, L.A. and Walker, G.C. (1988) RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation. *Proc.Natl.Acad.Sci.USA* **85**, 1816-1820.

Noyer-Weidner, M. and Trautner, T.A. (1993) Methylation of DNA in Prokaryotes. In: Jost, J.P. and Saluz, H.P., (Eds.) *DNA Methylation: Molecular Biology and Biological Significance*, 1 edn. pp. 39-108. Switzerland: Birkhauser Verlag Basel]

Noyer-Weidner, M., Jentsch, S., Kupsch, J., Bergbauer, M. and Trautner, T.A. (1985) DNA methyltransferase genes of *Bacillus subtilis* phages: structural relatedness and gene expression. *Gene* **35**, 143-150.

O'Neil, M., Chen, A. and Murray, N.E. (1997) The restriction-modification genes of *Escherichia coli* K-12 may not be selfish: They do not resist and are readily replaced by alleles conferring different specificities. *Proc.Natl.Acad.Sci.USA* **94**, 14596-14601.

Osterman, D.G., DePillis, G.D., Wu, J., Matsuda, A. and Santi, D. (1988) 5-fluorocytosine in DNA is a mechanism-based inhibitor of *HhaI* methylase. *Biochemistry* **27**, 5204-5210.

Peat, T.S., Frank, E.G., McDonald, J.P., Levine, A.S., Woodgate, R. and Hendrickson, W.A. (1996) Structure of the UmuD' protein and its regulation in response to DNA damage. *Nature* **380**, 727-730.

Perara, E. and Lingappa, V.R. (1985) A former amino terminal signal sequence engineered to an internal location directs translocation of both flanking protein domains. *Journal of Cell Biology* **101**, 2292

Perry, K.L., Elledge, S.J., Mitchell, B., Marsh, L. and Walker, G.C. (1985) *umuDC* and operons whose products are required for UV light- and chemical-induced mutagenesis: UmuD, MucA, and LexA proteins share homology. *Proc.Natl.Acad.Sci.USA* **82**, 4331-4335.

Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Research* **17**, 2421-2435.

Prakash, L., Sherman, F., Miller, M., Lawrence, C. and Tabor, W., (Eds.) 1973. Phenomenology of an inducible mutagenic DNA repair pathway in *E.coli*: SOS repair hyposthesis. **2**, 1 edn. 128 p. Springfield: Charles Thomas. (1974)

Radman, M. Phenomenology of an inducible mutagenic DNA repair pathway in *E.coli*: SOS repair hyposthesis. Prakash, L., Sherman, F., Miller, M., Lawrence, C., and Tabor, W. 2(1), 128-142. 1974. Springfield, Charles Thomas. Molecular and Enviromental Aspects of Mutagenesis. 1973.

Rajagopalan, M., Lu, A., Woodgate, R., O'Donnell, M., Goodman, M.F. and Echols, H. (1992) Activity of the purified mutagenesis proteins UmuC, UmuD', and RecA in replicative bypass of an abasic DNA lesion by DNA polymerase III. *Proc.Natl.Acad.Sci.USA* **89**, 10777-10781.

Rehrauer, W.M. and Kowalczykowski, S.C. (1996) The DNA Binding Site(s) of the *Escherichia coli* RecA Protein. *The Journal of Biological Chemistry* **271**, 11996-12002.

Reinisch, K., Chen, L., Verdine, G.L. and Lipscomb, W. (1995) The crystal structure of *Hae*III methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged basepairing. *Cell* **82**, 143-153.

Renbaum, P. and Razin, A. (1995) Footprint analysis of M.*Sss*I and M.*Hha*I methytransferases reveals extensive interactions with the substrate DNA backbone. *JMB* **248**, 19-26.

Reuven, N.B., Tomer, G. and Livneh, Z. (1998) The Mutagenesis Proteins UmuD' and UmuC Prevent Lethal Frameshifts While Increasing Base Substitution Mutations. *Molecular Cell* **2**, 191-199.

Richardson, F.C. and Richardson, K.K. (1990) Sequence-dependent formation of alkyl DNA adducts: a review of methods, results, and biological correlates. *Mutat.Res.* **233**, 127-138.

Rippe, K., Guthold, M., Hippel, P.H. and Bustamante, C. (1997) Transcriptional Activation via DNA-looping: Visualization of Intermediates in the Activation Pathway of *E. coli* RNA Polymerase σ^{54} Holoenzyme by Scanning Force Microscopy. *JMB* **270**, 125-138.

Roberts, R., Myers, P.A., Morrison, A. and Murray, K. (1976) A specific endonuclease from Haemophilus haemolyticus. *JMB* **103**, 199-208.

Roca, A.I. and Cox, M.M. (1997) RecR protein: Structure, Function, and Role in Recombinational DNA Repair. *Progress in Nucleic acid Research and Molecular Biology* **56**, 129-223.

Russell, D.W. and Zinder, N.D. (1987) Hemimethylation prevents DNA replication in *E. coli. Cell* **50**, 1071-1079.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.*, Second Edition edn. New York: Cold Spring Harbour laboratory Press.

Sancar, G.B. and Rupp, W.D. (1983) A novel repair enzyme: UvrABC Excision Nuclease of *E.coli* Cuts a DNA strands on both Sides of the Damaged Region. *Cell* **33** , 249-260.

Santi, D., Garrett, C. and Barr, P. (1983) On the mechanism of inhibition of DNA-cytosine methyltransferses by cytosine analogous. *Cell* **33**, 9-10.

Sassanfar, M. and Roberts, J.W. (1990) Nature of the SOS inducing signal in *E.coli*: the involvement of DNA replication. *JMB* **212**, 79-96.

Saveson, G.J. and Lovett, S.T. (1997) Enhanced Deletion by Aberrant DNA Replication in *Escherichia coli. Genetics* **146**, 457-470.

Service, F. (1998) Microchip Arrays Put DNA on the Spot. *Science* **282**, 396-399.

Setlow, R.B. (1966) Cyclobutane-type pyrimidine dimers in polynucleotides. *Science* **153**, 379-386.

Shinagawa, H., Iwasaki, H., Kato, T. and Nakata, A. (1988) RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. *Proc.Natl.Acad.Sci.USA* **85**, 1806-1810.

Skarstad, K. and Boye, E. (1993) Degradation of individual chromosomes in recA mutants of *Escherichia coli*. *Journal of Bacteriology* **175**, 5505

Smith, B.T. and Walker, G.C. (1998) Mutagenesis and More: *umuDC* and the *Escherichia coli* SOS Response. *Genetics* **148**, 1599-1610.

Smith, D.B. (1993) Purification of Glutathione S-Transferase Fusion Proteins. *Methods In Molecular and Cellular Biology* **4**, 220-229.

Smith, D.B. and Johnson, K.S. (1988) Single step purification of polypeptides expressed in *E.coli* as fusions with glutathione S-transferase. *Gene* **67**, 31-40.

Smith, H.O. (1979) Nucleotide sequence specificity of restriction endonucleases. *Science* **205**, 455

Smith, S.S., Kaplan, B.E., Sowers, L.C. and Newman, E.M. (1992) Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proc.Natl.Acad.Sci.USA* **89**, 4744-4748.

Som, S. and Friedman, S. (1994) Inhibition of Transcription *in Vitro* by Binding of DNA (Cytosine-5)-Methylases to DNA Templates Containing Cytosine Analogs. *The Journal of Biological Chemistry* **269**, 25986-25991.

Som, S., Bhagwat, A.S. and Friedman, S. (1987) Nucleotide sequence and expresion of the gene encoding the *Eco*RII modification enzyme. *Nucleic Acids Research* **15**, 313-323.

Sommer, S., Knezevic, J., Bailone, A. and Devoret, R. (1993) Induction of only one SOS operon, *umuDC* is required for SOS mutagenesis in *E. coli*. *Molecular and General Genetics* **239**, 137-144.

Steinborn, G. (1978) Uvm mutants of *E. coli* K12 deficient in UV mutagenesis. I. Isolation of *uvm* mutants and their phenotypical characterization in DNA repair and mutagenesis. *Molecular and General Genetics* **165**, 87-93.

Stenberg, E., Persson, B., Roos, H. and Urbaniczky, C. (1991) Quantitative determination of surface concentration of protein with surface plasmon resonance by using radiolabeled proteins. *J Colloid Interface Sci* **143**, 513-526.

Stewart, F.J. and Raleigh, E.A. (1998) Dependence of McrBC cleavage on distance between recognition elements. *Biol Chem* **379**, 611-616.

Story, R.M., Bishop, D.K., Kleckner, N. and Steitz, T.A. (1993) Structural relationship of bacterial RecA proteins to recombination proteins from bacteriophage T4 and yeast. *Science* **259**, 1892

Story, R.M., Weber, I.T. and Steitz, T.A. (1992) The structure of the *E. coli recA* protein monomer and polymer. *Nature* **355**, 318-325.

Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *JMB* **189**, 113-130.

Sutherland, E., Coe, L. and Raleigh, E.A. (1992) McrBC: a multisubunit GTP-depedent restriction endonuclease. *JMB* **225**, 327-348.

Szostak, J., Orr-Weaver, T., Rothstein, R. and Stahl, F. (1983) The double-Strand Break Repair Model for Recombination. *Cell* **33**, 25-35.

Szpilewska, H., Bertrand, P., Bailone, A., Dutreix, M. and Grzesiuk, E. (1995) In vitro inhibition of RecA-mediated homologous pairing by UmnD'C proteins. *Mutation Research - DNA Repair* **77**, 848-853.

Tang, M., Bruck, I., Eritja, R., Turner, J., Frank, E.G., Woodgate, R., O'Donnell, M. and Goodman, M.F. (1998) Biochemical basis of SOS-induced mutagenesis in *E. coli*: Reconstitution of *in vitro* lesion bypass dependent on the UmuD'$_2$C Mutagenic complex and RecA protein. *Proc.Natl.Acad.Sci.USA* **95**, 9755-9760.

Taylor, C., Ford, K., Connolly, B. and Hornby, D.P. (1993) Determination of the order of substrate addition to *MspI* DNA methyltransferase using a novel mechanism-based inhibitor. *Biochem.J.* **291**, 493-504.

Teo, I., Sedgwick, B., Demple, B., Li, B. and Lindahl, T. (1984) Induction of Resistance to Alkylating Agents in *E.coli*: The *ada* gene Product Serves Both as a Regulatory Protein and as an Enzyme for Repair of mutagenic Damage. *EMBO J* **3**, 2151-2157.

Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) DNA methyltransferase genes of *Bacillus subtilis* phage: comparison of their nucleotide sequences. *Gene* **42**, 89-96.

Trautner, T.A., Pawlek, B.P., Günthert, U., Canosi, U., Jentsch, S. and Freund, M. (1980) Restriction and modification in *Bacillus subtilius*: Identification of a gene in the temperate phage SPβ coding for a *Bsu*R specific modification methyltransferase. *Molecular and General Genetics* **180**, 361-367.

Wagner, R. and Meselson, M. (1976) Repair Tracts in Mismatched DNA Heteroduplexes. *Proc.Natl.Acad.Sci.USA* **73**, 4135-4139.

Walder, R.Y., Langtimm, C.J., Chatterjee, R. and Walder, J.A. (1983) *Journal of Biological Chemistry* **285**, 1235-1241.

Walker, G.C. (1998) Skiing the black diamond slope: Progress on the biochemistry of translesion DNA synthesis. *Proc.Natl.Acad.Sci.USA* **95**, 10348-10350.

Walter, J., Trautner, T.A. and Noyer-Weidner, M. (1992) High plasticity of multispecific DNA methyltransferases in the region carrying DNA target recognizing enzyme modules. *EMBO J* **11**, 4445-4450.

Watson, J.D. and Crick, F.H.C. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964-967.

Weissbach, A. (1993) A Chronicle of DNA methylation (1948-1975). In: Jost, J.P. and Saluz, H.P., (Eds.) *DNA Methylation: Molecular Biology and Biological Significance*, 1 edn. pp. 1-10. Switzerland: Birkhäuser]

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B.P., Behrens, B. and Trautner, T.A. (1988) Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J* **7**, 2601-2609.

Wilson, G.G. (1991a) Organization of restriction-modification systems. *Nucleic Acids Research* **19**, 2539-2566.

Wilson, G.G. and Murray, N.E. (1991b) Restriction and modification system. *Annual review of Genetics* **25**, 585-627.

Woodgate, R. and Ennis, D.G. (1991) Levels of chromosomally encoded Umu proteins and requirements for *in vivo* UmuD cleavage. *Molecular and General Genetics* **229**, 10-16.

Wu, J. and Santi, D. (1987) Kinetic and catalytic mechanism of *HhaI* methyltransferase. *Journal of Biological Chemistry* **262**, 4778-4786.

Wu, J. and Santi, D. (1988) High level expression and purification of *HhaI* methyltransferase. *Nucleic Acids Research* **16**, 703-717.

Wyszynski, M.W., Gabbara, S. and Bhagwat, A.S. (1991) Substitution of a cysteine conserved among DNA cytosine methylases results in a variety of phenotypes. *Nucleic Acids Research* **20**, 319-326.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Zubay, G. (1973) *In vitro* synthesis of protein in microbial systems. *Annual review of Genetics* **7**, 267-287.

APPENDIX ONE:

ABBREVIATIONS AND IUPAC AMBIGUOUS CODES FOR AMINO ACID

Amino acid	One-letter symbol	Three-letters abbreviation
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

IUPAC Ambiguous Codes

N	any base	R	A or G
B	not A	S	C or G
D	not C	V	not T
H	not G	W	A or T
K	G or T	Y	C or T
M	A or C		

APPENDIX TWO:

MAJOR GENETIC MARKERS

Major Genetic Markers Present in *E.coli* Strains used in This Study.

Symbol	Description	Effect
ara	Mutation in arabinose metabolism	Blocks arabinose catabolism.
*argE*3	Mutation in arginine metabolism	Requires arginine for growth on minimal media.
galk	Galactokinase mutation	Blocks catabolism of galactose.
glnV	Glutamine biosynthesis and activation	Glutamine-tRNA2
gpt	Mutation in guanine metabolism	Requires guanine for growth on minimal media.
hisG	Mutation in histidine metabolism	Requires histidine for growth on minimal media
*hsdS*20 (r_B^-,m_B^-)	Restriction minus, modification minus	Allows cloning without cleavage of transformed DNA by endogenous restriction endonuclease. DNA prepared from this strain is unmethylated by *hsdS*20 methylase.
ilv	Mutation in isoleucine and valine metabolism	
kdgk	Mutation in Ketodeoxygluconokinase	
*lacY*1	β-$_D$-galactosidase mutation	Blocks lactose utilization.
*leuB*6	β-isopropyl malate dehydrogenase mutation	Requires leucine for growth on minimal media.
mgl	Mutation in methyl-galactoside	
mtl-1	Mutation in mannitol metabolism	Blocks catabolism of mannitol.
*proA*2	Mutation in proline metabolism	Requires proline for growth on minimal media.
*recA*13	Mutation in recombination	Prevents recombination of introduced DNA with host DNA, ensuring stability of inserts.
rfbD	Mutation in cell wall synthesis	TDP-rhamnose synthesis.
rpsL	Mutation in subunit S12 of 30S ribosome	Confers resistance to streptomycin.
sulA	Suppressor of *lon*	
*supE*44	Suppressor mutations	Suppress amber (UAG) mutations.
thi-1	Mutation in thiamine metabolism	Thiamine required for growth in minimal media
thr-1	Mutation in threonine metabolism	Requires threonine for growth on minimal media.
tsx	Phage T6 and colicin K resistance	
UmuDC	Mutation in UmuDC gene	Sensitive to U.V.
xyl-5	Mutation in xylose metabolism	Blocks catabolism of xylose

APPENDIX THREE:

TABLE OF OLIGONUCLEOTIDES

Oligonucleotides used in this study and their features.

Oligos	Sequence 5`→3`	Features
ABF 19	GCGCGGATCCTTGGGTAAACTCCGGGTA ATG	forward primer for M.SPRI contain *BamHI* and *Msp*I sites
ABF 20	GCGCGCCGGCTACTCTAAAGCTTTGGCC AGGAGATCCTC	Reverse oligo encoding C78G mutation in M.SPRI gene contain *NgoM*I
ABF13	GCGCGAATTCTCTTTTCGTCTGAAATAG	Forward primer 100 bp upstream motif IV start from bp 383 in the M.*Msp*I gene *Eco*RI
ABF14	GCGCAAGCTTGAACATTGTTCCTTGAGT	primer 100 bp downstream motif IV start from bp 585 in the M.*Msp*I gene HindIII
ABF17	CCGGGATGAAACCTGAAATATTGAAATT GATTCGTAGT	prime in the 5' end of the M.*Msp*I gene Start
ABF18	CGCGAATTCTTAAACGAGTTCTAATTCA AAGTTTTCTT	prime in the 3' end of the M.*Msp*I gene Stop
ABF22	CCGGAGCATGACATTTTATGTGCAGGAT TTCCGGGCCAGC	Forward oligo encoding C174G in M.*Msp*I gene
ABF23	TCGTACTGTAAAATACACGTCCTAAAGG CCCGGTCGGTAC	Reverse oligo encoding C174G in M.*Msp*I gene
ABF29	**GGCGGATCCGAGCTCTAATACGACTCAC TATAGGAACAGACCACCATGTCCCCTAT ACTAGGTTAT**	forward primer contain *BamH*I, *Sac*I sites, T7 promoter and Start GST gene
AWSPR1	CGCGGGATCCTTGGGTAAACTACGTGTA ATGAGTC	forward primer for M.SPRI contain *BamH*I site
AWSPR2	CGCGACTCGAGTTATTCAGATTCTTTAT TAACGTATG	reverse primer for M.SPRI contain *Xho*I site
IP1	ACGAATTCCCATATGATTGAAATAAAAG AT	prime in the 5' end of the M.*Hha*I gene Start
IP4	AGGGGCTATGCCCCTTTCGTCGACATAT GGTTTGAAATTTAA	prime in the 3' end of the M.*Hha*I gene Stop
KANX3	GCGCTTCTAGATGAATCGCCCCATCA	reverse primer for Kanamycin^r gene with *Xba*I site
KANX5	GCGCTTCTAGAGTCGACACGTTGTGT	Forward primer for Kanamycin^r gene with *Xba*I site
BHA1	•BGAATGCTACAGTATCGTGGFCTCACGT ACAACATCCAG	Biotenylated oligonucleotide with *Hae*III site
BEC1	•BGAATGCTACAGTATCGTCFAGGTCACG TACAACATCCAG	Biotenylated oligonucleotide with *Eco*RII site
BMS1	•BGAATGCTACAGTATCGTCFGGTCACGT ACAACATCCAG	Biotenylated oligonucleotide with *Msp*I site

APPENDIX FOUR:

FULL SEQUENCE OF THE DELETION MUTANTS

1- Deletion in M.*Hha*I$_{C81G}$ gene aligned with the wild type.

```
GGATCCCCGAATTCCCATATGATTGAAATAAAAGATAAACAGCTCACAGGATTACGCTTTATTGAC
GGATCCCgGAATTCCCATATGATTGAAATtAAAGATAAACAGCTCACAGGATTACGCTTTATTGAC

CTTTTTGCAGGATTAGGTGGCTTTAGACTTGCTTTAGAATCTTGCGGTGCTGAGTGCGTTTATTCT
CTTTTTGCAGGATTAGGTGGCTTTAGACTTGCTTTAGAATCTTGCGGTGCTGAGTGCGTTTATTCT

AATGAATGGGATAAATATGCACAAGAAGTATATGAGATGAATTTTGGTGAAAAGCCTGAGGGCGAC
AATGAATGGGATAAATATGCACAAGAAGTATATGAGATGAATTTTGGTGAAAAGCCTGAGGGCGAC

ATTACCCAAGTAAATGAGAAAACCATTCCTGATCACGACATTTTATGTGCAGGGTTTCCGTGCCAA
ATTACCCAAGTAAATGAGAAAACCATTCCTGATCAnGACATTTTATGCGCAGGGTTTCCGgGCCAA

GCATTCTCTATTTCCGGAAAACAAAAAGGATTCGAGGACTCTAGAGGTACCCTCTTTTTTGATATC
GCATTCTCTATTTCCGGAAAACAAAAAGGATTCGAGGACT----------

GCACGTATTGTCCGTGAAAAAAAACCTAAAGTGGTTTTTATGGAAAATGTGAAAAATTTTGCATCG
----------

CATGATAATGGAAATACGTTAGAAGTTGTAAAAAATACAATGAATGAATTGGACTATTCTTTTCAT
----------

GCTAAAGTATTAAATGCTTTAGATTATGGGATTCCACAGAAAAGGGAACGTATCTATATGATTTGT
----------

TTTCGCAATGATCTCAATATTCAAAATTTCCAATTTCCAAAACCTTTTGAGCTTAATACTTTTGTG
----------

AAAGATTTGTTATTACCTGATAGCGAGGTGGAACACTTAGTTATTGATAGAAAAGATTTGGTAATG
----------

ACAAACCAAGAAATTGAGCAAACAACCCCCAAAACAGTTCGACTTGGTATTGTAGGAAAAGGTGGG
----------

CAAGGAGAACGAATTTATAGCACAAGAGGCATTGCAATTACCTTATCTGCTTATGGTGGCGGCATT
----------

TTCGCTAAGACAGGGGGATATTTAGTAAACGGGAAGACACGGAAATTACACCCTAGAGAGTGTGCT
----------

AGAGTAATGGGCTACCCAGATAGTTATAAAGTCCACCCGTCAACCAGCCAAGCATATAAACAATTT
---------GGCTACCCAGATAGTTATAAAGTCCACCCGTCAACCAGCCAAGCATATAAACAATTT

GGTAACTCAGTTGTTATCAATGTACTTCAATATATTGCTTATAACATTGGTTCATCATTAAATTTC
GGTAACTCAGTTGTTATCAATGTACTTCAATATATTGCTTATAACATTGGTTCATCATTAAATTTC

AAACCATATGTCGAC
AAACCATATGTCGAC
```

2- Full nucleotide sequencing for the M.SPRI$_{C78G}$ deletions

Refe. molecule: M.SPRI 1 - 1391 (1391 bps) Homology

Sequence 2:	SPR$_{C78G}$Δ6	1 - 1387	(1387 bps)	99%
Sequence 3:	SPR$_{C78G}$Δ32	1 - 1359	(1359 bps)	96%
Sequence 4:	SPR$_{C78G}$Δ6	1 - 746	(746 bps)	53%
Sequence 5:	SPR$_{C78G}$Δ6	1 - 673	(673 bps)	48%
Sequence 6:	SPR$_{C78G}$Δ6	1 - 642	(642 bps)	46%

```
GATCCTTGGGTAAACTACGTGTAATGAGTCTTTTTAGTGGGATCGGTGGATTTGAAGCTGCACTAAGAAA
GATCCTTGGGTAAACTCCGGGTAATGAGTCTTTTTAGTGGGATnGGTGGATTTGAAGCTGCACTAAGAAA
GATCCTTGGGTAAACTCCGGGTAATGAGTnTTTTTAGTGGGATCGGTGGATTTGAAGCTGCAnTAAGAAA
GATCCTTGGGTAAACTCCGGGTAATGAGTCTTTTTAGTGGGATCGGTGGATTTGAAGCTGCACTAAGAAA
GATCCTTGGGTAAACTCCGGGTAATGAGTCTTTTTAGTGGGATCGGTGGATTTGAAGCTGCACTAAGAAA
----------------------------------------------------------------------

CATTGGGGTTGGTTATGAGCTGGTTGGTTTAGTGAGATTGATAAATATGCNGTCAAATCTTTTTGTGCA
CATTGGGGTTGGTTATGAGCTGGTTGGTTTAGTGAGATTGATAAATATGCcGTCAAATCTTTTTGTGCA
CATTGGGGTTGGTTATGAGCTGGTTGGTTTAGTGAGATTGATAAATATGCcGTCAAATCTTTTTGTGCA
CATTGGGGTTGGTTATGAGCTGGTTGGTTTAGTGAGATTGATAAATATGCcGTCAAATCTTTTTGTGCA
CATTGGGGTTGGTTATGAGCTGGTTGGTTTAGTGAGATTGATAAATATGCcGTCAAATCTTTTTGTGCA
---------------------------------------------------------------------

ATTCACAACGTTGATGAGCAATTAAATTTTGGAGATGTAAGCAAGATTGATAAGAAAAAACTACCTGAAT
ATTCACAACGTTGATGAGCAATTAAATTTTGGAGATGTAAGCAAGATTGATAAGAAAAAACTACCTGAAT
ATTCACAACGTTGATGAGCAATTAAATTTTGGAGATGTAAGCAAGATTGATAAGAAAAAACTACCTGAAT
ATTCACAACGTTGATGAGCAATTAAATTTTGGAGATGTAAGCAAGATTGATAAGAAAAAACTACCTGAAT
ATTCACAACGTTGATGAGCAATTAAATTTTGGAGATGTAAGCAAGATTGATAAGAAAAAACTACCTGAAT
------------------------------------------------TGATAAG---------------

TTGATATTTTAGTTGGAGGATCTCCTTGTCAAAGCTTTAGTGTAGCCGGCCATCGAAAGGGATTTGAAGA
TTGATATcTTAGTTGGAGGATCTCCTgGcCAAAGCTTTAGTGTAcCCcGCCATCGAAAGGGATTTGAAGA
TTGATATTTTAGTTGGAGGATCTCCTgGcCAAAGCTTTAGTGTAGCCCGGgcTCGAAAGGGATTTGAAGA
TTGATATcTTAGTTGGAGGATCTCCTgGcCAAAGCTTTAGTGTAGC------------------------
TTGATcTTTTAGTTGGAGGATCTCCTgGcCAAAGCTTTAGTGTAGCCGGCCATCGAAAGGGATTTGAAGA
----------------------------------------------------------------------

TACAAGAGGTACCTTGTTTTTTCAATACGTTGAGACACTTAAGGAAAAGCAACCAAAGTTTTTTGTTTTT
TACAAGAGGTACCTTGTTTTTTCAATACGTTGAGAntCTTAAGGAAAAGCAACCAAAGTTTTTTGTTTTT
TACAAGAGGgACCTTGTTTTTTCAATACGTTGAGACtCTTAAGGAAAAGCAACCAAAGTTTTTTGTTTTT
----------------------------------------------------------------------
TACAAGAGGgACaTTGTTTTTTCAATACGTTGAGACtCTTAAGGAAAAGCAACCAAAGTTTTTTGTTTTT
----------------------------------------------------------------------

GAAAATGTTAAAGGGTTGATCAACCATGATAAAGGAAATACATTAAATGTTATGGCTGAAGCTTTCAGTG
GAAAATGTTAAAGGGTTGATCAACCATGATAAAGGAAATACATTAAATGTTATGGCTGAAGCTTTCAGTG
GAAAATGTTAAAGGGTTGATCAACCATGATAAAGGAAATACATTAAATGTTATGGCTGAAGCTTTCAGTG
----------------------------------------------------------------------
GAAAATGTTAAAGGGTTGATCAACCATGATAAAGGAAATACATTAAATGTTATGGCTGAAGCTTTCAGTG
----------------------------------------------------------------------

AAGTTGGGTACAGAATTGATCTAGAGCTCCTGAATTCAAAATTCTTTAATGTTCCACAAAATAGGGAGCG
AAGTTGGGTACAGAATTGATCTAGAGCTCCTGAATTCAAAATTCTTTAATGTTCCACAAAATAGGGAGCG
AAGTTGGGTACAGAATTGATCTAGAGCTCCTGAATTCAAAATTCTTTAATGTTCCACAAAATAGGGAGCG
----------------------------------------------------------------------
AAGTTGGGTACAGAATTGAcCTAGAGCTgCTtAATTCAAAATTCTTTAATGTTCCAC-------------
----------------------------------------------------------------------

ACTTTACATAATTGGAATTAGAGAAGATTTAATTAAAAATGAAGAATGGTCTTTGGATTTTAAAAGAAAG
ACTTTACATAATTGGAATTAGAGAAGATTTAATTAAAAATGAAGAATGGTCTTTGGATTTTAAAAGAAAG
ACTTTACATAATTGGAATTAGAGAAGATTTAATTAAAAATGAAGAATGGTCTTTGGATTTTAAAAGAAAG
----------------------------------------------------------------------
----------------------------------------------------------------------
----------------------------------------------------------------------
```

```
GATATACTTCAAAAAGGGAAACAGAGATTGGTAGAATTAGATATTAAAAGCTTTAATTTTAGATGGACAG
GATATACTTCAAAAAGGGAAACAGAGATTGGTAGAATTAGATATTAAAAGCTTTAATTTTAGATGGACAG
GATATACTTCAAAAAGGGAAACAGAGATTGGTAGAATTAGATATTAAAAGCTTTAATTTTAGATGGACAG
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------

CTCAATCGGCTGCTACGAAGAGGCTAAAAGATTTATTAGAAGAATACGTTGATGAAAAGTACTACTTGAA
CTCAATCGGCTGCTACGAAGAGGCTAAAAGATTTATTAGAAGAATACGTTGATGAAAAGTACTACTTGAA
CTCAATCGGCTGCTACGAAGAGGCTAAAAGATTTATTAGAAGAATACGTTGATGAAAAGTACTACTTGAA
------------------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------TACTTG--

TGAAGATAAAACAAACAGTTTGATCAAAGAGTTGTCTACAAGTCGACTTAATGAAAATCTTACTGTTGAG
TGAAGATAAAACAAACAGTTTGATCAAAGAGTTGTCTACAAGTCGACTTAATGAAAATCTTACTGTTGAG
TGAAGATAAAACAAACAGTTTGATCAAAGAGTTGTCTACAAGTCGA------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------------------AAATC--------cAG

CAAGTAGGTAACATTAATCCCTCTGGTAATGGAATGAATGGAAATGTTTATAATTCATCTGGATTAAGCC
CAAGTAGGTAACATTAATCCCTCTGGTAATGGAATGAATGGAAATGTTTATAATTCATCTGGATTAnGCC
--------TAACATTAATCCCTCTGGTAATGGAATGAATGGAAATGTTTATAATTCATCTGGATTAAGCC
------------------------------------------------------------------------
------------------------------------------------------------------------
CAAGTAGGTAACATTAATCCCTCTGGTAATGGAATGAATGGAAATGTTTATAATTCATCTGGATTAAGCC

CCACAATTACCACTAATAAAGGAGAGGGACTGAAAATTGCAGTTGAGTACTCCAGAAAAAGCGGGCTTGG
CCACAATTACCACTAATAAAGGAGAGGGACTGAAAATTGCAGTTGAGTACTCCAGAAAAAGCGGGCTTGG
CCACAATTACCACTAATAAAGGAGAGGGACTGAAAATTGCAGTTGAGTACTCCAGAAAAAGCGGGCTTGG
-----------------------------------------------------------CGGGCTTGG
------------------------------------------------------------------------
CCACAATTACCACTAATAAAGGAGAGGGACTGAAAATTGCAGTTGAGTACTCCAGAAAAAGCGGGCTTGG

ACGAGAACTTGCTGTATCTCATACGCTTTCTGCTTCTGACTGGAGAGGATTGAATAGGAACCAAAAACAA
ACGAGAACTTGCTGTATCTCATACGCTTTCTGCTTCTGACTGGAGAGGATTGAATAGGAACCAAAAACAA
ACGAGAACTTGCTGTATCTCATACGCTTTnTGnTTCTGACTGGAGAGGATTGAATAGGAACCAAAAACAA
ACGAGAACTTGCTGTATCTCATACGCTTTCTGCTTCTGACTGGAGAGGATTGAATAGGAACCAAAAACAA
------------------------------------------------------------------------
ACGAGAACTTGCTGTATCTCATACGCTTTCTGCTTCTGACTGGAGAGGATTGAATAGGAACCAAAAACAA

AATGCAGTTGTTGAGGTAAGGCCAGTATTAACCCCAGAAAGGGGGGAGAAGCGACAAAATGGAAGAAGAT
AATGCAGTTGTTGAGGTAAGGCCAGTATTAACCCCAGAAAGGGGGGAGAAGCGACAAAATGGAAGAAGAT
AATGCAGTTGTTGAGGTAAGGCCAGTATTAACCCCAGAAAGGGGGGnGcAGCGACAAAATGGAAGAAGAT
AATGCAGTTGTTGAGGTAAGGCCAGTATTAACCCCAGAAAGGGGGGAGAAGCGACAAAATGGAAGAAGAT
------------------------------------------------------------------------
AATGCAGTTGTTGAGGTAAGGCCAGTATTAACCCCAGAAAGGGGGGAGAAGCGACAAAATGGAAGAAGAT

TTAAAGATGACGGTGAACCAGCATTTACAGTAAACACAATTGACAGACACGGGGTAGCGGTTGGAGAGTA
TTAAAGATGACGGTGAACCAGCATTTACAGTAAACACAATTGACAGACACGGGGTAGCGGTTGGAGAGT-
TTAAAGATGACGGTGAACCAGCATTTACAGTAAACACAATTGACAGACACGGGGTAGCGGTTGGAGAGTA
TTAAAGATGACGGTGAACCAGCATTTACAGTAAACACAATTGACAGACACGGGGTAGCGGTTGGAGAGTA
------------------------------------------------------------------------
TTAAAGATGACGGTGAACCAGCATTTACAGTAAACACAATTGACAGACACGGGGTAGCGGTTGGAGAGTA

TCCAAAATACAGAATTAGAAGATTAACACCGTTAGAGTGCTTTAGGCTACAGGCTTTTGATGACGAAGA
-----AATACAGAATTAGAAGATTAACACCGTTAGAGTGCTTTAGGCTACAGGCTTTTGATGACGAAGA
TCCAAAATACAGAATTAGAAGATTAACACnGTTAGAGTGCTTTAGGCTACAGGCTTTTGATGACGAAGA
TCCAAAATACAGAATTAGAAGATTAACACCGTTAGAGTGCTTTAGGCTACAGGCTTTTGATGACGAAGA
------------------------------------------------------------------------
TCCAAAATACAGAATTAGAAGATTAACACCGTTAGAGTGCTTTAGGCTACAGGCTTTTGATGACGAAGA

TTTTGAAAAAGCTTTTGCTGCGGGAATAAGTAACTCACAATTATATAAGCAAGCCGGTAATTCAATTAC
TTTTGAAAAAGCTTTTGCTGCGGGAATAAGTAACTCACAATTATATAAGCAAGCCGGTAATTCAATTAC
TTTTGAAAAAGCTTTTGCTGCGGGAATAAGTAACTCACAATTATATAAGCAAGCCGGTAATTCAATTAC
TTTTGAAAAAGCTTTTGCTGCGGGAATAAGTAACTCACAATTATATAAGCAAGCCGGTAATTCAATTAC
```

```
------AAAAGCTTTTGCTGCGGGAATAAGTAACTCACAATTATATAAGCAAGCCGGTAATTCAATTAC
TTTTGAAAAAGCTTTTGCTGCGGGAATAAGTAACTCACAATTATATAAGCAAGCCGGTAATTCAATTAC

TGTAACTGTGCTTGAGTCAATATTCAAGGAATTAATACATACATACGTTAATAAAGAATCTGAATAACTC
TGTAACTGTGCTTGAGTCAATATTCAAGGAATTAATACATACATACGTTAATAAAGAATCTGAATAACTC
TGTAACTGTGCTTGAGTCAATATTCAAGGAATTAATACATACATACGTTAATAAAGAATCTGAATAACTC
TGTAACTGTGCTTGAGTCAATATTCAAGGAATTAATACATACATACGTTAATAAAGAATCTGAATAACTC
TGTAACTGTGCTTGAGTCAATATTCAAGGAATTAATACATACATACGTTAATAAAGAATCTGAATAACTC
TGTAACTGTGCTTGAGTCAATATTCAAGGAATTAATACATACATACGTTAATAAAGAATCTGAATAACTC

GACGCGTCTGCAGAAGCTTCTAGAATTCGAGCTCCCGGGGTACCATGGCATGCATCGATAGATC
GACGCGTCTGCAGAAGCTTCTAGAATTCGAcgTCCCGGGGTACCATGGCATGCATCGATAGATC
GACGCGTCTGCAGAAGCTTCTAGAAaTaGAGCaCCCGGGGTACCATGGCATGCATCGATAGATC
GACGCGTCTGCAGAAGCTTCTAGAATTCGAGCTCCCGGGGTACCATGGCATGCATCGATAGATC
GACGCGTCTGCAGAAGCTTCTAGAATTCGAGCTCCCGGGGTACCATGGCATGCATCGATAGATC
GACGCGTCTGCAGAAGCTTCTAGAATTCGAGCTCCCGGGGTACCATGGCATGCATCGATAGATC
```

APPENDIX FIVE

LIST OF FIGURES AND TABLES