Exploring The Structure And Function Of Bacterial Cytosine Specific DNA Methyltransferases Using Site-directed Mutagenesis

A thesis submitted to the
University of Sheffield
for the award of degree of Doctor of Philosophy
by

Qaiser Iftikhar Sheikh

Department of Molecular Biology and Biotechnology
University of Sheffield,
ENGLAND

May 2001

This Thesis is

Dedicated to my

Loving wife

ACKNOWLEDGEMENT

I owe special thanks to my supervisor Professor David Hornby who supported me whenever I was left in the middle of nowhere. In fact his continuous help in every respect and astute discussion throughout this study encouraged me to accomplish this piece of work.

I am grateful for my committee advisors Dr. Clive Price and Dr. Iain Murray for their guidance and useful discussions during my study. The assistance provided by Dr. Geoffrey Ford regarding the structural aspects of my experiments was priceless. I would also like to appreciate the work of Dr. Arthur Moir and Paul Brown for synthesizing oligonucleotides and nucleotide sequencing.

I am indebted to my friends in rooms D25, D27, and D29 who helped me over the course of study in the past and up until now. I am obliged for Dr. Greg Fowler for reading the draft of my thesis without any reluctance, Dr. Matthew Conroy for molecular modelling and diagrams, Mark Dickman, Maryam Matin, Adel Alghanim, John Ashby, Abdul-Aziz Aldukhyil, Abdul-Aziz Al-Sawailem, Ewan Murray, Jarawaree Snidwongse (Fah), Munsen Kan, Vahid Yasaee and Piyama Tasanasuwan, Vicki Harris, James Rose, Tracy Ware, Lindsy Hill and everyone else for creating a cordial atmosphere at present and in past in the lab. I would also like to acknowledge Mrs. Pam Smith for her help throughout.

I am most grateful to my sponsors and Transgenomic Inc. for their support and contribution in this project.

I cannot forget my parents, the only brother and two sisters whose prayers and love were always with me. I owe special thanks to my wife Farzana for her moral support and deep understanding, and my kids, Hassaan, Usaama and Nisha whose love made me feel strong.

(Qaiser Iftikhar Sheikh)

SUMMARY

Exploring The Structure And Function Of Bacterial Cytosine Specific DNA Methyltransferases Using Site-directed Mutagenesis

Submitted to

The University of Sheffield

For the degree of

Doctor of Philosophy

Qaiser Iftikhar Sheikh

Point mutations were engineered into the sequence of the multispecific DNA methyltransferase (Mtase) M.SPRI in motif IX, in order to mimic the corresponding motif IX of mono-specific Mtase. A similar approach was adopted to modify the sequence of the monospecific enzyme M.*Hha*I in motifs IX and X based on the available structure and as a consequence the enzyme regained methylation potential. It was thought that these changes might be sufficient to enable functional exchange of the target recognition domains (TRDs) between a mono- and a multispecific enzyme. However, insertion of various segments of TRD region from M.SPRI into the M.*Hha*I was not successful (Chapter 4). To establish whether mono- and multispecific Mtases are incompatible in terms of sequence exchanges, a systematic "swapping" of motifs was carried out (Chapter 5). These experiments suggested that there are some enzyme-specific structural interactions between different subunits within each class of Mtases.

In second half of this thesis a bacterial two-hybrid system based on the reversible assembly of an engineered form of M.SPRI was developed (Chapter 6). However the Mtase protein does not assemble into an active species until a DNA segment encoding a leucine zipper motif is fused to each of the two halves. Co-transformation of *E. coli* with the plasmids expressing the C-terminal and N-terminal domains respectively resulted in the abolition of colonies on double antibiotic plates, when an mcr^+ strain was used as host.

High performance liquid chromatography was used to estimate the extent of modification of plasmids indirectly. The extent of methylation at specific sequences within a plasmid molecule was readily detected by the corresponding differential susceptibility to digestion by specific restriction enzymes. Using this approach it proved possible to detect different levels of activity produced by wild type and mutant recombinant DNA Methyltransferases with sensitivity and in a semi quantitative manner.

In order to analyse the biochemical properties of Mtase, I have developed an *in vitro* translation-modification assay. Binary studies with the mutants (from Chapter 3 and 5) showed that there were no detectable sequence-specific recognition differences between these enzymes. Taken together, these results suggest that motif IX plays a role in general stabilisation of the enzyme core structure and has a less significant role in DNA recognition.

TABLE OF CONTENTS

CHAPTER THREE
REDESIGNING OF MOTIF IX OF MONOSPECIFIC Mtase M.*HhaI*

CHAPTER FOUR
CONSTRUCTION OF CHIMAERAS OF M.*Hha*I BY INSERTION MUTAGENESIS

CHAPTER SEVEN
GENERAL DISCUSSIONS AND FUTURE WORK

List of Tables

List of Appendices

Abbreviations

5mC	5-methylcytosine
Amp	Ampicillin
ATP	Adenosine tri-phosphate
bp	Base pair
BSA	Bovine serum Albumin
C5-mC	C5-methylcytosine
C5 Mtase	Cytosine (carbon-5) DNA methyltransferase
cDNA	Complementary deoxyribonucleic acid
CIP	Calf intestinal phosphatase
CPK	Corey, Pauling and Koltun
C-terminus	Carboxyl terminus
Da	Dalton(s)
dA	deoxyadenosine
dC	deoxycytidine
dG	deoxyguanosine
DMF	N, N-dimethyl formamide
DNA	Deoxyribonucleic acid
dNTPs	2'-deoxyribonucleoside-5'-triphosphate (where N is any nucleotide)
ds	Double stranded
dT	deoxythymidine
DTT	Dithiothreitol
E. coli	*Escherichia coli*
E.F.	Elution fraction
EDTA	Ethylenediamine tetra-acetic acid
Enase	Restriction endonuclease
EtBr	Ethidium bromide
EtOH	Ethanol
GST	Glutathion-S-transferase
IPTG	Isopropyl-β-thiogalactopyranoside
kb	Kilobase pair(s) of nucleotides

kDa	kilo Dalton(s)
*Lac*I	I, the gene for the lactose repressor protein
LB	Luria-Bertani media
M	Modification
mcr	Modified cytosine restriction
MOPS	(3-[N-Morpholino]propane sulfonic acid)
mrr	Methylated adenine recognition and restriction
mt	Mutant type
Mtase	Methyltransferase
MW	Molecular weight
N^4-mC	N^4-methylcytosine
N^6-mA	N^6-methyladenine
NEM	N-ethylmaleimide
N-terminus	Amino terminus
OD	Optical density
Oligo	Oligodeoxynucleotide
ORF	Open reading frame
Ori	Origin of replication
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenyl methyl sulphonyl fluoride
R	Restriction
R-M	Restriction and modification
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
RU	Response unit
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS- polyacrylamide gel electrophoresis

SPR	Surface plasmon resonance
T7	T7 Promoter
TAE	Tris-acetic acid-EDTA
TBE	Tris-boric acid-EDTA
TEAA	Triethylammonium acetate
TEMED	N,N,N',N'-tetramethyl-1,2-diaminoethane
TIR	Total internal reflection
TRD	Target recognition domain
Tris-base	(Tris[hydroxymethyl]amino methane)
UV	Ultraviolet
wt	Wild type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

All other abbreviations are explained in the text, or are the same as the internationally set to rules in "Biochemical Nomenclature and Related Documents" (1978).

CHAPTER ONE

INTRODUCTION

CHAPTER ONE

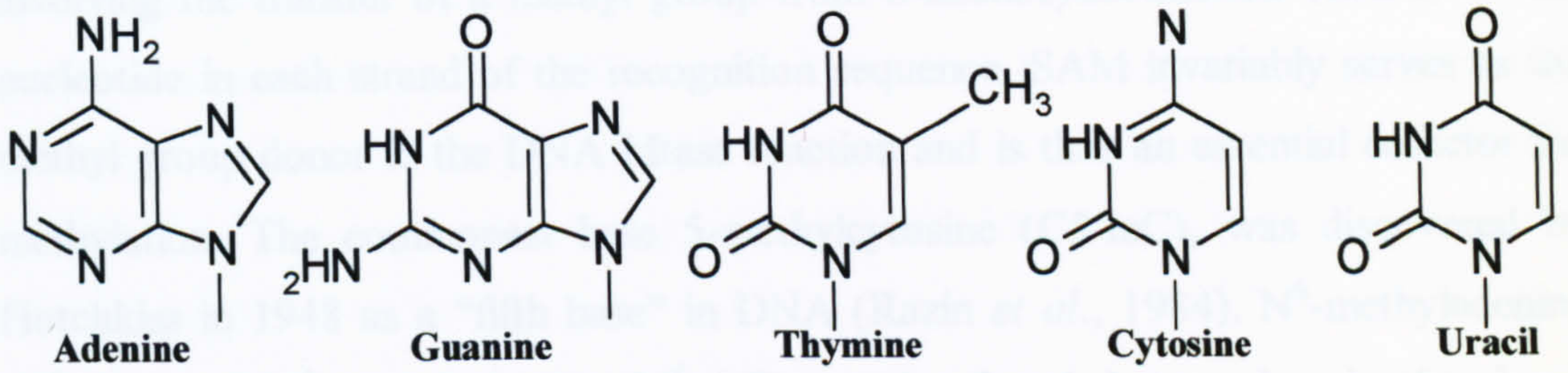
INTRODUCTION

1.1: DNA methylation

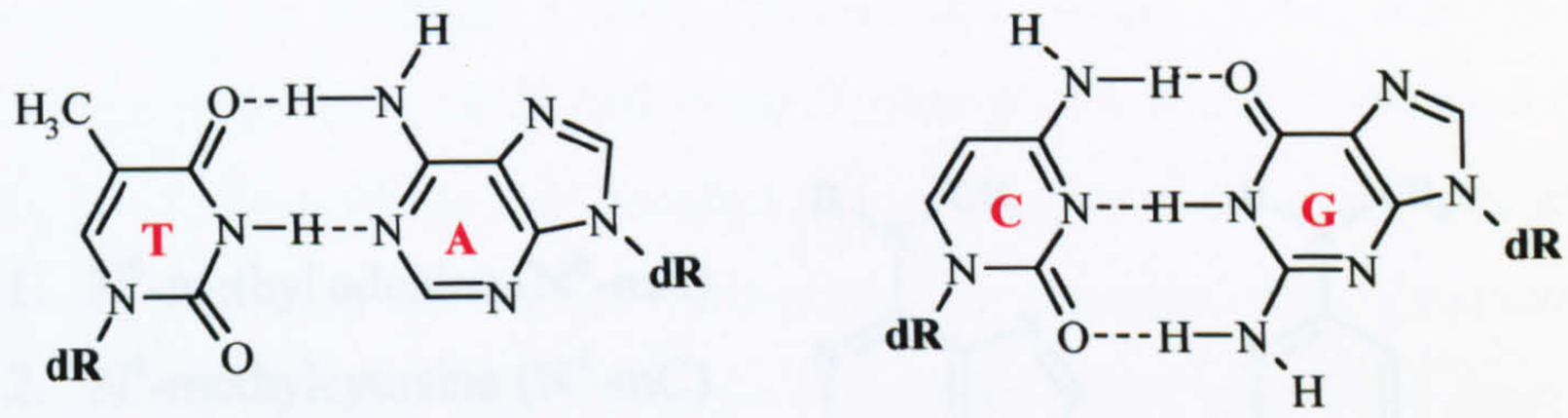
In all cells, nitrogenous bases that are strung together with sugars and phosphate molecules as polymers of nucleic acids (Figure 1.1) form the genetic material, which provides the information specifying all of the proteins necessary for life. The goal of understanding how each individual gene directs the formation and properties of cells irrespective of their origin remains one of the major challenges of science. The information obtained from recombinant DNA technology has already formed the basis for a new industry: biotechnology, and for new advances in medical treatment, particularly gene therapy and in the field of chemical diagnostics. However, a comprehensive understanding of information flow in biological systems at the molecular level remains a challenge for molecular biologists. The availability of sequence specific restriction enzymes and Methyltransferases (Mtases) was one of the key elements that led to the rapid developments in molecular biology (Sambrook *et al.*, 1989; Cheng, 1995).

Proteins and RNA are the working molecules of a cell, which carry out the program of activities encoded by genes. Some proteins build the structure of the cell, while others process materials by catalysing chemical reactions. For example, the enzymatic processes that are responsible for authentic bacterial replication include polymerisation, a range of base modification and DNA hydrolysis: specific enzymes or more often multiprotein enzyme complexes carry out each type of reaction. Modification enzymes protect a bacterium's own DNA from endonuclease-mediated hydrolysis or cleavage, by modifying it at or near each potential cleavage site. Modification enzymes typically add a methyl group to one or two bases usually within the restriction site at a specific nucleotide sequence. The DNA of most organisms contains one or more minor bases. Minor modified bases are not introduced as DNA is synthesised, but are produced after replication by modification of specific nucleotides already present in duplex DNA (Adams, 1990). DNA Mtases catalyse the addition of a methyl group to one nucleotide in each strand of the recognition sequence. DNA Mtases are classified on the basis of their ability to

a) Major bases found in DNA and RNA



b) The Watson-Crick base pairs



c) The primary structure of DNA

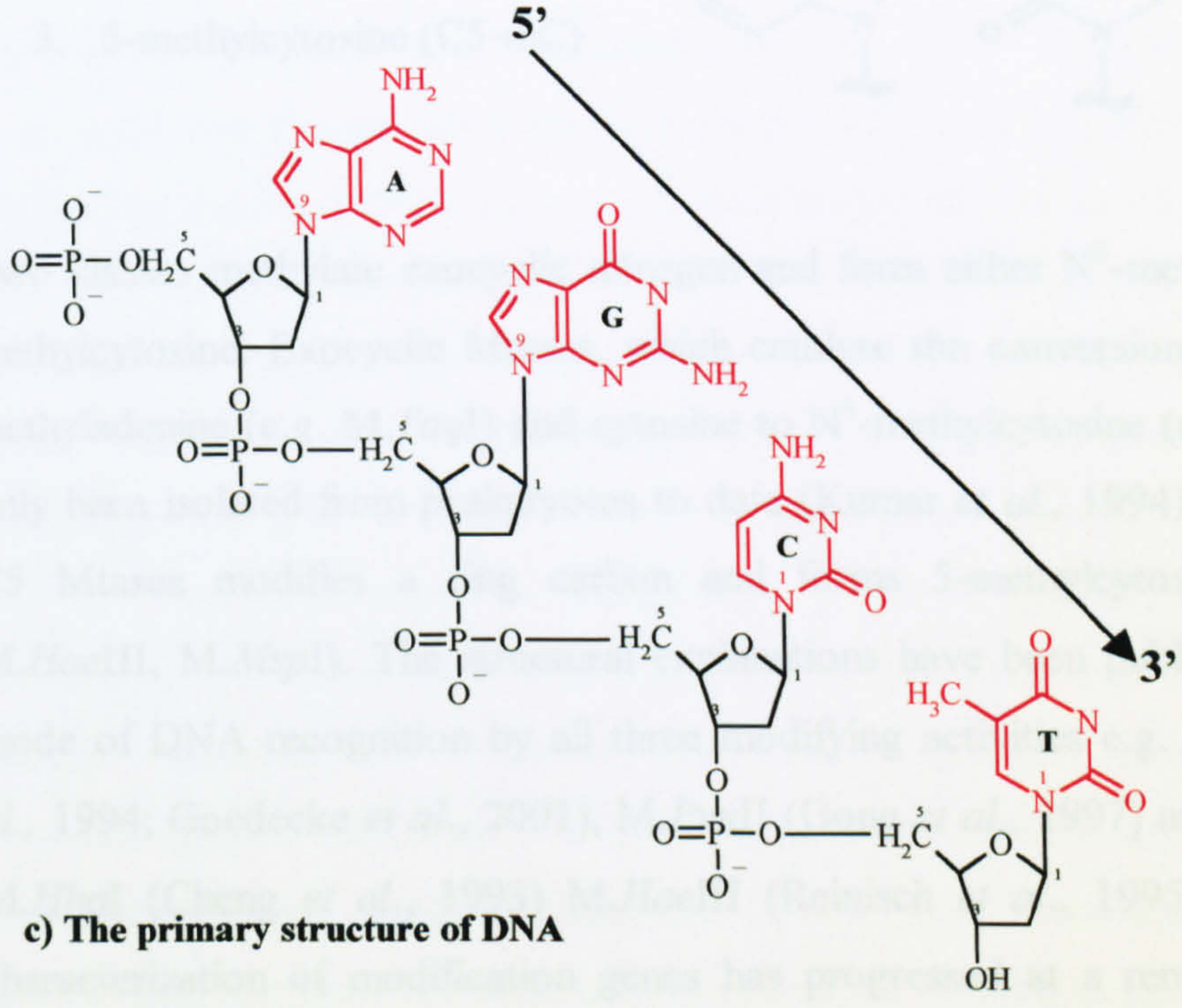
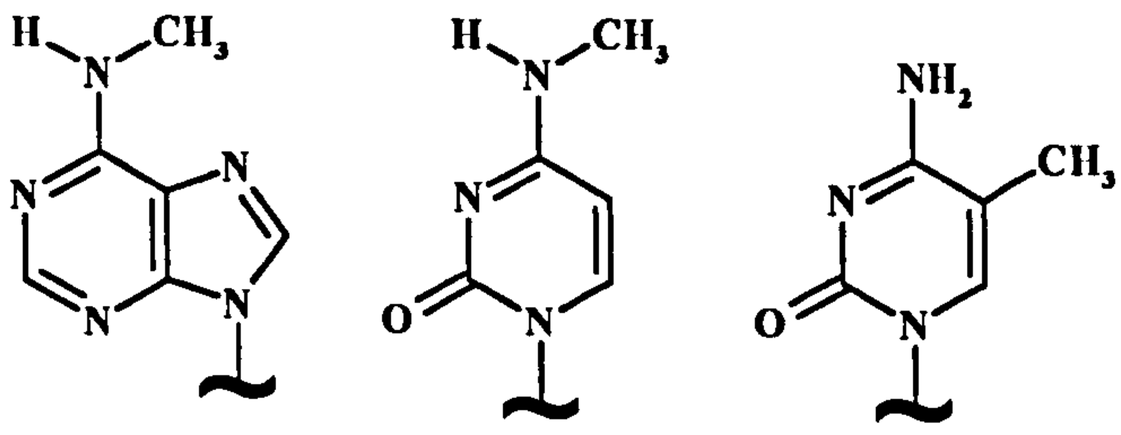


Figure 1.1: (a) Four major bases found in nucleic acids: Uracil is restricted to RNA while thymine is generally confined to DNA. (b) The regular hydrogen bonded base pairs found in DNA. (c) The polynucleotide arrangement of bases in a single strand of DNA showing the polar nature of the polymer with its sugar-phosphate backbone. **dR** represents 2'-deoxyribose.

modify bases in the DNA. Based on the chemical nature of modification of nucleotides, DNA Mtases fall into three classes (Wilson and Murray, 1991) each involving the transfer of a methyl group from S-adenosylmethionine (SAM) to one nucleotide in each strand of the recognition sequence. SAM invariably serves as the methyl group donor in the DNA Mtase reaction and is thus an essential cofactor for methylation. The commonest base 5-methylcytosine (C5-mC), was discovered by Hotchkiss in 1948 as a "fifth base" in DNA (Razin *et al.*, 1984). N^6-methyladenine (N^6-mA) and N^4-methylcytosine (N^4-mC) are also found, but are low in abundance by comparison. Therefore a total of three different modifications are produced by the action of DNA Mtases.

1. N^6-methyl adenine (N^6-mA)
2. N^4-methylcytosine (N^4-mC)
3. 5-methylcytosine (C5-mC)



Two classes methylate exocyclic nitrogen and form either N^6-methyladenine, or N^4-methylcytosine. Exocyclic Mtases, which catalyse the conversion of adenine to N^6-methyladenine (e.g. M.*Taq*I) and cytosine to N^4-methylcytosine (e.g. M.*Pvu*II), have only been isolated from prokaryotes to date (Kumar *et al.*, 1994). The third class of C5 Mtases modifies a ring carbon and forms 5-methylcytosine (e.g. M.*Hha*I, M.*Hae*III, M.*Msp*I). The structural explanations have been published regarding the mode of DNA recognition by all three modifying activities e.g. M.*Taq*I (Labahn *et al.*, 1994; Goedecke *et al.*, 2001), M.*Pvu*II (Gong *et al.*, 1997) and 5-methylcytosine M.*Hha*I (Cheng *et al.*, 1993) M.*Hae*III (Reinisch *et al.*, 1995). The cloning and characterization of modification genes has progressed at a remarkable rate. More than 300 putative Mtases have been cloned and sequenced (Roberts and Macelis, 1999) (see REBASE database at http://rebase.neb.com/rebase/rebase.html for complete listing available electronically). C5 Mtases, that form the basis of this study, transfer a methyl group from SAM to the 5-position of cytosine within a range of short DNA recognition sites (e.g. GCGC for M.*Hha*I, GGCC for M.*Hae*III and CCGG for M.*Msp*I).

When a methyl group is present at a specific position the cognate restriction endonuclease (Enase) is prevented from cutting the DNA. Together with the restriction Enase the methylating enzyme forms a Restriction-Modification (R-M) system that protects the host DNA while it destroys foreign DNA. Mostly in R-M systems the two activities are separate, however in some systems the two activities are combined in a single multisubunit enzyme (Wilson & Murray, 1991).

1.2: Restriction And Modification (R-M)

The phenomenon of restriction and modification (R-M) of bacterial viruses was first described fifty years ago (Luria and Human, 1952 and Bertani and Weigle in 1953) also reviewed by Heitman, (1993). Bertani and Weigle (1953) 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. Specific enzymes that catalyse the hydrolysis of specific DNA sequences form the basis of this underlying phenomenon of restriction (Arber, 1979). These enzymes are referred to as restriction Enases. Most restriction enzymes cleave unmethylated DNA and are blocked by modified DNA. In 1964 Gold and Hurwitz proposed that a DNA Mtase was responsible for the phenomenon of modification in *E. coli*. Mtase recognises and methylate the same DNA sequence to protect the host chromosome from Enase action. Resistance to restriction occurs when the Mtase modifies each phage borne recognition site prior to Enase action. Kuehnlein in 1969 isolated the enzyme from *E. coli* that was responsible for methylation of DNA and the enzyme was finally purified in 1972 (Kuehnlein and Arber, 1972). The molecular explanation has now become apparent: all restriction-modification systems include two opposing sequence specific enzymatic activities: (for a review, see Wilson and Murray, 1991, Bickle and Kruger, 1993) an Enase and a DNA Mtase. Enases and Mtases from the same system recognise the same sequences and are referred to as cognate enzymes.

1.2.1: NOMENCLATURE AND PROPERTIES OF R-M SYSTEM

Several kinds of RM enzymes have been discovered (Wilson and Murray, 1991). They appear to perform equivalent jobs but in opposing ways. More than two dozen bacterial and archaeal genomes have been completely sequenced within last 5 years

which has revealed a remarkable fact: >80% of the genomes appear to have at least one R-M system and most of these genomes appear to contain multiple R-M systems (Kong *et al.* 2000). The genes for restriction and modification enzymes appear to be closely linked. *Escherichia coli* K-12 contain at least four restriction systems to monitor the "origin" of invading DNA and determine its fate (Bickle and Kruger, 1993).

Bacterial restriction and modification (R-M) systems are traditionally divided into three classes, designated as type I, II and III on the basis of enzyme subunit composition, co-factor requirements, recognition sequence symmetry, cleavage position and reaction products (Wilson and Murray, 1991).

1.2.1.1: Type I

Type I restriction modification enzymes are the most complex and least diverse types known. Members of this group have similar enzymatic subunits, but different specificity subunits. The main enzyme comprises the products of the *hsdS*, *hsdM*, and *hsdR* genes (*hsd* for host specificity for DNA). The three polypeptide chains termed Restriction (R), Modification (M), & Specificity (S) and the resulting pentameric (R_2M_2S) multifunctional enzyme complexes catalyse both Enase and Mtase reactions. These complexes cleave at random sites distant from the site of recognition and cleavage, hydrolyse ATP, and are inactive after a single round of DNA scission. 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. These types of restriction enzymes have been found in *E. coli*, *Citrobacter* and *Salmonella* and have been grouped into three families.

1) IA (e.g. *Eco*KI) 2) IB (e.g. *Eco*AI) 3) IC (e.g. *Eco*R124I)

1.2.1.2: Type II

Type II R-M enzymes are the simplest and appear to be more common than the type III and I. Type II restriction and modification enzymes possess separate target recognition domains (TRD). The specificities of cognate enzymes are the same. Type II R-M enzymes possess varied specificities. Type II Enases and Mtases act independently and have simple requirements. The Enases require Mg^{+2} and the Mtases require SAM. Cleavage by type II Enases occurs symmetrically within the

recognition sequence. A Type II recognition sequence is essentially a symmetric sequence consisting of up to eight base pairs (bp).

1.2.1.3: Type IIs

Some Type II enzymes have been further sub classified as Type IIs because of their asymmetric recognition sequence: these enzymes usually cleave at some distance from the recognition sequence e.g. *Fok*I. The Type IIs R-M enzymes also have separate Enases and Mtases: their target sites, however, are asymmetric, non-palindromic sequences. The endonucleolytic scission catalyzed by type IIs Enases occurs at enzyme specific locations within 20 bp outside of the target sequence (only one active site is involved). Methylation of the two complementary, but non-identical target sequences require usually two different Mtases: one for each strand of the target.

1.2.1.4: Type III

The enzymes typical of this class are multi-subunit structures with the capacity to mediate both modification and restriction. SAM serves as the methyl group donor and is also required for cleavage by the complex. Adenine (specifically at N^6) is the only base that is known to be methylated by these enzymes.

Roberts, (personal communication) has proposed a new nomenclature for RM enzymes with some modifications to overcome several problems and limitations with the present system of nomenclature that cannot easily be fixed. In his proposal the original subdivisions of Type I, II and III were maintained with subdivisions within each major category to reflect either biochemical or sequence characteristics as outlined in table 1.1.

1.2.1.5: Genetics of R-M systems

The modification (M) and specificity (S) genes in type I enzymes are transcribed as a single operon and the restriction (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 opposite orientations: some transcription units diverge while others converge. Type III R and M gene are transcribed as a single operon in the same

RM Nomenclature		Definition	Example
Type I	I A	Contain R, M and S sub units, require ATP for their action and cleave at random positions away from the recognition sequence.	*Mpu*UI
Type II	II S	Recognise asymmetric sequences and cleave away from the recognition sequence.	*Hph*I
	II T	Resemble Type IIs, but are composed of heterodimeric subunits.	*Bsr*BI
	II B	Cleave on both sides of the recognition sequence	*Bcg*I
	II E	Interact with two copies of their recognition sequence, one being the actual target of cleavage, the other being the allosteric effector.	*Eco*RII
	II F	Interact with and cleave two copies of their recognition sequence but otherwise have the appearance of being normal.	*Sfi*I
	II G	Stimulated by SAM and have both R and M domains within a single polypeptide but otherwise resemble Type II.	*Eco*571
	II D	Recognise methylated DNA and cleave specifically.	*Dpn*I
	II M	Cleave methylated DNA but without the clean recognition specificity of IID.	McrBC
	II H	These are halfway between Type I and II in their genetic makeup, but look like Type II if only the enzymatic activity is considered	*Ahd*I
	II X	This is for oddballs that may ultimately need a separate category	
Type III		Contain two subunits for restriction and modification, require ATP for cleavage, cleave at a specific distance away from their recognition sequence and are easily recognised because of their apparent similarity at the sequence level.	*Eco*PI

Table 1.1: The three types of RM system (I, II, and III) with definition of each subdivision illustrated with an example for each. Roberts proposed nomenclature (personal communication).

orientation (Wilson, 1991). The general properties and the gene organisation of R-M systems are summarised in table 1.2.

1.2.1.6: Solitary Methylation by commonly used strains of *E. coli*

A biological function of post-replicative DNA methylation is observed in common strains of *E. coli*. Most laboratory strains of *E. coli* contain three site specific DNA Mtases. Cells contain enzymes that methylate DNA, e.g. Dam, Dcm and *Eco*KI Mtases.

The Dam Mtase (encoded by the *dam* gene) transfers a methyl group from SAM to the N^6 position of the adenine residue in the sequence "GATC". Several restriction sites contain this sequence (or part of it followed by nucleotides to complete "GATC") and therefore are unable to restrict Dam modified DNA. There are about 16 restriction sites that are blocked by Dam methylation and common examples are *Xba*I, *Nru*I and *Bcl*I. When it is necessary to cleave plasmid DNA containing these sequences which are blocked by Dam methylation, the DNA must be prepared from strains of *E. coli* that are *dam*$^-$ (New England Biolabs Catalogue, 2000, Sambrook *et al.*, 1989).

The Dcm Mtase (encoded by the *dcm* gene) methylates the internal cytosine residues in the sequence CCAGG and CCTGG at the C5 position (~1 site per 512 bp) in DNA of random sequence (New England Biolabs Catalogue, 2000, Sambrook *et al.*, 1989). There are about 20 restriction sites which are blocked by Dcm methylation, and therefore it is essential to prepare the plasmid DNA in a *dcm*$^-$ strain of *E. coli*. The common examples of restriction sites that can be blocked in this way are *Stu*I and *Alw*NI.

M.*Eco*KI (the *Eco*KI Mtase) modifies adenine residues in the sequences AAC(N6)GTGC and GCAC(N6)GTT. *Eco*KI sites are much less common than Dam or Dcm sites in a random DNA sequence. However, it has also been pointed out that DNA isolated from *E. coli* is not always methylated to the same extent. Few examples exists that these sites are not blocked by Dam or Dcm Mtase (New England Biolabs Catalogue, 2000).

The other reason to consider these Mtases is that the modification state of the plasmid DNA can affect the frequency of transformation in special situations. For example, when Dam-modified plasmid DNA is introduced into *dam*$^-$ *E. coli*, it has been shown that replication initiation is suppressed when plasmid DNA is

System and example	Genes	Specificity	Type of methylation	Cofactor Requirement	Gene Organisation	Location of TRD	Site of Cleavage with respect to Recognition sequence
Type I *Eco*K	RMS	AACN$_6$G<u>T</u>GC	N^6-mA	ATP, SAM, Mg^{+2}		Within specificity determining subunit	Remote from Target
Type II *Msp*I	RM	<u>C</u>CGG	C^5-mC N^4-mC N^6-mA	SAM, Mg^{+2}		Within indivdual enzymes	Within the target
Type IIs *Mbo*II	RM	GAAG<u>A</u>	C^5-mC N^6-mA	SAM, Mg^{+2}		Within indivdual enzymes	A defined location within 20bp on one side of the target
Type III *Eco*PI	RM	AG<u>A</u>CC	N^6-mA	ATP, SAM, Mg^{+2}		Within Mtase subunit	At a defined distance of 25-27 bp from the 3' site of the target

Table 1.2: Some general properties of Restriction-Modification Systems with target sequences and gene structures. The specificity of the RM systems shown is indicated, with the methylated bases underlined. The arrows describe gene for Mtase (Green), Enase (red) and blue (specificity determining subunits). The arrow's direction gives the direction of gene transcription. The numbers of amino acids for the corresponding proteins are given below the arrows. The map is not drawn to scale (Data adapted from Noyer-Weidner and Trautner, 1993; Wilson, G.G., 1991).

hemimethylated at Dam sites (Russell and Zinder, 1987). Dam-modified plasmids therefore replicate once in *dam*⁻ cells, and are unable to replicate again. The phenomenon of poor transformation of other bacterial strains may be due to the presence of a modification-dependent restriction system. See below for a detailed discussion on this topic.

1.2.1.7: Solitary Restriction of methylated DNA by commonly used *E. coli*

In addition to the capacity to protect DNA against cleavage by restriction Enases, *E. coli* has several mechanisms for identifying foreign DNA and destroying it. Methylation can also provide a signal for the restriction of foreign DNA. In this case, foreign DNA is acted on because it does contain methylated bases in a particular sequence, rather than because it does not. These systems only restrict modified DNA substrates. DNA is protected by the absence of particular modified sequences (Raleigh, 1992). At least four mechanistically distinct classes of modification-dependent restriction system are known. The first and best known of these is *Dpn*I from *Streptococcus pneumoniae* (previously called *Diplococcus pneumoniae*) (Lacks and Greenberg, 1977). The *Dpn*I Enase is encoded by one gene, and will cleave DNA with the appropriate modified base, N6-methyladenine, in its recognition sequence. Three mechanistically distinct sequence-specific, modification-dependent systems McrA, McrBC and Mrr have been identified in *E. coli* K12 (reviewed by Raleigh, 1992; Bickle and Kruger, 1993). Several prokaryotic restriction Enases have been identified (reviewed by Noyer-Weidner and Trautner, 1993) which cleave DNA invading a cell provided it is methylated at specific DNA sequences. It is evident that cells encoding this type of restriction activity must not methylate such sequences, and to this end such Enases are solitary enzymes. Such systems also have practical significance since cells harbouring such restriction activities display restriction *in vivo* when DNA (e.g. plasmid DNA) transforms the cell.

The observation that Mtase encoding genes could not be cloned in some *E. coli* strains led to the discovery of two *E. coli* K12 restriction systems directed against methylated DNA: Mcr (Modified Cytosine Restriction) and Mrr (Modified adenine Recognition and Restriction) (Raleigh and Wilson 1986; Heitman and Model 1987; Raleigh 1987; Waite-Rees *et al.* 1991; Raleigh 1992).

The Mrr system was discovered due to an observed activity that cut DNA that had been methylated by the *Hha*II or *Pst*I Mtase, both of which methylate adenosyl

residues to form N^6-mA in specific sequences (Heitman and Model, 1987). Later it was found that some sequences containing 5mC are also restricted and this has been referred to as McrF restriction (Kelleher and Raleigh, 1991; Waite-Rees *et al.*, 1991). Mrr is the only system known to specifically hydrolyse DNA containing either methyladenine or 5-methylcytosine in particular sequences (de Bruijn *et al.*, 1998).

The Mcr phenotype was the first restriction system reported (Luria and Human, 1952) and RglA and RglB were the original name given to this phenomenon (Raleigh *et al.*, 1989). In the 1980s, the Rgl systems were rediscovered in the course of cloning experiments (Raleigh and Wilson, 1986; Noyer-Weidner *et al.*, 1986). It was subsequently shown that the two Mcr phenotypes, McrA and McrBC were identical to RglA and RglB respectively (Raleigh *et al.*, 1991). McrA is specified by a single gene and restricts DNA with modified cytosine (hydroxymethylcytosine or methylcytosine) followed by guanine *in vivo*.

McrBC is the most extensively characterised of the modification dependent systems in *E. coli*, (reviewed by Raleigh, 1992). Like McrA, it acts on DNA containing modified cytosine (5-hydroxymethylcytosine or N^4-methylcytosine). This system acts as a multi-subunit nucleotide-dependent Enase. Unlike other nucleotide-dependent nucleases, it displays an absolute requirement for GTP rather than ATP. Restriction activity requires the products of two genes *mcrB* and *mcrC* that express three proteins for activity both *in vivo* and *in vitro*.

The *mcrB* gene produces two gene products. The complete *mcrB* open reading frame produces a 51-kDa protein ($McrB_L$) and a 33-kDa protein ($McrB_S$). The smaller McrB polypeptide is produced from an in-frame, internal translation start site in the *mcrB* gene. The $McrB_S$ sequence is identical to that of $McrB_L$ except that it lacks 161 amino acids present at the N-terminus of the latter protein. It has been suggested that $McrB_L$ is the DNA binding restriction subunit. The function of $McrB_S$ was unknown and challenging since it is produced in frame with $McrB_L$. There was speculation that it played a role in the modulation of McrBC restriction. The first experiments published suggested that $McrB_S$ was necessary for stabilization of the McrBC restriction complex *in vivo* (Beary *et al.*, 1997). It has also been suggested that the role of $McrB_S$ is to modulate McrBC activity by binding to McrC (Panne *et al.*, 1998).

The *mcrC* gene product, a single protein (39 kDa), is thought to be a specificity subunit and contains a leucine heptad repeat (leucine zipper) and a cationic region

that could play a role in protein-protein interaction and affords less specificity to the mcrB gene products (Ross, *et al.*, 1987, Raleigh, 1992, Beary, *et al.*, 1997).

1.3: DNA Methyltransferases (Mtases)

DNA Mtases are found in a wide variety of Prokaryotes and Eukaryotes. The widespread presence of methylated bases in DNA of different organisms, the sequence specificity of various Mtases and the even distribution of the methylated bases along the chromosomes strongly suggest that modified bases in DNA are biologically significant. Several kinds of R-M systems have been discovered. They appear to perform equivalent biological functions but in different ways (Wilson, 1991). Although in general the biological significance of methylated bases in DNA is still unresolved, a number of recent experiments strongly suggest that DNA methylation is involved in major biological processes some of which are described below.

1.3.1: DNA METHYLATION IN PROKARYOTES

In prokaryotes, Mtases have most often been identified as elements of restriction modification systems in which they act to protect DNA from cleavage by the corresponding Enases. DNA methylation is the modification product, which is found in almost all bacteria, and seems to protect cells from foreign DNA molecules, particularly viruses. Hence, the main function of methylation in prokaryotes is to give protection to cells from viral infection; it primarily serves as an "immune response" to phage infection. Enases digest foreign DNA that enters the cell, thereby protecting the cell from genetic subversion.

The biological role of R-M systems was recently a subject of considerable debate. Whilst it has been suggested that they provide a bacterium with a primitive immune system to combat phage infection, more recently it has been reported that R-M genes may be selfish (Naito *et al.*, 1995; Naito *et al.*, 1998).

As discussed above, many R-M systems have been demonstrated to restrict phage growth. However, for most R-M systems this property has not been tested due to the difficulties of working with unusual bacteria, which may or may not have characterised phage. It is certainly true that a large number of R-M systems do not

restrict phage growth when expressed in *E. coli*, even though in several cases the enzymes are sufficiently well expressed (Heitman, 1993).

Where as the Enase activity of a R-M system cannot usually exist in a cell in the absence of its cognate Mtase activity, since this would lead to suicidal degradation of the cell's DNA, the reciprocal situation, the solitary existence of a Mtase, is often encountered. Moreover, although the genes for restriction and modification systems appear to be closely linked (Wilson, 1991) sequence comparisons of a large collection of different cloned R-M systems reveal limited sequence homologies (Wilson and Murray, 1991; Wilson, 1991; Heitman, 1993). With the exception of type I R-M systems (see below), which share considerable homology within, but not between families, there has been little homology identified between Enases of different sources. In contrast, Mtases share a limited degree of homology that is related to enzymatic constraints rather than being a reflection of common origin (Wilson and Murray, 1991). Even in cases of isoschizomeric R-M systems, which recognise and cleave identical DNA sequences, the vast majority shares no homology between Enases.

Considering this information, the emerging view is that sequence-specific Enases and DNA Mtases arose independently in evolution to serve functions other than R-M or at least other than the 'cellular defence' hypothesis (Heitman, 1993). This model, which envisages the pairing of independent enzymes and the subsequent tailoring of Enase and Mtase specificities, is perhaps even more plausible in the light of work by Naito *et al.* (1995). Here plasmids carrying gene pairs encoding type II R-M systems (see below) were shown to have increased stability in *E. coli*. The descendants of cells that had lost these genes appeared unable to modify a sufficient number of recognition sites in their chromosomes to protect them from lethal attack by the remaining Enase molecules. The authors suggest that the enforced retention and functional integrity of such systems on plasmids, is as a result of the R-M systems behaving as a 'selfish' unit, which enforces its retention in hosts whose chromosome(s) bear the specific sites it recognises.

As an explanation for the widespread occurrence, diversity and specificity of R-M systems, such a 'selfish gene' or 'selfish symbiont' hypothesis would seem as plausible as the 'cellular defence' hypothesis. Both hypotheses are consistent with the findings outlined above, that R-M genes are tightly linked and widely distributed. However, the diversity and high degree of specialisation in sequence recognition, in

particular the existence of 'rare cutters' (e.g. Enases which have long recognition sequences, which are unlikely to be present in many bacterial viruses) is more consistent with a selfish gene hypothesis than one of cellular defence. This idea has been tested in type I R-M genes encoding *Eco*KI and has been shown that type I genes are not 'selfish' (O'Neil *et al.*, 1997). Clearly more experiments are required to resolve these issues.

1.3.2: DNA METHYLATION IN EUKARYOTES

DNA methylation in eukaryotes appears to have a different and more diverse role than in prokaryotes, and is implicated in the control of several cellular processes. However this study is concerned exclusively with the properties of two prokaryotic Mtases and therefore the functions of methylation in eukaryotes is briefly reviewed. Several excellent reviews are available on this subject (Mostoslavsky and Bergman, 1997; Colot and Rossingol, 1999; Hsieh, 2000; Robertson and Wolffe, 2000: Robertson and Jones, 2000).

Most eukaryotic DNA Mtases are ~1600 amino acids long, single subunit enzymes that preferentially methylate hemimethylated DNA (Bestor *et al.*, 1988). The N-terminal two third of the protein is considered the regulatory domain and the C-terminus is the catalytic domain. They have much smaller target recognition sequences than the bacterial enzymes (Adam, 1990). In vertebrates methylation of cytosines takes place when the cytosine residue is followed by a guanosine (CpG). In plants, methylcytosine occurs both in symmetrical and sequences, CpG and CpNpG, and in non-symmetrical sequences, for example, CpApTp and CpTpT. The relative importance of symmetric and asymmetric methylation in regulating gene expression is unknown but as methylation at symmetric sequences can be transmitted through cycles of DNA replication, it seems likely that methylation of these sequences will play an integral role in regulating gene expression (reviewed by Finnegan *et al.*, 2000). The CpG Mtases found in higher eukaryote is not known to be involved in restriction and modification. The patterns of CpG methylation are heritable, tissue specific, correlate with gene expression and also play role in differentiation. Studies have shown that cytosine methylation regulates the activity of a gene. Methylation of cytosine residues in higher eukaryotes appears to be involved in the control of gene expression and has been implicated in gene regulatory processes such as transcription, tissue specific expression, and genomic imprinting (Adams, 1995).

In eukaryotes, methylcytosine is made enzymatically after DNA is replicated, and about 5 percent of the cytosines in mammalian DNA are converted to 5-methylcytosine. Eukaryotic DNA Mtases were believed to have possible alternative functions such as regulation of gene expression (Wilson, 1991, Wilson and Murray, 1991). It has now become clear that DNA methylation is involved in a diverse range of cellular processes and there is a wider acceptance that DNA methylation plays a variety of roles in different cancer types and probably at different stages of oncogenesis (Laird and Jaenisch, 1996). Methylation of DNA is important in many organisms and is essential in mammals. In eukaryotes, methylation is the only known covalent modification of DNA. At least in mammals it is absolutely essential (Li *et al.*, 1992). There have been many suggestions as to how DNA methylation contributes to the differential expression of genes (Gilbert, 1994). The methylation of promoter sequences contributes to the temporal and spatial regulation of these genes encoding tissue-specific proteins. This has been intensively studied in vertebrates. DNA methylation is also thought to be responsible for the continued repression of the genes on one of the two X chromosome in each female mammalian cell. In many cases the methyl group tags a gene, thereby silencing its expression and an unnecessary protein product is not produced in a particular cell (Razin, 1998). Work on animals, plants and fungi leaves little doubt that gene silencing is a major biological consequence of DNA methylation (Colot and Rossingol, 1999). Experiments indicate that gene silencing by methylation involves the generation of a chromatin structure (Razin, 1998). Normally, a methyl-binding protein binds to methylated CpG units and brings together histone deacetylases and the co-repressor protein mSin3A. This complex contributes to the formation of repressive chromatin states and the deacetylation of histones. As a result, transcription machinery cannot bind to DNA, and the protein cannot be translated (Wolffe, 1998).

In prokaryotes, the methyltransferases are simpler and lack regulatory N-terminal domain compared to mammalian Mtase *Dnmt1*. The C-terminal domain of *Dnmt1* is clearly related to the prokaryotic C5-Mtases, since it harbours a set of ten motifs that are involved in the catalytic reaction and whose sequence is shared with over 100 prokaryotic Mtases (Margot *et al.*, 2000). Recently a proven connection has been made between DNA methylation and development in eukaryotic cells: the first human diseases caused by defects in the DNA methylation machinery have been reported such as "Rett Syndrome" (Amir *et al.*, 1999) and "ICF Syndrome (Xu *et al.*,

1999; Hansen *et al.*, 1999) (reviewed by Robertson and Wolffe 2000). These studies linked the observed methylation defect in ICF syndrome to mutations in the *DNMT3B* gene. Although no "hot spot" has been identified and many of the mutations are heterozygous, all seem to affect the carboxy-terminal catalytic domain of DNMT3B.

1.3.3: MONO AND MULTISPECIFIC DNA Mtases

Most Mtases recognise a single DNA sequence (Wilson and Murray, 1991), however, multispecific C5 DNA Mtases are known that methylate several unrelated sequences (Tran-Betcke *et al.*, 1986). The major difference between the mono- and multispecific Mtases occurs in the variable region between motifs VIII and IX. The region is 80-120 amino acids in the monospecific Mtases and 175-268 amino acids in the multispecific Mtases (Wilson and Murray, 1991; Trautner *et al.*, 1996). Multispecific Mtases possess a sequentially arranged, interchangeable region that can recognise up to 5 target DNA sequences (Behrens *et al.*, 1987; Wilke *et al.*, 1988; Lange *et al.*, 1991; Walter *et al.*, 1992; Schumann *et al.*, 1996; Trautner *et al.*, 1996; Lange *et al.*, 1996). See section 1.3.5 for the detailed discussion regarding the variable region.

1.3.4: THE CONSERVED MOTIFS OF C5-DNA Mtase

Comparative analysis has shown that these proteins share an ordered set of ten sequence motifs, which alternate with non-conserved regions. All ten motifs can be identified in the majority of the known sequences, including the C-terminal 500 amino acids of the eukaryotic CpG Mtases (Bestor *et al.*, 1988; Lauster *et al.*, 1989). In the original analysis of thirteen C5 Mtases (Posfai *et al.*, 1989), five motifs were considered highly conserved (I, IV, VI, VIII, and X), and the remaining five, moderately conserved. However, when 44 Mtases were compared, it was shown that motif IX is also highly conserved and resulted in the inclusion of motif IX with in the highly conserved set (Kumar *et al.*, 1994) (Figures 1.2 and Figure 1.3). The motifs may be of use in identifying putative Mtases among protein sequences of unknown function (Posfai *et al.*, 1989). A few sequences have been detected in which unambiguous assignments for some of the motifs are not possible. These motifs (II, III, IX, X) are either more tolerant to sequence variation or are nonessential for function in some of the proteins (Kumar *et al.*, 1994). The largest non-conserved or

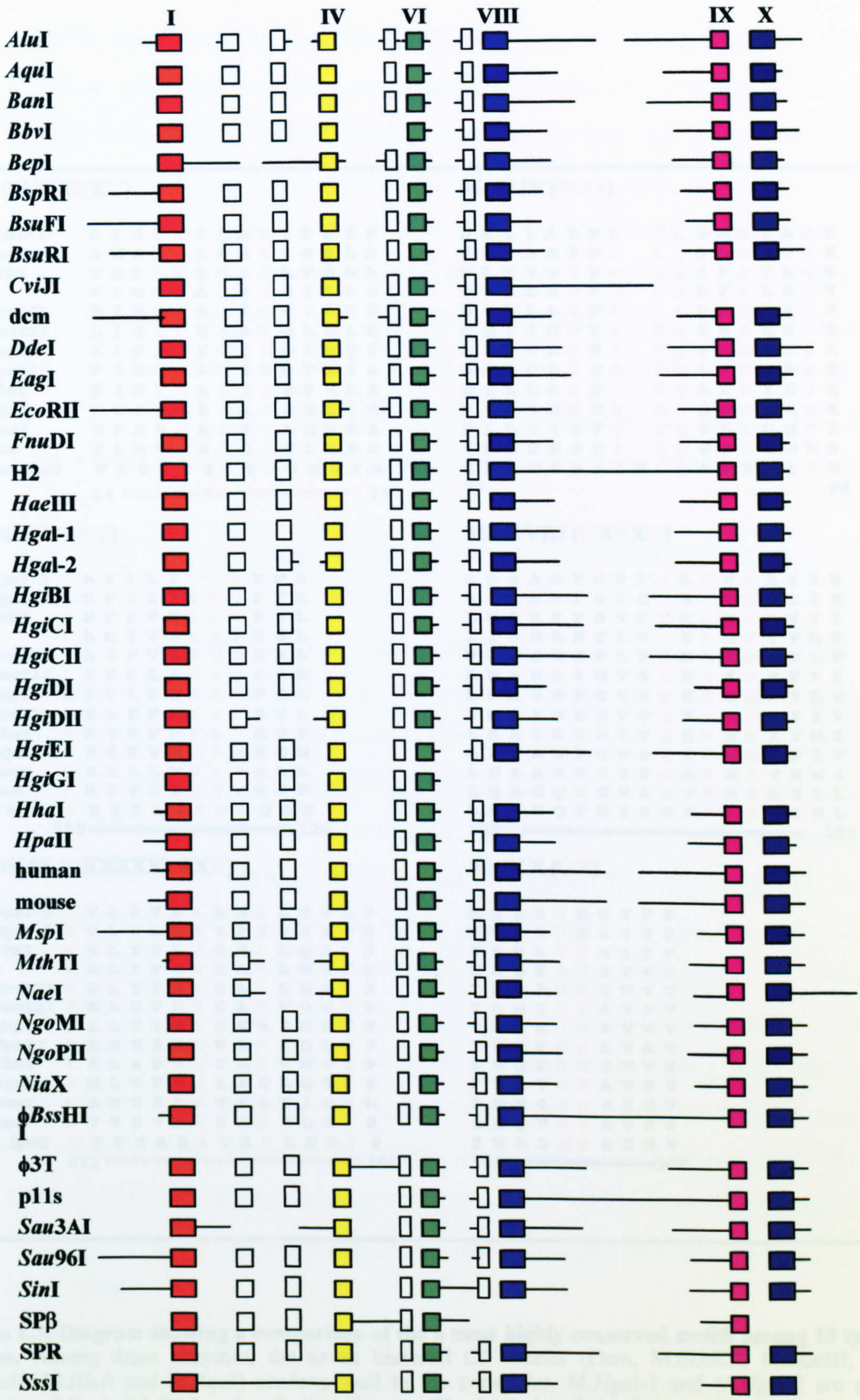


Figure 1.2 Diagrammatic representation showing the alignment of motifs of 44 C5-Mtases.
The six highly conserved motifs are coloured (red-motif I, yellow-motif IV, green-motif VI, cyan-motif VIII, magenta-motif IX, dark blue-motif X). Where boxes are missing, the motifs could not be equivocally assigned. Gaps are placed arbitrarily to allow alignment of motifs (Kumar *et al.*, 1994).

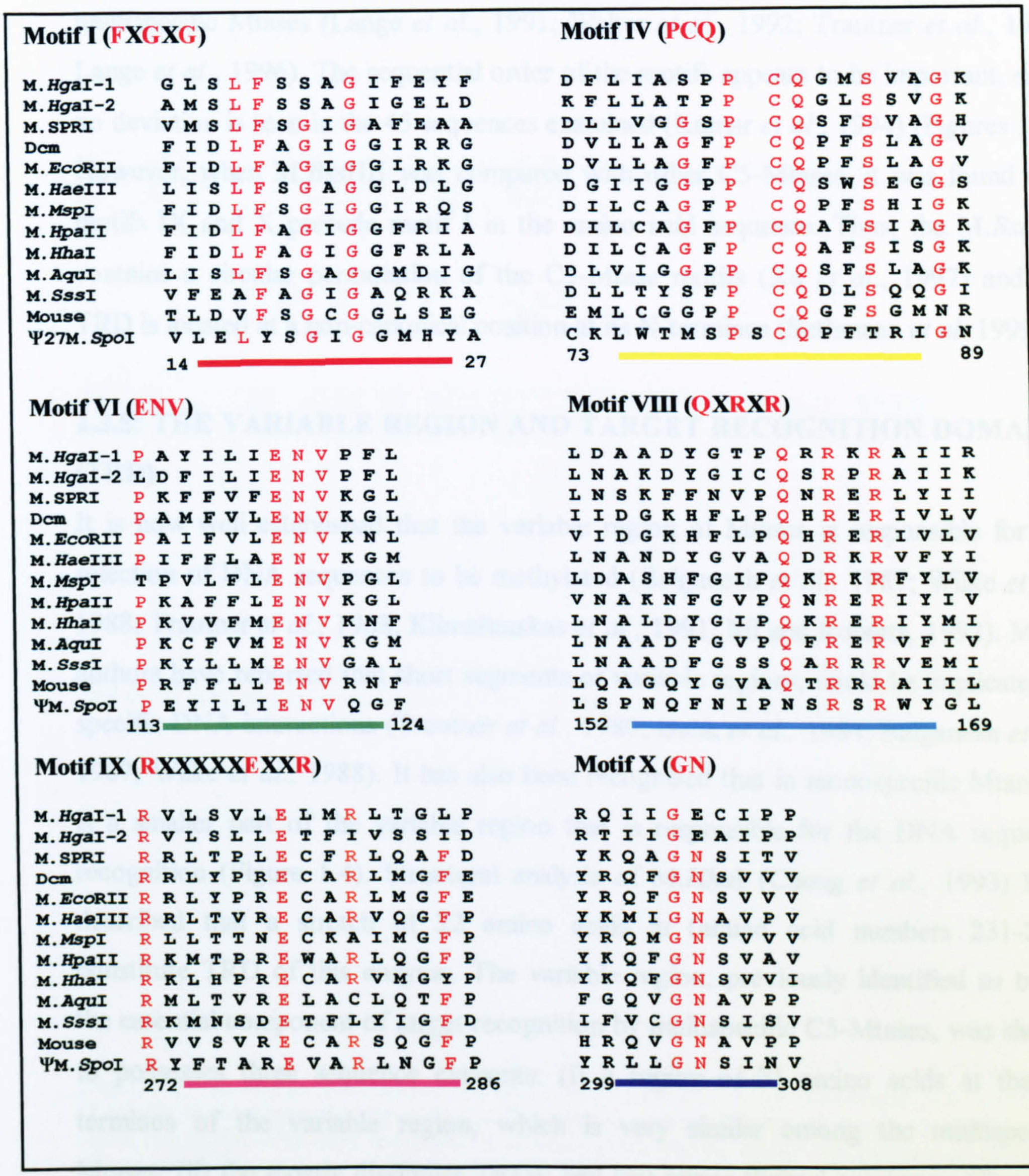


Figure 1.3: Diagram showing a comparison of the 6 most highly conserved motifs among 13 type II C5 Mtases. Among these enzymes, the seven bacterial C5 Mtases (Dcm, M.*Eco*RII, M.*Hae*III, M.*Msp*I, M.*Hpa*II, M.*Hha*I and M.*Aqu*I) conform well to the consensus. M.*Hga*I-1 and M.*Hga*I-2 are shown as examples of bacterial C5 hemi-methyltransferases. The M.*Sss*I is from the mycoplasm, *Spiroplasma*, the mouse sequence represents a eukaryotic Mtase and M.SPRI is a multispecific phage C5 Mtase. The amino acid numbering shown is from the M.*Hha*I sequence and thick coloured line represents (red-motif I, yellow-motif IV, green-motif VI, cyan-motif VIII, magenta-motif IX, dark blue-motif X). Residues, which with one or two exceptions, are invariably conserved are shown in red. Data taken from Wilson (1992) and Cheng *et al.* (1993). Sequence alignment of prokaryotic and eukaryotic Mtases are given separately in Appendices.

'variable' region lies between motifs VIII and IX and often varies greatly in size. Part of the variable region, which is responsible for target DNA recognition, is called the Target Recognition Domain (TRD) (Figure 1.4) that has been well defined in multispecific Mtases (Lange *et al.*, 1991; Walter *et al.*, 1992; Trautner *et al.*, 1996; Lange *et al.*, 1996). The sequential order of the motifs appears to be important, since no deviation is seen in the 45 sequences examined (Kumar *et al.*, 1994) (Figures 1.2). However, when M.*Bss*HII was compared with other C5-Mtases, it was found that motifs IX and X precede motif I in the amino acid sequence. Thus, the M.*Bss*HII contains a circular permutation of the C5-Mtase motifs (Xu *et al.*, 1997) and the TRD is located at a non-canonical position at its N-terminus (Sethmann *et al.* 1999).

1.3.5: THE VARIABLE REGION AND TARGET RECOGNITION DOMAINS (TRD)

It is now well established that the variable region of Mtases is responsible for the selection of DNA sequences to be methylated (Balganesh *et al.*, 1987; Wilke *et al.*, 1988; Trautner *et al.*, 1988; Klimašauskas *et al.*, 1991; Mi and Roberts, 1992). Many authors have reported that short segments of variable regions, could be implicated in specific DNA interactions (Trautner *et al.*, 1980; Buhk *et al.*, 1984; Balganesh *et al.*, 1987; Wilke *et al.*, 1988). It has also been recognised that in monospecific Mtases it is a smaller part of the variable region that is responsible for the DNA sequence recognition (Figure 1.4). Structural analysis of M.*Hha*I (Cheng *et al.*, 1993) have described that a stretch of 22 amino acids in (amino acid numbers 231-253) constitute TRD of this enzyme. The variable region, previously identified as being the essential component of target recognition by multispecific C5-Mtases, was shown to possesses three sequence elements: (i) a region of 71 amino acids at the N-terminus of the variable region, which is very similar among the multispecific Mtases; (ii) the mostly dissimilar TRDs; and (iii) minor 'linker' sequences (Trautner *et al.*, 1996).

It has been revealed that towards the C-terminus of the variable region of multi specific Mtases, there are ensembles of individual target recognition domains (TRDs), which are for example, composed of 39 (TRD *Eco*RII), 38 (TRD *Msp*I) and 40 (TRD *Hae*III) amino acids in M.SPRI (Trautner *et al.*, 1996) (Figures 1.4).

It has been found that a point mutation in the variable region is capable of abolishing single target specificity in a multispecific Mtase, while leaving the others intact. By

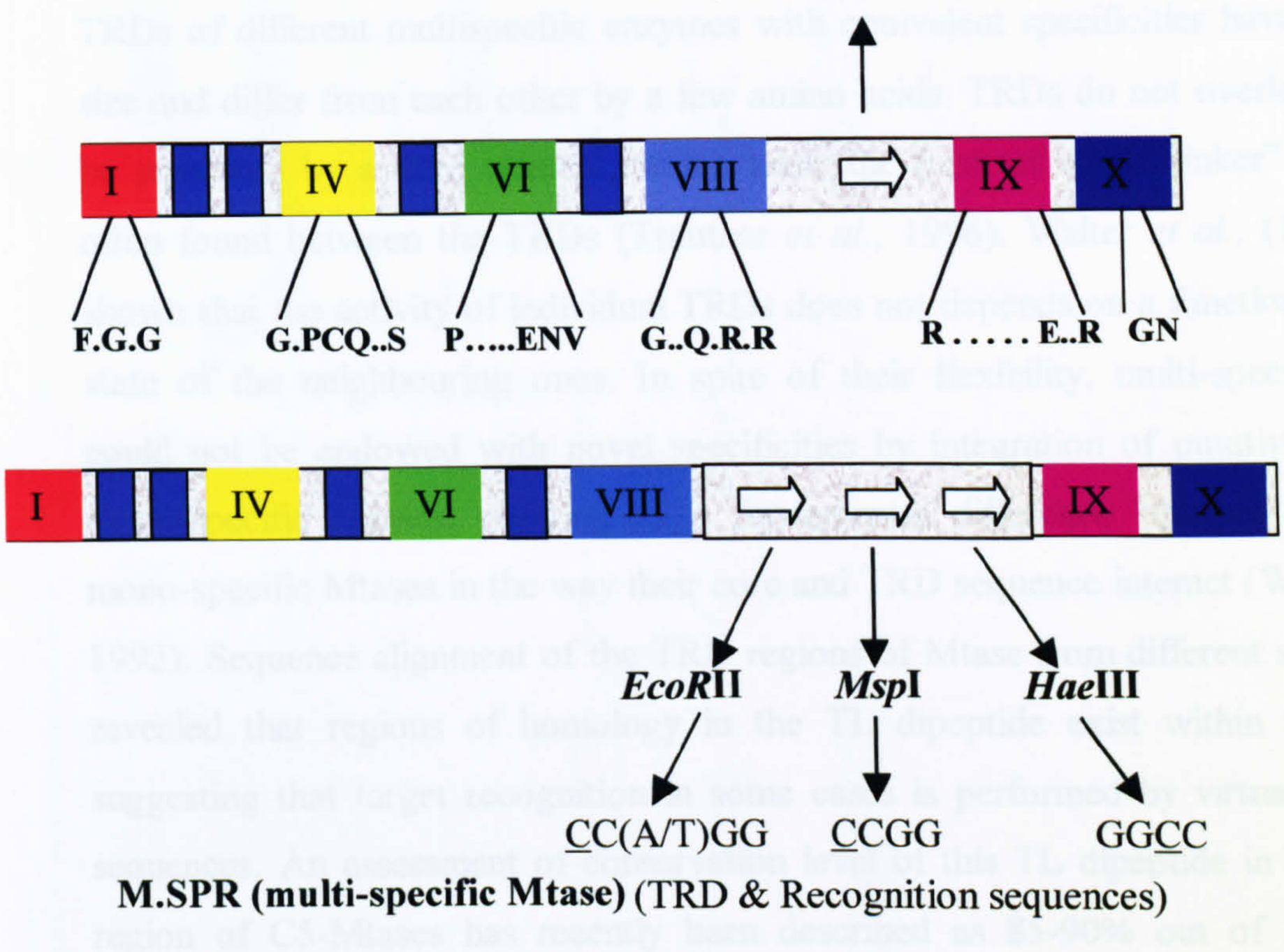


Figure 1.4: General architecture of mono- and multi-specific methyltransferases. The arrangement of motifs and conservation of amino acid sequence among C5 Mtases are illustrated. Six highly conserved motifs with their consensus sequence are shown.The motifs are coloured as (red-motif I, yellow-motif IV, green-motif VI, cyan-motif VIII, magenta-motif IX, dark blue-motif X). Target recognition sequences for Mtases SPRI and *Hha*I have been shown with the cytosine to be methylated underlined. Data taken from Adams, 1995.

mapping these mutations as well as determining the specificity of chimaeric multispecific Mtases, Trautner's group was able to define and eventually swap TRDs within the variable region (Balganesh *et al.,* 1987; Wilke *et al.,* 1988; Trautner *et al.,* 1988; Walter *et al.,* 1992; Trautner *et al.,* 1996).

TRDs of different multispecific enzymes with equivalent specificities have the same size and differ from each other by a few amino acids. TRDs do not overlap but may be separated by a third element, which takes the form of short "linker" sequences often found between the TRDs (Trautner *et al.,* 1996). Walter *et al.,* (1992) have shown that the activity of individual TRDs does not depends on a functionally active state of the neighbouring ones. In spite of their flexibility, multi-specific Mtases could not be endowed with novel specificities by integration of putative TRDs of mono-specific Mtases, pointing to a fundamental difference between multi- and mono-specific Mtases in the way their core and TRD sequence interact (Walter *et al.,* 1992). Sequence alignment of the TRD regions of Mtase from different sources, has revealed that regions of homology in the TL dipeptide exist within this region, suggesting that target recognition in some cases is performed by virtually identical sequences. An assessment of conservation level of this TL dipeptide in the variable region of C5-Mtases has recently been described as 85-90% out of 79 bacterial sequences analysed (Vilkaitis *et al.,* 2000).

1.3.6: THE DOMAINS OF C5 Mtase AND MONOMER STRUCTURE OF M.*Hha*I

M.*Hha*I is one of the smallest C5-Mtases containing 327 amino acids (molecular mass 37 kDa) that recognises the 5'-GCGC-3' sequence in double-stranded DNA and methylates the first cytosine of the recognition sequence. The crystal structure of M.*Hha*I has dimensions of 40 x 50 x 60 Å which is folded into two parts, a large domain and a small domain which are connected by a hinge region (Cheng *et al.,* 1993; Kumar *et al.,* 1994) (Figure 1.5).

The large domain contains a core beta sheet with six strands (five shown in ribbons, Figure 1.6). Four of the strands are parallel to one another, and the remaining two strands form a hairpin beta sheet turn at one end of the four parallel strands. The core beta sheet complex is sandwiched between two alpha helices and two alpha helices plus one beta strand. There is also an alpha helix that lies in front of the beta sheet

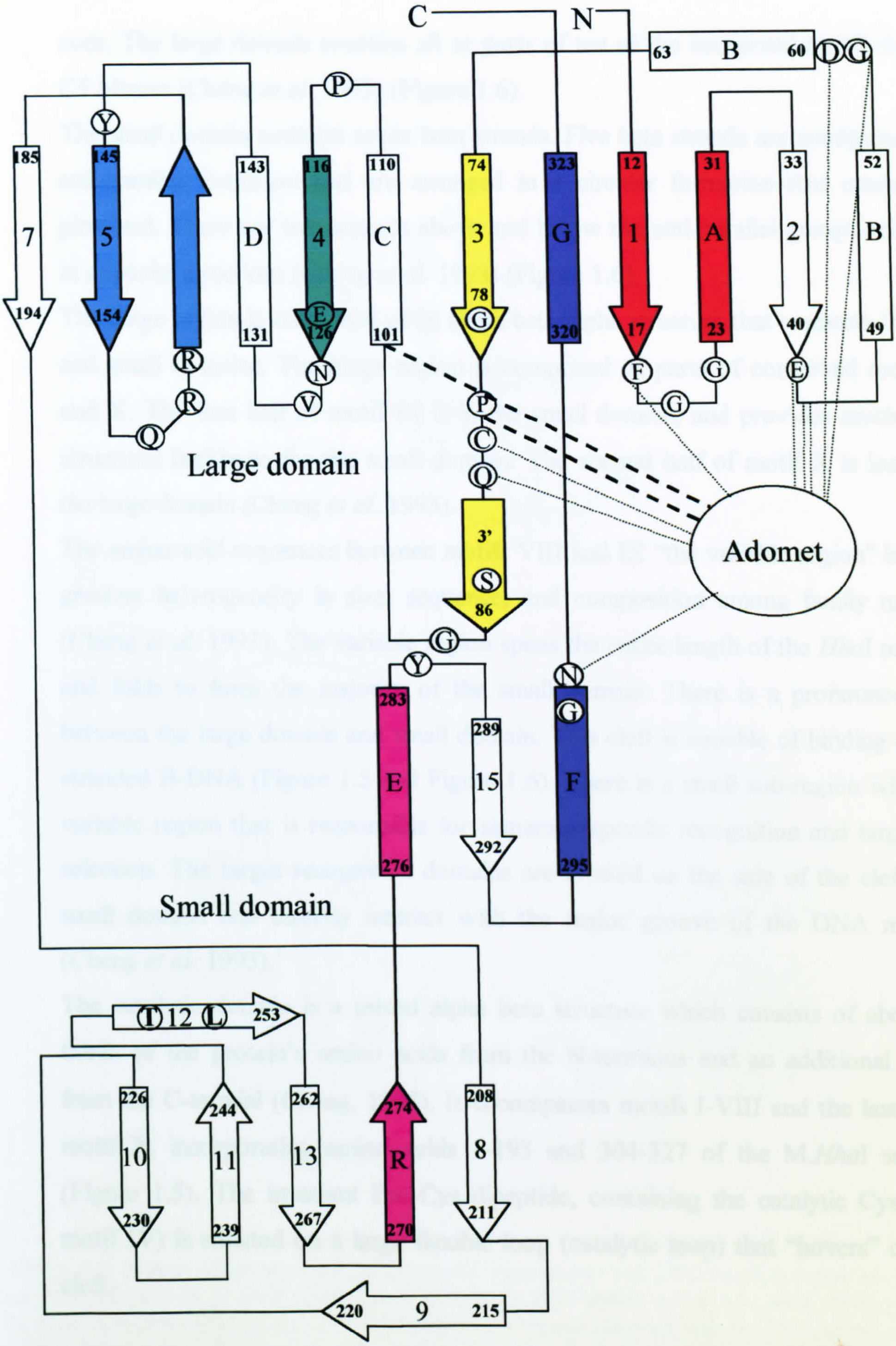


Figure 1.5: Schematic drawing of secondary structural elements of M.*Hha*I in relation to the ten conserved motifs. The large (amino acids 1-93 and amino acids 304-327) and small (amino acids 194-275) domains are marked. Rectangles indicate α-helices (lettered) and arrows indicate β-strands (numbered). Conserved or functionally important amino acids are circled.Cofactor S-adenosyl-L-methionine (SAM) binding site with its interactions with the protein is also shown. The thick dashed lines indicate hydrophobic interactions, and the thin dashed lines indicate electrostatic interactions or hydrogen bonds. The colours of various motifs are shown as (red-motif I, yellow-motif IV, green-motif VI, cyan-motif VIII, magenta-motif IX, dark blue-motif X). (Figure adapted from Cheng *et al.*, 1993)

core. The large domain contains all or parts of ten of the conserved motifs found in C5 Mtases (Cheng *et al.* 1993) (Figure 1.6).

The small domain contains seven beta strands. Five beta strands are configured in an anti-parallel formation and are arranged in a circular formation that resembles a pinwheel. There are two strands above and below the anti-parallel group, orientated in opposite directions (Cheng *et al.* 1993) (Figure 1.6).

The hinge region is composed of an alpha-beta-alpha structure that connects the large and small domains. The hinge region is comprised of parts of conserved motifs IX and X. The first half of motif IX is in the small domain, and provides much of the structural backbone for the small domain. The second half of motif X is located in the large domain (Cheng *et al.* 1993).

The amino acid sequences between motifs VIII and IX "the variable region" have the greatest heterogeneity in size, sequence, and composition among family members (Cheng *et al.* 1993). The variable region spans the entire length of the *Hha*I molecule and folds to form the majority of the small domain. There is a pronounced cleft between the large domain and small domain. This cleft is capable of binding double-stranded B-DNA (Figure 1.5 and Figure 1.6). There is a small sub-region within the variable region that is responsible for sequence specific recognition and target base selection. The target recognition domains are located on the side of the cleft in the small domain and directly interact with the major groove of the DNA molecule (Cheng *et al.* 1993).

The catalytic domain is a mixed alpha beta structure which consists of about two thirds of the protein's amino acids from the N-terminus and an additional α-helix from the C-termial (Cheng, 1995). It encompasses motifs I-VIII and the last half of motif X, incorporating amino acids 1-193 and 304-327 of the M.*Hha*I sequence (Figure 1.5). The invariant Pro-Cys dipeptide, containing the catalytic Cys 81 (in motif IV) is situated on a large flexible loop (catalytic loop) that "hovers" over the cleft.

1.3.7: THE THREE DIMENSIONAL STRUCTURE OF C5 Mtases

The crystal structure of binary M.*Hha*I-SAM (Cheng *et al.*, 1993), ternary M.*Hha*I-SAH-DNA (Klimašauskas *et al.*, 1994) and M.*Hae*III-DNA (Reinisch *et al.*, 1995) complexes has revealed the function of both the conserved motifs and the variable region (reviewed by Cheng, 1995). Since all C5 Mtases have a common structural

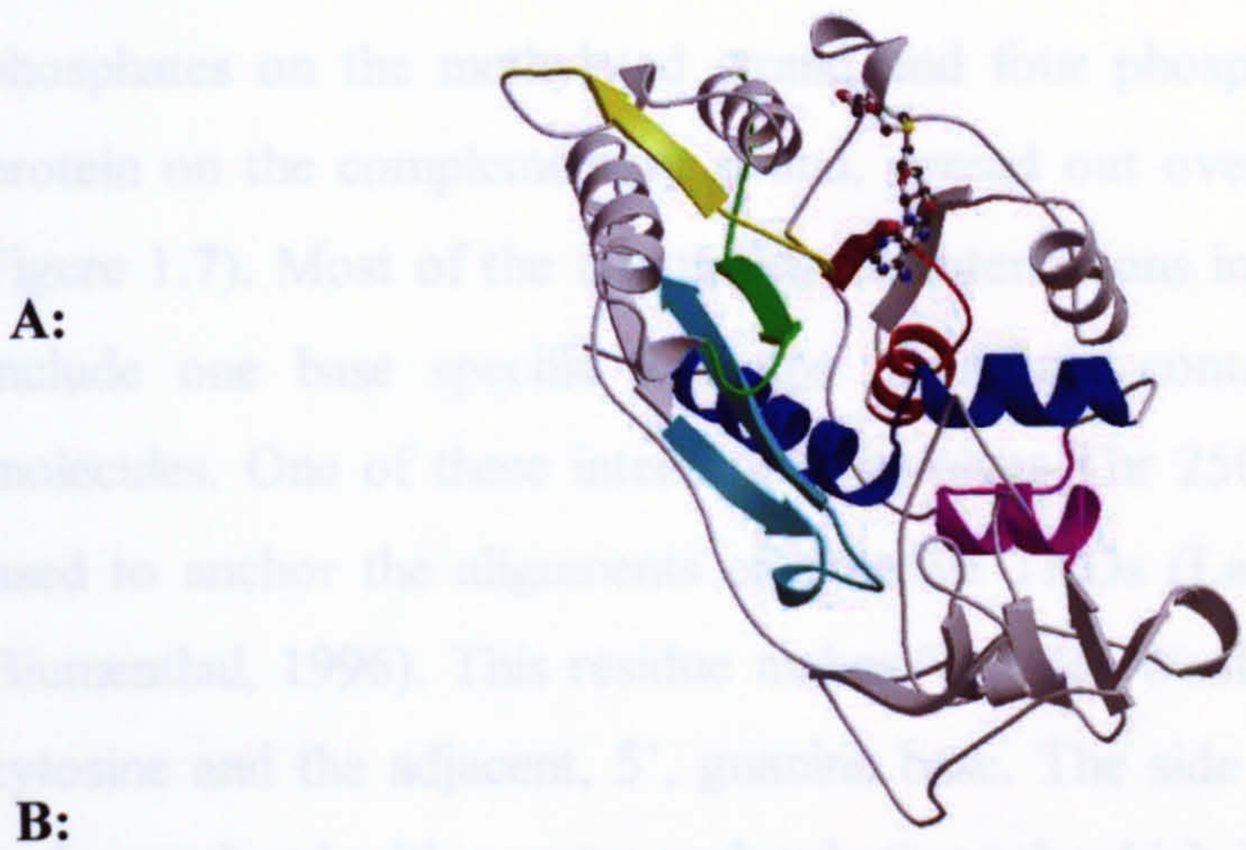


Figure 1.6: Ribbon diagram showing the domain organisation and DNA binding region of M.*Hha*I. (**A**) The structure of M.*Hha*I (PDB code, 1MHY). SAM is shown as ball and stick model. (**B**) The structure of M.*Hha*I (PDB code, 1MHT) illustrate the front and side view with the DNA as ball and stick model, showing the DNA binding cleft which is large enough to accommodate the DNA (Cheng *et al.*, 1993). The motifs are coloured as (red-motif I, yellow-motif IV, green-motif VI, cyan-motif VIII, magenta-motif IX, dark blue-motif X). The large and small domain are circled green. The image was created by Dr. Matthew J. Conroy. The figure was produced using Molscript (Kraulis (1991) *J. Appl. Cryst.* 24 946-950) and Raster3D, (Merritt & Murphy (1994) *Acta Cryst.* D50 869-873).

framework, the information obtained from these structures is likely to be very similar for the entire family of enzymes. The structure of M.*Hha*I-SAM has provided an opportunity to view the secondary structures of the conserved motifs as well as their interactions with each other and the rest of the protein. M.*Hha*I contacts six phosphates on the methylated strand and four phosphates are in contact with the protein on the complementary strand, spread out over ten residues (Figure 1.6 and Figure 1.7). Most of the intermolecular interactions involve direct contacts, but also include one base specific and one phosphate contact that is bridged by water molecules. One of these interactions involves Thr 250, which is a preferred residue used to anchor the alignments of putative TRDs (Lauster *et al.*, 1989; Cheng and Blumenthal, 1996). This residue makes Van der Waals contacts with the flipped out cytosine and the adjacent, 5', guanine base. The side chain hydroxyl group forms a hydrogen bond with a water molecule through which it contacts the phosphate of the same guanine base (Figure 1.7).

1.3.8: THE MECHANISM OF BASE FLIPPING AND THE METHYL TRANSFER REACTION

Proteins bind to DNA either to control transcription and replication, or to catalyse any chemical reactions on the DNA. Accessibility of proteins that need to interact with the bases and to perform chemistry on those bases was not clear until recently. Base flipping is the phenomenon whereby a base in normal B-DNA is swung completely out of the helix into an extra helical position to bring the target base close to the active site. There are two C5 Mtases whose complex with cognate DNA has been studied by X-ray crystallography which have provided a dramatic visual demonstration of base flipping. The structure of the M.*Hha*I complex revealed that the enzyme employs base flipping, a previously unknown mechanism, to access the target cytosine (Klimašauskas *et al.*, 1994). The structure of the M.*Hae*III-DNA complex showed that this enzyme also flips the target base (Reinisch *et al.*, 1995). Since then, the structures of a number of other protein-nucleic acid complexes that share this feature have been reported in many systems where enzymes need to gain access to a DNA base to catalyse chemical reaction on it (Roberts and Cheng, 1998). Two theories explain the mechanism of base flipping: "Active or Passive" (O'Gara *et al.*, 1998). "Active" base flipping suggests a process in which the base is first pushed out of the helix by appropriate amino acid(s) on the protein and then pulled into the

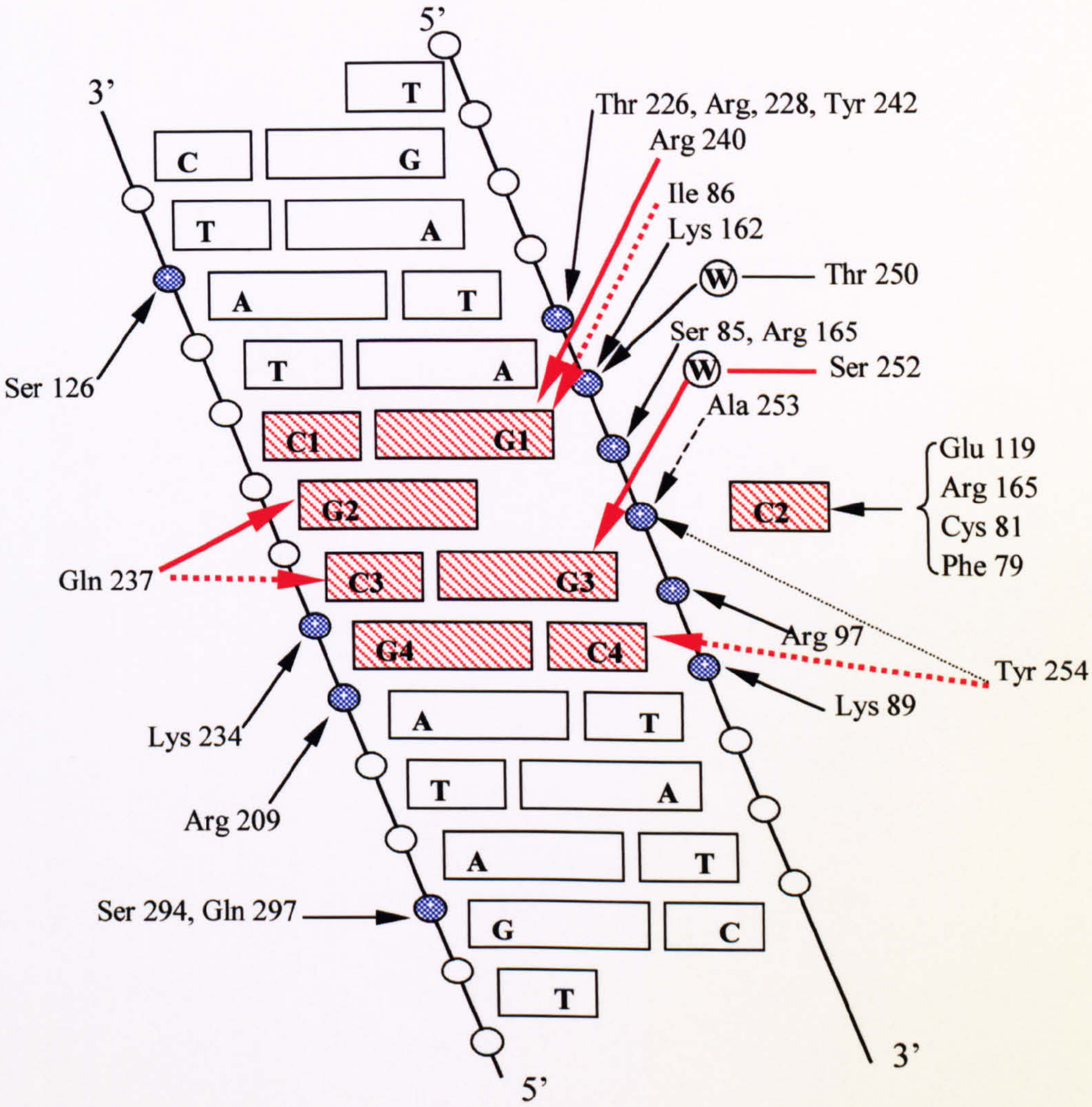


Figure 1.7: Schematic representation showing the specific base and phosphate contacts between M.*Hha*I and DNA.

The DNA is represented as a cylindrical projection. The recognition bases and contacted phosphates are coloured. Base contacts are shown with a thick red line, phosphate contacts with a thin line, and contacts through main chain atoms with a dashed line. The symbol (W) indicates the water-mediated contacts.
(Diagram adapted from Klimašauskas, *et al.,* 1994)

active site pocket of the enzyme where it remains trapped during the reaction. In the context of M.*Hha*I the "push" required could be supplied by the side chain of Gln 237, which penetrates into the DNA duplex and occupies the space left by the flipped nucleotide. The "Passive" theory suggests that during the normal "breathing" of DNA, the bases naturally spend some time in an extra-helical or spontaneously flipped-out position, or series of such positions, and it is this transient conformation in DNA that is recognised and caught by the protein (O'Gara *et al.*, 1998). The structure of M.*Hha*I complexed with substrates containing mismatches at the target base has been determined which provided an insight into the mechanism of base flipping (Figure 1.8). It has been proposed that, if the process involves the protein pushing the base out of the helix, then the push must take place not on the base, but rather on the sugar-phosphate backbone. Thus rotation of the DNA backbone is probably the key to base flipping (O'Gara *et al.*, 1998).

From studies of bacterial Mtases, the catalytic pathway to cytosine methylation is understood in some detail. The cytosine to be methylated forms a covalent link with a cysteine in conserved region IV. As this position in the cytosine ring is not readily accessible in duplex DNA, the enzyme has to catalyse a massive disruption of the DNA double helix with the cytosine being first flipped through about 180° out of the duplex and then rotated by 65°. This facilitates intimate contact with the protein; a section of which "wriggles" into the DNA duplex to fill the gap left by the cytosine in M.*HhaI* (Klimašauskas *et al.*, 1994). Covalent cytosine-cysteine bond formation activates the 5-position of the cytosine, which now accepts a methyl group from SAM which binds to the enzyme through interactions with conserved residues phenylalanine, leucine, glutamate, aspartate and tryptophan that stack against the adenine ring of SAM; the whole forms a pocket using parts of motifs I-V and X. Finally the product S-adenosyl homocysteine (SAH), is released and the covalent cytosine:cysteine bond is broken with concomitant release of the DNA from the enzyme. In this complex, un-stacking the substrate cytosine does not greatly perturb the nearby DNA structure, and a side chain from the enzyme fills most of the cavity left by the absent base. On the other hand the extrusion of the cytosine in *M.HaeIII* differs from *M.HhaI* in the mode of recognition. In the former case the bases in the recognition sequence undergo an extensive local rearrangement in their base pairing (Reinisch *et al.*, 1995). These bases are unstacked and a gap of 8 Å is formed in the DNA for the extrusion of cytosine. However, it seems likely that the features of this

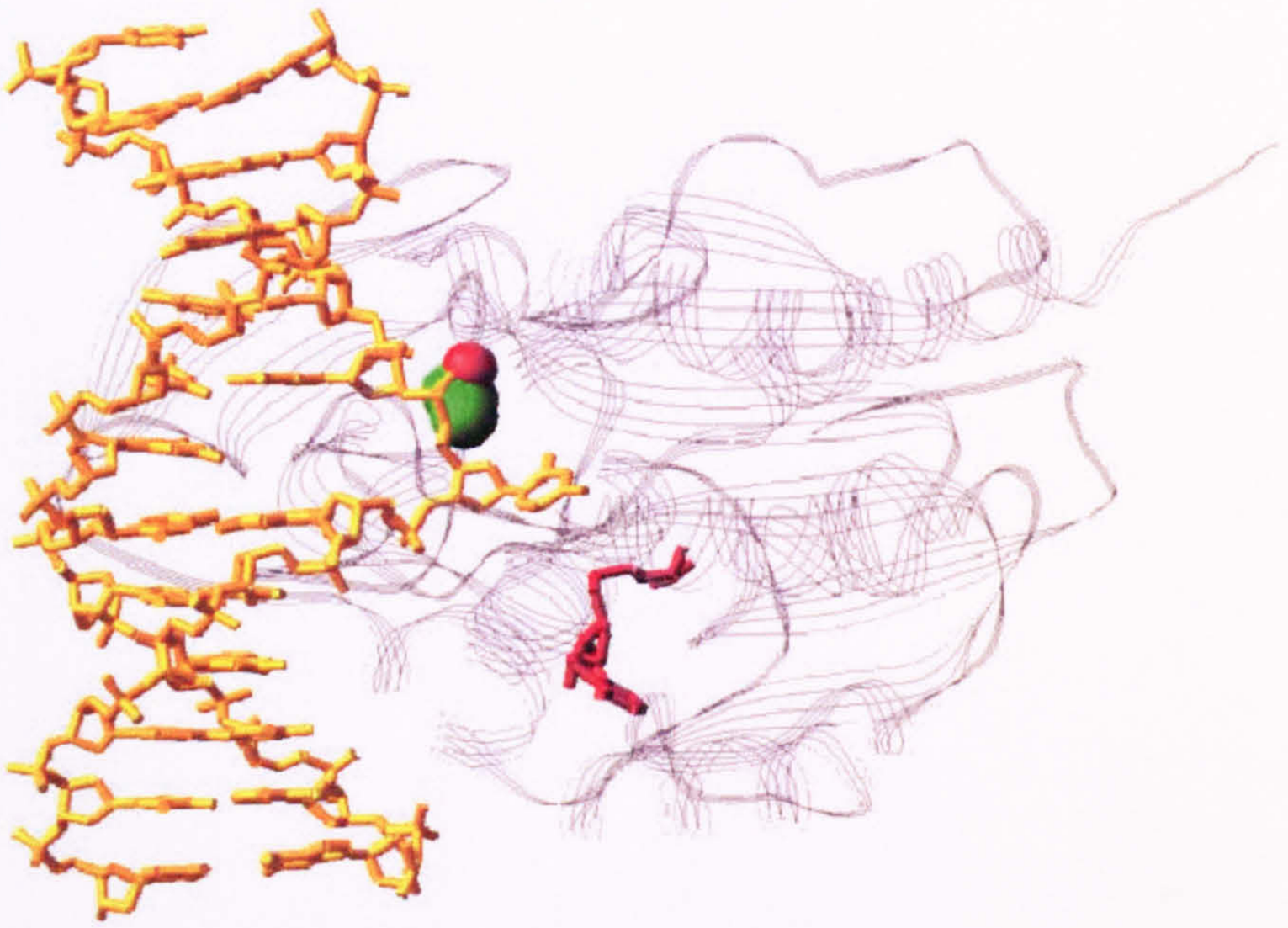


Figure 1.8: Structure of M.*Hha*I-DNA-AdoHcy complex (3MHT) illustrates the position of Thr-250 (space fill, green and red) relative to DNA (yellow sticks) and AdoHcy (red sticks) and the rest of the protein (shown as ribbons). Flipped out base is also shown. (Adapted from Vilkaitis *et al*. 2000)

class of enzymes for cytosine methylation are broadly conserved (Reinisch *et al.*, 1995).

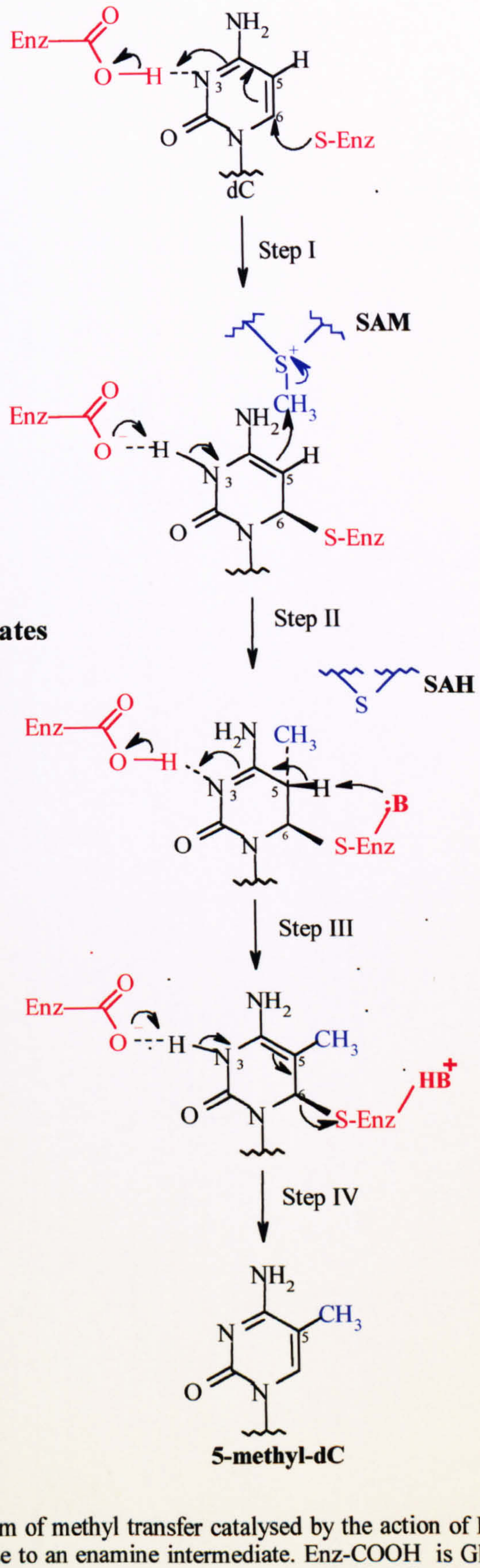
1.3.9: THE MECHANISM OF ACTION OF DNA METHYLATION

The classical mechanism proposed for enzyme mediated C5 DNA methylation involves the formation of a resonance-stabilised carbanion at C5 of cytosine as a consequence of nucleophilic attack at C6 by the protein (Santi *et al.*, 1983). A protein-DNA complex is formed by this covalent catalysis and therefore the C5-carbanion is available to attack at the methyl group of SAM. Following methyl group 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.

Successful attempts have been made to trap the intermediates formed by replacing the hydrogen atom at C5 position with a fluorine atom (Osterman *et al.*, 1988). Evidence in support of this catalytic mechanism comes from the use of C5 Mtase mechanistic based inhibitors such as 5-azadeoxycytidine (5-azaC) and 5-fluorodeoxycytidine (5-FdC). In 5-azaC, the carbon atom at the 5 position of the cytosine ring is replaced with a nitrogen atom, while in 5-FdC, the proton at C5 is replaced with a fluorine atom. When incorporated into DNA in place of the target cytosine, such substituted nucleotides cause an irreversible complex formation between the Mtase and the DNA (Santi *et al.*, 1984; Friedman, 1985; Osterman *et al.*, 1988; Chen *et al.*, 1991). Direct support for this mechanism has also been provided by the isolation and characterisation of covalent intermediates formed by M *Hae*III (Chen *et al.*, 1991), M.*Eco*RII (Friedman and Ansari, 1992), and human DNA Mtase (Smith *et al.*, 1992). Chen *et al.*, (1993) proposed a revised mechanism for enzymatic DNA methylation (Figure 1.9), and demonstrated that covalent addition of the active site Cys residue is dispensable for sequence specific DNA binding.

1.4: Aims Of The Thesis

DNA Mtase with varying sequence specificities provides an excellent model system for understanding the molecular mechanism of specific DNA recognition. Sequence comparisons of cloned genes and crystal structures of monospecific Mtases have clearly defined the functions of various conserved motifs. However, it has been revealed that although the conserved motifs are clearly similar but they are different



Enamine intermediates

Figure 1.9: Mechanism of methyl transfer catalysed by the action of DNA cytosine methyltransferase. Proton transfer to N3 give rise to an enamine intermediate. Enz-COOH is Glu 119 and Enz-S is Cys 81. Cofactor S-adenosylmethionine (SAM) and S-adenosylhomocystein (SAH) also shown as abbreviation. (Figure adapted from Bestor and Verdine, 1994).

in their target recognition domain. This could be explained with the assumption that each DNA Mtase has different recognition sequence while the similarity in catalytic unit is due to same mode of action and function of every Mtase.

Our focus of attention is variation in the TRD region and comparison between multi and monospecific Mtases. There are more subtle differences in the conserved motif IX that are essential for specificity swapping in monospecific Mtases but not in multispecific Mtases. I aim to address whether the conserved motifs are also critical for the observed incompatibility of TRDs between mono and multispecific Mtase enzyme. Based on the knowledge from a range of mutagenesis experiments we also aim to set out development of new techniques for the study of Protein-DNA interactions of Mtases in particular and other proteins in general. Development of a novel technique for the study of protein-protein interactions using the detailed information of the structure function relationship of M.SPRI is also an achievement in this regard.

CHAPTER TWO

MATERIALS AND METHODS

2.1: Materials

2.1.1: CHEMICALS, ENZYMES AND REAGENTS

Molecular biology grade or highest quality material and reagents were used for the experiments and purchased from Sigma, BDH and Fisons unless otherwise stated. Enzymes for use in general DNA manipulations were from New England Biolabs (N.E.B.), MBI Fermentas or Promega. Water was either glass distilled or purified by the MilliQ system (Millipore) and used after autoclaving at standard conditions. Oxoid and Difco Laboratories supplied growth media compounds for the culture of *Escherichia coli* (*E. Coli*). DNA Markers (1 kb ladder and 100 bp ladder) were obtained from MBI Fermentas and broad-range molecular weight marker for protein electrophoresis was from N.E.B. All other enzymes, chemicals and reagents were purchased from the sources listed in Table 2.1.

2.1.2: STANDARD BUFFERS AND SOLUTIONS

Following chemicals, buffers and reagents were prepared in the lab according to their composition given in the Table 2.2 in MilliQ water unless specified.

2.1.3: EQUIPMENT

The equipment used during the course of this study and their suppliers are listed in the Table 2.3.

2.1.4: BACTERIAL STRAINS

The *Escherichia coli* (*E. coli*) bacterial strains used as hosts in the experiments, their genotypes and suppliers are described in Table 2.4.

Table 2.1: List of chemicals and their suppliers

Reagent	Supplier
30% Acrylamide / Bisacrylamide (29:1)	BIO-RAD
Agarose	Boehringer Mannheim
Antibiotics: (Ampicillin/Chloramphenicol/Kenamycin)	Sigma
BIO-RAD protein assay kit	BIO-RAD
dNTPs	Promega
*Dpn*1	NEB
Gene-Clean kit	Bio-101 Inc.
Glutathione agarose beads	Sigma
Isopropylthio-β-D-Galactoside (IPTG)	Melford
Jetsorb™ gel extraction kit	Genomed Inc.
Lysozyme	Sigma
(3-[N-Morpholino] propane sulfonic acid) MOPS	Sigma
MobiSpin columns	Mo Bi Tec
Pfu DNA polymerase	Stratagene
Phenol pre equilibrated with Tris pH8.0	Sigma
Polaroid 667 film	Polaroid Ltd
T4 DNA ligase	MBI Fermentas
S-adenosyl-L-methionine (SAM)	NEB
Taq DNA polymerase	Promega
Taq Dyedeoxy™ Terminator Cycle sequencing kit	Applied Biosystems Inc.
Triethylammonium acetate (TEAA)	Transgenomic
Tetracycline	Sigma
TNT T7 Quick Coupled Transcription/Translation kit	Promega
Vent DNA Polymerase	NEB
Wizard® plus miniprep DNA purification system	Promega
X-Gal (5-Bromo-4-Chloro-3-indolyl-β-D-Galactoside)	Melford

Table 2.2: Composition of buffers/reagents used in this study

Buffer	Composition
Ammonium Per Sulphate (APS)	10% in MilliQ Water
Ampicillin	Stock 20 mg/ml in water, working concentration 200 µg/ml
Binding buffer 10X	500 mM Tris-HCl (pH 7.4), 500 mM NaCl, 100 mM EDTA (pH 8.0), 50 mM 2-mercaptoethanol
Chloramphenicol	Stock 34 mg/ml in Ethanol, working concentration 50 µg/ml
Coomassie Brilliant Blue Stain	230 ml methanol (46%), 230 ml water (46%), 40 ml acetic acid (8%) and 2 g Coomassie Brilliant Blue R250 for a total of 500 ml
Destainer	300 ml methanol (30%), 100 ml acetic acid (10%) and 600 ml water (60%) for one litre
Ethylene Diamine Tetra Acetic Acid (EDTA)	0.5 M (pH 8.0)
Ethidium Bromide	Stock solution 10 mg/ml in water, Working solution 0.5 µg/ml
Gel loading Buffer	0.25% (w/v) Bromophenol Blue and 50% (v/v) Glycerol
Glutathione Elution Buffer	10 mM reduced Glutathione, 50 mM Tris-HCl (pH 8.0)
IPTG	100 mM, working concentration 1 mM,
Kanamycin	Stock 10 mg/ml in water, working concentration 50 µg/ml
LA medium	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl and 1.5% (w/v) agar
LB medium	1% (w/v) Tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl.
Loading dye solution	0.2% Bromophenol blue, 0.2% Xylene cyanol, 60% Glycerol and 60 mM EDTA
NEM	10 mM (N-Ethylmaleimide)
PBS (10X)	1400 mM NaCl (8.816g), 27 mM KCl (2.0128 g), 100 mM Na_2HPO_4 (14.196 g), 18 mM KH_2PO_4 (2.4642 g) (pH 7.3)
Phenyl Methyl Sulfonyl-Fluoride (PMSF)	20 mM in Isopropanol

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) adjusted using 1M acetic acid. Sterilized by filtration through 0.22 μ membrane
RF2	10 mM RbCl, 75 mM $CaCl_2.2H_2O$, 10 mM MOPS and 15% Glycerol (w/v), final pH (6.8) adjusted using 10 M NaOH. Sterilized by filtration through 0.22 μ membrane
SDS Resolving Gel Buffer	375 mM Tris-HCl, 0.1% SDS (pH 8.8)
SDS Sample Buffer 1X	500 mM Tris-HCl (pH 6.8), 10% (v/v) Glycerol, 10% (w/v) SDS, 5% (v/v) 2-Mercaptoethanol, 0.005% (v/v) Bromophenol Blue
SDS Stacking Gel Buffer	125 mM Tris, 0.1% SDS (pH 6.8)
STE 1X	10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0), 100mM NaCl
Tetracycline	Stock 10 mg/ml in Ethanol, working concentration 50 μg/ml
Tris-Acetate-EDTA (TAE) 50X	Prepared as a 50X stock concentrate: 242g Tris base, 57.1 ml Glacial acetic acid and 100 ml 0.5 M EDTA (pH 8.0) per litre water (TAE) 1X = 40 mM Tris-acetate and 1 mM EDTA (pH 8.0)
Tris-EDTA (TE) 1X	10 mM Tris-HCl (pH 8.0) and 1 mM EDTA (pH 8.0)
X-Gal	20 mg/ml in N, N-Dimethylformamide (DMF) in glass bottle
Buffer A for Wave	0.1 M TEAA (pH 7.4)
Buffer B for Wave	0.1 M TEAA (pH 7.4), 25% Acetonitrile (CH_3CN)

Table 2.3: List of equipment used in this study

Equipment and Model	Manufacturer
Balance 0-100 g model AJ 100 0-800g model K7	Mettler
Biacore 2000	BIACORE
Biacore chips (SA and CM5)	BIACORE
Camera, Polaroid DS 34 camera	Polaroid
Centrifuges: a) J2 b) Centaur 2, Minorcentaur and Microcentaur	 Beckman MSE
DNAsep® Cartridge	Transgenomic Inc. Omaha
Gel Drier: Model 583 slab drier	BIO-RAD
Heating blocks Dri-Block DB1	Techne
Horizontal slab gel (mini- and wide mini-subcell)	BIO-RAD
Incubator Model S.I. 60	Stuart Scientific.
PC Scanner Flatbed Scanjet-4C	Hewlett-Packard
pH meter with Tris Electrode Denver Instruments	Denver Basic
Pipettes (P20, P200 and P1000 μL)	Gilson
Power supplies Model 200/20	BIO-RAD
Sonicator, Soniprep 350	M.S.E
Spectrophotometer Lambda 12	Perkin Elmer
Thermal cycler GeneE (Programmable)	Techne
UV Transilluminator, Model TM40	Genetic Research Instruments Ltd.
Vacuum Dryers Hetovac VR-I rotary evaporator	Heto-Intermed.
Vertical slab gel: Mini Protean II	BIO-RAD
Water baths	Grant Instruments, Cambridge
WAVE™ DHPLC System	Transgenomic Inc.

Table 2.4: Genotypes of *E. coli* strains used during the course of study

Strain	*Genotype*	Source
BL21 (DE3)	F⁻ *ompT hsdS$_B$ ($r_B^- m_B^-$) dcm gal λ* (DE3)	Novagen
DH5α	F⁻ */endA1 hsdR17 (r_K^- m_K^+) glnV44 thi-1 recA1 gyrA (Nal^r) relA1 Δ(lacIZYA-argF)U169 deoR (ϕ80dlacΔ(lacZ)M15)*	NEB catalogue
DH5αMCR	F⁻ *mcrA Δ(mrr-hsd RMS-mcrBC) (ϕ80dlacZΔM15) Δ(lacZYA-argF)U169 deoR recA1 endA1 phoA supE44 λ⁻ thi-1 gyrA96 relA1*	Gibco BRL, Life technologies catalogue.
GM2163	F⁻ *ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 rpsL136 dam13::Tn9 xylA5 mtl-1 thi⁻1 mcrB1 hsdR2*	NEB catalogue
INVαF	*endA1 recA1 hsdR17(rK-mK+) supE44 λ-thi-1 gyrA relA1 ϕ80lacZ Δ M15 Δ(lacZYA-argF) deoR+ F'*	Invitrogen Catalogue

2.1.5: PLASMIDS AND GENES

Cloning vectors used during the course of study for this work are listed in the Table 2.5 with their brief characteristics and their source. The plasmid DNAs were stored in MilliQ water at 20°C, or maintained in the appropriate bacterial strain and kept at –70°C.

The multi-specific Mtase gene M.SPRI was a generous gift from Professor T. A. Trautner (Berlin, Germany) and was used as the starting point for the all M.SPRI related studies. The gene encoding mono-specific Mtase gene M.*Hha*I was obtained from Dr. G. Wilson (New England Biolabs).

2.2: Methods

2.2.1: MANIPULATION OF BACTERIA

2.2.1.1: Growth of Bacterial Strains

All strains were grown using standard liquid culture Luria-Bertani (LB) medium and maintained on LA plates and, if appropriate, supplemented with antibiotic(s).

2.2.1.2: Storage of Bacterial Strains

Strains of *E. coli* were stored in glycerol (long term storage). Glycerol cultures of *E. coli* were prepared by adding 0.80 ml of a fresh culture to 0.2 ml of sterile 100% glycerol in a sterile 1.5 ml micro-centrifuge tube. The tubes were vortexed until the contents were homogenous and were then flash frozen and stored at -70°C until required. Bacteria were recovered by scraping the surface of the frozen glycerol with a sterile loop and streaked onto LB-agar plate. The tube was returned to the freezer before thawing had occurred. The plates were incubated at 37°C overnight for the growth and isolation of a single colony.

2.2.1.3: Preparation of Competent Cells

Competent cells and subsequent transformation of plasmid DNA was carried out by a method as described by Hanahan, 1985. This is a less time consuming method, and involves the preparation of frozen competent cells stored at -70° C.

Table 2.5: Plasmids acquired for use during this study

Plasmid	Characteristics	Source
pACYC184	This plasmid has a different origin of replication derived from p15A1 allowing it to be maintained in other transformants with ColE1 origin of replication for the expression of two recombinant proteins simultaneously.	Chang and Cohen, 1978. Transgenomic Ltd. formerly Kramel Biotech.
pBR322	General purpose cloning vector	New England Biolabs
pET14b, pET22b	*E. coli* inducible expression vector; containing 6 histidine and T7 promoter	Novagen
pGEX-KG	*E. coli* inducible expression vector; having GST gene	Guan and Dixon, 1991. Pharmacia Biotech
pLITMUS28, pLITMUS38	*E. coli* inducible expression vector containing both lac promoter and T7 promoter	New England Biolabs
pUC18, pUC19	General purpose cloning vectors	New England Biolabs
pUC4K	The pUC4K plasmid contains an aminoglicoside-3'phosphotransferase gene from Tn903. This gene confers resistance to Kanamycin in a wide variety of hosts and can be excised as a restriction size-mobilizing element (RSM).	Pharmacia Biotech

25ml of LB media in a 250ml conical flask was inoculated with a 250µl of overnight culture of the *E. coli* strain and growth was allowed to proceed at 37°C until the OD595 was approximately 0.4-0.6. Cells were harvested at 3,000 rpm in a M.S.E. Centaur-2 centrifuge for 10 minutes. Cell pellets were gently re-suspended in 8ml RF1 solution and incubated for 15 minutes on ice. Cells were harvested again and cell pellets re-suspended in 2ml RF2 solution followed by a 15-minute incubation on ice. Aliquots (100µl) of cells were added to micro-centrifuge tubes and flash frozen on liquid nitrogen before being stored at -70°C, where cells remained stable for approximately 6 months. The $RbCl_2$ method of making competent cells has an advantage that a large number of such cells can be prepared on one occasion, and then aliquots stored at –70°C and used for transformation for many months.
DH5αMCR strain of *E. coli* was used as host strain for usual transformations unless otherwise stated.

2.2.1.4: Transformation of *E. coli*

Frozen competent cells were thawed on ice and DNA was added (for plasmid miniprep the volume used was usually 1µl, while ligation mixture it was approximately up to 10µl). The cells were then incubated on ice for 20 minutes, heat-shocked at 42°C for 90 seconds and incubated on ice for a further 3 minutes. LB media (400µl) was added followed by a 60-minute incubation at 37°C with occasional shaking. 200 µl cells were then spread onto plates with a sterile spreader and the plates were incubated overnight.

2.2.2: PLASMID DNA ISOLATION FROM *E. COLI* & QUANTIFICATION

2.2.2.1: Small scale DNA preparation (miniprep)

Wizard™ Mini-preps system by Promega was used for small-scale preparation of plasmid DNA. This system is a modification of the alkaline lysis method. Cells harbouring the plasmid were grown overnight from a single colony in 4-10 ml LB with the appropriate antibiotic. The culture was centrifuged and the cell pellet was used to prepare the DNA. The cells were lysed and the DNA was allowed to bind with resin in a mini column. The DNA was then purified first by washing with

ethanol, and was eluted from the mini column in 50-100 µl of 1X TE buffer or MilliQ water.

The kit was purchased from Promega and was used as described in the manufacturer's instructions. Sufficient plasmid of good quality can be purified from 3-ml overnight culture and can be used directly for DNA sequencing and restriction digestion experiments without any further manipulation.

2.2.2.2: Medium scale DNA preparation (midiprep)

For large scale plasmid DNA purification, cells harbouring the plasmid were grown in 25-50 ml of LB with the appropriate antibiotic overnight, and plasmid DNA was then purified using QIAGEN® columns. DNA was separated from the lysate with a DNA-binding anion exchange column and then eluted in a high salt buffer. The DNA was precipitated with isopropanol, and washed with 70% ethanol to remove the salt, recentrifuged and dried under vacuum for 10 minutes and resuspended in 1X TE or MilliQ water.

Columns and buffers were purchased from QIAGEN® and used according to the manufacturer's instructions (except for an increase in the time of centrifugation in most steps). QIAGEN® kit produces a good yield and high quality pure DNA.

2.2.2.3: Spectrophotometric Quantification

The concentration of oligodeoxynucleotides was determined using the equation:

$$A260 / E = \text{Concentration (mM)}$$

Where:

A260 = absorbance at 260nm

E = the sum of all the mM extinction coefficients of the nucleotides present (nεdATP + nεdTTP + nεdGTP + nεdCTP)
Where n = number of bases present

The mM extinction coefficients for the nucleotides are as follows:

ε dATP	=	15.20
ε dTTP	=	8.40
ε dGTP	=	12.01
ε dCTP	=	7.05

The absorbance at 260nm was performed using a spectrophotometer in 1ml-quartz cuvettes. The path length was 1 cm in all cases. The concentrations of oligodeoxynucleotide were calculated by measuring the absorbance of a 500 fold

diluted sample at 260 nm and assuming that an absorbance reading of 1 corresponds to 33μg/ml single stranded oligodeoxynucleotide.

The concentrations of single stranded DNA were determined by reading the absorbance of samples at 260 nm. An OD_{260} of 1 corresponds to a concentration of 33 μg/ml for double stranded DNA. However, for miniprep and midiprep DNA the following method was used.

2.2.2.4: Ethidium Bromide Fluorescence Quantification

A more approximate determination of plasmid DNA concentration was obtained by running an appropriate amount of DNA in 1% agarose gel alongside a standard DNA marker (Figure 2.1 and Appendix I). After ethidium bromide staining, samples were visualized under UV and the approximate concentrations were determined by visual comparison between the unknown samples and standards.

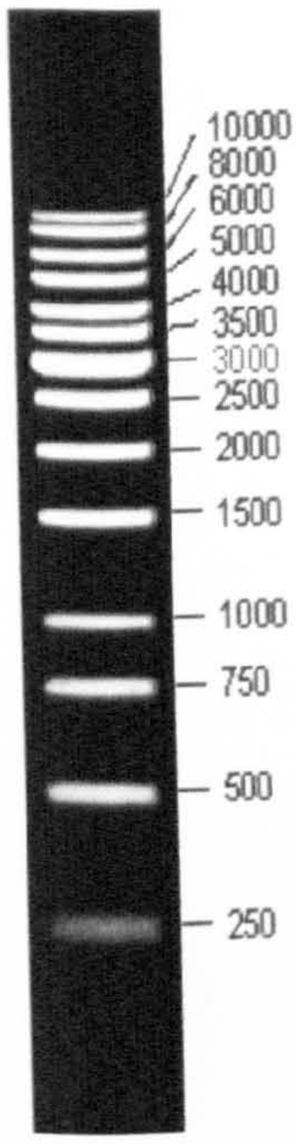
This technique was also adopted when the vector and insert fragments were recovered from the agarose gel and kept for ligation reaction in varying proportion.

2.2.3: DNA MANIPULATION TECHNIQUES

2.2.3.1: Deproteination of DNA Samples

Proteins were removed from DNA samples by phenol/chloroform extraction. Phenol and other organic solvents are used which precipitate the protein but leave the nucleic acids (DNA and RNA) in aqueous solution. Phenol pre-equilibrated with Tris pH 8.0 was mixed with chloroform and isoamyl alcohol (25:24:1 respectively) as described by Sambrook *et al.* (1989). The solution was stored in the dark at 4°C. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to the DNA solution, mixed by vortexing and then centrifuged at 13,000 rpm for 5 minutes. The precipitated protein molecules were left as a white coagulant mass at the interface between the aqueous and organic layers. The upper aqueous phase was transferred to another tube, with care taken not to disturb the aqueous/organic phase interface where the proteins are partitioned. If required, residual traces of phenol/chloroform/isoamyl alcohol were removed by the addition of an equal volume of diethyl ether, vortexing and centrifugation in a micro-centrifuge at 13,000 rpm for 1 minute. The upper ether layer was removed and discarded and traces of

Fragment	Base Pairs	DNA mass
1	10,000	20 ng
2	8,000	24 ng
4	6,000	30 ng
5	5,000	30 ng
6	4,000	40 ng
7	3,500	51 ng
8	3,000	154.5 ng
9	2,500	33.5 ng
10	2,000	35 ng
11	1,500	24.5 ng
12	1,000	10 ng
12	750	16.5 ng
13	500	14.5 ng
14	250	16.5 ng



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

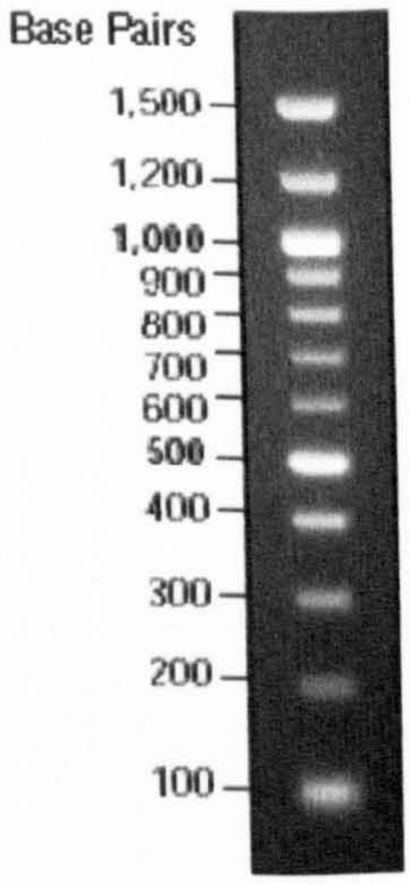


Figure 2.1: Electrophoretic molecular weight standards for DNA fragments separated on agarose.
(a) GeneRuler™ 1kb DNA Ladder (MBI) (14 fragments) with 0.5µg loading on 1.0% agarose gel and the (b) 100 bp ladder (MBI) (12 fragments) analysed on 1.8 % agarose gel. The approximate DNA mass has also been tabulated above for each fragment. (Data adapted from MBI Fermentas, 1999 Catalogue).

ether removed by gentle aspiration. The DNA was then concentrated by ethanol precipitation as described in the next section.

2.2.3.2: Ethanol Precipitation of DNA

DNA was precipitated by the addition of 1/10 volume of 3M potassium acetate solution pH 4.8 and 3 volumes of -20°C absolute ethanol. The solution was vortexed, incubated at -20°C for 20-30 minutes and then centrifuged in a bench-top micro-centrifuge at 13,000 rpm for 15 minutes. The pellet was washed in 70% (v/v) ethanol and spun again for 5 minutes. The pellet was dried under vacuum for 5 minutes and then dissolved in an appropriate volume of TE buffer or MilliQ water.

2.2.3.3: Restriction Enzyme Digestion of DNA

Plasmids, DNA fragments and oligodeoxynucleotides were specifically cleaved using the appropriate restriction Enase(s). Standard digestion of DNA (0.1-2μg) was carried out in a total volume of 20-50μl with 5-10 units of restriction Enase(s) (one unit of restriction enzyme activity being defined as the amount of enzyme required to digest 1μg of λ DNA to completion in 1 hour at 37°C). Buffer was added to a final concentration of 1X according to manufacturer's recommendations.

Digestion reactions were incubated at 37°C for 2-6 hours unless otherwise stated.

When digests requiring two enzymes with different buffer preferences were carried out, either,

1. Using activity tables supplied by the manufacturers to decide the most suitable buffer and the amount of enzymes to be used for the double digestion, or
2. Sequentially, upon completion of the first digest, the digestion mixture was diluted to a volume of 200μl with MilliQ water and extracted 3 times with phenol / chloroform / isoamyl alcohol and the digested DNA precipitated by ethanol. Once resuspended, the DNA was subjected to further restriction analysis.

2.2.3.4: Recovery of DNA through Agarose gel

DNA fragments were recovered from agarose gel slices using a Jetsorb™ gel extraction kit (Genomed Inc.). The DNA is adsorbed onto the glass beads and washed repeatedly before being eluted by incubation in 1X TE or MilliQ water. The

recommended protocol from the manufacturer was observed to carry out the procedure.

2.2.3.5: Filling in Recessed 3′-ends

The Klenow fragment is the large fragment of DNA polymerase I. It carries the 5′-3′ polymerase and the 3'-5' exonuclease activities of intact DNA polymerase I, but lacks the 5'-3' exonuclease activity of the native enzyme. The enzyme catalyse the addition of mononucleotides from deoxynucleoside-5'-triphosphates to the 3'-hydroxyl terminus of template DNA. This property is used to synthesize DNA complementary to single-stranded DNA templates. Klenow enzyme was used to convert the 3'-recessed ends of restricted DNA fragments to blunt ends by a fill-in process.

2.2.3.6: Annealing of Oligodeoxynucleotides

Complementary oligodeoxynucleotides were annealed by mixing equimolar concentrations of each oligodeoxynucleotide together and incubating at 95°C for 10 minutes and allowing to cool slowly to room temperature. Annealed oligodeoxynucleotides were then used as linker for ligation in plasmid.

2.2.3.7: DNA Ligation

DNA fragments or oligodeoxynucleotides (inserts) were ligated into plasmid vectors (containing compatible cohesive termini) using the reaction catalysed by bacteriophage T4 DNA ligase (as described in Sambrook *et al.*, 1989). Ligations were performed in 20µl volumes using an approximate molar ratio of 1:1 and 1:3 (plasmid DNA: insert DNA). Buffer, which was supplied by the manufacturer, was added to a final concentration of 1X. Finally T4 DNA ligase was added and the reaction was incubated at room temperature for 2-3 hours or at 16°C for 4-6 hours.

2.2.4: GEL ELECTROPHORESIS

2.2.4.1: Agarose Gel Electrophoresis of DNA fragments

DNA fragments were separated by agarose gel electrophoresis through 0.8-1.5% agarose gels run in 1X TAE buffer for 1 hour at 100V. Samples were run on either mini sub-cell (50ml (1X TAE) gel solution) with 8 lanes or on larger (100ml (1X TAE gel solution) wide mini sub-cells where 15-20 lanes were available. A sixth volume of loading dye solution was added to each sample prior to loading onto the

gel. Following electrophoresis, the gels were stained for 10-15 minutes in a solution of ethidium bromide (0.5 μg/ml). DNA was visualized by illuminating with UV light on a transilluminator at 302 nm. The sizes of DNA fragments were estimated by comparison to Gene Ruler™ 1 kb DNA ladder or Gene Ruler™ 100 bp DNA ladder run on the same gel. Photographs of the gels were taken using a Polaroid DS-34 direct screen instant camera with an orange filter and Polaroid 667 film. Photographs were then scanned into a PC and the image files stored on floppy disc.

2.2.4.2: Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein expression and purification were analyzed by the SDS-PAGE method described in Sambrook *et al* (1989). Protein samples were denatured by boiling for 10 minutes at 100°C with one half volume of 1X SDS loading buffer. Electrophoresis was carried out under denaturing conditions on 10% polyacrylamide gels with 4.5% stacking gels. In order to prepare a 10% gel, 5ml of acrylamide-bisacrylamide (29% - 1%), 3.75 ml 4X resolving gel buffer, 6ml dH2O, 100μl 10% ammonium persulphate and 20 μl TEMED were mixed and poured. Propane-1-ol was layered on top of the gel. Once set, propane-1-ol was removed and the stacking gel (2.25ml acrylamide-bisacrylamide 29% - 1%, 3.75 ml stacking gel buffer, 9 ml dH_2O, 120 μl 10% ammonium persulphate and 30 μl TEMED) was poured. Electrophoresis was carried out at 100V through the stacking gel and 120V through the resolving gel in 1X SDS running buffer (5 mM Tris-HCl pH 8.3, 38 mM glycine and 0.1% (w/v) SDS). Gels were stained with Coommassie Brilliant Blue R (46% methanol, 46% dH_2O, 3% acetic acid and 3g Coommassie Blue R250) for 30 minutes to 1 hour and then destained in 30% methanol, 10% acetic acid and 60% dH_2O. Broad range molecular marker was run on gels alongside the protein samples to enable their sizes to be determined (Figure 2.2 and Appendix II).

2.2.5: POLYMERASE CHAIN REACTION (PCR)

DNA segments can be hugely amplified using the polymerase chain reaction (PCR). It has an elegant simplicity. Two synthetic oligonucleotides are synthesized, each complementary to sequences on opposite strands of the target DNA at position just beyond the ends of segment to be amplified. The oligonucleotides serve as replication primers, with the 3' ends oriented towards each other and positioned to prime DNA synthesis across the desired DNA segment. The DNA segment to be

Fragment	Protein	Source	Calculated MW (Da)
1	Myosin	rabbit muscle	212,000
2	MBP-β-galactosidase	*E. coli*	158,194
3	β-galactosidase	*E. coli*	116,351
4	Phosphorylase b	rabbit muscle	97,184
5	Serum albumin	bovine	66,409
6	Glutamic dehydrogenase	bovine liver	55,561
7	MBP2	*E. coli*	42,710
8	Lactose dehydrogenase	porcine muscle	36,487
9	Triose phosphate isomerase	rabbit muscle	26,625
10	Trypsin inhibitor	soybean	(20,040-20,167)
11	Lysozyme	chicken egg white	14,313
12	Aprotinin	bovine lung	6,517
13	Insulin A, B chain	bovine pancreas	(2,340-3,400)

Figure 2.2: Protein broad range marker. It is a mixture of purified proteins with known amino acid sequences. They are resolved to 13 sharp bands when analysed by SDS-PAGE and stained with Coomassie Blue R-250. Protein bands numbered 1-10 are visible in this gel. Two bands (BSA, MW 66.4 kDa and triphosphate isomerase, MW 26.6 kDA) are at double intensity to serve as reference point. MBP is abbreviation for maltose binding protein (Data adapted from NEB, 1999 Catalogue).

amplified is heated briefly to denature it and then cooled in the presence of large excess of the synthetic oligonucleotide primers. In the presence of four dNTPs and heat-stable DNA polymerase, the primed DNA segment is replicated selectively. The cycle of heating, cooling and replication is repeated 25-30 times in an automated thermal cycler, amplifying the DNA segment flanked by the primers until it can be readily analysed and/or cloned.

2.2.5.1: Standard polymerase chain reaction (PCR)

Taq DNA polymerase was used to catalyse the reaction with the buffer (10 X PCR buffer) provided by the supplier. Ultra-pure dNTPs were stored as a mixture of all four at a concentration of 5 mM at -20°C and were used at a final concentration of 0.2 mM. Primers were stored at -20°C as 5 μM stocks and were used at a concentration of 0.2 μM. A DNA template concentration of 10 ng / μl was generally employed. The reaction was set up as follows:

The reagents were mixed at the concentration described in a thin-walled 0.5 ml micro-centrifuge tube prior to addition of the polymerase in a final volume of 49 μl. 1 μl (2U) of the polymerase enzyme was added and the contents were mixed gently. The thermal cycler program 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 25-30 cycles. The concentration of the amplified DNA was analysed on agarose gel. If required the PCR product was cut and recovered from the gel as described earlier section 2.2.3.

2.2.5.2: PCR for Site Directed Mutagenesis (SDM)

Site-directed mutagenesis (SDM) can be used to alter the gene coding for a protein and thus to make point mutations, switch amino acids and delete or insert single or multiple amino acids (Stratagene, 1997). The method of site-directed mutagenesis used here employs about 50-100 ng of double-stranded DNA (plasmid) and equal amount (100-125 ng) of two complementary synthetic oligonucleotide primers containing the desired mutation in the middle flanked by 10-15 bases on each side. The following criteria were observed for the design of primers.

a) Both the mutagenic primers contained the desired mutations and would anneal to the same sequence on opposite strand of the plasmid.
b) Primers were between 25 and 45 bases in length, and the melting temperature (T_m) of the primers was ~10°C above the extension temperature of 68°C.
The formula used for calculating the (T_m) of primers was:

$T_m = 81.5 + 0.41$ (% GC) – 675/N – % mismatch.

Where N is the primer length in base pairs.

c) The desired mutation (deletion or insertion) was in the middle of the primers with 10~15 bases of correct sequence on both sides.
d) The primers ideally had a minimum GC content of 50% and terminated in one or more G or C bases.

The primers synthesis is explained in the following section. The other components were as described in the manufacturer's protocol. The recommended thermal cycling program for site-directed mutagenesis is as follows:

- Rapid thermal ramp to 96°C
- 96°C for 30 seconds
- Rapid thermal ramp to 55°C
- 55°C for 60 seconds
- Rapid thermal ramp to 68°C
- 68°C for 15-20 minutes (depending on the product size)

All these steps were repeated for 18 cycles in Progene Techne thermocycler with a heated lid.

The oligonucleotide primers, each complementary to opposite strands of the vector are then extended during temperature cycling by *Pfu* DNA polymerase. On incorporation of the oligonucleotide primers, a mutated plasmid containing staggered nicks is generated. After temperature cycling, the product is treated with *Dpn*I. The *Dpn*I restriction endonuclease (target sequence 5'-G^{m}ATC-3') is used to digest the parental DNA template leaving behind mutagenised DNA synthesized by PCR. Since template DNA used was isolated from most *E. coli* strains is dam methylated, it is susceptible to *Dpn*I digestion, which is specific for methylated and hemimethylated DNA. 30 µl of the PCR product (after *Dpn*I digestion) was normally run on agarose gel to check if the reaction has worked. The nicked vector DNA incorporating the desired mutation(s) is then used to transform *E. coli* DH5αMCR. Colonies arising as a result of transformation were screened for the presence of desired mutation in their

DNA by restriction digestion. Sequencing was also carried out for selected mutants especially when multiple mutants were generated.

2.2.6: DNA SYNTHESIS

The chemical synthesis of oligonucleotides with a known sequence has paved the way for many biochemical advances, and had a powerful impact on all areas of molecular biology. In this study oligodeoxynucleotides were used as primers for the amplification of DNA and for site directed mutagenesis by polymerase chain reaction (PCR). Oligodeoxynucleotides were synthesized by Paul Brown using a 381A DNA Synthesizer (Applied Biosystem) using the standard phosphoramidate method followed by deprotection in 30% ammonia overnight. Oligodeoxynucleotides were ethanol precipitated and resuspended in MilliQ water and quantified using the spectrophotometric method described in section 2.2.2.3.

2.2.7: DNA SEQUENCING

A DNA molecule's most important property is its nucleotide sequence. Automated DNA sequencing uses a variation of Sanger's sequencing method in which the dideoxynucleotides used for each reaction labelled with a differently coloured fluorescent tag. The DNA to be sequenced is used as the template strand in a DNA polymerase reaction, and a short primer is annealed to it. The 3'-hydroxyl group of the primer reacts with incoming deoxynucleoside triphosphate (dNTP) to form a new phosphodiester bond. The Sanger sequencing procedure uses dideoxynucleoside triphosphate (ddNTP) analogs to interrupt DNA synthesis. When a ddNTP is inserted in place of a dNTP, strand elongation is halted because the ddNTP lacks the necessary 3'-hydroxyl group. By addition of small amounts of each ddNTPs over dNTPs, there exist a probability of acquiring at least one ddNTP at some point during synthesis and the strand synthesis will be prematurely terminated. The resulting solution contains a mixture of different-sized labelled fragments, which are then separated by electrophoresis; the colour associated with each peak is detected using a laser beam. The DNA sequence is read by determining the sequence of colours in the peaks as they pass the detector.

Polymerase chain reaction (PCR) sequencing was carried out as described below.

Template DNA was isolated from *E. coli* and purified using a Wizard® column as described in section in 2.2.2. Cycle sequencing reactions were performed using the

Taq Dyedeoxy Terminator Cycle sequencing kit as described in the manufacturer's protocols using 0.2-0.5 μg template DNA and 3.2 pmol of oligodeoxynucleotide primer.

The recommended thermal cycling programme that was used was:

- Rapid thermal ramp to 96°C
- 96°C for 30 seconds
- Rapid thermal ramp to 50°C
- 50°C for 15 seconds
- Rapid thermal ramp to 60°C
- 60°C for 4 minutes.

All these steps were repeated for 25 cycles in a Progene Techne thermocycler with a heated lid.

The extension products were concentrated using sodium acetate and ethanol precipitation as described in the manufacturer protocols. Samples were dried under vacuum. The Biomolecular Synthesis Laboratory in the Krebs Institute at Sheffield carried out the DNA sequencing on an Applied Biosystems 373A DNA sequencer and nucleotide-sequencing data was obtained as a computer file. The sequencing facility of Lark Technologies was also utilized in some cases.

2.2.8: *IN VIVO* PROTEIN EXPRESSION

2.2.8.1: Cell Growth and Induction

Small scale inductions for initial screening of protein expression were performed by inoculating 5 ml of LB media with 200 μl of an overnight culture of which had been inoculated from a colony on agar plate. For large-scale inductions, 30 ml of media was inoculated with a colony and grown overnight. This was then used to inoculate 500 ml-1L of fresh media. The protein was expressed in *E. coli* cells, which were induced by the addition of IPTG to actively growing cells in LB at an OD_{600} of 0.6-1. Incubation was continued for a further 4-6 hours for M.*Hha*I and overnight for M.SPRI. The induction conditions varied depending on the vector and strain being used and it was found that bacterial strain of *E. coli* GM2163 was the best for expression of M.SPRI as well as for M.*Hha*I expression when the vector used was pGEX-KG (personal communication, Adel AlGhanim and Dr. Mun Sen Kan).

2.2.8.2: Cell Harvesting and Lysis

Small-scale cultures (up to 5ml) were harvested by centrifugation at 13,000 rpm for 5 minutes in a micro centrifuge. Cell pellets were resuspended in 200 µl of PBS Buffer 1X containing EDTA 1mM, PMSF 0.2 µM, Lysozyme 1mg/ml and DTT 0.05 mM. Cells were kept on ice for 15-20 minutes before sonication at a full setting for 2x 5 seconds bursts. Insoluble protein was separated from soluble protein by centrifugation at 13,000 rpm for 10 minutes in a micro-centrifuge at 4°C. The insoluble pellet was resuspended in 200 µl of PBS Buffer.

Larger volume cultures (200 ml) were harvested by centrifugation at 5,000 rpm for 10 minutes at 4°C in a Sorvall RC5 or Beckman J2 centrifuge. Cell pellets were washed and resuspended in 2ml of PBS Buffer containing PMSF, EDTA, DTT and lysozyme as above. The cell pellets were kept on ice for 30 minutes followed by sonication on ice using five 10-second bursts at full setting. Soluble and insoluble fractions were separated by centrifugation at 12,000 rpm for 20 minutes at 4°C in a Sorvall RC5 or Beckman J2 centrifuge. 7-10 µl volumes of each soluble and insoluble sample were subjected for SDS-PAGE analysis.

2.2.8.3: Protein Purification by Glutathione Affinity Chromatography

Recombinant proteins were expressed as in-frame fusions with GST using the pGEX-2T vector (Pharmacia) and were purified by glutathione affinity chromatography (Smith and Johnson, 1988; Frangioni and Neel, 1993). Purification of the protein from 500 ml induced cultures was carried out as follows: The sonicated cell extract containing the soluble fusion protein in 10 ml Buffer X /1% TritonX-100 was gently shaken at 4°C for 30 minutes with 1 ml of glutathione agarose beads (pre-washed with three 10 ml volumes of Buffer X) in a 15 ml Sterilin tube. The Sterilin tubes were briefly spun (10-20 seconds) in a M.S.E. micro-centrifuge to collect the agarose beads. The supernatant was removed and the beads were resuspended in 1 ml of Buffer X and transferred to a 50 ml Falcon tube. The beads were washed three times with 50 ml volumes of Buffer X, the beads being pelleted after each wash by centrifugation for 3 minutes in a M.S.E. centaur2 centrifuge at 4°C. Following a further 50 ml wash with 50 mM Tris-HCl pH 8.0, and recentrifugation, beads were resuspended in 2 ml of 50mM Tris-HCl pH 8.0 and applied to a 2.5 ml disposable plastic column (Pierce and Warriner). Bound protein

was eluted at 4°C with 3 washes of 1 bed volume of freshly prepared 50 mM Tris-HCl pH 8.0 containing 10 mM reduced glutathione. Aliquots of from each elution (20 µl) were analyzed by SDS-PAGE (see Section 2.2.4.2) to ascertain purity. 15% (w/v) glycerol was added to each elution sample and 100 µl aliquots were stored at -70°C. Modifications to the basic purification procedure are detailed in the text.

2.2.8.4: Concentration of Purified Protein

The protein samples were concentrated either by reverse dialysis against PEG or by centrifugation in the centricon microconcentrator. For reverse dialysis method, samples were placed in dialysis tubing, laid on a tray and covered in solid PEG (MW 15,000-20,000) for 1-2 hours. For the second method, samples were centrifuged in a centricon concentrator at 3000 rpm for 30-90 min.

2.2.8.5: Protein Quantification

Protein concentration was determined using the method of Bradford (1976) and the Bio-Rad protein assay kit. A standard curve was prepared using bovine serum albumin (BSA) supplied by the manufacturer and the concentration of the unknown protein sample was determined from this. All protein samples were assayed in 1 ml disposable plastic cuvettes using a spectrophotometer at 595 nm.

2.2.9: *IN VITRO* PROTEIN EXPRESSION

Coupled transcription-translation in the prokaryotic cell free system manufactured by Promega was used to express proteins from the wild type and mutant constructs. The reaction components in a 0.5 ml micro centrifuge tube were as follows.

T7 S30 extract for Circular DNA	15 µl
S30 Premix without amino acids	20 µl
Amino Acid mixture	5 µl
Rnasin ® Ribonuclease Inhibitor (40 units/ µl)	1 µl
DNA Template (≈1 µg)	x µl
Nuclease-Free Water to a final volume 50 µl	y µl
Total =	50 µl

The lysate was gently mixed by pipetting and then centrifuged briefly where necessary to return the reaction contents to the bottom of the tubes. The micro-centrifuges tubes were then incubated for 90-120 minutes at 37°C. A control experiment was also carried out using luciferase control DNA. The results of

translation were analyzed using an *in vitro* DNA Methyltransferase activity assay and by using Biacore as described (see section 2.2.11).

2.2.10: DETERMINATION OF 5-METHYLCYTOSINE IN DNA SAMPLES

Methods already in practice for the determination of 5-methylcytosine (m5C) are Maxam Gilbert sequencing and sodium bisulfite treatment methods.

2.2.10.1: Maxam-Gilbert sequencing

Maxam-Gilbert procedure of nucleotide sequencing can differentiate modified m^5C and normal C and is a standard way to check methylated cytosine (Ohmori *et al.*, 1978; Maxam and Gilbert, 1980).

2.2.10.2: Sodium bisulfite treatment

The bisulfite genomic sequencing technique has gained a wider acceptance for the generation of DNA methylation maps with single base resolution. The method is based on the selective deamination of cytosine to uracil (and subsequent conversion to thymine via PCR), whereas 5-methylcytosine residues remain unchanged. Sequencing of the PCR amplified DNA can detect the sites at which the cytosine is methylated (Frommer *et al.*, 1992; Feil *et al.*, 1994; Xiong and Laird, 1997).

2.2.11: DNA METHYLTRANSFERASE ACTIVITY ASSAY

The activity of the encoded protein by the cloned wild type and mutant methyltransferases were analysed by the methods as described below.

2.2.11.1: *In vivo* Mcr restriction assay

This procedure is based on the exclusion of certain modified forms of DNA from *Escherichia coli* strains harbouring the *mcr* genes (Raleigh and Wilson, 1986). Plasmids carrying methyltransferase genes transformed into *mcr*A^+BC^+ strains of *E. coli* (like DH5α or INVα F') will result in the form of no colonies. It has been anticipated that Mcr assays would be the most sensitive, because of the continuous production of methylated DNA, which should constantly provide new targets for the *mcr* systems. However, (Mi and Roberts, 1992) suggest that testing plasmid sensitivity to restriction enzyme is preferable.

2.2.11.2: *In vivo* DNA Methyltransferase Activity Assay (Sensitivity to Restriction Endonuclease)

The evaluation of the activity of a C5 Mtase involves digestion of DNA extracted from the cells containing the Mtase, with a restriction endonuclease. For example when a plasmid carries a DNA Mtase gene, that plasmid is fully methylated at the sites on the plasmid that are recognized by the enzyme. This phenomenon facilitates the analysis of active and inactive mutant enzymes e.g. pGEX-KG-2T carrying the M.SPRI gene has 21 *Msp*I, 23 *Hae*III and 11 *Eco*RII recognition sites. All sites are protected from restriction digestion by these restriction enzymes when the plasmid is recovered from the cells. Similarly M.*Hha*I can methylate all of its recognition sites and therefore *Hha*I cannot restrict the plasmid. However, certain mutant gene carrying plasmids are no longer resistant to cleavage and show smaller fragments when subjected to restriction digestion by any of these enzymes. Therefore, making it qualitatively possible to analyse the activity check for the cloned gene. The digestion profile is then checked by agarose gel electrophoresis. However when site-specific mutagenesis was carried out at various locations in the M.*Hha*I or M.SPRI gene, the activity of encoded protein was altered and various different levels of methylation was observed during this study which were difficult to analyse on the agarose gel.

To establish various levels of methylation potential for an accurate digestion profile I have analysed the restricted fragments by a more sensitive technique by using DHPLC.

2.2.12: DNA FRAGMENT SEPARATION WITH THE DHPLC (WAVE™) SYSTEM

2.2.12.1: Introduction

Although the WAVE™ system works like high performance/pressure liquid chromatography (HPLC), there are two important factors, which make it unique, and they are:

a) Buffers used in combination with Ion pairing reagent(s)

b) The DNAsep® column

DNA molecules carry a negative charge due to their phosphate backbone. Long DNA fragments carry more net negative charges than shorter DNA fragments. The

stationary phase of the DNAsep® cartridge in the WAVE™ instrument is electrically neutral and hydrophobic, and therefore DNA fragments themselves cannot be absorbed to the column. An ion-pairing reagent, triethylammonium acetate (TEAA), is needed as a bridging molecule to help the adsorption of DNA onto the column.

The sizing of DNA fragments is based on the conversion of the hydrophobic stationary phase of the column into a dynamic ion-exchanger by adsorption of TEAA, which acts as a bridge between the fragments and the column. The positively charged ammonium ion of TEAA interacts with the negatively charged phosphate ions of DNA. The alkyl chains of the TEAA molecule interact with the hydrophobic surface of the column and therefore the longer DNA fragments will be adsorbed more strongly to the column than the shorter fragments. This interaction is reduced with the increase of the amount of acetonitrile in buffers A/B, which are used as mobile phase in the system. This means that the separation on the WAVE™ is size dependent and not sequence dependent.

Optimization of column and buffer gradient

In order to calibrate the DNAsep column, standard DNA markers such as pUC 18 Hae*III* digest and the GeneRuler™ 100 bp DNA ladder were used for separation by DHPLC. These fragments were separated according to their sizes by DHPLC. The gradient of buffers for DHPLC was optimized so that in most of the cases the largest fragment is eluted within 14 minutes and any uncut DNA (protected due to methylation) will be eluted at the end of the chromatogram with the acetonitrile washing of the column. The conditions for separation of DNA fragments on the DHPLC were optimized and used throughout (shown in Table 2.6). 2.5 µg of the each marker DNA was injected onto a preparative column (Transgenomic) and the retention time obtained for each standard DNA fragments is also tabulated (Table 2.7).

Table 2.6: The gradient of buffers used for DHPLC.

Time (min)	%Buffer A (0.1 M TEAA)	%Buffer B (0.1 M TEAA + 25% ACN)	Flow Rate ml/min	Temperature °C
0.0	65.0	35.0	0.75	50
2.0	45.0	55.0		
10.0	35.0	65.0		
12.0	35.0	65.0		
17.0	28.0	72.0		
18.0	28.0	72.0		
19.0	0.0	100.0		
20.0	0.0	100.0		
21.0	65.0	35.0		

Table 2.7: Size of DNA marker fragments and their retention time

	Size of DNA (in base pairs)	Retention time (in minutes)
1	80	4.39
2	100	4.82
3	102	4.87
4	174	6.29
5	200	6.89
6	257	8.37
7	267	8.65
8	298	9.36
9	300	9.29
10	400	11.12
11	434	11.78
12	458	12.15
13	500	12.17
14	587	13.57
15	600	13.18
16	700	13.87
17	800	14.75
18	900	15.89
19	1000	16.58

2.2.12.2: Experimental procedure using DHPLC (Wave™ system)

After overnight growth at 37°C, plasmids were purified using a Wizard Plus Minipreps kit (Promega) and then analysed by restriction digestion and separation of DNA fragments by DHPLC. The Wave™ system workstation handles complete DNA fragment analysis. The Wave™ built-in auto sampler injects the restricted DNA samples into DNAsep® column and a high-resolution micro-pellicular matrix rapidly separates DNA fragments (see above). The gradient of buffer A and B was used as shown in Table 2.6. The UV detector was set to 260 nm wavelength and all the DNA separations were carried out at 50°C with a flow rate of 0.75 ml per minute. Data obtained was analyzed using Hitachi model D-7000 Chromatography Data Station Software and Transgenomic Wave viewer software.

2.2.13: *IN VITRO* DNA METHYLTRANSFERASE ACTIVITY ASSAY

A) The activity of *in vivo* expressed and purified protein was checked by methylation of 1 μg of bacteriophage λ DNA, in the presence of excess of S-adenosyl-L-methionine (SAM). Reaction was carried out in a final volume of 20 μl with 2 μl of SAM (1 mM), 2 μl (10X) Mtase buffer (50 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 mM DTT & 5 mM β-mercaptoethanol) and the protein sample. The mixture was incubated at 37°C for one hour. The reaction was stopped by adding 2 μl of 10 mM N-ethylmaleimide (NEM).

Each reaction mixture was then subjected to digestion with 10 units of *Hae*III (in case of M.SPRI) and *Hha*I in case of M.*Hha*I protein. The volume of reaction mixture is increased up to 30 μl with appropriate restriction enzyme buffer and water. The mixture was further kept for one hour at 37°C before loading the samples onto a 1% agarose gel. A control experiment was also carried out in parallel using commercially available respective Mtase.

B) The activity of *in vitro* protein was checked first by binding the protein with glutathione agarose beads. Freshly prepared 20-25 μl beads (bed volume 10 μl) were used to bind with 10 μl *in vitro* translated mixture and kept for binding for some time at 4°C and then washed once with 1X Mtase buffer to remove unbound proteins from the beads. The beads were then mixed with S-adenosyl-L-methionine (SAM), 600 ng λ DNA and Mtase buffer for incubation at 37°C for 1 hour. The mixture was

centrifuged to recover the supernatant in a separate tube before the reaction was stopped by adding 2 µl of 10 mM NEM. Each reaction mixture was then subjected to restriction analysis and run on a 1% agarose gel.

2.2.14: EXPERIMENTAL PROCEDURE USING SURFACE PLASMON RESONANCE (SPR) (QUANTIFICATION AND BINDING ASSAYS ON BIACORE)

A BIAcore instrument (BIAcore 2000) was used with two sensor chips (both research grade) for the biochemical analysis of *in vitro*-translated products of wild type and mutant methyltransferases.

The BIAcore chips that were used were:

- SA (pre-coated with streptavidin) and
- CM5 (pre-coated with carboxy-methyl dextran)

HBS BIA buffer from BIAcore was used for running the samples and all experiments were carried out at 25°C.

2.2.14.1: Immobilization of Biotinylated Oligonucleotides

5'-biotinylated oligonucleotide solution were quantified spectrophotometrically and 100 nM solutions were prepared by diluting the oligonucleotide in the appropriate running buffer. The BIAcore, sensor chip SA was first washed 3X using 0.1 M NaOH. The biotinylated oligonucleotide was then passed over the streptavidin sensor chip surface at a flow rate of 10 µl/min, until the desired level was achieved.

2.2.14.2: Immobilization of anti-GST

The GST capture kit from BIAcore was used for high affinity capture of the GST fusion protein on the surface of a CM5 chip. Carboxyl groups on the matrix of CM5 chip were activated to form active esters by passing 30 µl 1:1 mixture of 100mM N-hydroxysuccinimide (NHS) and 400 mM N-ethyl-N'(dimethylaminopropyl) carbodiimide (EDC) at a flow rate of 5 µl/min. The active ester groups react spontaneously with amine groups on the ligand. The diluted goat anti-GST antibody 30 µg/ml in coupling solution (10 mM Sodium Acetate) (3 µl anti-GST + 77 µl Coupling solution) was passed over CM5 chip for immobilization by activated esters. After ligand coupling with matrix a solution of 1M-ethanolamine hydrochloride (pH

8.5) is added to deactivate remaining active esters. CM5 chip was now used for standard curve formation and quantification of unknown samples.

2.2.14.3: Protein DNA Interactions

Various wild type and mutant plasmids DNA were used to produce protein using Promega's T7 S30 extract System for circular DNA. This method was chosen simply because it reduces the time required to obtain *in vitro* translation results and it was easy to handle many samples at one time. *In vitro* translated products were first used for site-directed immobilization of fusion proteins through capture on immobilized anti-GST, for quantification. A positive control in this case was a standard recombinant GST fusion protein, which allowed for its quantification. Then the diluted *in vitro* products with running buffer were used for kinetic analysis of the protein-DNA interaction using a constant flow rate at 20 µl/minute. Bound protein was eluted from the DNA using a short pulse of 1 M NaCl. This regeneration procedure did not alter to any measurable extent the ability of the immobilized DNA to bind protein in subsequent cycles. All analysis on the sensogram data was performed using the BIAevaluation software.

CHAPTER THREE

REDESIGNING OF MOTIF IX OF MONOSPECIFIC Mtase M.*Hha*I

3.1: M.SPRI multi-specific Mtase

3.1.1: INTRODUCTION

A number of bacteriophages from a variety of phylogenetically different hosts encode type II Mtases that recognise targets similar to those of the restriction modification (RM) system of the host that they infect. SPR is a *Bacillus subtilus* phage, which encodes a 'multispecific' Mtase, which has the capacity to methylate several different DNA target sequences (Trautner *et al.,* 1980; Buhk *et al.,* 1984; Wilke *et al.,* 1988). One of these multiple specificities may be identical to the RM system present in the bacterial host lysogenised by the bacteriophage.

M.SPRI is a type II multispecific Mtase that catalyses the transfer of a methyl group from SAM to the C5 position of the underlined cytosine within the nucleotide sequences C̲CA/TGG (*Eco*RII), C̲CGG (*Msp*I) and GGC̲C (*Hae*III). The three recognition sequences for M.SPRI are due to the activity of TRDs E, M and H, which are present in the order shown (Trautner *et al.,* 1980; Behrens *et al.,* 1983; Tran-Betcke *et al.,* 1986; Wilke *et al.,* 1988; Trautner *et al.,* 1996). The TRDs of multispecific Mtases are structurally and functionally independent (Walter *et al.,* 1992). The gene coding for the sequence-specific DNA Mtase of the *Bacillus subtilus* phage SPR has been cloned from a genomic library of *Bacillus subtilus* phage DNA (Behrens *et al.,* 1983) and its nucleotide sequence determined (Buhk *et al.,* and Posfai *et al.,* 1984). The DNA sequence of M.SPRI is shown in Figure 3.1 and is from the GenBank™/EMBL Data bank with the accession number X01670. The M.SPRI gene encodes a 49.8 kDa protein that shows a primary structure typical of a C5 Mtase (Tran-Betcke *et al.,* 1986; Lauster *et al.,* 1989; Posfai *et al.,* 1989; Wilson and Murray, 1991; Kumar *et al.,* 1994).

The gene encoding M.SPRI was obtained from Professor T. Trautner (Berlin) and was sub-cloned into pGEX-KG by Dr. Alan Whitmarsh. In order to expedite a mutational analysis of M.SPRI, a set of unique restriction sites were introduced

```
1     ATG GGT AAA CTA CGT GTA ATG AGT CTT TTT AGT GGG ATC GGT GGA TTT
        M   G   K   L   R   V   M   S   L   F   S   G   I   G   G   F
49    GAA GCT GCA CTA AGA AAC ATT GGG GTT GGT TAT GAG CTG GTT GGT TTT
        E   A   A   L   R   N   I   G   V   G   Y   E   L   V   G   F
97    AGT GAG ATT GAT AAA TAT GCC GTC AAA TCT TTT TGT GCA ATT CAC AAC
        S   E   I   D   K   Y   A   V   K   S   F   C   A   I   H   N
145   GTT GAT GAG CAA TTA AAT TTT GGA GAT GTA AGC AAG ATT GAT AAG AAA
        V   D   E   Q   L   N   F   G   D   V   S   K   I   D   K   K
193   AAA CTA CCT GAA TTT GAT ATC TTA GTT GGA GGA TCT CCT TGT CAA AGC
        K   L   P   E   F   D   I   L   V   G   G   S   P   C   Q   S
241   TTT AGT GTA GCC GGC CAT CGA AAG GGA TTT GAA GAT ACA AGA GGG ACA
        F   S   V   A   G   H   R   K   G   F   E   D   T   R   G   T
289   TTG TTT TTT CAA TAC GTT GAG ACT CTT AAG GAA AAG CAA CCA AAG TTT
        L   F   F   Q   Y   V   E   T   L   K   E   K   Q   P   K   F
337   TTT GTT TTT GAA AAT GTT AAA GGG TTG ATC AAC CAT GAT AAA GGA AAT
        F   V   F   E   N   V   K   G   L   I   N   H   D   K   G   N
385   ACA TTA AAT GTT ATG GCT GAA GCT TTC AGT GAA GTT GGG TAC AGA ATT
        T   L   N   V   M   A   E   A   F   S   E   V   G   Y   R   I
433   GAT CTA GAG CTC CTG AAT TCA AAA TTC TTT AAT GTT CCA CAA AAT AGG
        D   L   E   L   L   N   S   K   F   F   N   V   P   Q   N   R
481   GAG CGA CTT TAC ATA ATT GGA ATT AGA GAA GAT TTA ATT AAA AAT GAA
        E   R   L   Y   I   I   G   I   R   E   D   L   I   K   N   E
529   GAA TGG TCT TTG GAT TTT AAA AGA AAG GAT ATA CTT CAA AAA GGG AAA
        E   W   S   L   D   F   K   R   K   D   I   L   Q   K   G   K
577   CAG AGA TTG GTA GAA TTA GAT ATT AAA AGC TTT AAT TTT AGA TGG ACA
        Q   R   L   V   E   L   D   I   K   S   F   N   F   R   W   T
625   GCT CAA TCG GCT GCT ACG AAG AGG CTA AAA GAT TTA TTA GAA GAA TAC
        A   Q   S   A   A   T   K   R   L   K   D   L   L   E   E   Y
673   GTT GAT GAA AAG TAC TAC TTG AAT GAA GAT AAA ACA AAC AGT TTG ATC
        V   D   E   K   Y   Y   L   N   E   D   K   T   N   S   L   I
721   AAA GAG TTG TCT ACA AGT CGA CTT AAT GAA AAT CTT ACT GTT GAG CAA
        K   E   L   S   T   S   R   L   N   E   N   L   T   V   E   Q
769   GTA GGT AAC ATT AAT CCC TCT GGT AAT GGA ATG AAT GGA AAT GTT TAT
        V   G   N   I   N   P   S   G   N   G   M   N   G   N   V   Y
817   AAT TCA TCT GGA TTA AGC CCC ACA ATT ACC ACT AAT AAA GGA GAG GGA
        N   S   S   G   L   S   P   T   I   T   T   N   K   G   E   G
865   CTG AAA ATT GCA GTT GAG TAC TCC AGA AAA AGC GGG CTT GGA CGA GAA
        L   K   I   A   V   E   Y   S   R   K   S   G   L   G   R   E
913   CTT GCT GTA TCT CAT ACG CTT TCT GCT TCT GAC TGG AGA GGA TTG AAT
        L   A   V   S   H   T   L   S   A   S   D   W   R   G   L   N
961   AGG AAC CAA AAA CAA AAT GCA GTT GTT GAG GTA AGG CCA GTA TTA ACC
        R   N   Q   K   Q   N   A   V   V   E   V   R   P   V   L   T
1009  CCA GAA AGG GGG GAG AAG CGA CAA AAT GGA AGA AGA TTT AAA GAT GAC
        P   E   R   G   E   K   R   Q   N   G   R   R   F   K   D   D
1057  GGT GAA CCA GCA TTT ACA GTA AAC ACA ATT GAC AGA CAC GGG GTA GCG
        G   E   P   A   F   T   V   N   T   I   D   R   H   G   V   A
1105  GTT GGA GAG TAT CCA AAA TAC AGA ATT AGA AGA TTA ACA CCG TTA GAG
        V   G   E   Y   P   K   Y   R   I   R   R   L   T   P   L   E
1153  TGC TTT AGG CTA CAG GCT TTT GAT GAC GAA GAT TTT GAA AAA GCT TTT
        C   F   R   L   Q   A   F   D   D   E   D   F   E   K   A   F
1201  GCT GCG GGA ATA AGT AAC TCA CAA TTA TAT AAG CAA GCC GGT AAT TCA
        A   A   G   I   S   N   S   Q   L   Y   K   Q   A   G   N   S
1249  ATT ACT GTA ACT GTG CTT GAG TCA ATA TTC AAG GAA TTA ATA CAT ACA
        I   T   V   T   V   L   E   S   I   F   K   E   L   I   H   T
1297  TAC GTT AAT AAA GAA TCT GAA TAA
        Y   V   N   K   E   S   E Stop
```

Figure 3.1: Nucleotide and amino acid sequences of M.SPRI. The amino acids present in the six highly conserved motifs of C5-Mtases are coloured as follows: Motif I (red), motif IV (brown), motif VI (green), motif VIII (light blue), motif IX (pink) and motif X (dark blue). The TRD region between motif VIII and IX is shown with underlined amino acids. Data from GenBank™/EMBL Data Bank: accession number K02124/X01670.

which would not lead to a change in the enzyme's primary structure. To achieve this, four sets of oligonucleotides were designed (Figure 3.2) for the sequential replacement of the 3' end of the gene, carrying unique restriction sites that are not present in the starting plasmid pGEXKG-SPR. These linker duplex oligos were annealed and ligated sequentially into pLitmus 28 and were subsequently transferred to pGEXKG-SPR by exchanging the original *Mun*I/*Xho*I fragment. The construct is named pQIS 6 (Figure 3.3). The presence of the newly inserted DNA was also confirmed by nucleotide sequence analysis. The protein encoded by construct pQIS 6 was active and the plasmid preparation of pQIS 6 showed resistance to cleavage by *Msp*I and *Hae*III.

Although the gene is expressed as a GST fusion protein, it has been observed that a small C-terminal moiety is partially removed by *in vivo* proteolysis (personal communication from Adel Al-Ghanim). In order to facilitate purification of the intact fusion protein an oligonucleotide duplex encoding six histidines was introduced at the 3' end of the gene before the stop codon. Two complementary oligodeoxynucleotides were designed in order to produce cohesive ends for *Nde*I (see below). This annealed synthetic duplex, NIS 7/NIS 8, carries six histidine codons along with a new restriction site, *Xba*I, for diagnostic purposes (see below). A two base pair overhang (TA) has been introduced at both ends, and after insertion of the linker at the *Nde*I site in pQIS 6 a unique *Nde*I site was introduced into the plasmid. The last two bases after the *Xba*I site in the linker has been changed from CA to TC and therefore the linker can be ligated with any *Nde*I digested overhang but will not be cleaved by *Nde*I after ligation.

Sequence of oligos for the 6X histidine codon insertion into the M.SPRI gene

Top oligo NIS 7: 5'-TATGTTAATAAAGAATCTGAA**CATCATCATCATCATCAT**TAA**TCTAGA**TC-**3'**

Bottom oligo NIS 8: 5'-TAGA**TCTAGA**TTA**ATGATGATGATGATGATG**TTCAGATTCTTTATTAACA-**3'**

NIS7/NIS8 linker duplex with *Nde*I site overhang:

*Xba*I

5'-TATGTTAATAAAGAATCTGAA**CATCATCATCATCATCAT**TAA**TCTAGA**TC-3'
3-' ACAATTATTTCTTAGACTT**GTAGTAGTAGTAGTAGTA**ATT**AGATCT**AGAT-5'
His His His His His His Stop

Linker 1

*Mun*I *Eco*RI *Xba*I *Xho*I

5′--**AATTG**ACAGACACGGGGTAGCGGTTGGAGAGTATCCAAAATACA**GAATTCTCTAGAC**--3′

3′--**C**TGTCTGTGCCCCATCGCCAACCTCTCATAGGTTTTATGT**CTTAAGAGATCTGAGCT**--5′

Oligo Linker 1 = Q1T/Q1B duplex

Linker 2

*Eco*RI *Stu*I *Xho*I

5′–**AATTC**GAAGATTAACACCGTTAGAGTGCTTTAGGCTACAGGCTTTTGATGACGAAGATTTCGAAA**AGGCCTC**–3′

3′–**G**CTTCTAATTGTGGCAATCTCACGAAATCCGATGTCCGAAAACTACTGCTTCTAAAGCTTT**TCCGGAGAGCT**–5′

Oligo Linker 2 = Q2T/Q2B duplex

Linker 3

*Stu*I *Bsp*DI/*Cla*I *Xho*I

5′–**CCT**TTGCTGCGGGAATAAGTAACTCACAGTTGTACAAGCAAGCCGGTAATTCAATTACTGTAACTGTGCTTGA**ATCGATC**–3′

3′–**GGA**AACGACGCCCTTATTCATTGAGTGTCAACATGTTCGTTCGGCCATTAAGTTAATGACATTGACACGAACT**TAGCTAGAGCT**—3’

Oligo Linker 3 = Q3T/Q3B duplex

Linker 4

*Bsp*DI/*Cla*I *Nde*I *Xho*I

5′–**CGAT**ATTCAAGGAATTAATACATA**CATATG**TTAATAAAGAATCTGAATAA**C**–3′

3′–**TA**TAAGTTCCTTAATTATGTAT**GTATAC**AATTATTTCTTAGACTTATT**GAGCT**–5′

Oligo Linker 4 = Q4T/Q4B duplex

M.SPRI motif IX and X

(239 bp)

Figure 3.2: A set of four Oligonucleotide linkers 1-4 used for cloning and introduction of new restriction sites into the gene for M.SPR. Oligonucleotides are shown with their complementary oligonucleotides as duplex. The new restriction sites are shown in bold. The sticky or blunt-ended restriction sites used for cloning are shown in red. A general map of the linker regions relative to the original M.SPRI gene is shown at the bottom. (Map size of linkers are not shown according to scale).

The wild type M.SPRI gene cloned in pGEXKG-SPR with the new restriction sites, (pQIS 6) (see above) was cleaved with *Nde*I and the resultant 6327 bp fragment was purified after gel electrophoresis. In order to inhibit the religation of vector, 5' terminal phosphates were removed using calf intestinal phosphatase (CIP). CIP treated and purified DNA was ligated with the NIS 7 / NIS 8 linker duplex. A construct pQIS 7 was finally made which encoded an active Mtase. Restriction digestion with *Xba*I confirmed the presence of an insert carrying six histidine codons (Figure 3.3). Nucleotide sequencing of the final construct pQIS 7 also confirmed that it contain the newly introduced sites and six histidine codons. Part of the chromatograms showing sequencing results in the motif IX and X of M.SPRI is shown in Figure 3.4.

3.1.2: STRUCTURAL FLEXIBILITY

Proteins show remarkably high tolerance to point mutations (Matthews, 1987). This is due to the high plasticity of protein structures in which only a limited subset of amino acids is critical for folding, stability and activity. The tertiary structure of some proteins can permit the incorporation of additional amino acids into a regular structural element without causing structural perturbations as long as the vital internal interactions are preserved (Heinz *et al.,* 1994).

Multi-specific Mtases, such as M.SPRI and M.*Bss*HII, recognise several different DNA sequences and possess two to five TRDs. Each TRD recognises a unique sequence of bases but utilises the same catalytic site for the methyl transfer reaction. Multi-specific Mtase genes can also accommodate DNA of non-Mtase origin within the coding section of the TRDs without loss of enzyme activity (Walter *et al.*, 1992). This remarkable plasticity suggests that the enzyme core sequences preceding or following the TRD form separable functional domains (Walter *et al.,* 1992). It has been shown that the domain in the SPR Mtase, determining CCGG methylation, can be deleted without affecting the capacity of the mutant enzyme to methylate other target sequences (Wilke *et al.,* 1988). Matin (2000) have modified the variable region of M.SPRI thereby producing a novel method for the detection and characterisation of nonsense mutations.

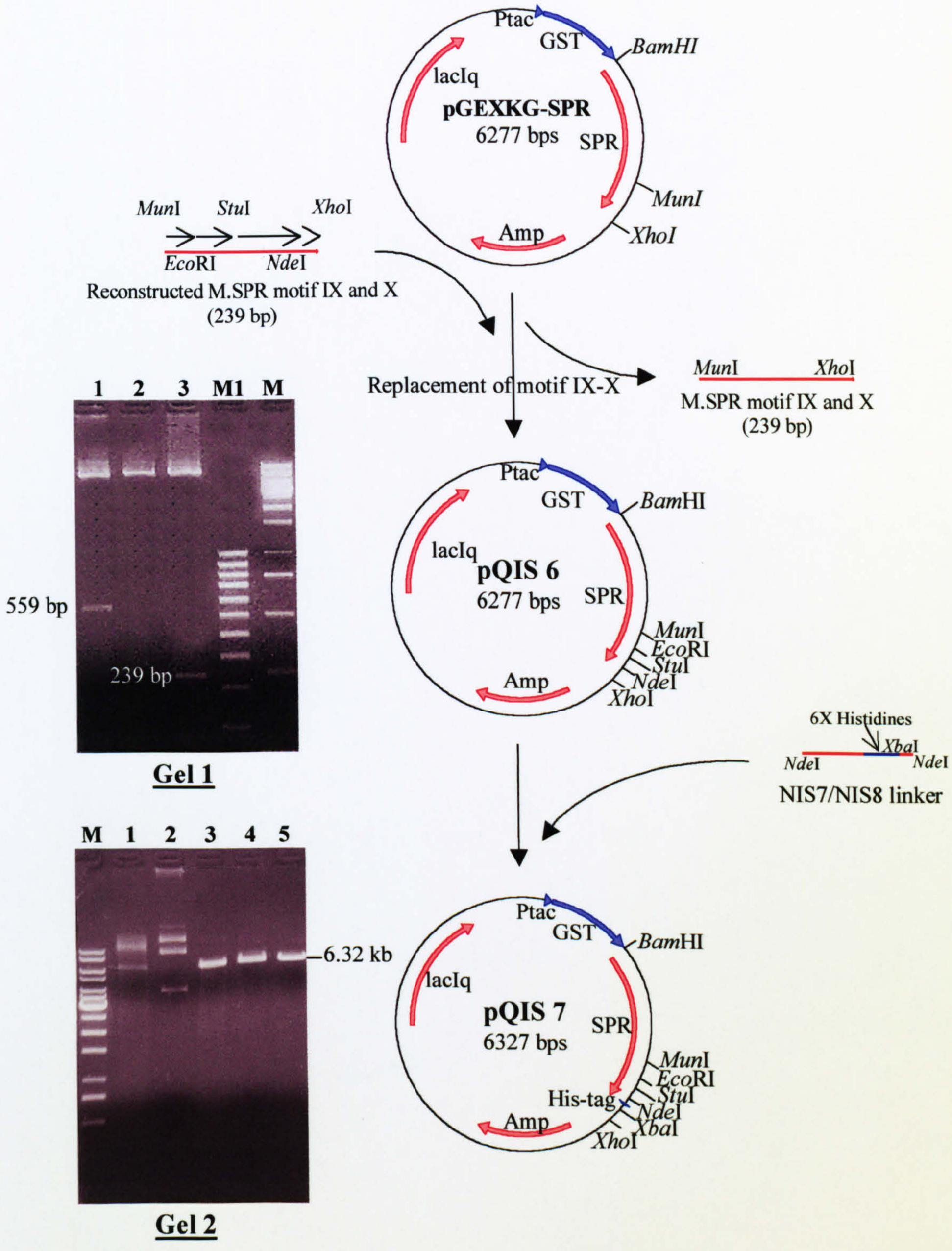


Figure 3.3: Schematic representation showing the construction of pQIS 6 & pQIS 7. The newly introduced restriction sites are shown in plasmids pQIS 6 and pQIS 7. **Gel 1** shows restriction analysis of pQIS 6, **Lane 1:** *Sal*I/*Nde*I digest (559, 5718 bp) **2:** *Stu*I cut (linear 6277 bp) **3:** *Mun*I/*Xho*I digest (239, 6038 bp) **M1:** 100 bp DNA ladder **M:** GeneRuler™ 1 kb DNA ladder. **Gel 2** shows methylation status and restriction analysis of pQIS 7. **Lane M:** GeneRuler™ 1 kb DNA ladder. **1:** pQIS 7 *Msp*I digest **2:** pQIS 7 *Hae*III digest. **3:** *Eco*RI cut (linear 6327 bp) **4:** *Mun*I cut (linear 6327 bp) **5:** *Xba*I cut (linear 6327 bp). (For markers fragment sizes, see Chapter 2).

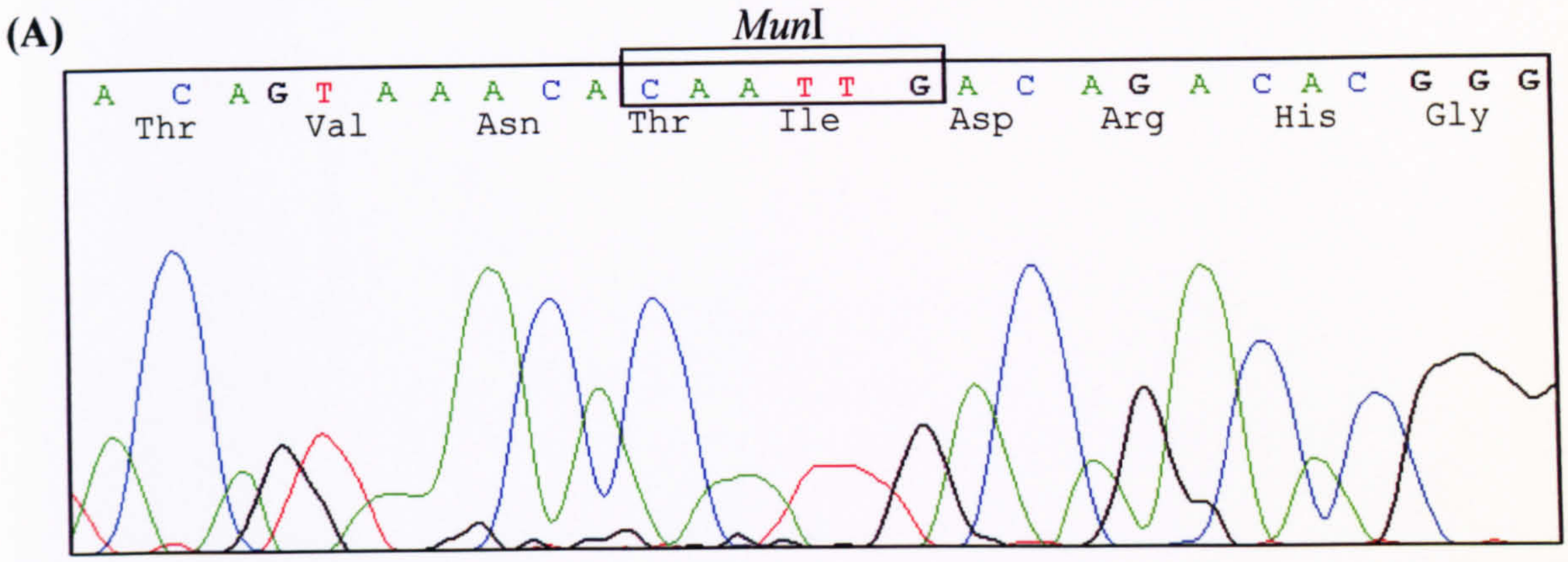
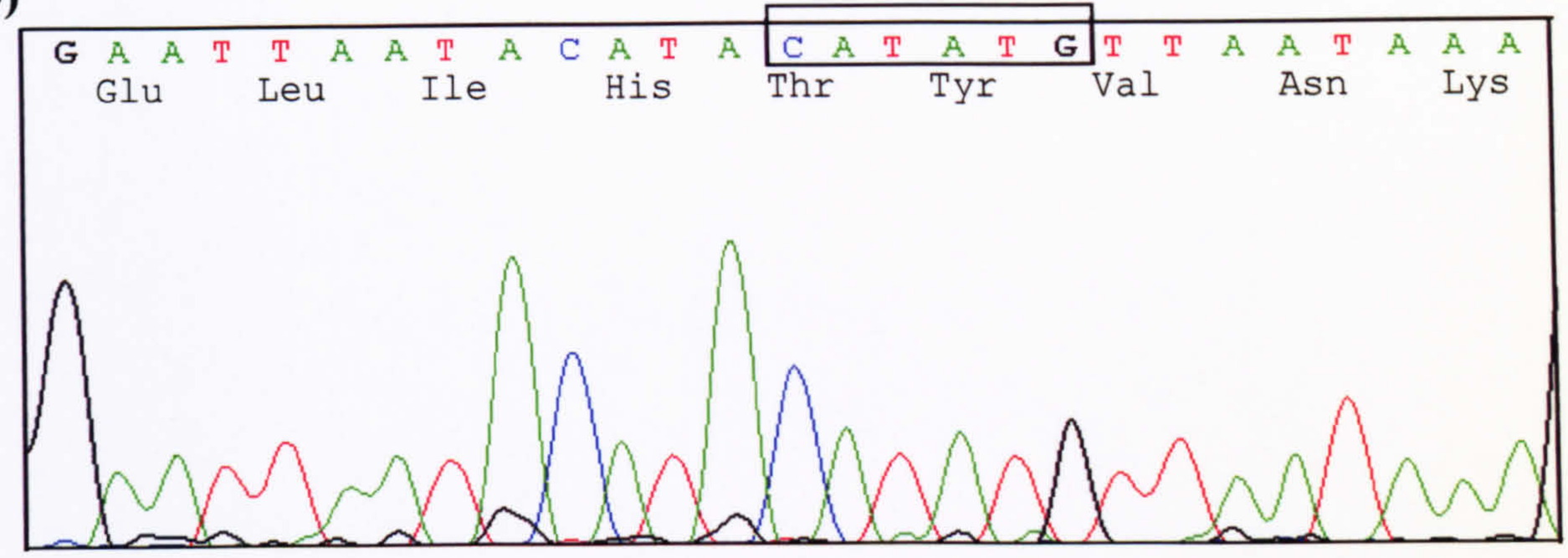
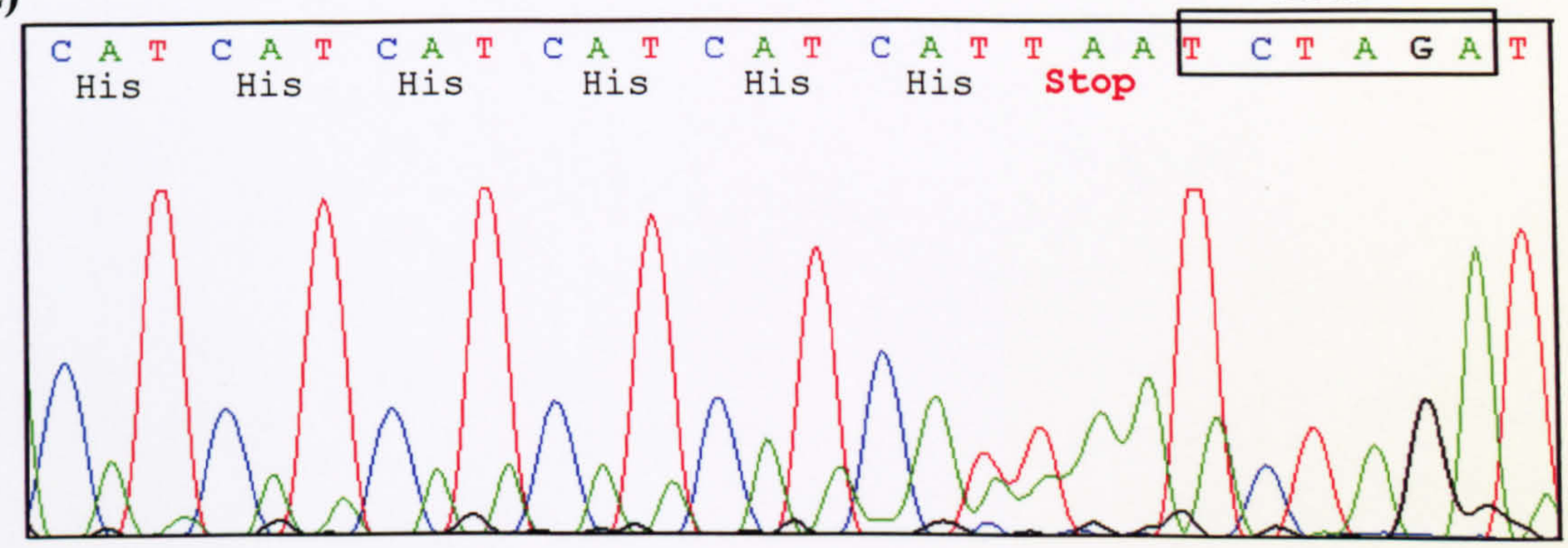


Figure 3.4: Part of the DNA sequencing analysis for wild type M.SPRI gene (pQIS 7) after introduction of new restriction sites and insertion of codons for six histidines. Data shows the presence of the *Mun*I restriction site **(panel A)**, *Eco*RI site **(panel B)**, *Stu*I site **(panel C)**, *Nde*I site **(panel D),** codons for six histidines, stop codon followed by *Xba*I site **(panel E),** and *Xho*I site **(panel F).**

(D)



(E)



(F)

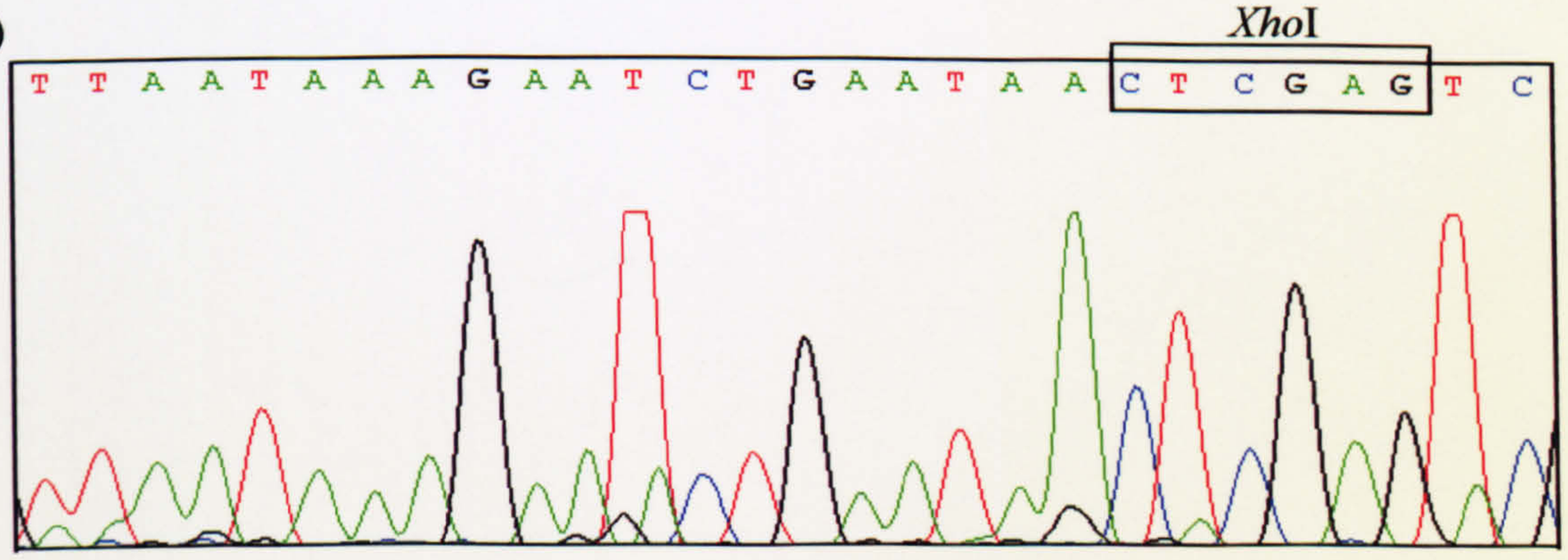


Figure 3.4: Part of the DNA sequencing analysis for wild type M.SPRI gene (pQIS 7) after introduction of new restriction sites and insertion of codons for six histidines. Data shows the presence of the *Mun*I restriction site **(panel A)**, *Eco*RI site **(panel B)**, *Stu*I site **(panel C)**, *Nde*I site **(panel D),** codons for six histidines, stop codon followed by *Xba*I site **(panel E),** and *Xho*I site **(panel F).**

3.1.3: MOTIF IX MODIFICATION IN M.SPRI GENE

The determination of the 3D structure of M.*Hha*I revealed a two-domain organisation of this typical mono-specific Mtase. This structure confirmed many of the predictions made by (Behrens *et al.,* 1987; Wilke *et al.,* 1988; Posfai *et al.,* 1989; Walter *et al.*, 1992).

Mutagenesis studies by Trautner (1988) have established that motif IX is the "core structure", i.e. the region needed for the methylation reaction. The conserved motif IX of C5-MTases has the consensus sequence **R**XXXXX**E**XX**R** (where X may be any amino acid). The sequences obtained from Genbank for several multi-specific DNA-Mtases contain similar sequences in this region (Figure 3.5) with the exception of M.*Bss*HII, which possesses unusual target recognition properties (Schumann *et al.,* 1996) and also exists in an isoform (Sethmann *et al.*, 1999). We have also calculated the percentage conservation of M.SPRI and M.*Hha*I at each amino acid level present in motif IX (Table 3.1). M.SPRI differs only at amino acid position numbers 379 and 386 compared to the other members of the same family (Figure 3.5). In general, the sequence of M.SPRI shows that there is some room for flexibility: Arg at position 379 can be accommodated instead of Lys (Figure 3.5). Similarly, Phe at position 386 in place of Trp in other multi-specific Mtases also suggests some flexibility in the structure (Figure 3.5).

In order to understand more fully the relationship between mono-specific and multi-specific Mtases the following are three possible experimental approaches that I have considered.

i) Alteration of the sequences of motif IX of M.SPRI to conform to other multi-specific Mtases. This involves mutation of Phenylalanine (F) 386 to Tryptophan (W) and Arginine (R) 379 to Lysine (K).

ii) Alterations to motif IX of M.SPRI towards the sequences of mono-specific Mtases.

iii) Finally it may be interesting to see whether random mutagenesis of the region-encoding motif IX can generate proteins with novel sequences that retain activity.

Trautner *et al.,* 1988 has carried out random mutagenesis of M.SPRI and found all the mutants in motif IX and X belongs to class I mutants in which the general methylation capacity of the Mtases is reduced or destroyed. Therefore a rational approach for mutagenesis was chosen instead of a random mutagenesis technique. It

Position of A.A of SPR	378	80		84	86		90	92	95	Genbank Accession number
1. M.SPRI	R	R	L	T	P	L	E	C	F	K02124
2. M.ρIIs	R	K	L	T	P	L	E	C	W	X05242
3. M.H21	R	K	L	T	P	L	E	C	W	M72412
4. M.ϕ3TII	R	K	L	S	P	L	E	C	W	M13488
5. M.*Bss*HII	R	R	F	T	V	R	E	C	L	X90873
Consensus	R						E		R	
Decreasing Order		K				W		A	D	
		R				F/L		S	P	

Position of A.A of SPR	378	80		84	86		90	92	95	Genbank Accession number
M.SPR	Y	P	K	Y	R	I	R	R	L	K02124
M.*Alu*I	I	V	G	P	L	E	R	R	L	Z11841
M.*Aqu*I	I	H	P	F	E	D	R	M	L	M28051
M.*Ban*I	V	Q	N	N	V	P	R	R	I	D00704
M.*Bep*I	E	L	H	L	P	Q	R	R	L	X13555
M.*Bsp*RI	D	G	E	E	N	H	R	R	L	X15758
M.*Bsu*FI	D	G	E	T	G	L	R	L	F	X51515
M.*dcm*	N	Q	Q	H	R	P	R	R	L	X13330
M.*Dde*I	I	H	P	F	Y	N	R	N	F	Y00449
M.*Dsa*V	Q	A	N	K	N	P	R	V	L	U10528
M.*Eco*RII	N	Q	A	H	R	P	R	R	L	X05050
M.*Hae*III	G	K	E	H	L	Y	R	R	L	M24625
M.*Hga*I	G	T	Y	S	D	A	R	V	L	D90363
M.*Hha*I	L	V	N	G	K	T	R	K	L	J02677
M.*Hpa*II	V	N	R	E	G	I	R	K	M	X51322
M.*Msp*I	D	G	E	T	G	I	R	L	L	X14191
M.*Mth* TI	E	S	P	K	P	Y	R	R	L	M97222
M.*Ngo*I	G	K	E	T	L	Y	R	R	M	U42459
M.*Ngo* V	Y	H	F	P	E	P	R	A	F	U43735
M.φ3TII	S	Q	P	K	A	P	R	R	F	X80202
M.*Sin*I	A	H	P	D	E	L	R	P	L	J03391

Figure 3.5: Comparison of sequence of amino acids in and around motif IX among 5 multispecific and 20 monospecific methyltransferases. All these enzymes conform well to the consensus sequence (shown in red) except M.*Bep*I (*Brevibacterium epidermidis*) and M.*Sin*I (*Salmonella infantis)*. Among these enzymes, only one is (multispecific Mtase SPR), whereas M.*Hga*I is a bacterial C5 hemimethyltransferase. The consensus sequence is shown in red. The amino acids shown at the bottom of alignment at position 379, 386, 390, and 392 of M.SPRI is the sequence in decreasing order of conservation at these positions.

was decided to mutate the SPR motif IX at three different positions. Three amino acids positioned at 386, 390 and 392 were replaced with the most frequently found amino acids present in monospecific Mtases, as shown in Figure 3.5.

When 20 mono-specific Mtases of different origins were compared with respect to motif IX, it was observed that this region is not as stringently conserved as that in the multi-specific Mtases. Additionally these Mtases show a lower degree of primary structure conservation than multi-specific enzymes. Mono-specific Mtases usually have an alanine at position 386 whereas at position 390 and 392 the most commonly found amino acids are glycine and proline (Figure 3.5, Table 3.1). Therefore, we opted to alter M.SPRI by introducing these amino acids and to examine the effects of these changes upon the activity of the mutant Mtases. Overviews of the mutations generated in this experiment are shown in figure 3.5 (a). The figures for the schematic representation of formation of mutants were collectively placed at the end of the chapter.

3.1.4: MUTAGENESIS OF WILD TYPE MOTIF IX OF M.SPRI TOWARDS M.*Hha*I

The oligonucleotides were designed for these three mutations (Phe 386 Ala, Ala 390 Gly and Asp 392 Pro) by site directed mutagenesis and were synthesised by the central facility of the department.

3.1.4.1: Phe 386 Ala mutagenesis in wild type M. SPRI

Primer/oligonucleotide NIS 9 and NIS 10 were designed to mutate Phe 386 to Ala of M.SPRI, and the duplex also carries a new *Fsp*I site for diagnostic purposes. The

NIS 9/NIS10 duplex oligonucleotide:

***Fsp*I**

5′-CCGTTAGAG**TGCGCA**AGGCTACAGGC-3′

3′-GGCAATCTC**ACGCGT**TCCGATGTCCG-5′

P L E C **A** R L Q

NIS 9: 5'—CCGTTAGAG**TGCGCA**AGGCTACAGGC—3'

NIS 10: 5'—GCCTGTAGCCT**TGCGCA**CTCTAACGG—3'

codons changed are underlined and the new *Fsp*I restriction site is shown in red.

Amino Acid Residues with numbers		Percentage Conservation within their group		Amino acids within 4.0 Å vicinity in M.*Hha*I
M.SPR	**M.*Hha*I**	**Multispecific Mtases**	Monospecific **Mtases**	
R378	**R**272	100	100	Y265, V202, V267, E278, L196
R379	**K**273	40	15	D208
L380	**L**274	80	65	T262, Y299
T381	**H**275	80	5	R277, E278, C279
P382	**P**276	80	35	E278, C279, A280
L383	**R**277	80	75	H275, Y289, C279, A280, R281
E384	**E**278	100	100	H275, P276, R272, L196, L197, A280, R281, V282
C385	**C**279	100	50	H275, P276, R277, T262, Y299, R281, V282, M283
F386	**A**280	20	75	P276, R277, E278, V282, M283, G284, Y285
R387	**R**281	100	95	R277, E278, C279, D287, L195, M283, G284
L388	**V**282	80	10	E278, C279, A280, G284, I159, Q161,
Q389	**M**283	100	20	C279, A280, R281, Q161, F302, Y285, I308,
A390	**G**284	80	70	A280, R281, V282, M1
F391	**Y**285	80	5	A280, M283, G23, S305
D392	**P**286	80	95	M1, S288, Y289
D393	**D**287	60	55	R281, Y289
E394	**S**288	80	20	P286
D395	**Y**289	80	40	P286, D287, N52, L21, R277

Table 3.1: Motif IX Residues of M.SPR & M.*Hha*I, percentage conservation, and M.*Hha*I interaction with other amino acids. The amino acids shown in red are conserved. The percentage similarity was calculated on the basis of Figure 3.5. Similar amino acids in M.SPR and M.*Hha*I are shown blue.

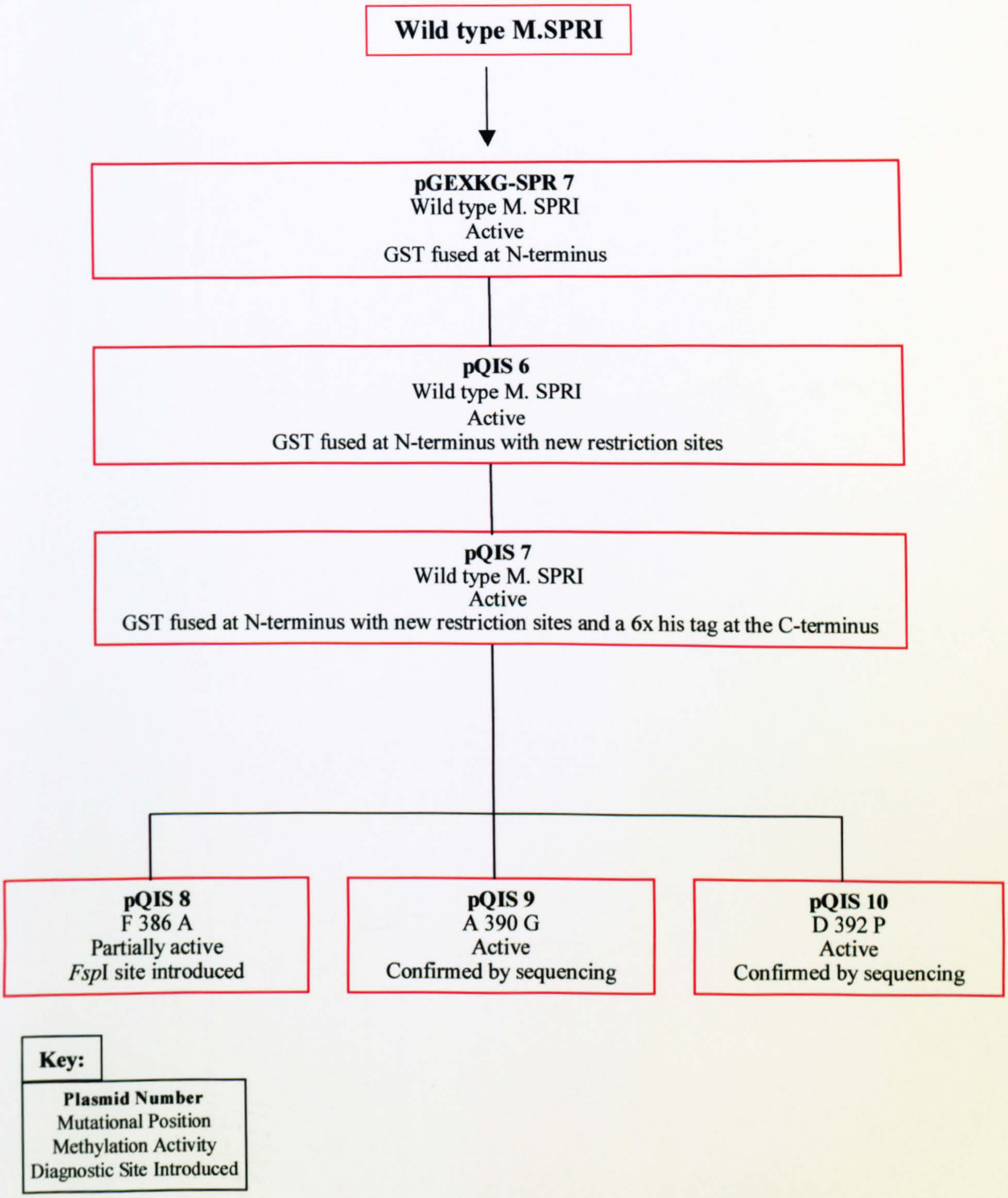


Figure 3.5 (a): An overview of the formation of mutants of M.SPRI. Wild type M.SPRI in this figure already carries GST fusion at its N-terminus and a six histidine tag at the C-terminus (page 51-52) and figure 3.3.

3.1.4.2: Ala 390 Gly mutagenesis in wild type M. SPRI

It was not possible in this case to introduce a diagnostic site for the Ala 390 Gly mutation. The sequences of the primers/oligonucleotides designed for the Ala 390 to Gly mutation were NIS 11 and NIS 12 as shown below. The mutated bases are shown in blue.

NIS 11/NIS12 duplex oligonucleotide:

```
5′-GCTTTAGGCTACAGGGTTTTGATGACGAAG-3′
3′-CGAAATCCGATGTCCCAAAACTACTGCTTC-5′
      F  R  L  Q  G  F  D  D  E
```

NIS 11: **5'—**GCTTTAGGCTACAG**GGT**TTTGATGACGAAG**—3'**

NIS 12: **5'—**CTTCGTCATCAAA**ACC**CTGTAGCCTAAAGC**—3'**

3.1.4.3: Asp 392 Pro mutagenesis in wild type M. SPRI

The primers used for the introduction of the third mutation in M.SPRI, Asp 392 Pro were NIS 13 and NIS 14. The sequences are shown below with the changes underlined and in blue.

NIS 13/NIS14 duplex oligonucleotide:

```
5′-GGCTACAGGCTTTTCCGGACGAAGATTTTG-3′
3′-CCGATGTCCGAAAAGGCCTGCTTCTAAAAC-5′
      L  Q  A  F  P  D  E  D  F
```

NIS 13: 5'—GGCTACAGGCTTTT**CCG**GACGAAGATTTTG—3'

NIS 14: 5'—CAAAATCTTCGTC**CGG**AAAAGCCTGTAGCC—3'

Site-directed mutagenesis was carried out to generate each mutation separately using the construct pQIS 7 (which encodes wild-type M.SPRI) as a template and the newly synthesized mutated plasmids are sketched in Figure 3.6.

The first mutant was identified by the presence of a new *Fsp*I site. The original wild type sequence of M.SPRI contains two *Fsp*I sites whereas the construct pQIS 8 when

digested with *Fsp*I gave three fragments of size 2738, 2380, and 1209 bps as shown in Figure 3.6. The other mutants were identified using DNA sequencing. The complete sequences of the mutant plasmids pQIS 8, pQIS 9, and pQIS10 contained only the desired mutations in motif IX (Figure 3.7). Part of the sequencing analysis data, which confirm the formation of mutants, is shown in Figure 3.8.

3.1.5: *IN VIVO* PLASMID RESTRICTION PROTECTION ASSAY

In order to determine the effect of the three mutations upon the activity of M.SPRI, a plasmid protection assay was performed. This involves assessing the sensitivity of the plasmid to restriction with *Msp*I and *Hae*III endonuclease. As described earlier (Chapter Two, section 2.2.11.2), modification of DNA sequence by a DNA Mtase results in the resistance of that sequence to cleavage by the cognate endonuclease. If the cloned Mtase gene is active, all *Msp*I, *Hae*III and *Eco*RII sites on the plasmid DNA will be methylated and therefore will be refractory to cleavage. In contrast an inactive Mtase would be signified by digested plasmid DNA following analysis by agarose gel electrophoresis.

Plasmid DNA isolated from *E. coli* cultures was subjected to restriction using an excess of enzyme *Msp*I. It was found that the pQIS 8 (which encodes for Phe 386 Ala in M.SPRI) was not fully protected (See lane 1, Gel 2 of Figure 3.6). Another batch of plasmid was made from an overnight culture (originally the plasmid was obtained from a six hour culture) that showed the DNA to be fully protected against *Msp*I and *Hae*III digestion. Further experiments on pQIS 8 revealed that if the culture was grown overnight either at room temperature or at 37°C, fully protected DNA was produced. It was therefore concluded from these experiments that Phe 386 Ala mutant of M.SPRI encoded by pQIS 8, had not lost its ability to methylate DNA completely, but needed an extended growth time for complete activity of the protein to be expressed. This suggests that the mutant is catalytically or structurally compromised. Such mutants are often referred to as partially active, owing to the reduced or incomplete plasmid methylation that is observed *in vivo* (Wilke *et al.*, 1988; Vilkaitis *et al.*, 2000; Matin, 2000). Therefore, when a restriction digestion of the plasmid encoding a Mtase gene shows this kind of pattern, it will be referred to as "partially active" throughout this study.

The mutants Ala 390 Gly and Asp 392 Pro were fully protected and not digested by *Msp*I and *Hae*III restriction digestion and therefore produced a fully active enzyme,

Construct	Mutation	Amino Acid Sequence	Activity[a]
		370 380 390	
pQIS 7	W/T M.SPRI	GEYPKYRIRRLTPLECFRLQAFDDED	+
pQIS 8	F386A	GEYPKYRIRRLTPLECARLQAFDDED	+/-
pQIS 9	A390G	GEYPKYRIRRLTPLECFRLQGFDDED	+
pQIS 10	D392P	GEYPKYRIRRLTPLECFRLQAFPDED	+

Figure 3.7: Amino acid sequence of Motif IX from wild type and mutant of M.SPRI. The amino acids shown in red are conserved motif IX residues. The mutated amino acids are shown in blue. The amino acids residues number are shown on the top.
a: Activity shown on the right panel of diagram = Full protection (+), Partial protection (+/-)

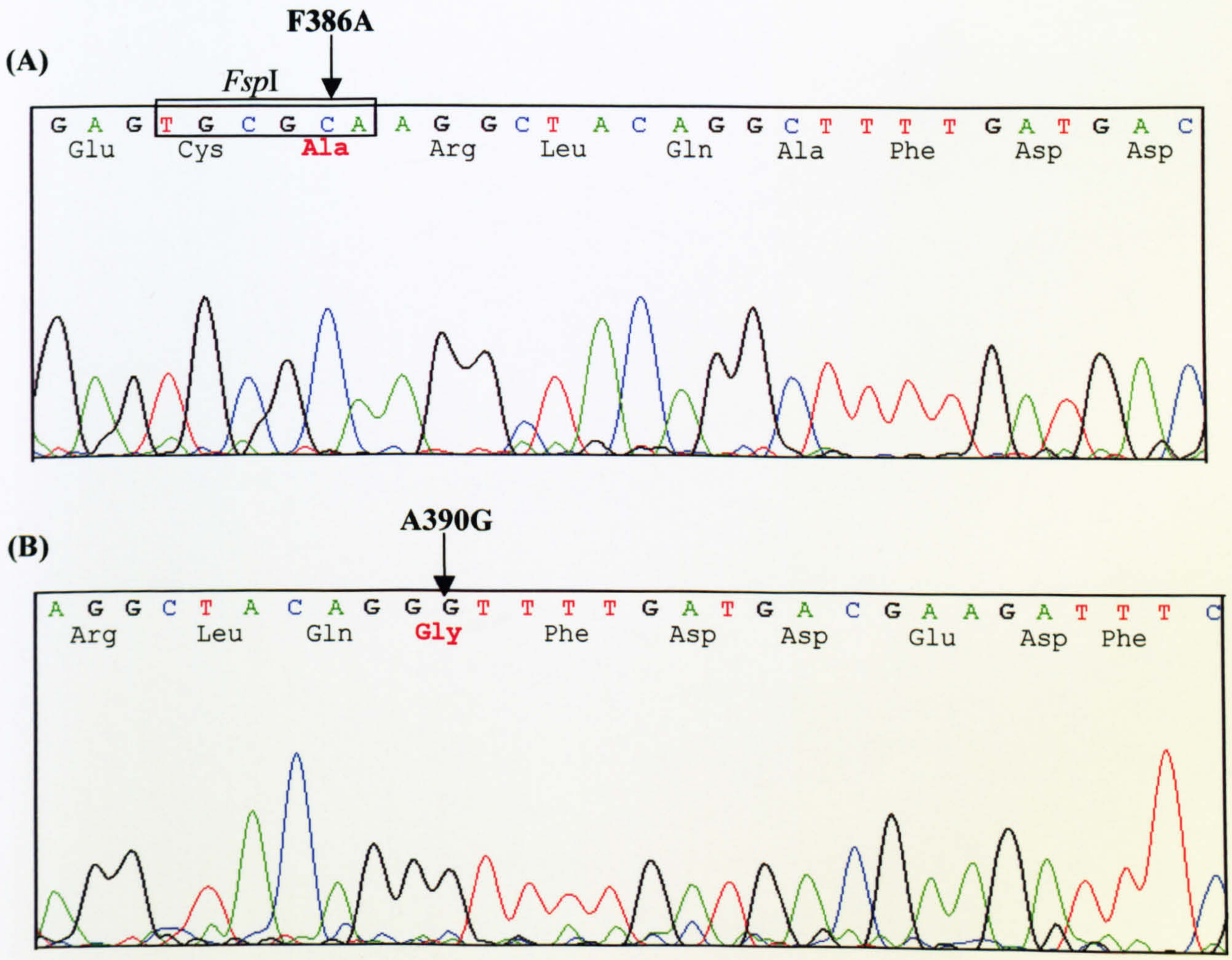


Figure 3.8: Part of the DNA sequencing analysis of plasmids pQIS 8 and pQIS 9. (A) Data confirms the formation of construct pQIS 8, which encodes for the Phe 386 Ala mutation in M.SPRI and **(B)** the formation of construct pQIS 9 which encodes for the Ala 390 Gly mutation in M.SPRI.

in spite of the changes to the amino acid sequence (See lane 2 and 3 of Gel 2 in Figure 3.6).

3.1.6: PROTEIN EXPRESSION AND PURIFICATION OF WILD TYPE AND MUTANT PHE 386 ALA M.SPRI

Plasmids pQIS 7 (wild type M.SPRI) and pQIS 8 (mutant Phe 386 Ala of M.SPRI) were used for expression of the corresponding proteins in *E. coli* GM2163, for biochemical characterisation purposes. The proteins were expressed and analysed by SDS polyacrylamide gel electrophoresis (PAGE) and were purified as shown in Figure 3.9.

The M.SPRI polypeptide is 50 kDa. However when it is fused to GST, it has a combined molecular weight of 76 kDa. SDS-PAGE analysis (Figure 3.9) of the expressed fusion protein showed that the major induced species was of this molecular weight. The fusion protein was purified to greater than 90% purity, as assessed using a Coomassie stained gel. The yield was estimated to be ~40 μg of protein after purification from a 50 ml culture (see Chapter two, section 2.2.8).

3.1.7: PLASMID PROTECTION ASSAY USING THE WILD TYPE AND MUTANT PHE 386 ALA M.SPRI PURIFIED FROM BACTERIA

3.1.7.1: Dose response

Purified enzymes were assayed for their *in vitro* DNA methylation activity. This assay is an *in vitro* mimic of biological restriction and modification, and involves pre-incubation of DNA with the Mtase enzyme in the presence of the methyl donor SAM, prior to addition of the cognate restriction enzyme (usually in excess). If the recombinant protein is active, DNA will be methylated and will therefore be resistant to digestion with *Hae*III. Bacteriophage lambda DNA was used as substrate in a parallel set of reactions. (Lambda DNA contains 328 *Msp*I and 149 *Hae*III recognition sites). In order to establish the dose response, varying amounts of recombinant wild type and Phe 386 Ala mutant M.SPRI were used in a protection assay carried out on λ DNA. The recombinant wild type and mutant Phe 386 Ala Mtases were incubated at 37°C with 1 μg bacteriophage λ DNA in Mtase buffer (50 mM Tris-HCl, pH 7.5, 25°C, 50 mM NaCl, 10 mM EDTA, 1 mM DTT) supplemented with 100 μM SAM to give a final reaction volume of 20 μl. This

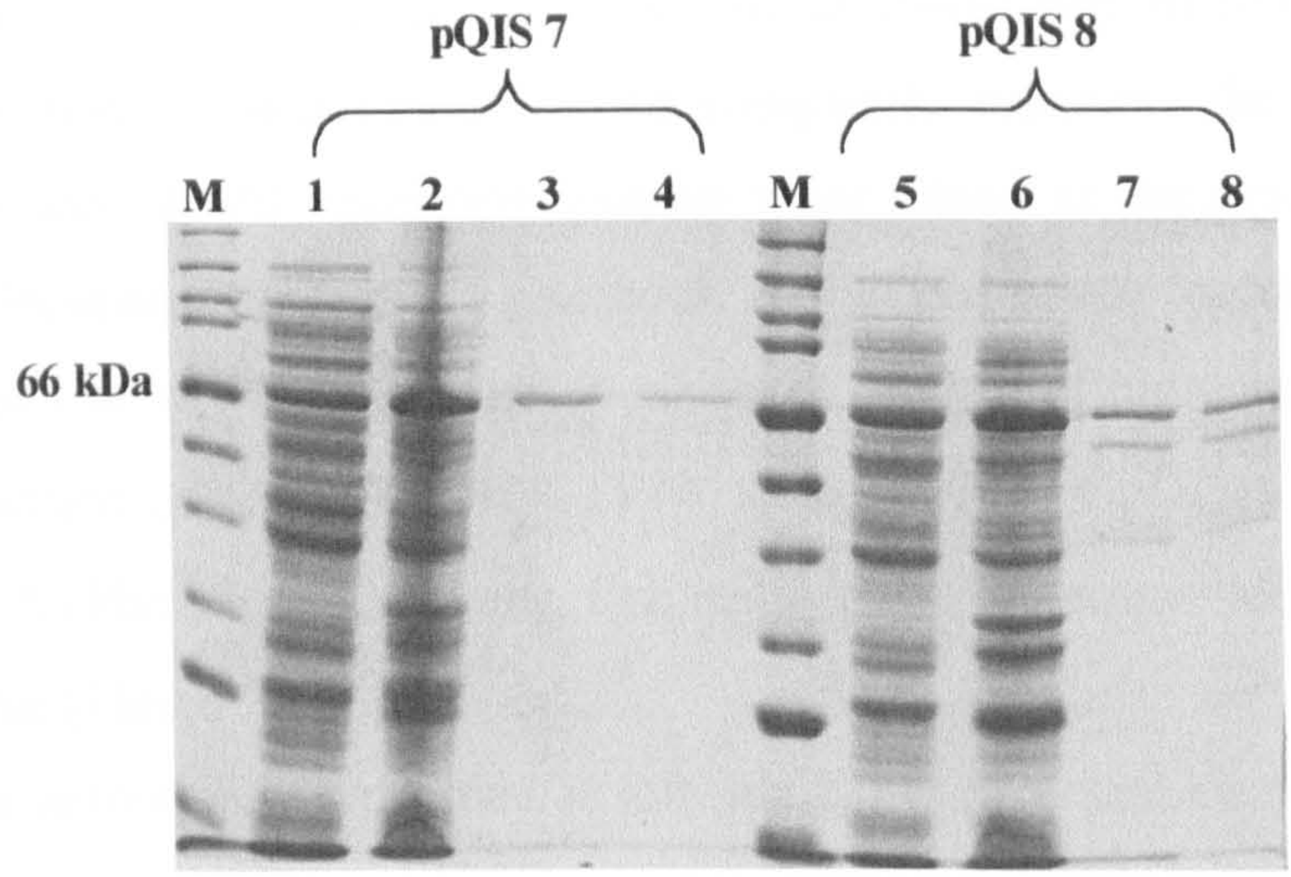


Figure 3.9: SDS PAGE showing the protein expression and purification after expression pf plasmids pQIS 7 (wild type M.SPRI), and pQIS 8 (mutant F 386 A of M.SPRI). The proteins were over-expressed, purified and analysed on 10% SDS-polyacrylamide gels. **M:** Broad range protein molecular weight marker; **1:** 8 µl of Soluble fraction; **2:** 8 µl of Insoluble fraction; **3, 4:** Samples from the first and second elution fractions (6 µl containing 240 ng and 120 ng respectively); **M:** Broad range protein molecular weight marker; **5:** 8 µl Soluble fraction; **6:** 8 µl Insoluble fraction; **7, 8:** Sample from the of first and second elution fractions (6 µl containing 300 ng and 150 ng respectively). The gel was developed with coommassie blue stain followed by destaining.

incubation initiated the required methylation reactions. The reactions were stopped with 1 mM N-Ethylmaleimide (NEM – a modification reagent specific for cysteine residues in proteins) after one hour. An additional step of heat inactivation at 65°C for 20 minutes was carried out to completely inactivate the enzyme. 10X *Hae*III buffer and *Hae*III restriction enzyme were added to the reaction mixtures, which were incubated at 37°C for one hour. At the end of the reaction, the products were analysed on a 1% agarose gel to assess the degree of protection. Figure 3.10 shows the pattern of protection of λ DNA for recombinant pQIS 7 (wild type SPR) and pQIS 8 (Phe 386 Ala mutant). Full methylation protection was observed with 5µl of enzyme (Figure 3.10).

Mtase activity was quantified as follows: 1.0 Unit recombinant SPR Mtase (M.SPRI) is the amount of enzyme required to fully methylate 1.0 µg λ DNA within 60 minutes in a 20 µl reaction volume at 37°C. The M.SPRI concentration varied from batch to batch of protein preparation. In the experiment discussed (and as shown in Figure 3.10) 1.0 unit was equal to 200 ng (at a concentration of 40 ng / µl) of purified wild type M.SPRI. Under these conditions the mutant Phe 386 Ala offered no protection against restriction (compare lanes 6 and 13 in Figure 3.10).

3.1.7.2: Time course assay

As observed from the *in vivo* protection assays, the recombinant mutant Phe 386 Ala was found to be active when the culture had been grown overnight. It was decided to carry out a time course methylation protection assay with both the wild type and mutant proteins. Equimolar concentrations of proteins were added to 1 µg bacteriophage λ DNA in Mtase buffer (50 mM Tris-HCl, pH 7.5, 25°C, 50 mM NaCl, 10 mM EDTA, 1 mM DTT) supplemented with 100 µM SAM. The reaction mixtures were incubated for 30, 60, 90, 120, 150, 300 minutes and overnight to initiate DNA methylation. Reactions were terminated with 1 mM NEM followed by a heat inactivation step at 65°C for 20 minutes. *Hae*III buffer and restriction enzyme were added and the mixture incubated at 37°C for one hour. All the samples were analysed on a 1% agarose gel (Figure 3.11). As expected the wild type M.SPRI was fully functional, methylating λ DNA and rendering it fully refractory to *Hae*III. On the other hand the Phe 386 Ala mutant was only active after a prolonged incubation (compare lane 3 and 8, Figure 3.11). After 30 minutes, the wild type enzyme had

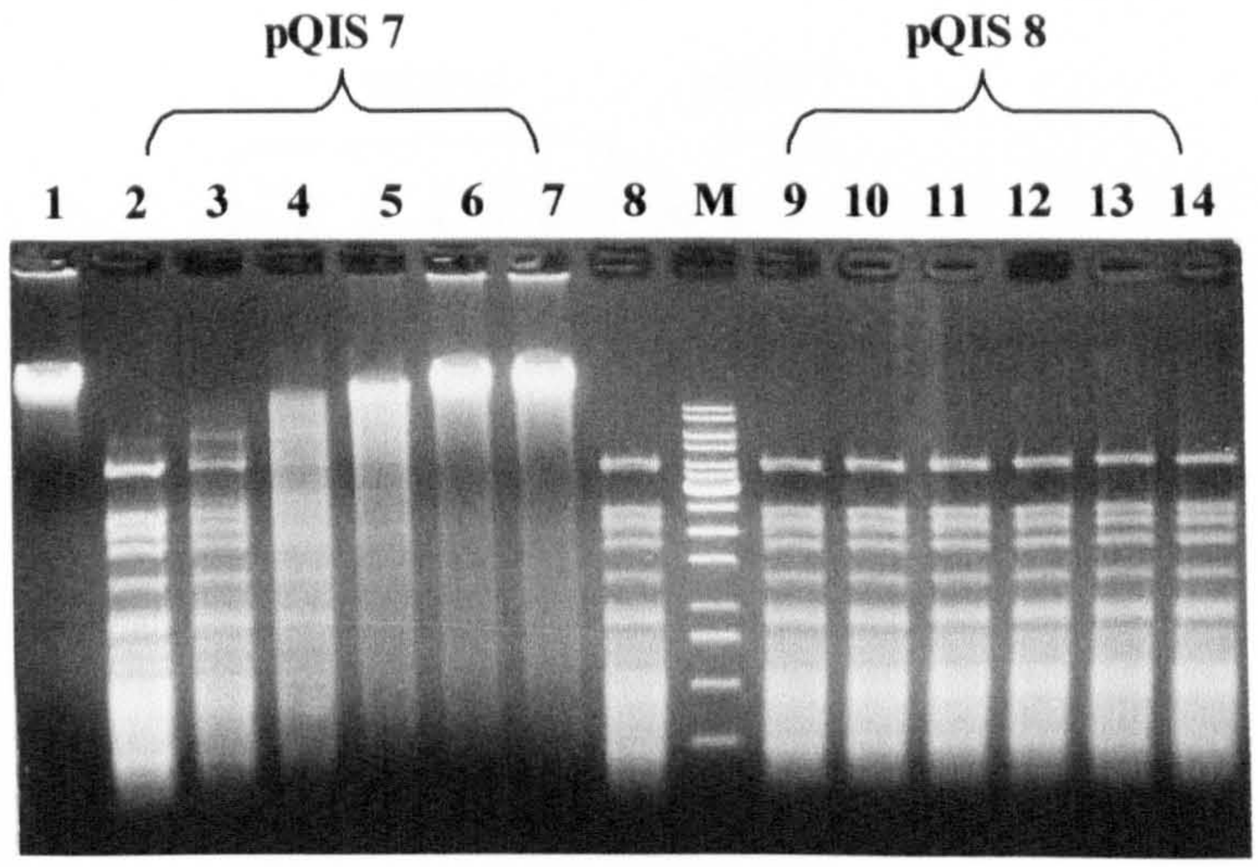


Figure 3.10: An experiment to show the methylation "dose response" of M.SPRI enzymes, using an assay for the protection of λ DNA followed by restriction digestion with *Hae*III.
All the contents of the reaction tubes were analysed on a 1% gel. The reaction tubes were set up as follows.

Lane **1:** 1 μg λ DNA only (no protein and no *Hae*III)
Lane **8:** 1 μg λ DNA + *Hae*III (no protein)
Other lanes contain 1 μg λ DNA, *Hae*III and M.SPRI enzymes.
Lanes **2-7** = wild type M.SPRI
Lanes **9-14** = mutant M.SPRI
Lanes **2, 9** = **8** ng M.SPRI
Lanes **3, 10** = **20** ng M.SPRI
Lanes **4, 11** = **40** ng M.SPRI
Lanes **5, 12** = **80** ng M.SPRI
Lanes **6, 13** = **200** ng M.SPRI
Lanes **7, 14** = **400** ng M.SPRI

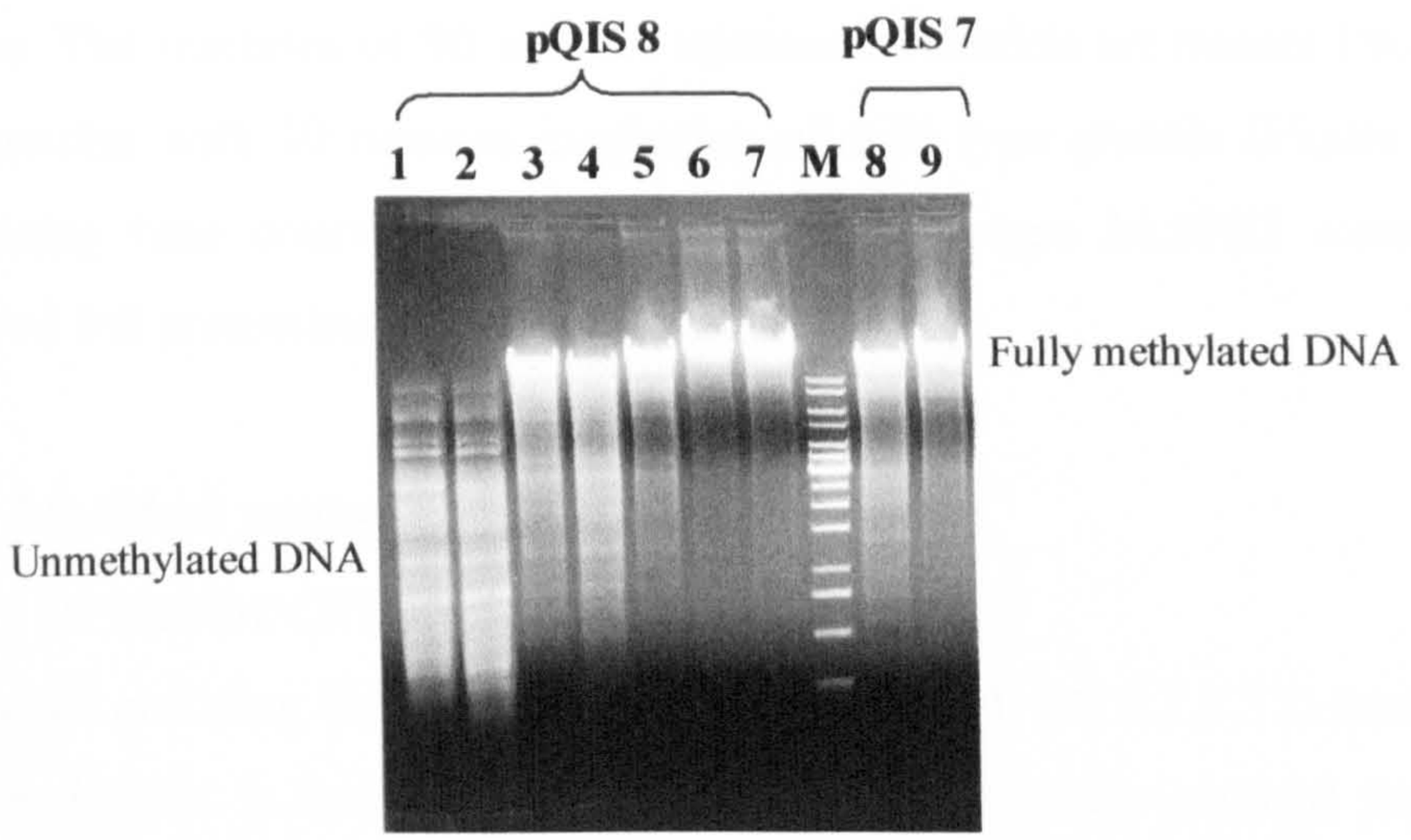


Figure 3.11: Time course protection assay for pQIS 7 (wild type M.SPR) and pQIS 8 (mutant F 386 A) using 1 μg lambda DNA and (200 ng) of purified protein for methylation followed by restriction digestion with *Hae*III. Lane 1-7: 30, 60, 90, 120, 150, 300 minutes and overnight incubation of λ DNA with mutant F 386 A (pQIS 8) protein. **M:** GeneRuler™ 1 kb DNA ladder. **Lane 8-9:** 30 minutes and overnight incubation of λ DNA with wild type M.SPR (pQIS 7) protein. (All other reaction tubes between these timings have same digestion profile as lane 9).

fully methylated the substrate, while the mutant enzyme at the same concentration took more than 4X longer to complete the reaction. Reactions of less than 60 minutes were not sufficient for this mutant to methylate all *Hae*III sites in phage λ DNA. Methylation protection was first observed after 90 minutes incubation and therefore compared to the wild type there is a significant delay in methylation with the mutant Mtase. The reactions of 90 and 120 minutes incubation for mutant Phe 386 Ala were comparable with 30 minutes incubation of wild type protein (Figure 3.11). All the remaining time course assay reactions for wild type M.SPRI were constant and showed full protection (not shown in figure).

3.2: M.*Hha*I mono-specific Mtase

3.2.1: INTRODUCTION

The gene encoding the C5-cytosine Mtase (M.*Hha*I, EC 2.1.1.37) from *Haemophilus heaemolyticus*, is encoded by 981 bp and was originally isolated from a library of plasmid clones carrying *Hind*III digested DNA fragments (Caserta *et al.,* 1987). M.*Hha*I is one of the smallest of the C5-Mtases, containing 327 amino acids and having a molecular weight of 37 kDa. Recombinants carrying the M.*Hha*I gene from the library were isolated and then cloned into plasmid pNW2801 for analysis. The full DNA sequence and translation product is shown in Figure 3.12. M.*Hha*I recognises the palindromic sequence 5'---GCGC---3' and methylates the inner cytosine to produce 5'---GmCGC---3'. The three dimensional structure of M.*Hha*I and implications for functions are described in Chapter One. The gene encoding M.*Hha*I was obtained from Dr. Geoffrey Wilson (New England Biolabs) and was sub-cloned into two different vectors (pGEX-KG and pLITMUS-28) to improve expression and to facilitate purification (Kan, 1999). Both plasmids carrying the *Hha*I Mtase gene were propagated in the *E. coli* strain DH5α *mcr*A$^-$BC$^-$ and were used in the following experiments.

The amino acid sequence of M.SPRI and M.*Hha*I were aligned and are presented in figure 3.12 (a).

```
1     ATG ATT GAA ATA AAA GAT AAA CAG CTC ACA GGA TTA CGC TTT ATT GAC
        M   I   E   I   K   D   K   Q   L   T   G   L   R   F   I   D
49    CTT TTT GCA GGA TTA GGT GGC TTT AGA CTT GCT TTA GAA TCT TGC GGT
        L   F   A   G   L   G   G   F   R   L   A   L   E   S   C   G
97    GCT GAG TGC GTT TAT TCT AAT GAA TGG GAT AAA TAT GCA CAA GAA GTA
        A   E   C   V   Y   S   N   E   W   D   K   Y   A   Q   E   V
145   TAT GAG ATG AAT TTT GGT GAA AAG CCT GAG GGC GAC ATT ACC CAA GTA
        Y   E   M   N   F   G   E   K   P   E   G   D   I   T   Q   V
193   AAT GAG AAA ACC ATT CCT GAT CAC GAC ATT TTA TGT GCA GGG TTT CCG
        N   E   K   T   I   P   D   H   D   I   L   C   A   G   F   P
241   TGC CAA GCG TTT TCT ATT AGT GGA AAA CAA AAA GGA TTC GAG GAC AGC
        C   Q   A   F   S   I   S   G   K   Q   K   G   F   E   D   S
289   AGA GGT ACG CTC TTT TTT GAT ATT GCA CGT ATT GTC CGT GAA AAA AAA
        R   G   T   L   F   F   D   I   A   R   I   V   R   E   K   K
337   CCT AAA GTG GTT TTT ATG GAA AAT GTG AAA AAT TTT GCA TCG CAT GAT
        P   K   V   V   F   M   E   N   V   K   N   F   A   S   H   D
385   AAT GGA AAT ACG TTA GAA GTT GTA AAA AAT ACA ATG AAT GAA TTG GAC
        N   G   N   T   L   E   V   V   K   N   T   M   N   E   L   D
433   TAT TCT TTT CAT GCT AAA GTA TTA AAT GCT TTA GAT TAT GGG ATT CCA
        Y   S   F   H   A   K   V   L   N   A   L   D   Y   G   I   P
481   CAG AAA AGG GAA CGT ATC TAT ATG ATT TGT TTT CGC AAT GAT CTC AAT
        Q   K   R   E   R   I   Y   M   I   C   F   R   N   D   L   N
529   ATT CAA AAT TTC CAA TTT CCA AAA CCT TTT GAG CTT AAT ACT TTT GTG
        I   Q   N   F   Q   F   P   K   P   F   E   L   N   T   F   V
577   AAA GAT TTG TTA TTA CCT GAT AGC GAG GTG GAA CAC TTA GTT ATT GAT
        K   D   L   L   L   P   D   S   E   V   E   H   L   V   I   D
625   AGA AAA GAT TTG GTA ATG ACA AAC CAA GAA ATT GAG CAA ACA ACC CCC
        R   K   D   L   V   M   T   N   Q   E   I   E   Q   T   T   P
673   AAA ACA GTT CGA CTT GGT ATT GTA GGA AAA GGT GGG CAA GGA GAA CGA
        K   T   V   R   L   G   I   V   G   K   G   G   Q   G   E   R
721   ATT TAT AGC ACA AGA GGC ATT GCA ATT ACC TTA TCT GCT TAT GGT GGC
        I   Y   S   T   R   G   I   A   I   T   L   S   A   Y   G   G
767   GGC ATT TTC GCT AAG ACA GGG GGA TAT TTA GTA AAC GGG AAG ACA CGG
        G   I   F   A   K   T   G   G   Y   L   V   N   G   K   T   R
817   AAA TTA CAC CCT AGA GAG TGT GCT AGA GTA ATG GGC TAC CCA GAT AGT
        K   L   H   P   R   E   C   A   R   V   M   G   Y   P   D   S
865   TAT AAA GTC CAC CCG TCA ACC AGC CAA GCA TAT AAA CAA TTT GGT AAC
        Y   K   V   H   P   S   T   S   Q   A   Y   K   Q   F   G   N
913   TCA GTT GTT ATC AAT GTA CTT CAA TAT ATT GCT TAT AAC ATT GGT TCA
        S   V   V   I   N   V   L   Q   Y   I   A   Y   N   I   G   S
961   TCA TTA AAT TTC AAA CCA TAT TAA
        S   L   N   F   K   P   Y Stop
```

Figure 3.12: Nucleotide and amino acid sequences of M.*Hha*I. The amino acids present in the six highly conserved motifs of C5-Mtases are coloured as follows: Motif I (red), motif IV (brown), motif VI (green), motif VIII (light blue), motif IX (pink) and motif X (dark blue). The TRD region between motif VIII and IX is shown with underlined amino acids. Data from GenBank™/EMBL Data Bank: accession number J02677.

Wild Type *M.HhaI*

```
1  MIEIKDKQLTGLRFIDLFAGLGGFRLALESCGA  ECVYSNEWDKYA  QEVYEMNFGE  KPEGDITQVNEKTIPDHDILCAGF
                                                                                      69
```

Wild Type *M.SPR*

```
1  MGKLRVMSLFSGIGAFEAALRNIGVGYELVGFSEIDKYAVKSFCAIHNVDEQLNFGDVSKIDKKKLPEFDLLVGGS
                                                                                  66
```

I II III

Motif IV-VIII

```
80  PCQAFSISGKQKGFEDSRGTLFFDIARIVREKKPKVVFMENVKNFASHDNGNTLEVVKNTMNELDYSFHAKVLNALDYGIPQKRERIYMICFRND
150
77  PCQSFSVAGHRKGFEDTRGTLFFQYVETLKEKQPKFFVFENVKGLINHDKGNTLNVMAEAFSEVGYRIDLELLNSKFFNVPQNRERLYIIGIRED
150
```

IV V VI VII VIII

Variable and TRD region

TRD of *M.HhaI*

```
175  LNIQNFQFPKPFELNTFVKDLLLPDSEVEHLVIDRKDLVMTNQEIEQTTPKTVRLGIVGKGGQGERIYSTRGIAITLSAYGGGIFAKTG
250
172  LIKNEEWSLDFKRKDILQKGKQRLVELDIKSFNFRWTAQSAATKRLKDLLEEYVDEKYYLNEDKTNSLIKELSTSRLNENLTVEQVGNINPS
240
```

M.SPR *Eco*RI TRD | M.SPRI *Msp*I TRD | M.SPR *Hae*III TRD

```
264  GNGMNGNVYNSSGLSPTITTNKGEGLKIAVEYSRKSGLGRELAVSHTLSASDWRGLNRNQKQNAVVEVRPVLTPERGEKRQNGRRFKDDGEP
369
356  AFTVNTIDRHGVAV
369
```

Motif IX & X

```
264  GYLVNGKTRKLHPRECARVMGYPDSYKVHPST  SQAYKQFGNSVVINVLQYIAYNIGSSLNFKPY
420
370  GEYPKYRIRRLTPLECFRLQAFDDEDFEKAFAAGISNSQLYKQAGNSITVTVLESIFKELIHTYVNKESE
420
```

IX X

Figure 3.12 (a): Sequence alignment of of M.*Hha*I and M.SPRI. The motifs are underlined whereas the conserved consensus sequences in each motif is shown in red. Colours are shown as in secondary structure of M.*Hha*I (motif I red; motif IV, yellow; motif IX, magenta; and motif X, blue). TRD region is shown in rectangles.

3.2.2: MUTAGENESIS OF WILD TYPE MOTIF IX OF M.*Hha*I TOWARDS M.SPRI

The consensus sequence of motif IX in C5-Mtases is **RXXXXXEXXR**. In monospecific Mtases this region is moderately conserved whereas in multispecific Mtases, motif IX is strongly conserved (Figure 3.5 and Table 3.1). However, our mutagenesis experiment on motif IX of multispecific Mtase confirmed that although multispecific Mtase have a strongly conserved motif IX, they have the capability to accommodate certain mutations.

We decided to carry out an experiment for mutagenesis of wild type motif IX of M.*Hha*I in view of the results obtained from the mutagenesis of multispecific Mtase M.SPR. As a starting point Ala 280 was chosen which aligns with Phe 386 of M.SPRI. To understand the nature of the experiment, a summary of the mutants constructed is shown in figure 3.12 (b) and the schematic diagrams are placed at the end of the chapter.

3.2.2.1: Ala 280 Phe mutagenesis in wild type M.*Hha*I

The region of M.*Hha*I targeted for mutation was motif IX-X (as with M.SPRI). Some amino acids of this region are strongly conserved, such as Ala 280. Although Ala 280 in M.*Hha*I is strongly conserved in other monospecific Mtases, it aligns with Phe 386 in M.SPRI and this residue was therefore chosen for mutation. In order to mutagenise Ala 280 of M.*Hha*I it was convenient to introduce a new restriction site into this region of the gene. Analysis of the sequence of the gene showed that the introduction of a new *Sty*I site was possible without changing the amino acid sequence. In addition a single base change would also change the codon for alanine to phenylalanine. The oligonucleotides NIS 15 and NIS 16 (shown below) were used

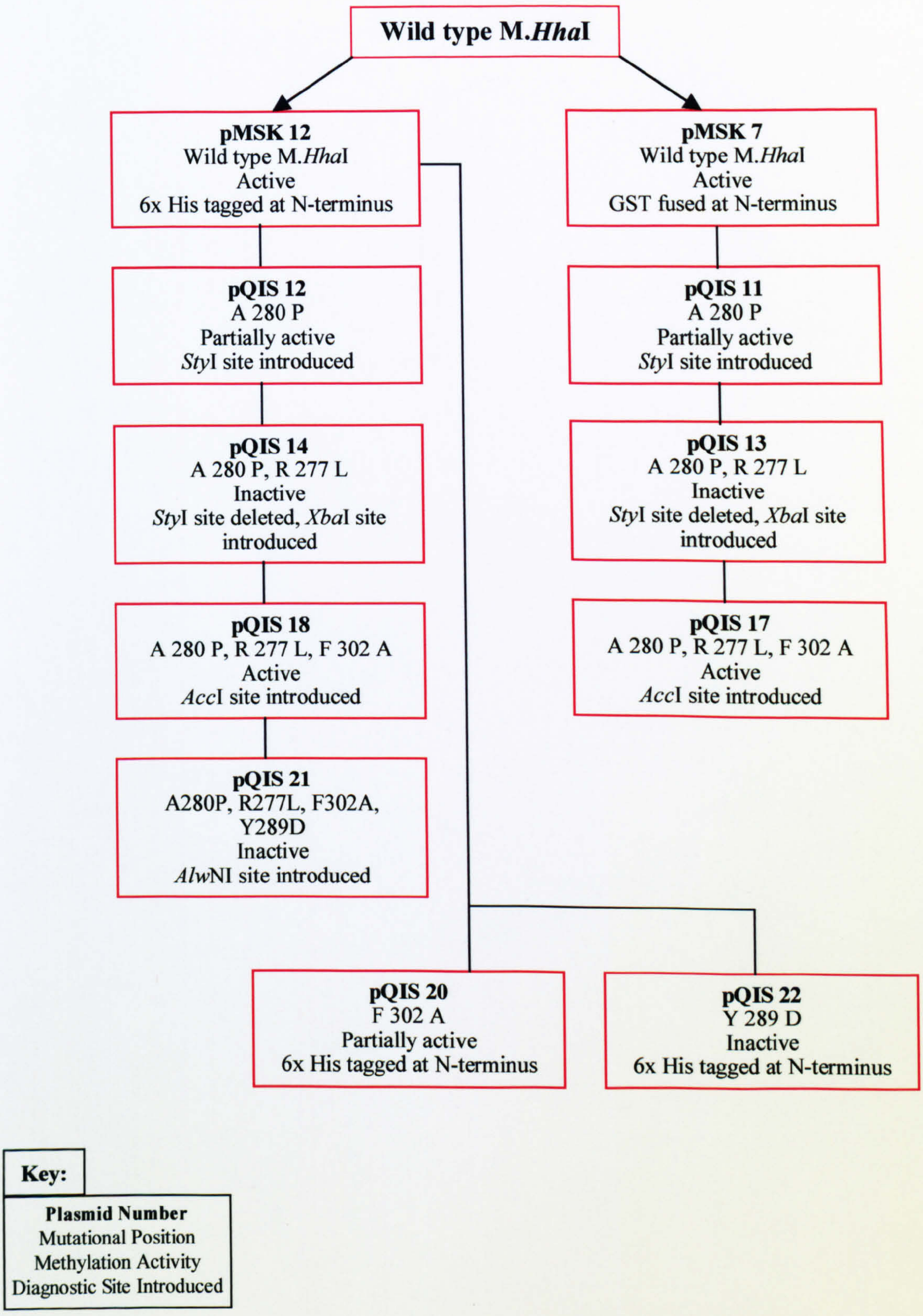


Figure 3.12 (b): An schematic representation of the formation of mutants of M.*Hha*I. Two constructs of M.*Hha*I one containing GST fusion and the other with 6x histidine tag were simultaneously used for the experiment.

NIS 15/NIS 16 duplex oligonucleotide:

*Sty*I

5'—GGAAATTACAC**CCTAGG**GAGTGT**TTT**AGAGTAATGGGCTAC—**3'**

3'—CCTTTAATGTG**GGATCC**CTCACA**AAA**TCTCATTACCCGATG—**5'**

K L H P R E C F R V M G Y

NIS 15: **5'**—GGAAATTACACCCTAG**G**GAGTGT**TTT**AGAGTAATGGGCTAC—**3'**

NIS 16: **5'**—GTAGCCCATTACTCT**AAA**ACACTC**C**CTAGGGTGTAATTTCC—**3'**

for this mutagenesis experiment. Two templates were used for site directed mutagenesis, namely pMSK 7 (wild type M.*Hha*I in pGEX-4T-3 GST fused) and pMSK 12 (wild type M.*Hha*I in pLITMUS-28 with 6 histidines).

The resulting constructs were called as pQIS 11 and pQIS 12 as shown in Figure 3.13. Both mutant constructs pQIS11 and pQIS 12 were checked for the presence of a new *Sty*I site and for M.*Hha*I activities. There is no *Sty*I site in pMSK 7, while three *Sty*I sites exist in pMSK 12. The mutant construct pQIS 11 acquired a new *Sty*I site and was linearised by *Sty*I digestion. pQIS 12 also contained a new *Sty*I site. The 1085 bp DNA fragment was cut into two fragments of sizes 887 bp and 198 bp, which confirmed the presence of a new restriction site (Figure 3.13). Restriction digestion of plasmids with *Hha*I established that both mutants encode an inactive Mtase. However, when plasmids obtained from overnight cultures that were induced with 0.1 mM IPTG were analysed, it was observed that the proteins encoded by these plasmid possessed partial activity. It is known that even a single alteration may change multiple properties of a protein, including its structure, folding pathway, flexibility, stability, and enzymatic properties (Coombs and Corey, 1998). The availability of a crystal structure of M.*Hha*I (Cheng *et al*., 1993) allowed us to have a closer look (with the assistance of Dr. G. C. Ford) at the region surrounding these residues. It was also instructive to identify the amino acids within the vicinity of these residues that constitute motif IX. We analysed the amino acids within a 4.0 Å radius of each amino acid in motif IX of M.*Hha*I and these are given in Table 3.1. This distance is sufficient to cover interactions such as hydrogen bonds, ionic and Van der Waals interactions, and hydrophobic packing. Therefore this analysis includes all amino acids, which may be directly disturbed by mutation. From the

above analysis it was found that the oxygen of Arg 277 interacts both with the nitrogen and the β carbon of Ala 280. It became apparent that by altering the Ala 280 to Phe of M.*Hha*I the interaction between Ala 280 and Arg 277 could well be disturbed, perhaps accounting for a partial loss of catalytic activity. Phenylalanine carries an aromatic side chain; the volume of which is 190 $Å^3$: more than double to that of Ala (88.6 $Å^3$). It seems that as a consequence of this mutation, the Arg 277 present in the same alpha helix may be sterically hindered in M.*Hha*I and as a result M.*Hha*I becomes partially active.

3.2.2.2: Arg 277 Leu mutagenesis in the mutant Ala 280 Phe of M.*Hha*I

Arg 277 appears to be highly conserved in motif IX of mono-specific Mtases (Figure 3.5 and Table 3.1). Leu is proposed as a possible replacement because it is found in all multi-specific Mtases at this position (with the exception of M.ϕ*Bss*HII). A pair of oligonucleotides was synthesised for this mutagenesis experiment, which would lead to the replacement of the Arg codon with that of Leu. These oligonucleotides NIS 17 and NIS 18 (shown below), introduce a unique *Xba*I site into the gene, for diagnostic purpose. In addition, due to this change, the *Sty*I site that had been introduced earlier would be deleted (see above). Substitution of Arg 277 with Leu was carried out by site directed mutagenesis using pQIS 11 and pQIS 12 as templates (Figure 3.14).

NIS17/NIS18 duplex oligonucleotide

*Xba*I

5'—CGGAAATTACACCC**TCTAGA**GTGTTTTAGAG—**3'**

3'—GCCTTTAATGTGGG**AGATCT**CACAAAATCTC—**5'**

R K L H P **L** E C F R

NIS 17: **5'**—CGGAAATTACACCC**TCTAGA**GTGTTTTAGAG—**3'**

NIS 18: **5'**—CTCTAAAACAC**TCTAGA**GGGTGTAATTTCCG—**3'**

Site-directed mutagenesis experiments to generate the two new double mutants pQIS 13 and pQIS 14, were carried out simultaneously and the presence of a new *Xba*I restriction site confirmed the replacement of the arginine codon with that of leucine.

When subjected to a restriction analysis with *Hha*I, both the recombinant constructs were deduced to be completely inactive and no significant difference was observed after overnight growth in the presence of IPTG (Figure 3.14).

Once again the structure of M.*Hha*I was examined in more detail in and around position 277 and 280, and it was found that phenylalanine at position 302 is in close proximity to the side chain of residue 280 where the Ala 280 Phe mutation had been introduced. Phe 302 is present in the α-helix running parallel to the α-helix of motif IX and its bulky side chain is positioned towards the side chain of Phe 280 in the mutant. Therefore, assuming that the two bulky groups of phenylalanine are vying for the same space, it was decided to replace the phenylalanine at position 302 with a smaller amino acid such as alanine.

3.2.2.3: Phe 302 Ala mutagenesis in the double mutant Arg 277 Leu, Ala 280 Phe of M.*Hha*I

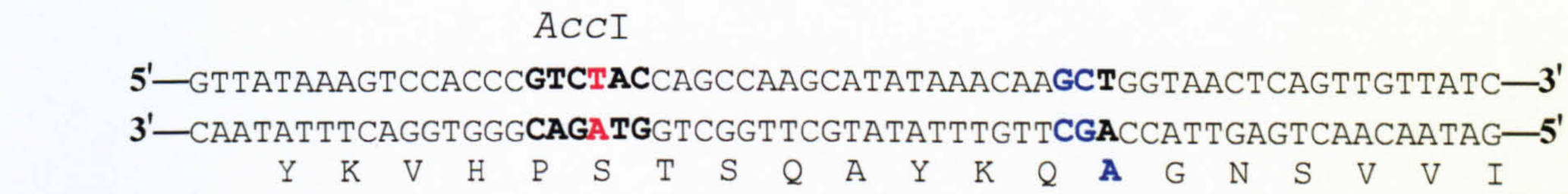
Sequence of motif X of M.*Hha*I revealed that Phenylalanine at position 302 is located just before the consensus amino acid sequence **GNXV**. Comparison of various mono-specific Mtases sequences showed that Phe 302 of M.*Hha*I is not conserved (Figure 1.4).

Amino acid sequence of Motif X of M.SPRI and M.*Hha*I

M.SPR:	Y	K	Q	A	**G**	**N**	S	I	T	V
M.*Hha*I:	Y	K	Q	F	**G**	**N**	S	**V**	V	I
				302						

When the sequence of M.*Hha*I was compared with that of M.SPRI in motif X (shown above) it was observed that there was an alanine instead of a phenylalanine. In order to alter Phe 302 of M.*Hha*I into Ala two oligonucleotides NIS 19 and NIS 20 were synthesised. The duplex is shown below, with changes shown in red, and the new restriction site *Acc*I (silent mutation) is shown in bold. In addition, a *Hin*cII site was deleted in this region, which provided an additional diagnostic test. The oligonucleotides are shown below with the mutated bases in red (silent mutation for *Acc*I site) and blue for the desired Phe 302 Ala codon change.

NIS 19/NIS 20 duplex oligonucleotide



Site-directed mutagenesis was carried out to construct a triple mutant of M.*Hha*I using the templates pQIS 13 and pQIS 14. The new mutant constructs now carry three mutations: Ala 280 Phe, Arg 277 Leu and Phe 302 Ala (Figure 3.15). The mutant pQIS 17 was confirmed on the basis of the loss of a *Hinc*II site. The mutant pQIS 18 was confirmed by the presence of a new, unique *Acc*I site (Figure 3.15). When the triple mutant plasmids were tested with *Hha*I for their methylation status, it was found that the mutant pQIS 18 expressed a protein that has fully regained Mtase activity (Figure 3.15). The mutant construct pQIS 18 was sequenced and was found to contain all the expected mutations and no other changes. Part of the sequencing analysis showing all the three mutations is shown in Figure 3.16.

3.2.2.4: Phe 302 Ala mutagenesis in wild type M.*Hha*I

The same oligonucleotides NIS 19 and NIS 20 (shown above) were also used for constructing the single mutant Phe 302 Ala in wild type M.*Hha*I to see if this alone had any effect on the activity of M.*Hha*I. The single mutant pQIS 20 was generated using pMSK 12 (that encodes wild type M.*Hha*I) as a template (Figure 3.17). The construct pQIS 20 encoded a protein that shows partial activity. Graphical analysis of the region surrounding residue Phe 302 revealed that the oxygen atom of Phe is close to the ε carbon atom of Met 283, in addition to its neighbouring atoms such as Ala 298, Lys 300 and Asn 304 which may have been disturbed by the substitution and this therefore may account for the reduction in catalytic activity of the enzyme.

3.2.2.5: Tyr 289 Asp mutagenesis in the triple mutant (Arg 277 Leu, Ala 280 Phe, Phe 302 Ala) of M.*Hha*I

It has been observed from the crystal structure of M.*Hha*I (Cheng *et al.*, 1993) that Tyr 289 makes close interaction with the NH1 and NH2 groups of Arg 277, in addition to other residues such as Leu 21, Asn 52, Pro 286 and Asp 287. Because

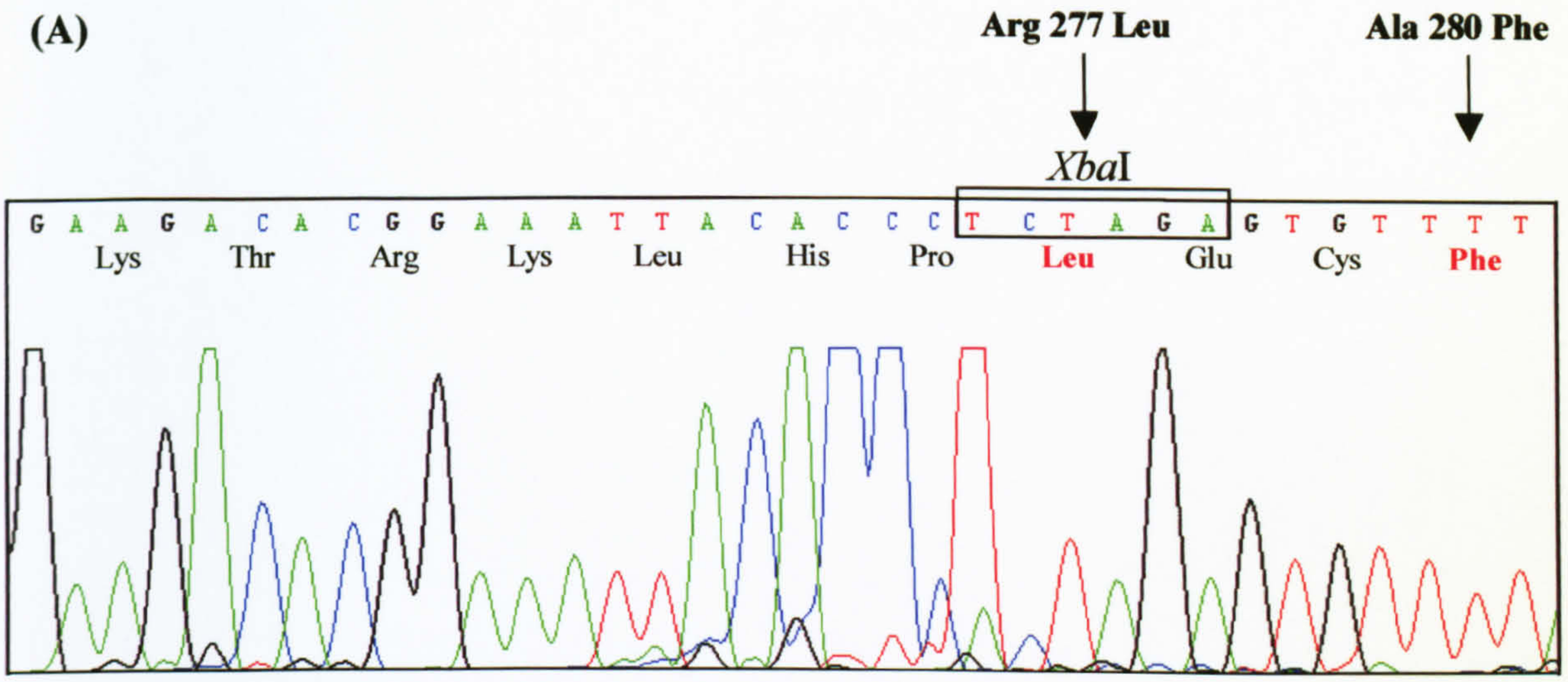


Figure 3.16: Part of the DNA sequencing analysis chromatogram for pQIS 18 (encoding a triple mutant of M.*Hha*I). (A): Arg 277 Leu (with new *Xba*I restriction site) and Ala 280 Phe. **(B):** Phe 302 Ala with new *Acc*I site.

Arg 277 has already been mutated to Leu, it was assumed that there might be a positive effect on catalytic activity if Tyr 289 was replaced with an amino acid found in M.SPRI. Therefore another mutagenic experiment was designed to change Tyr 289 (in the triple mutant of M.*Hha*I, Arg 277 Leu, Ala 280 Phe and Phe 302 Ala) to Asp, to determine its effect upon the activity of the enzyme. Two oligonucleotides

NIS 21/NIS 22 duplex oligonucleotide

*Alw*NI

5'-GGCTACC**CAGATTCTGACAAAGTCCACCCGTC-3'**

3'-CCGATGG**GTCTAAGACTGTTTCAGGTGGGCAG-5'**

G Y P D S **D** K V H P

NIS 21: **5'-**GGCTACCCAGAT**TC**T**G**A**C**AAAGTCCACCCGTC**-3'**

NIS 22: **5'-**GACGGGTGGACTTT**G**T**C**A**GA**ATCTGGGTAGCC**-3'**

were designed and synthesized for this site-directed mutagenesis experiment. There is one *Alw*NI site already present in the sequence of the triple mutant of M.*Hha*I (encoding for Arg 277 Leu, Ala 280 Phe and Phe 302 Ala) construct number pQIS 18. Therefore the new construct can easily be detected on the basis of a newly introduced *Alw*NI site. The oligonucleotides NIS 21 and NIS 22 are shown below with mutated bases indicated in red (silent mutation for *Alw*NI site) and blue for the desired Tyr 289 Asp codon mutation.

The mutagenesis experiment was carried out, and the new construct was called pQIS 21, a quadruple mutant of M.*Hha*I carrying the following mutations (Arg 277 Leu, Ala 280 Phe Tyr 289 Asp and Phe 302 Ala) (Figure 3.18). The authenticity of the construct was confirmed first on the basis of a newly introduced *Alw*NI site (Figure 3.18). To determine the effect of this mutation upon the activity of the enzyme, the recombinant plasmid was challenged with an excess of *Hha*I. The quadruple mutant as determined in this manner was found to be inactive. It appears that the triple mutant pQIS 18 (Arg 277 Leu, Ala 280 Phe and Phe 302 Ala) has altered the enzyme in terms of regional interactions and structural integrity but this mutation Tyr 289 Asp has a more pronounced effect on the stability of M.*Hha*I. The triple mutant provides a satisfactory alternative constellation of amino acid side chains for the activity of M.*Hha*I to be retained, but that the fourth mutant in which Tyr 289 is substituted by Asp is incompatible with biological activity of the enzyme.

For completeness it was important to characterize the Tyr 289 Asp single mutant of M.*Hha*I to see if this mutation alone had any effect on the stability/activity of the enzyme.

3.2.2.6: Tyr 289 Asp mutagenesis of wild type M.*Hha*I

The same oligonucleotides NIS 21/NIS 22 that were used for the above experiment were used for this mutagenesis (Figure 3.19). The new construct pQIS 21, which is wild type M.*Hha*I in which Tyr 289 is replaced by Asp was verified by the presence of an *Alw*NI site. The activity of the resultant plasmid pQIS 22 was checked by *Hha*I digestion: the modified Mtase M.*Hha*I was found to be completely inactive. The result shows that this Tyr 289 residue is also critical for catalytic activity. Tyr 289 is partially buried in the M.*Hha*I molecule and the analysis of the crystal structure of the enzyme revealed that -OH group of Tyr is hydrogen bonded with Oxygen of Leu 21 (in the absence of SAM). In addition, Tyr 289 also makes interactions with Pro 286, Asp 287, Asn 52 and Arg 277. In the presence of SAM, the Tyr 289 makes interactions with Pro 286, Leu 21 and Arg 277. Although Tyr 289 is not the core of the enzyme but mutation to Asp has a strong adverse effect on catalytic activity. Asp is much smaller in size than Tyr and its replacement leaves very large space in the middle of the protein and in addition its acidic nature leads to drastic effects that render the enzyme inactive. A representation of the mutations introduced into M.SPRI and M.*Hha*I in this study are presented in Figure 3.20.

3.3: The use of WAVE DHPLC system to analyse the activity of wild type and mutant DNA Mtases

The activity of both wild type and mutant M.*Hha*I and M.SPRI can be assessed using an *in vivo* methylation protection assay. The methylation protection assay examines the ability of methylation by restriction with relevant restriction enzyme to digest the plasmid encoding the Mtase (for example, digestion of M.SPRI encoding plasmid with *Hae*III or *Msp*I). However, there are a number of factors that can limit the amount of information that is obtained.

a) The digestion of the plasmid can only be informative in terms of whether the DNA can be digested by a restriction enzyme or not. Although the presence of a partially active Mtase in the methylation protection assay can lead to a

Figure 3.20: Amino Acid sequence of motif IX & X from wild type M.SPRI & M.*Hha*I in and their mutants. Red blocks are made to highlight the two motifs under study. Those amino acids of the wild type enzyme that have been exchanged by others are indicated by arrows. The mutants of M.SPRI are shown above its amino acid sequence and the mutants for M.HhaI are shown below its sequence. Activities of mutant enzymes are shown with **+**, **+/-** and **-** indicating **high**, **reduced** and **absence of methylation activity** respectively. Amino acids shown in red are conserved in all Mtases.

diagnostically significant DNA fragment smear on the agarose gel, this does not give any quantitative measure of the Mtase activity.

b) With a partially active Mtase there is a trade off between using enough DNA to obtain a clear result and using too much (and creating a situation where not all of the DNA can be cut by the restriction enzyme).

c) In addition, if there is any uncut DNA present in the tube because of reduced methylation activity of mutant Mtase, it is difficult to visualize that uncut DNA because it is present in very small amounts. Therefore, the information regarding the reduced level of activity is difficult to extract from the gel assays.

An alternative to gel analysis of the DNA fragments is the use of HPLC. Therefore, an experiment was designed to analyse and compare the activity of three different mutant Mtases by HPLC, in order to attempt to quantify differences in plasmid methylation profiles.

3.3.1: COMPARISON OF THE ACTIVITY OF MUTANT Mtases USING HPLC

Three mutants with different levels of enzymatic activity (active, partially active and inactive) were chosen and their activity was analysed by HPLC. The mutant plasmid pQIS 18 that codes for a triple mutant of M.*Hha*I (Arg 277 Leu, Ala 280 Phe, Phe 302 Ala) was used as a control for an active enzyme. The mutant construct pQIS 12 that was constructed by exchanging Ala 280 to Phe and showed partial activity was selected as an example of partial active Mtase. The plasmid pQIS 14 (encoding for the Ala 280 Phe and Arg 277 Leu in M.*Hha*I) expressed an inactive enzyme and was selected as an example of inactive Mtase. The comparative agarose gel profile for the three plasmids pQIS 18, pQIS 14 and pQIS 12 after *Hha*I cleavage is shown in Figure 3.21.

In order to compare the three plasmids by HPLC, samples were injected onto the WAVE system after *Hha*I cleavage, using the optimised gradient as described in section 2.2.12 (Table 2.6). The resulting chromatograms for the three plasmids were overlaid and are shown in Figure 3.22. The digestion profile for pQIS 18 showed that it has not been cleaved by the action of *Hha*I. As a result of this, the intact DNA is eluted at the end of elution with in the high acetonitrile wash phase (See the chromatogram shown in purple, Figure 3.22). This shows that the plasmid encodes an active enzyme and protected against restriction with its cognate restriction enzyme

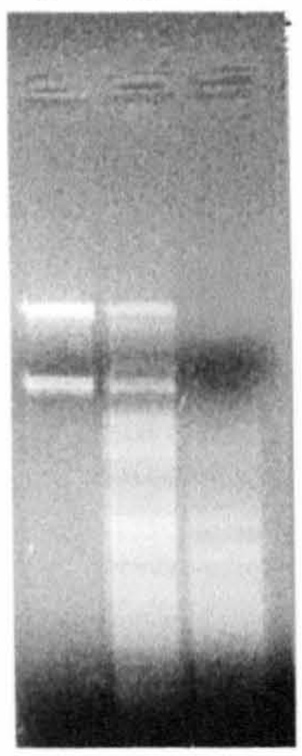


Figure 3.21: Assessment of methyltransferase activity using an *Hha*I restriction analysis on 1% Agarose Gel.
Lane 1: pQIS 18 (Active methyltransferase).
Lane 2: pQIS 12 (Partially active Mtase).
Lane 3: pQIS 14 (Inactive Mtase).
Methylation of the plasmids *Hha*I sites prevents digestion by the *Hha*I restriction enzyme.

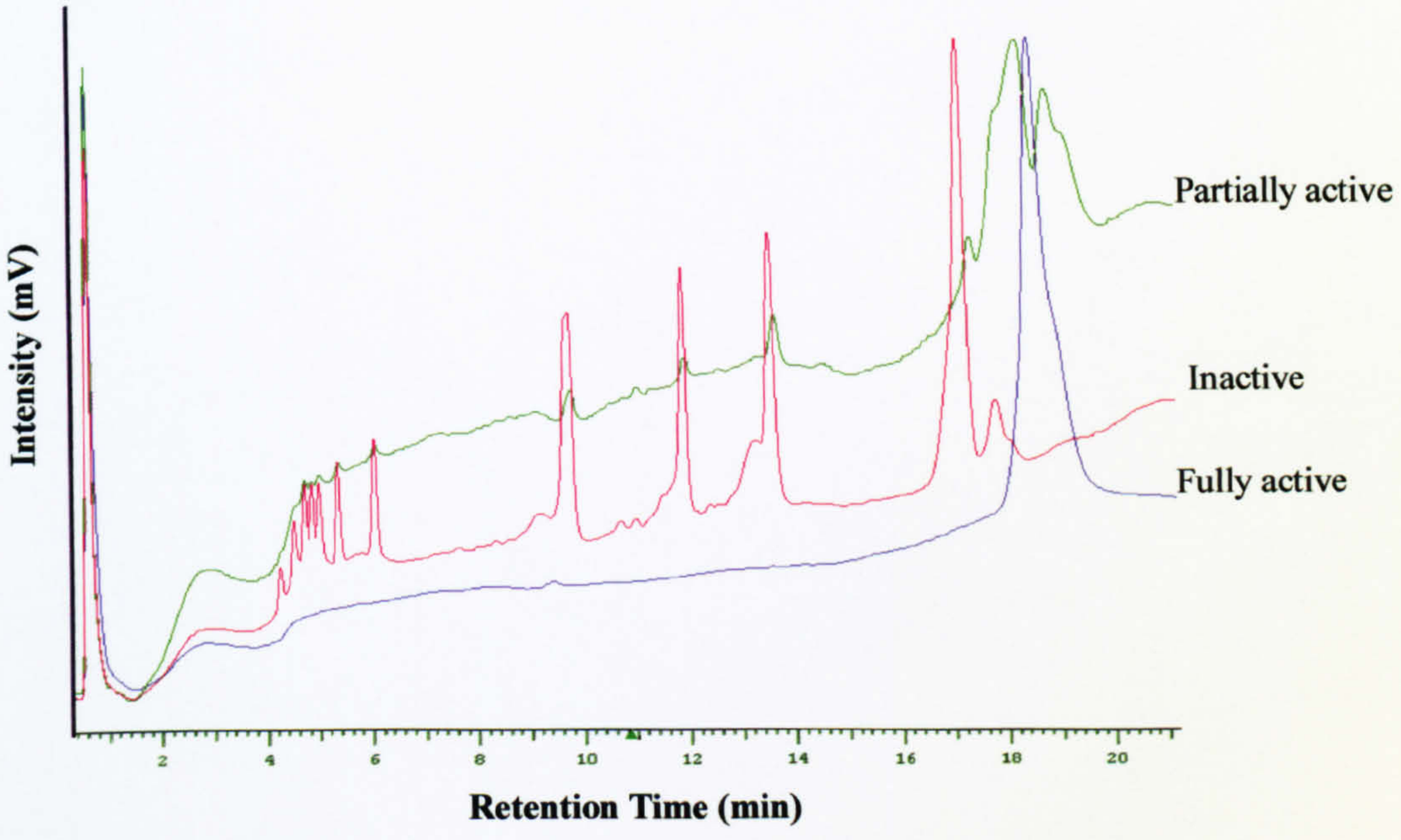


Figure 3.22: Assessment of methyltransferase activity using DHPLC analysis of *Hha*I restriction digests.
Blue Trace: pQIS 18 (Active methyltransferase).
Green Trace: pQIS 12 (Partially active Mtase).
Red Trace: pQIS 14 (Inactive Mtase).
The results are consistent with our previous *in vivo* methylation protection assays analysed on agarose gel (see Section 3.1.5 and 3.2.2). The slight increase in base line is due to the increased concentration of Acetonitrile.

*Hha*I. HPLC analysis for pQIS 14 separates the DNA fragments in clearly visible peaks (See the chromatogram shown in red, Figure 3.22). The peaks in the chromatogram clearly indicate that the construct is completely degraded and not protected. Therefore the protein encoded by this construct is catalytically inactive. HPLC analysis of pQIS 12 showed that it is only partially restricted resulting in the appearance of very small amount of peaks. Most of the sample was only eluted at the end of HPLC run, indicating a partially active construct (shown in green) (Figure 3.22). Hence, the difference in the methylation profile is clearly comparable by this HPLC analysis.

3.4: Discussion

In this chapter the mutagenesis experiments on multi-specific Mtase SPRI and mono-specific Mtase *Hha*I are presented. The summary of the mutagenesis experiments with the amino acid sequences of mutant constructs and the observed methylation activity is given in Figure 3.23. Mutagenesis experiments were focused on motif IX and X to evaluate the structure function relationship of the enzyme in general and the TRD region and motif IX in particular. The nucleotide sequences of the mutants are also aligned with the sequence of wild type sequence and shown with the mutated residues in blue and conserved residues in red. In addition we have attempted to correlate and establish the digestion profile of the mutant and wild type Mtases by HPLC.

In the first experiment Phe in M.SPRI was altered to Ala to mimic motif IX of M.*Hha*I. Phe 386 of M.SPRI aligns with Ala 280 of M.*Hha*I and many other mono-specific Mtases and therefore we have chosen to mutate Phe to Ala. This single mutation led to a partial loss of M.SPRI activity. The principal difference between Phe/Ala is that the former side chain is much bigger than the methyl group of Ala (Figure 3.24).

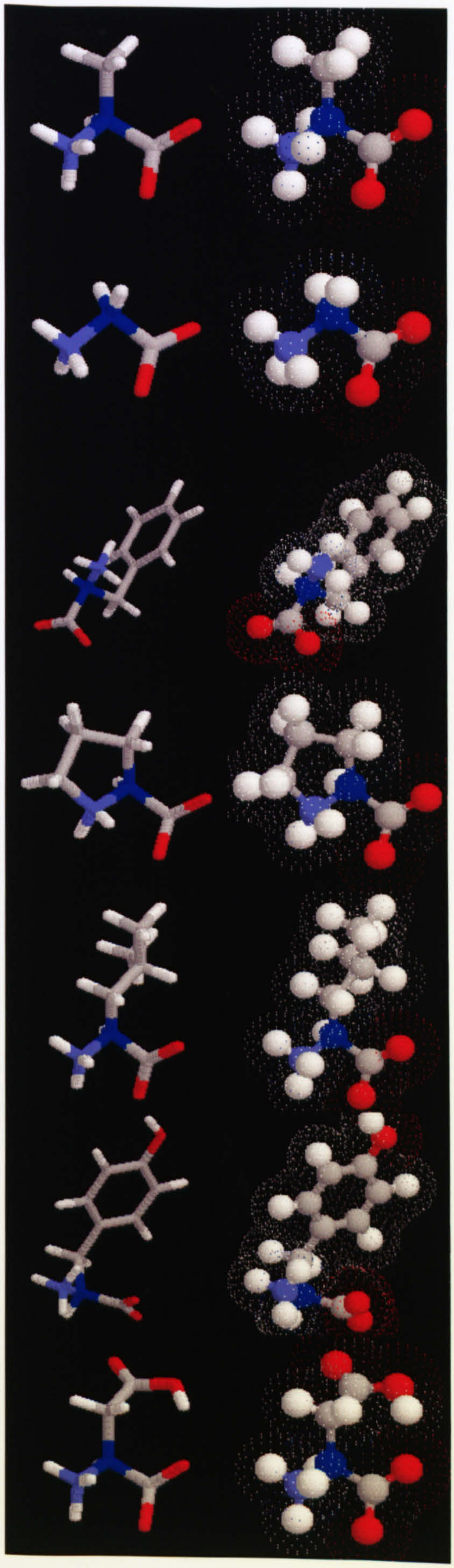
The crystal structure of M.SPRI has not yet been obtained and therefore we cannot predict with any certainty the interaction of any of the amino acid residues in this enzyme. However on the basis of information obtained from the crystal structure of M.*Hha*I (Cheng *et al.*, 1993), it is clear that Ala 280 of M.*Hha*I makes interactions with its neighbouring residues. Further analysis of these interactions revealed that Ala 280 in M.*Hha*I makes an interaction with residues 276-278 and 282-285 in the presence of DNA. In the absence of DNA, M.*Hha*I makes interactions with residues

Summary of Mutants of M.*Hha*I & M.*SPR* in Motif IX & X

Construct	Amino Acid Sequence	Mutation Position	Activity[a]
	264 270 280 290 300 310		
Wild Type M.*Hha*I	GYLVNGKTRKLHPRECARVMGYPDSYKVHPST SQAYKQFGNSVVINVL		+
pQIS 11 & 12	GYLVNGKTRKLHPRECFRVMGYPDSYKVHPST SQAYKQFGNSVVINVL	A280F	+/-
pQIS 13 & 14	GYLVNGKTRKLHPLECFRVMGYPDSYKVHPST SQAYKQFGNSVVINVL	R277L, A280F	-
pQIS 17 & 18	GYLVNGKTRKLHPLECFRVMGYPDSYKVHPST SQAYKQAGNSVVINVL	R277L, A280F, F302A	+
pQIS 20	GYLVNGKTRKLHPRECARVMGYPDSYKVHPST SQAYKQAGNSVVINVL	F302A	+/-
pQIS 21	GYLVNGKTRKLHPLECFRVMGYPDSDKVHPST SQAYKQAGNSVVINVL	R277L, A280F, Y289D, F302A	-
pQIS 22	GYLVNGKTRKLHPRECARVMGYPDSDKVHPST SQAYKQFGNSVVINVL	Y289D	-
	370 380 390 400 410 420		
Wild Type *M.SPR*	GEYPKYRIRRLTPLECFRLQAFDDEDFEKAFAAGISNSQLYKQAGNSITVTVL		+
pQIS 8	GEYPKYRIRRLTPLECARLQAFDDEDFEKAFAAGISNSQLYKQAGNSITVTVL	F386A	+/-
pQIS 9	GEYPKYRIRRLTPLECFRLQGFDDEDFEKAFAAGISNSQLYKQAGNSITVTVL	A390G	+
pQIS 10	GEYPKYRIRRLTPLECFRLQAFPDEDFEKAFAAGISNSQLYKQAGNSITVTVL	D392P	+

Figure 3.23: Summary of all the mutants constructed in motif IX and X of M.*Hha*I and M.SPR. Conserved residues in both regions are shown in red. The amino acids replaced for making mutants constructs are shown in blue. The number of amino acids are shown on the top of the amino acid sequence. The position of mutation and the activity of the encoded protein from final construct is shown on the panel shown at right hand side.

a: Activity level: Protected (+), Partial protection (+/-), No protection (-)



Alanine (88.6 Å^3)

Glycine (60.01 Å^3)

Phenylalanine (189.9 Å^3)

Proline (112.7 Å^3)

Leucine (166.7 Å^3)

Tyrosine (193.6 Å^3)

Aspartic acid (111.1 Å^3)

Figure 3.24: Comparison of amino acid sizes which were used for mutagenesis studies. Amino acids shown as Stick models on the left hand side and Ball and stick model with their Van der Waals radii on the right hand side. The residue volume are shown with amino acid names on the right panel: http://prowl.rockfeller.edu/aainfo/volume.htm.

276 and 283-285. It would appear that the replacement of Phe with a much smaller hydrophobic amino acid Ala leads to a loss in some of the steric interference, which restores methylation capacity of the enzyme.

The second mutation Ala 390 Gly, did not affect the activity of the enzyme at all, as determined by our assays. The larger size of the methyl side chain is therefore not a critical factor in the enzyme structure.

In the third mutagenic experiment (Asp 392 Pro) in M.SPRI, whilst the mass difference in the two amino acid residues is very small (Figure 3.24), Asp has an acidic side chain whereas Pro is hydrophobic and changes directionality of back bone (See Appendix III for amino acid chemistry) but it appears that this amino acid plays a less important role in terms of global structural interactions in the enzyme.

Despite the impressive ability of natural enzymes to catalyse a diverse set of reactions, attempts to make even minor modification to substrate specificity and to improve catalytic efficiency have proven to be difficult. The function of enzymes is not only dependent on the chemical properties of the side chains of the amino acid residues that contact substrates and cofactor; it is also influenced by residues distant from the active site. Such residues may play a critical role in holding the catalytic residues in their required orientations. Charge distribution throughout a whole enzyme molecule may facilitate substrate binding by electro statically guiding the substrate into the active site. A common example for selective substrate guidance and to catalyze the hydrolysis of proteins is the family of serine proteases that are found in prokaryotes and Eukaryotes. Despite the fact that proteins folds are quite different, the geometric orientations of these are closely similar between families. Thus the peptidases of the chymotrypsin, subtilisin and carboxypeptidase C clan have a common "catalytic triad" of the three amino acids: Serine (nucleophile), aspartate (electrophile), and histidine (base). Chymotrypsin (mammalian digestive enzyme) is found in the small intestine and has a similar catalytic site (Stryer, 1995). In this enzyme all charged groups are present on the surface of the molecule except for the three that play a critical role in catalysis. Serine 195 is adjacent to histidine 57 and carboxylate group of aspartate 102, and these residues are buried in the protein and form a catalytic triad. Crystallographic studies of complexes of chymotrypsin with substrate analogs have shown the location of the site of specific recognition and the likely orientation of the susceptible peptide bond. Tryptophan binds to chymotrypsin with its indole side chain fitted neatly into a pocket near serine. This deep cleft

accounts for the specificity of chymotrypsin for aromatic and other bulky hydrophobic side chains. Therefore, it has been difficult to demonstrate the importance of residues that are remote from the active site because the number of these residues is large and the contribution of each residue is difficult to detect individually. In general, it can be concluded from these experiments that there is some flexibility in motif IX of M.SPRI and it can tolerate the mutations to amino acids residues present at the same position as in mono-specific Mtase M.*Hha*I.

The interaction of the TRDs of the variable region in multi-specific Mtases with DNA can be assumed to be similar to those described for monospecific Mtases. However the unconventional placement of a functional TRD at the N-terminus of M.(ϕ)*Bss*HII shows that the interactions between large and small domains and the target DNA must be different in this case (Sethmann *et al.*, 1999). It looks that a different steric relationship between TRD, DNA and other enzyme regions involved in the methylation process must exists in M.(ϕ)*Bss*HII. In view of the established three-dimensional structures of Mtases folded into a large (catalytic) and small (recognition) domain, it is conceivable that the N-terminal TRD becomes placed at the join between these domains in the neighbourhood of the variable region. It is also proposed that specific DNA recognition may occur in a similar way as with the TRDs of the variable region (Sethmann *et al.*, 1999). This supports the idea that the interactions in multi-specific enzymes can be more flexible than in mono-specific Mtases.

In the light of the experiments made in multi-specific Mtase M.SPRI, experiments have been designed to mutate mono-specific Mtase M.*Hha*I. The sequence of M.*Hha*I was mutated from Ala 280 to Phe (as was found in M.SPRI). In the crystal structure of mono-specific C5-Mtases M.*Hha*I (Cheng *et al.*, 1993; Kumar *et al.*, 1994) and M.*Hae*III (Reinisch *et al.*, 1995) motifs I to VIII and X form the large domain while the variable region plus motif IX comprise the small domain (Figure 3.25). This mutagenesis made the enzyme partially active which upon analysis revealed that this residue Ala 280 when replaced with a larger amino acid (with bulky side chain) Phe, sterically hindered neighbouring residues (Figure 3.26).

Considering Arg 277 as being disturbed with earlier mutation was mutated into Leu. This mutation further worsens the situation and the enzyme became inactive. However, when another possible interacting residue Phe 302 was mutated into Ala, it

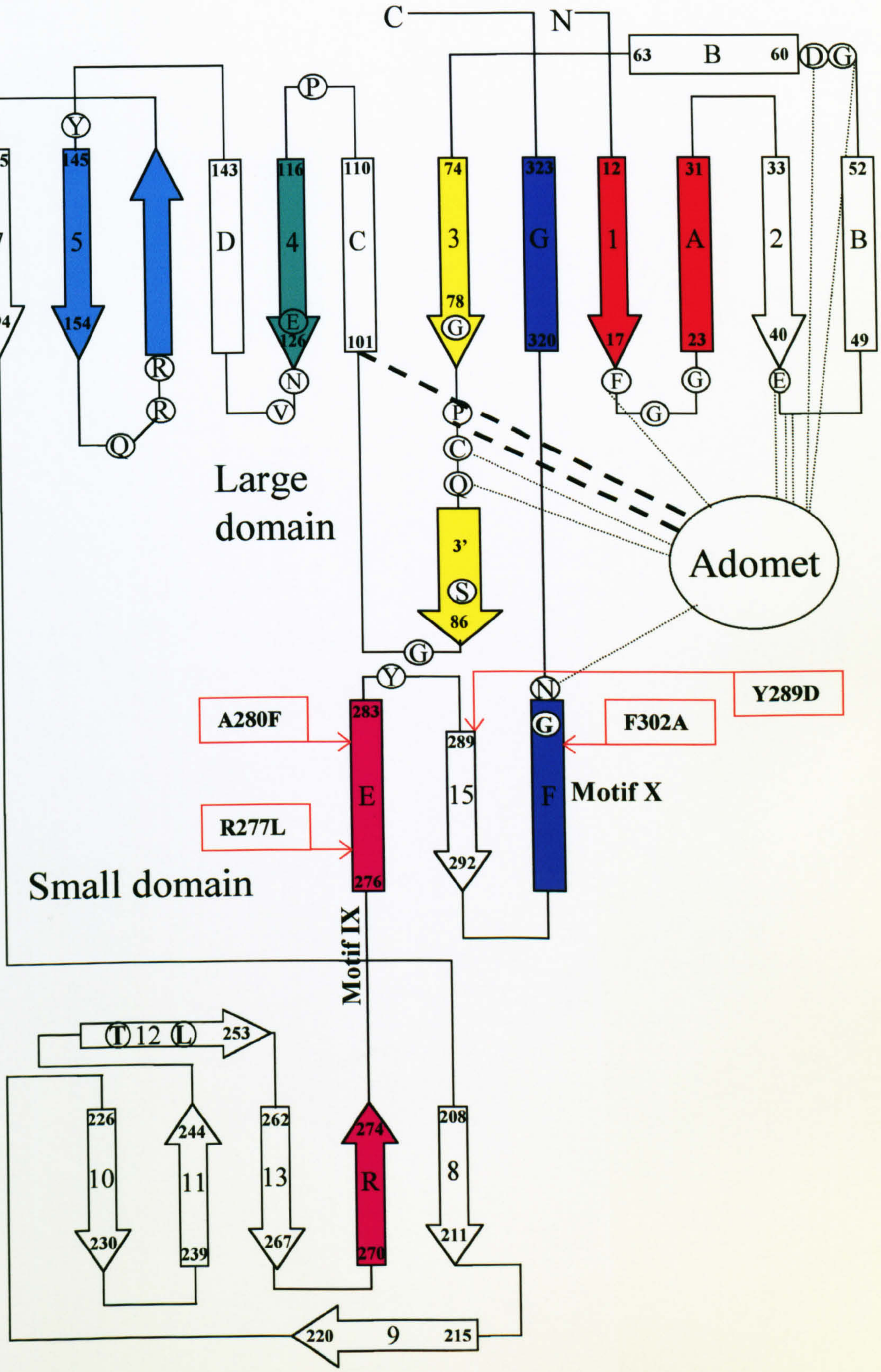


Figure 3.25: Schematic drawing of the secondary structural elements of M.*Hha*I. The large (amino acids 1-93 and amino acids 304-327) and small (amino acids 194-275) domains are marked. Conserved or functionally important amino acids are circled. The cofactor S-adenosyl-L-methionine (SAM) binding site, with its interactions with the protein, is also shown. The thick dashed lines indicate hydrophobic interactions, and the thin lines indicate electrostatic interactions. The red boxes represent the amino acid residues that were changed in the experiment. The colours of various motifs are as shown in figure 1.5 (Figure adapted from Cheng *et al.*, 1993).

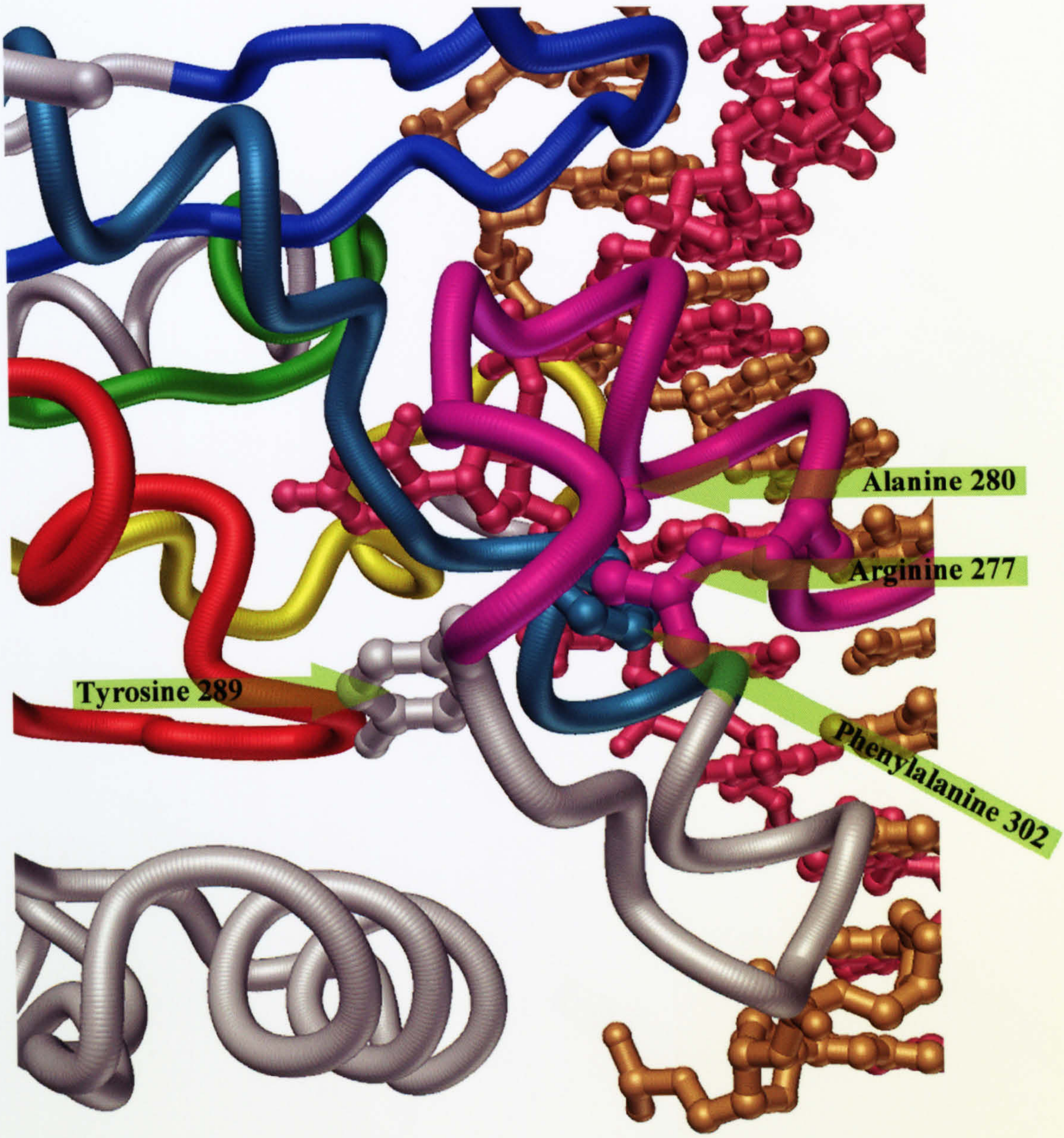


Figure 3.26: Tertiary structure of Motif IX of M.*Hha*I. The protein backbone is shown as a worm plot. Colours are shown as in secondary structure of M.*Hha*I (motif I red; motif IV, yellow; motif IX, magenta; and motif X, blue). The DNA is shown as a "ball and stick" representation. The side chains of four amino acids which were targeted for mutations are also shown as a Ball and Stick representation.

restored the activity of the enzyme. We have presented a model for this triple mutation (Figure 3.27). From the model it appears that the two helices that are parallel to each other and can only accommodate amino acids that do not disfigure its orientation. Therefore it is only possible to swap amino acids from motif IX from a multi-specific Mtase to mono-specific Mtase and redesign M.*Hha*I provided that if the structure is not distorted. However, further sequence exchanges are incompatible between mono-specific and multi-specific Mtases. To conclude from this set of experiments I would like to add that these enzymes are clearly superficially related but mechanistically/structurally different. A 3D-structure analysis of M.SPRI should help to explain these differences in structure / function relationship.

The analysis of active, partial active and inactive Mtase by HPLC (Figure 3.22) have also given a new way of profiling the methylation potential which we have used in our future experiments (See Chapter Six). We believe that this technique is more reliable and efficient as well as more sensitive than the agarose gel electrophoresis.

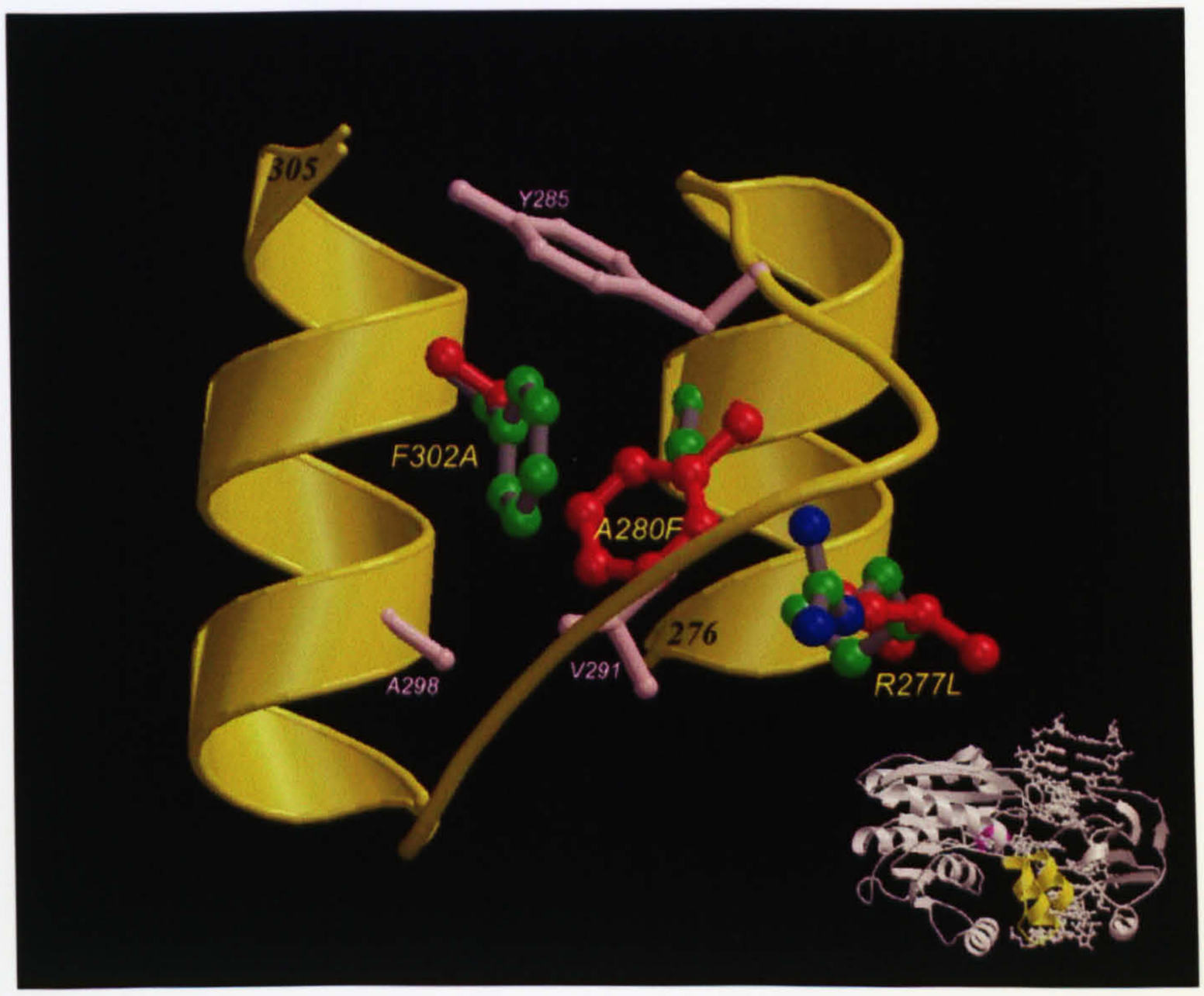


Figure 3.27: A model for the Triple mutant structure of M.*Hha*I (A280F, R277L & F302A). The wild type M.*Hha*I protein is shown as yellow ribbon. The residues that were altered are shown as "ball and stick" representation with CPK colour atoms. The three mutant side chains are shown superimposed on this structure in red ball and stick representation. Other side chains within 4 Å in the wild type molecule are shown in pink wire frame. The whole molecule is shown in the bottom right with residues 376-305 coloured yellow in order to locate this region in the overall structure. The mutant image was created by Dr. Matthew J. Conroy, using Insight (MSI inc.) and subsequently energy minimised in XPLOR (Brünger 1992 XPLOR version 3.1:A system for crystallography and NMR. Yale University, New Haven). The figure was produced using Molscript (Kraulis (1991) *J. Appl. Cryst.* **24** 946-950) and Raster3D, (Merritt & Murphy (1994) *Acta Cryst.* D50 869-873).

Schematic diagrams showing the plasmid sketches with gel pictures

(Including Figures 3.6, 3.13 ~ 3.15, 3.17 ~ 3.19)

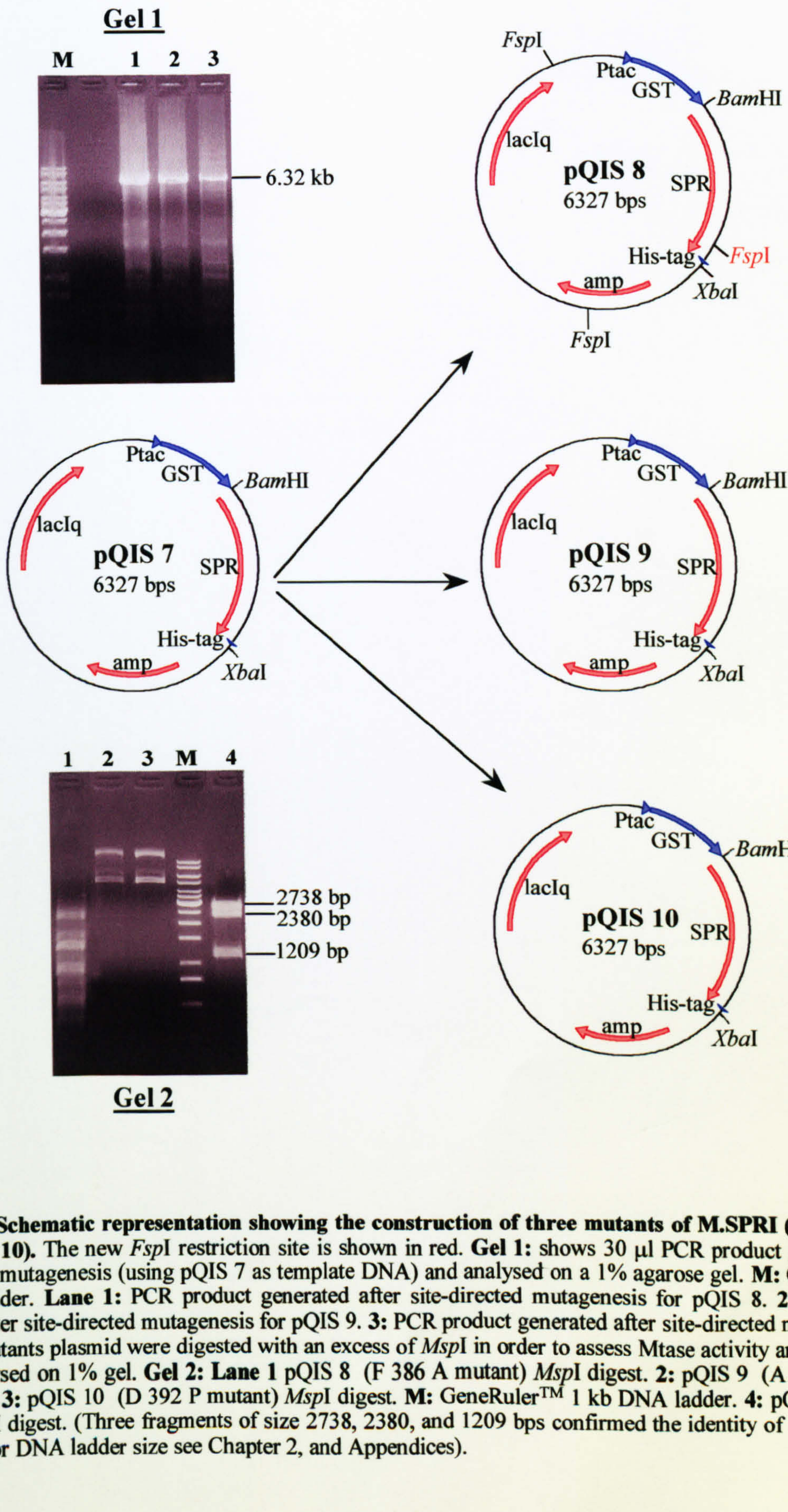


Figure 3.6: Schematic representation showing the construction of three mutants of M.SPRI (pQIS 8, pQIS 9 and pQIS 10). The new *Fsp*I restriction site is shown in red. **Gel 1:** shows 30 µl PCR product generated after site-directed mutagenesis (using pQIS 7 as template DNA) and analysed on a 1% agarose gel. **M:** GeneRuler™ 1 kb DNA ladder. **Lane 1:** PCR product generated after site-directed mutagenesis for pQIS 8. **2:** PCR product generated after site-directed mutagenesis for pQIS 9. **3:** PCR product generated after site-directed mutagenesis for pQIS 10. Mutants plasmid were digested with an excess of *Msp*I in order to assess Mtase activity and the resulting digests analysed on 1% gel. **Gel 2: Lane 1** pQIS 8 (F 386 A mutant) *Msp*I digest. **2:** pQIS 9 (A 390 G mutant) *Msp*I digest. **3:** pQIS 10 (D 392 P mutant) *Msp*I digest. **M:** GeneRuler™ 1 kb DNA ladder. **4:** pQIS 8 (F 386 A mutant) *Fsp*I digest. (Three fragments of size 2738, 2380, and 1209 bps confirmed the identity of construct pQIS 8). (**Note**: For DNA ladder size see Chapter 2, and Appendices).

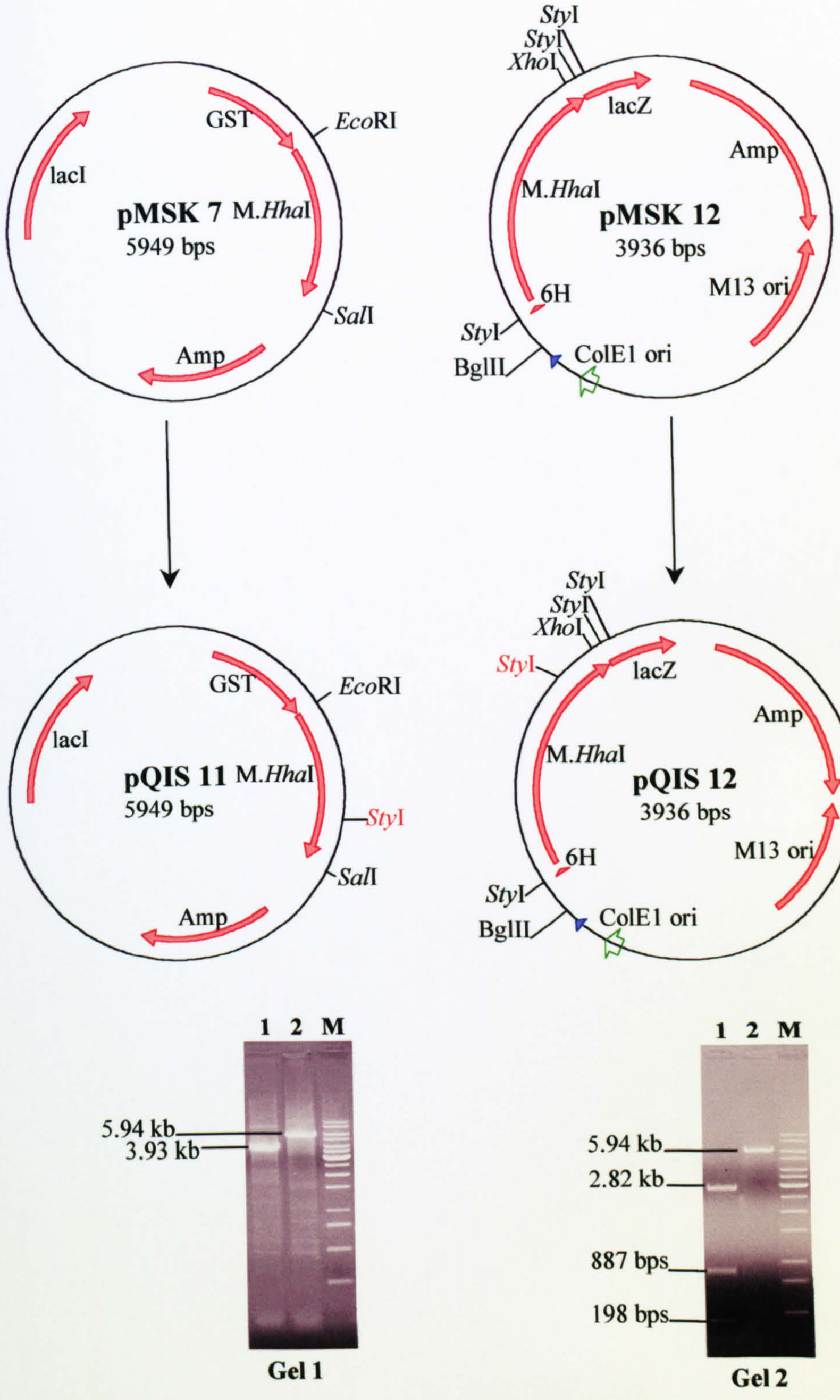


Figure 3.13: Schematic representation showing the construction of pQIS 11 and pQIS 12 (containing the gene encoding the Ala 280 Phe mutant of M.*Hha*I). The new restriction site *Sty*I is shown in red. **Gel 1: Lane 1** PCR product generated during site directed mutagenesis from pMSK 12. **Lane 2** PCR product generated during site directed mutagenesis from pMSK 7. **Gel 2:** Digestion analysis of products shown using *Sty*I restriction enzyme. **Lane 1**, pQIS 12 cut with *Sty*I . (Fragment sizes 2824, 887, 198 and 27 bp). **Lane 2:** pQIS 11 cut with StyI (linear DNA, 5949 bp). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

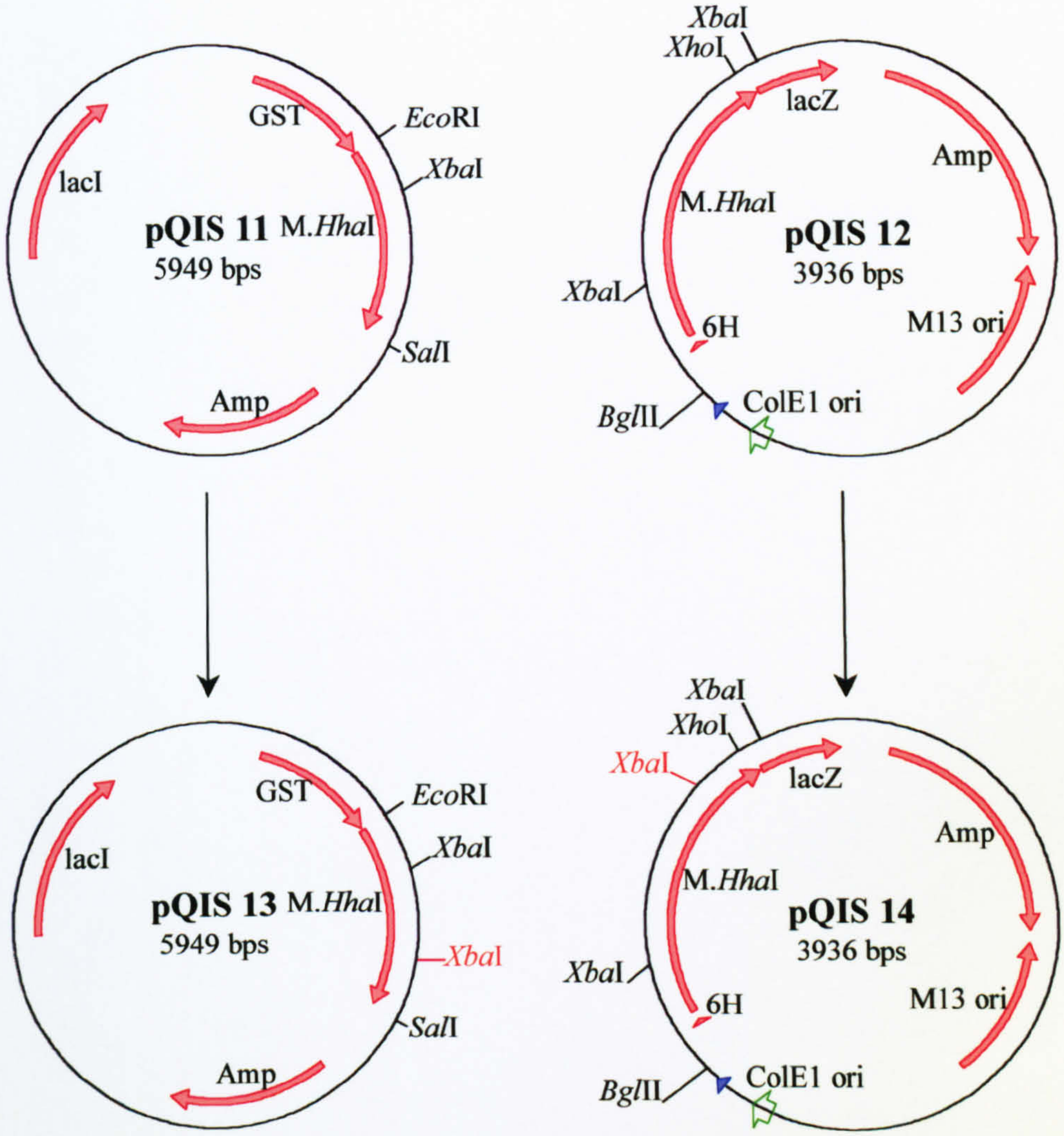


Figure 3.14: Schematic representation of the construction of double mutants (A280F, A277L) pQIS 13 and pQIS 14 from pQIS 11 and pQIS 12. New restriction site of *Xba*I shown in red. Mutants were made by site directed mutagenesis (SDM) (**See Gel 1**). **1:** PCR product generated during site directed mutagenesis for pQIS 13 **2:** PCR product generated during site directed mutagenesis for pQIS 14. **Gel 2: Lane 1:** pQIS 13 double mutant (A280F, A277L) *Hha*I digest. (Cleaved DNA, thus encoded enzyme was inactive). **2:** pQIS 13 double mutant (A280F, A277L) *Xba*I digest.(Fragment sizes 542 and 5407 bps) **3:** pQIS 14 double mutant (A280F, A277L) *Hha*I digest. **4:** pQIS 14 double mutant (A280F, A277L) *Xba*I digest. (Fragment sizes 215, 542 and 3179 bps) **M:** GeneRuler™ 1 kb DNA ladder in both gels.

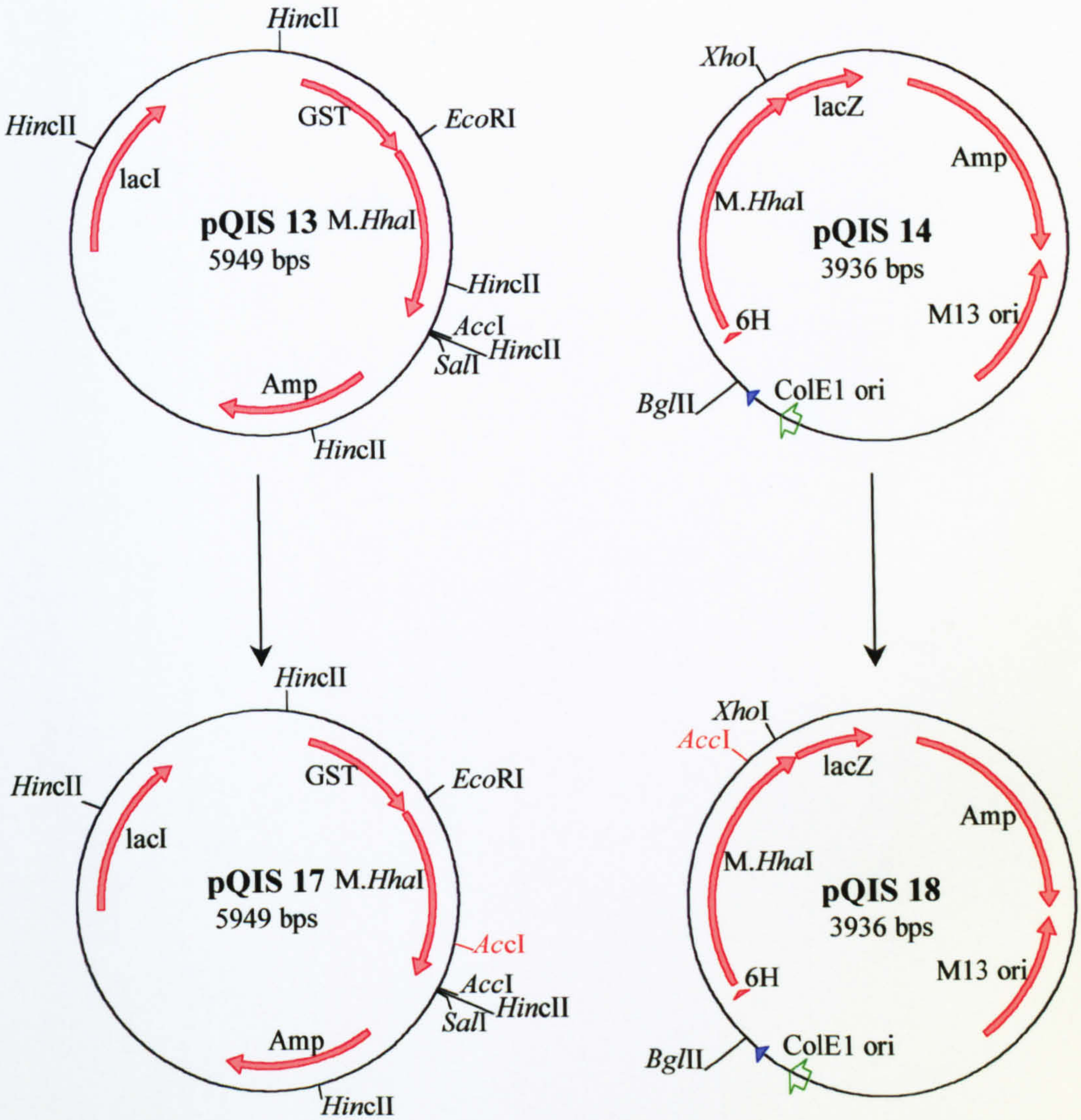


Figure 3.15: Schematic representation of the construction of triple mutants (A280F, A277L & F302A) pQIS 17 and pQIS 18 from pQIS 13 and pIS14. New restriction site of *Acc*I shown in red in plasmid constructs pQIS 17 and pQIS 18. Mutants were made by site directed mutagenesis (SDM). **Gel 1: Lane 1** and **2;** 30 µl PCR product after SDM for both new constructs. pQIS 17 was confirmed by the disappearance of *Hinc*II site. **Gel 2 (1.5% agarose gel): Lane M 1:** 100 bp DNA ladder. **Lane 2:** pQIS 17 cleaved with *Hinc*II (2552, 1747, 979, 671 bp). pQIS 18 was digested with *Acc*I for the presence of unique restriction site. **Gel 3: Lane 1:** pQIS 18 digested with an excess of *Hha*I for its activity. (Uncut DNA reveals that the plasmid encodes for an active Mtase *Hha*I). **Lane 2:** pQIS 18 cleaved with *Acc*I (linear 3936 bp). **M:** GeneRuler™ 1 kb DNA ladder in all three gels. (For exact sizes of standard 100 bp and 1 kb DNA ladder please see Chapter 2 and Appendices).

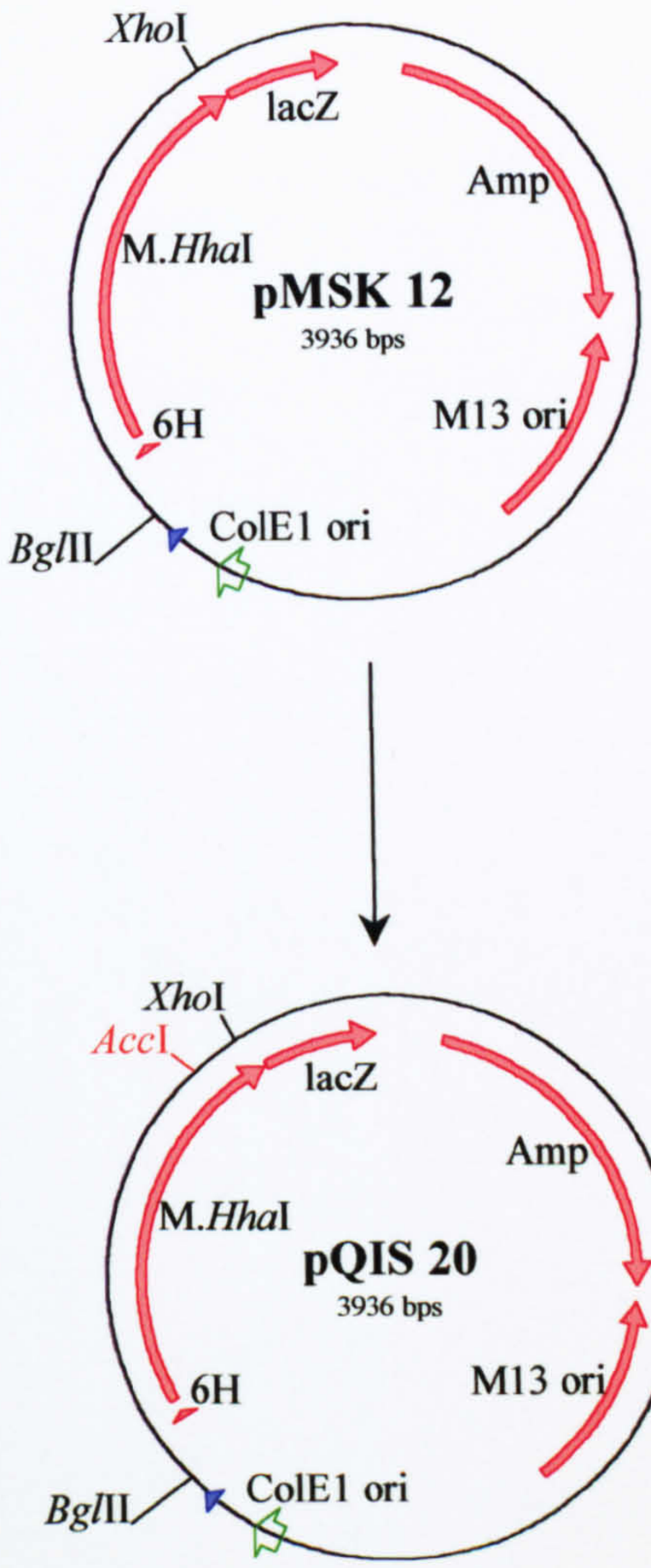
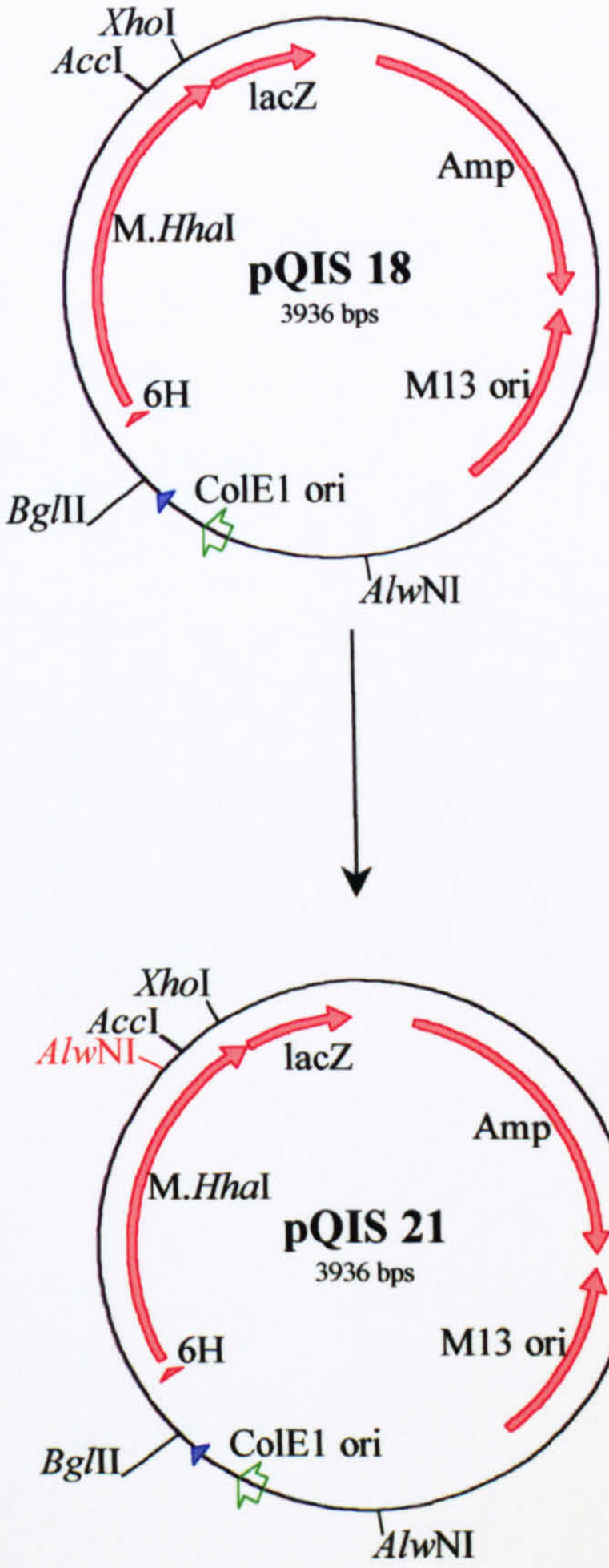


Figure 3.17: Schematic representation of the construction of pQIS 20 (single mutant Phe 302 Ala) using pMSK 12 (which encodes for wild type M.*Hha*I). Substitution of Phe 302 into Ala in motif IX by site directed mutagenesis. The new restriction site of *Acc*I shown in red in the plasmid pQIS 20 map. **Gel 1: Lane 1** PCR product generated during site directed mutagenesis using pMSK 12 as template. **Gel 2: lane 1** represents the construct pQIS 20 digested with *Acc*I (linear 3936 bp). This digestion verifies the formation of pQIS 20. **Gel 3 lane 1,** pQIS 20 digested with *Hha*I. (Partial activity is shown by the protein encoded by pQIS 20) **M:** GeneRuler™ 1 kb DNA ladder in all gels.



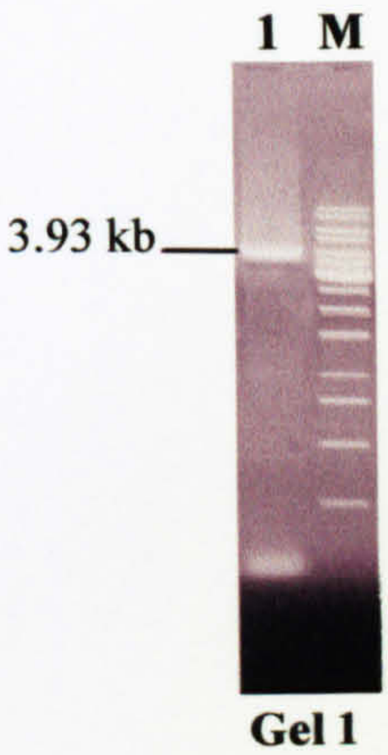


Figure 3.18: Schematic representation of the construction of quad mutant (Ala 280 Phe, Arg 277 Leu, Phe 302 Ala, & Try 289 Asp) pQIS 21 from active triple mutant (Ala 280 Phe, Arg 277 Leu & Phe 302 Ala) pQIS 18 . In plasmid map new restriction site of *Alw*NI is shown in red. **Gel 1: lane 1:** shows PCR product generated during site directed mutagenesis using pQIS 18 as template. **Gel 2: Lane 1:** pQIS 21 cleaved with *Alw*NI (2342, 1594 bp) and verifies the formation of pQIS 21. **M:** GeneRuler™ 1 kb DNA ladder in both gels.

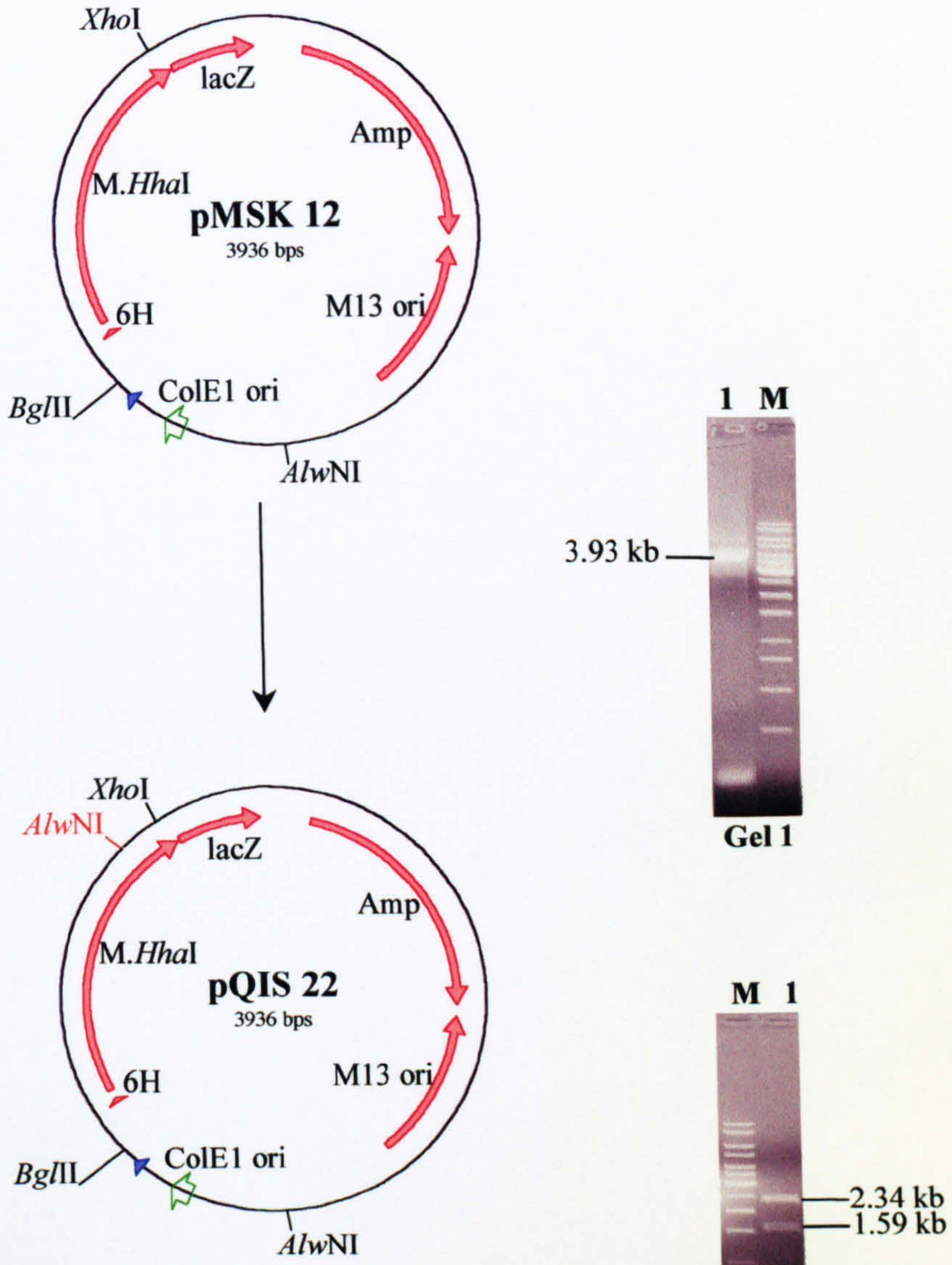


Figure 3.19: Schematic representation of the construction of single mutant pQIS 22 (Tyr 289 Asp) from wild type M.*Hha*I (pMSK 12). Substitution of Tyr 289 into Asp in motif IX introduces new restriction site of *Alw*NI shown in red in the plasmid map. **Gel 1: lane 1:** PCR product generated during site directed mutagenesis using pMSK 12 as template. **Gel 2: Lane 1:** pQIS 22 cleaved with *Alw*NI (2342, 1594 bp). This digestion verifies the formation of pQIS 22. **M:** GeneRuler™ 1 kb DNA ladder in both gels.

CHAPTER FOUR

CONSTRUCTION OF CHIMAERAS OF M.*Hha*I BY INSERTION MUTAGENESIS

CHAPTER FOUR

CONSTRUCTION OF CHIMAERAS OF M.*Hha*I BY INSERTION MUTAGENESIS

4.1: Introduction

It is known from the crystal structure of M.*Hha*I that the variable region (amino acids 171-271) begins as a long stalk on the surface of the protein (Cheng *et al.*, 1993). Many experiments have been carried out to explore the structural flexibility of methyltransferases (Mtases) (Wilke *et al.*, 1988; Trautner *et al.*, 1988; Walter *et al.*, 1992; Trautner *et al.*, 1996; Matin, 2000) (see also Chapter 3, section 3.1.2.) most of which focus on the TRD region. Experiments concerning the exchange of TRDs regions among multispecific Mtases have shown that chimaeric constructs can be made successfully (Balganesh *et al.*, 1987; Trautner *et al.*, 1988; Walter *et al.*, 1992; Trautner *et al.*, 1996). These experiments showed the interchange-ability of the core structures of different phage enzymes. However, the only case in which the exchange of TRDs between the mono and multispecific Mtases has been productive is in the chimaeric constructs involving M.(ϕ)BssHII and the highly related monospecific Mtase M.ϕ3TII (Sethmann *et al.*, 1999). In such a situation one should be able to delete or add target recognising modules in the variable region, causing either loss or the acquisition of novel methylation capacities. In support of this expectation, an experiment has been planned to insert various segments from the multispecific Mtase M.SPRI into the monospecific M.*Hha*I that has been modified as described in chapter three.

The appropriate interaction between the large and small domains is essential for C5-Mtases to catalyse sequence specific methylation (Sethmann *et al.*, 1999). In particular, complementary interactions between the variable region, TRD and motifs IX and X appear to be crucial for optimal enzymatic activity. It would seem therefore possible that the enzyme could tolerate insertions only if the extra polypeptide is placed on the surface of the protein structure without disturbing the "core" interactions.

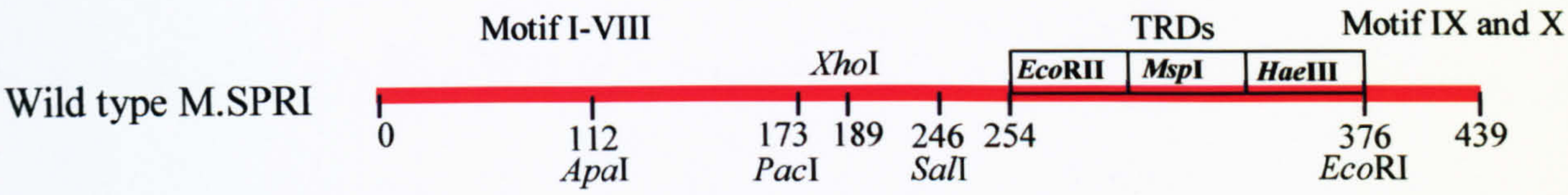
4.2: Construction of a Chimaera in which the TRD region of M.*Hha*I has been exchanged with segment from M.SPRI

Previous work has shown that specificity determinants of the mono-specific Mtases are located within the variable region (see Chapter 1). In addition a structural correlation exists between conserved motif IX and the variable region (Klimašauskas *et al.*, 1991; Mi and Roberts, 1992). It seems that hybrid Mtases retain greater activity when the variable region and the flanking motif IX derive from the same parent enzyme. These results suggest that there may be some direct structural interactions between the variable region and motif IX (Mi and Roberts, 1992) that play a role in catalysis.

In order to study the degree of tolerance to structural changes and to explore the relationship between the variable region and motif IX of M.*Hha*I, I have carried out a series of TRD insertion and exchange experiments. As discussed earlier, in the light of the successful TRD swapping experiments between multi-specific Mtases, it was hypothesised that the triple mutant plasmid pQIS 18 (Ala 280 Phe, Arg 277 Leu and Phe 302 Ala) and the quadruple mutant plasmid pQIS 21 (Ala 280 Phe, Arg 277 Leu, Phe 302 Ala and Tyr 289 Asp) might provide more suitable frameworks for TRD exchanges between M.*Hha*I and one of the TRDs from M.SPRI.

The wild type M.*Hha*I gene has been used as a "host" for the *Msp*I specific TRD of M.SPRI: the resulting construct (pHhaI-TRDM) was catalytically inactive (Matin, 2000). In this experiment the region encoding the *Msp*I specific TRD was inserted as an *Xho*I-*Eco*RI fragment into wild type M.*Hha*I (Matin, 2000). In the following set of experiments mutants of M.*Hha*I (described in Chapter 3), which mimic motif IX of the multi-specific Mtase M.SPRI, were chosen for insertion of those segments from M.SPRI shown in Figure 4.1.

During the course of insertion of segments from M.SPRI it was vital to introduce certain restriction sites both in M.SPRI and M.*Hha*I to facilitate the cloning and for the formation of chimaeras. The mutants prepared in this experiment are summarised in figure 4.1 (a) and 4.1 (b). The schematic diagrams showing the plasmid sketches with gel pictures are placed at the end of the chapter.



Various segments of M.SPRI which were used for insertion into M.*Hha*I in the TRDs encoding region.

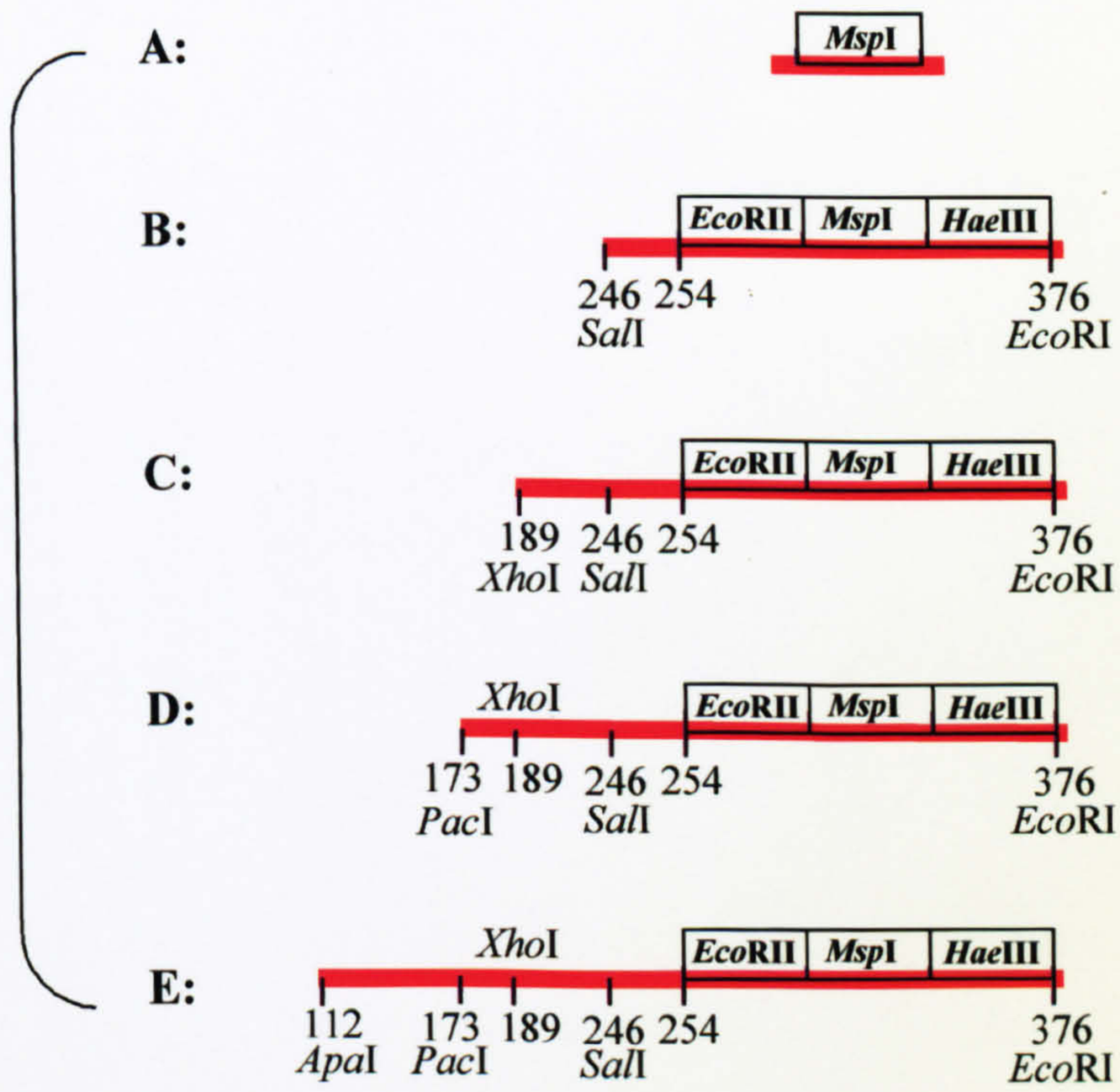


Figure 4.1: The approximate size and position of segments from the variable region of wild type M.SPRI which were used for cloning into the triple mutant of M.*Hha*I. Wild type M.SPRI and segments from M.SPRI are shown in red. The numbers are the coordinates of amino acids in M.SPRI with unique restriction sites. (Some restriction sites were not present in the wild type sequence and were generated later, see text for details). The TRD region are described in boxes. Map is not drawn to scale.

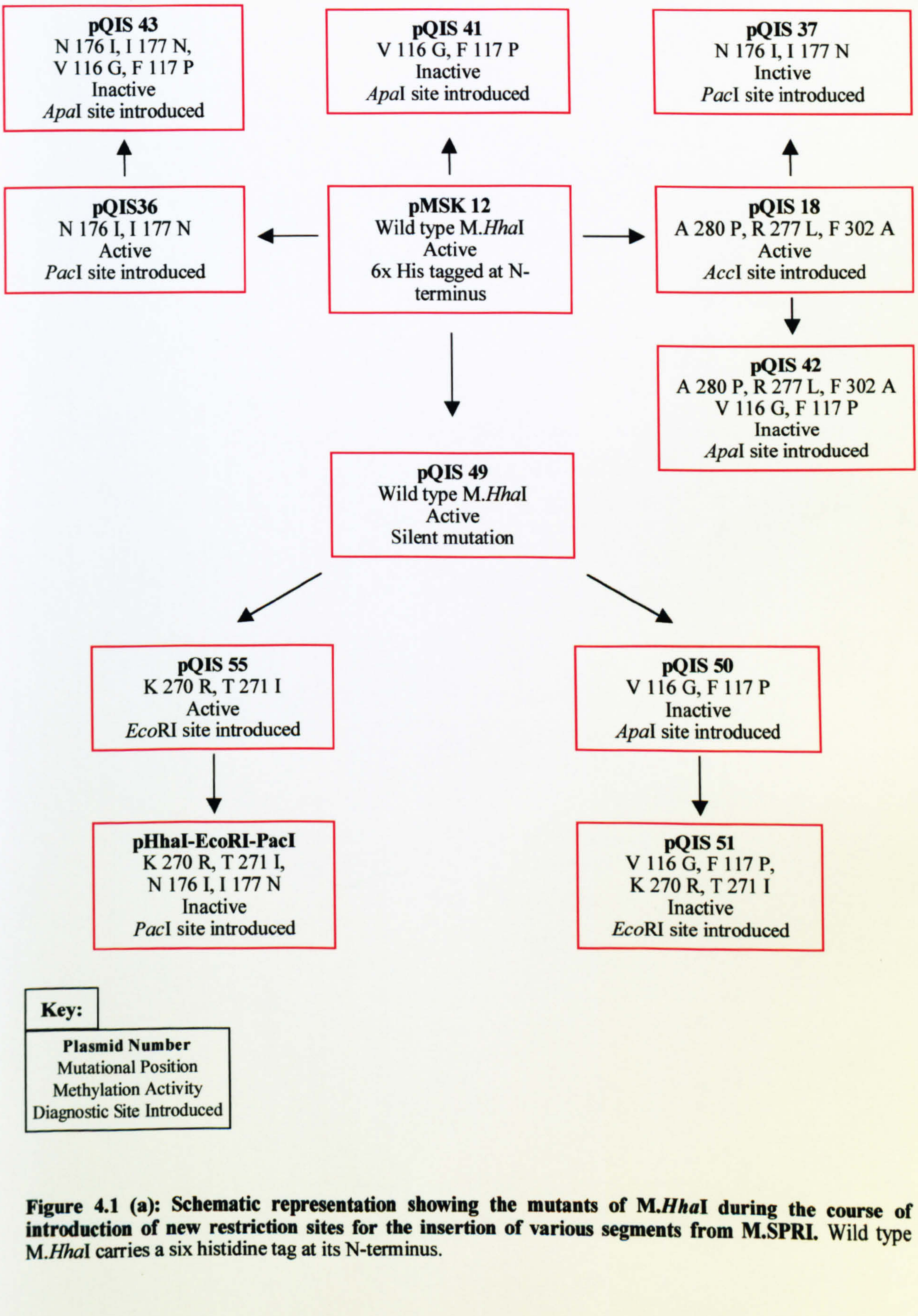


Figure 4.1 (a): Schematic representation showing the mutants of M.*Hha*I during the course of introduction of new restriction sites for the insertion of various segments from M.SPRI. Wild type M.*Hha*I carries a six histidine tag at its N-terminus.

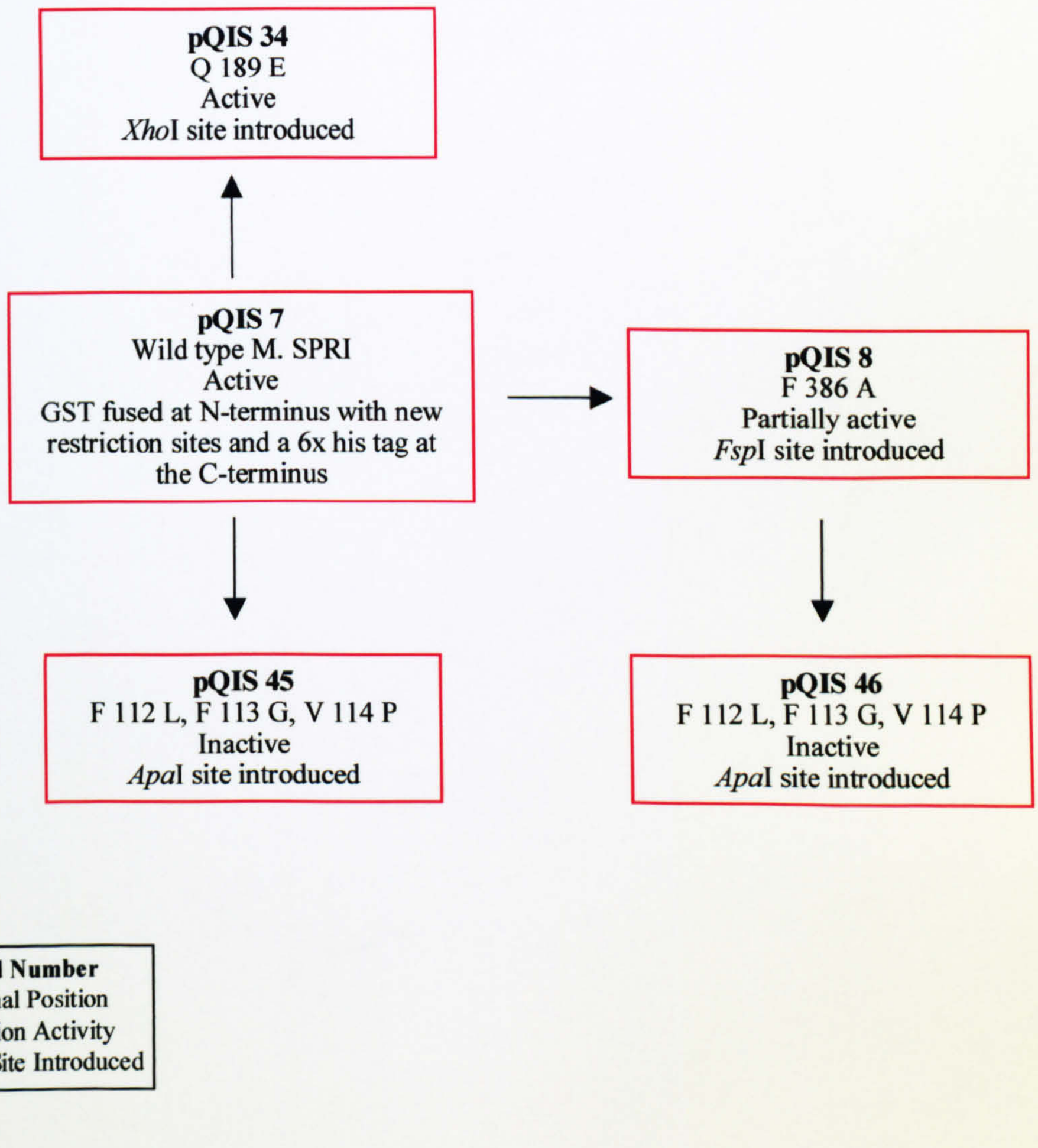


Figure 4.1 (b): Schematic representation showing the mutants of M.SPRI during the course of introduction of new restriction sites for the insertion of various segments from M.SPRI into M.*Hha*I.

4.2.1: INSERTION OF THE *Msp*I SPECIFIC TRD FROM M.SPRI INTO THE M.*Hha*I TRIPLE MUTANT

In the first instance plasmid pQIS 18, carrying three mutations (Ala 280 Phe, Arg 277 Leu and Phe 302 Ala) in M.*Hha*I was selected for the insertion of an *Msp*I specific TRD coding sequence (Segment A, see Figure 4.1) from M.SPRI. It was convenient to use the pHhaI-TRDM construct (see above) for swapping the entire TRD region, which carries both TRDs (*Msp*I and *Hha*I) arranged sequentially.

The enzymes chosen for the cloning of the desired region of DNA from p*Hha*I-TRDM to pQIS 18 were *Bcl*I and *Bbs*I. *Bcl*I requires the DNA to be prepared in a Dam^- strain of *E. coli*. The enzymes *Bcl*I and *Bbs*I utilize the same buffer but require a different temperature for incubation. Therefore the DNA was prepared using the *E. coli* strain GM2163 and was then subjected to sequential restriction digestion with *Bcl*I and *Bbs*I. The experiment is summarised in Figure 4.2. The final construct was called pQIS 23 (Triple mutant + *Msp*I TRD) and was identified by restriction analysis. The appearance of a new *Bgl*II site confirmed that the new constructs carried the *Msp*I TRD, along with the *Hha*I TRD region (Figure 4.2). The sequence of M.*Hha*I DNA showing the deduced amino acid sequence before and after the *Msp*I TRD insertion is shown in Figure 4.3. The activity of the mutant enzyme encoded by the pQIS 23 was evaluated by *Hha*I and *Msp*I restriction analysis; it was found to be inactive (Figure 4.2).

4.2.2: INSERTION OF THE *Msp*I SPECIFIC TRD FROM M.SPRI INTO THE M. *Hha*I QUADRUPLE MUTANT

The quadruple mutant of M.*Hha*I pQIS 21 carrying four mutations (Ala 280 Phe, Arg 277 Leu, Phe 302 Ala and Tyr 289 Asp) was also chosen for an experiment involving the insertion of the *Msp*I TRD coding sequence (Segment A, see Figure 4.1) from the M.SPRI gene. The same strategy was used as was carried out for the triple mutant pQIS 18 (see above). *Bcl*I and *Bbs*I were used to create a 3332 bp (vector) DNA from pQIS 21, and a 733 bp (insert) DNA from p*Hha*I-TRDM. Recombinant plasmids were screened to identify pQIS 24 (quadruple mutant + *Msp*I TRD). The starting plasmids pQIS 21 and pHhaI-TRDM both have a unique *Bgl*II site, as shown in Figure 4.4. In addition, the inserted *Msp*I specific TRD region also contains *Xho*I and *Eco*RI sites. Restriction analysis with *Xho*I, *Eco*RI and *Bgl*II confirmed the authenticity of the new plasmid pQIS 24. *Msp*I and *Hha*I restriction digestion revealed that the new construct pQIS 24 was enzymatically inactive (Figure 4.4).

A

```
     AAA GAT TTG TTA TTA CCT GAT AGC GAG GTG GAA CAC TTA GTT ATT GAT
193   K   D   L   L   L   P   D   S   E   V   E   H   L   V   I   D
     AGA AAA GAT TTG GTA ATG ACA AAC CAA GAA ATT GAG CAA ACA ACC CCC
209   R   K   D   L   V   M   T   N   Q   E   I   E   Q   T   T   P
     AAA ACA GTT CGA CTT GGT ATT GTA GGA AAA GGT GGG CAA GGA GAA CGA
225   K   T   V   R   L   G   I   V   G   K   G   G   Q   G   E   R
     ATT TAT AGC ACA AGA GGC ATT GCA ATT ACC TTA TCT GCT TAT GGT GGC
241   I   Y   S   T   R   G   I   A   I   T   L   S   A   Y   G   G
     GGC ATT TTC GCT AAG ACA GGG GGA TAT TTA GTA AAC GGG AAG ACA CGG
257   G   I   F   A   K   T   G   G   Y   L   V   N   G   K   T   R
     AAA TTA CAC CCT AGA GAG TGT GCT AGA GTA ATG GGC TAC CCA GAT AGT
273   K   L   H   P   R   E   C   A   R   V   M   G   Y   P   D   S
     TAT AAA GTC CAC CCG TCA ACC AGC CAA GCA TAT AAA CAA TTT GGT AAC
289   Y   K   V   H   P   S   T   S   Q   A   Y   K   Q   F   G   N
     TCA GTT GTT ATC AAT GTA CTT CAA TAT ATT GCT TAT AAC ATT GGT TCA
305   S   V   V   I   N   V   L   Q   Y   I   A   Y   N   I   G   S
```

B

```
     GTT GAG TAC TCC AGA AAA AGC GGG CTT GGA CGA GAA CTT GCT GTA TCT
       V   E   Y   S   R   K   S   G   L   G   R   E   L   A   V   S
     CAT ACG CTT TCT GCT TCT GAC TGG AGA GGA TTG AAT AGG AAC CAA AAA
       H   T   L   S   A   S   D   W   R   G   L   N   R   N   Q   K
     CAA AAT GCA GTT GTT GAG
       Q   N   A   V   V   E
```

C

```
     AAA GAT TTG TTA TTA CCT GAT AGC CTC GAG GTT GAG TAC TCC AGA AAA
       K   D   L   L   L   P   D   S   L   E   V   E   Y   S   R   K
     AGC GGG CTT GGA CGA GAA CTT GCT GTA TCT CAT ACG CTT TCT GCT TCT
       S   G   L   G   R   E   L   A   V   S   H   T   L   S   A   S
     GAC TGG AGA GGA TTG AAT AGG AAC CAA AAA CAA AAT GCA GTT GTT GAG
       D   W   R   G   L   N   R   N   Q   K   Q   N   A   V   V   E
     CAT ATG GAA TTC GTG GAA CAC TTA GTT ATT GAT AGA AAA GAT TTG GTA
       H   M   E   F   V   E   H   L   V   I   D   R   K   D   L   V
     ATG ACA AAC CAA GAA ATT GAG CAA ACA ACC CCC AAA ACA GTT CGA CTT
       M   T   N   Q   E   I   E   Q   T   T   P   K   T   V   R   L
     GGT ATT GTA GGA AAA GGT GGG CAA GGA GAA CGA ATT TAT AGC ACA AGA
       G   I   V   G   K   G   G   Q   G   E   R   I   Y   S   T   R
     GGC ATT GCA ATT ACC TTA TCT GCT TAT GGT GGC GGC ATT TTC GCT AAG
       G   I   A   I   T   L   S   A   Y   G   G   G   I   F   A   K
     ACA GGG GGA TAT TTA GTA AAC GGG AAG ACA CGG AAA TTA CAC CCT CTA
       T   G   G   Y   L   V   N   G   K   T   R   K   L   H   P   L
     GAG TGT TTT AGA GTA ATG GGC TAC CCA GAT AGT TAT AAA GTC CAC CCG
       E   C   F   R   V   M   G   Y   P   D   S   Y   K   V   H   P
     TCA ACC AGC CAA GCA TAT AAA CAA GCT GGT AAC TCA GTT GTT ATC AAT
       S   T   S   Q   A   Y   K   Q   A   G   N   S   V   V   I   N
     GTA CTT CAA TAT ATT GCT TAT AAC ATT GGT TCA
       V   L   Q   Y   I   A   Y   N   I   G   S
```

Figure 4.3: Nucleotide and amino acid sequences of: A) "Triple mutant" of M.*Hha*I, showing the variable region, as well as motif IX (purple) and motif X (blue), **B)** M.SPRI *Msp*I specific target recognition domain (TRD), **C)** "Triple mutant" of M.*Hha*I with the *Msp*I specific TRD from M.SPRI gene inserted (the insertion shown in red). Extra amino acids flanking the inserted *Msp*I specific TRD are shown in green.

4.3: Replacement experiments in M.*Hha*I

4.3.1: REPLACEMENT OF THE M.*Hha*I TRD WITH M.SPRI TRDs

As shown by the above experiments, M.*Hha*I cannot tolerate the insertion of the *Msp*I TRD, and this insertion abolishes M.*Hha*I activity even when motif IX has been modified to resemble that of multi-specific Mtase M.SPRI. All the insertions described in the present experiment were carried out in the variable region towards the C-terminal region of the protein, which is essential for interactions with motif IX. It was assumed that because the sequence of motif IX has been modified, and had now become more like multi-specific Mtase M.SPRI, it might accommodate a larger variable region carrying all three TRDs. Therefore it was decided to swap the TRD region in construct pQIS 23 (see above) with the three TRDs of M.SPRI. However, from cloning point of view, this was not a straightforward experiment. In order to ensure that the cloning sites in the M.*Hha*I mutant gene are unique, some restriction sites must be removed. In addition, similar or compatible sites were required in both M.*Hha*I and M.SPRI genes for error-free swapping. The construct pQIS 23 was used first for the abolition of few of the restriction sites from the multiple cloning sites at 3' end of the M.*Hha*I gene as follows.

4.3.1.1: Removal of an 83 bp fragment downstream of the 3′end of the M.*Hha*I gene

The 3′ end of the gene contained in construct pQIS 23 carries part of a multiple cloning site, which proved to be a significant hurdle in obtaining the recombinant plasmid. Removal of this sequence was achieved by cleavage of the construct pQIS 23 with *Stu*I and *Pst*I followed by an end-filling reaction. The linearised 3982 bp plasmid was recovered from the gel and the cohesive end generated by *Pst*I was converted into a blunt end. This was carried out by using $Vent_R^{®}$ DNA polymerase. This enzyme can be used for filling-in a 5' overhang due to 5′-3′ polymerase activity, or to "chew back" a 3′ overhang via its 3′-5′ exonuclease activity to create blunt ends. Restriction digestion with *Stu*I yields blunt-ended DNA products. After treatment with $Vent_R^{®}$ DNA polymerase the linear plasmid was purified by phenol extraction followed by ethanol precipitation. The DNA was circularised by self-ligation. The resultant construct was called pQIS 26 and its identity was confirmed on the basis of the disappearance of various restriction sites (Figure 4.5).

4.3.1.2: Deletion of *Xho*I and *Eco*RI sites downstream of the 3′end of the gene encoding M.*Hha*I

The *Xho*I and *Eco*RI restriction sites that are present at the 3′ end of the M.*Hha*I gene were deleted as follows. Two complementary mutagenic oligodeoxynucleotides NIS

NIS23/ NIS 24 duplex mutagenic oligonucleotide:

Δ *Xho*I / Δ *Eco*RI

5′-TTCAAACCATATG**CTCGTGGGATTC**CCCTTAAGTGAG-3′
3′-AAGTTTGGTATAC**GAGCACCCTAAG**GGGAATTCACTC-5′
F K P Y A R G I P L K

NIS 23: 5′-TTCAAACCATATG**CTCGTGGGATTC**CCCTTAAGTGAG-3′

NIS 24: 5′-CTCACTTAAGGG**GAATCCCACGAG**CATATGGTTTGAA-3′

23 and NIS 24 were designed and synthesised (with the mutant base shown in red in the primer sequence below). Mutagenesis was achieved using the Quick-Change (Stratagene) method and the resultant construct was named pQIS 28 (Figure 4.6). The identity of the construct was confirmed by the restriction analysis as shown (Figure 4.6).

4.3.1.3: Construction of M.*Hha*I lacking a TRD region by the insertion of Linker NIS 25/NIS 26 into the variable region

The next step in this cloning strategy was insertion of a short linker, replacing the entire TRD coding region up to the start of the sequence-encoding motif IX that lies on an *Xho*I-*Bbs*I restriction fragment in pQIS 28. A pair of oligonucleotides were designed and synthesized for this purpose. The NIS 25/ NIS 26 linker duplex, which was obtained by annealing the two oligonucleotides, has cohesive ends for *Xho*I and *Bbs*I. The linker also contains unique *Sal*I, *Pst*I, *Mun*I and *Eco*RI sites (shown below in bold or in a different colour for distinction).

NIS 25/NIS 26 Linker duplex:

***Xho*I *Sal*I *Pst*I *Mun*I *Eco*RI *Bbs*I**
5′-**TCGAG**C**GTCGACCTGCAGCAATTGGAATTC**G**GAAGACAC**-3′
3′-**C**G**CAGCTGGACGTCGTTAACCTTAAG**C**CTTCTGTGCCTT**-5′

NIS 25: 5′-TCGAGCGTCGACCTGCAGCAATTGGAATTCGGAAGACAC-3′

NIS 26: 5′-TTCCGTGTCTTCCGAATTCCAATTGCTGCAGGTCGACGC-3′

pQIS 28 was cleaved with *Xho*I and *Bbs*I to remove a 342 bp fragment encoding the variable region and *Msp*I and *Hha*I TRDs, and, after recovering the larger band (3636 bp) from the agarose gel, the NIS 25/NIS 26 linker was inserted into this vector. Digestion of the mutant constructs with various restriction enzymes confirmed the formation of the new construct, which consists of the M.*Hha*I gene without any TRD region (Figure 4.7). The construct was called pQIS 30.

4.3.1.4: Cloning of the *Eco*RII, *Msp*I & *Hae*III specific TRDs of M.SPRI into the M.*Hha*I triple mutant

An experiment was planned to see if the triple mutant of M.*Hha*I (Ala 280 Phe, Arg 277 Leu and Phe 302 Ala) (which contains a motif IX that is like that of multi-specific Mtases) could tolerate an insert (Segment B, consisting of all the three TRDs) from multi-specific M.SPRI. The vector DNA was prepared by cleavage of pQIS 30 with *Sal*I and *Eco*RI for the insertion of the TRDs coding sequence. The resulting vector DNA (3657 bp) was recovered from an agarose gel. The insert (390 bp) DNA comprising the three TRD coding sequences prepared by using the same restriction enzyme digestion of pQIS 7 (encoding wild type M.SPRI) and was then recovered from the gel (Figure 4.8). The insert and the vector DNA were kept for ligation and the resulting constructs were analysed by restriction digestion. The *Pst*I site that was present in the linker region had now disappeared (because the insertion of the sequence encoding three TRDs from M.SPRI does not contain this site). When the new molecule was cleaved with *Mun*I it gave a linear DNA fragment of 4.0 kb, which also confirmed the presence of the insert. The new construct was called pQIS 31. Restriction analysis with *Msp*I and *Hae*III were used to check the activity of the enzyme encoded by recombinant construct pQIS 31. The enzyme encoded by the resultant construct was found inactive (Figure 4.8).

4.3.1.5: Removal of three amino acids to give an exact alignment with the consensus motif IX

The alignment containing the new construct (shown below) revealed that there are three extra amino acids Lys, Thr and Arg present because of a slight difference in the relative positions of *Eco*RI site in the two Mtases. Although the last Arg present could be equivalent to the conserved Arg, this assumption would change the space

between the last TRD (*Hae*III TRD) and motif IX. The data indicate that the amino acids beyond the C-terminus of the ensemble of TRDs, although not necessary for individual TRD activity, may very well modulate the overall methylation potential of these enzymes (Trautner *et al.*, 1996).

Therefore the next experiment was designed to remove these three amino acids, which were Lys, Thr and Arg. It was decided to introduce another *Eco*RI site after three amino acids. Finally the DNA was cut with *Eco*RI in order to remove the extra three codons (followed by self-ligation).

Name	Swiss Prot	Sequence
M.*Hha*I	P05102	L V N G K T R - - - K L H P R E C A R V M G Y P D
M.SPRI	P00476	Y P K Y R I R - - - R L T P L E C F R L Q A F D D
pQIS 31	——	Y P K Y R I R K T R K L H P L E C F R V M G Y P D

Two mutagenic primers NIS 27 and NIS 28 were designed to introduce a new *Eco*RI site 9 bp away from the existing *Eco*RI site. The introduction of a new *Eco*RI site leads to the deletion of a *Bbs*I restriction site, which was used for diagnostic purposes. The sequences of the primers and of the duplex are shown below, with the mutated bases shown in red.

NIS 27/ NIS 28 duplex mutagenic oligonucleotide:

*Eco*RI *Eco*RI

5′-CCAAAATACA**GAATTC**GGA**GA**A**TT**C**G**GAAATTACACCCTC-3′

3′-GGTTTTATGT**CTTAAG**CCT**CT**T**AA**G**C**CTTTAATGTGGGAG-5′

P K Y R I R R I R K L H P

NIS 27: 5′-CCAAAATACA**GAATTC**GGA**GA**A**TT**C**G**GAAATTACACCCTC-3′

NIS 28: 5′-GAGGGTGTAATTTCC**GA**A**TT**C**T**CC**GAATTC**TGTATTTTGG-3′

Site-directed mutagenesis was carried out using *Pfu* polymerase, and the PCR product was digested with *Dpn*I for one hour before being used to transform *E. coli* (Figure 4.9). The resulting constructs were screened by restriction analysis with *Bbs*I. The correct construct (numbered pQIS 32) was used for digestion with *Eco*RI followed by religation.

The two *Eco*RI sites in pQIS 32 are 9 bp apart. According to the manufacturer (NEB) *Eco*RI requires one base at the end of a DNA fragment to effect cleavage with an efficiency greater than 90% after two hours of incubation. Therefore the DNA was incubated with concentrated *Eco*RI for 6 hours and the products were run on an agarose gel in order to recover the linearised DNA fragment (Figure 4.9). The DNA was then retained for religation and was then used to transform *E. coli*. Plasmid DNA was prepared from the colonies and the corrected construct was named pQIS 33 (Figure 4.9).

4.3.2: CLONING OF AN ORF FROM M.SPRI CONSISTING OF PART OF THE VARIABLE REGION AND THREE TRDs INTO THE M.*Hha*I TRIPLE MUTANT

Although the redesigned motif IX of mono-specific Mtase M.*Hha*I resembled that of the multispecific Mtases and the distance between the TRD region and motif IX had been altered, the chimaeric construct was still inactive. Previous data have demonstrated that the there is a need to preserve the variable region for enzyme activity (Wilke *et al.*, 1988, Trautner *et al.*, 1996). Because we have exchanged the TRD region of M.*Hha*I with that of three specific TRDs of M.SPRI, it was assumed that the TRDs of M.SPRI may require their own variable region for activity.

In the mono specific C5-Mtase M.*Hha*I, the variable region of some 60 amino acids represents part of the large catalytic domain and the connector region to the small domain involved in target recognition (Klimašauskas *et al.*, 1994).

Using this information we decided to insert the ORF for the variable region to a limited length of some 65 amino acids from M.SPRI so that it should not exceed the size of larger catalytic domain as in the original monospecific Mtase M.*Hha*I. Therefore the final experiment planned was the insertion of the sequence of the variable region and the three specific TRDs (Segment C, see Figure 4.1) from M.SPRI gene into the triple mutant M.*Hha*I gene.

4.3.2.1: Introduction of a new *Xho*I site into the gene encoding wild type M.SPRI

For the cleavage of the sequence of variable region and three TRDs from M.SPR, it was essential to generate a new site in pQIS 7 (encoding wild type M.SPR) at the appropriate location. It was decided to introduce a *Xho*I site and to use the *Xho*I and

*Eco*RI sites for the proposed experiment. Therefore a new *Xho*I site was introduced into the variable region of the M.SPR gene, in order to facilitate subsequent removal of the sequence encoding the variable region and the three TRDs by *Xho*I-*Eco*RI cleavage. Unfortunately it was not possible to introduce a new *Xho*I site without introducing an amino acid change in the open reading frame of M.SPR: Gln 189 Glu. A set of mutagenic oligonucleotides; NIS 31 and NIS 32 were designed and synthesized as shown below.

NIS 31/ NIS 32 duplex mutagenic oligonucleotide:

*Xho*I/AvaI

5′-AGAAAGGATATA**CTCGAG**AAAGGGAAACAGAG-3′

3′-TCTTTCCTATAT**GAGCTC**TTTCCCTTTGTCTC-5′

R K D I L **E** K G K Q

NIS 31: 5′-AGAAAGGATATA**CTCGAG**AAAGGGAAACAGAG-3′

NIS 32: 5′-CTCTGTTTCCCTTT**CTCGAG**TATATCCTTTCT-3′

pQIS 34 was obtained from pQIS 7 and carries two *Xho*I sites and was therefore subsequently used for the excision of *Xho*I/*Eco*RI fragment for insertion into the M.*Hha*I encoding chimaeric construct pQIS 32 (Figure 4.10). The construct pQIS 34 encodes wild type M.SPRI containing a new *Xho*I site in the coding sequence specifying the variable region and has one amino acid codon Gln 189, mutated to a Glu codon. Since pQIS 34 was protected against cleavage with *Msp*I (Figure 4.10), it was therefore concluded that the protein encoded by pQIS 34 is enzymatically active.

4.3.2.2: Cloning part of the sequence encoding the variable and TRD regions from M.SPRI into the M.*Hha*I triple mutant gene

The appropriate vector and insert were prepared by cleavage of pQIS 32 and pQIS 34 with *Xho*I and *Eco*RI. Both the vector DNA (3.6 kb) from pQIS 32 and insert DNA (550 bp) from pQIS 34 were recovered from an agarose gel, and were ligated together. The resultant construct was called pQIS 35. The presence of the insert was confirmed by *Xho*I/*Eco*RI double digestion, which yielded an expected 565 bp fragment. The activity of the construct was checked by *Hae*III and *Msp*I digestion which revealed that the protein encoded by the construct was enzymatically inactive.

The bacterial cells carrying the new construct, pQIS 35, were grown in the presence of IPTG, and plasmid DNA was recovered. It was checked again to see if the plasmid was protected against *Hae*III and *Msp*I digestion. This experiment confirmed that the plasmid coded for a protein that was enzymatically inactive (Figure 4.11).

4.3.3: CLONING OF AN ORF FROM M.SPRI, CONSISTING OF THE ENTIRE VARIABLE REGION AND THREE TRDs INTO THE M.*Hha*I TRIPLE MUTANT

The general architecture of all C5-Mtases is well documented and it has been noted that two regions are important for base flipping. One is at the C-terminus of the highly conserved catalytic motif, while the second lies in the variable region that is responsible for DNA sequence recognition (Roberts, 1995). All variable regions are bipartite. They contain at their N-terminal side a very similar sequence of 71 amino acids. This sequence is highly conserved among the multi-specific Mtases, and the integrity of this sequence must be maintained in order to preserve enzyme activity (Trautner *et al.*, 1996). In the mono-specific C5-Mtase M.*Hha*I, the equivalent region of 60 amino acids represents part of a large catalytic domain, as well as the connector region to the small domain involved in target recognition (Klimašauskas *et al.*, 1994). Hence, it is plausible that target recognition by multi-specific Mtases requires a 71-conserved amino acid sequence (i.e. not the equivalent region of M.*Hha*I). It was conjectured, furthermore, that if the conserved 71 amino acid sequence of the multi-specific Mtase M.SPRI was replaced with the equivalent region of 60 amino acids from the mono-specific Mtase M.*Hha*I, it might restore the activity to the enzyme. The next experiment was designed to introduce the conserved 71 amino acid sequence of M.SPRI with the three TRDs into the redesigned M.*Hha*I gene. In pQIS 7 (which encodes for wild type M.SPRI) there is an existing *Pac*I site at the start of the coding sequence for the variable region. However, there was a need to introduce a similar or compatible site at an equivalent position in construct carrying the M.*Hha*I gene.

4.3.3.1: Introduction of a *Pac*I site into various constructs carrying the M.*Hha*I gene

It was decided to introduce a *Pac*I site into pMSK 7 (encoding wild type M.*Hha*I), into pQIS 18 (expressing the redesigned triple mutant of M.*Hha*I) as well as pQIS 35

(in which we have introduced an *Eco*RI site at an equivalent position to that in the gene encoding M.SPRI) (See sections 4.1.5.2 and 4.1.5.3). The mutagenic oligonucleotides used for this purpose were NIS 33 and NIS 34 (the new restriction site is shown below in bold, with the mutation in red).

NIS 33/ NIS 34 duplex mutagenic oligonucleotide:

*Pac*I

5′-GATTTGTTTTCGCAATGAT**TTAATTAA**TCAAAATTTCCAATTTCC-3′

3′-CTAAACAAAAGCGTTACTA**AATTAATT**AGTTTTAAAGGTTAAAGG-5′

I C F R N D L **I N** Q N F Q F

NIS 33: 5′-GATTTGTTTTCGCAATGAT**TTAATTAA**TCAAAATTTCCAATTTCC-3′

NIS 34: 5′-GGAAATTGGAAATTTTGA**TTAATTAA**ATCATTGCGAAAACAAATC-3′

Three plasmids: pMSK 12, pQIS 18 and pQIS 35 were used as templates for site-directed mutagenesis to introduce the *Pac*I site and to effect two codon changes: Asn 176 Ile and Ile 177 Asn (Figure 4.12). The constructs were checked by restriction analysis, and numbered pQIS 36, pQIS 37 and pQIS 38 (as shown in Figure 4.12). The protein coded by the construct pQIS 36 (a mutant of wild type M.*Hha*I (Asn 176 Ile and Ile 177 Asn)) retains enzymatic activity since the pQIS 36 is not cleaved by *Hha*I. The construct pQIS 37 was generated from pQIS 18 (encoding the redesigned triple mutant of M.*Hha*I) and was found to express an inactive Mtase; it appears that the Asn 176 Ile and Ile 177 Asn mutations inactivate the triple mutant of M.*Hha*I (pQIS 18). The third construct pQIS 38 was inactive, as expected, because the template DNA pQIS 35 was inactive. This experiment indicated that there may be a initial interaction between motif IX and the variable region in M.*Hha*I. This inference is based on the observation that when the redesigned motif IX (active in the triple mutant state) is used for mutation (Asn 176 Ile and Ile 177 Asn) in the variable region, it is unable to methylate.

The analysis of the crystal structure of M.*Hha*I for Asn-Ile at position 176-177 revealed that these two amino acids reside at the surface of the protein at the starting curl of the variable loop region (see Figure 4.21). Although these two amino acids

two triplets, involving the Pro-Cys region and the Thr-Leu residues that are found in the TRD which lies in the small domain, on the other side of the cleft (Cheng *et al.*, 1993). The side-chain nitrogen of Asn 120 forms a hydrogen bond to the backbone oxygen of residue 79 next to Pro 80. Glu 119 also interacts with Arg 165 through ion pairing. A number of hydrogen bonds link the large and small domains through amino acids 160-163 and the region around Thr250-Leu251 (Cheng *et al.*, 1993). Because the overall architecture of this set of enzymes (C5-Mtases) is so well conserved, it is highly likely that they will all have similar structures and will use the same catalytic mechanism (Roberts, 1995). However, it is likely that the physical interactions between the conserved dipeptide T (L, V, I) found in the TRD region and the core catalytic domain will be arranged in a different way in multi-specific Mtases because of the presence of multiple TRDs. Based on this assumption, we once again decided to carry out another set of experiments that involve the transfer of a region starting from motif VI and ending at the TRDs (from the multi-specific Mtase M.SPRI). The region before motif VI contains the invariant triplet Glu119-Asn120-Val121 (**ENV**) in M.*Hha*I and M.SPRI were aligned to find a suitable place for the introduction of a unique site, which would facilitate exchange of the desired region. The alignment is shown below with the conserved residues shown in red.

Alignment of M.*Hha*I and M.SPRI in motif VI

```
                    116
                     |
M.HhaI:     REKKPKVVFMENVKNFASHDNGN
M.SPRI:     KEKQPKFFVFENVKGLINHDKGN
                     |
                    112
```

In insertion mutagenesis experiments, two problems are often encountered, and this was the case in planning a strategy for exchanging segments of M.SPRI and M.*Hha*I. For example, in this experiment the only suitable unique restriction site present in both plasmids was *Apa*I, which can be introduced at codons Val 116 and Phe 117 of M.*Hha*I and in M.SPRI at codons Phe 112, Phe 113 and Val 114. This is a drawback, as the resulting site generates a mutation in itself but at the same time provides an opportunity to study the individual mutations. Therefore, it was decided to introduce an *Apa*I site in both genes, complete the fragment exchange and, if required restore

the original codons by mutation to maintain the structural integrity of the protein. The second problem that we faced was the maintenance of the reading frame with respect to the newly introduced restriction sites. The ORF of M.*Hha*I is in a different frame to that of M.SPRI, and this had to be taken into consideration.

4.3.4.1: Introduction of a new *Apa*I site into the gene encoding M.SPRI

The mutagenic oligonucleotides designed and synthesized for the introduction of this *Apa*I site were NIS 35 and NIS 36 (as shown below, with the mutated bases shown in red). This mutagenesis experiment led to the following changes: Phe 112 Leu, Phe 113 Gly and Val 114 Pro. It was decided to study these mutations using two plasmids as templates, which are as follows.

A: Wild type M. SPRI (encoded by construct number pQIS 7) (See Section 3.1.1).

B: Mutant M.SPRI F386A (encoded by construct number pQIS 8) (See Section 3.1.4.1).

NIS 35/ NIS 36 duplex mutagenic oligonucleotide:

```
                        ApaI
        5′-CAAAGCAACCAAAGTTGGGCCCTTTTGAAAATGTTAAAG-3′
        3′-GTTTCGTTGGTTTCAACCCGGGAAAACTTTTACAATTTC-5′
             K  Q  P  K  L  G  P  F  E  N  V  K
```

NIS 35: 5′-CAAAGCAACCAAAGTTGGGCCCTTTTGAAAATGTTAAAG-3′

NIS 36: 5′-CTTTAACATTTTCAAAAGGGCCCAACTTTGGTTGCTTTG-3′

(A): Introduction of an *Apa*I site in pQIS 7

The construct carrying the wild type M.SPRI gene (pQIS 7) was used as the template for the Phe 112 Leu, Phe 113 Gly and Val 114 Pro mutagenesis experiment and the resulting construct was verified by *Apa*I digestion. pQIS 7 contains a unique *Apa*I site: by this experiment a second site was introduced. Upon restriction digestion with *Apa*I, two fragments were obtained of the correct size (3962+2365 bps), which confirmed that the mutagenesis had been successful. The new construct, pQIS 45, upon restriction digestion with *Apa*I and *Eco*RI yielded a 787 bp fragment, which also confirmed its authenticity. The plasmid was checked for the activity of the

encoded mutant enzyme by *Msp*I restriction analysis. Once again it was found that the protein encoded by the mutated protein of M.SPRI was inactive (Figure 4.14).

(B): Introduction of an *Apa*I site into pQIS 8

The construct pQIS 8 (encoding the Phe 386 Ala mutant of M.SPRI) was used as a template for the site-directed mutagenesis experiment, for the introduction of the new *Apa*I site. The resulting construct, when digested with *Apa*I yielded two fragments of DNA (3962+2365 bps), which confirmed that the plasmid carried a new *Apa*I site (Figure 4.14). The new construct was called pQIS 46 (a Phe 386 Ala mutant of M.SPRI which also now carried other mutations Phe 112 Leu, Phe 113 Gly and Val 114 Pro in the region just before the ENV region (motif VI)). When the methylation potential of the mutant protein was measured, it was observed that the protein was enzymatically inactive (Figure 4.14).

4. 3.4.2: Introduction of a new *Apa*I site into the gene encoding M.*Hha*I

For mutagenesis and introduction of a new site into the gene encoding M.*Hha*I (and its mutant constructs) oligonucleotides NIS 37 and NIS 38 were designed and are as shown below with the base changes shown in red. This mutation experiment led to changes of two amino acid residues (from the wild type M.*Hha*I) namely Val 116 Gly and Phe 117 Pro.

NIS 37/ NIS 38 duplex mutagenic oligonucleotide:

*Apa*I

5′-AAAAAACCTAAAGT**GG**GCCCTATGGAAAATGTG-3′

3′-TTTTTTGGATTTCA**CC**CGGGATACCTTTTACAC-5′

K K P K V G P M E N V

NIS 37: 5′-AAAAAACCTAAAGT**GG**GCCCTATGGAAAATGTG-3′

NIS 38: 5′-CACATTTTCCATAGGGC**CC**ACTTTAGGTTTTTT-3′

The methyltransferase *Hha*I-carrying constructs selected for these experiments were as follows.

A: pMSK 12 (encoding for wild type M.*Hha*I)

B: pQIS 18 (encoding for redesigned triple mutant of M.*Hha*I)

C: pQIS 36 (encoding for Asn 176 Ile, Ile 177 Asn mutant of M.*Hha*I)

A: Introduction of a new *Apa*I site into pMSK 12

The substitutions Changes Val 116 Gly and Phe 117 Pro were carried out using site-directed mutagenesis (for the synthesis of new construct pQIS 41). pQIS 41 was identified by the presence of a unique *Apa*I site, digestion with which produced linearised plasmid (Figure 4.15). The mutations Val 116 Gly and Phe 117 Pro caused the encoded protein to be enzymatically inactive (Figure 4.15).

B: Introduction of a new *Apa*I site into pQIS 18

Using oligonucleotides NIS 37 and NIS 38, a site-directed mutagenesis experiment was carried out on pQIS 18 and the resulting new construct, pQIS 42 was checked for the presence of this new *Apa*I site. Restriction analysis confirmed the identity of the new construct pQIS 42 (Figure 4.15). *Hha*I degraded the plasmid, demonstrated that the plasmid is not protected and therefore encodes an inactive Mtase (Figure 4.15).

C: Introduction of a new *Apa*I site into pQIS 36

pQIS 36 was also used as template DNA for the introduction of mutations Val 116 Gly and Phe 117 Pro. The identity of the resulting construct pQIS 43 was confirmed on the basis of the presence of the unique *Apa*I site. pQIS 43 was checked for susceptibility to *Hha*I and was found to encode an inactive enzyme (Figure 4.15).

D: Introduction of motif VI, the variable regions and the TRDs region from M.SPRI into the M.*Hha*I triple mutant

This experiment was performed in two steps. In the first step the insert (segment E, see Figure 4.1) from construct pQIS 45 (encoding M.SPRI), which had been recovered by cleavage with *Apa*I and *Eco*RI, was ligated with the vector generated from pQIS 41 (cut with same two enzymes (See Figure 4.15)). The identity of the resulting construct pQIS 47 (Figure 4.16) was confirmed by restriction analysis, and then, using *Alw*NI and *Eco*RI restriction enzymes, the missing sequence encoding the C-terminal part of the triple mutant enzyme M.*Hha*I (region containing motif IX and motif X) was swapped from pQIS 40. The final construct pQIS 48 (Figure 4.16) now

possessed a larger coding region of M.SPRI and includes conserved motif VI (ENV region) and all three TRDs up to the start of motif IX, flanked by the redesigned triple mutant of M.*Hha*I. The identity of this construct was confirmed by restriction analysis, as shown in Figure 4.16. The activity of the encoded mutant Mtase was also checked by *Msp*I digestion of the plasmid and it was found that the protein encoded by the gene is inactive (Figure 4.16). However, this was to be expected because the mutation introduced by the extra *Apa*I site had already inactivated the enzyme (See section above). Therefore, as planned earlier, it was necessary to restore the original amino acid codons to see if this could make any difference to the methylation capability of the enzyme.

E: Restoration of original codon of motif VI in the chimaeric construct encoded by pQIS 48

The sequence of the chimaeric construct pQIS 48 was aligned with that of wild type M.*Hha*I and M.SPRI, and on examination it was decided to mutate the Gly into Val and Pro into Phe (to maintain the amino acids up to this point similar to that of M.*Hha*I). Mutagenic primers designed for site-directed mutagenesis were NIS 41 and NIS 42, as shown below (with the changed bases shown in bold and red).

Alignment of M.*Hha*I, M.SPRI and chimaeric construct pQIS 48 in motif VI

M.*Hha*I: **REKKPKVVFMENVKNFASHDNGN**
M.SPRI: **KEKQPKFFVFENVKGLINHDKGN**
pQIS 48: **REKKPKVGPFENVKGLINHDKGN**

NIS 41/ NIS 42 duplex mutagenic oligonucleotide:

ΔApaI site
5′-GAAAAAAAACCTAAAGTGG**T**C**TT**TTTTGAAAATGTTAAAGGG-3′
3′-CTTTTTTTTGGATTTCACC**A**G**AA**AAAACTTTTACAATTTCCC-5′
E K K P K V V F F E N V K G

NIS 41: 5′-GAAAAAAAACCTAAAGTGG**T**C**TT**TTTTGAAAATGTTAAAGGG-3′

NIS 42: 5′-CCCTTTAACATTTTCAAAA**AA**G**A**CCACTTTAGGTTTTTTTTC-3′

These primers were designed to mutate the Gly-Pro dipeptide into Phe-Val, and to abolish the *Apa*I site. Site-directed mutagenesis-generated plasmids were prepared and characterised and it was found that the *Apa*I site had been deleted. The construct was named pQIS 52, as shown in Figure 4.17. However, the enzyme generated by the encoded protein was still inactive (Figure 4.17).

4.3.4.3: Introduction of motif VI, the variable regions and the TRDs region from M.SPRI into wild type M.*Hha*I

A: Deletion of the *Eco*RI site from pMSK 12 (encoding the wild type M.*Hha*I)

To achieve this objective an experiment was designed in which the existing *Eco*RI site would be deleted from the wild type M.*Hha*I gene, and a second *Eco*RI site introduced before the sequence encoding motif IX. Since an *Eco*RI site exists outside the ORF of M.*Hha*I (pMSK12) it was deleted by cleavage with *Eco*RI followed by treatment with $Vent_R^{®}$ DNA polymerase to make the DNA blunt-ended (see section 4.1.4.1). Blunt-ended DNA was self-ligated and a construct with a deleted *Eco*RI site was obtained. The construct was named pQIS 49 (wild type M.*Hha*I gene without an *Eco*RI site). The construct expressed protein, which subsequently protected the plasmid against *Hha*I cleavage (Figure 4.18).

B: Introduction of an *Apa*I site in pQIS 49

The oligonucleotides used for mutagenesis and introduction of the *Apa*I site in pQIS 49 (wild type M.*Hha*I with deleted *Eco*RI site) were NIS 37 and NIS 38 (See Section 4.1.7.2). A Site-directed mutagenesis experiment with wild type pQIS 49 was carried out to introduce changes: Val 116 Gly and Phe 117 Pro produced the new construct pQIS 50. pQIS 50 was characterised on the basis of its newly acquired unique *Apa*I site (Figure 4.18). The changes Val 116 Gly and Phe 117 Pro rendered the encoded protein enzymatically inactive (Figure 4.18).

C: Introduction of an *Eco*RI site into pQIS 50 (Val 116 Gly and Phe 117 Pro mutant of M.*Hha*I gene)

pQIS 50 was used as the template for the introduction of an *Eco*RI site (at the start of the sequence encoding motif IX, at an equivalent position to that in M.SPRI) using the mutagenic primers shown below, with changes shown in red.

NIS 43 / NIS 44 duplex mutagenic oligonucleotide:

*Eco*RI

5′-TATTTAGTAAACGGGA**GAATTC**GGAAATTACACCCTAG-3′

3′-ATAAATCATTTGCCCT**CTTAAG**CCTTTAATGTGGGATC-5′

Y L V N G **R I** R K L H P

NIS 43: 5′-TATTTAGTAAACGGGA**GAATTC**GGAAATTACACCCTAG-3′

NIS 44: 5′-CTAGGGTGTAATTTCC**GAATTC**TCCCGTTTACTAAATA-3′

Two changes were engineered using site-directed mutagenesis namely Lys 270 Arg and Thr 271 Ile. The construct pQIS 51 was made and was shown to contain both *Apa*I and *Eco*RI sites. pQIS 51 was found to encode inactive protein: the construct was not protected from *Hha*I digestion (Figure 4.18). pQIS 51 was now suitable for a desired swap of the coding region from M.SPRI (Motif VI to the start of motif IX).

D: Cloning of an *Apa*I-*Eco*RI gene fragment from the M.SPRI gene into the wild type M.*Hha*I gene

In this experiment an insert (Segment E, see Figure 4.1) from construct pQIS 45 (encoding M.SPRI) was recovered by cleavage with *Apa*I and *Eco*RI and was then ligated with the vector generated from pQIS 51 (already cut with same two enzymes). The resulting construct pQIS 53 (4267 bps) contained wild type M.*Hha*I expressing an inserted region from M.SPRI (from motif VI to the end of the TRDs) and this was confirmed on the basis of restriction analysis (Figure 4.19). The protein encoded by the construct pQIS 53 was found to be enzymatically inactive.

E: Restoration of the original codon of motif VI in the chimaeric construct encoded by pQIS 53

The chimaeric construct pQIS 53 encoded a similar motif VI sequence to that of pQIS 48. Therefore a similar experiment was designed to mutate Gly into Val and Pro into Phe. Mutagenic primers NIS 41 and NIS 42 (section 4.1.7.4), which had been previously used to mutate Gly-Pro residues into Val-Phe residues and to delete the *Apa*I site, were again used for site-directed mutagenesis. The new construct pQIS 54 was prepared (which encodes wild-type M.*Hha*I sequence from motif VI to the end of the TRDs from M.SPRI) (Figure 4.19). However, when the activity of the

protein encoded by the construct pQIS 54 was measured, it was found that it was enzymatically inactive.

4.3.5: THE RELATIONSHIP BETWEEN THE VARIABLE REGION AND MOTIF IX OF WILD TYPE M.*Hha*I

4.3.5.1: Introduction of an *Eco*RI site into pQIS 49

Plasmid pQIS 49 (encoding wild type M.*Hha*I) in which the existing *Eco*RI site had been abolished (Section 4.1.7.5:A) was used to engineer two changes, namely Lys 270 Arg and Thr 271 Ile using site-directed mutagenesis. This mutagenesis experiment introduces an *Eco*RI site at a position equivalent to the *Eco*RI site in M.SPRI. The primers NIS 43 and NIS 44 (Section 4.1.7.5:C) were used for this mutagenesis experiment. The mutant construct pQIS 55 (K270R, T271I) was identified by the presence of a unique *Eco*RI restriction site (Figure 4.20). However these changes affected the activity of the encoded M.*Hha*I. It was observed that the methylation activity of the mutant enzyme encoded by pQIS 55 was reduced (Figure 4.20).

4.3.5.2: Introduction of *Pac*I site into pQIS 55

pQIS 55 (containing the K270R and T271I mutations (and an *Eco*RI site)) was used as the template for Asn 176 Ile and Ile 177 Asn mutagenesis (for the introduction of a *Pac*I site). The primers NIS 33 and NIS 34 (section 4.3.3.1) were used for this experiment. The resulting construct, p*Hha*I-*Eco*RI-*Pac*I, was tested for the activity of the encoded protein. It was found that it had lost enzymatic activity and thus the plasmid was not protected against *Hha*I cleavage (Figure 4.20).

This experiment supports our previous observation (for the formation of pQIS 37) that if mutations are carried out in motif IX and the TRD region simultaneously, methylation activity is lost (Section 4.3.3.1). This shows a clear relationship between the two regions and supports previous author's findings (Mi and Roberts, 1992).

4.4: Discussion

In all C5-DNA Mtases from prokaryotes and eukaryotes, a set of conserved amino acid sequence elements are associated with general enzymatic functions. These sequence elements are arranged in canonical order. Crystallographic studies on two

bacterial Mtases have shown that general enzymatic functions such as cofactor (SAM) binding and catalysis are provided by a large "catalytic" domain formed primarily by the N-terminal part of the enzyme, (motifs I-VIII) in conjunction with motif X. The second, smaller "recognition" domain of the Mtases is composed of the central variable region containing the target recognition domain(s) (TRDs) and part of the C-terminal motifs IX and X (Klimašauskas *et al.*, 1994; Reinisch *et al.*, 1995). The specificity of methylation is achieved by the selective recognition of defined DNA target sequences, within which methylation occurs by the TRDs (Trautner *et al.*, 1996). The integrity of the small domain, providing a correct interplay between the variable TRD region and the conserved motifs IX and X, appears to be crucial for the enzymatic activity. This assumption followed experiments in which chimaeric monospecific Mtases were analysed. Activity of such chimaeric Mtases was only observed when the entire small domains (variable region, TRD and motif IX-X) were exchanged, whereas an exchange of only the variable region or only motif IX-X resulted in inactive enzymes (Mi and Roberts, 1992).

On the other hand, Walter *et al.*, (1992) have demonstrated the modular character of multispecific Mtases by deleting, rearranging and exchanging several TRDs. It has been shown that target recognition is independent of a particular TRD or core sequence context. It is also documented that multispecific Mtases can accommodate inert material of non-Mtase origin within their variable region without losing their activity (Walter *et al.*, 1992, Matin, 2000). This suggests that the enzyme core sequences preceding or following it form separable functional domains.

The enzyme M.*Hha*I regained activity after redesigning of motif IX by the introduction of three point mutations and it was thought that these changes may be sufficient to enable functional exchange of TRDs between a mono- and a multi-specific Mtase. The variable region in M.*Hha*I starts from amino acid 171 and lies on the surface of protein (Figure 4.21) spanning about 100 amino acids (Cheng *et al.*, 1993). Experiments were carried out in the loop at the N-terminus of the variable region of M.*Hha*I in the hope that a functional Mtase that would be re-targeted to alternative DNA sequences. The *Msp*I specific TRD sequence from M.SPRI was inserted preceding the *Hha*I TRD sequence at amino acid number 201-205 in both triple and quadruple mutants of M.*Hha*I. Both mutants mimic motifs IX-X of the multispecific Mtase, M.SPRI. However, both chimaeric construct pQIS 23 and pQIS 24 encoded inactive enzymes.

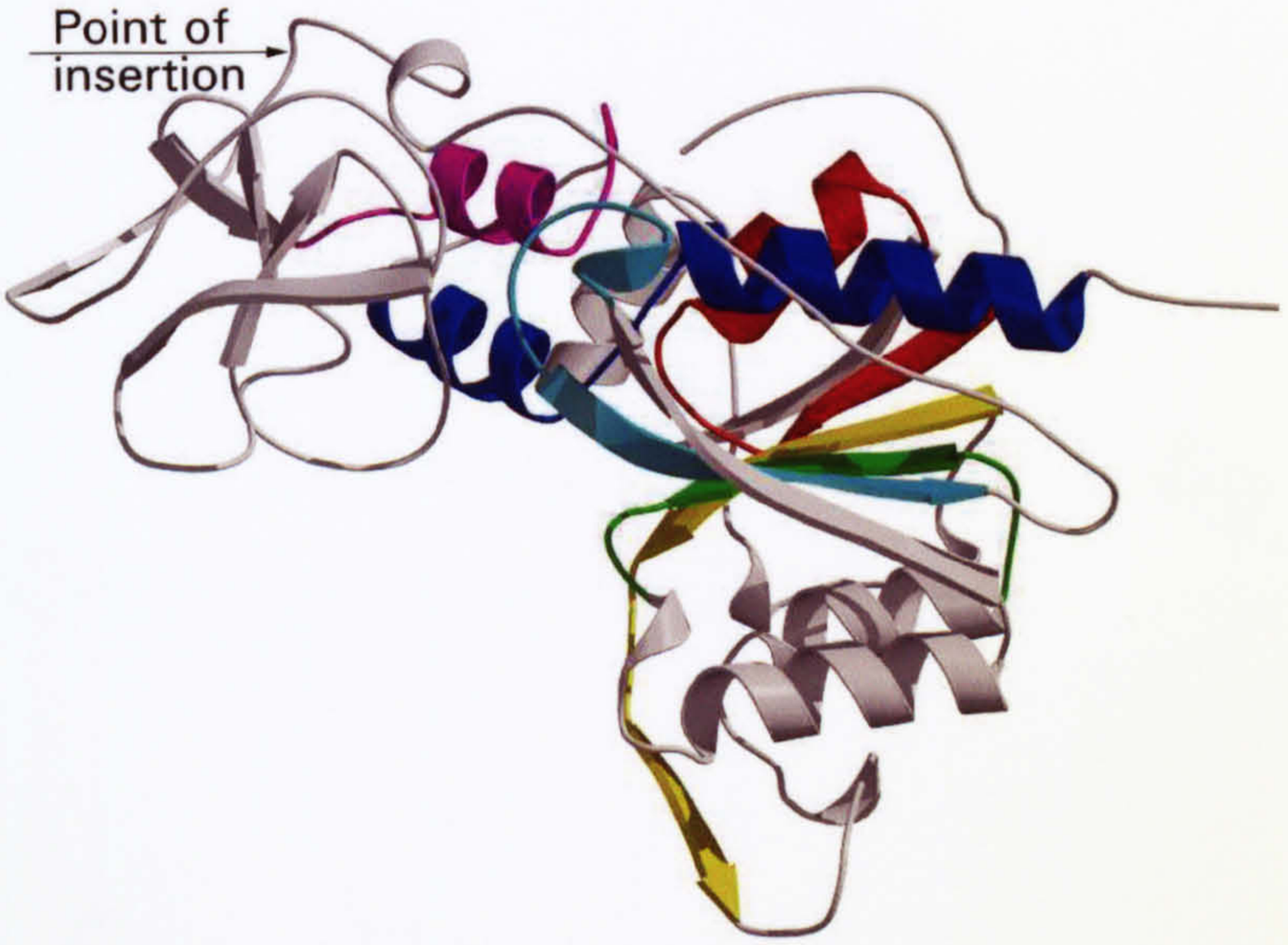


Figure 4.21:Ribbon diagram of M.*Hha*I coloured by motif as shown in Figure 1.6. The location of insertions is indicated with an arrow which is amino acid 201 as described in Figure 4.3. The image was created by Dr. Matthew J. Conroy. The figure was produced using Molscript (Kraulis (1991) *J. Appl. Cryst.* 24 946-950) and Raster3D, (Merritt & Murphy (1994) *Acta Cryst.* D50 869-873). (Figure from Cheng *et al.*, 1993, pdb code: 1hmy).

A second approach was adopted initially by exchanging the TRD of *Hha*I with that of M.SPRI. A series of other cloning experiments that included the exchange of the variable and the TRDs region of M.SPRI in M.*Hha*I followed this exchange strategy. From these insertion and exchange experiments (summarised in Figure 4.22 and 4.23) it became clear that it was difficult to convert a mono-specific Mtase into a multi-specific Mtase.

Various mutagenesis experiments in selected mutants and the wild type forms of both M.*Hha*I and M.SPRI were also carried out to determine the effect of individual and multiple mutations on the catalytic activity of the corresponding enzymes. A summary of mutations introduced in M.SPRI and M.*Hha*I are given in Figure 4.24. It has been observed that mutation in the variable region at or near the N-terminus of either M.SPRI or M.*Hha*I (construct pQIS 34 and pQIS 36) is tolerated by both Mtases. However, when pQIS 18 (encoding the triple mutant of M.*Hha*I) was used as template for the construction of pQIS 37, it was observed that the encoded enzyme had lost its ability to methylate. Another similar example can be seen where pQIS 55 was used to construct pHhaI-EcoRI-PacI. Although, the Mtase expressed from pQIS 55 already possesses a reduced level of enzyme activity, it is completely inactive when further mutations were introduced at the beginning of the variable region. This clearly confirms that close coupling of the variable region and motif IX of monospecific Mtase *Hha*I exists as observed by Mi and Roberts (1992).

It has also been observed that the loop of the variable region in monospecific M.*Hha*I, which starts from amino acid number 171, is not flexible and cannot support the insertion of polypeptide of any length.

In these experiments it has proved not possible to replace a mono-specific TRD region with that of a multispecific Mtase and in order for a multi-specific TRD to work in the framework of a monospecific Mtase a deeper understanding of the structure-function relationship of these enzymes is required. From this set of experiments it can be concluded that these enzymes are clearly superficially related but exhibit mechanistic and structural differences. A 3D-structure analysis of M.SPRI will be invaluable in attempting to explain these differences.

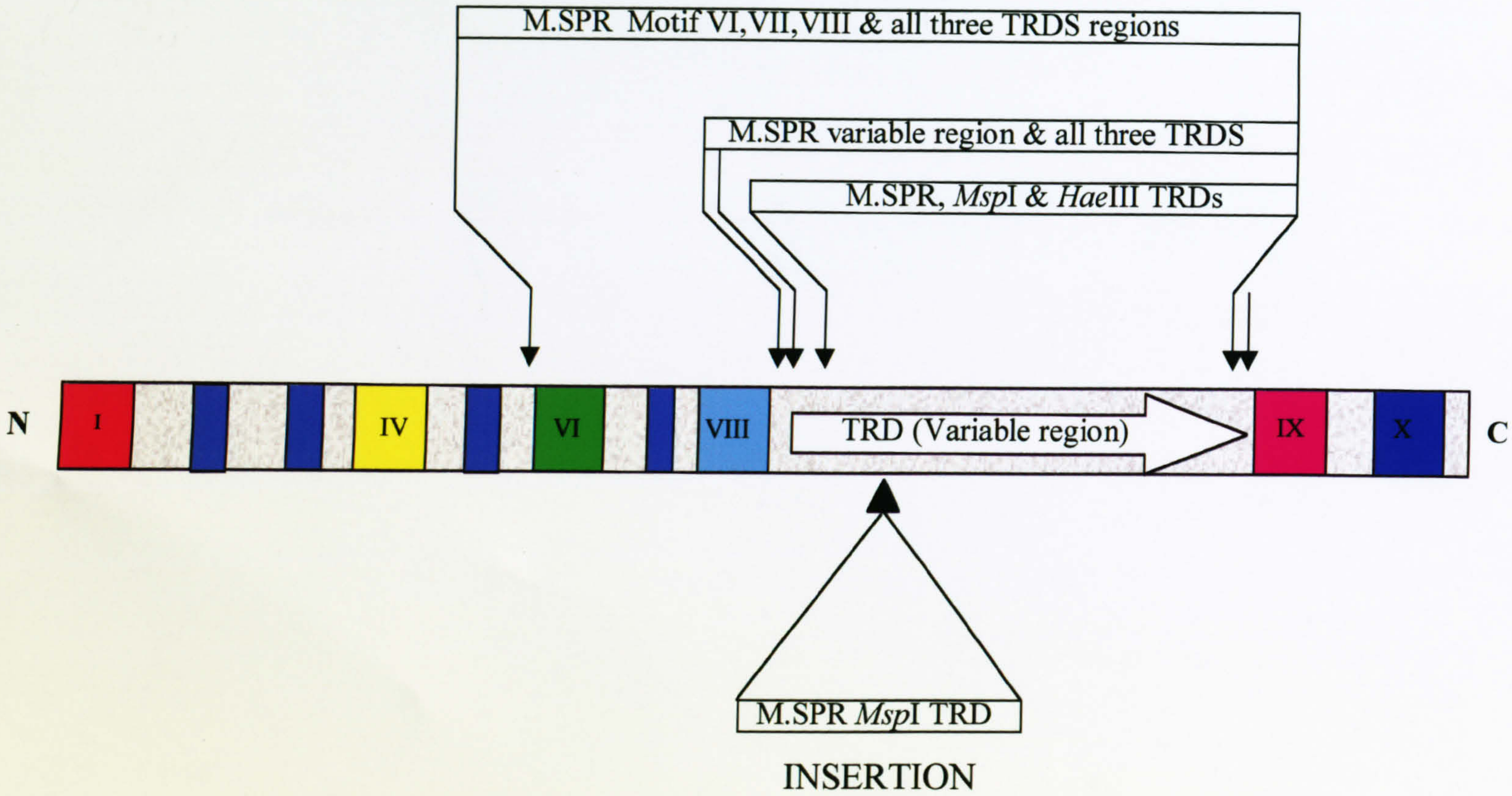


Figure 4.22: A general plan showing the construction of chimaeras by the insertion and replacement with various segments from M.SPRI into the engineered triple mutant of M.*Hha*I. The backbone architecture of C5-methyltransferase and their motifs are coloured as in Figure 1.4. The map is not according to scale. For further details of segments also consult Figure 4.23and text.

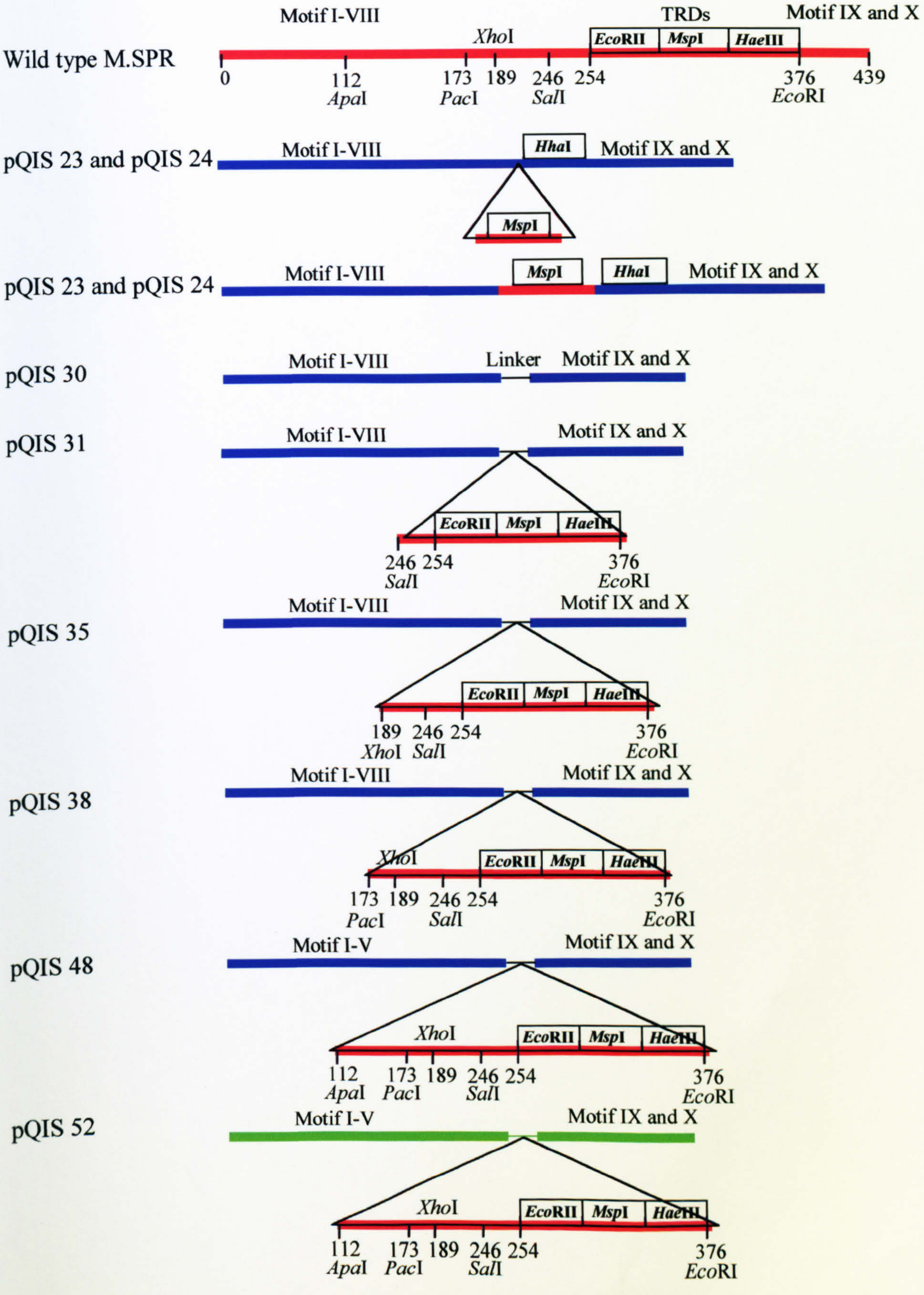
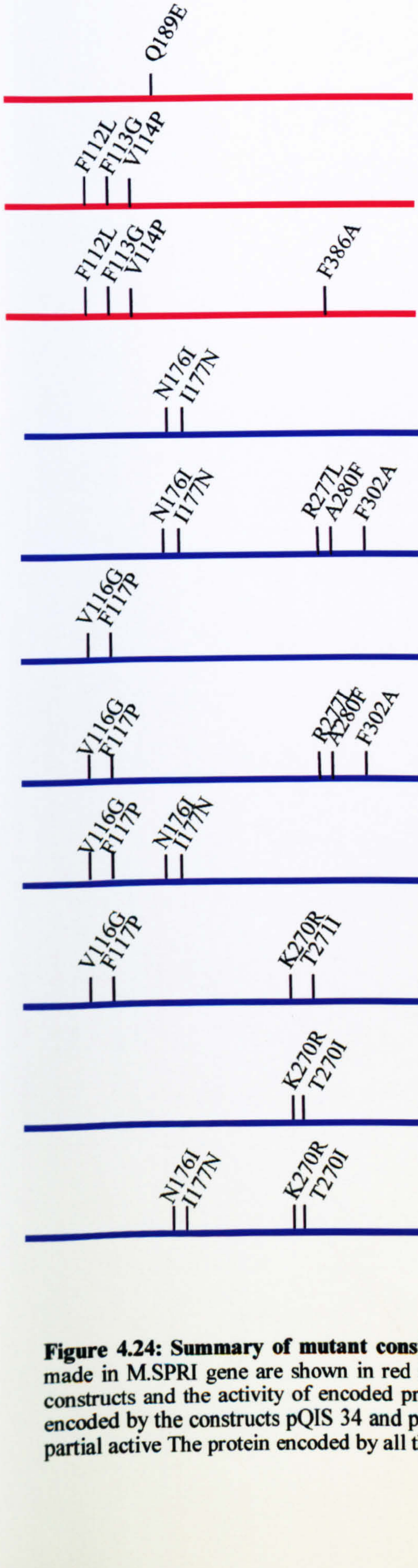


Figure 4.23: Schematic presentation of approximate size and position of inserts in the variable region of wild type and redesigned triple mutant of M.*Hha*I. Wild type M.SPR and the insert from M.SPR are shown in red. The positions of amino acids where mutations have been carried out for the introduction of unique restriction sites have been shown with amino acid residue number in M.SPR. The TRD region are shown in boxes. Wild type M.*Hha*I is shown in green and redesigned M.*Hha*I in blue. The map is not drawn to scale.



Construct No	Description
pQIS 34:	Single point mutation for the introduction of *Xho*I site in pQIS 7 encoding for wild type M.SPRI.
pQIS 45:	Triple mutant with the introduction of *Apa*I site in pQIS 7 encoding for wild type M.SPRI.
pQIS 46:	Triple mutant with the introduction of *Apa*I site in pQIS 8 encoding for single mutant of M.SPRI.
pQIS 36:	Double mutant with the introduction of *Pac*I site in construct encoding for wild type M.*Hha*I.
pQIS 37:	Double mutant with the introduction of *Pac*I site in mutant construct pQIS 36 encoding for triple mutant of M.*Hha*I.
pQIS 41/pQIS 50:	Double mutant with the introduction of *Apa*I site in construct encoding wild type M.*Hha*I.
pQIS 42:	Double mutant with the introduction of *Apa*I site in pQIS 18 encoding for triple mutant of M.*Hha*I.
pQIS 43:	Double mutant with the introduction of *Apa*I site in pQIS 36 encoding for double mutant of M.*Hha*I.
pQIS 51:	Double mutant with the introduction of *Eco*RI site in pQIS 50 encoding for double mutant of M.*Hha*I.
pQIS 55:	Double mutant with the introduction of *Eco*RI site in wild type M.*Hha*I.
pHhaI-EcoRI-PacI:	Double mutant with the introduction of *Pac*I site in construct pQIS 55 encoding for double mutant of M.*Hha*I.

Figure 4.24: Summary of mutant constructs made by point mutations. The map of mutants constructs made in M.SPRI gene are shown in red and in M.*Hha*I gene are shown in blue. A brief description of the constructs and the activity of encoded protein by these construct is shown in right hand panel. The protein encoded by the constructs pQIS 34 and pQIS 36 is active. The protein encoded by the constructs pQIS 55 is partial active The protein encoded by all the other constructs was inactive. The map is not according to scale.

Schematic diagrams
showing the plasmid sketches
with gel pictures

(Including Figures 4.2, 4.4 ~ 4.20)

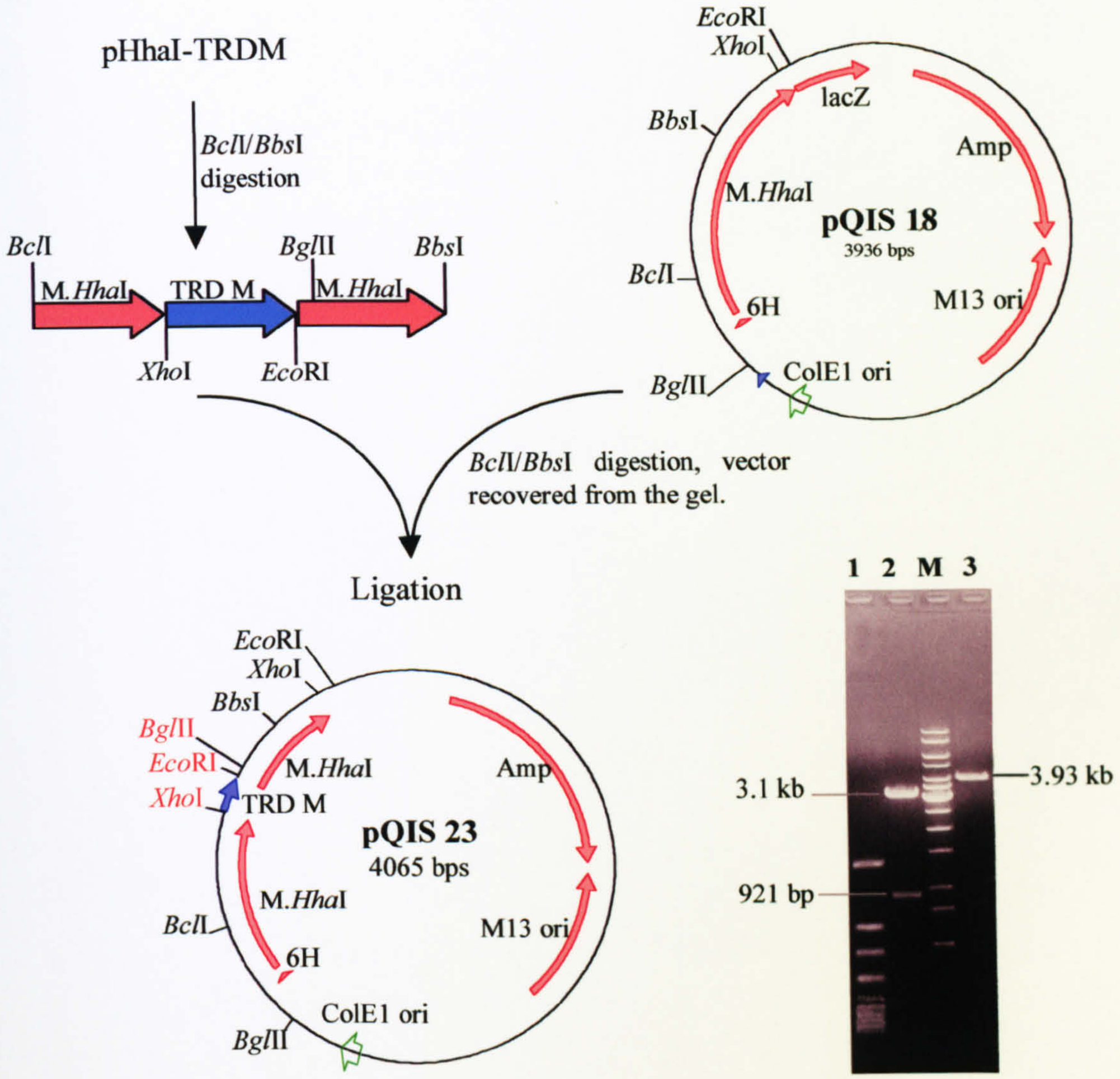


Figure 4.2: Schematic representation of the construction of pQIS 23. The TRD from *Msp*I was cloned into pQIS 18 (encoding the Ala 280 Phe, Arg 277 Leu & Phe 302 Ala mutations). New restriction sites *BglII, Xho*I *and EcoR*I are shown in red. **Gel** shows a restriction analysis of the new construct pQIS 23. **Lane 1:** pQIS 23 digested with *Msp*I. (Since the plasmid *Msp*I sites are unmethylated, pQIS 23 cannot code for an active methyltransferase). **Lane 2:** pQIS 23 digested with *Bgl*II. (it verifies the presence of another *Bgl*II site, fragment size are 3144 and 921 bps) **Lane 3:** pQIS 18 digested with *Bgl*II. (Linear DNA, 3936 bps). **M:** GeneRuler™ 1 kb DNA ladder. (For GeneRuler™ 1 kb DNA ladder size, see Chapter Two and appendices). Map is not drawn to scale.

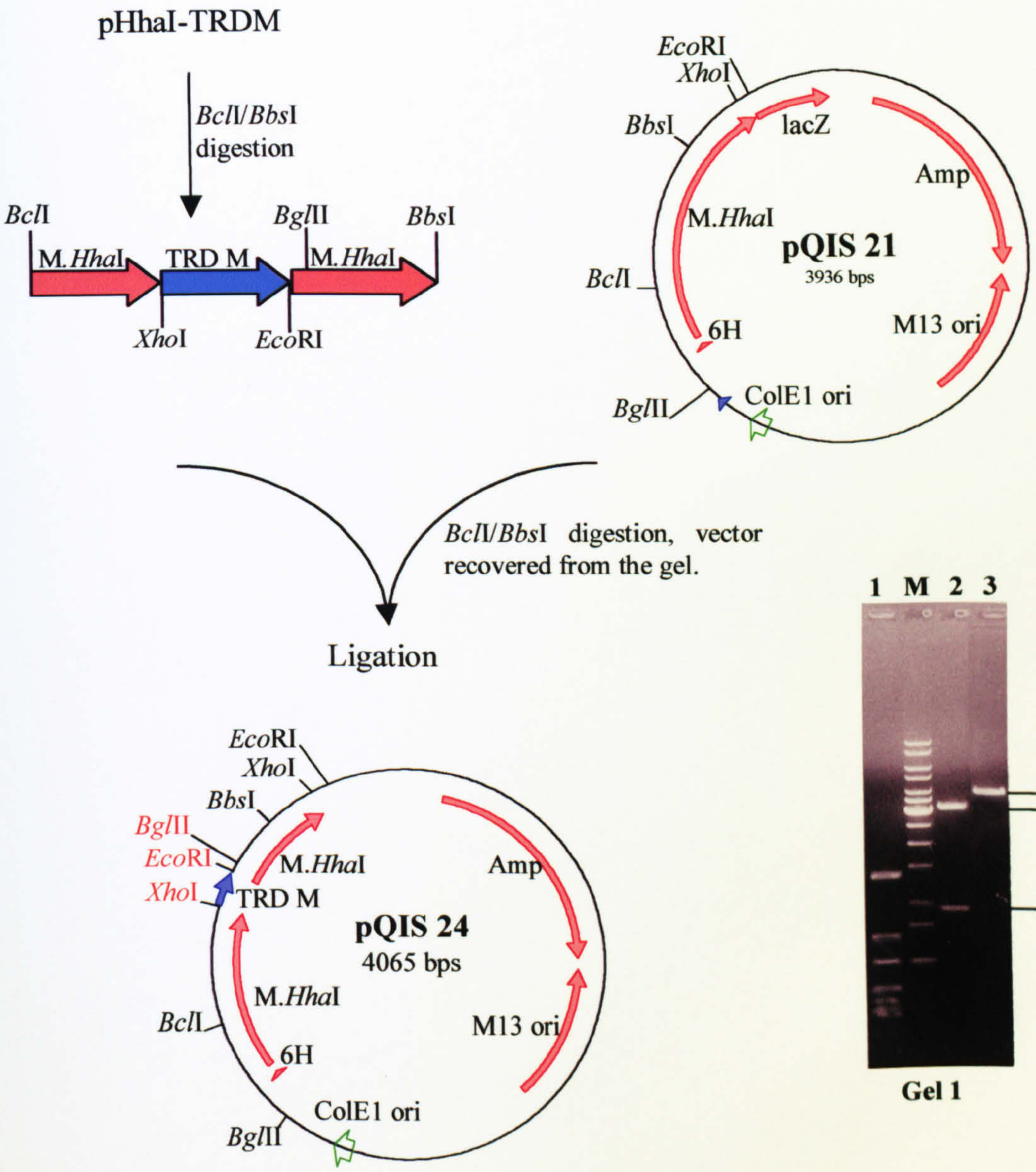


Figure 4.4: Schematic representation of the construction of pQIS 24. The *Msp*I specific TRD was cloned into pQIS 21 (encoding for Ala 280 Phe, Arg 277 Leu, Tyr 289 Asp & Phe 302 Ala mutations). New restriction sites of *BglII, Xho*I *and EcoR*I *are* shown in red. **Gel** shows a restriction analysis of new construct pQIS 24. **Lane 1:** pQIS 24 digested with *Msp*I. (Since the plasmid *Msp*I sites are unmethylated, pQIS 24 cannot code for an active methyltransferase). **Lane 2:** pQIS 24 digested with *Bgl*II. (To verify the presence of another *Bgl*II site, fragment size are 3144 and 921 bps) **Lane 3:** pQIS 21 cut with *Bgl*II. (Linear DNA, 3936 bps). **M:** GeneRuler™ 1 kb DNA ladder.

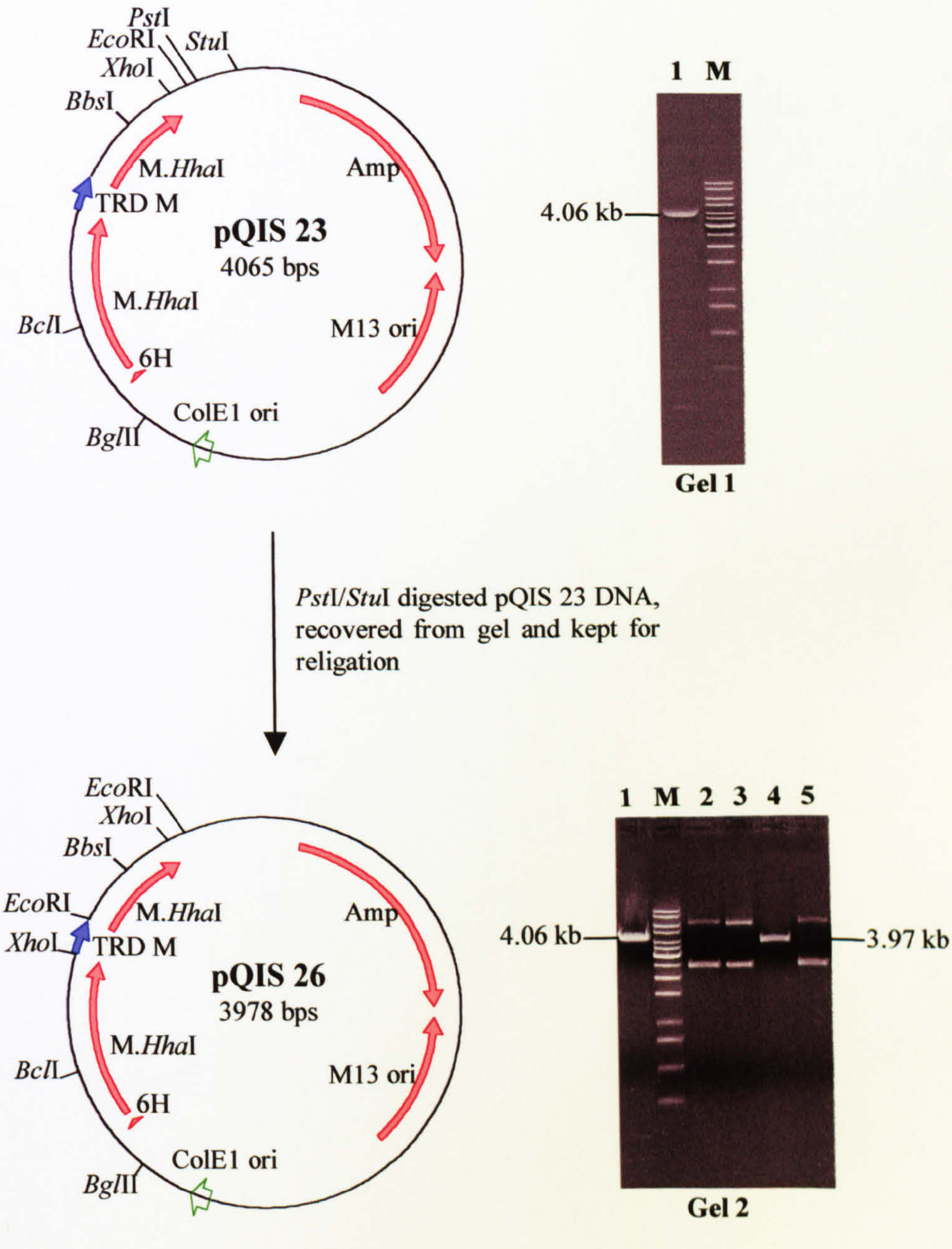


Figure 4.5: Schematic representation of the construction of pQIS 26 (Triple mutant + *Msp*I TRD) with deleted multiple cloning site. Gel 1:Lane 1, pQIS 23 cut with *Pst*I and *Stu*I and gel purified. **M:** GeneRuler^{TM} 1 kb DNA ladder. **Gel 2:** Restriction analysis for pQIS 26 to verify that the mutiple cloning site has been deleted. **Lane 1,** Control pQIS 23 digested with *Pst*I. **M:** GeneRuler^{TM} 1 kb DNA ladder. **2:** pQIS 26 incubated with *Pst*I (uncut DNA), **3:** pQIS 26 incubated with *Stu*I (uncut DNA). **4:** pQIS 26 digested with *Alw*NI (Linear DNA 3978 bps). **5:** pQIS 26 incubated with *Bam*HI (uncut DNA).

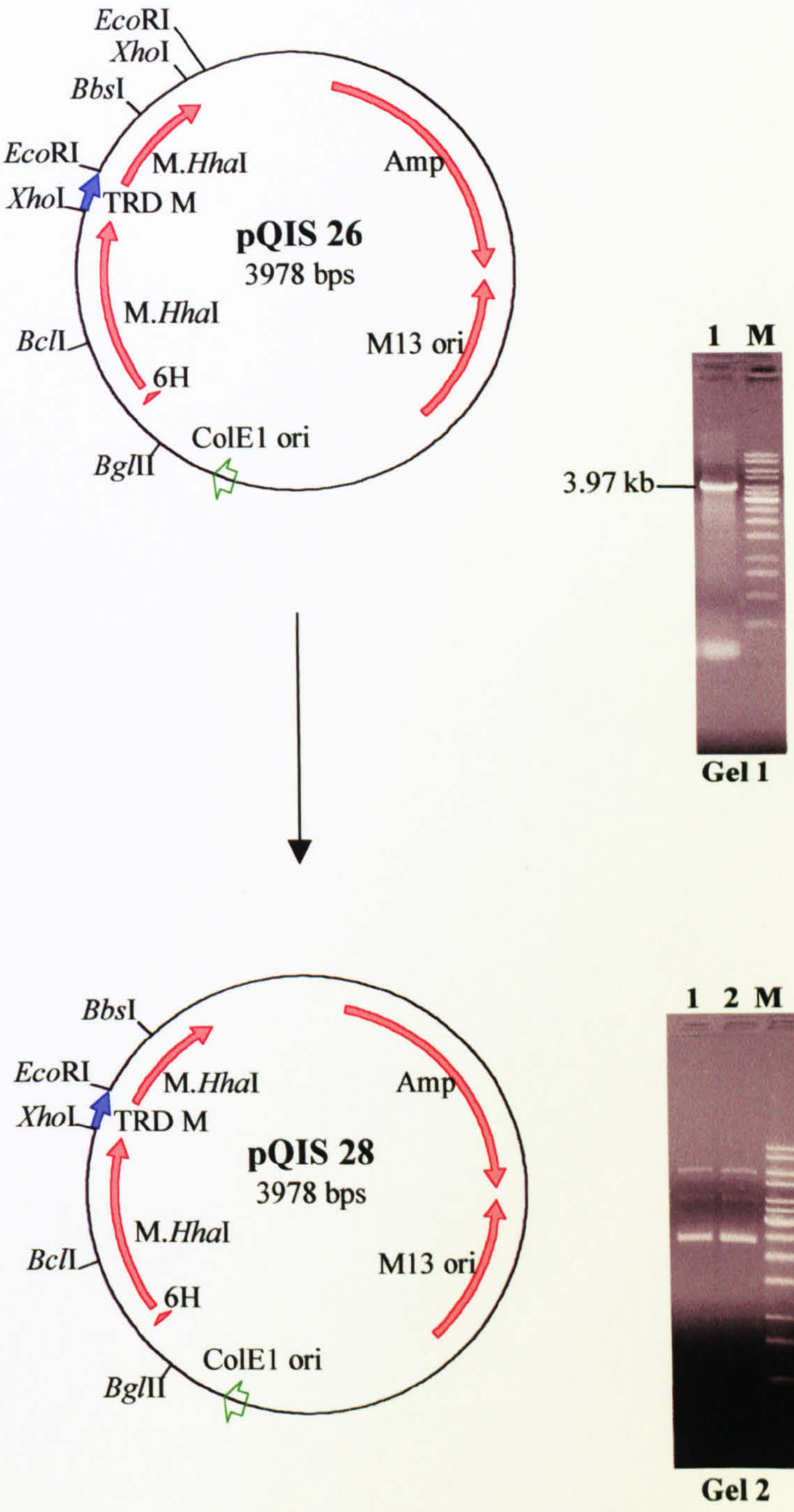


Figure 4.6: Schematic representation showing the deletion of the *Eco*RI and *Xho*I sites from pQIS 26. Gel 1:Lane 1: PCR product generated after site-directed mutagenesis using pQIS 26 as template. **Gel 2:** Restriction analysis for pQIS 28 to verify that the *Xho*I and *Eco*RI sites has been deleted. **Lane 1:**pQIS 28 incubated with *Eco*RI. (Uncut DNA) **2:** pQIS 28 incubated with *Xho*I (Uncut DNA). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

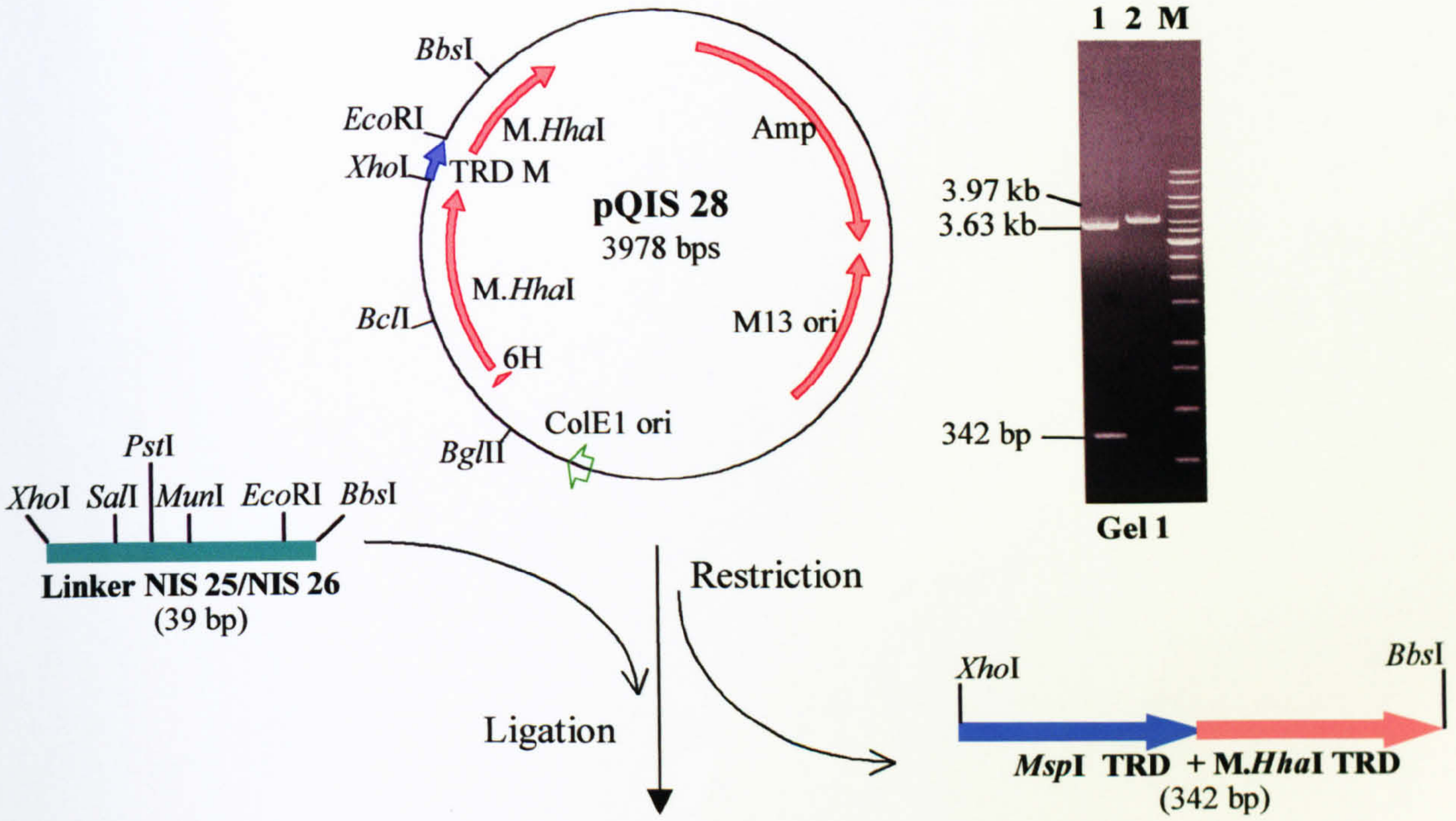
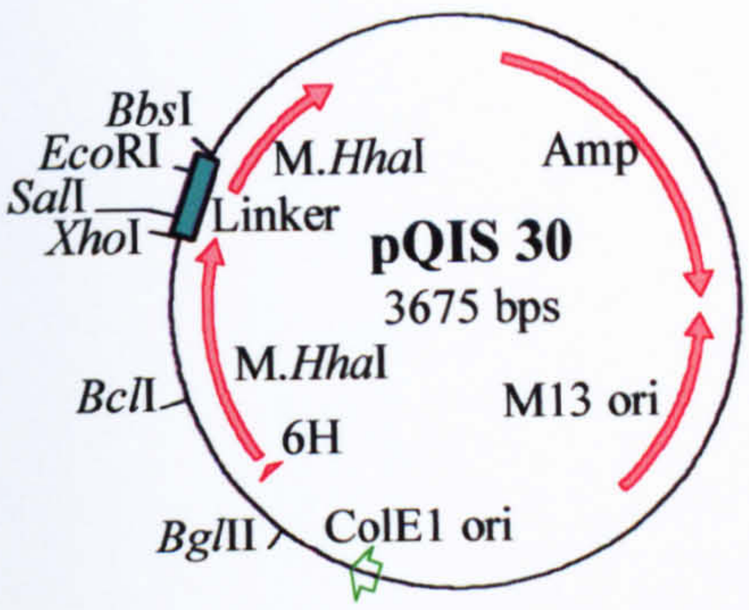
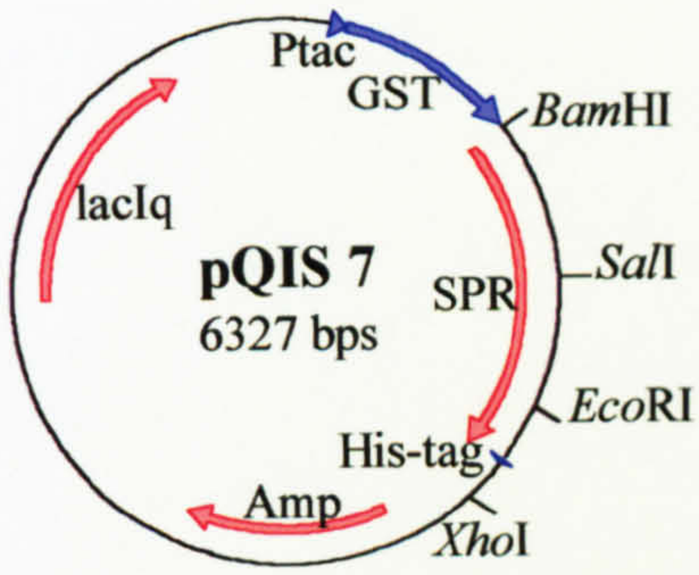


Figure 4.7: Schematic representation showing the construction of pQIS 30 (M.*Hha*I without any TRD region). Gel 1:Lane 1: pQIS 28 digested with *Xho*I and *Bbs*I (3636+ 342 bp) (for the gel purification of vector) **2:** pQIS 28 digested with *Alw*NI (linear 3978 bp) **M:** GeneRuler™ 1 kb DNA ladder **Gel 2:** Restriction analysis for the confirmation of pQIS 30. Presence of unique sites *Mun*I and *Sal*I confirmed the formation of new construct and the insertion of linker. **Lane 1:** pQIS 30 digested with *Mun*I. (Linear) **2:** pQIS 30 digested with *Sal*I (Linear) **3:** pQIS 30 digested with *Pst*I (Linear) **M:** GeneRuler™ 1 kb DNA ladder.





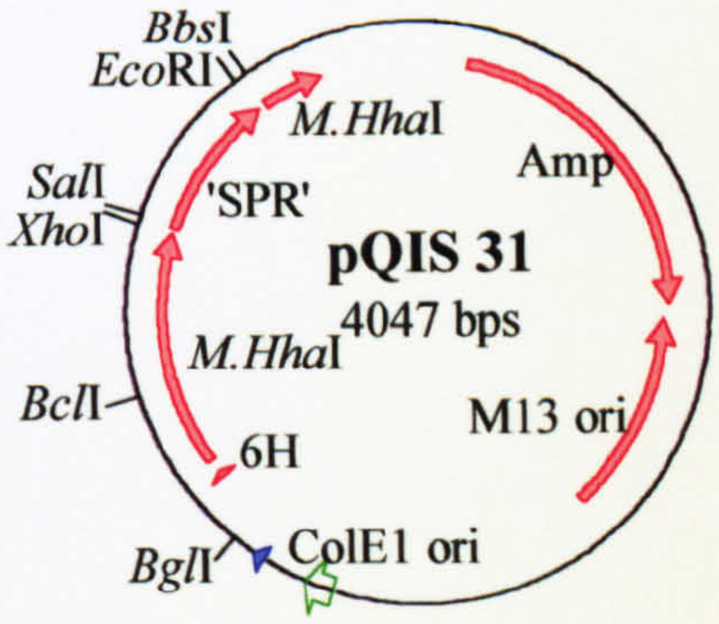
Vector

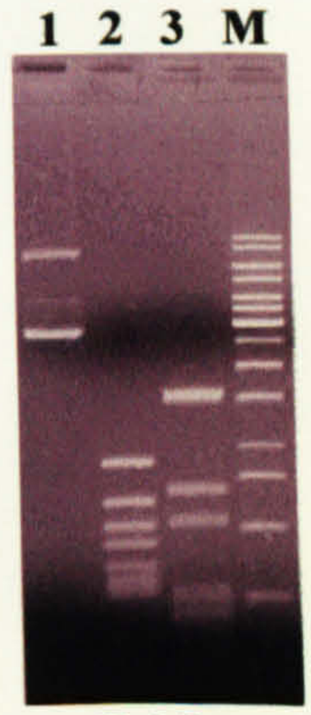
*Sal*I/*Eco*RI digestion, vector & 390 bps insert recovered from the gel.

Insert

Ligation

Gel 1

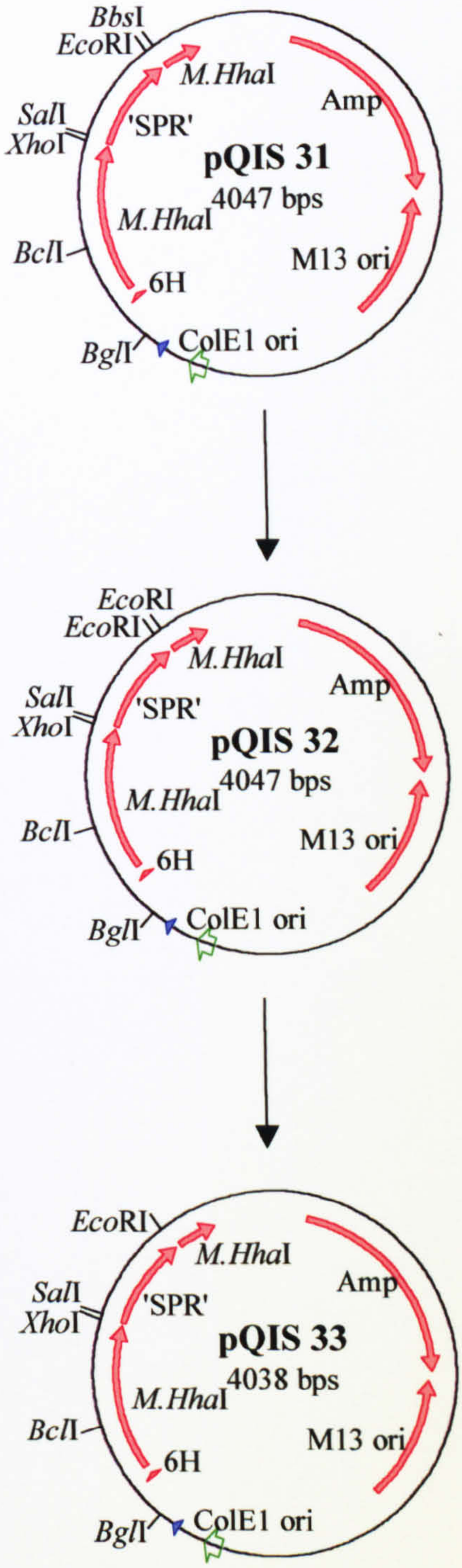




Gel 2

Gel 3

Figure 4.8: Schematic representation of the construction of pQIS 31 (swapping of M.*Hha*I TRD with three TRD encoding region "segment B" specific for M.SPRI).
Gel 1: Lane 1: pQIS 7 digested with *Eco*RI and *Sal*I to gel recover the insert (fragment size 5937 and 390 bp).
Gel 2: Lane 1: pQIS 31 digested with *Pst*I. (Uncut DNA signifies the removal of linker insert). **2:** pQIS 31 digested with *Hae*III (No protection, inactive). **3:** pQIS 31 digested with *Msp*I (No protection, inactive).
Gel 3: Lane 1: pQIS 31 digested with *Mun*I (Linear DNA, 4.0 kb). **M:** GeneRuler™ 1 kb DNA ladder in all three gels.



Gel 1

Gel 2

Figure 4.9: Schematic representation of the construction of pQIS 32 & pQIS 33 (M.*Hha*I with corrected motif IX encoding three TRDs specific for M.SPRI). Gel 1:Lane 1: PCR product generated after SDM for the introduction of a new *Eco*RI site 9 bp apart from the existing *Eco*RI site. **Gel 2: Lane 1:** pQIS 32 cut with *Bbs*I (Uncut DNA, due to abolition of *Bbs*I site). **Lane 2:** pQIS 31 cut with *Bbs*I and *Sty*I. (as control digestion, fragment size 2974 and 1073 bps). **Lane 3:** pQIS 32 cut with *Bbs*I and *Sty*I. (Only *Sty*I cut and linearized the DNA, 4047 bps). **Lane 4:** pQIS 32 cut with concentrated *Eco*RI for gel purification and religation to construct pQIS 33. **M:** GeneRuler™ 1 kb DNA ladder in both gels.

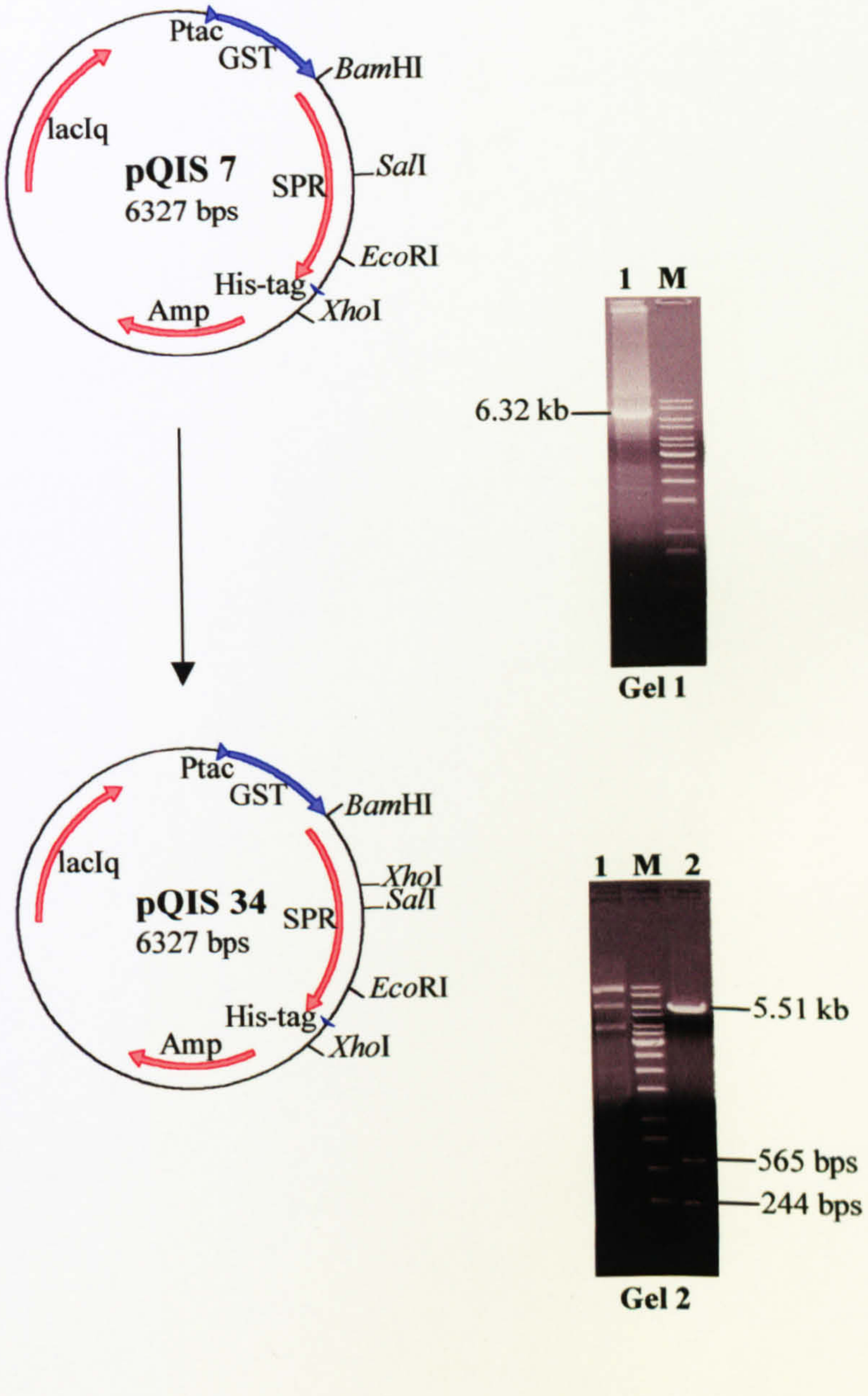
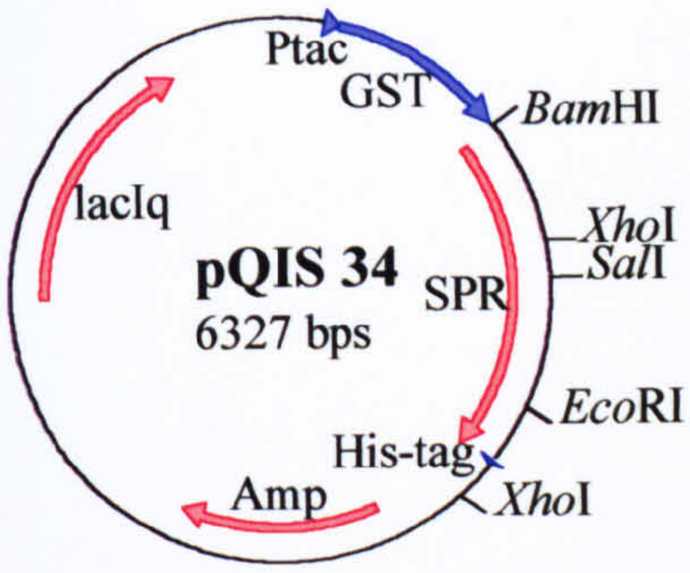
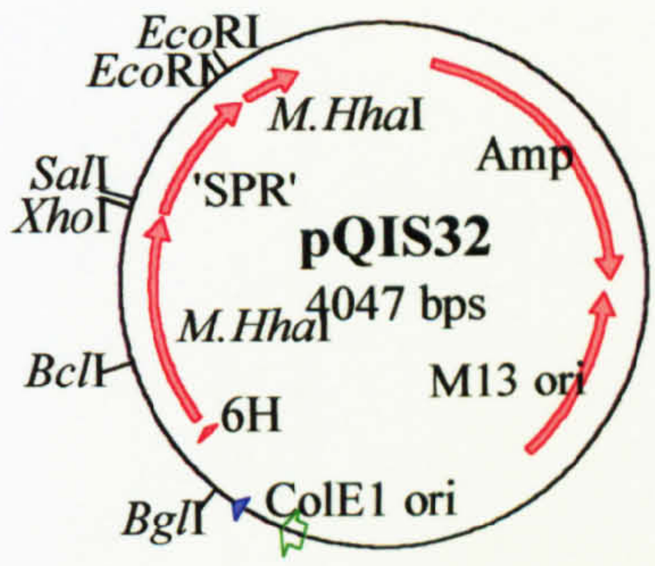
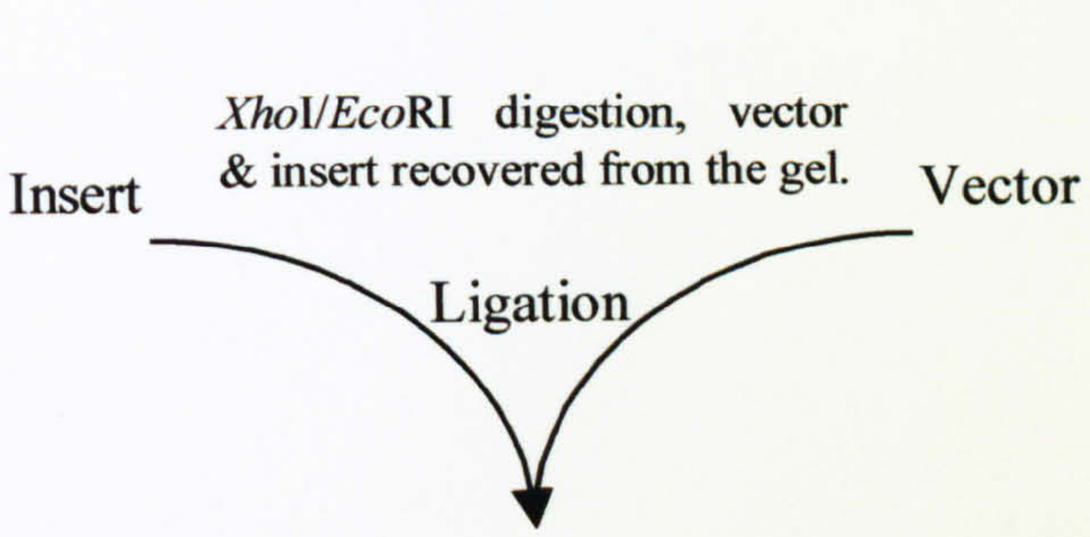
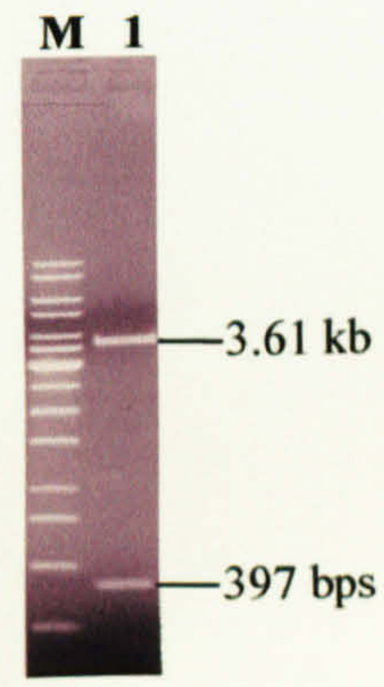


Figure 4.10: Schematic representation of the construction of pQIS 34. (This introduces Gln 189 Glu mutation in order to introduce a new *Xho*I site into wild type sequence of M.SPRI i.e. in pQIS 7). **Gel 1: Lane 1:** Shows the PCR product formed after site directed mutagenesis using pQIS 7 as template. **Gel 2: Lane 1**, pQIS 34 cut with *Msp*I. (Uncut DNA, construct pQIS 34 encodes for an active Mtase enzyme). **Lane 2:** pQIS 34 cut with *Xho*I and *Eco*RI. (Fragment size, 5518, 565 and 244 bps verifies the presence of new *Xho*I site). **M:** GeneRuler™ 1 kb DNA ladder in both gels.



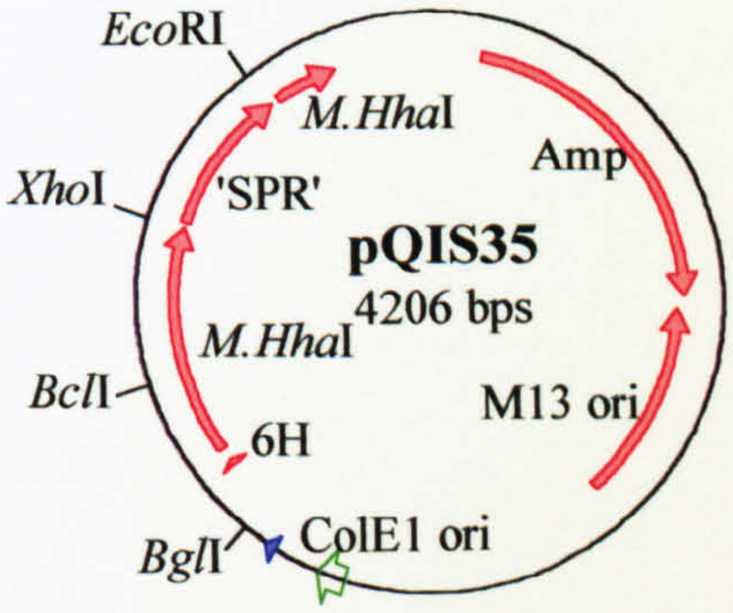


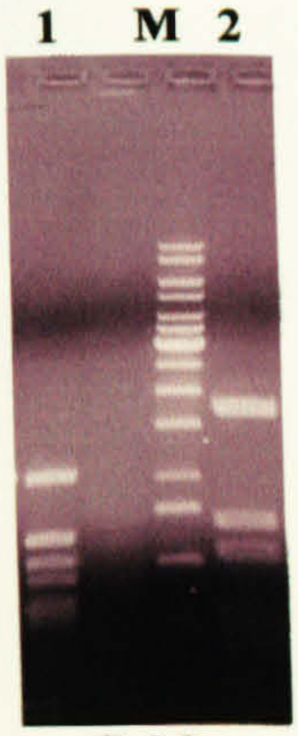




Gel 1

Gel 2





Gel 3

Figure 4.11: Schematic representation showing the construction of pQIS 35 (by swapping *Xho*I-*Eco*RI "segment C" from M.SPRI into redesigned triple mutant of M.*Hha*I). Gel 1, Lane 1: pQIS 32 cut with *Xho*I and *Eco*RI for the gel purification of vector DNA. (Fragment size, 3641, 397 and 9 bp). **Gel 2: Lane 1**, pQIS 35 cut with *Xho*I and *Eco*RI. (Fragment sizes, 3641 and 565 bps). The presence of 565 bp insert confirms the formation of pQIS 35. **Gel 3: Lane 1**, pQIS 35 cut with *Hae*III. **Lane 2**, pQIS 35 cut with *Msp*I. Digestion with *Hae*III and *Msp*I showed that the chimaeric protein encoded by construct pQIS 35 is inactive and unable to methylate its specific sites. **M:** GeneRuler™ 1 kb DNA ladder for all three gels.

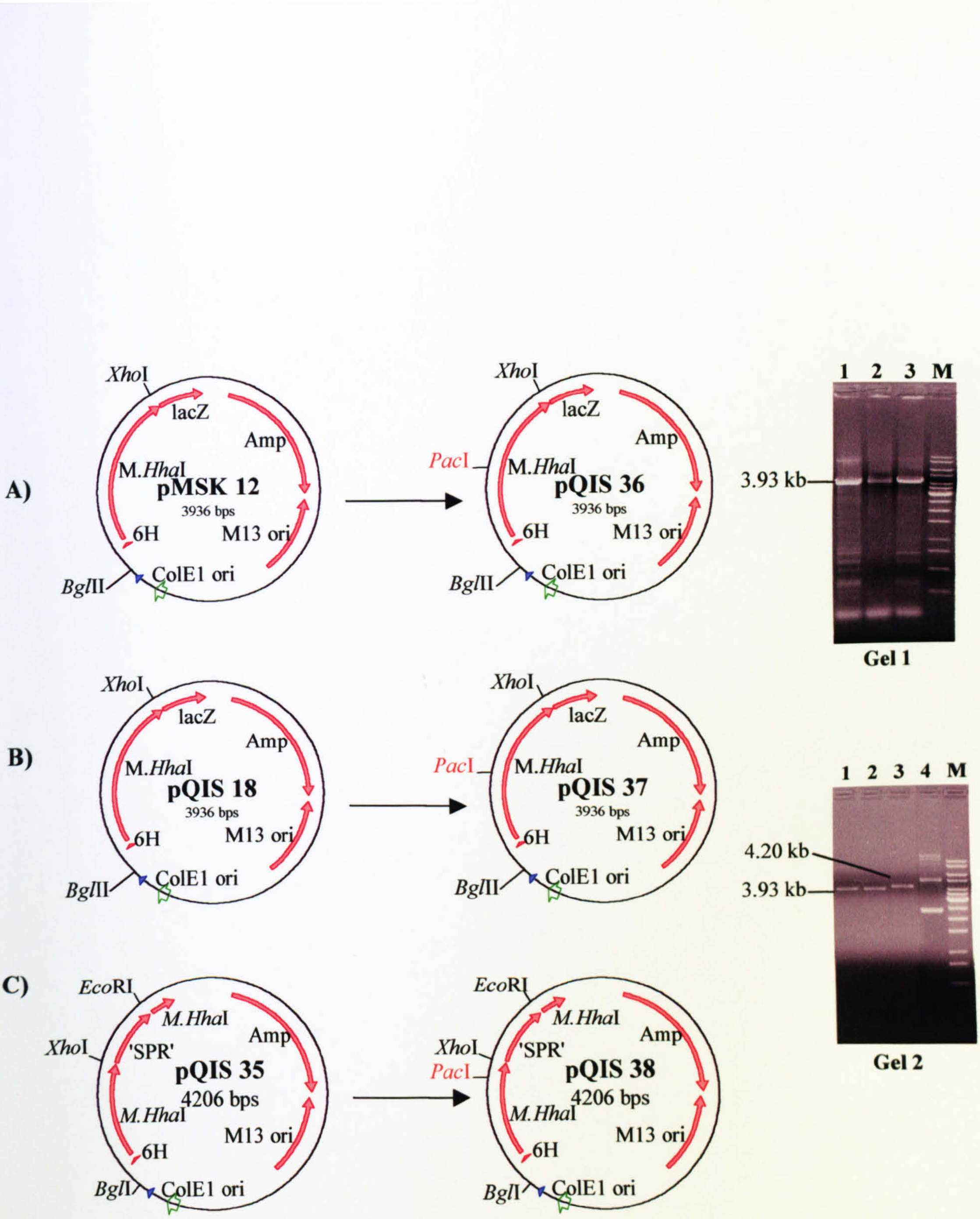


Figure 4.12: Schematic representation showing the construction of pQIS 36, pQIS 37 & pQIS 38.
A) Formation of pQIS 36 (encoding for Asn 176 Ile and Ile 177 in wild type M.H*ha*I). **B)** Formation of pQIS 37 (encoding for Asn 176 Ile and Ile 177 in triple mutant M.H*ha*I). **C)** Formation of pQIS 38 (encoding for Asn 176 Ile and Ile 177 in chimaeric construct pQIS 35). **Gel 1: Lane 1-3,** shows the PCR products generated after site directed mutagenesis for three plasmids. **Gel 2:** Restriction analysis for newly constructed plasmids pQIS 36, pQIS 37 and pQIS 38 that verifies the introduction of new *Pac*I site. **Lane 1**, pQIS 36 cut with *Pac*I (linear DNA 3936 bps, indicated the presence of unique restriction site *Pac*I). **2:** pQIS 37 cut with *Pac*I (linear DNA, 3936 bps) **3:** pQIS 38 cut with *Pac*I. (linear DNA 4206 bps) **M:** GeneRuler™ 1 kb DNA ladder in both gels.

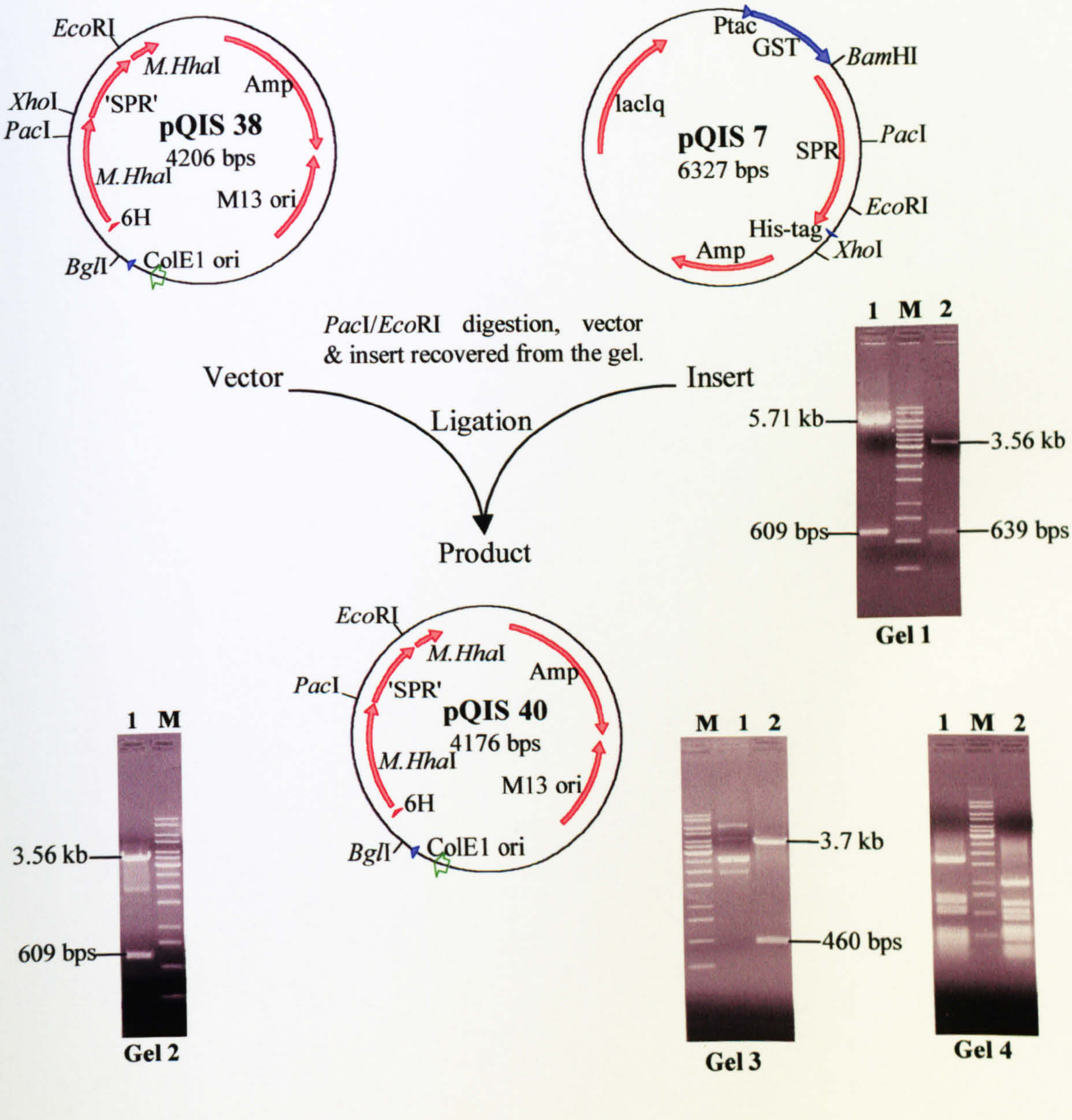


Figure 4.13: Schematic representation showing the construction of pQIS 40. This was achieved by swapping *Pac*I-*Eco*RI "segment D" from M.SPRI into redesigned triple mutant of M.*Hha*I. **Gel 1: Lane 1,** pQIS 7 cut with *Pac*I and *Eco*RI (5718 + 609 bp) (to gel purify the 609 bp insert). **2:** pQIS 38 cut with *Pac*I and *Eco*RI (3567 + 639 bps) (to gel purify the 3567 bp vector). **Gel 2: Lane 1,** pQIS 40 cut with *Pac*I and *Eco*RI. (Fragment sizes, 3567 and 609 bp) **Gel 3: Lane 1**, pQIS 40 cut with *Xho*I (Uncut DNA). Abolition of *Xho*I site confirms the new insert from pQIS 7. **2:** pQIS 40 cut with *Acc*I. (Fragments sizes, 3708, 460 and 8 bps). Another *Acc*I site is introduced as a result of swapping and also confirmed the insertion of "segment D" from M.SPRI. **Gel 4: Lane 1**, pQIS 40 incubated with *Msp*I. (Inactive). **2:** pQIS 40 incubated with *Hae*III. (Inactive). **M:** GeneRuler™ 1 kb DNA ladder in all gels.

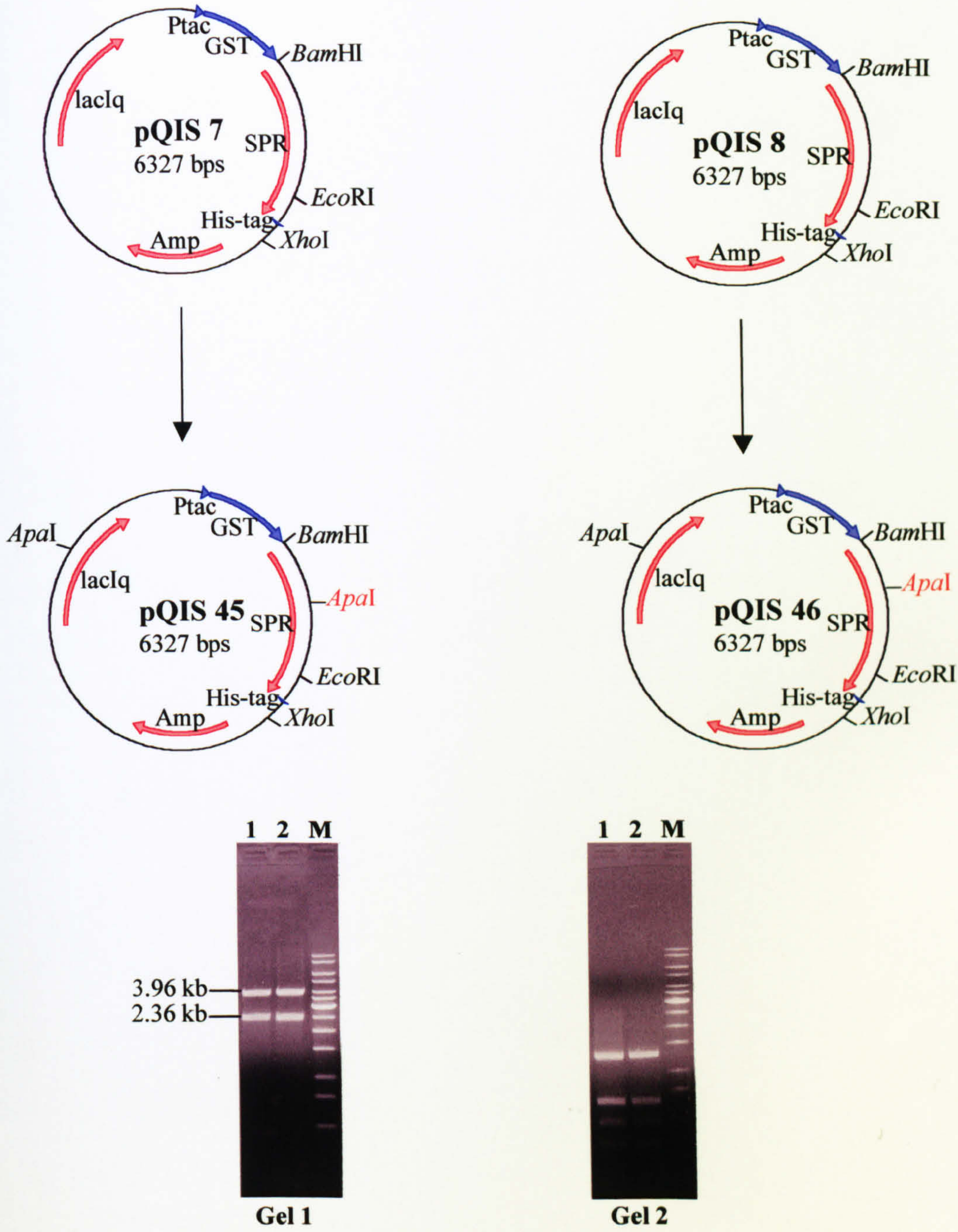
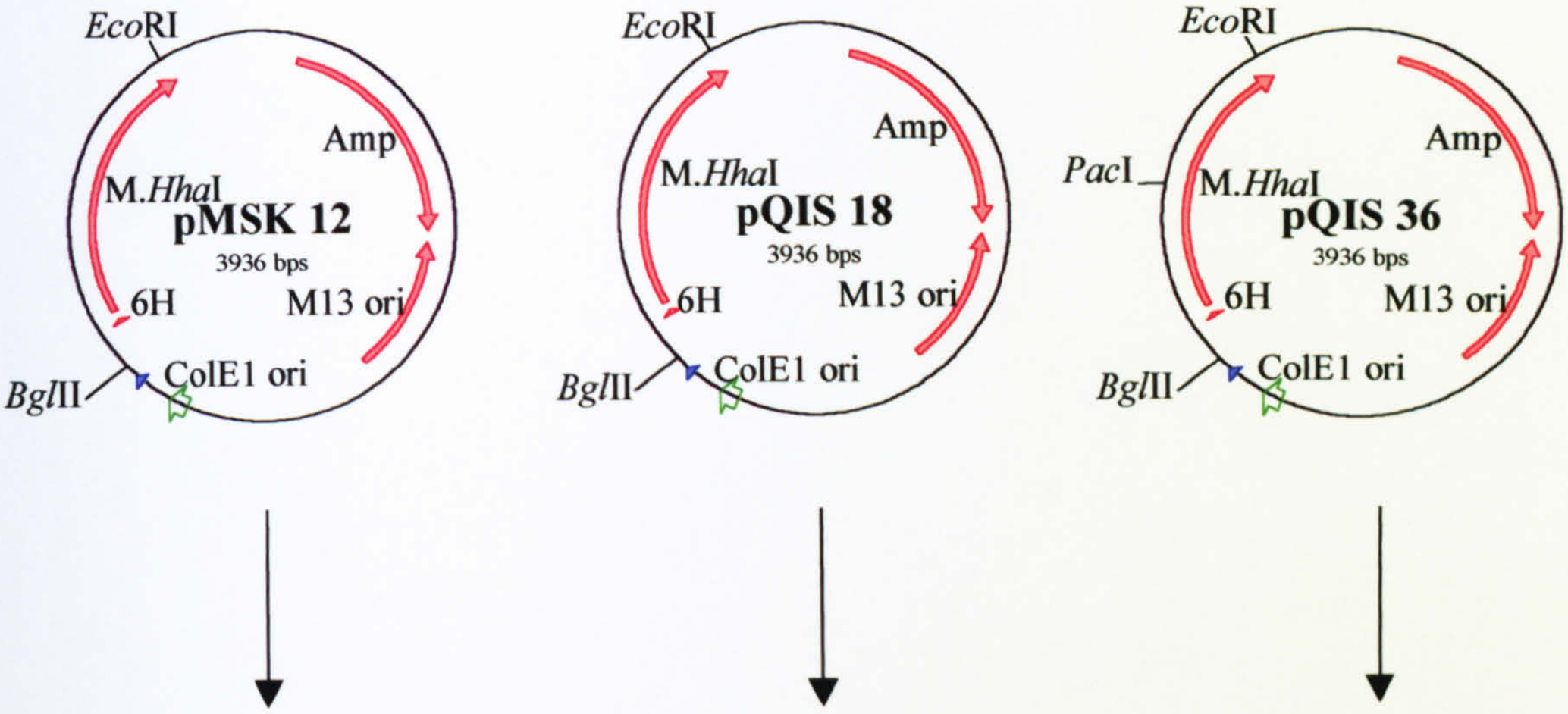


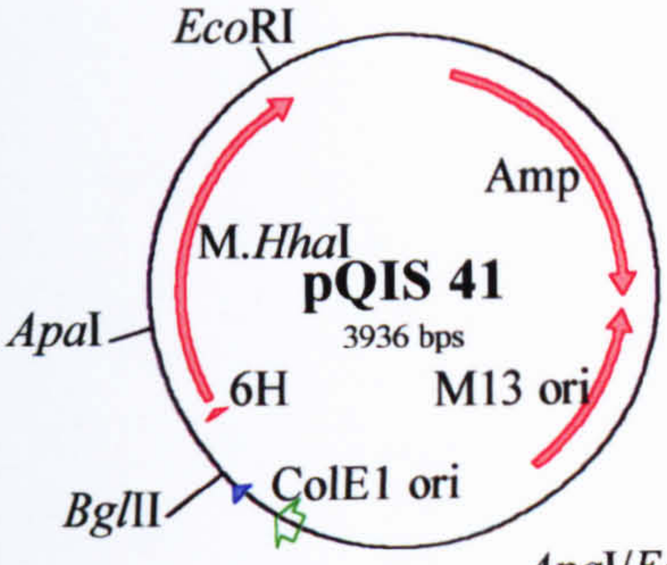
Figure 4.14: Schematic representation showing the construction of pQIS 45 and pQIS 46 (which encodes for Phe 112 Leu, Phe 113 Gly and Val 114 Pro mutation in the sequence of wild type M.SPRI pQIS 7 and mutant M.SPRI pQIS 8). This mutation also introduces *Apa*I site shown in red. **Gel 1: Lane 1,** pQIS 45 cut with *Apa*I. (Fragment size, 3962 and 2365 bp). **2:** pQIS 46 cut with *Apa*I. (Fragment size, 3962 and 2365 bp). **Gel 2: Lane 1**, pQIS 45 cut with *Msp*I. (Inactive). **2:** pQIS 46 cut with *Msp*I. (Inactive) **M:** GeneRuler™ 1 kb DNA ladder in both gels.

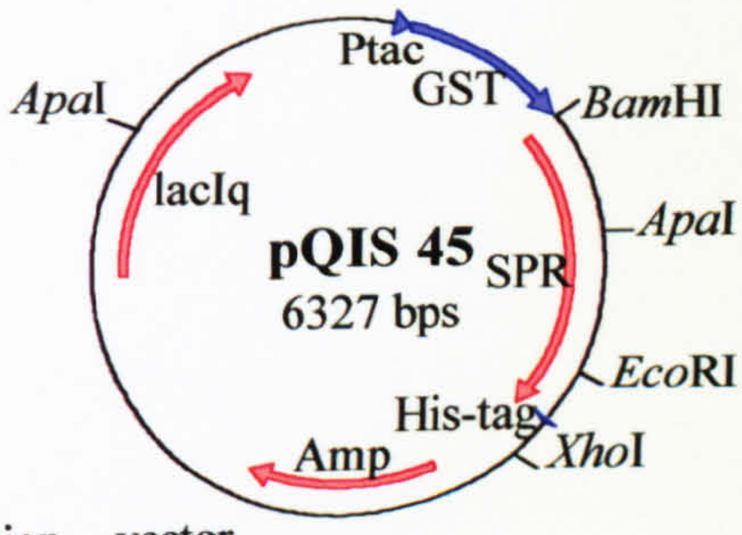


Gel 1

Gel 2

Figure 4.15: Schematic representation showing the construction of pQIS 41, 42 and pQIS 43.
A) Formation of pQIS 41 (which encodes for Val 116 Gly and Phe 117 Pro mutation in wild type M.*Hha*I). **B)** Formation of pQIS 42 (which encodes for Val 116 Gly and Phe 117 Pro mutation in triple mutant of M.*Hha*I) **and C)** Formation of pQIS 43 (which encodes for Val 116 Gly and Phe 117 Pro mutation in pQIS 36). This mutation also introduces *Apa*I site for diagnostic purpose shown in red. **Gel 1: Lane 1-3,** pQIS 41, pQIS 42 and pQIS 43 double digested with *Apa*I and *Eco*RI. (Fragment sizes 3296 and 640 bps confirms the presence of the newly introduced site *Apa*I). **Gel 2: Lane 1-3,** pQIS 41, pQIS 42 and pQIS 43 digested with *Hha*I. (the presence of completely degraded DNA indicates that there was no restriction protection and thus the encoded enzyme is inactive). **M:** GeneRuler™ 1 kb DNA ladder in both gels.





*Apa*I/*Eco*RI digestion, vector & insert recovered from the gel.

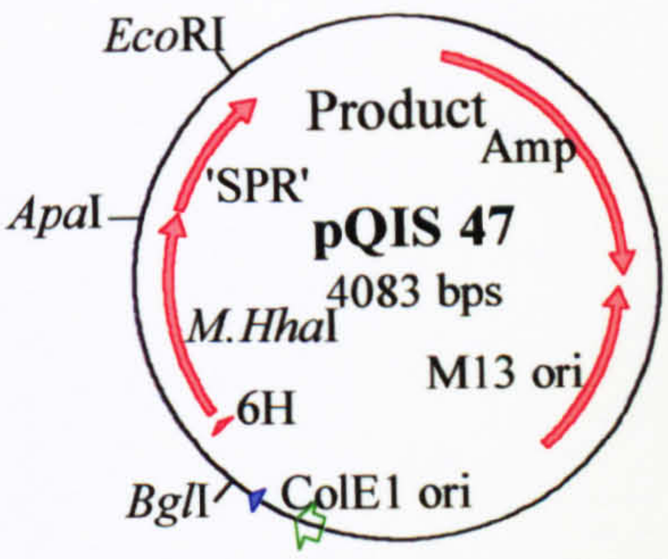
Vector

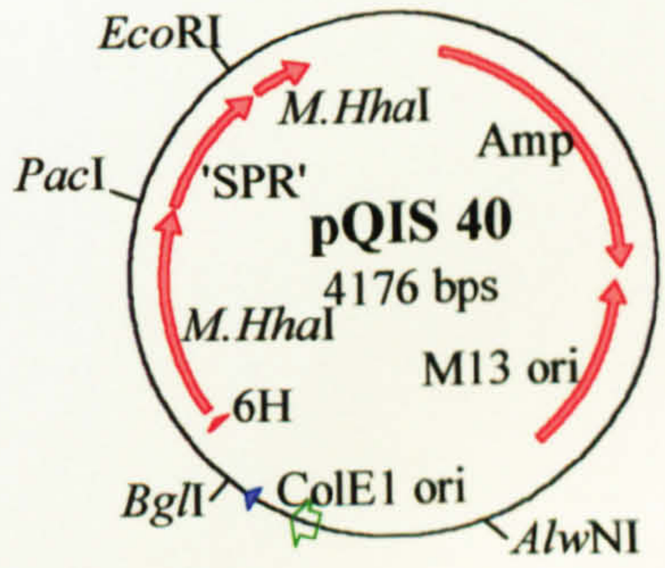
Insert

Gel 1

Ligation

Gel 2





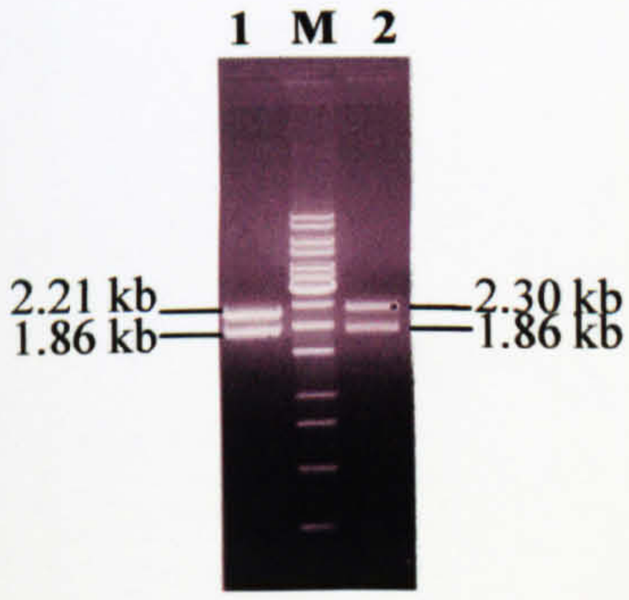
*Alw*NI/*Eco*RI digestion, vector & insert recovered from the gel.

Insert

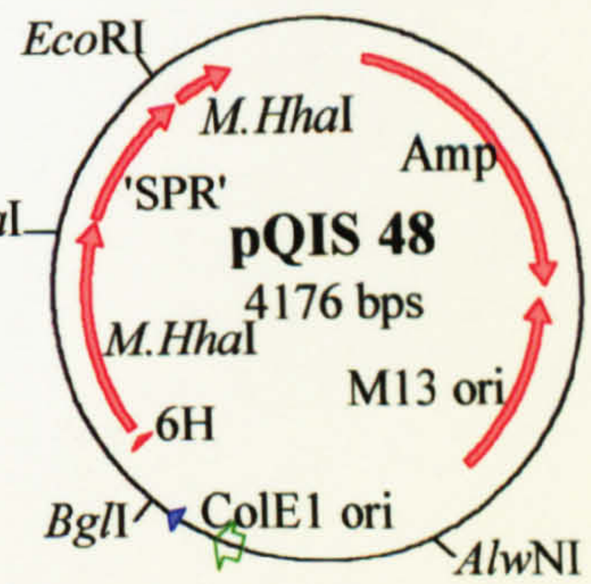
Vector

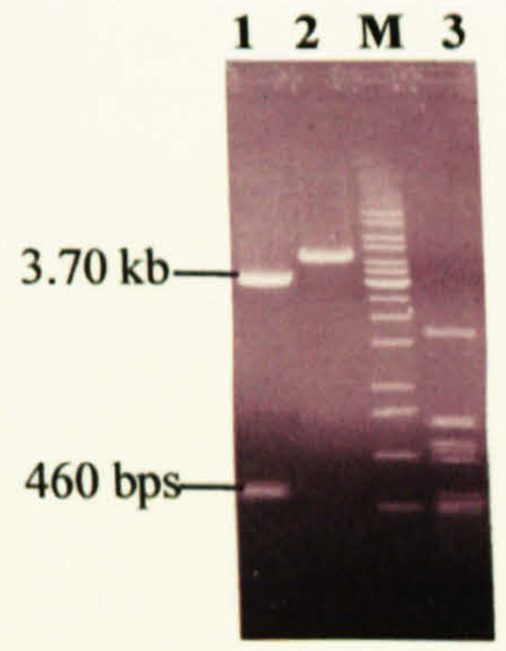
Ligation

Product



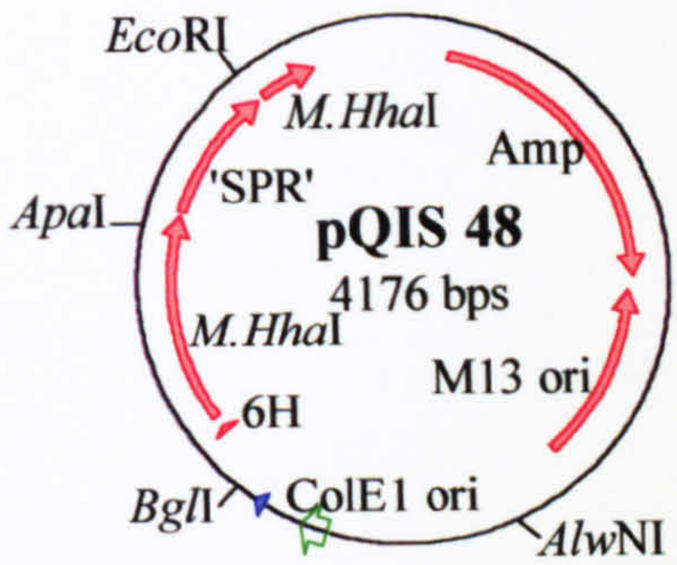
Gel 3

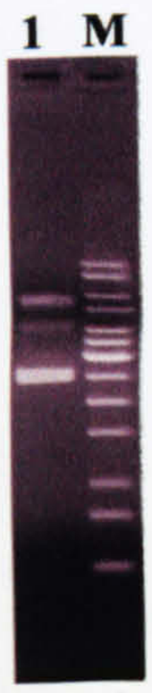




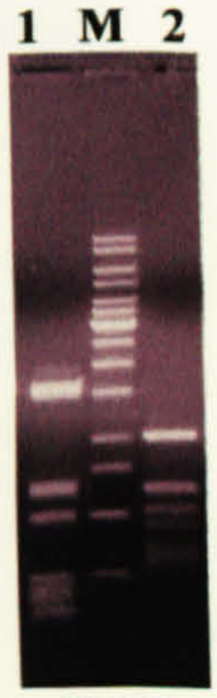
Gel 4

Figure 4.16: Schematic representation showing the construction of pQIS 47 and pQIS 48. The final chimaeric construct pQIS 48, contains M.SPRI "segment E" (figure 4.1) into the (redesigned triple mutant) M.*Hha*I. **Gel 1:** shows the digestion of pQIS 45 with *Apa*I and *Eco*RI to gel purify 787 bp insert sequence. (Fragment sizes are 3175, 2365 and 787 bps) **Gel 2:** shows the digestion of pQIS 47 with *Apa*I and *Eco*RI to verify the insert size.(3296 + 787 bps) **Gel 3: Lane 1** show the digestion of pQIS 47 with *Alw*NI and *Eco*RI to recover insert sequence. (Fragment sizes, 2215 and 1868 bps). **Lane 2** show the digestion of pQIS 40 with *Alw*NI and *Eco*RI to recover vector sequence. (Fragment sizes, 2308 and 1868 bps). **Gel 4: Lane 1** shows the digestion of pQIS 48 cut with *Acc*I for verification of this construct. (Fragment sizes are 3708, 460 and 8 bps). **Lane 2:** pQIS 48 cut with *Apa*I (Linear 4176 bps) for confirmation of this construct. **Lane 3:** pQIS 48 cut with *Msp*I (Inactive enzyme, DNA completely cleaved). **M:** GeneRuler™ 1 kb DNA ladder in all gels.



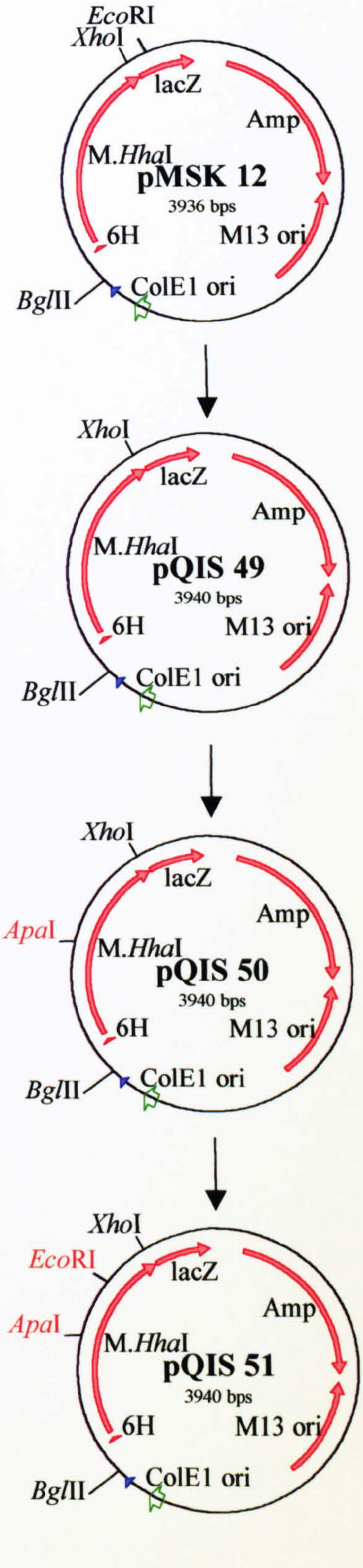


Gel 2



Gel 3

Figure 4.17: Schematic representation showing the construction of pQIS 52 from chimaeric construct pQIS 48. Gel 1: shows the PCR product generated after SDM using pQIS 48 as template. **Gel 2:** shows the digestion of pQIS 52 with ApaI. (Uncut DNA verifies the abolition of existing *Apa*I site and construction of pQIS 52). **Gel 3: Lane 1** shows digestion of pQIS 52 with *Msp*I. (DNA cleaved, therefore the plasmid encodes for inactive enzyme). **Lane 2** shows the digestion of pQIS 52 with *Hae*III. (DNA completely cleaved, thus the encoded recombinant protein is inactive). **M:** GeneRuler™ 1 kb DNA ladder in all three gels.



*Eco*RI site deleted by digestion of pMSK 12 with *Eco*RI and treatment with Vent Polymerase. The construct pQIS 49 encodes for wild type and an active M.*Hha*I.

SDM to create *Apa*I site in pQIS 49. The construct pQIS 50 encodes an inactive active M.*Hha*I.

SDM to create *Eco*RI site in pQIS 50. The construct pQIS 51 encodes an inactive active M.*Hha*I.

Figure 4.18: Schematic representation showing the construction of pQIS 51 from wild type M.*Hha*I (pMSK 12).

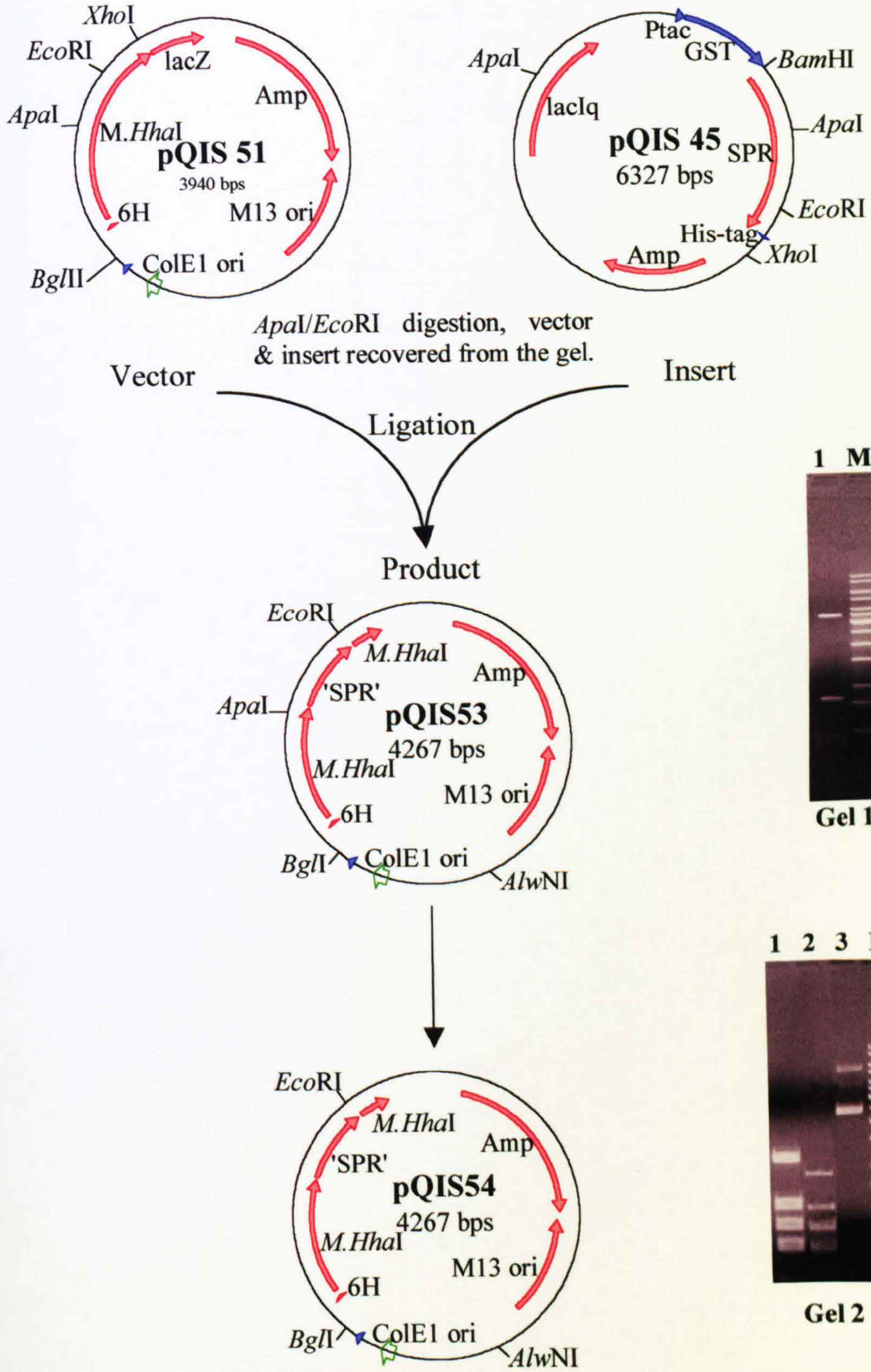
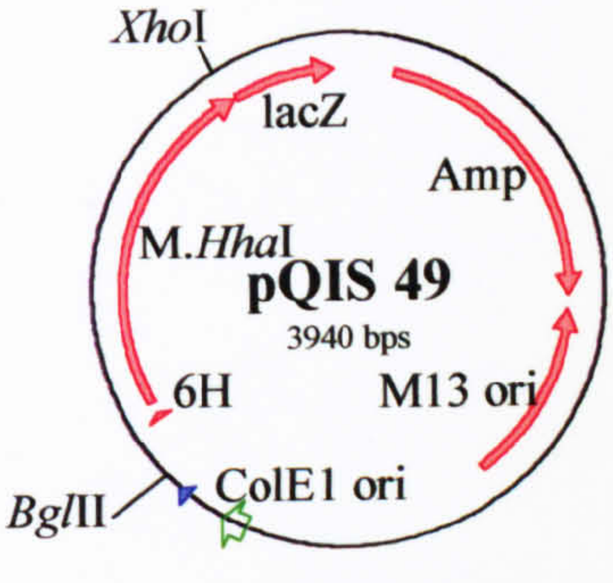
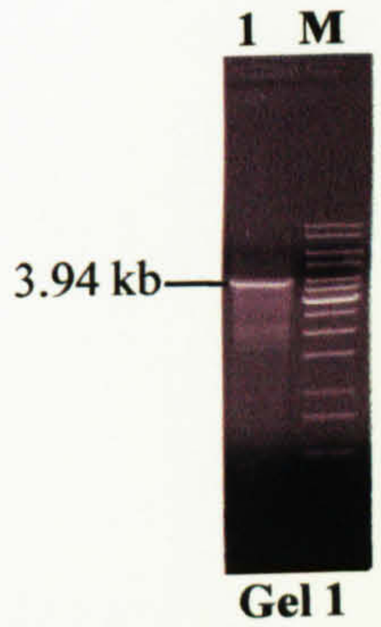
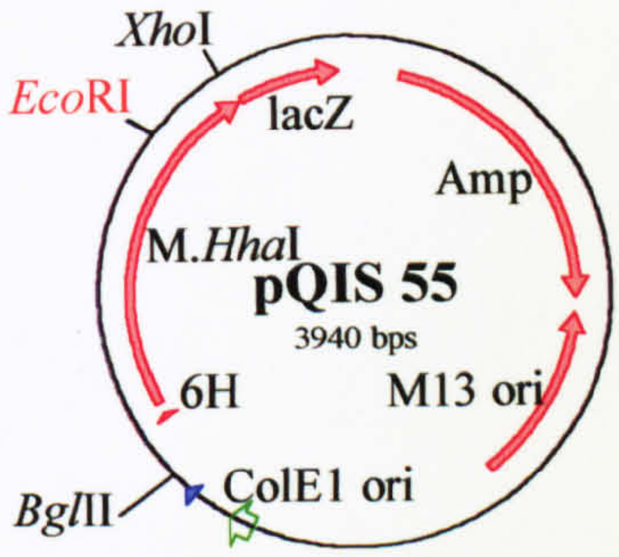


Figure 4.19: Schematic representation showing the construction of pQIS 53 and pQIS 54. This was achieved by introducing the *Apa*I-*Eco*RI fragment "segment E" (Figure 4.1) from M.SPR into wild type M.*Hha*I to generate pQIS 53. **Gel1, Lane 1**: pQIS 53 cut with *Apa*I and *Eco*RI. (Fragment sizes are 3480 abd 787 bps). Restoration of original codon of motif VI to Val and Phe as in M.*Hha*I generated pQIS 54. Gel 2, Lane 1: pQIS 54 cut with *Msp*I (Completely degraded DNA, Inactive chimaeric Mtase is encoded). Lane 2: pQIS 54 cut with *Hae*III (Completely digested DNA, inactive Mtase). Lane 3: pQIS 54 cut with *Apa*I. (Uncut DNA confirms the deletion of *Apa*I site).





Gel 1



Gel 2

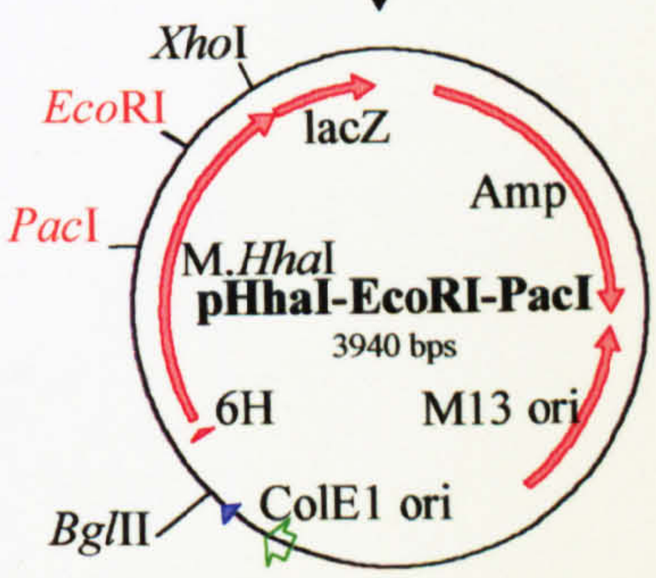


Figure 4.20: Schematic representation showing the construction of pQIS 55 and pHhaI-EcoRI-PacI. pQIS 55 is M.*Hha*I that encodes for K270R, T271I mutation and p*Hha*I-*Eco*RI-*Pac*I encodes for K270R, T271I + N176I, I177N mutation. **Gel 1:** shows the PCR product generated after SDM. **Gel 2: Lane 1** shows digestion of pQIS 55 cut with *Eco*RI. (Linear band verifying the introduction of new *Eco*RI site and formation of new construct pQIS 55). **2:** pQIS 49 cut with *Hha*I. (Uncut DNA thus encoded protein is active). **3:** pQIS 55 cut with *Hha*I. (Partial protection, thus encoded protein activity is reduced). **M:** GeneRuler^{TM} 1 kb DNA ladder in both gels.

CHAPTER FIVE

CONSTRUCTION OF CHIMAERIC DNA METHYLTRANSFERASES BY EXCHANGE OF MOTIFS

CHAPTER FIVE

CONSTRUCTION OF CHIMAERIC DNA METHYLTRANSFERASES BY EXCHANGE OF MOTIFS

5.1: Introduction

All prokaryotic DNA C-5 Mtases possess a common structural organisation. Comparison of the amino acid sequences of this class of enzymes has shown the presence of some highly conserved regions within a large background of seemingly unrelated sequences (Posfai *et al.*, 1984). It is reasonable to hypothesize that the conserved regions in such sequences represent sub-domains, which are involved in equivalent general biochemical steps in the methylation reaction in all such enzymes. Behrens *et al.* (1987) compared the primary structure of M.ρ11s with the amino acid sequences of two other multi-specific DNA Mtases from the temperate *Bacillus subtilis* bacteriophages SPR and φ3T. From this analysis, it was proposed that several regions with sequence diversity evoke specificity-recognising domains. Lauster *et al.* (1989) concluded that four homologous regions occur in the same order within the core of 11 prokaryotic enzymes, and designated them as core elements (CEs) I, II, III and IV (reading from the NH_2 terminus to COOH terminus). Motifs I-X, containing an invariant amino acid sequence, are found in the same order in all enzymes of this class except *Aqu*I (Posfai *et al*, 1989). Among these ten conserved motifs there are six most strongly conserved motifs (I, IV, VI, VIII, IX & X) shown in Figure 5.1. These motifs form a rigid structural framework separated by a flexible connecting loop region (Cheng *et al.*, 1993). Therefore it has long been proposed that the conserved sequences are involved in the biochemical steps of the methylation reaction while the variable region represents the target recognition domains.

Attempts to construct functional chimaeric enzymes were found to be successful after the introduction of unique restriction sites at equivalent locations within the gene encoding the two multi-specific methyltransferase M.SPRI and M.φ3T, followed by the swapping of these regions (Balganesh *et al.*, 1987). The chimaeric constructs were found to express a Mtase with an activity comparable to that of the

M*.SPR *(multi-specific MTase)

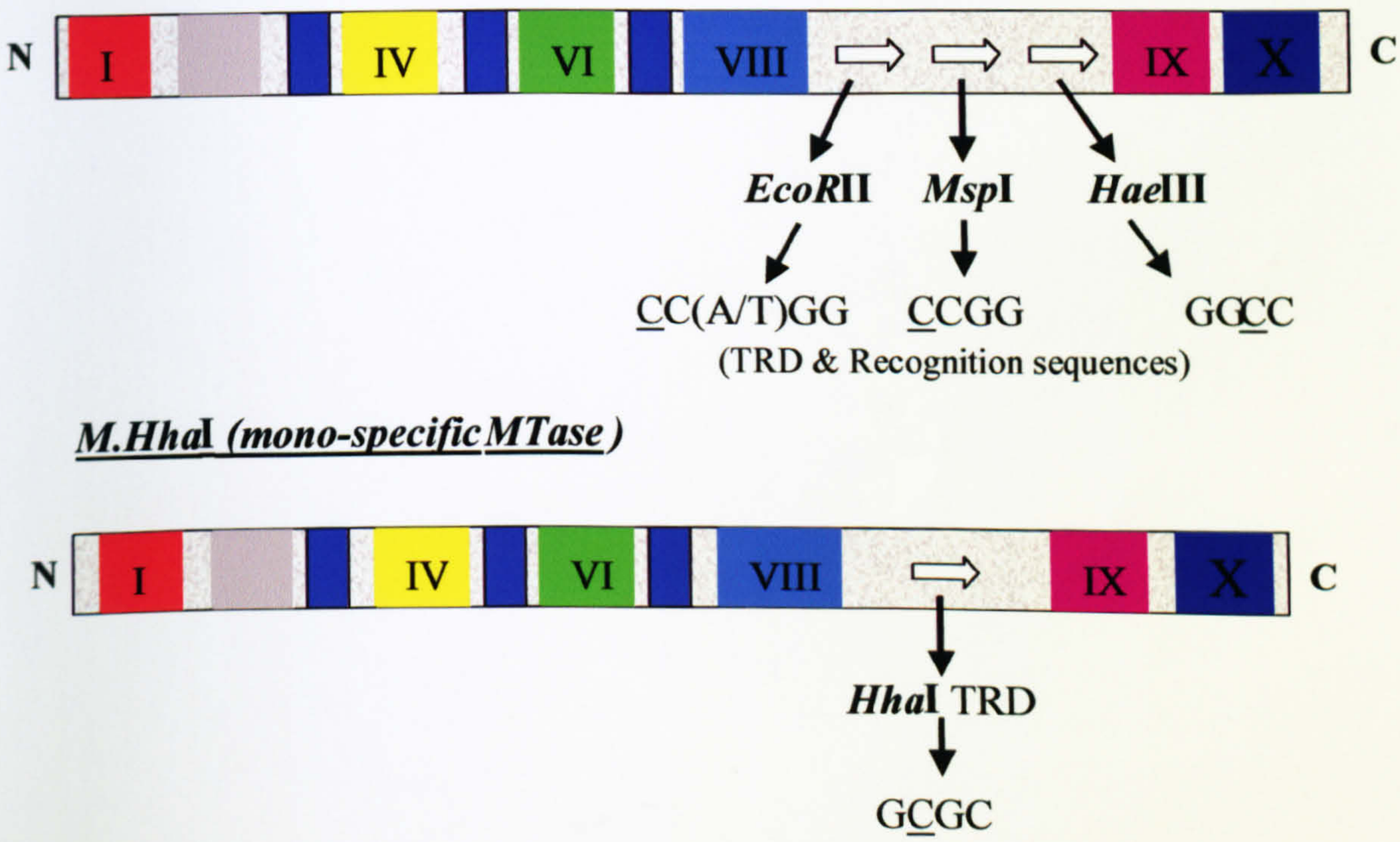


Figure 5.1: Schematic representation of the primary structure of mono and multispecific methyltransferases. Six of the most strongly conserved structural motifs of the methyltransferases (I, IV, VI, VIII, IX & X) are shown. Colour scheme is as shown in Figure 1.2. Target recognition sequences (TRDs) for both multispecific MTase SPR and *Hha*I are shown with the cytosine that can be methylated is underlined.

parental M.SPRI, and to that of M.φ3T. These experiments revealed the compatibility of the core elements of different bacteriophage enzymes. Target recognising domains from the Mtases SPR and φ3T have also been combined with those of $\rho 11_S$ to give enzymes with novel combinations of target recognition properties (Trautner *et al.* 1988). However, all these experiments were carried out on these closely related *Bacillus subtilis* bacteriophage Mtases.

In spite of the conservation between C5-DNA Mtases, M.*Hha*I & M.SPRI have a different recognition sequence, and therefore different specificities. We have attempted to convert mono-specific Mtase into multi-specific Mtase in two steps. In the first step, this was attempted by changing motif IX of a mono-specific Mtase to resemble that of M.SPRI (a multi-specific Mtase), and this experiment was successful. This was believed to be necessary as direct structural interactions between the variable region and motif IX have been proposed (Mi and Roberts 1992). This conclusion is also confirmed by the mutagenesis experiments described here, in which mutation was engineered in motif IX followed by the creation of another mutation in the variable region; these changes make the enzyme inactive (see Chapter 4, section 4.3.5). In the second set of experiments attempts were made to insert the TRDs of M.SPRI and its amino-terminal flanking region in a range of constructs that extend to motif V. However, the redesigned M.*Hha*I would not tolerate any polypeptide insertion and consequently lost methylation activity. From these data two possible conclusions were drawn: the first was that the redesign of motif IX was sub-optimal. Secondly the structure of this mono-specific Mtase M.*Hha*I is not flexible enough to tolerate any insertions in the TRD region, in contrast to multi-specific Mtase M.SPRI, which can tolerate such insertions in the variable region (Walter *et al.,* 1992; Matin, 2000). To determine whether the mono-specific Mtase M.*Hha*I is flexible or not, we carried out motif exchange experiments and to see whether M.*Hha*I and M.SPRI can accept motifs from each other.

Aims

The initial aim for this experiment was to determine whether the conserved motifs I-IV is interchangeable between multi-specific and mono-specific Mtases as has been observed for bacteriophage Mtases (Balganesh *et al.*, 1987). It was hoped that this would reveal whether there was any compatibility at all between a mono- and

multispecific Mtase that would lead to accommodation of a region carrying motif-I and motif I-IV from Mtase M.*Hha*I within M.SPRI, and vice versa.

5.2: Engineering of chimaeric constructs by motif swapping

5.2.1: EXPERIMENTAL DESIGN

Motif swapping of M.*Hha*I and M.SPRI was carried out sequentially:

The first engineered protein contained an exchange of motif I, and the second construct contained the region from motif I to motif IV. Finally reciprocal chimaeras were constructed.

For swapping the motifs there were certain limitations; the main constraints were

i. The restriction site(s) introduced into each gene had to be unique site.

ii. Mutagenesis had to be neutral both with respect to the reading frame of the gene and the amino acid sequence of parental enzymes.

iii. The open reading frame (ORF) for the two plasmids had to be conserved.

All of these limitations were considered and restriction sites were chosen at appropriate locations for the swap of motifs as described in the following section.

5.2.2: REPLACEMENT OF MOTIF I OF M.*Hha*I WITH MOTIF I OF M.SPRI AND VICE VERSA

The primary structures of two enzymes M.*Hha*I and M.SPRI were aligned from motif I to motif IV (shown in Figure 5.2). A stretch of 13 amino acid residues is also shown below in which the conserved amino acid residue Glu of motif II is shown in red. It was decided to introduce a *Nru*I restriction enzyme site into the gene encoding M.*Hha*I just before the conserved codon for Glu, by mutating Asn at position 39 into Arg. The decision to change the uncharged polar residue Asn at position 39 into a charged (basic) hydrophilic residue did not represent an ideal solution. However, since there was no choice it was decided to mutate the Arg amino acid residue back to the original Asn amino acid if required, after the desired swap had been completed.

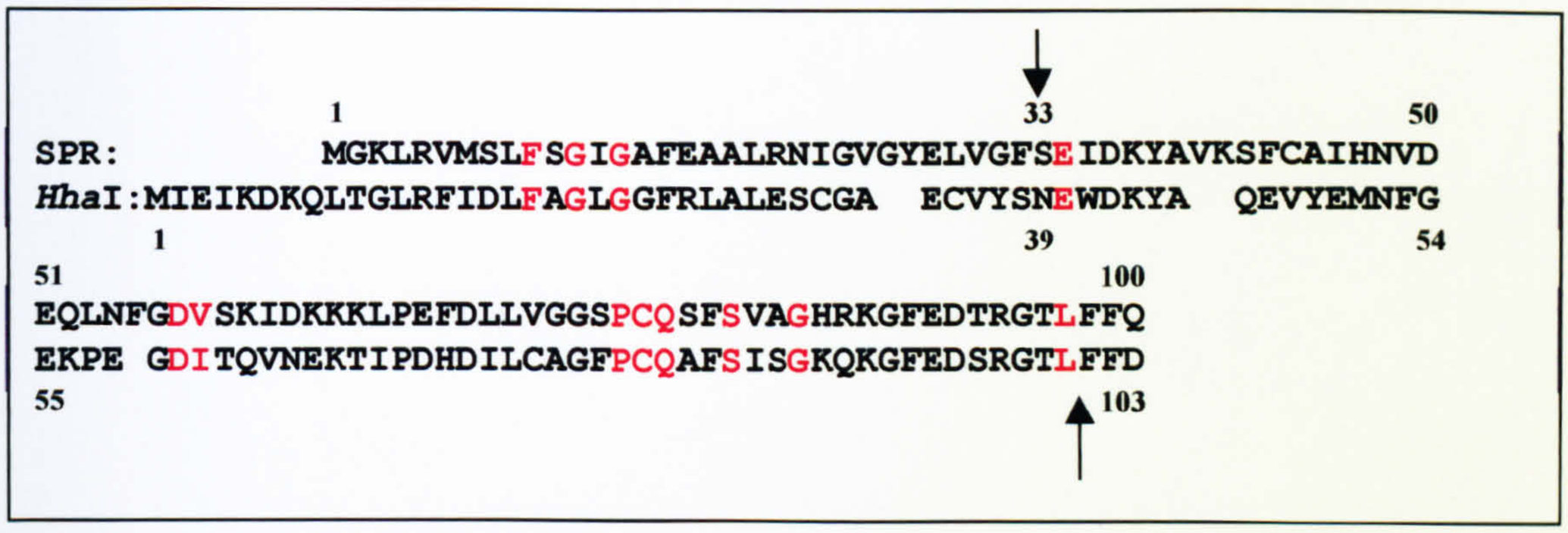


Figure 5.2: Alignment of M.SPR and M.*Hha*I amino acid sequences from motif I to motif IV. The conserved amino acids are shown in red. Arrows represent the point where the gene was cleaved.

	34	39	46
M.*Hha*I:	ECVYSN**E**WDKYAQ		
M.SPRI:	ELVGFS**E**IDKYAV		
	28	33	40

5.2.2.1: Introduction of a unique *Nru*I site into pQIS 64 (which encodes wild type M.*Hha*I)

In M.*Hha*I amino acid residue Asn 39 was chosen for mutation into Arg (for the introduction of a *Nru*I site). Two mutagenic oligonucleotides NIS 51 and NIS 52 were synthesized for this codon change as shown below, with the mutation shown in red.

NIS 51 / NIS 52 duplex mutagenic oligonucleotide:

*Nru*I

5′-GCTGAGTGCGTTTATTC**TCGCGA**ATGGGATAAATATGC-3′

3′-CGACTCACGCAAATAAG**AGCGCT**TACCCTATTTATACG-5′

A E C V Y S R E W D K Y

NIS 51: 5′-GCTGAGTGCGTTTATTC**TCGCGA**ATGGGATAAATATGC-3′

NIS 52: 5′-GCATATTTATCCCAT**TCGCGA**GAATAAACGCACTCAGC-3′

The primers NIS 51 and NIS 52 were used for site-directed mutagenesis and pQIS 64 was used as the template DNA. The correct construct formed was identified by the presence of a new *Nru*I site (Figure 5.3, and related figures for the plasmids sketches with gel diagrams are placed at the end of the chapter). The plasmid encoding the mutant M.*Hha*I (Asn 39 Arg) was called pQIS 74 and the protein encoded by this mutant was found to be inactive. The region specific interactions of Asn 39 were analysed in the structural model of M.*Hha*I. This residue is close to the Oxygen of Asp 16, the Nitrogen of Phe 18, the Nitrogen of Ala 19 and the Oxygen of Gly 59 (Cheng *et al.*, 1993). The distance between these residues was less than 3.0 Å. It has also been noted that residues Asp 16, Phe 18, Ala 19 are present in β1 and β2 sheets, and run parallel to the position of Asn 39. In addition, Gly 59, which is present in the helix B that runs across the β1 and β2 sheets. Electrostatic interactions or hydrogen bonds also exist between S-adenosylmethionine (SAM) and Glu 40 and surrounding residues (Cheng *et al.*, 1993). It seems that the disturbance of these analogous interactions has caused a major destabilisation of the enzyme core structure, rendering the enzyme inactive.

5.2.2.2: Introduction of *Nru*I site in pQIS 7 (which encodes wild type M.SPRI)

In M.SPRI the serine residue at position 33 was chosen for mutation into Arg; the aim being to introduce a unique *Nru*I site into the gene. The oligonucleotides used for the mutation were NIS 53 and NIS 54. In this mutagenesis experiment, two bases were changed as shown below in red.

NIS 53 / NIS 54 duplex mutagenic oligonucleotide:

*Nru*I

5′-GAGCTGGTTGGTTT**TCGCGA**GATTGATAAATATGCC-3′
3′-CTCGACCAACCAAA**AGCGCT**CTAACTATTTATACGG-5′
E L V G F **R** E I D K Y A

NIS 53: 5′-GAGCTGGTTGGTTT**TCGCGA**GATTGATAAATATGCC-3′

NIS 54: 5′-GGCATATTTATCAATC**TCGCGA**AAACCAACCAGCTC-3′

The template DNA used for this experiment was pQIS 7 and the construct generated as a result was pQIS 76. The identity of the new construct was confirmed by the presence of a new unique *Nru*I site (Figure 5.4). When tested for its activity, it was revealed that the protein encoded by the mutant of M.SPRI (Ser 33 Arg) had lost its ability to methylate (Figure 5.4). It appears that the instability in the enzyme core structure, which might be similar to that found in M.*Hha*I may occur here as well, and may therefore explain the enzyme's inactivity.

5.2.2.3: Exchange of motif I from M.*Hha*I the corresponding region in M.SPRI

The plasmids pQIS 74 and pQIS 76 were cleaved with *Apa*I and *Nru*I, and the insert and vector DNA were recovered from the gel as shown in Figure 5.5. They were ligated prior to transformation of *E. coli* and the resulting colonies were screened for the presence of the constructs carrying the insert from the M.*Hha*I. Restriction analysis with *Eco*RI confirmed the presence of the insert. There is a single *Eco*RI site in pQIS 76 and pQIS 74 also carries an *Eco*RI site before the start of M.*Hha*I gene. Therefore, the discovery of the two expected *Eco*RI fragment sizes confirmed the identity of pQIS 78 (Figure 5.5). The activity of the protein encoded by the plasmid pQIS 78 was checked by *Msp*I cleavage and it was shown to be inactive.

5.2.2.4: Arg 39 Asn mutagenesis in pQIS 78 (which encode motif I of M.*Hha*I and motif II to X of M.SPRI)

As noted above the mutation of Asn 39 to Arg leads to inactivation of the enzyme during the synthesis of pQIS 74. The Asn 39 lies in the (β1-αA-β2) structural segment that forms part of the binding site for SAM contains motif I, which is highly conserved among C5-Mtases (Cheng *et al.*, 1993). Therefore, it was decided to restore the codon sequence so that it contained the original wild type amino acid. The mutagenic primers NIS 55 and NIS 56 were synthesized to engineer this mutation in pQIS 78, which also deleted the *Nru*I site. The sequence of the oligonucleotides is shown below with the mutation shown in red and the deleted *Nru*I restriction site in bold.

NIS 55 and NIS 56 duplex mutagenic oligonucleotide:

```
                                  ΔNruI
5′-GCTGAGTGCGTTTATTCTAACGAGATTGATAAATATGCCGTC-3′
3′-CGACTCACGCAAATAAGATTGCTCTAACTATTTATACGGCAG-5′
     A  E  C  V  Y  S  N  E  I  D  K  Y  A  V
```

NIS 55: 5′-GCTGAGTGCGTTTATTCTAACGAGATTGATAAATATGCCGTC-3′

NIS 56: 5′-GACGGCATATTTATCAATCTCGTTAGAATAAACGCACTCAGC-3′

Site-directed mutagenesis generated the construct pQIS 82 carrying Arg to Asn mutation which was confirmed by the loss of the *Nru*I site. However, the enzyme did not regain methylation activity (Figure 5.6) confirming that the engineered protein containing motif I of M.*Hha*I and motif II to X of M.SPRI is not compatible, and unable to perform enzymatic activity.

5.2.2.5: Exchange of motif I from M.SPRI to M. *Hha*I

The plasmids pQIS 74 and pQIS 76 were cleaved with *Nru*I and *Xho*I, and the insert (871 bps) and vector DNA (5054 bps) were recovered from the gel as shown (Figure 5.7). The fragments were ligated together prior to transformation of *E. coli* and the resulting colonies were screened for the presence of the constructs carrying the insert from M.SPRI.

There is one *Eco*RI site in M.SPRI before motif IX and M.*Hha*I also carries an *Eco*RI site before the start codon. The final construct, pQIS 80, has lost both *Eco*RI sites, and this property was used to confirm its identity. Restriction analysis with

*Nru*I/*Xho*I also confirmed the presence of a 871 bps insert (Figure 5.7). The activity of the protein encoded by the chimeric construct was checked with *Hha*I cleavage and it was found that it was inactive. This experiment confirmed that that the engineered protein is incompatible.

5.2.2.6: Arg 33 Ser mutagenesis in pQIS 80

In order to maintain the sequence of M.SPRI up to the point of swap, it was decided to bring back the original amino acid Ser at position 33. A pair of oligonucleotide NIS 57 and NIS 58 was made to engineer this mutation of Arg to Ser in construct pQIS 80 and to delete *Nru*I site. The sequence of oligonucleotide is shown below with the mutation shown in red and the deleted *Nru*I site in bold.

NIS 57 and NIS 58 duplex mutagenic oligonucleotide:

```
                        ΔNruI
5′-GAGCTGGTTGGTTTTAGCGAATGGGATAAATATGCAC-3′
3′-CTCGACCAACCAAAATCGCTTACCCTATTTATACGTG-5′
     E   L   V   G   F   S   E   W   D   K   Y   A
```

NIS 57: 5′-GAGCTGGTTGGTTTTAGCGAATGGGATAAATATGCAC-3′

NIS 58: 5′-GTGCATATTTATCCCATTCGCTAAAACCAACCAGCTC-3′

The mutagenesis experiment generated pQIS 83 with one amino acid changed (Arg to Ser) and with the *Nru*I restriction site abolished (Figure 5.8). The chimaera (which contained motif I of M.SPRI and Motif II and onward from M.*Hha*I) was found to be inactive.

5.2.3: REPLACEMENT OF MOTIF I-IV OF M.*Hha*I WITH MOTIF I-IV OF M.SPRI AND VICE VERSA

The second experiment planned was to generate a pair of chimaeras by fusion between two Mtases M.*Hha*I and M.SPRI containing motif I-IV of one and the rest of the molecule from the other. A *Kpn*I restriction site already exists in plasmid pQIS 64 (which contain sequence of wild type M.*Hha*I). It was decided to use this site as a fusion point to construct the planned chimaeras. However, there is a need to create a similar site at same position in the plasmid pQIS 7 carrying the M.SPRI gene. The primary structure of M.*Hha*I and M.SPRI is shown in Figure 5.2 and an arrow represents the point of interchange.

5.2.3.1: Introduction of a *Kpn*I site into pQIS 7

The primers used for the introduction of *Kpn*I site in pQIS 7 (containing the sequence for wild type M.SPRI) are shown below. This represents a silent mutation and the resulting construct pQIS 72 encodes an active Mtase enzyme M.SPRI (Figure 5.9). The construct pQIS 72 was confirmed by the presence of unique *Kpn*I site.

NIS 47 and NIS 48 duplex mutagenic oligonucleotide:

*Kpn*I

5′-GAAGATACAAGA**GGTACC**TTGTTTTTTCAATACG-3′
3′-CTTCTATGTTCT**CCATGG**AACAAAAAAGTTATGC -5′

NIS 47: 5′-GAAGATACAAGA**GGTACC**TTGTTTTTTCAATACG-3′

NIS 48: 5′-CGTATTGAAAAAACAA**GGTACC**TCTTGTATCTTC-3′

5.2.3.2: Swapping of motif I-IV from M.*Hha*I to M.SPRI

Ligation of vector (5258 bps) generated from pQIS 64 (encoding wild type M.*Hha*I) with the insert (1084 bps) from pQIS 72 (encoding wild type M.SPRI) both of which had been cut with *Kpn*I and *Xho*I and gel purified (Figure 5.9). The new construct pQIS 79, that is a chimaera of two Mtases M.*Hha*I and M.SPRI now contains motif I-IV of M.*Hha*I and V-X from M.SPRI (Figure 5.10).

5.2.3.3: Swapping of motif I-IV from M.SPRI to M.*Hha*I

Ligation of pQIS 72 (wild type M.SPRI) (5243 bps) with the insert from pQIS 64 (wild type M.*Hha*I) (691 bps) both cut with *Kpn*I and *Xho*I and recovered from the gel (Figure 5.10). The construct pQIS 81 is a chimaera of the two Mtases M.*Hha*I and M.SPRI encoding for motif I-IV of M.SPRI and V-X from M.*Hha*I (Figure 5.11). The graphic representation summarising all of the constructs made in this motif exchange experiment is given in Figure 5.12.

Although mutagenesis experiment that includes point mutations (chapter three), insertion mutagenesis (chapter four) and exchange of regions (chapter five) the enzyme remains inactive, it was not clear whether the enzyme had also lost the sequence specific binding function of the inserted TRD. Surface Plasmon Resonance (SPR) experiments were designed to address this question. Surface Plasmon Resonance (SPR) allows a direct comparison to be made between the binding affinity for the DNA of wild type and mutant Mtases.

Figure 5.12: Summary of the chimaeric constructs made in this chapter.
A) Motif I of M.*Hha*I fused with motifs II - X of M.SPR.
B) Motifs I - IV of M.*Hha*I fused with motifs V - X of M.SPR.
C) Motif I of M.SPR fused with motifs II - X of M.*Hha*I.
D) Motifs I - IV of M.SPR fused with motifs V - X of M.*Hha*I.
E) Motif I of M.*Hha*I fused with motifs II - X of M.SPR (restored Asn 39).
F) Motif I of M.SPR fused with motifs II - X of M.*Hha*I (restored Ser 33).

5.3: Surface Plasmon Resonance (SPR)

5.3.1: INTRODUCTION

SPR is an optical technique that detects refractive index changes at the surface of a sensor chip. In fact, Surface Plasmon Resonance (SPR) is a technique that can be used for the measurement of biomolecular interactions in real time, with no labelling of the molecules necessary. SPR is an optical phenomenon that occurs when light illuminates a thin conducting film under specific conditions, which can best be described by first understanding the principle of total internal reflection (TIR). TIR occurs at an interface between non-absorbing media. When a light beam propagating in a medium of higher refractive index meets an interface at a medium of lower refractive index at an angle of incidence above a critical angle, the light is totally reflected at the interface and propagates back into the high refractive index medium. When TIR occurs an electromagnetic evanescent wave travels from the interface into the lower refractive index medium. If the interface is coated in a thin layer of conducting material such as a metal of suitable thickness, SPR can arise (Figure 5.13). In a metallic solid, the plasma consists of a 'gas' of free electrons. Plasma oscillations that propagate at the interface of a metal and a dielectric medium are called Surface Plasmons (SPs). In an SPR sensor, light is coupled resonantly into electric oscillations, or surface plasmons, at a metal surface. Such oscillations give rise to a non-propagating evanescent wave, which extends from the metal surface into the sample solution, decaying exponentially as a function of distance. Macromolecular complexes formed at the metal-liquid interface, resulting in a change in refractive index of the liquid media at the interface, perturb the evanescent wave and alter the propagation characteristics of the plasmons. Changes in the propagation characteristics of the plasmons in turn alter the characteristics of the internally reflected light. Such changes are ultimately detected and quantified. In general, increasing the bio-layer thickness, results in an increase in the resonance angle. The thicker the bio-layer on the conductive film surface, the greater the SPRI minimum shifts to higher angles. This shift in angle is shown on the sensorgram of the BIACORE® apparatus as a change in the response signal, and the larger the change in angle, the greater will be the response signal.

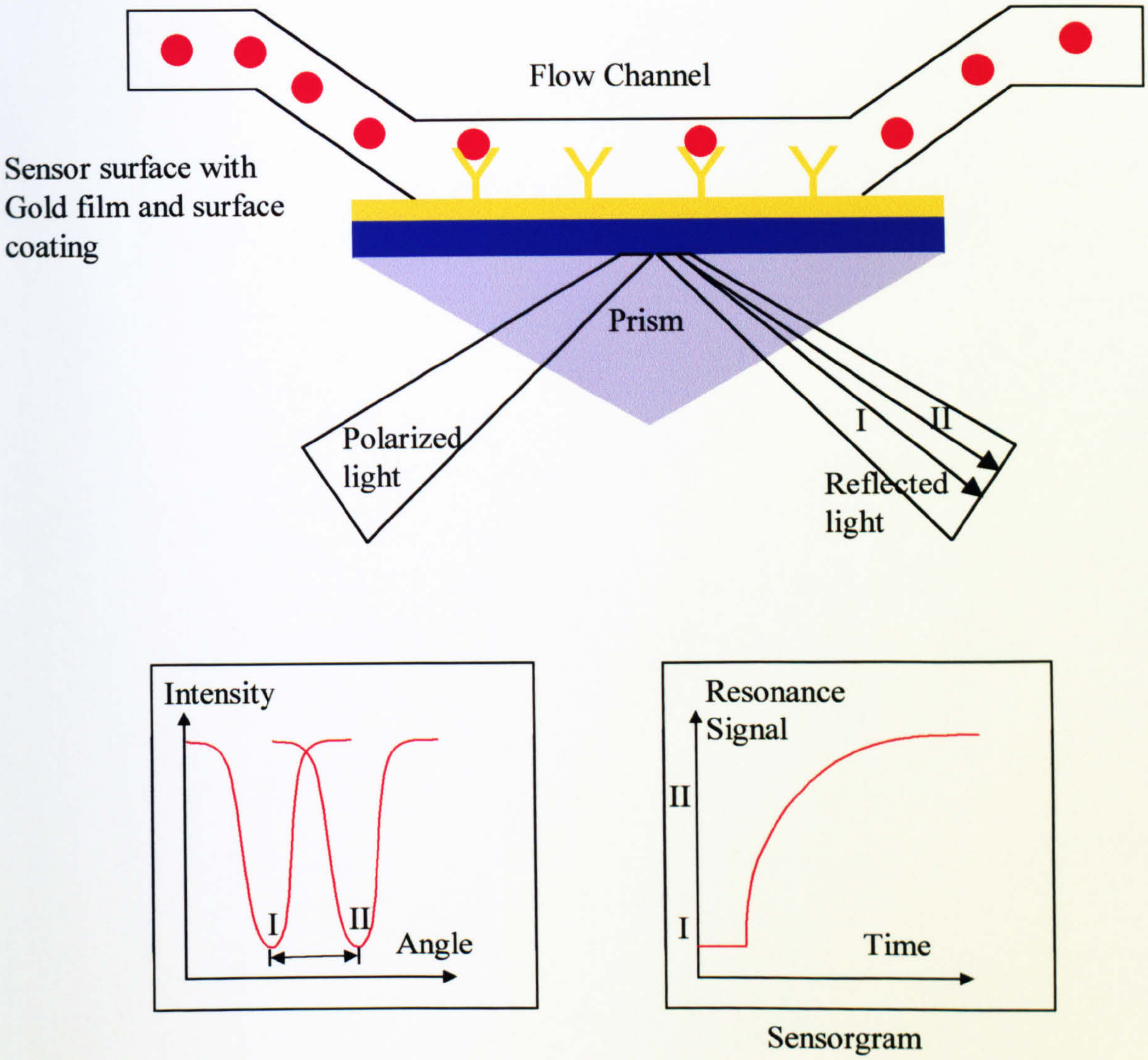


Figure 5.13: Surface Plasmon Resonance (SPR) Detection System.
The polarized light is focused into a wedge-shaped beam providing a continuous interval of light wavevectors k_x. This covers the working range for the plasma wavevector k_{sp} during biomolecular interaction analysis. An increased sample concentration in the surface coating of the sensor chip causes a corresponding increase in refractive index which alters the angle of incidence required to create the SPR phenomenon (the SPR angle). This SPR angle is monitored as a change in the detector position for the reflected intensity dip (from I to II). By monitoring the SPR-angle as a function of time the kinetic events in the surface can be displayed in a sensorgram. (Figure adapted from BIACORE **BIAtechnology Note 107).**

5.3.2: BIACORE SYSTEM

One of today's frontlines in biology is to investigate macromolecular interactions in a quantitative way. Such analysis using BIACORE® instruments has been established as a powerful technique. The BIACORE® system is a biosensor instrument employing surface plasmon resonance detection for molecular interaction analysis. The first commercial biosensor was first developed in 1990 by BIACORE Inc. formally Pharmacia Biosensors (Myszka, 1997). This system monitors the refractive index change as a molecule interacts at the sensor surface. The signals generated are recorded in arbitrary response units (RU). While BIACORE instruments are routinely used for kinetic measurements and for the determination of binding constants, the immobilization of a ligand onto the sensor chip surface has to be individually optimised for every system. In the basic design of a SPR biosensor-based experiment, one reactant, which is often called the ligand, is attached to the sensor surface. The other reactant, which is referred to as the analyte, flows past this surface in solution. When the analyte (A) and the ligand (B) interact to form a complex (AB), a response is generated (Myszka, 1997). Schematic diagram of the molecular interactions that occur at the sensor surface are shown in Figure 5.14. As the binding responses occur in real time, it is possible to obtain kinetic information about the interaction.

A real-time sensorgram provide two kinds of information (BIAtechnoloogy Handbook, 1994): (a) the rate of interactions (association, dissociation or both) providing a kinetic rate constant and (b) analyte concentration. It is important to note that the SPRI detection systems signal is related to the change in surface concentration of the analyte. Therefore it depends on the molecular weight of the analyte and the valence of the interaction. A higher molecular weight analyte will give a proportionally larger response.

5.3.2.1: Stoichiometric Determination

The following equations can be used on the assumption that the relationship between response and mass is the same for both ligand and analyte (1000 RU = 1ng/mm^2 for protein).

Ligand sites (pmole/mm2) $\propto$ ligand response/ ligand molecular weight x valence

Where valence is the number of analyte molecules, which can bind to the ligand (i.e. 2:1 stoichiometry has a valence of 2) substituting ligand sites for analyte molecules.

R_{max} = Analyte MW / ligand MW x ligand response x valence

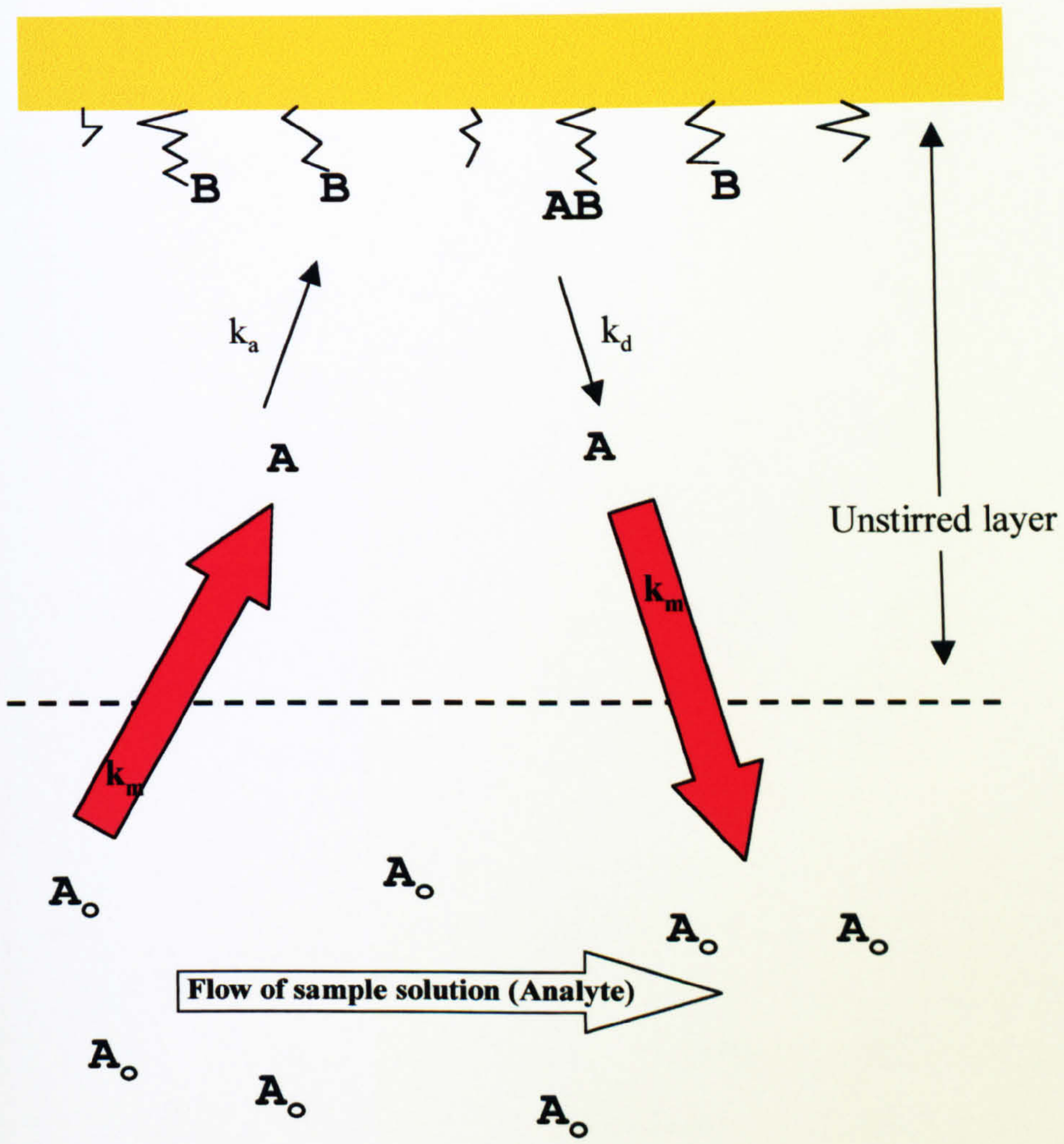


Figure 5.14: A schematic representation showing molecular interactions at the biosensor surface.
The ligand B is attached to the surface via a flexible linker. During the association phase, analyte (A_o) is flowed past the surface. k_m is the mass transport coefficient used to describe the diffusion of analyte through the unstirred solvent layer. k_a and K_d are the intrinsic reaction rate constants, which describe the formation of (AB) complex. Figure adapted from Myszka (1997).

Where R_{max} is the maximum binding capacity of the surface ligand for this analyte in RU.

5.3.2.2: General Rate equation

The on and off rates for binary complex formation are described by

$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} AB$$

Association rate

$d[AB]/dt = k_a [A][B]$

Dissociation rate

$-d[AB]dt = k_d [AB]$

At equilibrium, dissociation rates and association rates are equal, therefore:

$k_d / k_a = K_D$ (the equilibrium (affinity) dissociation constant)

or

$k_a / k_d = K_A$ (the equilibrium (affinity) association constant)

The equilibrium constant defines the balance between a complex made up of the bound and the free components at equilibrium, while the rate constants define how fast the interaction occurs. An interaction is fully characterised in kinetic terms only when both rate constants and equilibrium constants are determined. If one rate constant and the equilibrium constant can be measured, the other rate constant can be deduced (alternatively, the equilibrium constant can be calculated from measured rate constants).

Typical ranges of parameters for equilibrium and rate constants measurable with BIACORE® are shown in the following table. An important feature is that because SPR is sensitive to molecular mass, there is a lower limit (≈200 Da) for detection (Karlsson and Stahlberg, 1995).

Equilibrium constant K_D	$10^{-4} - 10^{-12}$ M
Association rate constant K_a	$10^{3} - 10^{7}$ $M^{-1}s^{-1}$
Dissociation rate constant K_d	$10^{-6} - 10^{-1}$ s^{-1}
Temperature range	4 – 40°C
Minimum molecular weight	200 Da
Flow rates	1– 100 µl min^{-1}

In kinetic experiments, a major parameter that must be controlled is the diffusion of the analyte across the sensor surface, where it will interact with the ligand. This diffusion can cause mass transport limitations that compromise the determination of rate constants for fast binding reactions. These effects can be minimised by using low amounts of ligand bound to the surface (<150RU) and high flow rates (>20 µl min^{-1}). At the same time a similar problem appears for the opposite situation, if the analyte is not transported away from the surface quickly enough, it may re-bind thereby slowing the apparent dissociation process. Another important aspect to consider when designing a BIACORE experiment is the use of a reference surface on which no ligand is immobilised. This removes instrument artefacts, such as matrix effects and non-specific binding. However, in this study preliminary comparisons are made between wild type and mutants (which only differ in one or two amino acids), and detailed kinetics analysis was not involved, it was not necessary to use small amounts of ligand or high flow rates.

5.3.2.3: Analysis of Binding Data

For the establishment of reliable kinetic parameters in an SPR experiment it is important to carry out a number of experiments simultaneously. Any response curve can be described by the sum of two or more integrated rate equations and therefore there is not enough information in a single response curve to discriminate between different binding mechanisms. However, it has been demonstrated that fitting association and dissociation phases from a series of analyte concentrations (global) analysis can be used to discriminate between different reaction mechanisms (Myszka, 1997).

5.3.2.4: Aims

Whilst DNA Mtases do not exhibit the same level of affinity in non-catalytic recognition of their target sites, they still bind with sufficiently high affinity to be investigated by surface plasmon resonance (Kan, 1999; Matin, 2000). In this experiment it was proposed to explore nucleic acid-protein interactions using surface plasmon resonance, which provides a convenient means of analysing complex formation at the surface of the BIACORE® chip. The interactions between recombinant wild type and mutant proteins with synthetic oligonucleotides carrying the recognition sequence for the protein were studied.

5.3.3: *IN VITRO* TRANSLATION OF WILD TYPE AND MUTANT M.*Hha*I

It was decided to express protein from plasmids carrying the wild type and mutant M.*Hha*I gene. However, considering the number of mutants it was thought convenient to express the proteins *in vitro*. The cell free system originally devised by Zubay (1973) exploits DNA fragments, plasmids or isolated genes, which are transcribed by the endogenous RNA polymerase of the bacterial extract coupled with translation of the transcript by ribosomes. Specific proteins synthesised can then be assayed by their enzymatic or their biological activity. The *E. coli*T7 S30 extract system for circular DNA from Promega offers a single tube, coupled transcription/translation system. The *E. coli*T7 S30 extract system also simplifies the process and reduces the time required to obtain *in vitro* translation results (Promega Technical Bulletin No 219). It has been reported that protein translated *in vitro* by *E. coli*T7 S30 extract system can synthesize up to several hundred micrograms of protein (Davis *et al.*, 1998). Therefore the protein expressed in this way is sufficient for the study of DNA-protein interactions using the BIACORE® system. The *E. coli*T7 S30 extract system can be used for transcription and translation of genes cloned downstream from the T7 RNA polymerase promoter and therefore, it was necessary to transfer all the mutant constructs from the original plasmid to a plasmid with a T7 promoter. The pET system is the most powerful system yet developed for the cloning and expression of recombinant proteins in *E. coli* (Novagen Catalogue 1998-99).

It was therefore necessary to swap all the wild type and mutant constructs from previous constructs into pET-22b. In the first experiment the wild type M.*Hha*I was transferred from pMSK 7 to pET-22b. After that, selected mutants of M.*Hha*I were swapped into pET-22b.

5.3.3.1: Transfer of the wild type M.*Hha*I gene from pMSK-7 to pET-22b

The wild type M.*Hha*I gene (derived from pMSK 7) was cleaved by *Eco*RI and *Xho*I and recovered from an agarose gel. It was ligated into gel-purified pET-22b, which was pre-cut with the *Eco*RI and *Xho*I. Those transformants arising from *E. coli* DH5α mcr^{-}, derived from the ligation mixture, were screened for the presence of the M.*Hha*I gene. The resulting construct pQIS 56 contained the wild-type M.*Hha*I gene in an appropriate reading frame and is under the influence of the bacteriophage T7 promoter. In addition, the expressed protein will have a six-histidine tag attached at

the C-terminus. However, it was not possible to check the activity of the construct pQIS 56 simply by challenging the plasmid with *Hha*I because DH5α mcr^{-} *E. coli* cells lack the T7 RNA polymerase gene. The target gene was therefore virtually "off" *in vivo* and did not express the M.*Hha*I enzyme. The activity of M.*Hha*I was checked by the *mcr* response of *E. coli* cells BL 21 (DE3). BL21 (DE3) is the most widely used host strain that is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter. The genes cloned in pET plasmid are under the control of strong bacteriophage T7 transcription and translation signals, which provide a source of T7 RNA polymerase in the host cells induces expression. Once it had been established, that the construct contained the correct insert, the recombinant expression plasmid was transferred into an expression host containing a chromosomal copy of the T7 RNA polymerase gene under *lacUV5* control, and expression was induced. However owing to the presence of the active Mtase gene, hydrolysis of the methylated plasmid takes place because of the *mcr*$^{+}$ genotype of BL21(DE3), and as a result no colonies appear on transformation plates.

After the wild type M.*Hha*I had been transferred into pET-22b, it was convenient to swap the mutants into pET-22b. The selected constructs made in pET 22b were translated *in vitro* (See Chapter Two) and the protein expressed was used for BIACORE® SPR experiments. Quantification of protein was necessary in order to ensure equivalent amounts of protein were used in all binding assays. It had been observed that the quantification using NTA chip in the BIACORE® 2000 was not a good choice, as poor quality binding was observed at the NTA surface. It has been reported that stable binding can be obtained for a monomeric hexahistidine peptide but not for monomeric His-tagged proteins (Nieba *et al.*, 1997). Proteins with only one His-tag only bind very weakly to the Ni^{2+}-NTA surface at neutral pH. In conclusion, the only limitation is that one His tag on a monomeric protein is usually not sufficient for stable binding, except in special cases (Nieba *et al.*, 1997).

As an alternative assay, *in vitro* methylation of DNA by wild type M.*Hha*I expressed *in vitro* was carried out to see whether the protein was expressed at sufficient level to detect activity. The assay is described in material and methods (Chapter Two). Typically 10 µl of *in vitro* expressed protein was incubated with 25 µl of IMAC resin (50% bed volume) (for binding of M.*Hha*I with beads) for one hour. The beads were

then centrifuged, washed once with Mtase buffer, and were then used to catalyse the methylation of 600 ng (2 µl) λ DNA in the presence of 5 µl Mtase buffer (described in section 3.1.7.1) supplemented with 100 µM of SAM overnight, at 37°C. After this time, the contents of the tube were centrifuged and the supernatant was transferred into another tube. The reaction was stopped by addition of NEM followed by the addition of *Hha*I buffer and the restriction enzyme followed by a further incubation for one hour. Two control experiments were also carried out in parallel to this experiment. The first used a purified DNA methyltransferase as a positive control and the second used luciferase expressed *in vitro* as a negative control. At the end of the reaction the products were analysed on a 1% agarose gel (Figure 5.15). It was observed that *in vitro* expressed M.*Hha*I could methylate λ DNA so that it was completely protected from *Hha*I restriction (Figure 5.15). We can see a slight smear (see lane one) compared to the positive control (see lane three), which is single band of λ DNA, which could be due to the presence of various other proteins present in the lysate which were not removed by a single washing. It was also observed in this experiment that one to two hours incubation was not sufficient (data not shown) for the *in vitro* translated product to complete the methylation reaction. It appears that the either the protein yield is very low in this experiment or the presence of other proteins affect the methylation reaction.

In order to improve the quality of the data the selected constructs have been "normalised" by sub cloning them into pET-22b as GST fusions. The constructs were made and checked for their activity (by transformation of *E. coli* BL21 (DE3)) via the *mcr* response. Table 5.1 summarises the properties of parent plasmid and the new constructs made in pET-22b. It was decided to carry out an *in vitro* translation for wild type M.*Hha*I as a GST fusion and to check its encoded protein by in vitro methylation protection assay as was carried out with IMAC resin (explained above). The only difference this time was glutathione agarose beads instead of IMAC resin. The result (Figure 5.16) showed that the encoded protein by pQIS 65 is catalytically active and protection of λ DNA was observed against restriction with *Hha*I.

5.3.3.2: Preparation of CM5 chip

The quantification of recombinant fusion protein was carried out using a GST capture kit (Biacore). The GST capture kit provides site directed immobilization of GST fusion proteins through capture on immobilized Goat anti-GST antibody at the

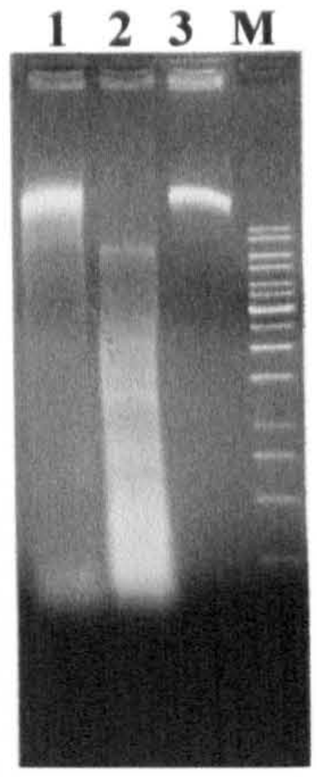


Figure 5.15: *In vitro* methylation protection by wild type M.*Hha*I expressed *in vitro*. The Gel above shows an alternative assay for *in vitro* methylation protection of lambda DNA by *in vitro* expressed M.*Hha*I. **Lane 1:** 600 ng of λ DNA incubated with *in vitro* translated M.*Hha*I followed by restriction with *Hha*I (Protected λ DNA without cleavage). **Lane 2:** Negative control. 600 ng of λ DNA incubated with luciferase followed by restriction with *Hha*I. (No protection, λ DNA completely cleaved). **Lane 3:** 600 ng of λ DNA incubated with *in vivo* expressed and purified M.*Hha*I followed by restriction with *Hha*I (Protected λ DNA, no cleavage). **M:** GeneRuler™ 1 kb DNA ladder. (See Chapter Two and appendices for the DNA marker ladder size).

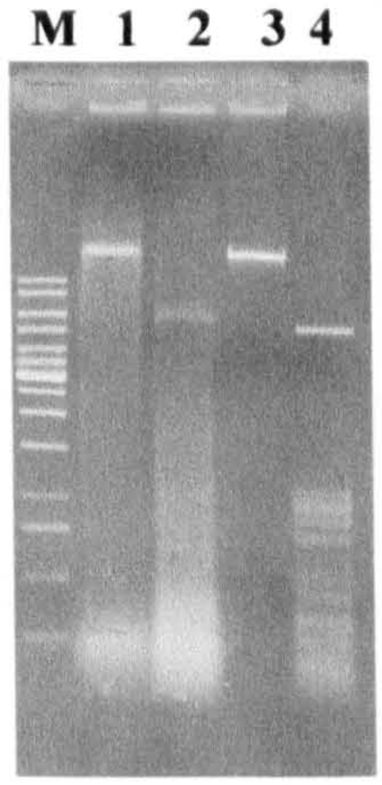


Figure 5.16: *In vitro* methylation protection by wild type M.*Hha*I expressed *in vitro*. The Gel above shows an alternative assay for *in vitro* methylation protection of lambda DNA by *in vitro* expressed M.*Hha*I purified by glutathion agarose beads. **M:** GeneRuler^{TM} 1 kb DNA ladder. **Lane 1:** 400 ng of λ DNA incubated with *in vitro* translated M.*Hha*I followed by restriction with *Hha*I (Protected λ DNA without cleavage). **Lane 2:** Negative control. 400 ng of λ DNA incubated with luciferase followed by restriction with *Hha*I. (No protection, λ DNA completely cleaved). **Lane 3:** 400 ng of λ DNA incubated with *in vivo* expressed and purified M.*Hha*I followed by restriction with *Hha*I (Protected λ DNA, no cleavage). **4.** 400 ng λ DNA restricted with *Hha*I (negative control, λ DNA completely cleaved). (See Chapter Two and appendices for the DNA marker ladder size).

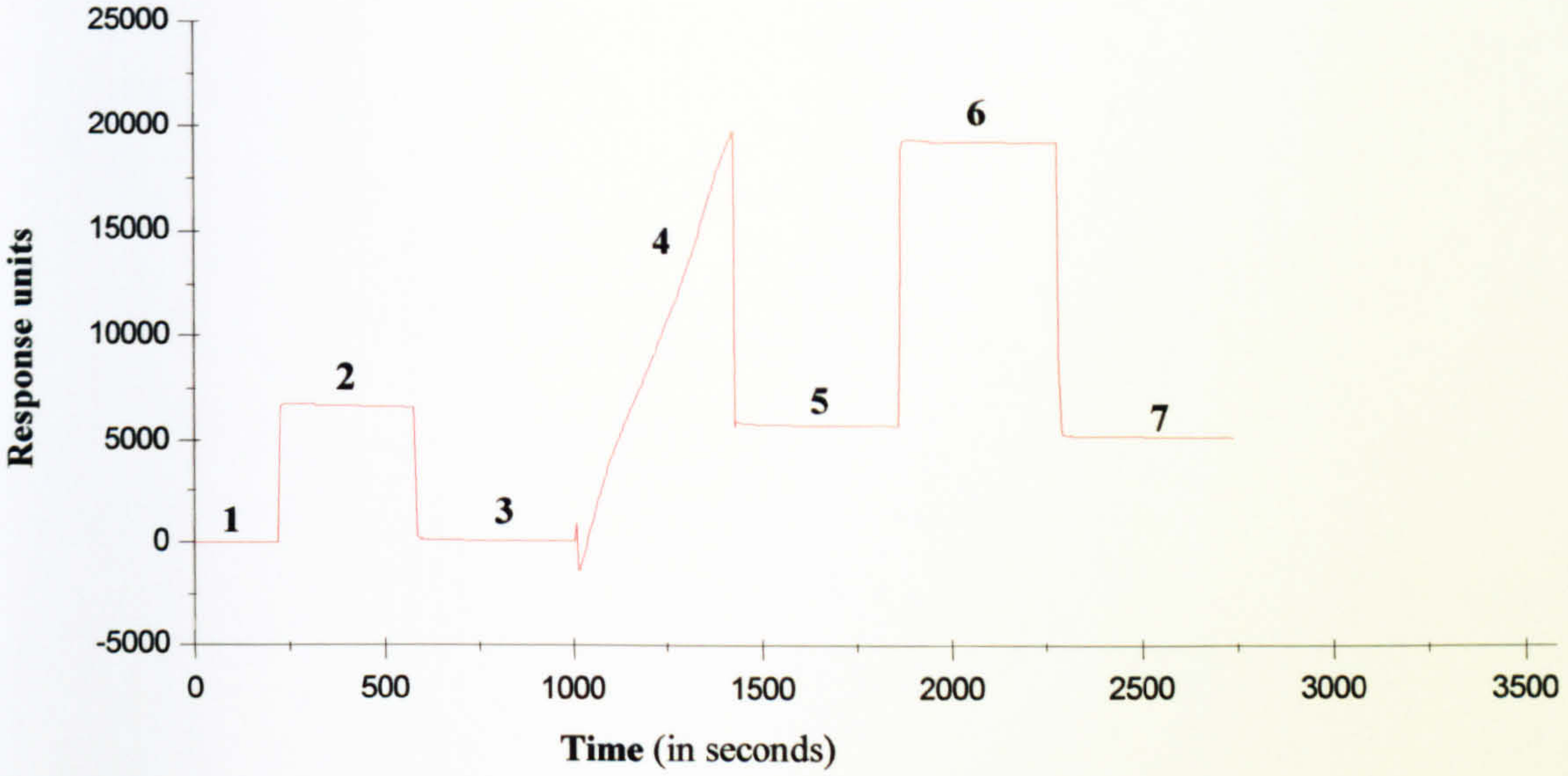
Table 5.1: Summary and properties of the new constructs made in pET-22b

	Description	Parent construct	*In vivo* methylation protection response on gel	New construct in pET 22b	Mcr^+ response
1	Wild type M.*Hha*I	pMSK 7	Active	pQIS 65	Active
2	Single mutant (A280F)	pQIS 12	Partial Active	pQIS 66	Inactive
3	Double mutant (A280F, R277L)	pQIS 14	Inactive	pQIS 67	Inactive
4	Triple mutant (A280F, R277L, F302A)	pQIS 18	Active	pQIS 68	Active
5	Double mutant (K270R, T271I)	pQIS 55	Partial Active	pQIS 69	Active

surface of a CM5 chip. These types of chips were prepared manually in the laboratory. Anti-GST was first diluted to 30 µg/ml in coupling solution (10 mM Na-acetate, pH 5.0), and immobilised onto a CM5 sensor chip using the amine coupling method. This method involves activation of the carboxyl groups on the matrix by the addition of a mixture (1:1) of succinimide (NHS, N-hydroxysuccinimide, 11.5 mg/ml) and carbodiimide (EDC, N-ethyl-N'-(3-dimethylaminopropyl) carbodiimide hydrochloride, 75 mg/ml) to form active esters which react with amine groups on the ligand (anti-GST) to form a covalent bond. After ligand coupling, ethanolamine is added to deactivate the remaining active esters (see Figure 5.17). After the successful coupling of the antibody, it was equilibrated with 1 x PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3). Following were the observations during preparation of CM5 chip.

	Description	Relative Response (RU)
1	Non activated surface	0
2	Activation with NHS/EDC	137.4
3	Injection of anti-GST	4164.2
4	Deactivation with ethanolamine	-329.3

A)



B)

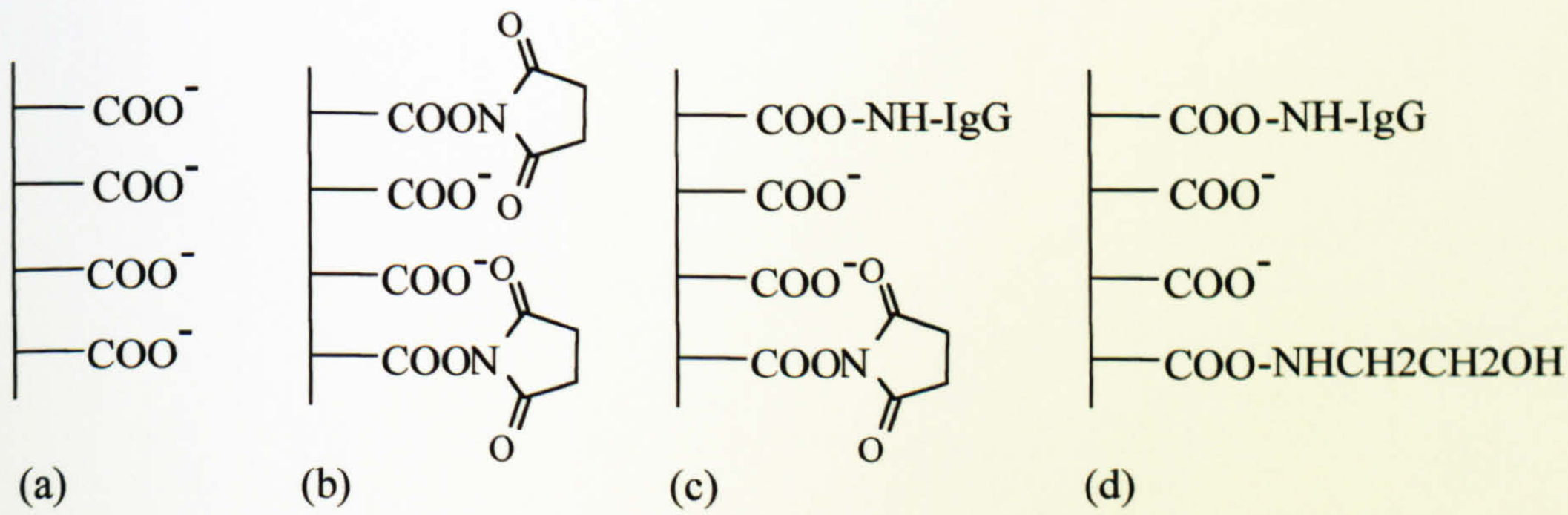


Figure 5.17: A) Graphic representation of Coupling of anti-GST on the surface of CM5 chip. The interactions relating to each part of the sensorgram are shown above and also explained below, more information is given in the text. **1)** Base line showing non-activated carboxyl groups of the carboxymethylated dextran layer; **2)** Activation of the carboxyl groups on the matrix by adding a mixture of carbodiimide (EDC) and succinimide (NHS) to form active esters that can react spontaneously with amine groups on the ligand; **3)** After activation; **4)** Coupling of ligand: Goat anti GST IgG is covalently bound; **5)** Buffer washes away loosely associated ligand; **6)** Residual active esters are deactivated with ethanolamine; 7) Difference in 3 and 7 reflects the amount immobilized. **B) Chemistry of reactions taking place on the surface of the CM5 sensor chip during ligand immobilization. a)** Non-activated carboxyl groups on the surface of the chip **b)** Some carboxyl groups are activated with NHS/EDC. **c)** IgG is covalently bound **d)** Residual active esters are deactivated with ethanolamine.

The CM5 sensor chip is now ready for co-eluting GST by injecting onto the surface of the sensor chip. A range of recombinant GST (*Schistosoma japonicum*) dilutions from 4 ng/µl to 100 ng/µl was used for the preparation of a calibration curve.

The plasmids listed in Table 5.1 were used to direct *in vitro* translation and the *in vitro* translated proteins were then first quantified. In this step, typically 2, 5 and 10 µl of *in vitro* expressed protein diluted with 1X methyltransferase buffer up to a final volume of 50 µl were injected at a flow rate of 5 µl per minute on to the surface of CM5 sensor chip. Response curves obtained were used for the determination of unknown concentration of *in vitro* expressed recombinant proteins.

5.3.3.3: Analysis of the M.*Hha*I binary complex with DNA

An experiment was designed to analyse the wild type M.*Hha*I and some selected mutants (shown in the Table 5.1) for binary complex formation with an oligodeoxynucleotide containing the *Hha*I recognition site using surface plasmon resonance. In order to do so, Biotinylated DNA sequence/oligonucleotides specific for binding to *Hha*I, *Hae*III and *Msp*I specific sites were designed and synthesized of

Name of Oligonucleotide	Feature	Sequence
A1 = 5'BHH	Top Biotinylated oligo for *HhaI* with (GCGC) recognition sequence	**5'-B**GAATGCTACAGTATCGT**GCGC**TCA CGTACAACATCCAG-**3'**
A2 = 3'HHM	Bottom oligo for A1	**5'-**CTGGATGTTGTACGTGA**GMGC**ACG ATACTGTAGCATTC-**3'**
A13 = 5'BSP	Top Biotinylated oligo for *MspI* with (CCGG) recognition sequence	**5'-B**GAATGCTACAGTATCGT**CCGG**TCA CGTACAACATCCAG-**3'**
ASP2 = 3'SPM	Bottom oligo for A13	**5'-**CTGGATGTTGTACGTGA**MCGG**ACG ATACTGTAGCATTC-**3'**
A14 = 5'BHA	Top Biotinylated oligo for *HaeIII* with (GGCC) recognition sequence	**5'-B**GAATGCTACAGTATCGT**GGCC**TCA CGTACAACATCCAG-**3'**
BHA2 = 3'HAM	Bottom oligo for A14	**5'-**CTGGATGTTGTACGTGA**GGMC**ACG ATACTGTAGCATTC-**3'**

B = Biotinylated
M = 5' Methylcytosine

Table 5.2: List of oligonucleotides used as ligands for binding experiments with SA chip of BIACORE

equivalent length differing by a single base pair at the target site (See Table 5.2). Only the top oligonucleotide was 5'-biotinylated. The sequences detailed in Table 5.2 comprise one recognition site after annealing and mimic the natural substrates.

In the M.*Hha*I:DNA binding analysis using SPR, two types of controls must be considered, the first one is the non-specific binding between the *in vitro* protein and the ligand. This can be determined by using a non-Mtase protein, translated in the same way as the other proteins were translated such as luciferase. Luciferase is used to check whether the kit 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 achieved by using the blank flow cell that contains streptavidin without DNA.

An SA sensor chip was normalised with Bianormalising solution (40% W/W glycerol in water) to compensate for the variations in reflectance characteristics between different sensor chips. It was then washed with 1 M NaCl in 50 mM NaOH prior to the assay. The biotinylated substrate (800 nM) was prepared in methyltransferase buffer (50 mM Tris-HCl, pH 8.5, 50 mM NaCl, 10 mM dithiothreitol), and immobilised in molar excess on the surface of the SA sensor chip utilising the streptavidin-biotin interaction. Flow cell 1 was kept blank, with no pre-attached oligonucleotide. Flow cell 2 contained immobilised biotinylated *Hha*I binding site-specific oligonucleotide. Flow cell 3 contained an immobilised biotinylated *Msp*I binding site-specific oligonucleotide (in order to measure non-specific interactions).

The protein under investigation was passed over the "flow cells" of the SA chip in equimolar concentration to initiate the molecular interaction. The injection flow rate was set at 5 µl/min in order to achieve maximum binding in the shortest period of time. Enzyme-substrate binding was detected soon after the enzyme solution was injected. Response curves show steady state formation of enzyme substrate complex. When enzyme-substrate coupling had reached a maximum value, and the enzyme solution injection was terminated, the sensor chip surface was flushed thoroughly with an excess of *Hha*I methyltransferase buffer. The molecular interactions in the form of response units for variants of *Hha*I was recorded for blank (flow cell 1), *Hha*I specific (flow cell 2) and non-specific (flow cell 3).

The DNA binding affinity of the GST-M.*Hha*I wild type and its mutant variants were compared by SPR. Interestingly the association phase is almost the same for all mutants and wild type M.*Hha*I.

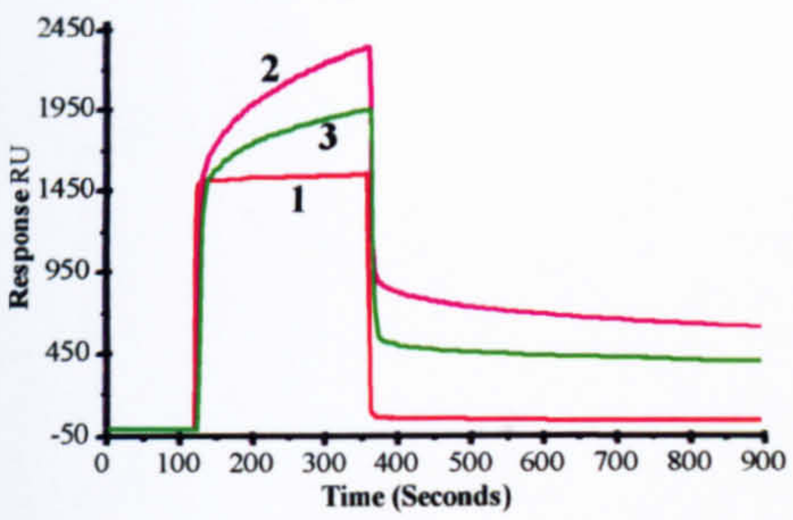
All experiments were carried out at 25°C. Wild type and mutant protein were injected over substrate in a total injection volume of 20µl. The background resulting from non-specific binding of the enzyme to a blank flow cell was subtracted from the data before analysis, which was performed using the Bia-evaluation program. The results are presented in table below and shown in Figure 5.18.

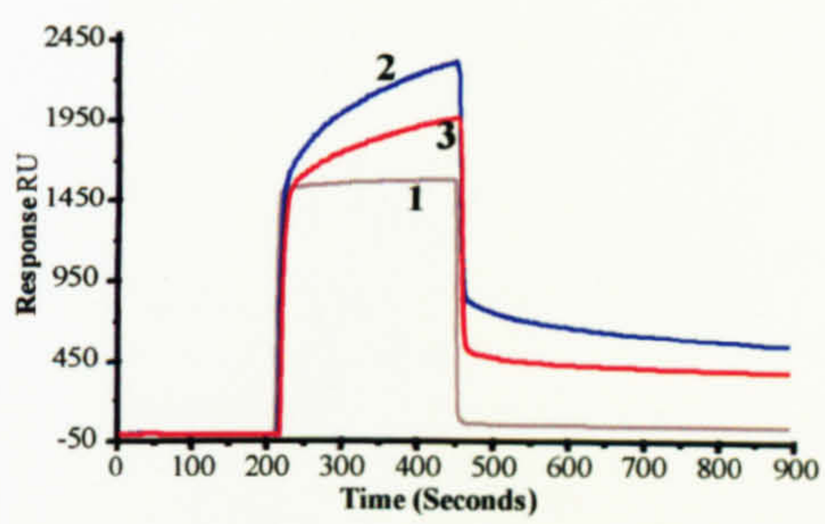
	Description	Construct number	Rmax (RU)
1	Wild type M.*Hha*I	pQIS 65	966
2	Single mutant (A280F)	pQIS 66	883
3	Double mutant (A280F, R277L)	pQIS 67	930
4	Triple mutant (A280F, R277L, F302A)	pQIS 68	936
5	Double mutant (K270R, T271I)	pQIS 69	889

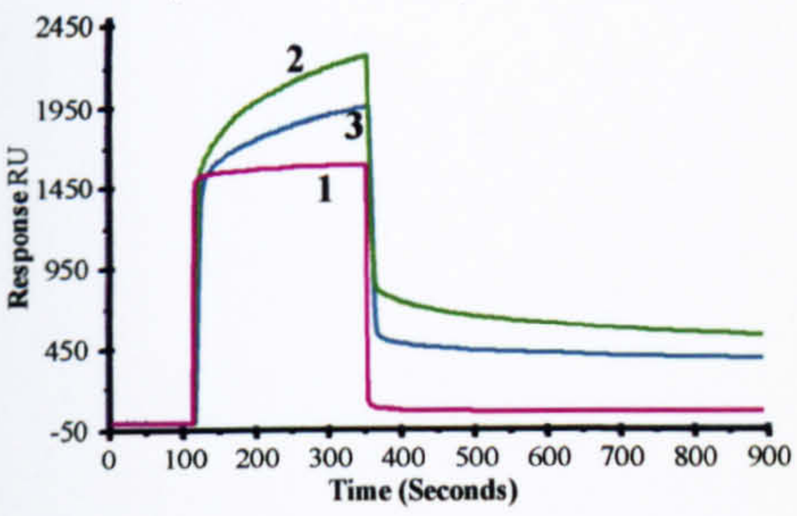
The result showed that there is no significant difference in the association properties of mutants with that of wild type of M.*Hha*I. Although the mutants showed different methylation profile by *in vitro* protection assay, these results indicate that the DNA recognition is unaffected by these mutations.

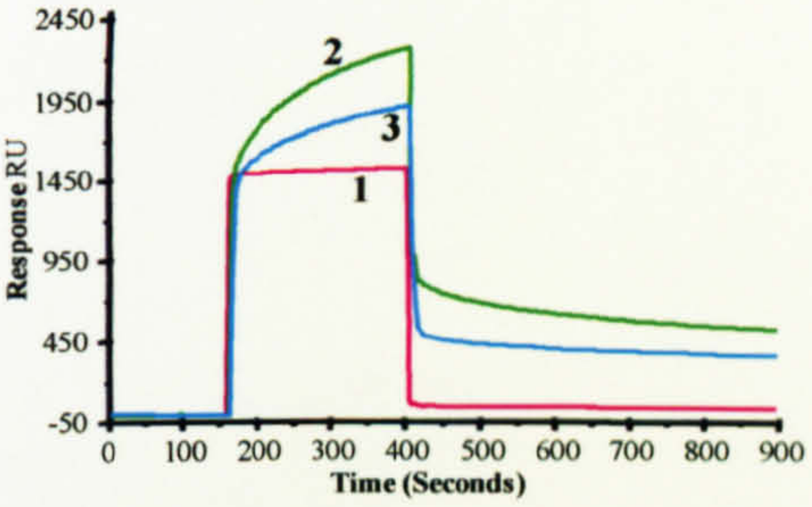
5.4: Discussion

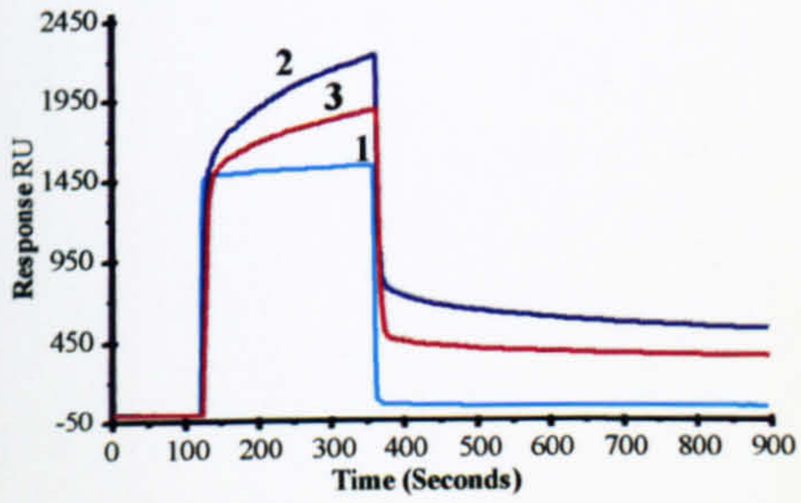
It has been established for the monospecific and multispecific Mtases that the variable region is wholly responsible for the selection of DNA sequences to be methylated (Wilke *et al.*, 1988; Mi and Roberts, 1992; Trautner *et al.*, 1996). We have attempted to dissect the variable region of M.*Hha*I by mutational analysis (see chapter three and four) and by constructing various hybrids by insertion of segments from multispecific Mtase M.SPRI (chapter four) to establish the relationship between variable region and motif IX of M.*Hha*I. We managed to redesign motif IX of M.*Hha*I that mimics to multispecific Mtase M.SPRI. However, all other attempts to exchange the monospecific variable region with the TRDs of M.SPRI were unsuccessful. A list of all the mutant constructs made in this study, by the codon changes at single or multiple sites in M.SPRI is given in Table 5.3. Table 5.4 summarises the mutant made at single or multiple sites in M.*Hha*I.











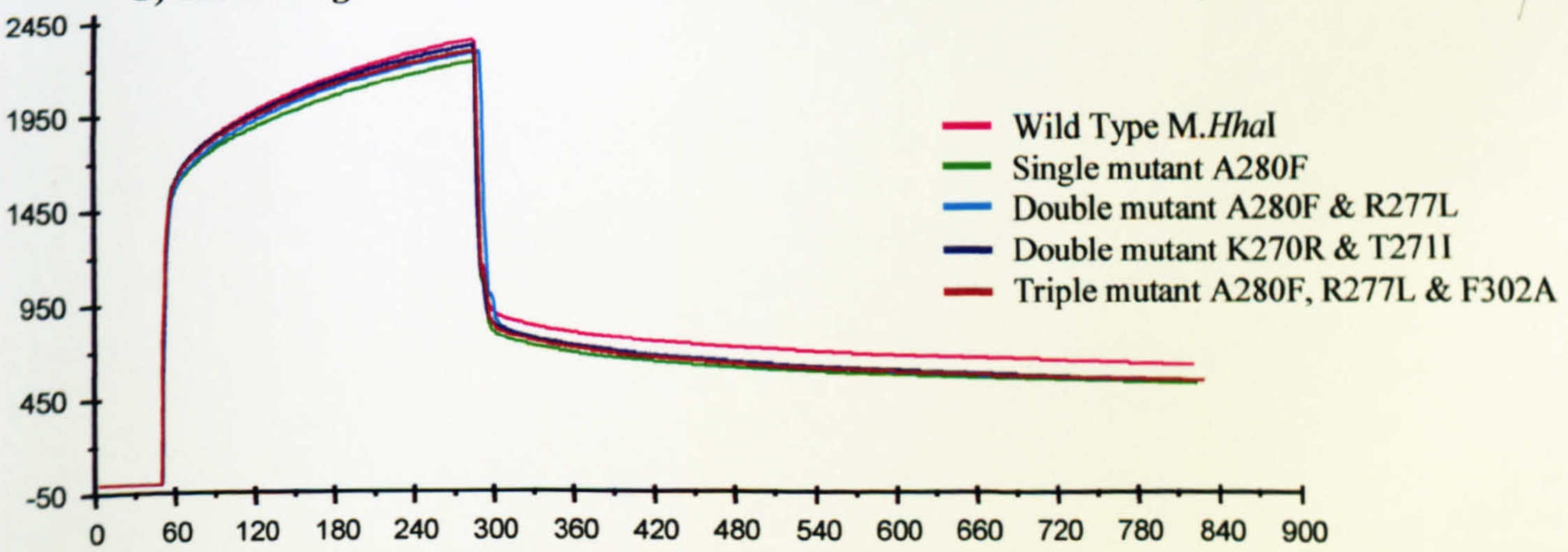


Figure 5.18: Biacore analysis of the interaction between an *Hha*I binding site-specific oligonucleotide and variant of the *Hha*I methyltransferase. The results showed in figure A-F represents molecular interactions as follows. The protein under investigations were passed over three "flow cells" of the Biacore chip and interactions were measured and shown for each flow cell as labelled. **Flow cell 1** was a blank (without substrate) with a no pre-attached oligonucleotide. **Flow cell 2** contained an immobilized biotinylated *Hha*I binding site specific oligonucleotide. **Flow cell 3** contained an immobilized biotinylated *Msp*I binding site-specific oligonucleotide (in order to measure non specific interactions). After subtracting the effects due to non-specific interactions, the chromatograms were overlaid and are shown in panel G.

Table 5.3: Summary of mutant constructs made in M.SPRI gene

Construct	Mutation	Description	Activity
pQIS 8	Phe386Ala	Single point mutation with the introduction of *Fsp*I site	Partial
pQIS 9	Ala390Gly	Single point mutation	Active
pQIS 10	Asp392Pro	Single point mutation	Active
pQIS 34	Gln189Glu	Single point mutation with the introduction of *Xho*I site	Active
pQIS 45	Phe112Leu Phe113Gly Val114Pro	Triple mutant with the introduction of *Apa*I site	Inactive
pQIS 46	Phe112Leu Phe113Gly Val114Pro Phe386Ala	Quadruple mutant with the introduction of *Apa*I site	Inactive
pQIS 76	Ser33Arg	Single point mutation with the introduction of *Nru*I site	Inactive

Table 5.4: Summary of mutant constructs made in M.*Hha*I gene

Construct	Mutation	Description	Activity
pQIS 11/ pQIS 12	Ala280Phe	Single point mutation with the introduction of *Sty*I site	Partial
pQIS 13/ pQIS 14	Arg277Leu Ala280Phe	Double mutation with the introduction of *Xba*I site and deletion of *Sty*I site	Inactive
pQIS 17/ pQIS 18	Arg277Leu Ala280Phe Phe302Ala	Triple mutation with the introduction of *Acc*I site and deletion of *Hin*cII site	Active
pQIS 20	Phe302Ala	Single point mutation with the introduction of *Acc*I site	Partial
pQIS 21	Arg277Leu Ala280Phe	Quadruple mutation with the introduction of *Alw*NI site	Inactive

	Tyr289Asp Phe302Ala		
pQIS 22	Tyr289Asp	Single point mutation with the introduction of *Alw*NI site	Inactive
pQIS 36	Asn176Ile Ile177Asn	Double mutant with the introduction of *Pac*I site	Active
pQIS 37	Asn176Ile Ile177Asn Arg277Leu Ala280Phe Phe302Ala	Double mutation with the introduction of *Pac*I site in triple mutant of M.*Hha*I gene encoded by pQIS 18	Inactive
pQIS 41/ pQIS 50	Val116Gly Phe117Pro	Double mutant with the introduction of *Apa*I site	Inactive
pQIS 42	Val116Gly Phe117Pro Arg277Leu Ala280Phe Phe302Ala	Double mutant with the introduction of *Apa*I site in triple mutant of M.*Hha*I gene encoded by pQIS 18	Inactive
pQIS 43	Val116Gly Phe117Pro Asn176Ile Ile177Asn	Double mutant with the introduction of *Apa*I site in double mutant of M.*Hha*I gene encoded by pQIS 36	Inactive
pQIS 51	Val116Gly Phe117Pro Lys270Arg Thr271Ile	Double mutant with the introduction of *Eco*RI site in double mutant of M.*Hha*I gene encoded by pQIS 50	Inactive
pQIS 55	Lys270Arg Thr271Ile	Double point mutation with the introduction of *Eco*RI site	Partial
pQIS 74	Asn39Arg	Single point mutation with the introduction of *Nru*I site	Inactive

As a final attempt to check the flexibility of monospecific Mtase M.*Hha*I an experiment was carried out to exchange motifs in M.*Hha*I and M.SPRI. During this experiment site directed mutagenesis was also carried out to generate new restriction sites. Furthermore, protein expressed using in vitro transcription/translation was used for biochemical experiments including an *in vitro* methylation protection assay. The experiment leading to real time binding interactions between variants of M.*Hha*I and double stranded oligonucleotides has also been discussed.

The Asn 39 Arg mutagenesis in wild type M.*Hha*I made the enzyme catalytically inactive. It has been found from the analysis of crystal structure of M.*Hha*I (Cheng *et al.*, 1993) that Asn 39 makes close contacts with oxygen of Asp 16 and nitrogen of both Phe and Ala at amino acids position 18 and 19 in the presence of SAM. In the presence of DNA Asn 39 is found to make contacts to SAH and the oxygen of Gly 59 in addition to the contacts already stated in the presence of SAM. Asn is uncharged polar amino acid while Arg is positively charged amino acid. Introduction of a positively charged amino acid in the hydrophobic core structure of protein leads to an adverse structural arrangement. Molecular modelling software was used to predict the altered position of Arg 39 in M.*Hha*I and it appears that the side chain of Arg could well point towards SAM/SAH because of the orientation of the backbone of the protein. Therefore, this destabilization of the enzyme mass result in the activity of the enzyme being lost. A tentative model for the Asn 39 Arg mutation is shown in Figure 5.19.

The codon for Arg 39 has been exchanged with the codons for Asn in the protein encoded by construct pQIS 78. However, the activity of the encoded protein has not been restored. Although the contacts of Asn 39 with amino acids towards N-terminal end of the enzyme have been restored, however the hydrogen bond between Gly 59 and Asn 39 do not exists any more because of the exchange of sequence from the M.SPRI gene. Therefore, the restoration of methylation capacity of hybrid enzyme is not achieved. In the same manner, I believe that similar situation should have been arisen with M.SPRI and restoration of codons for Ser 33 in construct pQIS 80 did not encode an active chimaeric Mtase.

Among the hybrid Mtases that we prepared (Chapter Five) none of them showed Mtase activity. This could be due to a lower level of expression of hybrid genes or some inherent instability in the protein products. This was tested by expression of protein on a small scale in 5 ml culture from each of the hybrid Mtases and then

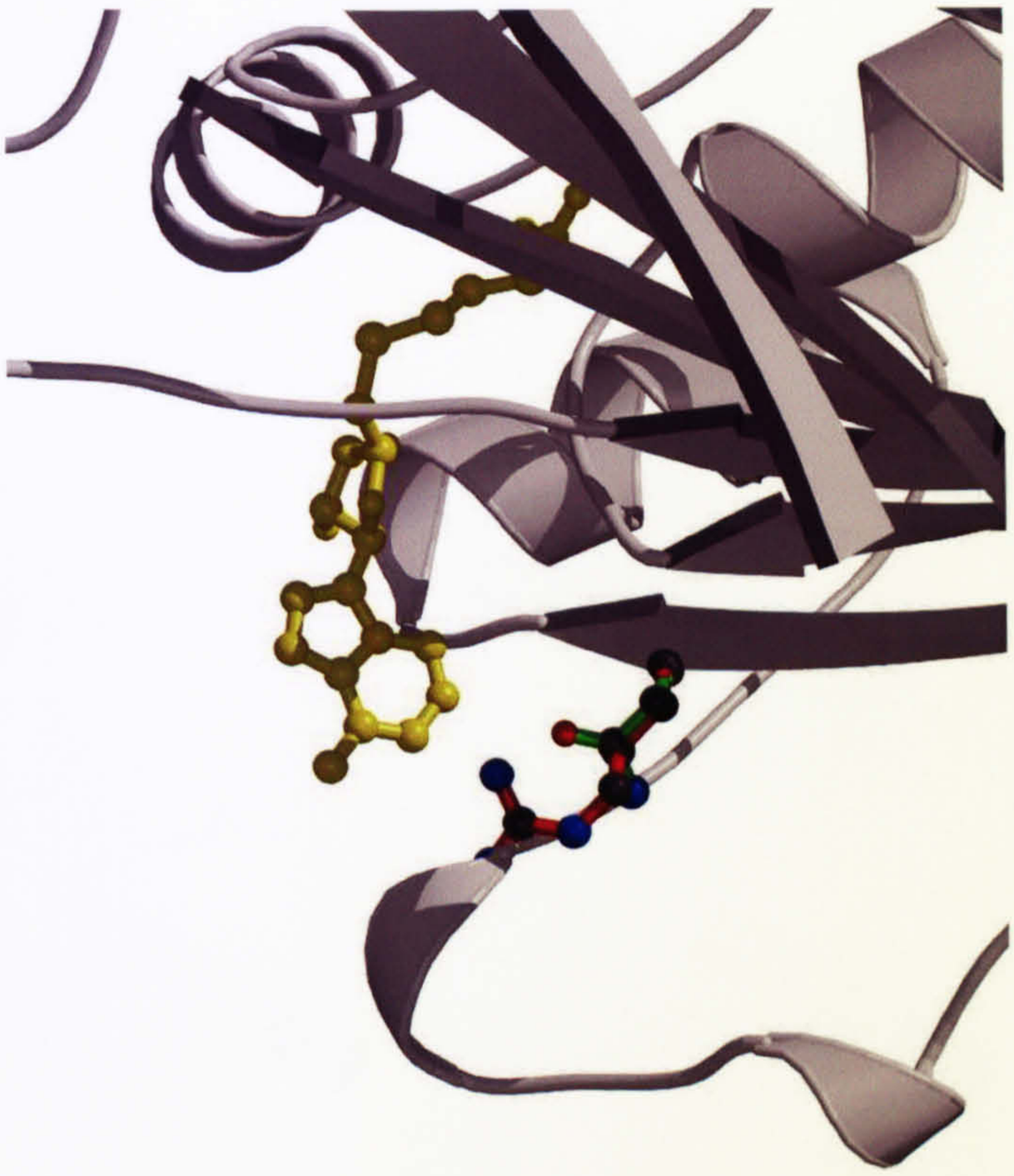


Figure 5.19: A model for the Asn 39 Arg in the structure of M.*Hha*I. SAH, Asn 39 and the modelled mutant Arg are shown in ball-and-stick with CPK coloured atoms and green and red sticks respectively. The Arg side chain comes very close to the SAH co-factor, the positive charge possibly affecting binding. The longer side chain of the Arg may cause steric problems as can be seen in the diagram where it clashes with a short helical region. Some secondary structure elements have been removed in this diagram for clarity. The image was created by Dr. Matthew J. Conroy. The figure was produced using Molscript (Kraulis (1991) *J. Appl. Cryst.* 24 946-950) and Raster3D, (Merritt & Murphy (1994) *Acta Cryst.* D50 869-873).

subjecting a crude cell extract to SDS PAGE. Four chimaeric constructs (i.e. pQIS 79, pQIS 81, pQIS 82 and pQIS 83) were induced for the expression of hybrid protein. In all cases, following induction, the hybrid proteins were clearly visible and the amounts were comparable to those of the wild type Mtases (Figure 5.20). This indicated that poor expression was not the cause of inactive protein. Therefore it is assumed that due to exchange of motifs between two different types of Mtases, the protein instability has occurred. This suggests that there may be some direct structural interactions between various regions within each class of Mtases.

In vitro translated wild type enzyme were shown to be competent in a methylation protection assay. The aim of these *in vitro* transcription/translation experiments was to establish whether the mutants retained some degree of structural integrity, as indicated by the retention of DNA recognition capacity. It was concluded from the experiments in which binary complex formation was monitored between the variants of M.*Hha*I enzyme and oligonucleotides, that there is no detectable sequence-specific difference in DNA association. The first step in DNA methylation is binding of DNA with the Mtase. The TRD region is mainly responsible for the recognition of the specific target site, which is then followed by sequence specific methylation. All the mutants of M.*Hha*I tested for DNA recognition carry mutations in or around motif IX. Different mutants that were enzymatically active, partially active and inactive all exhibited similar association profiles with the specific oligonucleotide. This suggests that motif IX plays a role in general stability of the enzyme core structure but have less role in the recognition of DNA sequences.

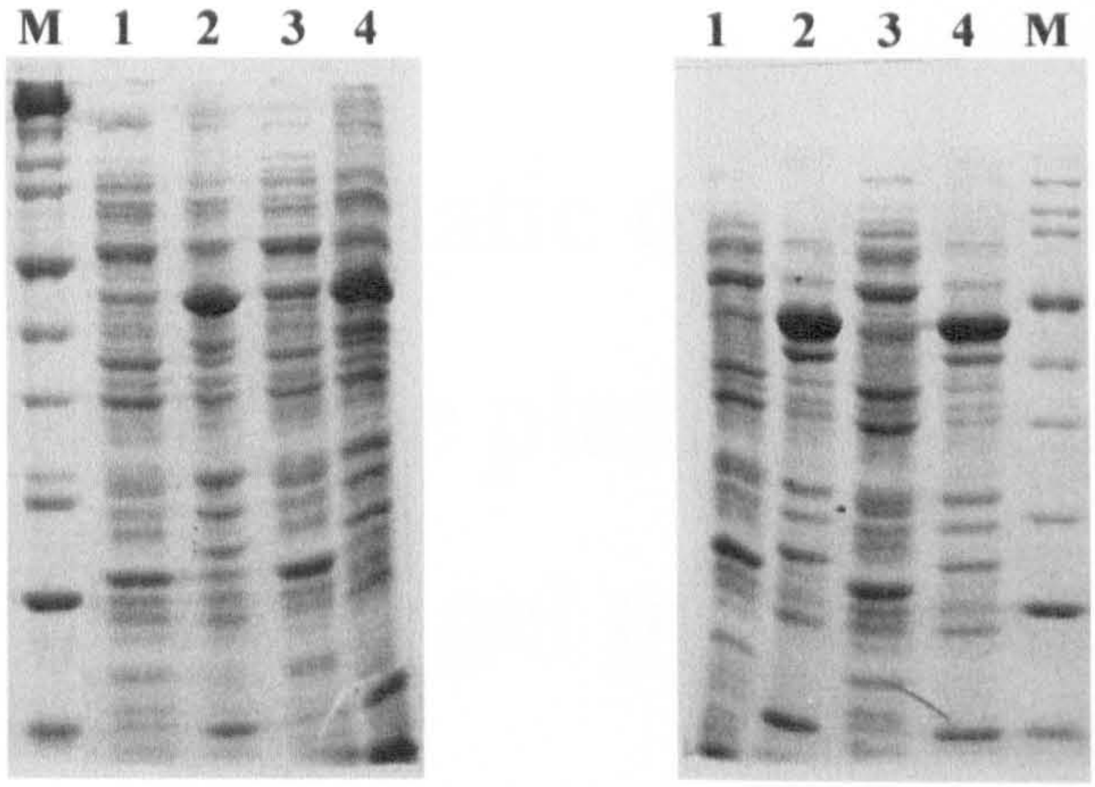


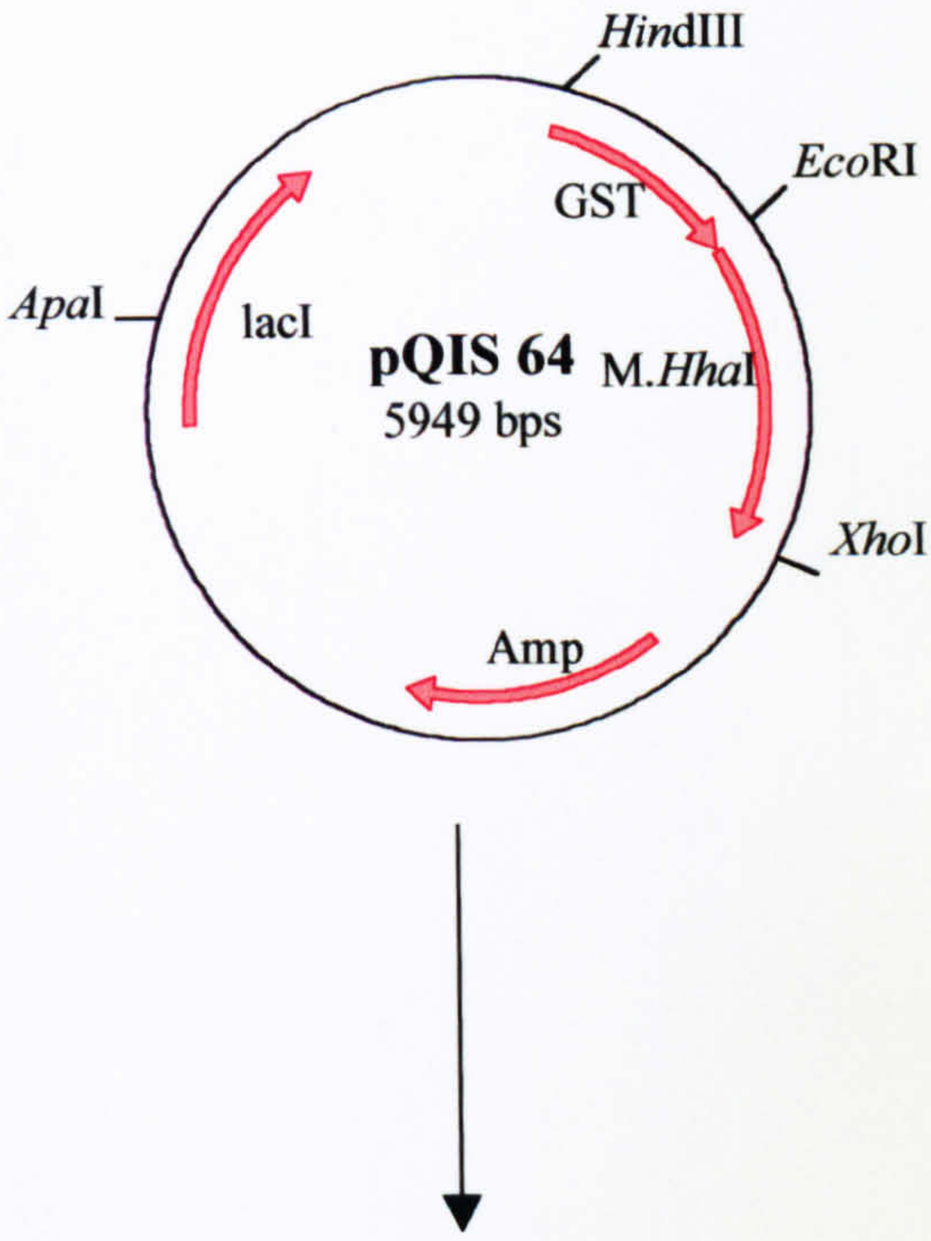
Figure 5.20: Protein expression for the chimaeric constructs pQIS 79, pQIS 81, pQIS 82 and pQIS 83. PAGE analysis after the protein induction and small scale extraction for the chimaeric constructs. The gel confirms that the protein is intact without any truncation and can be expressed by induction. **M** is broad range protein molecular weight marker as shown in chapter two and appendix in both gels. The lane order is as follows.

Gel 1:
pQIS 79 Soluble fraction
pQIS 79 Pallet fraction
pQIS 81 Soluble fraction
pQIS 81 Pallet fraction

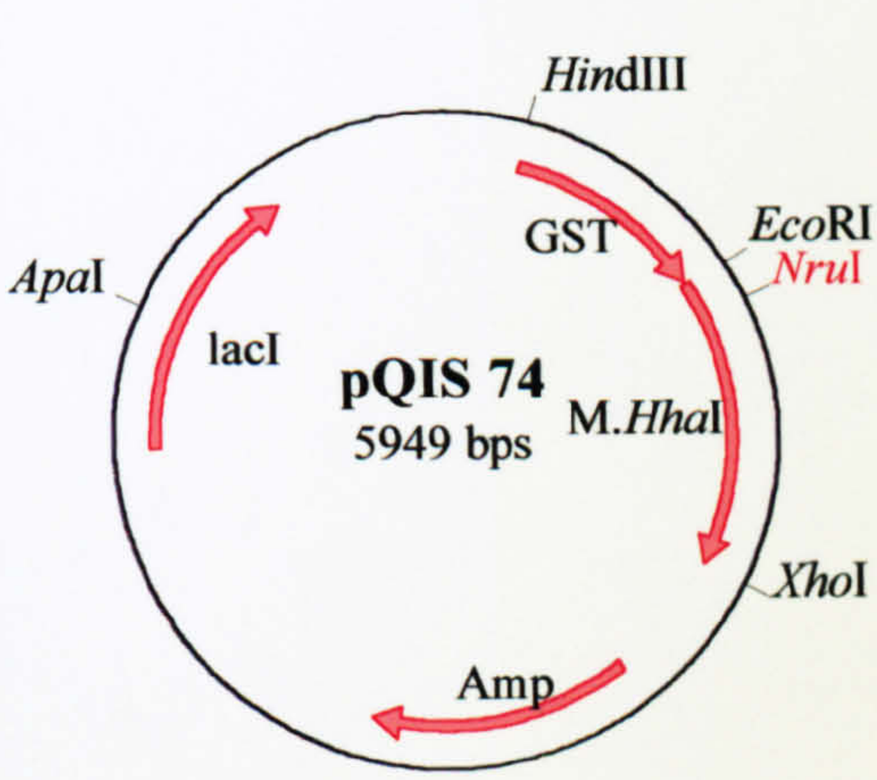
Gel 2:
pQIS 82 Soluble fraction
pQIS 82 Pallet fraction
pQIS 83 Soluble fraction
pQIS 83 Pallet fraction

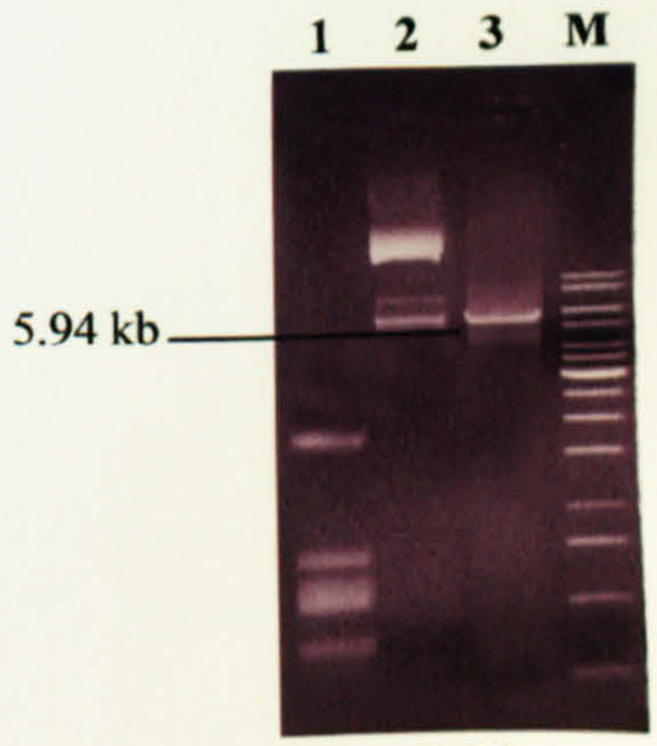
Schematic diagrams showing the plasmid sketches with gel pictures

(Including Figures 5.3 ~ 5.11)



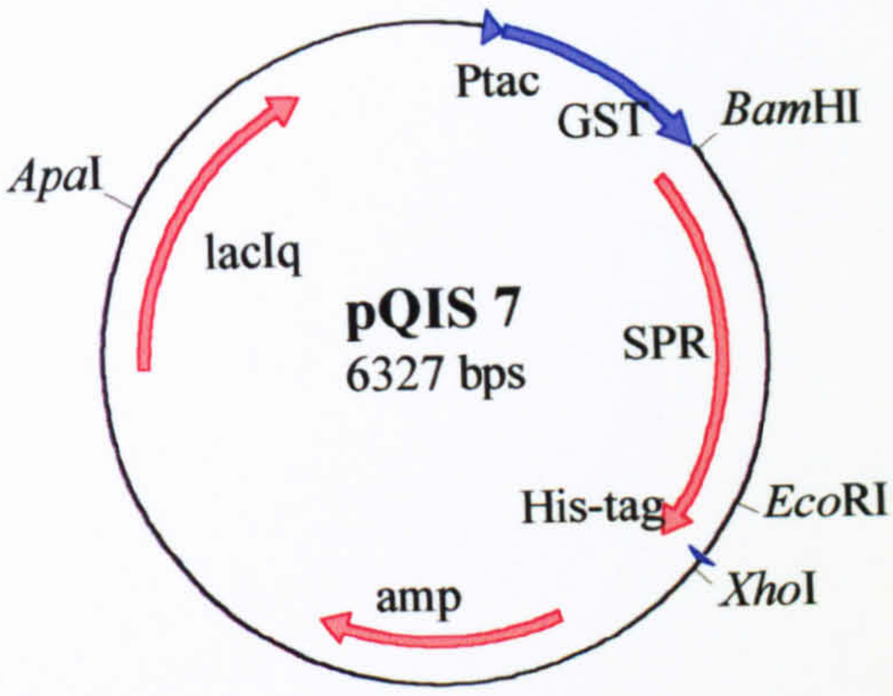
Gel 1

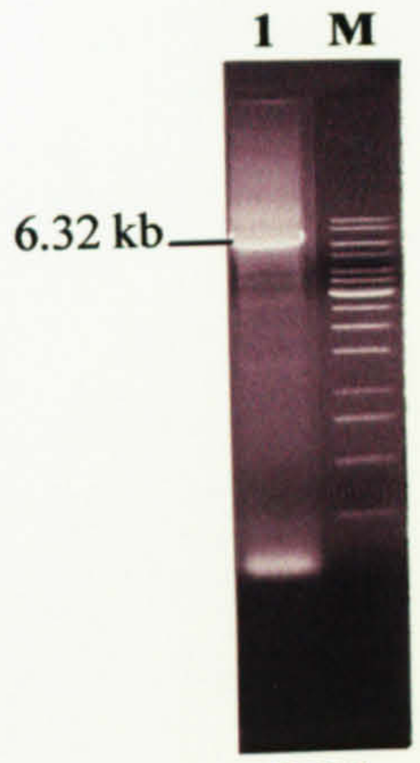




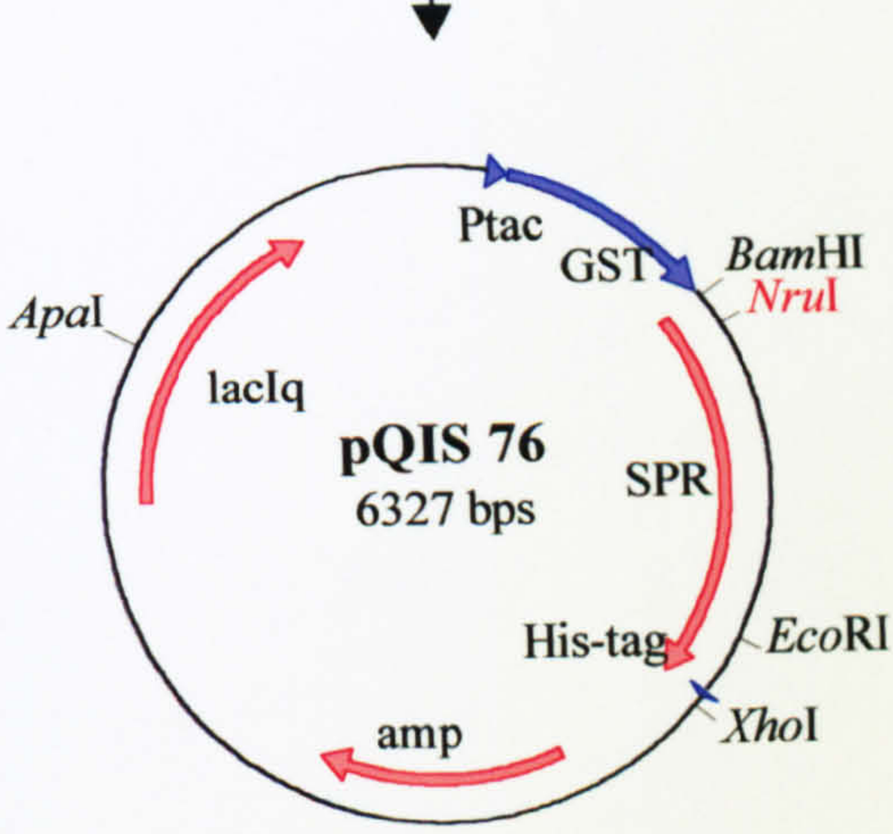
Gel 2

Figure 5.3: Schematic representation showing the construction of pQIS 74 (mutant of M.*Hha*I encoding for Asn 39 Arg). The newly introduced *Nru*I site is shown in red. **Gel 1** shows the PCR product generated after site-directed mutagenesis. **M:** GeneRuler™ 1 kb DNA ladder. **Gel 2, Lane 1:** pQIS 74 digested with *Hha*I (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this plasmid). **Lane 2:** Control pQIS 64 cut with *Nru*I (Uncut DNA shows that *Nru*I site does not exist in this plasmid) **Lane 3:** pQIS 74 cut with *Nru*I to verify the identity of the new construct. (Linear DNA due to unique *Nru*I site confirms the formation of pQIS 74). **M:** GeneRuler™ 1 kb DNA ladder.



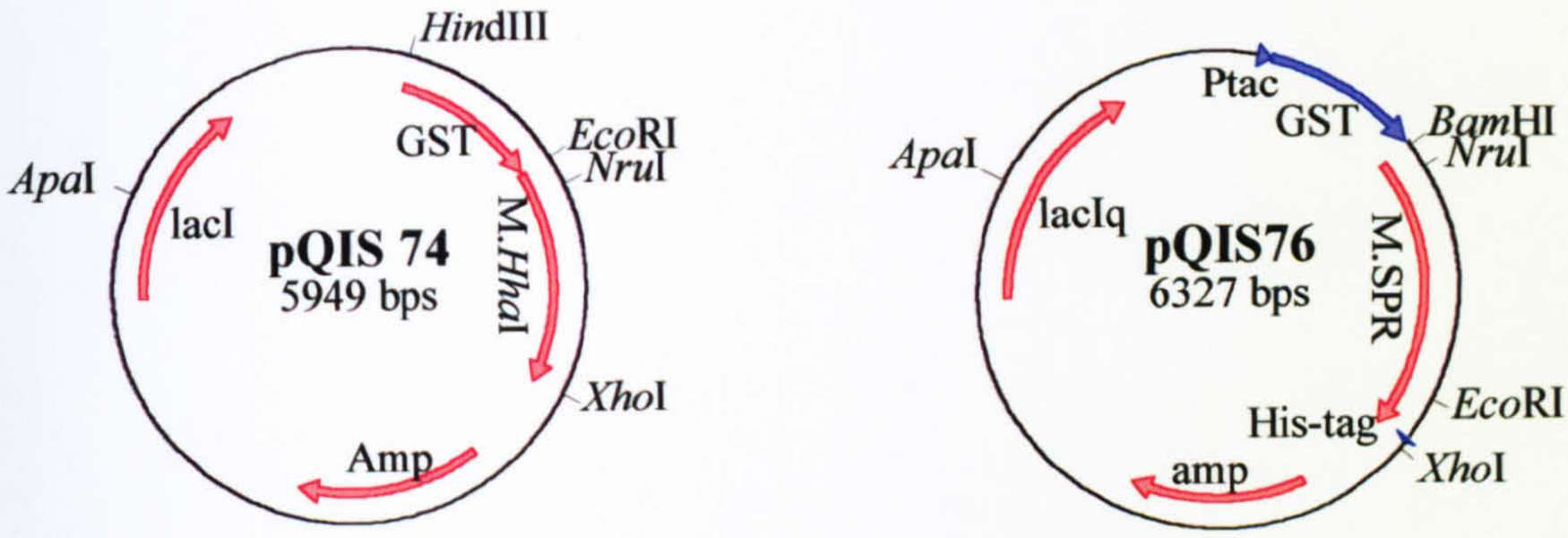


Gel 1



Gel 2

Figure 5.4: Schematic representation showing the construction of pQIS 76 which encodes for Ser 33 Arg codon in M.SPRI. The newly introduced *Nru*I site is shown in red. **Gel 1** shows the PCR product generated after site-directed mutagenesis. **Gel 2, Lane 1:** pQIS 76 digested with *Hae*III (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this plasmid). **Lane 2:** pQIS 76 digested with *Nru*I for verification of the identity of new construct. (DNA has been linearised due to the presence of unique *Nru*I site). **M:** GeneRuler™ 1 kb DNA ladder in both gels.



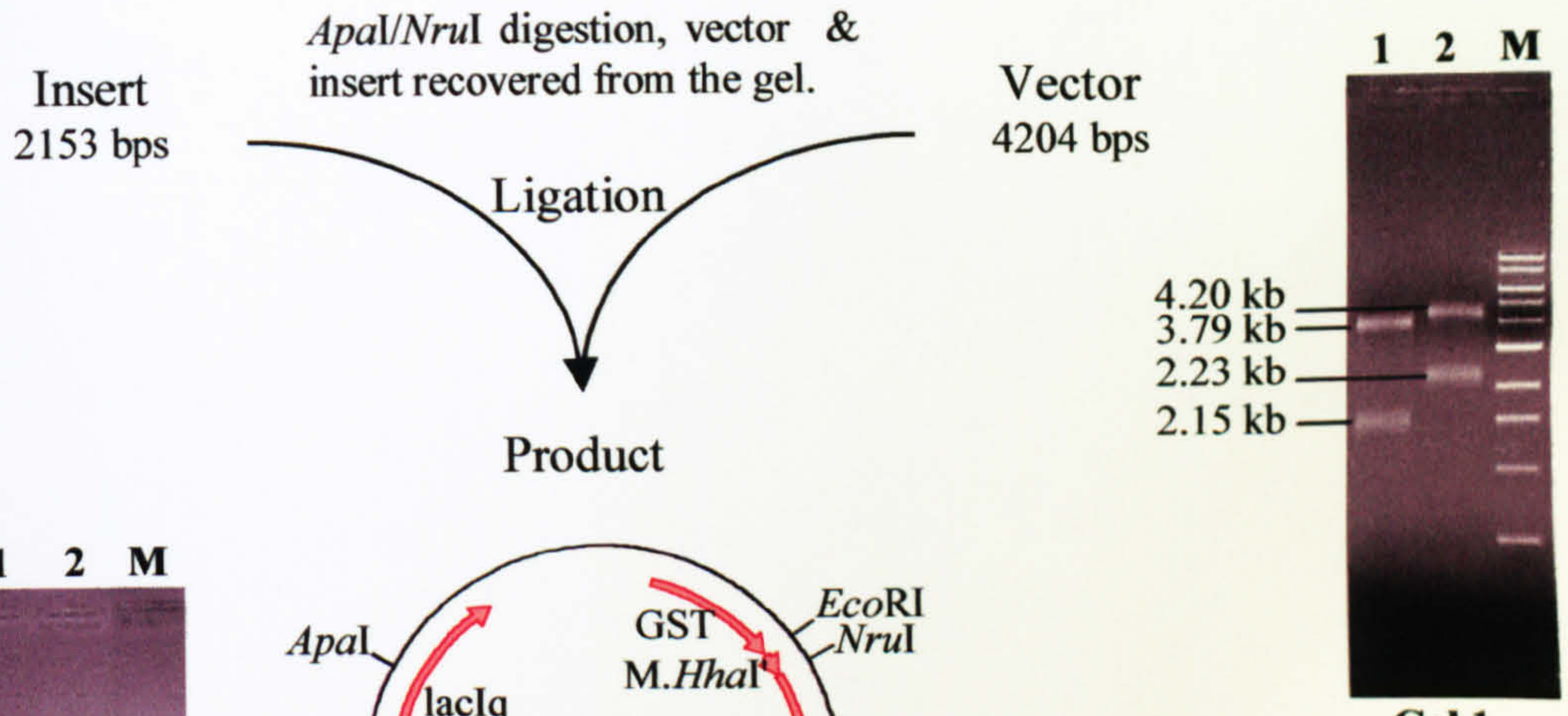
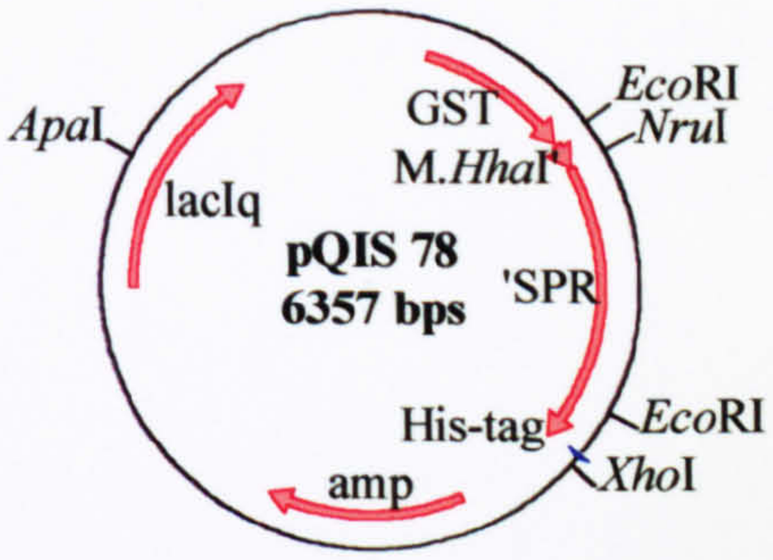


Figure 5.5: Schematic representation showing the construction of pQIS 78 (chimaera of M.SPRI and M.*Hha*I). Gel 1 shows the digestion of pQIS 74 and pQIS 76 to gel purify the vector and insert DNA for the construction of pQIS 78**. Lane 1:** pQIS 74 digested with *Apa*I and *Nru*I. (Fragment sizes are 3796 and 2153 bps). **Lane 2:** pQIS 76 digested with *Apa*I and *Nru*I. (Fragment sizes are 4204 and 2123 bps). **Gel 2: Lane 1:** pQIS 78 cut with *Msp*I for activity check. (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this plasmid). **Lane 2:** pQIS 78 cut with *Eco*RI for the identity of verification of new construct. (Fragment sizes are 5203 and 1154 bps). **M:** GeneRuler™ 1 kb DNA ladder in both gels.



Gel 1

Gel 2

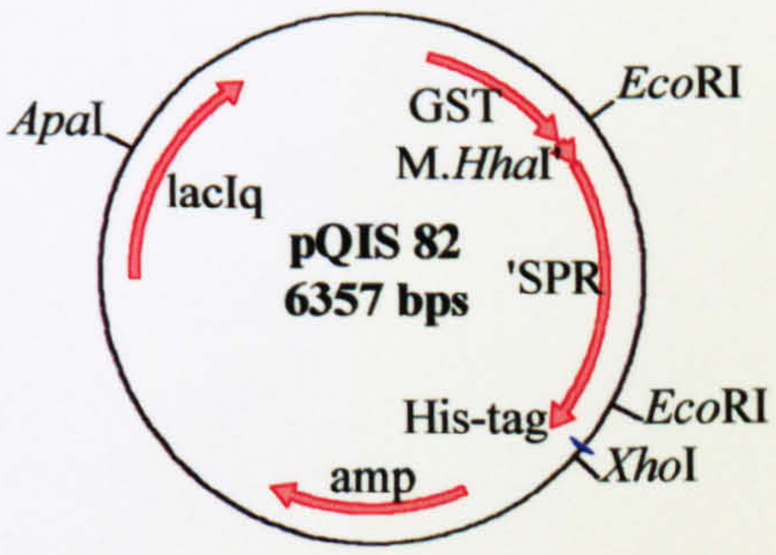


Figure 5.6: Schematic representation showing the construction of pQIS 82 (chimaera of M.SPRI and M.*Hha*I). This construct is equivalent to pQIS 78 made earlier but with the restoration of original codon sequence of Asn of M.*Hha*I by SDM at the fusion point of M.*Hha*I and M.SPRI sequence**. Gel 1** shows the generation of PCR product after SDM. **Gel 2: Lane 1:** pQIS 78 cut with *Nru*I as control digestion. (Linear DNA). **Lane 2:** pQIS 82 digested with *Nru*I. (Uncut DNA signifies that the mutagenesis has been successful and *Nru*I site has been abolished). **Lane 3:** pQIS 82 cut with *Msp*I for activity check. (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this plasmid). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

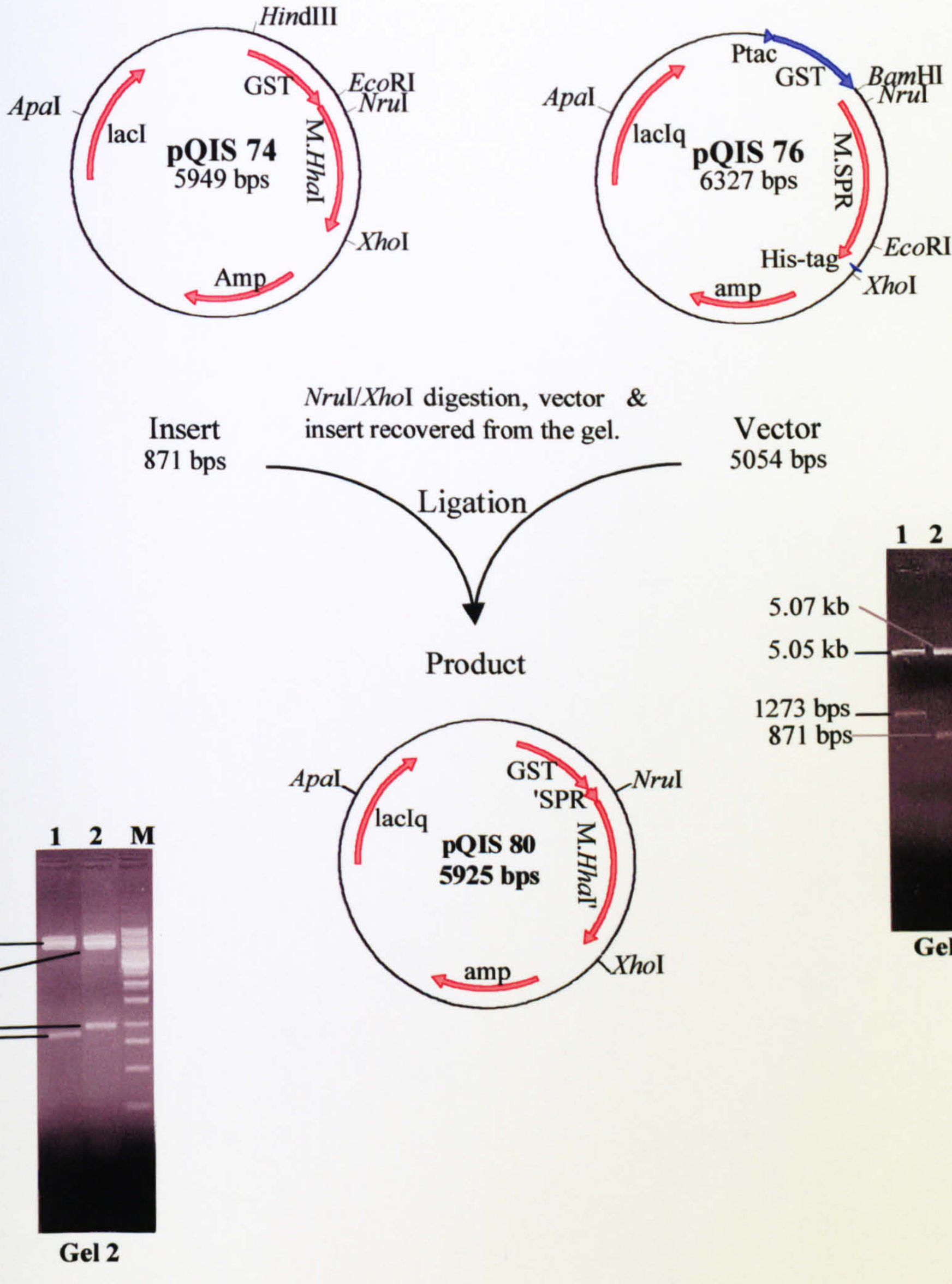
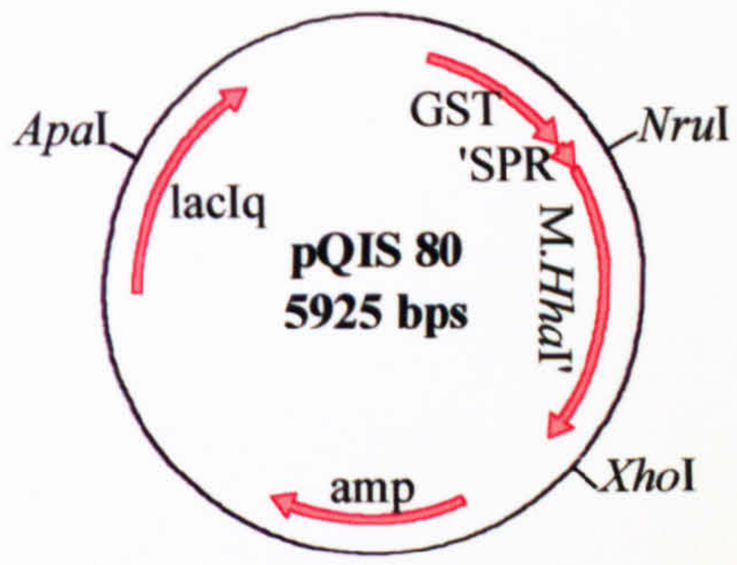
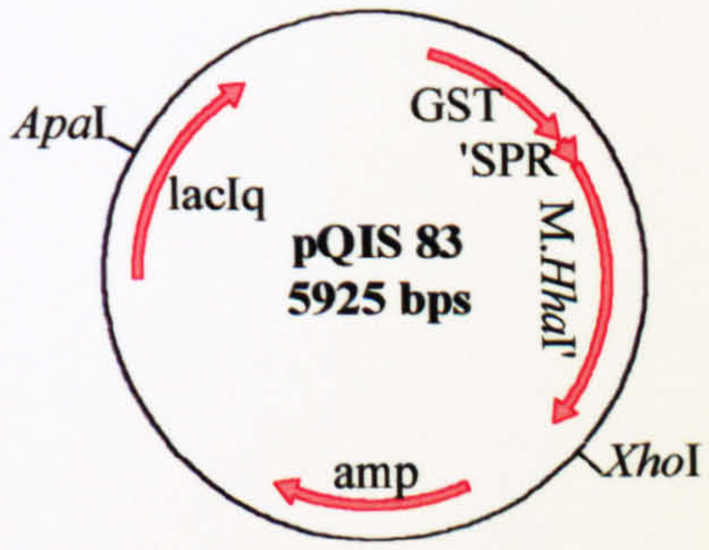


Figure 5.7: Schematic representation showing the construction of chimaeric construct pQIS 80 which contains sequence of motif I of M.SPR and the sequence of motif II-X of M.*Hha*I. Gel 1: Lane 1 shows the digestion of pQIS 76 with *Nru*I and *Xho*I. (Fragment sizes are 5054 and 1273 bps). **Lane 2:** pQIS 74 cut with *Nru*I and *Xho*I. (Fragment sizes are 5078 and 871 bps). **Gel 2: Lane 1:** pQIS 80 cut with *Nru*I and *Xho*I. (Fragment sizes are 5054 and 871 bps). **Lane 2:** pQIS 80 cut with *Bam*HI and *Xho*I. (Fragment sizes are 4951 and 974 bps). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

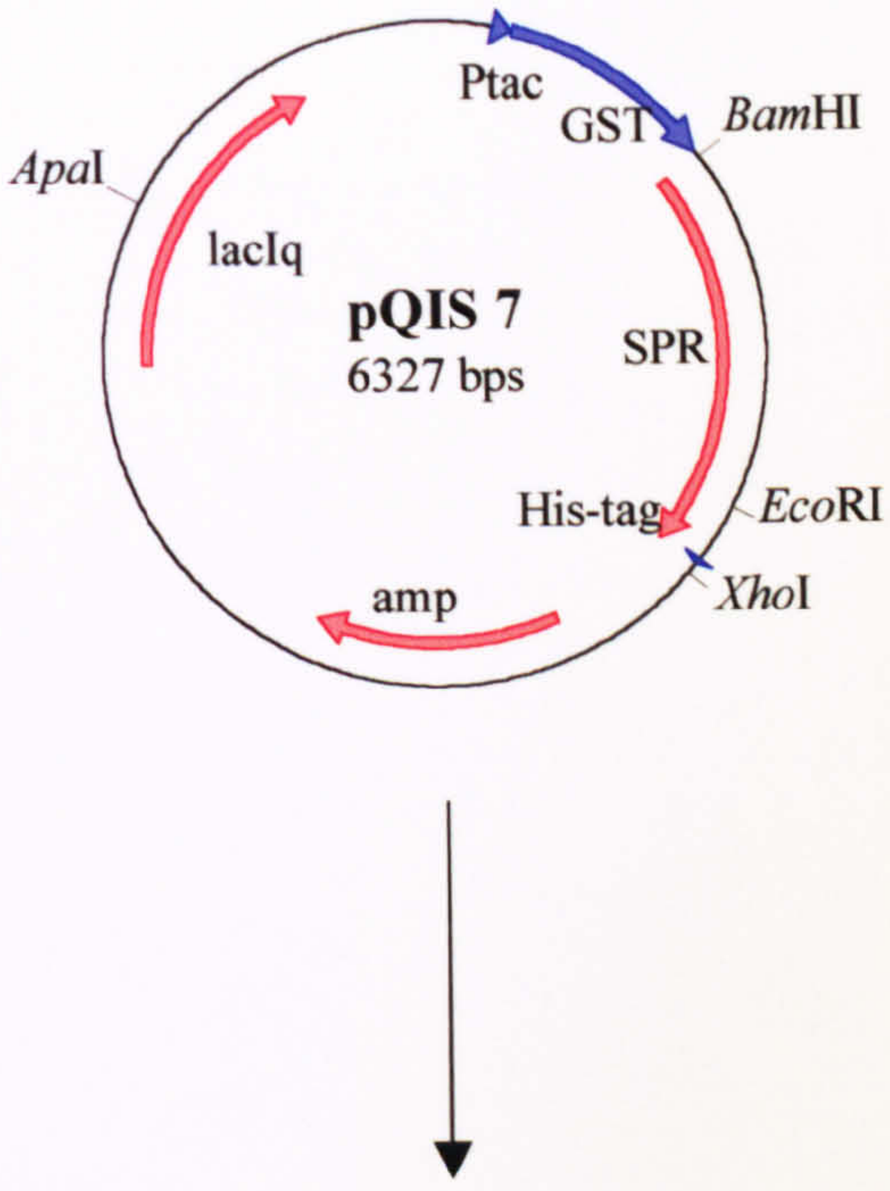


Gel 1

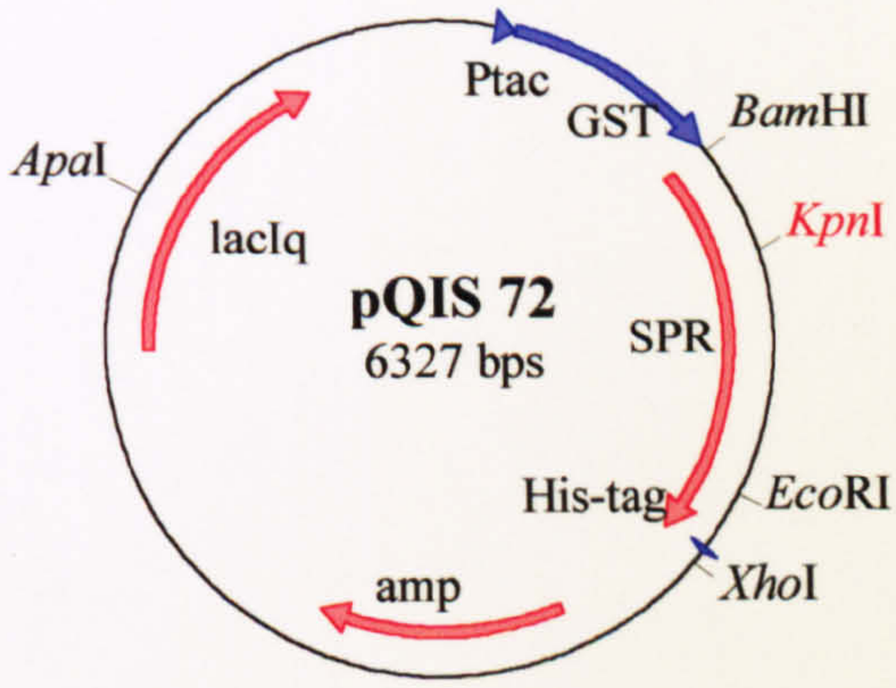


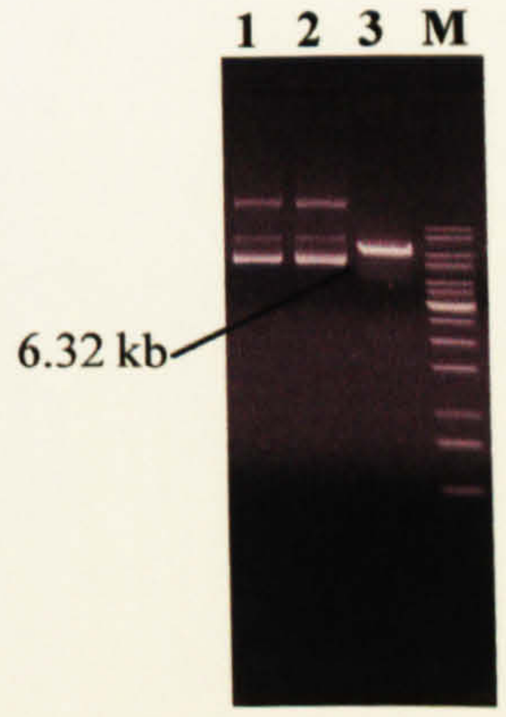
Gel 2

Figure 5.8: Schematic representation showing the construction of pQIS 83 (Chimaera of M.SPRI and M.*Hha*I). This construct is equivalent to pQIS 80 made earlier but now restore the original codon sequence of Ser of M.SPRI by SDM at the fusion point of M.SPRI and M.*Hha*I sequence**. Gel 1** shows the generation of PCR product after SDM. **Gel 2: Lane 1:** pQIS 83 cut with *Nru*I (Uncut DNA signifies that the mutagenesis has been successful and *Nru*I site has been abolished). **Lane 2:** pQIS 83 cut with *Msp*I for activity check. (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this plasmid). **M:** GeneRuler^{TM} 1 kb DNA ladder in both gels.



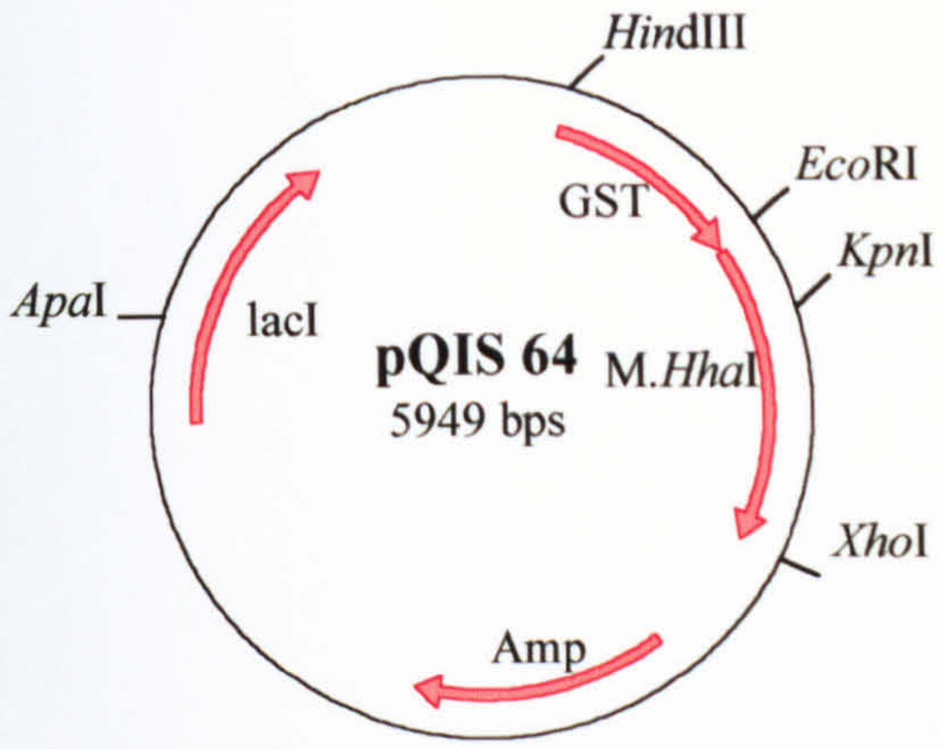
Gel 1

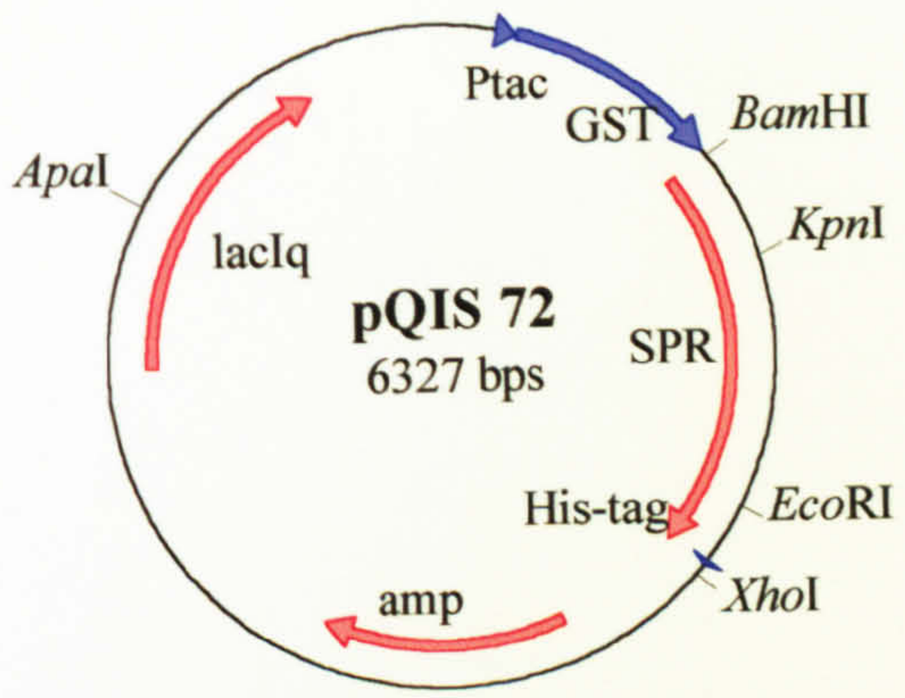




Gel 2

Figure 5.9: Schematic representation showing the construction of pQIS 72 which encodes for wild type M.SPRI. Gel 1 shows the PCR product generated (after site-directed mutagenesis for the introduction of unique *Kpn*I site, silent mutation). **Gel 2, Lane 1:** pQIS 72 digested with *Msp*I (The DNA is protected from restriction digestion indicating that an active protein is encoded by this plasmid). **Lane 2:** pQIS 72 cut with *Hae*III (The DNA is protected from restriction digestion indicating that an active protein is encoded by this plasmid). **Lane 3:** pQIS 72 cut with *Kpn*I to verify the new construct. (Linear DNA due to unique *Kpn*I site confirms the formation of pQIS 72). **M:** GeneRuler™ 1 kb DNA ladder in both gels.



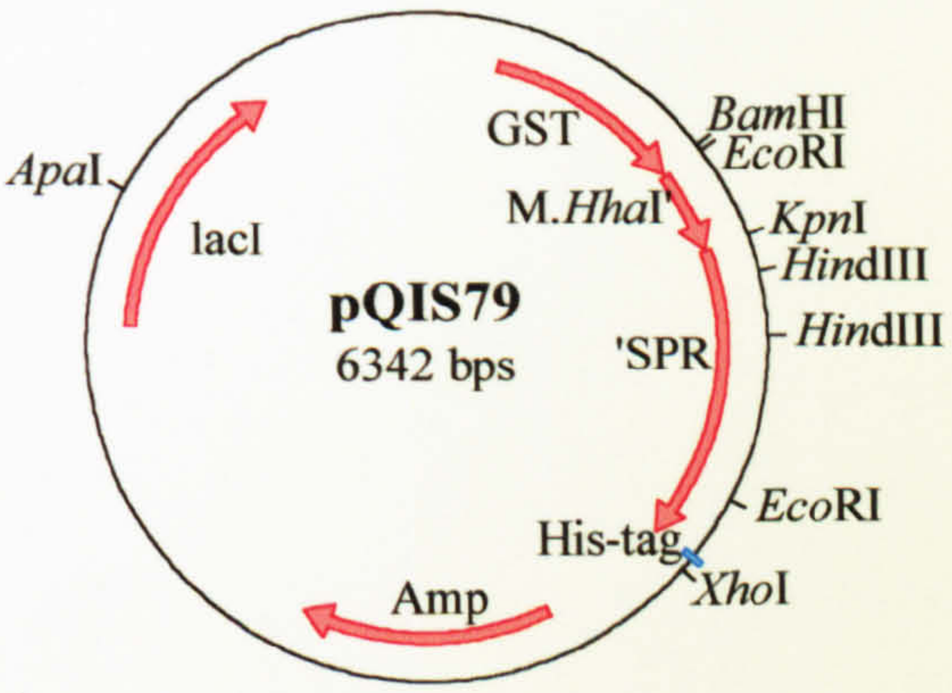


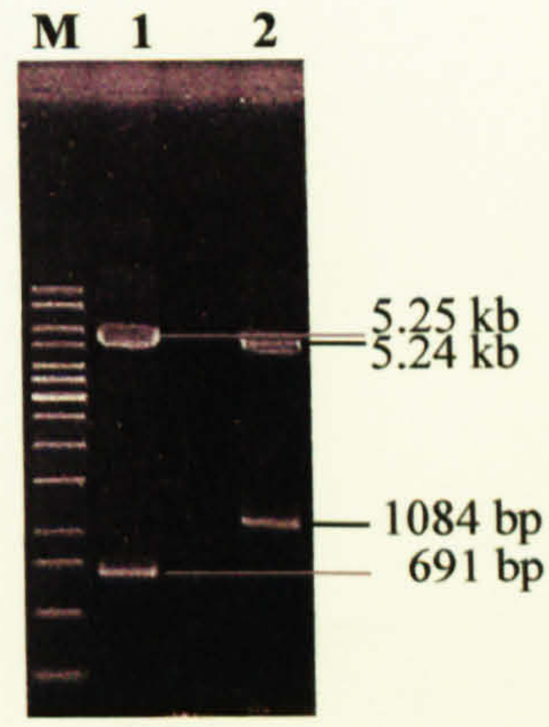
Vector
5258 bps

*Kpn*I/*Xho*I digestion, vector & insert recovered from the gel.

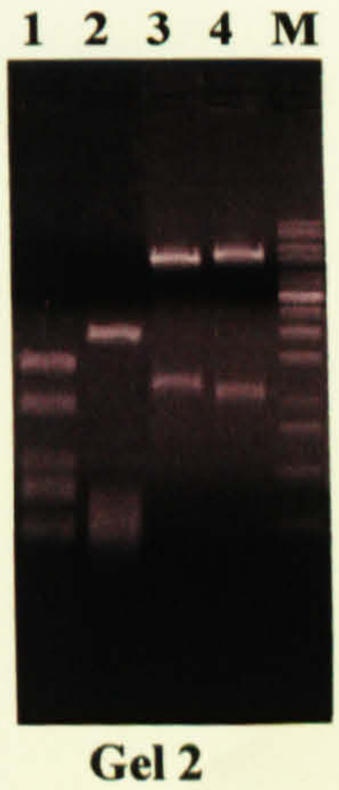
Ligation

Insert
1084 bps



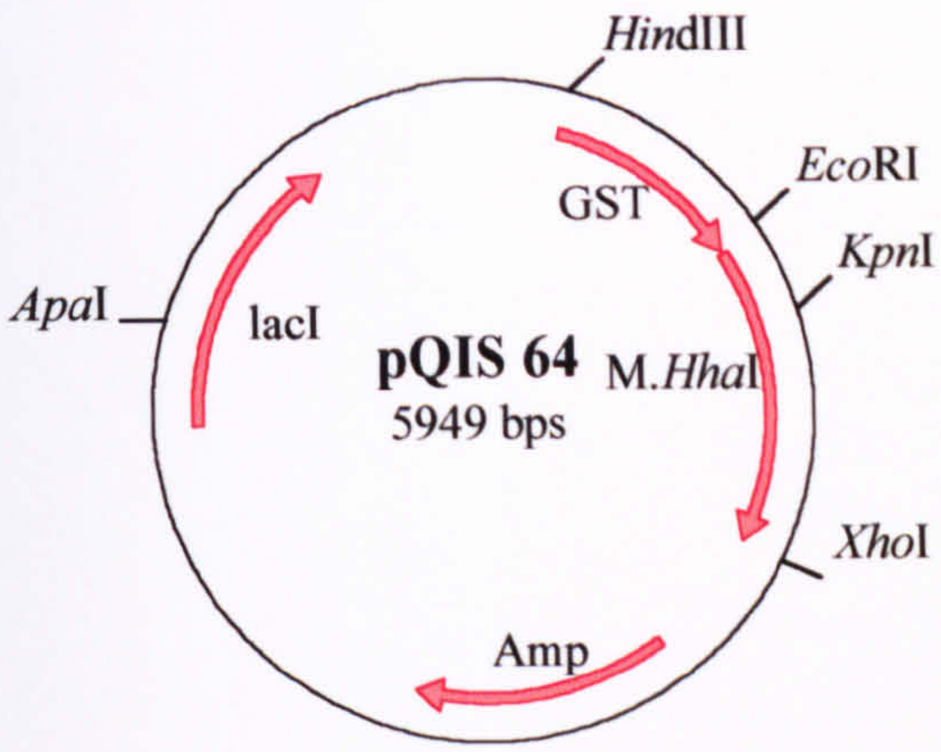


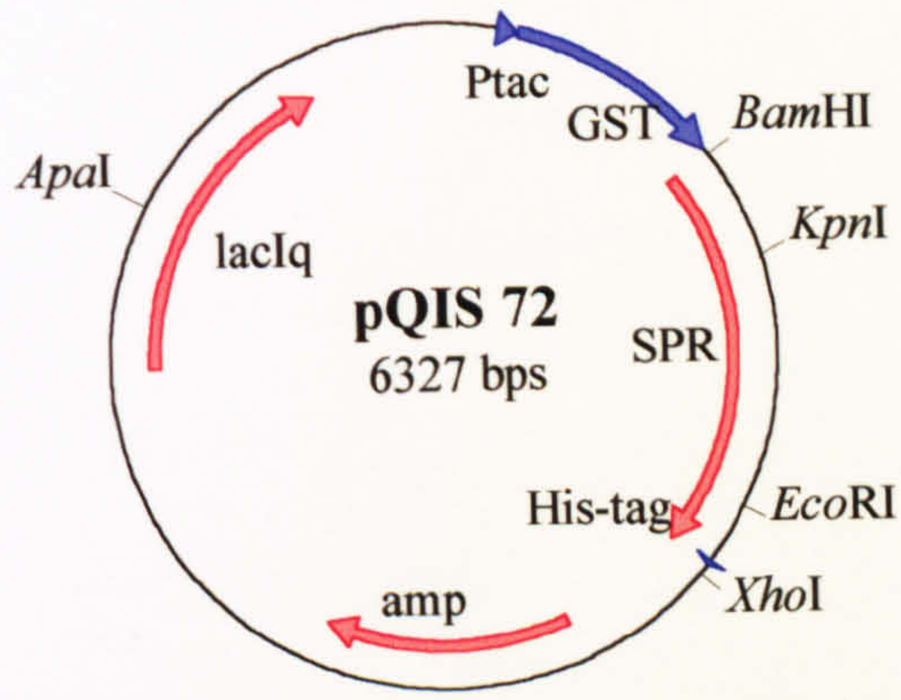
Gel 1



Gel 2

Figure 5.10: Schematic representation showing the construction of pQIS 79. Gel 1 shows the restricted DNA fragments for the recovery through gel. **Lane 1**: pQIS 64 cut with *Kpn*I and *Xho*I to extract the 5258 bps vector. **Lane 2**: pQIS 72 cut with *Kpn*I and *Xho*I to gel extract the 1084 bp insert. **Gel 2, Lane 1:** pQIS 79 digested with *Msp*I (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this plasmid). **Lane 2:** pQIS 79 cut with *Hha*I (The DNA is cleaved indicating that an inactive protein is encoded by this plasmid). **Lane 3:** pQIS 79 cut with *Eco*RI to verify the new construct. (Fragment sizes are 1145 and 5197 bps). **Lane 4:** pQIS 79 cut with *Kpn*I and *Xho*I to verify the new construct. (Fragment sizes are 1084 and 55258 bps). **M:** GeneRuler^{TM} 1 kb DNA ladder in both gels.



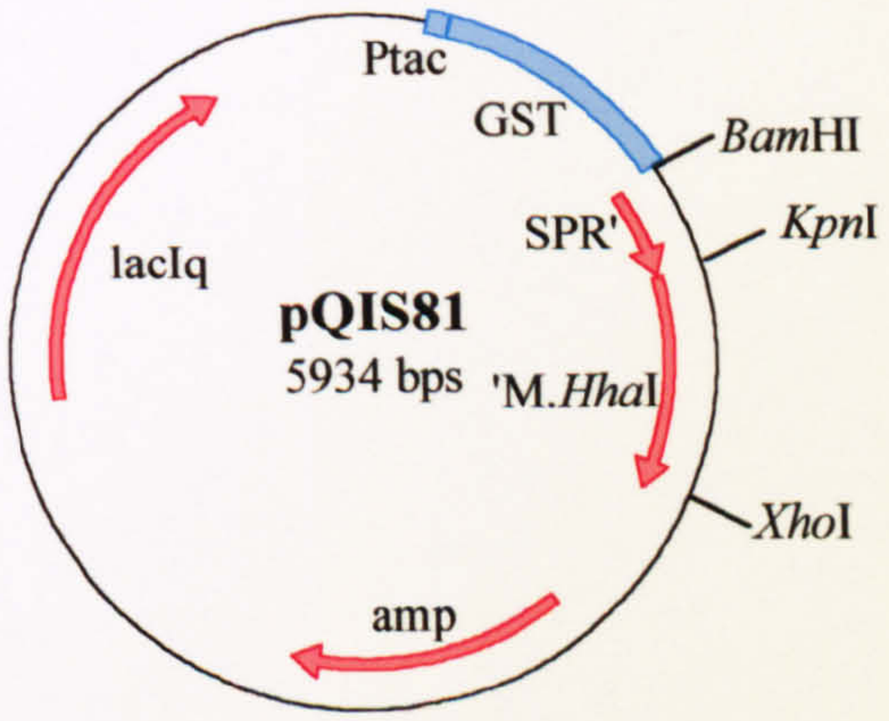


Insert
691 bps

*Kpn*I/*Xho*I digestion, vector & insert recovered from the gel.

Vector
5243 bps

Ligation



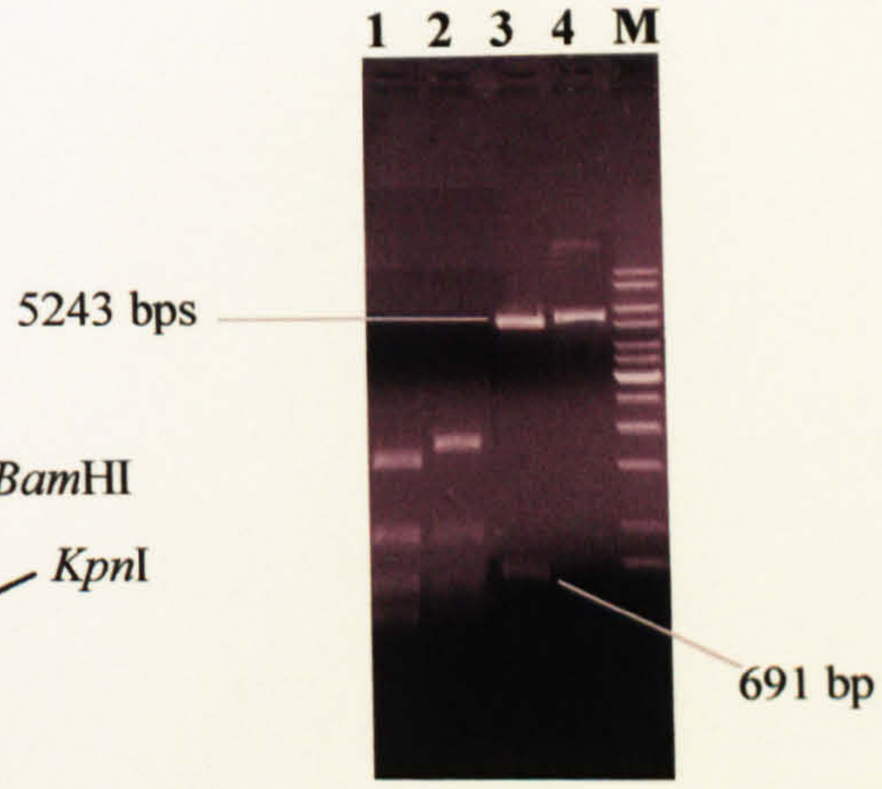


Figure 5.11: Schematic representation showing the construction of pQIS 81. The insert and vector DNA fragments for this ligation were recovered from the gel 1 shown in figure 5.10. **Gel 1, Lane 1:** pQIS 81 digested with *Msp*I (The DNA is not protected from restriction digestion indicating that an inactive protein is encoded by this chimaeric construct). **Lane 2:** pQIS 81 cut with *Hha*I (The DNA is cleaved indicating that an inactive protein is encoded by this construct). **Lane 3:** pQIS 81 cut with *Kpn*I and *Xho*I to verify the new construct. (Fragment sizes are 691 and 5243 bps). **Lane 4:** pQIS 81 cut with *Eco*RI to verify the new construct (uncut DNA, no *Eco*RI site). **M:** GeneRuler™ 1 kb DNA ladder.

CHAPTER SIX

DEVELOPMENT OF A "BACTERIAL TWO-HYBRID SYSTEM" BASED ON THE REVERSIBLE ASSEMBLY OF AN ENGINEERED FORM OF M.SPRI

CHAPTER SIX

DEVELOPMENT OF A "BACTERIAL TWO-HYBRID SYSTEM" BASED ON THE REVERSIBLE ASSEMBLY OF AN ENGINEERED FORM OF M.SPRI

6.1: The Two-Hybrid System

6.1.1: INTRODUCTION

Two-hybrid systems are based on a genetic method that uses transcriptional activity as an indirect measure of protein-protein interaction (Phizicky and Fields, 1995). It relies on the modular nature of many site specific Transcriptional Activators which typically consist of a DNA Binding Domain (X) (targets the activator to the gene that will be expressed) and a Transcriptional Activation domain (Y) (contacts other proteins of transcriptional machinery to enable Transcription to occur) (Phizicky and Fields, 1995; Drees, 1999). In the Two-Hybrid system the two domains of the activator need not be covalently linked and can be brought together by the interaction of two proteins (Phizicky and Fields, 1995).

This system requires two chimaeric proteins to be constructed:

1. DNA binding domain fused to some protein (X)
2. Transcription activation domain fused to some protein (Y)

These two hybrids are expressed in a cell containing one or more reporter genes. If protein X and Y interact, they create a functional transcriptional activator by bringing the activation domain into close proximity with the DNA-binding domain. This can be detected by expression of the reporter gene (Phizicky and Fields, 1995) (Figure 6.1).

The Yeast Two-Hybrid system has been used extensively since its original description almost 10 years ago to identify protein-protein interactions (Drees, 1999). Many modifications have been introduced in the original Two-Hybrid system to expand the range of possible investigation of macromolecular interactions (Drees, 1999). These powerful genetic selection methods are proving useful in proteomic projects aimed at generating macromolecular interaction maps.

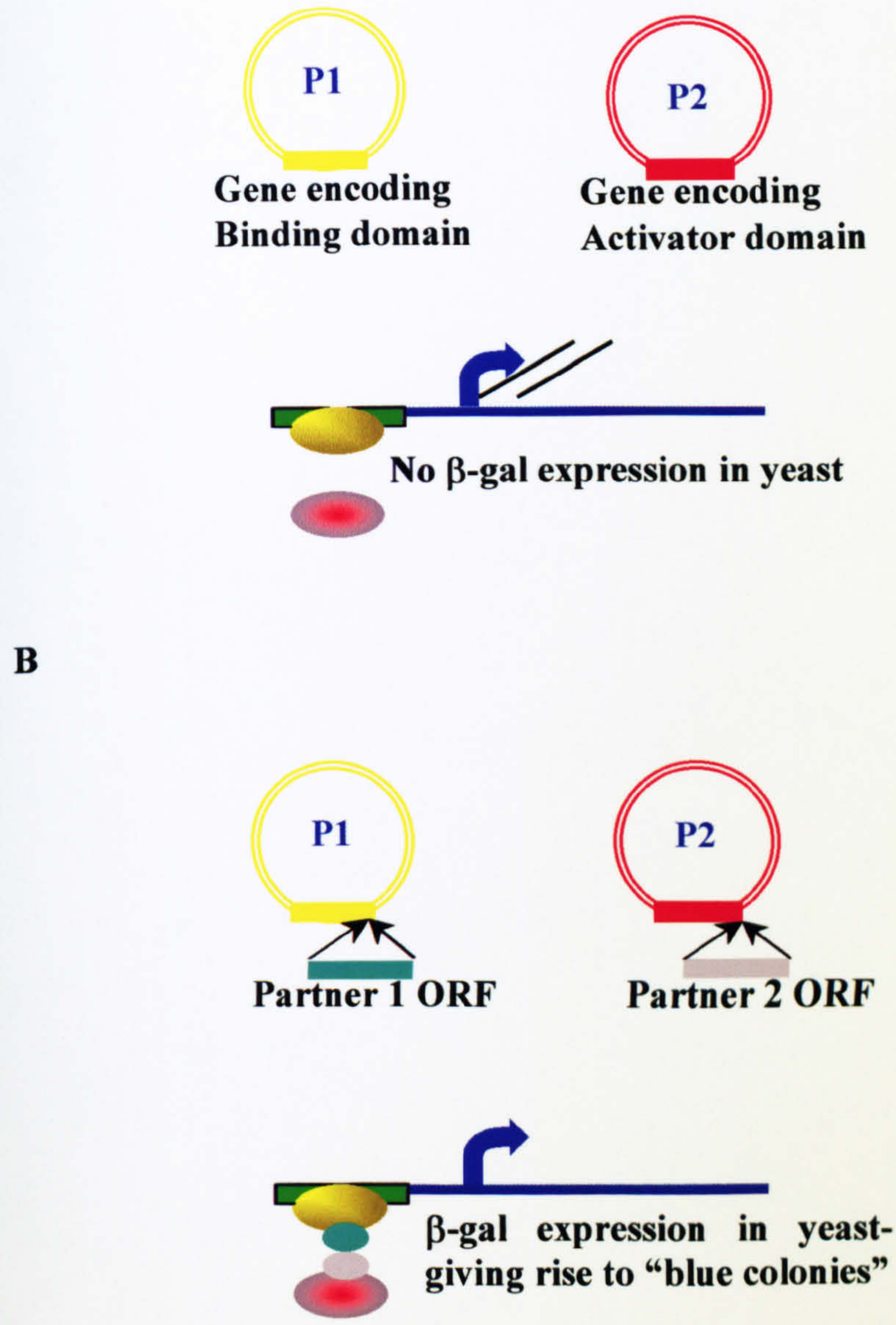


Figure 6.1: A simplified model for the current basis of the yeast two hybrid system (Fields, 1989).
A: In the absence of interacting partner, the two domains are unable to come close enough to activate the reporter gene. **B:** The partners 1 and 2 promote the interaction of activation domain and binding domain and as a result of this, activation of a reporter gene (β-galactosidase in the above example) takes place and this can be easily detected using the chromophore X-Gal in the growth medium ("blue colonies").

The Two-Hybrid system has certain advantages over conventional methods for studying protein-protein interactions, such as co-immuno-precipitation, co-fractionation and affinity chromatography. In this system, there is no need to purify any of the candidate proteins or produce antibodies. No prior knowledge of protein function is required. It has ultimate flexibility for enabling the detection of completely novel sets of interacting proteins. This system can be used for testing interactions of pairs of proteins, groups of proteins in pair wise combination or entire libraries of genes against either individual proteins or against another library. Furthermore, the nucleotide sequences of those clones that are defined as positive in a two hybrid screens can be readily obtained by recovering the corresponding plasmids.

Gene regulatory proteins generally bind to specific DNA sequences. Several DNA binding motifs have been described and transcription factors often are classified according to the type of DNA binding domain they contain e.g. Helix-Turn Helix, C_2H_2 Zinc Finger, and Homeodomain (Nelson and Cox, 2000). Such regulatory proteins contain domains not only for DNA binding but also for protein-protein interaction. Several well-defined structural domains have been described that are devoted to the interactions required for dimer formation, which is generally a prerequisite for DNA binding. Like DNA-binding motifs, these structural motifs, which promote oligomer formation, contain characteristic consensus sequences such as the Leucine Zipper motif and the basic Helix-Loop Helix motif (Nelson and Cox, 2000).

6.1.2: LEUCINE ZIPPER PROTEINS

One protein motif whose function can be measured effectively using the Two-Hybrid system is the "leucine zipper". The leucine zipper is a structure that mediates the function of several eukaryotic gene regulatory proteins. The zipper consists of a periodic repetition of leucine residues at approximately every seventh position, and regions containing them typically span 8 turns of alpha helix (Landschulz, 1988). Initially it was thought that the Leu residues interdigitated and hence they were named "zippers". Now it is known that the leucine side chains that extend from one helix interact with those from a similar helix, hence facilitating dimerisation in the form of a coiled-coil (Nelson and Cox, 2000). Leucine zippers are present in many gene regulatory proteins, and examples include the CREB proteins, Jun/AP1 transcription factors, fos oncogene and fos-related proteins, C-myc, L-myc and N-

myc oncogenes, and so on. Leucine zipper proteins can be divided into several families on the basis of the properties of polypeptide region adjacent to the zipper. One such leucine zipper family contains a region of dense positive charge immediately adjacent to the amino terminus of the zipper domain (Figure 6.2). This family is generally referred as the B-zip family, which can be further subdivided into the two sub-families according to observed degree of similarity between amino acids in the leucine zipper and basic region (Kerppola and Curran, 1991):

A: The Fos related subfamily e.g. Fos, FosB, Fra1, Fra2 etc.

B: The Jun related subfamily e.g. Jun, JunB, JunD etc.

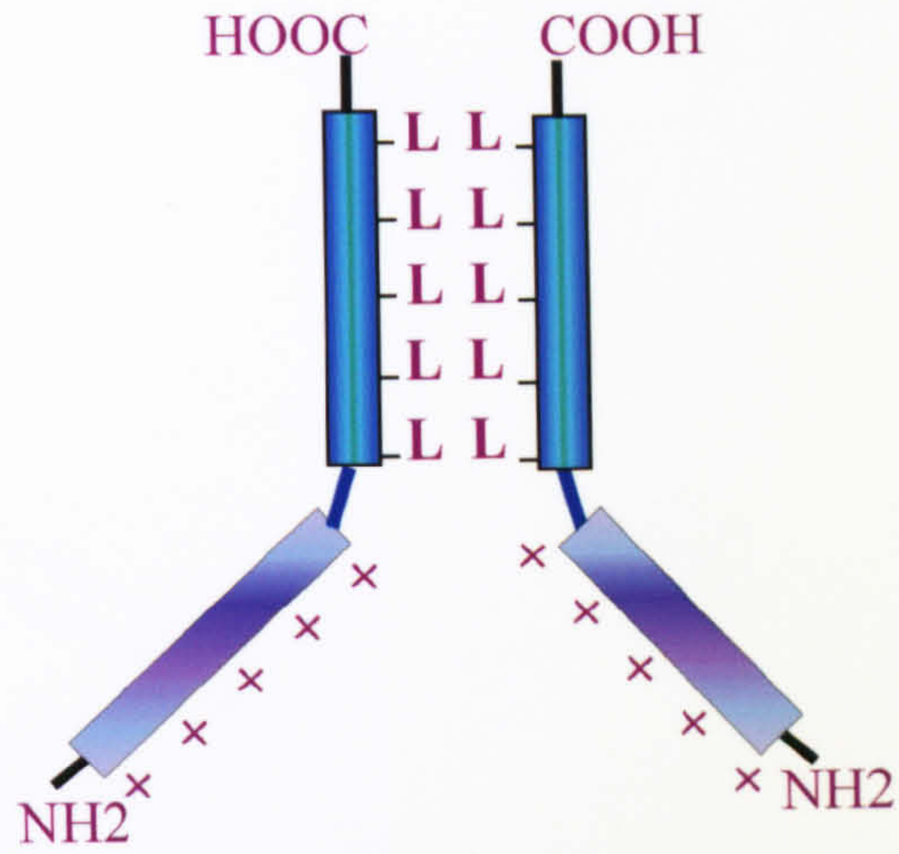
At present in the SWISSPROT database roughly 100 of the 10000 human protein sequences listed contain a potential leucine zipper motif, and of these about one third are described as "potential" or putative leucine zippers. Computer modelling studies have shown that it is difficult to predict the presence of an active leucine zipper from sequence data alone (Bornberg-Bauer *et al.*, 1998); it was pointed out in this particular study that of 3398 sequences predicted to contain leucine zippers or containing a leucine repeat frame, only 408 could be unequivocally identified as leucine zippers. Therefore it can be seen that a rapid functional test is needed for the study of leucine zippers in genome research.

6.1.3: THE YEAST TWO-HYBRID SYSTEM

It is apparent that many functions in the cell are performed by multi-component protein systems (Blundell and Srinivasan, 1996; Alberts, 1998). In addition, a number of complete genome sequences (for example, from *Saccharomyces cerevisiae, Drosophila melanogaster* and *Escherichia coli*) contain open reading frames that have not been completely characterised (Uetz *et al.*, 2000; Hodgkin, 2000). A complete understanding of functional genomics (also known as "the physiology of cells") involve identifying how gene products interact in carrying out their biochemical functions.

A potentially useful test for protein–protein interactions is the *S. cerevisiae* Two-Hybrid approach (Fields and Song, 1989; Phizicky and Fields, 1995; Bartel and Fields, 1997). This has been characterized and modified by several groups, and is available in various commercial forms (Uetz *et al.*, 2000; Gietz *et al.*, 1997). Cells provide the apparatus of a biochemical assay for the partners in a potential protein–protein interaction, sometimes known as "bait" and "prey". Since each cell in a library can contain a different pair of bait and prey, in theory a large number of pair

A:



B:

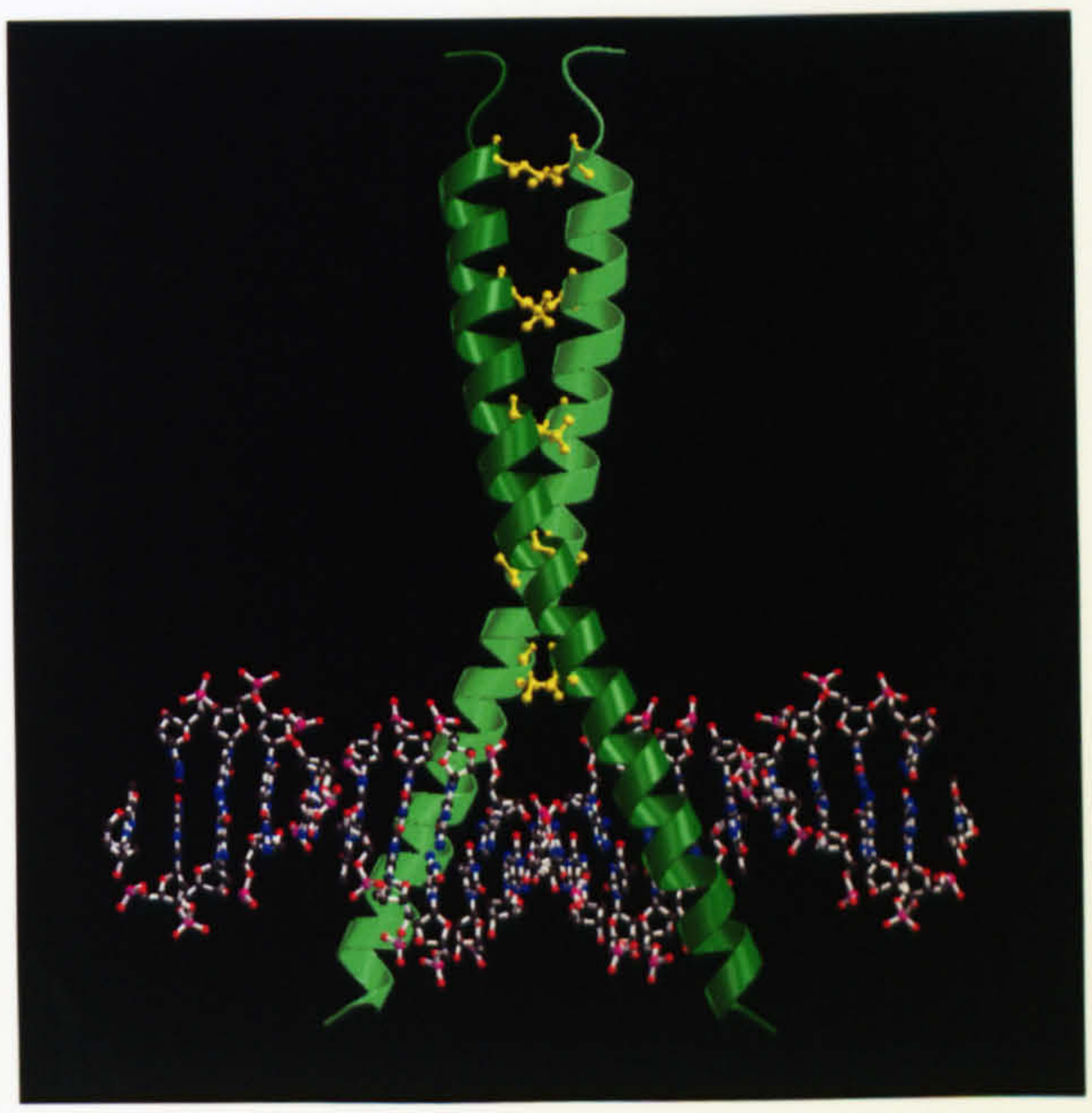


Figure 6.2:
(A) Structural model for leucine zipper.
(B) Crystal structure of leucine zipper complexed with DNA (PDB code 1dgc). DNA is shown as a "ball and stick" model and two helices of the interacting leucine zipper proteins are shown in green, with the leucines shown in yellow. The image was created by Dr. Matthew J. Conroy. The figure was produced using Molscript (Kraulis (1991) *J. Appl. Cryst.* 24 946-950) and Raster3D, (Merritt & Murphy (1994) *Acta Cryst.* D50 869-873).

wise interactions can be examined on comparatively few agar plates with appropriate selections or screens. Reviews are available that describe the use, interpretation, and pitfalls of two hybrid experiments in *S. cerevisiae*, and it has been predicted that this type of assay will provide a high throughput system that might be needed for automated screening for the all of the protein–protein interactions encoded by a genome (Hu *et al.*, 2000, Colas and Brent, 1998).

6.1.4: THE NEED FOR A BACTERIAL TWO HYBRID SYSTEM

In the past few years the genomes of *C. elegans, Escherichia coli and S. cerevisiae* have been sequenced (The *C. elegans* Sequencing Consortium, 1998; Blattner *et al.*, 1997; Hu *et al.*, 2000). The sequence of the human genome has also been completed recently and initial sequencing and analysis of the human genome has been released (International human genome sequencing consortium, 2001). The crucial challenge for biologists in the future will be to make sense of this information, much of which will encode protein sequences. To fully understand the function of proteins in living cells, a detailed knowledge of their structures and their role will be required. Despite the widespread use of the yeast two-hybrid system, there are certain limitations in handling yeast and there are reasons why similar genetic assays in *E. coli* would be useful. It is more convenient to use prokaryotes, which are scientifically "friendly" to molecular biologists and biotechnologists. A great advantage of *E. coli* is that the speed with which large numbers of interactions could be tested owing to relatively greater efficiency of transformation of bacteria over yeast. The development of a user-friendly system with increased speed of analysis represents an interesting challenge that has important commercial implications. *E. coli* grows faster than yeast and, with present methods, can be transformed with higher efficiency, allowing better coverage in library-based screens. In addition, since the cloning steps in yeast often involve passing libraries through *E. coli* hosts, using *E. coli* would remove a step from each cycle in a high throughput-screening programme. The smaller genome complexity and greater evolutionary distance from higher eukaryotes mean that *E. coli* systems should not generate as many false positives and negatives due to interactions between endogenous proteins and eukaryotic baits and prey as found in *S. cerevisiae*. With respect to drug screening, another potential advantage is that the cell envelope of *E. coli* seems to be somewhat more permeable to small molecules than the cell wall of yeast. Finally, yeast two hybrid systems require nuclear localization of the hybrid proteins and the presence of localization signals for other

cellular compartments could prevent bait or prey from reaching the nucleus, a problem that should not occur in *E. coli*.

6.1.5: AIMS

The main aim of this experiment is to develop a two-hybrid system for *E. coli* drawing on a knowledge of the structural features of C5-DNA Mtases, which can be used to positively identify leucine zipper encoding genes originating from genomic libraries, which may eventually provide the basis for a more general system for the analysis of protein-protein interactions.

6.2: The use of M.SPRI as the core of a bacterial two hybrid system

6.2.1: INTRODUCTION

It has been observed that individually inactive N- and C-terminal encoding parts of the M.*Bsp*RI gene can complement each other, resulting in specific, *in vivo* methylation of DNA (Posfai *et al.*, 1991). The DNA encoding the N- and C-terminal parts of this enzyme have been cloned in compatible plasmids which have then been used to co-transform *E. coli*. Both non-overlapping and partially overlapping fragments were found to be capable of complementation. If the detachment position occurred in motif VIII or C-terminal to it, the resultant enzyme showed no activity, but if the C-terminal part was longer than this, the resulting proteins were active or partially active (Posfai *et al.*, 1991).

A similar example was observed when a clinically isolated *Escherichia coli* strain HK31 was found to possess a gene encoding the *Eco*HK31I restriction modification system (Lee *et al.*, 1995). M. *Eco*HK31I consists of two polypeptides α and β with sizes 309 and 176 amino acids respectively. All the conserved motifs in C5-Mtases can be found in polypeptide α except motif IX that is present in polypeptide β (Lee *et al.*, 1995). This group reconstituted a functional M.*Eco*HK311 by fusing polypeptides α (all the elements in the structure of C5 Mtases apart from motif IX) and β (motif IX) together. Polypeptide α and β were separately synthesized in a T7-promoter controlled over-expression system and *in vitro* methylation occurred only when the two extracts were mixed. This experiment confirms that the two polypeptides are required for methylation (Lee *et al.*, 1995).

Another such variant type, M.*Aqu*I is also known to complement for its full methylase activity from two partly overlapping ORFs of 248 and 139 codons (Karreman and Waard, 1990).

It has also been established that it is possible to insert genetic material into the gene encoding the cytosine-specific DNA methyltransferase M.*Msp*I whilst still retaining activity of the gene product (Walter *et al.*, 1992; Matin, 2000). The tolerance of M.SPRI to polypeptide insertions has been evaluated and it has been found that the variable region of M.SPRI can accommodate an insert of varying lengths (Matin, 2000). The target recognition domain for *Msp*I, (*Msp*I TRD), from M.SPRI was replaced with a polylinker region. This new plasmid was named pSPR(X) and is refractory to *Hae*III restriction (Matin, 2000).

We have also observed during the creation of new restriction sites in the M.SPRI gene that when a Kanamycin cassette (approximately 1.25 kb in size) was inserted in the variable region, it changed the open reading frame (ORF) of the gene thereby splitting the encoded protein into two polypeptides. This might have made the Mtase gene inactive, but it did not (data not shown). The plasmid was not cleaved by *Hae*III because the two encoded polypeptides probably re-assembled to produce an active enzyme. Based on these experiments it was hypothesised that two separate polypeptides from the N- and C-terminal parts of M.SPRI could produce an active enzyme by complementation, if by some means they could be brought close to each other. The present work takes this approach a stage further in that it is shown that it is possible to "cut" the gene of the multispecific methyltransferase M.SPRI into two parts, to fuse each part to a DNA segment coding for a leucine zipper motif, and then, on expression of the genes, to produce a fully functional enzyme whose structural domains are "held together" by leucine zipper motifs. Therefore the M.SPRI enzyme can work as a hetero-dimer, that is to say, the two halves of the enzyme can be expressed separately and the two peptides will reconstitute to produce methylation activity *in vivo*.

6.2.2: CLONING OF TWO FRAGMENT OF THE M.SPRI GENE ENCODING THE N- AND C-TERMINAL DOMAINS INTO COMPATIBLE PLASMIDS

In order to develop a co-expression system, an experiment was designed to clone the N-terminal and C-terminal encoding fragments of M.SPRI into two compatible plasmids, and to express the recombinant proteins to test the above hypothesis. It was

convenient to use pSPR(X) (see above) for cloning the two regions into two compatible plasmids owing to the presence of a linker region in the central coding region of the gene containing several unique restriction sites.

In general, unrelated plasmids are compatible because their replication systems are different enough for them not to be able to co-exist in *E. coli*. A compatible partner plasmid in the co-expression system requires a different *ori* to prevent segregation and ultimate loss of one plasmid in culture (Johnston *et al.*, 2000 and references therein). In addition, the other requirements include a different antibiotic resistance gene for selection and most importantly, a variety of unique restriction sites that are shared with other plasmid for efficient cloning (Johnston *et al.*, 2000). We have chosen pACYC184 (Chang and Cohen, 1978) as second vector to be used for co-expression. The pSPR(X) / pACYC 184 co-expression vectors were found convenient as two vector bacterial co-expression system. The plasmid pSPR(X) derived from pGEXKG-2T (Guan and Dixon, 1991) originally carry ColE1 replicon and is fully compatible with those plasmids that carry p15A replicon such as plasmid pACYC 184 (MBI Fermentas catalogue, 2000-2001).

The cloning steps needed to make the system as explained above are as follows. All the diagrams pertaining to the sketches and gel pictures for the new constructs are placed at the end of the chapter.

6.2.2.1: Construction of pQIS 84 in which the "linker" is replaced by a sequence encoding a leucine zipper

The aim of this experiment was to insert a leucine zipper region in place of the "linker" in M.SPRI encoded by pSPR(X). The leucine zipper chosen for this experiment was from the human transcription factor CCAAT/Enhancer Binding Protein Alpha (CEBA_HUMAN), and has the amino acid sequence shown below (also shown in Appendix IV) (Landschulz *et al.*, 1988). For simplicity I will use numbering of leucine zipper sequence encoded amino acids starting from first leucine as number one and onward as shown with the sequence below.

A pair of complementary oligodeoxynucleotides, NIS 59 and NIS 60 were synthesised to create the appropriately coding DNA duplex. Two restriction sites were incorporated into the oligonucleotides to produce *Xho*I and *Eco*RI "ends" for subsequent cloning in pSPR(X) (shown in bold and red). The oligonucleotides were annealed, and the resulting double-stranded DNA duplex was cloned as a *Xho*I-

*Eco*RI fragment between the *Xho*I and *Eco*RI sites in the SPRI polylinker of pSPR(X). The cloning of the leucine zipper gene also introduced a unique *Nru*I site into the plasmid pSPR(X), and this restriction site was subsequently used to diagnose the success of the cloning experiment. The resulting plasmid was named pQIS 84 and coded for an active form of M.SPRI (Figure 6.3).

NIS 59/60: Leucine zipper C/EBP duplex oligonucleotide

```
                    XhoI
5'-TCGAGTTGAC TAGTGACAAT GACCGCCTGC GCAAACGCGT
   3'-CAACTG ATCACTGTTA CTGGCGGACG CGTTTGCGCA
       E  L   T  S  D  N   D  R  L   R  K  R
          1  2   3  4  5  6   7  8   9  10 11
                    NruI
GGAACAGCTG AGTCGCGAAC TGGACACGCT GCGGGGCATC
CCTTGTCGAC TCAGCGCTTG ACCTGTGCGA CGCCCCGTAG
V  E  Q  L   S  R  E   L  D  T   L  R  G  I
12 13 14 15  16 17 18  19 20 21  21 22 23 24
                    EcoRI
TTCCGCCAGC TGG-3'
AAGGCGGTCG ACCTTAA-5'
  F  R  Q  L
  25 26 27 28
```

NIS 59: Leucine zipper C/EBP top strand oligo

```
    XhoI
5'-TCGAGTTGACTAGTGACAATGACCGCCTGCGCAAACGCGTGGAA
          NruI
CAGCTGAGTCGCGAACTGGACACGCTGCGGGGCATCTTCCGCCAGCTGG-3'
```

NIS 60: Leucine zipper C/EBP bottom strand oligo

```
   EcoRI                                        NruI
5'-AATTCCAGCTGGCGGAAGATGCCCCGCAGCGTGTCCAGTTCGCGA

CTCAGCTGTTCCACGCGTTTGCGCAGGCGGTCATTGTCACTAGTCAAC-3'
```

6.2.2.2: Construction of pQIS 98 (pMETC-LZ) containing sequences encoding for leucine zipper followed by *Hae*III TRD and motif IX-X of M.SPRI

As was mentioned earlier, wild type M.SPRI has a structure that consists of 10 well-conserved motifs surrounding a variable region containing the *Eco*RII, *Msp*I and *Hae*III TRDs (Behrens *et al.*, 1987; Wilke *et al.*, 1988; Posfai *et al.*, 1989; Lauster *et*

al., 1989). The aim of this cloning experiment was to remove the nucleotide sequence encoding motifs I to VIII and the *Eco*RII TRD from the M.SPRI gene in pQIS 84.

The region coding for M.SPRI motifs I to VIII and the *Eco*RII TRD lies between bases 936-1820 in the plasmid pQIS 84, and these were removed from the pQIS 84 plasmid as a *Bam*HI-*Bgl*II fragment. (The *Bgl*II site is present in the "SPR polylinker" described above.) The remaining plasmid, which contained a gene fusion consisting of GST, the leucine zipper, the *Hae*III TRD and SPRI motifs IX and X, was re-circularised by ligation. The resulting plasmid was named pMETC-LZ (pQIS 98) (Figure 6.3).

6.2.2.3: Construction of three plasmids (encoding sequences for M.SPRI with leucine zipper, wild type M.SPRI and motif I-VIII of M.SPR with leucine zipper) in pACYC 184

The aim of these experiments was to construct three plasmids in pACYC 184 as follows. The first one comprised DNA sequence encoding M.SPRI containing an insert encoding a leucine zipper, in pACYC 184, equivalent to pQIS 84. The second construct contained the sequence of wild type M.SPRI in pACYC184. The third construct contained motifs I to VIII, the *Eco*RII TRD and leucine zipper region sequences, in cloning vector pACYC184.

The construction of these plasmids is described below. For this set of experiments the cloning vector pACYC184 was used (Gene bank accession number X06403, New England Biolabs (UK) Ltd., Hitchin, Herts). pACYC 184 contains a tetracycline resistance gene (bases 1581-2771) which contains within it two unique restriction sites, namely *Bam*HI (position 1870) and *Nru*I (position 2467). It has a p15A origin of replication and is a low copy number plasmid.

(A) Construction of pQIS 92 containing sequence of M.SPRI with leucine zipper insert sequence in pACYC 184

pACYC 184 was linearised by a *Bam*HI / *Nru*I double digest and the DNA fragment consisting of the central portion of the tetracycline resistance gene (599 bp fragment) was discarded. The 3646 bps fragment was recovered from the gel.

Plasmid pQIS 84 (described earlier) was then linearised with a *Bam*HI / *Ecl*136I double digest. *Ecl*136II recognises the site GAG↑CTC, and is an isoschizomer of *Sac*I (MBI Fermentas, Amherst NY). This enzyme also generates blunt-ended DNA, which can be joined with *Nru*I (also a blunt-end generating restriction enzyme). This produced a 1382bp (*Bam*HI / *Ecl*136I ended) fragment, which was then ligated into the *Bam*HI / *Nru*I linearised pACYC 184 plasmid, producing final plasmid pQIS 92 (containing the M.SPRI I-VIII/*Eco*RII TRD, leucine zipper region, *Msp*I TRD, motif IX-X of M.SPRI) in pACYC 184 (Figure 6.4). The construct pQIS 92 was confirmed by restriction analysis with *Nco*I and *Eco*RV. There is no site for *Nco*I in pQIS 84 and this unique site is present only in the chloramphenicol gene in pACYC 184. *Eco*RV generated fragments from pQIS 92 also match with the expected fragments. When tested for *in vivo* methylation potential by restriction with *Hae*III, it was revealed that the plasmid was partially protected. This can be partly explained by the fact that the Mtase gene is expressed from a plasmid with a p15A origin of replication and this has a low copy number, which results in reduced "gene dosage", and therefore partial methylation protection is observed. It was decided to check the Mcr response by transformation of *E. coli* strain DH5α. When used to transform *E. coli* strain DH5α, no colonies appeared, thus indicating that the Mtase is still active. Therefore, it was also decided to clone wild type M.SPRI sequence in pACYC 184 to see if the wild type has similar methylation potential while in a low copy number plasmid. This will also remove any doubts concerning the use of M.SPRI which carries an insert encoding a leucine zipper (in the TRD region) did not reduce the methylation activity.

(B) Construction of plasmid pQIS 99 encoding wild type M.SPRI sequence in pACYC 184

This experiment was designed to maintain wild type M.SPRI in pACYC 184. The construct pQIS 92 (see above) was cleaved with *Kpn*I and *Mun*I and the resulting vector (4182 bps) was ligated with the *Kpn*I and *Mun*I cleaved insert (846 bps) already cleaved from wild type M.SPRI (construct pQIS 72, see chapter four) and recovered from the gel (Figure 6.4). The final construct pQIS 99 now contains a complete M.SPRI sequence cloned in pACYC 184. However, when checked with *Msp*I and *Hae*III, it was found that the enzyme encoded by the plasmid was partially

active. It appears that the level of expression of the enzyme was not efficient to provide full protection.

(C) Construction of plasmid pQIS 100 encoding sequences of M.SPRI from motif I-VIII (N-Terminal fragment) and leucine zipper in pACYC 184

Using a *Bam*HI / *Stu*I double digest the genes coding for the SPRI motifs I-VIII, the *Eco*RII TRD region and the sequence encoding for leucine zipper (see above) were cut from plasmid pQIS 92 to give a 1009 bp DNA fragment. The pACYC 184 plasmid was again linearised with a *Bam*HI / *Nru*I double digest and the DNA fragment consisting of the central portion of the tetracycline resistance gene discarded (597 bp fragment). The 1009 bp *Bam*HI / *Stu*I ended fragment was then ligated into *Bam*HI / *Nru*I linearised pACYC 184 (3646 bps), producing plasmid pQIS 100. The construct was confirmed by restriction analysis with *Kpn*I. The plasmid pACYC 184 does not contain a *Kpn*I site but the recombinant plasmid pQIS 100 does. Another double digestion with *Bam*HI and *Xho*I also confirmed the presence of the correct insert (Figure 6.5).

6.2.2.4: Construction of pQIS 101 by the deletion of the *Eco*RI restriction site from pQIS 100

Since it was intended that experiments would be carried out in which the sequence encoding leucine zipper in pQIS 100 would be replaced by other sequences (such as mutant leucine zippers, see below), and since the leucine zipper coding sequence was cloned as an *Eco*RI-*Xho*I fragment, it was necessary to remove a second *Eco*RI site from within the chloramphenicol resistance gene in the plasmid. (The second *Eco*RI site is at position 1 in both the pACYC184 and pQIS 100 plasmids.) To do so, a pair of complementary oligodeoxynucleotides NIS 63 and NIS 64 (35bp) were synthesised which removed the *Eco*RI restriction site, using the Quick-change method (Stratagene).

*Eco*RI

Wild type sequence: 5′-CTTTCATTGCCATACG**GAATTC**CGGATGAGCATTC-3′

NIS 63: 5′-CTTTCATTGCCATACG**AAATTC**CGGATGAGCATTC-3′

NIS 64: 5′-GAATGCTCATCCG**GAATTT**CGTATGGCAATGAAAG-3′

This was a silent mutagenesis and the sequence of amino acid residues was not changed by this mutation. In fact, the chloramphenicol gene is in a complementary direction and the change in codon from **TTC** to **TTT** did not alter the phenylalanine residue as shown (see primer NIS 64 with mutation shown in red). The resulting plasmid was named pQIS 101. The digestion of the plasmid with *Eco*RI makes it linear. The construct was further confirmed by restriction digestion with various enzymes as shown in Figure 6.6.

6.2.2.5: Introduction of another ATG (start codon) after the *Bam*HI site in pMETN-LZ (pQIS 101)

Originally the M.SPRI gene had been cloned as a GST fusion. The N-terminal sequence of M.SPRI containing motifs I-VIII and the *Eco*RII TRD region were cloned into pACYC 184 as a *Bam*HI / *Stu*I fragment, without the GST open reading frame. During this cloning experiment the open reading frame of M.SPRI was lost and the nearest ATG start codon was at base pair position 1894 instead of base pair number 1870 in all constructs made in pACYC 184. The original and the new sequence of amino acids are aligned and shown below.

Original sequence in wild type M.SPR:	M	G	K	L	R	V	M	S	L	**F**	S	**G**	I	**G**	A
Present sequence after cloning in pACYC 184:							M	S	L	**F**	S	**G**	I	**G**	A

It was decided to introduce another start codon ATG at the correct position so that a complete sequence of M.SPRI at the N-terminal region would be translated. A forward primer NIS 77 was designed to amplify a segment of the M.SPRI gene with a start codon corresponding to the position of the first methionine. In addition to the *Bam*HI site another site for *Nco*I was introduced for diagnostic purposes. The new

start codon ATG is shown in red whereas the original ORF is shown blue. The reverse primer NIS 73 starts from the *Eco*RI site within the leucine zipper-coding region. The sequences of the primers are as shown below with the restriction sites shown in bold.

Sequence of oligonucleotides used for introduction of a start codon in constructs in pACYC 184

*Bam*HI/*Nco*I

NIS 77: 5′-GCGC**GGATCCATGG**GTAAACTACGTGTA**ATG**-3′

R G S **M** G K L R V **M**

*Eco*RI

NIS 73: 5′-GCGC**GAATTC**CAGCTGGCGGAAGATGC-3′

Using pQIS 101 as a template and primers NIS 77 and NIS 73 a PCR experiment was carried out (using *Pfu* polymerase) to generate a 999 bp fragment. The PCR product was recovered from the gel and restriction digested with *Bam*HI and *Eco*RI to create sticky ends. The vector DNA was also prepared by cleavage of pQIS 101 with *Bam*HI and *Eco*RI and recovered from the gel. Both the vector and insert DNA were ligated to produce a final construct pMETN-LZ (pQIS 101-N). The construct pMETN-LZ was confirmed by restriction analysis with *Nco*I (Figure 6.7).

6.2.3: *IN VIVO* COMPLEMENTATION OF N- AND C-TERMINAL FRAGMENTS OF M.SPRI

The two constructs pMETC-LZ and pMETN-LZ were used (in equimolar concentrations) to co-transform competent cells of the Mcr^+ strain of *E. coli* DH5α, and were plated on media supplemented with ampicillin and chloramphenicol. The same cells were also plated out on single antibiotic plates containing either ampicillin or chloramphenicol to determine the transformation efficiency of the cells for individual constructs. A control experiment was also carried out in parallel using a construct (p19nh) containing the C-terminal fragment of M.SPRI without any leucine zipper coding sequence attached at the end of the gene. In each experiment the competent DH5αcells gave a transformation efficiency of $\geq 10^7$ per μg of pLITMUS28 DNA.

In this experiment M.SPRI was expressed from two plasmids by expressing the gene encoding motifs I-VIII from one plasmid and that encoding motifs IX-X from

another compatible plasmid. The genes coding for either the N-terminal or for the C-terminal regions of M.SPRI were fused with the gene encoding the dimerisation domain of leucine zipper. This enabled both motifs, after their expression in *E. coli*, to drive dimerisation of the expressed polypeptide, thereby forming an active Mtase. Co-transformation of pMETC-LZ and pMETN-LZ resulted in the formation of no colonies on double antibiotic plates (see Table below and Figure 6.8). However, in a control experiment where one plasmid did not contain the leucine zipper region, there was no observable dimerisation and therefore many colonies were formed on the double antibiotic plate. A proposed model for the *in vivo* complementation of Mtase M.SPRI is shown in Figure 6.9.

	Bait (pMETN based plasmid)	**Prey (pMETC based plasmid)**	**Observation**
1	pQIS 100 (pMETN-LZ)	pQIS 98 (pMETC-LZ)	Active Mtase (No colonies on Amp+Cm plate)
2	pQIS 101N (pMETN-LZ, with corrected start codon)	pQIS 98 (pMETC-LZ)	Active Mtase (No colonies on Amp+Cm plate)
3	pQIS 101N (pMETN-LZ, with corrected start codon)	p19nh (pMETC, no leucine zipper)	Inactive Mtase (Many colonies on Amp+Cm plate, approx. 800~1000)

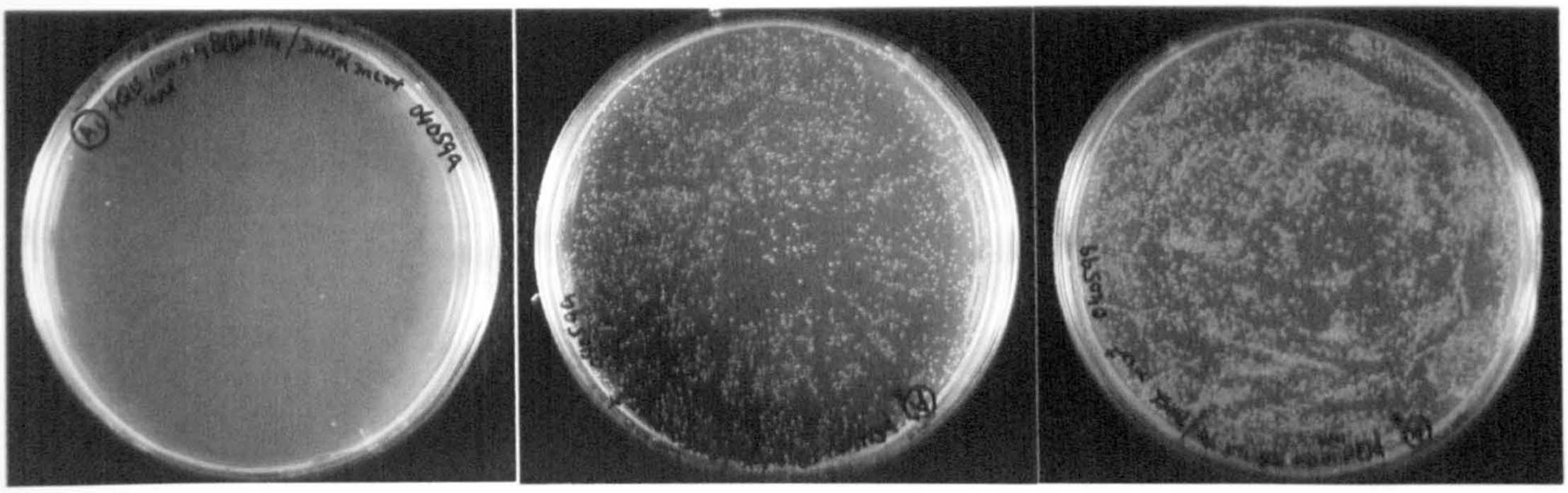
After the initial success of the above complementation experiment involving the N- and C-terminal fragments of M.SPRI *in vivo* I set out to design a robust prokaryotic two-hybrid system.

The first problem to consider is that the interaction of the two halves of the Mtase enzyme leads to the formation of no transformants and therefore the system should be considered as a reverse two-hybrid system. It was also necessary to establish whether the abolition of the leucine zipper interaction did competitively abolish M.SPRI activity.

Finally, experiments are described to address whether a selection of known interacting proteins can also bring the two peptides of Mtase M.SPRI close enough to make the enzyme active.

6.2.4: CONSTRUCTION OF PLASMIDS CONTAINING SEQUENCES FOR YEAST INTERACTING PROTEINS MLC1P AND IQGAP

It is known that budding yeast IQGAP-like protein, Cyk1/Iqg1p, has multiple roles in the assembly and contraction of the actomysin ring. Shannon and Li (2000) have



A1 A2 A3

Figure 6.8: Results from the *E. coli* assay after co-transformation of pMETC-LZ and pMETN-LZ. Agar media plates supplemented with (**A1**) Ampicillin and Chloramphenicol (**A2**) Chloramphenicol and (**A3**) Ampicillin. These experiments show the results of plasmid co-transformation into a *mcr*+ methylation-intolerant DH5α *E. coli* strain. A lack of bacterial colonies in plate A1 indicates that M.SPR-mediated methylation of the plasmid has occurred. The presence of colonies on plate A2 indicates that the cells harbour pMETN-LZ. The cells which acquired plasmid pMETC-LZ alone appear on plate A3.

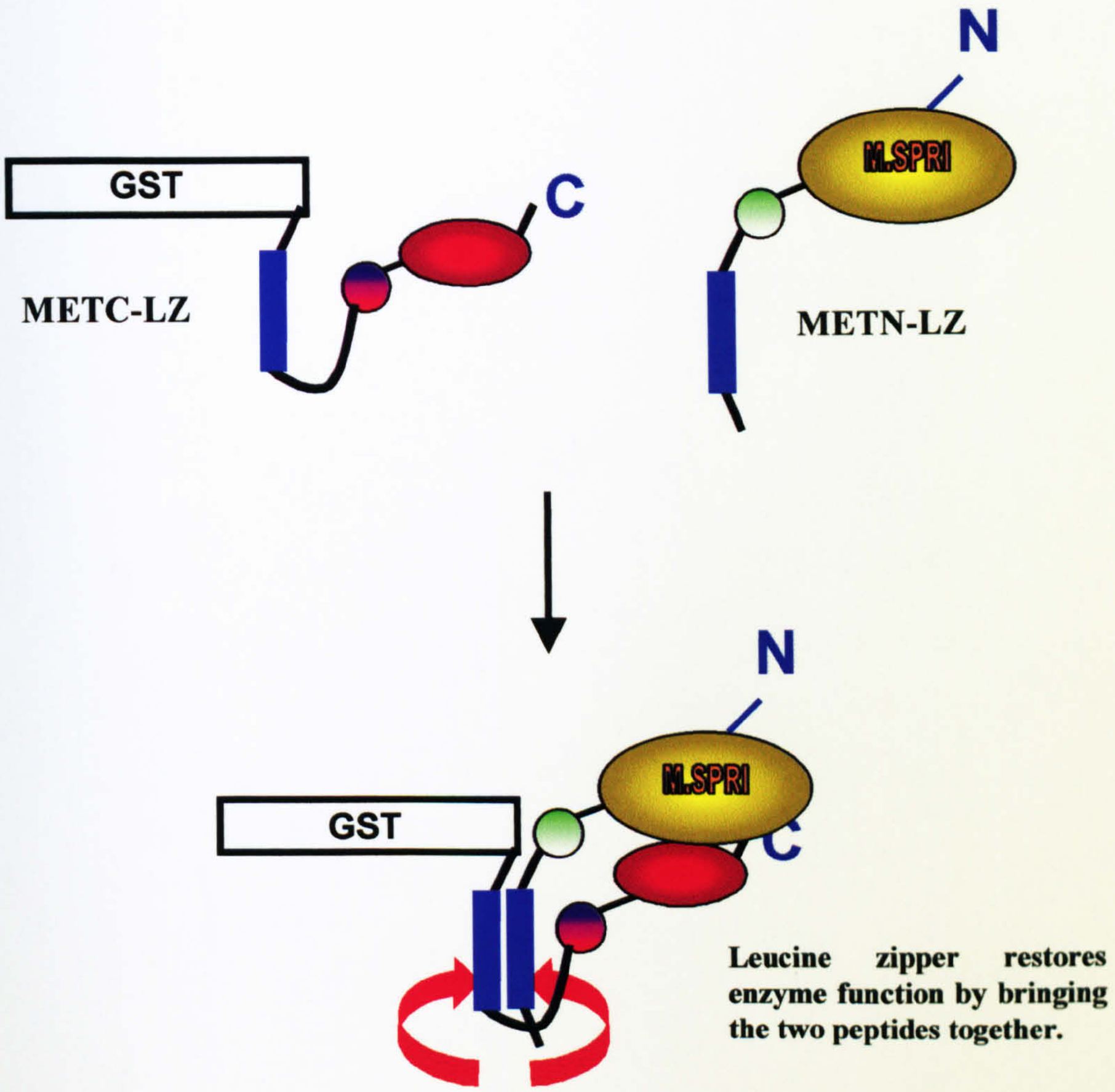


Figure 6.9: A proposed model of the leucine zipper mediated assembly of an active M.SPRI from METC-LZ and METN-LZ. When plasmid pMETN-LZ (containing sequence of block I-VIII of M.SPRI fused with Leucine zipper) and plasmid pMETC-LZ (containing sequence of Leucine zipper and block IX-X of SPR) are co-transformed, an active methyltransferase can be identified by an *mcr*+ host response, in the form of no colonies, on plates containing Ampicillin and Chloramphenicol (see figure 6.8).

shown that Mlc1p binds with the IQ motifs of Cyk1/Iqgp1 and have presented evidence that this interaction recruits Cyk1/Iqg1p to the bud neck. They demonstrated that Mlc1p is important for the assembly of the actomysin ring in budding yeast and that this function is mediated through interaction with Cyk1/Iqg1p. Boyne *et al*. (2000) have also reported that Mlc1 functions to recruit Iqg1p and in turn actin to the actomyosin ring and that it is also required for Myo1p function during ring contraction. An experiment was designed to test the interaction of these two proteins i.e. N-terminal fragment of Iqg1p and Mlc1 by using the Mtase two-hybrid system.

6.2.4.1: Construction of pQIS 103 containing the ORF of Mlc1 in pMETN-LZ by replacing the leucine zipper containing sequence

For the amplification of the Mlc1 gene sequence, two oligodeoxynucleotides QY 1 and QY 2 were synthesized. The forward primer QY 1 containing a *Xho*I site whereas the reverse primer QY 2 contained an *Eco*RI site (shown below) in order to facilitate the subsequent cloning of the gene in pMETN-LZ plasmid.

QY 1 and QY 2: Forward and reverse primer for Amplification of Mlc1 gene

*Xho*I

QY 1: 5′-GGCC**CTCGAG**ATGTCAGCCACCAGAGCCAATAAAG-3′

*Eco*RI

QY 2: 5′-GGCC**GAATTC**TCATTGTCTCAAAACATCTCGATGAACTTC-3′

The Mlc1 gene sequence, known to contain 450 bp was amplified from yeast genomic DNA (a kind gift from Dr. Clive Price). The amplified PCR product (450 bps) was cleaved with *Xho*I and *Eco*RI (to generate sticky ends) before ligation with pMETN-LZ that was precut (for excision of the leucine zipper coding sequence) and gel purified with the same enzymes. The final construct that contained Mlc1 gene was confirmed by restriction analysis. There is no *Stu*I site in the parent plasmid: but this site is present in the Mlc1gene. Following *Stu*I digestion a linear DNA molecule confirmed the presence of the inserted sequence. *Xho*I and *Eco*RI digestion of the construct pQIS 103 also resulted in excision of a 450 bp fragment, which confirmed insertion of the Mlc1 gene (Figure 6.10).

```
      ATGGCTGATCAACTGACAGAAGAGGAGATTGCAGAGTTCAAAGAGGCTTTTTCT
        M  A  D  Q  L  T  E  E  E  I  A  E  F  K  E  A  F  S

CTGTTTGACAAGGATGGTGATGGTACCATCACTACAAAGGAGTTGGGTACTGTGATGCGT
  L  F  D  K  D  G  D  G  T  I  T  T  K  E  L  G  T  V  M  R

TCTCTTGGTCAGAACCCGACAGAAGCAGAATTACAGGACATGATCAACGAAGTCGACGCT
  S  L  G  Q  N  P  T  E  A  E  L  Q  D  M  I  N  E  V  D  A

GATGGCAACGGCACAATTGACTTCCCAGAGTTTCTGACAATGATGGCGCGCAAAATGAAA
  D  G  N  G  T  I  D  F  P  E  F  L  T  M  M  A  R  K  M  K

GATACAGATAGCGAAGAAATTCGTGAAGCGTTCCGTGTGTTTGACAAGGATGGTAACGGT
  D  T  D  S  E  E  I  R  E  A  F  R  V  F  D  K  D  G  N  G

TACATTTCTGCTGCAGAACTTCGTCATGTGATGACAAACCTTGGTGAGAAGCTGACAGAT
  Y  I  S  A  A  E  L  R  H  V  M  T  N  L  G  E  K  L  T  D

GAAGAAGTTGATATGATTCGTGAAGCAGACATCGATGGTGATGGTCAAGTAAACTATGAA
  E  E  V  D  M  I  R  E  A  D  I  D  G  D  G  Q  V  N  Y  E

GAGTTTGTACAGATGATGACAGCGAAGTAA
  E  F  V  Q  M  M  T  A  K  STOP
```

Figure 6.12: DNA and amino acid sequence of human calmodulin.

6.2.4.2: Construction of pQIS 104 containing the ORF of Iqg1p in pMETC-LZ by replacing the sequence encoding a leucine zipper

The Iqg1p gene sequence was also amplified by PCR using two oligodeoxynucleotides QY 3 and QY 4. The forward primer QY 3 containing a *Xho*I site whereas the reverse primer QY 4 contained an *Eco*RI site (shown below) in order to facilitate the subsequent cloning of the gene in pMETC-LZ plasmid.

QY 3 and QY 4: Forward and reverse primer for amplification of Iqg1 gene

		*Xho*I
QY 3:	5′-GGCC**CTCGAG**GGATCCATGACAGCATATTCAGGCTCT-3′	
		*Eco*RI
QY 4:	5′-GGCC**GAATTC**GTCGACTTCTGGCTCTTTCAAATACAGC-3′	

The 2.5 kb N-terminal fragment of Iqg1p gene was amplified from yeast genomic DNA (Figure 6.11). However, when digested with *Xho*I and *Eco*RI to mobilise the PCR product, the presence of an internal *Eco*RI site was observed: two fragments of 1.5 and 1.0 kb approximately were produced. I therefore cloned the two fragments sequentially. In the first step the 1.5 kB *Xho*I/*Eco*RI fragment was cloned into pMETC-LZ by replacing the leucine zipper coding sequence and the product was identified by restriction analysis. The 1.0 kB *Eco*RI fragment was then inserted into pQIS 102 to produce pQIS 104 as shown in Figure 6.11. The correct orientation of 1 kb fragment was confirmed by restriction digestion by restriction digestion with *Bam*HI and *Sal*I: digestion with *Eco*RI and *Xho*I confirmed this conclusion.

6.2.4.3: *In vivo* complementation of N- and C-terminal fragments of M.SPRI

The two constructs pQIS 103 and pQIS 104 were used (in equimolar concentrations) to co-transform competent cells of *E. coli* DH5α. No complementation was observed: there were many colonies on double antibiotic plates.

6.2.5: CONSTRUCTION OF PLASMIDS ENCODING CALMODULIN AND CALMODULING BINDING PROTEINS

Calcium is a well-established second messenger (Corcoran and Means, 2001). The calcium receptor calmodulin (CaM) is involved in many changes through its effects

on a variety of CaM-binding proteins. Among these, the multifunctional Ca^{2+} / calmodulin-dependent kinases (CaMKs) are notable for their effects on components of transcription complexes, directly connecting Ca^{2+} with changes in gene expression (Corcoran and Means, 2001).

The crystal structure of calmodulin complexed with a calmodulin-binding peptide (a synthetic peptide based on the *Gallus gallus* (chicken) (**KMLS_CHICK**/P11799 in Swiss-Prot)) from Smooth Muscle Myosin Light Chain Kinase (smMLCK) is available (PDB code 1cdl)(Meador *et al.*, 1992). The calmodulin cDNA sequence and primary structure are shown in Figure 6.12. The structure of a calmodulin:peptide complex is presented in Figure 6.13.

The aim of this cloning experiment was to insert the human calmodulin gene (as bait) in place of the leucine zipper in pMETC-LZ, and a potential calmodulin-binding region (RS20) of chicken myosin light-chain kinase (MLCK) gene (as prey) in place of the leucine zipper in pMETN-LZ. Co-transformation was subsequently used to test whether the two proteins could promote assembly of the separate N- and C-termini of M.SPRI to create an active Mtase enzyme response.

6.2.5.1: Construction of pQIS 113 containing the ORF of calmodulin binding peptide in pMETN-LZ by replacing the leucine zipper ORF

A pair of complementary oligodeoxynucleotides, NIS 71 and NIS 72 (85 bps) were synthesised to create the calmodulin-binding peptide RS20 coding sequence. The sequence of the oligonucleotide with the amino acid sequence of the corresponding peptide is shown below. Four extra amino acids towards the N-terminus and two extra amino acids towards the C-terminus from the original sequence were added in order to facilitate the formation of the helix of calmodulin binding polypeptide by providing a short linker for fusion with Mtase N- or C-terminus fragments. The *Xho*I and *Eco*RI restriction sites for cloning into pMETN-LZ are shown in red. A restriction site for *Nco*I was added for diagnostic purposes without changing the amino acid codon and is shown in bold.

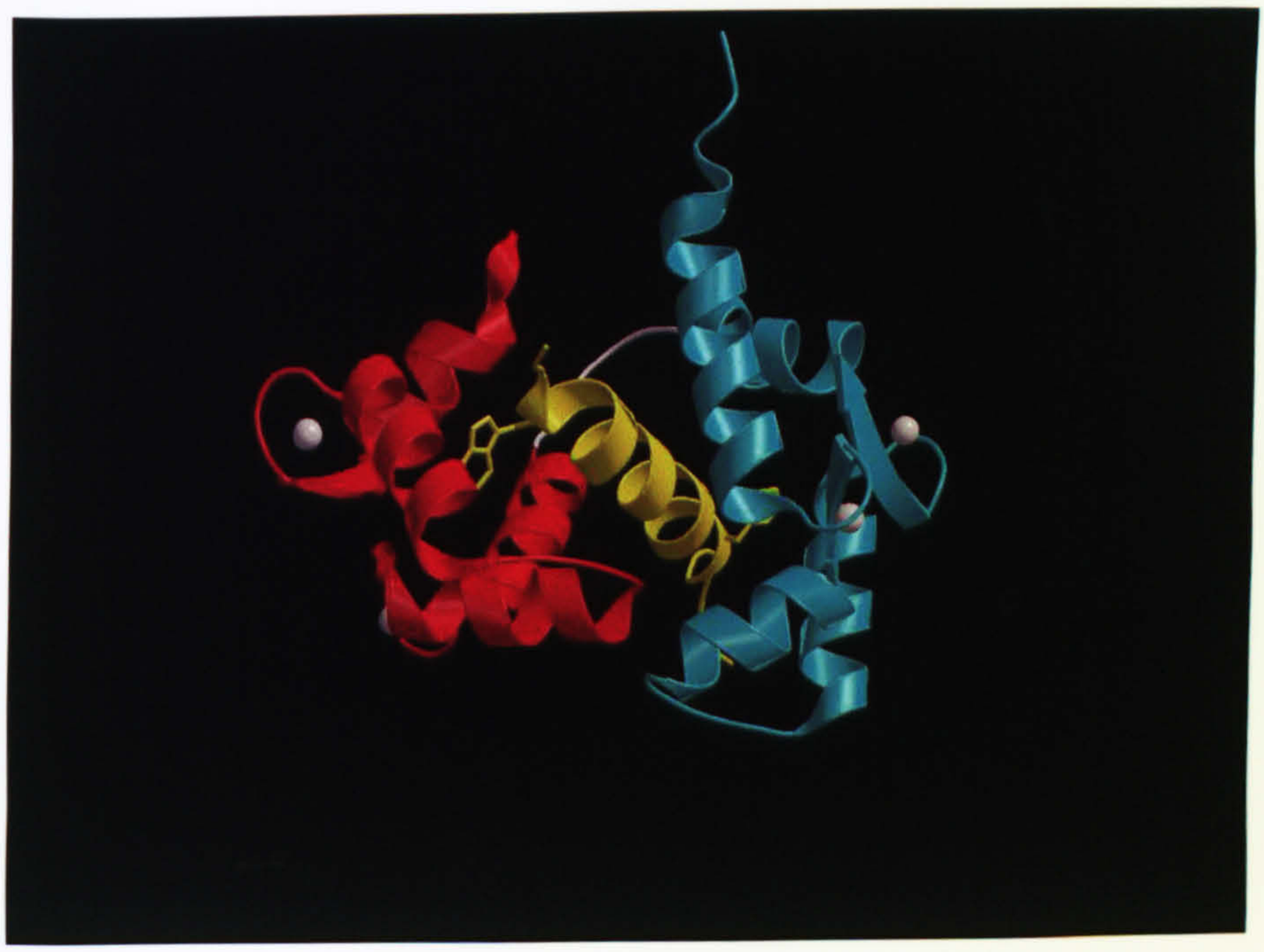


Figure 6.13: Calmodulin (CaM) complexed with a synthetic peptide from smooth muscle Myosin Light Chain Kinase (MLCK) (PDB code 1cdl). The CaM domain is shown in red and blue. Ca^{+2} ions are shown as white spheres whereas the smMLCK CaM binding peptide is shown in yellow. Hydrophobic residues of smMLCK which bind into the CaM binding pocket are shown in wire frame.

NIS 71 / NIS 72 duplex oligodeoxynucleotides:

```
     XhoI
5'-TCGAGAAATATATGGCCAGAAGAAAATGGCAGAAAACAGGCCATGCTGTCCGAGCA
   3'-CTTTATATACCGGTCTTCTTTTACCGTCTTTTGTCCGGTACGACAGGCTCGT
       E  K  Y  M  A  R  R  K  W  Q  K  T  G  H  A  V  R  A

                          NcoI          EcoRI
ATAGGAAGACTGTCATCCATGGCAATGG-3'
TATCCTTCTGACAGTAGGTACCGTTACCTTAA-5'
  I  G  R  L  S  S  M  A  M
```

NIS 71:

```
5'-TCGAGAAATATATGGCCAGAAGAAAATGGCAGAAAACAGGCCATGCTGTCCGAGCA
ATAGGAAGACTGTCATCCATGGCAATGG-3'
```

NIS 72:

```
5'-AATTCCATTGCCATGGATGACAGTCTTCCTATTGCTCGGACAGCATGGCCTGTTTTC
TGCCATTTTCTTCTGGCCATATATTTC-3'
```

The oliogonucleotides NIS 71 and NIS 72 were annealed to give an 85 bp duplex fragment with *Xho*I and *Eco*RI ends, and this was ligated into pMETN-LZ, which had already been cleaved with *Xho*I and *Eco*RI to allow the replacement of the leucine zipper open reading frame. The identity of pQIS 113 was confirmed by restriction digestion with *Nco*I. The duplex oligonucleotide contained two *Nco*I sites whereas the plasmid pMETN-LZ contained one *Nco*I site. The resulting fragments were of an expected size (Figure 6.14).

6.2.5.2: Construction of pQIS 114R containing sequence of calmodulin in pMETC-LZ by replacing sequence of leucine zipper

The Calmodulin gene was obtained from Prof. John Waltho (University of Sheffield). For the amplification of calmodulin gene, a pair of complementary oligodeoxynucleotides NIS 74 and NIS 76 were synthesized, which carry *Xho*I and *Eco*RI sites for subsequent cloning in pMETN-LZ or pMETC-LZ plasmids (shown below).

NIS 74 and NIS 76 Forward and reverse primers for amplification of calmodulin

*Xho*I *Bcl*I

NIS 74: 5′-GCGC**CTCGAG**ATGGC**TGATCA**ACTGACAGAAGAG-3′

*Eco*RI

NIS 76: 5′-GCGC**GAATTC**CTTCGCTGTCATCATCTGTACAAAC-3′

NIS 74 and NIS 76 (34 and 35bp) were used to amplify the gene from a plasmid containing a calmodulin gene. The PCR product (Figure 6.15) was purified and cleaved with *Xho*I and *Eco*RI to generate sticky ends. This gave a 447 bp fragment, which was then ligated into pMETC-LZ (Figure 6.15). The final product pQIS 114R now carried an insert of approximately 450 bps in size and this was confirmed by restriction digestion with *Xho*I and *Eco*RI (Figure 6.15).

6.2.5.3: *In vivo* Complementation of the N- and C-terminal fragments of M.SPRI fused to calmodulin and a calmodulin recognising peptide

The next step after construction of the two plasmids that carried (a) calmodulin and (b) calmodulin binding peptide fused to the N- and C-terminus fragments of M.SPRI respectively, was to express them *in vivo* in *E. coli* DH5α. Therefore the two plasmids were used in equimolar amounts to co-transform *E. coli* DH5α. However, on an antibiotic plate containing ampicillin and chloramphenicol growth of colonies were observed. This indicated that *in vivo* complementation of M.SPRI had not taken place. The observations recorded were as follows.

S. No	Bait (pMETN based plasmid)	Prey (pMETC based plasmid)	Observation
1	pQIS 113 (CaM peptide)	pQIS 114R (CaM)	Inactive Mtase (Many colonies on Amp+Cm plate)
2	pQIS 113 (CaM peptide)	pQIS 98 (pMETC-LZ)	Inactive Mtase (Many colonies on Amp+Cm plate)

A negative control experiment was performed in parallel using pQIS 98 (carrying sequences encoding a leucine zipper and motif IX-X of M.SPRI) and pQIS 113 (carrying sequences encoding motif I-VIII of M.SPRI and a calmodulin binding

peptide). The length of the sequence encoding leucine zipper (93 bps) is almost equivalent as that of calmodulin peptide (85 bps). The idea was to examine the two fusion protein having similar molecular weight but without any known interactions as negative control. The result of this co-transformation was as expected: complementation was not observed and consequently many colonies were obtained on double antibiotic plates.

In order to establish whether Ca^{+2} ions are required for binding of peptide with calmodulin, another experiment was carried out by growing colonies picked from a double antibiotic plate in a 5 ml culture containing ampicillin and chloramphenicol in the presence of varying amount of $CaCl_2$. Four colonies picked from double antibiotic plates were transferred into four tubes containing 5 ml LB, ampicillin, chloramphenicol and $CaCl_2$ with 1 mM, 5 mM, 10 mM and 0.0 mM in each tube. The cells were grown for 6-8 hours to observe any difference in growth. There was no significant difference in growth in all four cultures. Miniprep DNA prepared from the culture was cleaved with *Xho*I to linearize the plasmid before analysis on agarose gel, and this showed that both plasmids were present in the culture. In another experiment, co-transformation was carried out and the cells were plated onto agar media plates containing 1 mM $CaCl_2$. The appearance of colonies on a double antibiotic plate confirmed that the proteins encoded by the two plasmids did not complement *in vivo* and the presence of calcium had no effect.

6.2.6: ANALYSIS OF THE ASSEMBLY OF M.SPRI DRIVEN BY MUTANT LEUCINE ZIPPERS

In order to investigate the role of long amino acid side chains on the ability of leucine zipper peptides to interact, two experiments were carried out.

1. The conserved Asparagine at position 5 was substituted by Lysine
2. The conserved Leucine at position 8 was substituted by Proline

6.2.6.1: Construction of pQIS 115 containing a mutant (Asn 5 Lys) leucine zipper motif in pMETN-LZ

Complementary oligodeoxynucleotides, NIS 80 and NIS 81 were used to introduce a site-directed change (Asn 5 Lys) into the leucine zipper ORF in pMETN-LZ using the Quick-Change method (Stratagene). One base pair change from AAT to AA**A** changed the Asn codon to that specifying Lys. A new *Cla*I restriction site was also

introduced in order to facilitate screening for the mutant. The presence of this *Cla*I site does not alter the final amino acid sequence. The sequences of oligonucleotides are shown below, with the new restriction site shown in boldface and the mutation in red.

NIS 80 / NIS 81 duplex oligonucleotide:

*Cla*I
5′-GTTGACTAGTGACAA**AGATCGAT**TGCGCAAACGCGTGG-3′
3′-CAACTGATCACTGTTTC**TAGCTA**ACGCGTTTGCGCACC-5′
L T S D **K** D R L R K R V
*Cla*I
NIS 80: 5′-GTTGACTAGTGACAAAG**ATCGAT**TGCGCAAACGCGTGG-3′

*Cla*I
NIS 81: 5′-CCACGCGTTTGCGCA**ATCGAT**CTTTGTCACTAGTCAAC-3′

The resulting construct was named pQIS 115 and encoded the N-terminus of M.SPRI fused to the Asn 5 Lys mutant leucine zipper. Colonies were screened for plasmids containing an additional *Cla*I site (Figure 6.16).

6.2.6.2: Construction of pQIS 116 encoding the Leu 8 Pro mutant leucine zipper

Complementary oligodeoxynucleotides NIS 82 and NIS 83 (36 bp) were used to introduce site-directed changes into pMETN-LZ using the Quick-Change method (Stratagene). The sequence of primers is shown below with a single base mutation from CTG to C**C**G (to change leucine into proline) and another silent mutation (CGC to CG**G**) to generate the *Sac*II restriction site.

NIS 82 / NIS 83 duplex oligonucleotide:

*Sac*II
5′-AGTGACAATGACCGC**CCGCGG**AAACGCGTGGAACAG-3′
3′-TCACTGTTACTGGCG**GGCGCC**TTTGCGCACCTTGTC-5′
S D N D R **P** R K R V E Q
*Sac*II
NIS 82: 5′-AGTGACAATGACCGC**CCGCGG**AAACGCGTGGAACAG-3′

*Sac*II
NIS 83: 5′-CTGTTCCACGCGTTT**CCGCGG**GCGGTCATTGTCACT-3′

The resulting plasmids were screened by restriction digestion. The oligonucleotides used for mutagenesis for Leu 8 Pro contained a *Sac*II site to be used for diagnostic purposes. pMETN-LZ already contained one *Sac*II site whereas the final construct (pQIS 116) now contained two *Sac*II sites. The cleavage pattern obtained by a restriction digest analysis of the new construct pQIS 116 confirmed the presence of mutation (Figure 6.17).

6.2.6.3: Construction of pQIS 117 containing sequence encoding the Leu 8 Pro leucine zipper mutant

The mutagenic complementary oligodeoxynucleotides NIS 82 and NIS 83 (36 bp) synthesized earlier (see above) were used to carry out the site-directed change of leucine to proline in pMETC-LZ using the Quick-Change method (Stratagene). The new construct pQIS 117 also attained a unique *Sac*II site that was not present in its parent plasmid pMETC-LZ (Figure 6.18).

6.2.6.4: *In vivo* Complementation of N- and C-terminal fragments of M.SPRI using a mutated leucine zipper motif

The next experiment after construction of the three mutant plasmids pQIS 115 (pMETN-LZ Asn 5 Lys), pQIS 116 (pMETN-LZ-Leu 8 Pro), and pQIS 117 (pMETC-LZ Leu 8 Pro) was one to test *in vivo* complementation. Co-transformation was carried out in different combinations to measure the Mcr response. The results of these experiments are shown in the table below.

S. No	Bait (pMETN based plasmid)	Prey (pMETC based plasmid)	Observation
1	pQIS 115 (Asn 5 Lys mutant)	pQIS 98 (pMETC-LZ)	Active Mtase (No colonies on Amp+Cm plate)
2	pQIS 116 (Leu 8 Pro mutant)	pQIS 98 (pMETC-LZ)	Active Mtase (No colonies on Amp+Cm plate)
3	pQIS 101N (pMETN-LZ)	pQIS 117 (Leu 8 Pro mutant)	Active Mtase (No colonies on Amp+Cm plate)
4	pQIS 116 (Leu 8 Pro mutant)	pQIS 117 (Leu 8 Pro mutant)	Active Mtase (No colonies on Amp+Cm plate)

The set of results from the co-transformation experiments (see table above) demonstrated that the complementation of N- and C-terminal fragments *in vivo* led to the formation of an active methyltransferase. As a result there were no colonies on double antibiotic plates. When the results from another two co-transformations experiments were obtained it was confirmed that the Leu 8 Pro mutant had not lost the ability to promote assembly of an active Mtase. A fourth co-transformation experiment also gave the same result.

Mutational analysis of leucine zippers have shown that alteration of one of the leucine residues in the zipper motif is generally tolerated except when it is replaced by a proline, which introduces a kink into the α-helix, or by a charged amino acid, which disrupts the hydrophobic spine (Kouzarides and Ziff, 1988; Gentz *et al.*, 1989). It has also been reported that the simultaneous substitution of leucines at either the amino- or the carboxy terminus of both the Fos and Jun leucine zippers has a less deleterious effect than the substitution of leucines at the amino terminus of one and the carboxy terminus of the other zipper (Gentz *et al.*, 1989). According to the results of Schuermann *et al.* (1989) it appears that all of the five leucines participate in the interaction with Jun, but that the different leucine residues have different functional importance. We have mutated the second leucine in the series (from N-terminal) and it could be argued that the resulting successful complementation is due to the compensation of the mutation by the adjacent leucine residues. It is also probably true that the mutation did not prevent the two N- and C- terminal fragments interacting to make an active Mtase. In other words the sensitivity of the experiment is probably good enough to be used for other experiments. However, it was necessary to measure the level of association of the leucine zipper motifs by measuring the methylation potential of the resulting Mtase.

We decided to carry out *in vivo* methylation assays by growing two plasmids in *mcr*$^-$ cells and then digestion with *Hae*III followed by agarose gel analysis and analysis by Wave^{TM}. This has been described in section 6.3.

6.2.7: CONSTRUCTION OF PLASMIDS CONTAINING SEQUENCES FOR WBP1 AND OST1

It was decided to carry out another experiment using a pair of known interacting yeast protein (WBP1 and OST1). Yeast genomic DNA was obtained from Dr. Clive

Price (University of Sheffield). The WBP1 and OST1 sequences were amplified from genomic DNA using PCR.

6.2.7.1: Construction of pQIS 118 containing sequence encoding for WBP1

For the amplification of the WBP1 gene, two oligodeoxynucleotides NIS 86 and NIS 87 were synthesized. The forward primer NIS 86 containing a *Xho*I site whereas the reverse primer NIS 87 contained a *Mfe*I site (shown below) in order to allow the subsequent cloning of the gene in pMETC-LZ plasmids. *Mfe*I is an isoschizomer of *Mun*I and the site generated by this enzyme is compatible for ligation with DNA cut with *Eco*RI. This was essential as the sequence of WBP1 carried an internal *Eco*RI site. The WBP1 sequence is estimated to be 1293 bps. The sequence of amino acids of WBP1 is shown in Appendix IV.

NIS 86 and NIS 87 Forward and reverse primers for the amplification of WBP1

	*Xho*I
NIS 86:	5′-GCGC**CTCGAG**ATGCGGACCGATTGGAATTT-3′

	*Mfe*I
NIS 87:	5′-GCGC**CAATTG**GTTTGTTTTTTTTGAATGTTT-3′

Pfu polymerase was used for the amplification of DNA sequence from yeast genomic DNA and the size of WBP1 PCR product was checked by running an aliquot on an agarose gel. The PCR product was purified by passing through a Mobispin column and was then subjected to restriction analysis using *Xho*I and *Mun*I. The resulting DNA was heated at 65°C for 20 minutes before ligation to the vector pMETC-LZ that had been precut with *Xho*I and *Eco*RI. The colonies appearing as a result of transformation to *E. coli* were screened for the correct insertion of the 1293 bp DNA fragment with various restriction enzymes such as *Xho*I/*Eco*RI, *Bam*HI, *Xho*I/*Ecl*136 and *Xba*I and shown in Figure 6.19.

*Xho*I/*Eco*RI should generate a 1032 bp fragment because WBP1 has been cloned as a *Xho*I/*Mun*I fragment and therefore the original *Eco*RI site was lost. The resulting fragment is shorter because of the presence of the internal *Eco*RI site. A restriction digestion analysis also confirmed the presence of the correct insert (Figure 6.19).

6.2.7.2: Construction of pQIS 119 containing sequence encoding for OST1

OST1/YJL002C gene sequence was amplified by using two oligodeoxynucleotides NIS 84 and NIS 85 (shown below). The forward primer NIS 84 contained *Xho*I site and reverse primer contained an *Eco*RI sites in order to allow subsequent cloning in pMETN-LZ. The sequence of OST1 was estimated to be 1431 bps. The sequence of amino acids for OST1 is shown in Appendix IV.

NIS 84 and NIS 85 Forward and reverse primers for the amplification of OST1

*Xho*I

NIS 84: 5′-GCGC**CTCGAG**ATGAGGCAGGTTTGGTTCTC-3′

*Eco*RI

NIS 85: 5′-GCGC**GAATTC**TCAGTTAGTTACGTTCATGT-3′

Pfu polymerase was used for the amplification of DNA sequence from yeast genomic DNA and the size of the OST1 PCR product was checked by running an aliquot on an agarose gel. The PCR product was purified by passing through a Mobispin column and was then cut with *Xho*I and *Mun*I. The resulting DNA was heated at 65°C for 20 minutes before ligation to the vector pMETC-LZ that had been precut with *Xho*I. The colonies appearing as a result of transformation were screened for the correct insertion of 1431 bp DNA fragment. Restriction analysis with *Mun*I, *Kpn*I/*Mun*I, *Bam*HI/*Eco*RI, *Xho*I/*Eco*RI confirmed the presence of correct insert (Figure 6.20).

6.2.7.3: *In vivo* Complementation of N- and C-terminal fragments of M.SPRI with WBP1 and OST1

After construction of these two plasmids they were tested for complementation as before.

S. No	Bait (pMETN based plasmid)	Prey (pMETC based plasmid)	Observation
1	pQIS 119 (OST1)	pQIS 118 (WBP1)	Inactive Mtase (Many colonies on Amp+Cm plate)

The summary of all constructs made so far and the results of co-transformation are presented in Table 6.1.

Table 6.1 Summary of transformation experiments based on *mcr*$^+$ response (for the complementation of N- and C-terminal fragments in bacterial two-hybrid system). The transformation efficiency of E. coli DH5α mcr$^+$ was always ≈10^7 or >10^7.

	Bait (pMETN based plasmid)	Prey (pMETC based plasmid)	Observation
1	pQIS 101N (pMETN-LZ)	pQIS 98 (pMETC-LZ)	Active Mtase (No colonies on Amp + Cm plate)
2	pQIS 101N (pMETN-LZ)	p19nh (Motif IX-X without LZ)	Inactive Mtase (Many colonies on Amp + Cm plate)
3	pQIS 113 (CaM peptide)	pQIS 98 (pMETC-LZ)	Inactive Mtase (Many colonies on Amp + Cm plate)
4	pQIS 104 (yeast gene)	pQIS 103 (yeast gene)	Inactive Mtase (Many colonies on Amp + Cm plate)
5	pQIS 113 (CaM peptide)	pQIS 114R (CaM)	Inactive Mtase (Many colonies on Amp + Cm plate)
6	pQIS 115 (Asn 5 Lys mutant)	pQIS 98 (pMETC-LZ)	Active Mtase (No colonies on Amp + Cm plate)
7	pQIS 116 (Leu 8 Pro mutant)	pQIS 117 (Leu 8 Pro mutant)	Active Mtase (No colonies on Amp + Cm plate)
8	pQIS 119 (OST1)	pQIS 118 (WBP1)	Inactive Mtase (Many colonics on Amp + Cm plate)

6.3: Comparison of Methylation potential by three different methods

So far the bacterial two-hybrid hypothesis has been tested by co-transformation and sequential transformation in parallel and the observations were recorded in the form of presence or absence of bacterial colonies. Sequential transformation was carried out for each combination as shown in the summary Table 6.1. In sequential transformations the bacteria carrying one plasmid were made competent to accept another plasmid. In this study, to keep the experiment procedures constant, we have always first transformed bacteria with a low copy number plasmid i.e. originating from pACYC 184 (pMETN-series plasmids). Then the second plasmid was used to transform bacteria and the results were recorded as active or inactive due to the Mcr response of *Escherichia coli* DH5α. However, to find out more regarding the level of interactions in leucine zippers or interacting proteins, it was decided to carry out *in vivo* methylation protection assays (as described in Chapter Two). *Escherichia coli* DH5α *mcr*$^{-}$ harbouring plasmids for analysis were grown on plates supplemented with two antibiotics i.e. ampicillin and chloramphenicol. The DNA was extracted, purified and the identity of both plasmids was confirmed by restriction analysis with *Eco*RI or *Xho*I, after which the DNA was cut with *Hae*III in order to analyse its methylation status. An aliquot of reaction mixture (10 µl) for each restriction digestion was analysed on agarose gels (Figure 6.21). The remaining 10 µl of the reaction mixture was used for the analysis by DHPLC.

The result of the restriction protection assay for all the two-hybrid plasmids, analysed on an agarose gel are shown in Figure 6.21. The DHPLC analysis after restriction with *Hae*III for all two-hybrid plasmids was also analysed and the resulting chromatograms are shown in Figure 6.22. The summary for the two-hybrid experiment compared by three different means is presented in Table 6.2. A visual rating of (+ and -) (active and inactive respectively) was provisionally assigned to these results from visual interpretation of the gel and DHPLC traces.

The data obtained from the DHPLC analysis were subsequently manipulated further to extract more information. The traces obtained by DHPLC were overlaid in order to identify uncut DNA (retention time, 15 minutes and later) (Figure 6.23). A visible fraction of uncut DNA persists in the case of the plasmids encoding active Mtases (see chromatograms for A, C and E). In addition there is some uncut DNA found in the chromatogram F (see red profile). It could be due to the presence of larger

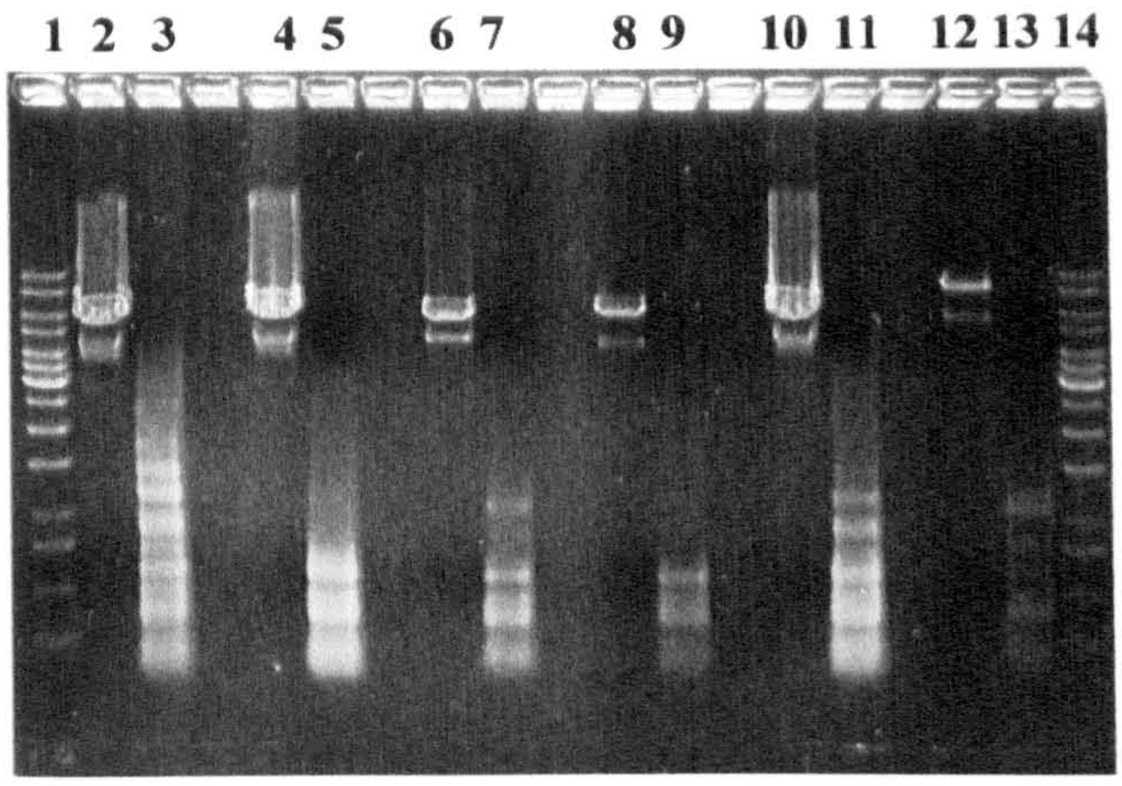


Figure 6.21: Two hybrid plasmid restriction protection assay: Plasmids DNA after recovery from *E. coli* DH5α *mcr*$^{-}$ were restricted with *Xho*I to confirm the presence of both Bait and Prey plasmids, and with *Hae*III to establish the digestion profile. The lane order (in 1% agarose gel) is as follows;

1. GeneRuler^{TM} 1 kb DNA ladder.
2. pQIS 101N (W/T LZ) + pQIS 98 (W/T LZ) cut with *Xho*I
3. pQIS 101N (W/T LZ) + pQIS 98 (W/T LZ) cut with *Hae*III

4. pQIS 113 (CaM pep) + pQIS 98 (W/T LZ) cut with *Xho*I
5. pQIS 113 (CaM pep) + pQIS 98 (W/T LZ) cut with *Hae*III

6. pQIS 115 (Asn 5 Lys) + pQIS 98 (W/T LZ) cut with *Xho*I
7. pQIS 115 (Asn 5 Lys) + pQIS 98 (W/T LZ) cut with *Hae*III

8. pQIS 113 (CaM Pep) + pQIS 114 (CaM) cut with *Xho*I
9. pQIS 113 (CaM Pep) + pQIS 114 (CaM) cut with *Hae*III

10. pQIS 116 (Leu 8 Pro) + pQIS 117 (Leu 8 Pro) cut with *Xho*I
11. pQIS 116 (Leu 8 Pro) + pQIS 117 (Leu 8 Pro) cut with *Hae*III

12. pQIS 119 (OST1) + pQIS 118 (WBP1) cut with *Xho*I
13. pQIS 119 (OST1) + pQIS 118 (WBP1) cut with *Hae*III
14. GeneRuler^{TM} 1 kb DNA ladder.

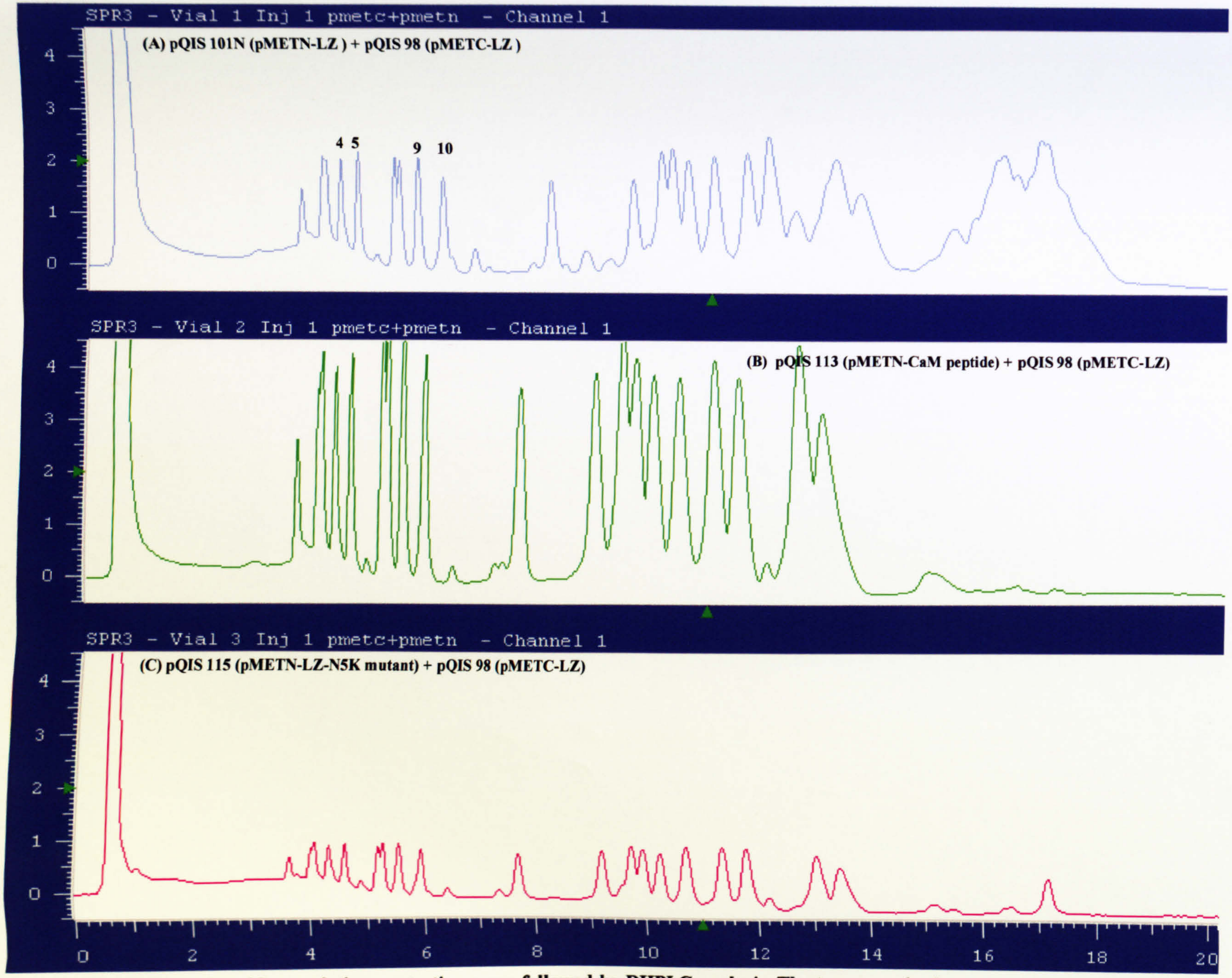


Figure 6.22: Two-hybrid plasmids restriction protection assay followed by DHPLC analysis. The traces are in the same order as shown in table 6.2 and in the gel in figure 6.21. The peaks labelled with numbers (4, 5, 9 and 10) were used to evaluated by comparing the ratios of the areas of similar peaks in each graph.

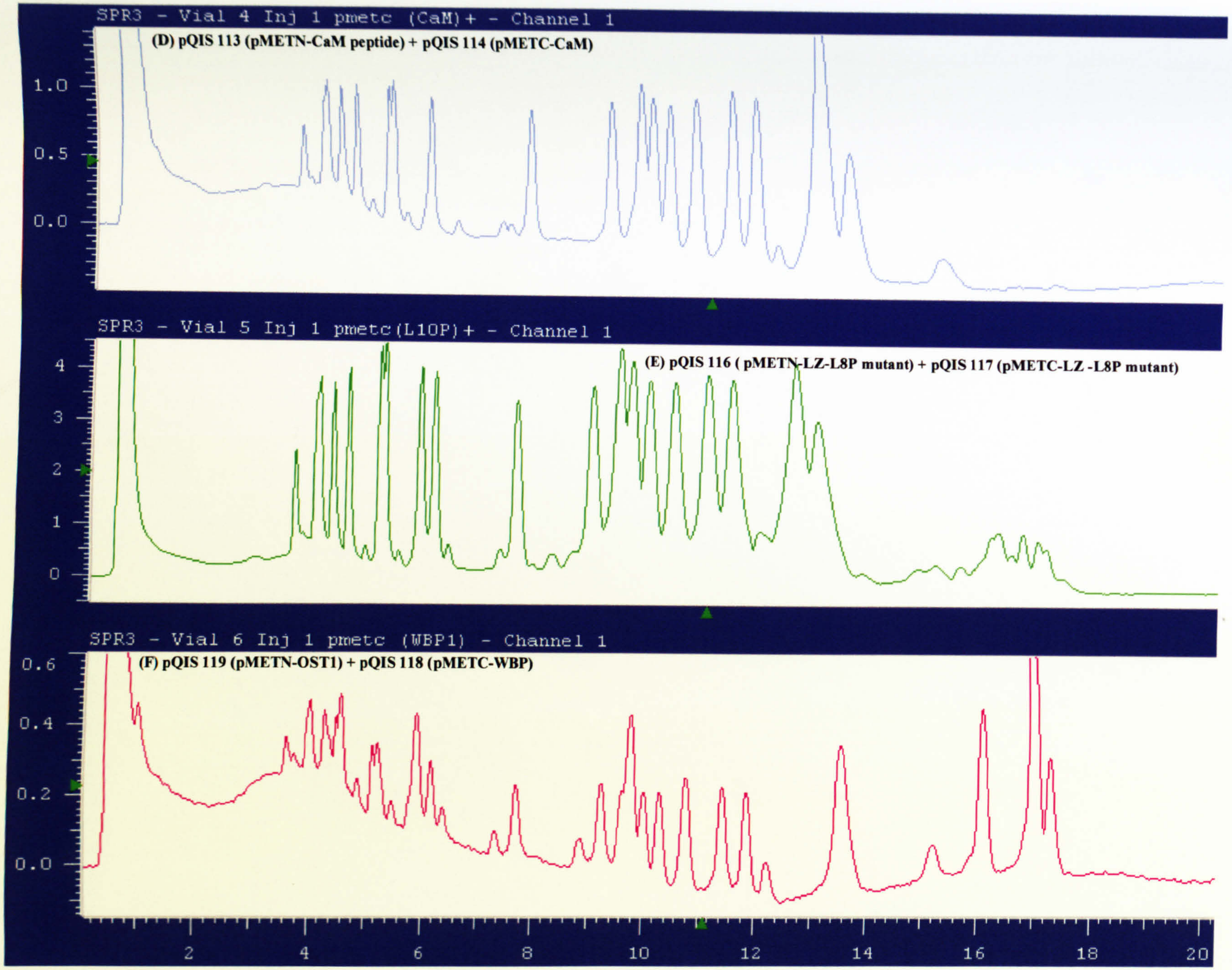


Figure 6.22: Two-hybrid plasmids restriction protection assay followed by DHPLC analysis. The traces are in the same order as shown in table 6.2 and in the gel in figure 6.21.

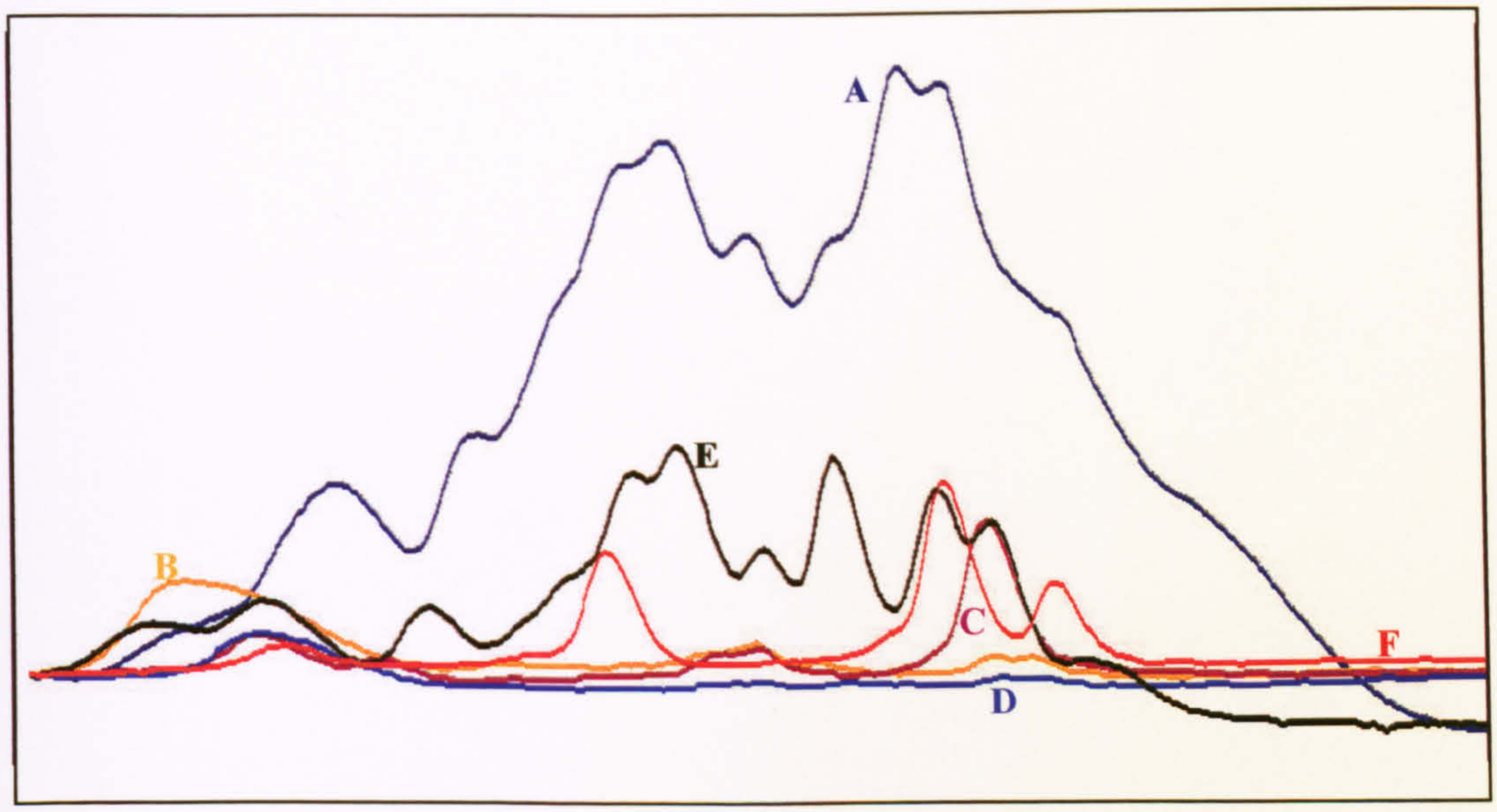


Figure 6.23: The overlaid chromatogram of two-hybrid plasmid restriction protection assay analysed by DHPLC shown after 15 minutes of retention time. Plasmids DNA after recovery from *E. coli* DH5α *mcr*$^{-}$ were restricted with *Hae*III to establish the digestion profile. The order is as follows;
A. pQIS 101N (W/T LZ) + pQIS 98 (W/T LZ) cut with *Hae*III
B. pQIS 113 (CaM pep) + pQIS 98 (W/T LZ) cut with *Hae*III
C. pQIS 115 (Asn 5 Lys) + pQIS 98 (W/T LZ) cut with *Hae*III
D. pQIS 113 (CaM Pep) + pQIS 114 (CaM) cut with *Hae*III
E. pQIS 116 (Leu 8 Pro) + pQIS 117 (Leu 8 Pro) cut with *Hae*III
F. pQIS 119 (OST1) + pQIS 118 (WBP1) cut with *Hae*III

Table 6.2: Summary of the Joint Results obtained by co-transformations, restriction protection and by DHPLC A visual rating of (+ and -) (active and inactive) has been assigned for these results after viewing the gel and DHPLC traces.

	Bait (pMETN)	Prey (pMETC)	Mcr response	Restriction Protection on gel	DHPLC profile
A	pQIS 101n (W/T LZ)	pQIS 98 (W/T LZ)	Active	+++	+++
B	pQIS 113 (CaM pep)	pQIS 98 (W/T LZ)	Inactive	+	-
C	pQIS 115 (N 5 K)	pQIS 98 (W/T LZ)	Active	+	+
D	pQIS 113 (CaM Pep)	pQIS 114 (CaM)	Inactive	-	-
E	pQIS 116 (L 8 P)	pQIS 117 (L 8 P)	Active	++	++
F	pQIS 119 (OST1)	pQIS 118 (WBP1)	Inactive	+	++

restriction fragments in these constructs (Appendix V). There might be a possibility that a weak interaction of yeast Wbp1 and Ost1 protein would lead to this kind of result. At the moment I cannot speculate on any reason for this result.

6.3.2: Nucleotide sequence analysis of the Mtase genes used in the two-hybrid experiments

Twelve primers were designed that facilitated the nucleotide sequencing of the complete 5' or 3' terminal sequence of the Mtases M.SPRI gene including the inserted sequence of interacting protein. The sequences of primers THI-TH12 that

were used for sequencing reactions are shown in Table 6.3. The primers used for sequencing against each construct are given in Table 6.4.

Table 6.3: Sequence of Primers used for sequencing two-hybrid plasmids

Primer Number	Sequence 5' → 3'	Length
TH 1	AGGGCTGGCAAGCCACGTT	19 bp
TH 2	CAGATGATGACAGCGAAGG	19 bp
TH 3	CCAGGTGCTGTTCCGAAT	18 bp
TH 4	TGCGGTTCATTCACATGCG	19 bp
TH 5	CGTTACGACTTCCTCTGT	18 bp
TH 6	CCAGTCCTGCTCGCTTCG	18 bp
TH 7	TAGGGAGCGACTTTACAT	18 bp
TH 8	GACGACAATAGACGTGTC	18 bp
TH 9	GTTTATTCCCACAATGCCC	19 bp
TH 10	AACGGCCCACCGGACACT	18 bp
TH 11	CAGAATTGACCTAGAGCTG	19 bp
TH 12	GAGAGGGACTGAAAATTGCAG	21`bp

Table 6.4: Two-hybrid plasmids and Primers used for them

	Construct No	Gene length to be sequenced (in bps)	Primers used
1	pQIS 98	478	TH1
2	pQIS 101	999	TH6, TH7
3	pQIS 113	990	TH6, TH7
4	pQIS 114	832	TH1, TH2
5	pQIS 115	999	TH6, TH7
6	pQIS 116	999	TH6, TH7
7	pQIS 117	478	TH1
8	pQIS 118	1681	TH1, TH3, TH4, TH5
9	pQIS 119	2343	TH6, TH7, TH8, TH9, TH10
10	pSPR(X)	1315	TH1, TH11, TH12

6.4: Discussion

As we pass into the post-genomic era, our attention naturally shifts to the products of the genomes that have been solved i.e. proteins. While the initial focus of this new age will be the description of the three dimensional-structure of all these gene products, not far behind must be the determination of how all of these proteins integrate to produce generalised phenotypes. The general principals for both genetic and biochemical tests for protein-protein interactions have been reviewed by Phizicky and Fields (1995). Genetic assays are thought of as analogous to biochemical assays (Hu *et al.*, 2000). The yeast two-hybrid system represents a powerful *in vivo* approach to analyse interactions between macromolecules and to screen for polypeptides that bind to a given "bait" protein (Field and Song, 1989).

Bacterial equivalents to the yeast two-hybrid system have not been developed until recently (Karimova *et al.*, 1998; also reviewed by Hu *et al.*, 2000). In this chapter we describe a novel bacterial two-hybrid system that allows easy *in vivo* screening of functional interactions between two proteins. The two-hybrid system that we have set up is based on the reconstitution of an active multi-specific Mtase M.SPRI. It takes advantage of the modular structure of M.SPRI, which has perhaps a similar two domain structural organisation to that of monospecific Mtases *Hha*I, and *Hae*III. The multi-specific Mtases are characterised by an unprecedented structural plasticity (Walter *et al.*, 1992). We have used individually inactive N- and C-terminal containing segments of M.SPRI (detached at the *Msp*I specific TRD region) to complement and form an active enzyme with the help of an appended leucine zipper motif.

Two compatible plasmids were constructed each carrying the ORFs of the N- or C-terminal components of M.SPRI fused to a leucine zipper motif. When expressed separately in *E. coli*, these polypeptides do not assemble into an active enzyme unless the interacting polypeptides are genetically fused to these fragments. In the bacterial two-hybrid system, interaction between the two chimaeric proteins results in functional complementation between the two fragments and restoration of enzymatic activity. The resulting active Mtase triggers the *E. coli* resident *mcr* gene that challenges the growth of colonies by biological restriction of methylated DNA. A functional methyltransferase is revealed by a positive mcr^+ host response, in the form abolition of plasmid transformants.

The experiments described in this chapter have demonstrated that a leucine zipper can mediate the formation of a heterodimer of the multi-specific methyltransferase (M.SPRI). The 28 amino acid leucine zipper motif of C/EBP dimerizes with high affinity. Leucine zipper proteins exist in an alpha helical conformation and the dimer interface of a leucine zipper is formed by the interaction of the side chains of the residues at the a, d, e, and g position of the $(abcdefg)_n$ heptad repeat (Zeng *et al.*, 1997). In a control experiment using the bacterial two-hybrid system it was found that the deletion of the leucine zipper domain is sufficient to abolish M.SPRI assembly. It has therefore been shown that this bacterial two-hybrid system is capable of "reporting" on the interactions between small peptides (such as, for example the C/EBP leucine zipper).

I have extended this analysis and studied the interactions between two yeast proteins namely mlc1p and the N-terminal fragment of Iqg1p. In this case however, no interaction was observed. It was thought that this might represent a limitation of the bacterial system. However, when a similar experiment was carried out with yeast two-hybrid system, using the same two proteins, no interaction was recorded (Personal communication, Boyne, J). When a full-length Iqg1 protein (172.729 kDa in size instead of 98.744 kDa) added to mlc1, an interaction was detected by co-immuno-precipitation (Boyne *et al.*, 2000).

In addition, I investigated whether point mutations in a leucine zipper could abolish the homodimer formation. It was found that such point mutation(s) did not affect dimerisation, since no colonies were obtained on multiple antibiotic plates arising from complementation of the expressed fragments of M.SPRI. Kouzarides and Ziff (1988) have reported that the mutation of individual leucine(s) in the leucine zipper protein fos, do not affect complex formation with Jun.

Therefore either these mutations have not affected the binding properties at all or it might be possible that the mutation has reduced the binding properties. If it is the later case, the significance of these relatively small changes in binding activity is difficult to judge. Therefore we can claim that the system is powerful enough to detect even weak interaction, or alternatively it is not capable of distinguishing between strong and weak interactions by microbiological means.

I have also tried using the bacterial two-hybrid system to analyse the dimerisation capacity of a calmodulin-binding peptide with calmodulin. The interaction between calmodulin and the calmodulin-binding peptide is well documented. However, I did

not find any evidence for interactions using this bacterial two-hybrid system. In addition experiments were performed to study the interactions between two yeast membrane proteins Wbp1 and Ost1. The *Saccharomyces cerevisiae* oligosaccharyltransferasese (OST) is a protein complex composed of six polypeptides (Kelleher and Gilmore, 1994). Oligosaccharyltransferases of this type catalyse the transfer of a high mannose oligosaccharide from a lipid linked oligosaccharide donor onto an Asn acceptor site within Asn-X-Ser/Thr consensus motifs in newly synthesised proteins. This Asn linked glycosylation of proteins via OST, takes place in the lumen of the endoplasmic reticulum. In native rough microsomal membranes, components of the OST complex are in close proximity to membrane-bound ribosomes, since they can be chemically cross-linked to the 60S ribosomal subunit. Mammalian OST48L and RIL are homologous to yeast Ost1 and Wbp1 and have been shown to interact by a yeast two-hybrid approach and by biochemical assays (Fu *et al.*, 1997). In this work yeast Wbp1 and Ost1 did not show an interaction using the bacterial two-hybrid system. Such inconsistencies between the genetic and biochemical methods make it essential that all putative interactions are verified by more than on method. However it is unclear why this experiment failed but it could be due to various factors including the basic differences in *E. coli* and yeast. The other possible reasons might be interference of large protein with Mtase, poor folding, susceptibility to proteolysis or the inappropriate ratio of expression of the two proteins. Living cells need globular proteins in correct quantities at appropriate times. It is therefore as important to be able easily to degrade these proteins, as it is to be able to synthesise them. Globular proteins in living cells usually have a rather rapid turnover and their native states have therefore evolved to be only marginally stable (Branden and Tooze, 1999). Hence, there could be a possibility that very large proteins (especially membrane proteins) could not be used for interaction studies by this bacterial two-hybrid assay. Further experiments on bacterial two-hybrid system are currently underway in our lab.

Schematic diagrams
showing the plasmid sketches
with gel pictures

(Including Figures 6.3 ~ 6.7, 6.10 ~ 6.11, 6.14 ~ 6.20)

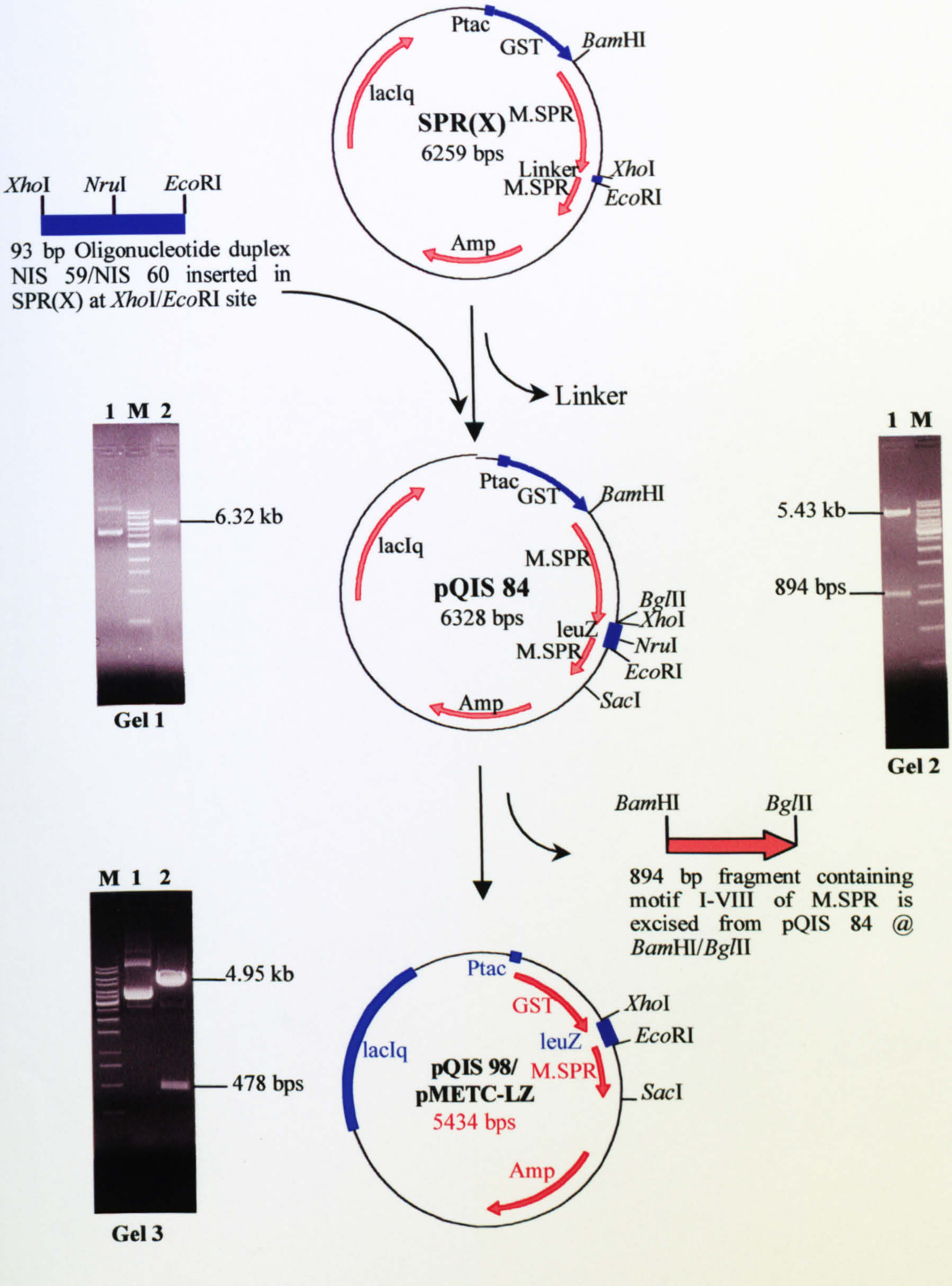


Figure 6.3: Schematic representation showing the construction of pMETC-LZ (pQIS 98) encoding a leucine zipper fused to motif IX and X of M.SPRI. Gel 1: Lane 1: pQIS 84 incubated with *Hae*III. (DNA is protected by *Hae*III restriction digestion indicating that an active protein is encoded by this plasmid). **Lane 2:** pQIS 84 digested with *Nru*I. (Linear DNA digested with *Nru*I verifies the insertion of linker and formation pQIS 84). **Gel 2: Lane 1** pQIS 84 digested with *Bam*HI and *Bgl*II. This procedure was used in order to excise the N-terminal M.SPR fragment containing motif I-VIII. (Fragment sizes are 5434 and 894 bps) **Gel 3: Lane 1** pQIS 98/pMETC-LZ digested with *Bam*HI. Uncut DNA indicated the loss of *Bam*HI site. **Lane 2** pQIS 98/pMETC-LZ digested with *Xho*I and *Sac*I (fragments sizes are 4956 and 478 bp). **M:** GeneRuler™ 1 kb DNA ladder in all gels.

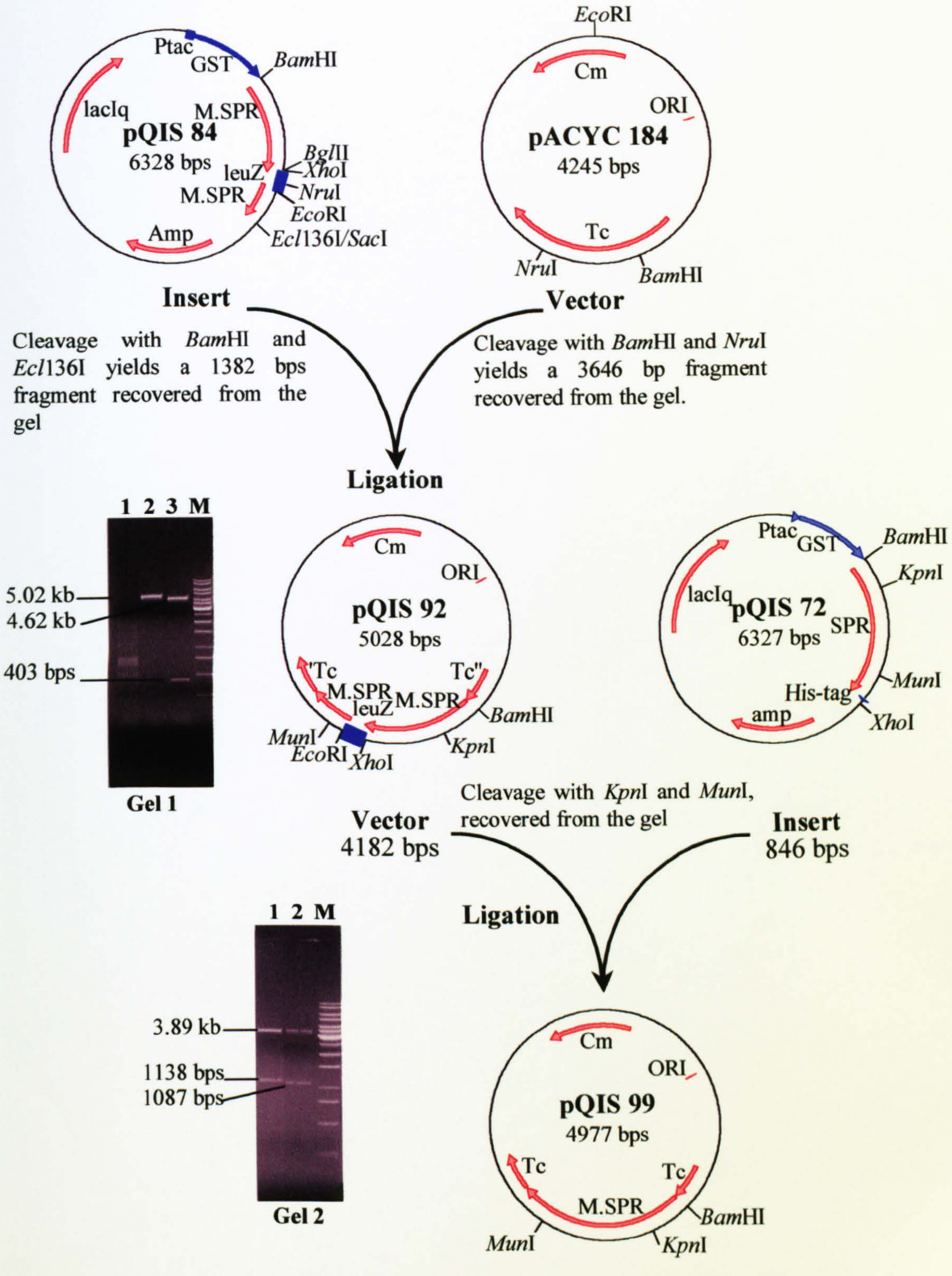
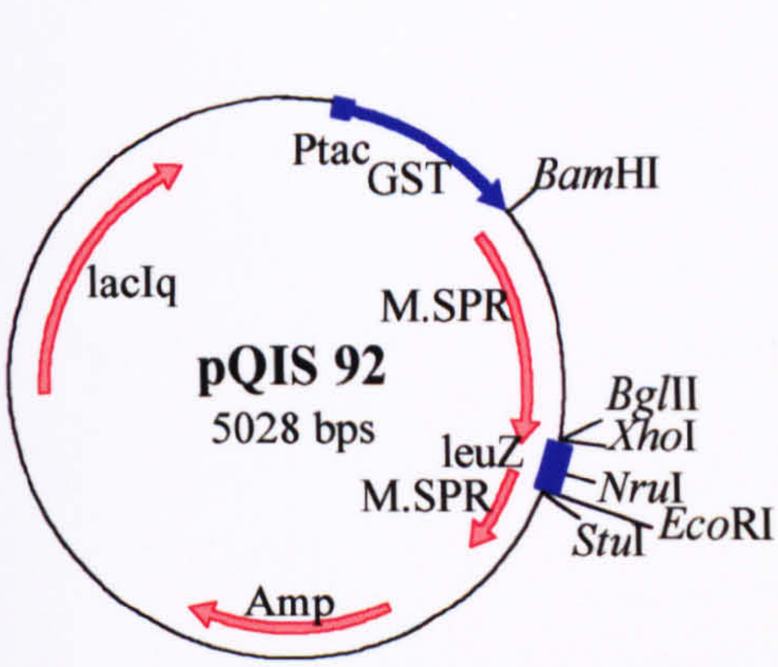
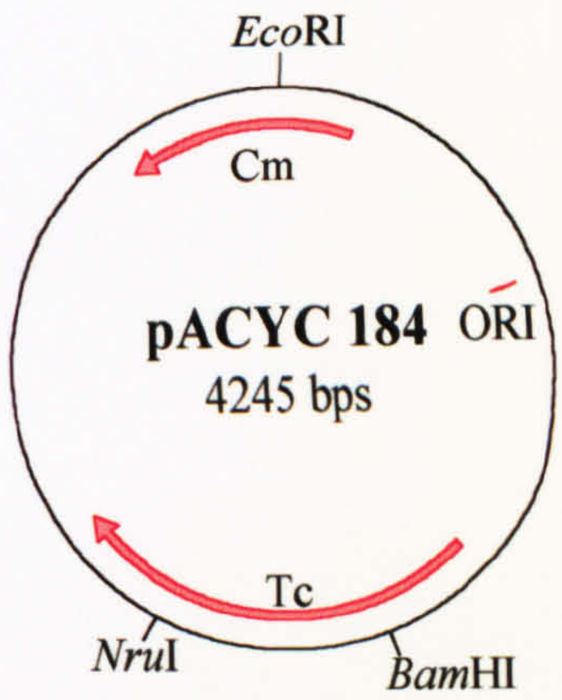


Figure 6.4: Schematic representation showing the construction of pQIS 92 and pQIS 99. The construct pQIS 92 encodes for a version of M.SPRI in which the *Msp*I TRD is replaced with a leucine zipper region. The construct pQIS 99 represents wild type M.SPRI gene cloned into the plasmid pACYC 184. **Gel 1: Lane 1:** pQIS 92 digested with *Hae*III (showing partial protection). **Lane 2:** pQIS 92 digested with *Nco*I. (Linear DNA fragment). There is no *Nco*I site in pQIS 84 whereas pACYC 184 contains one *Nco*I site. **Lane 3:** pQIS 92 digested with *Eco*RV. (Fragments of sizes 4625 and 403 bps). One *Eco*RV site is present in M.SPRI whereas the other is in pACYC 184 and the fragments yielded are as expected. **Gel 2: Lane 1** pQIS 92 digested with *Bam*HI and *Mun*I. (Fragments sizes are 1138 and 3890 bp). **Lane 2:** pQIS 99 cut with *Bam*HI and *Mun*I (Fragments sizes are 1087 and 3890). A comparison of the difference in size of the lower bands in lane 1 and 2 illustrates the difference between pQIS 92 and pQIS 99. **M:** GeneRuler™ 1 kb DNA ladder in both gels.





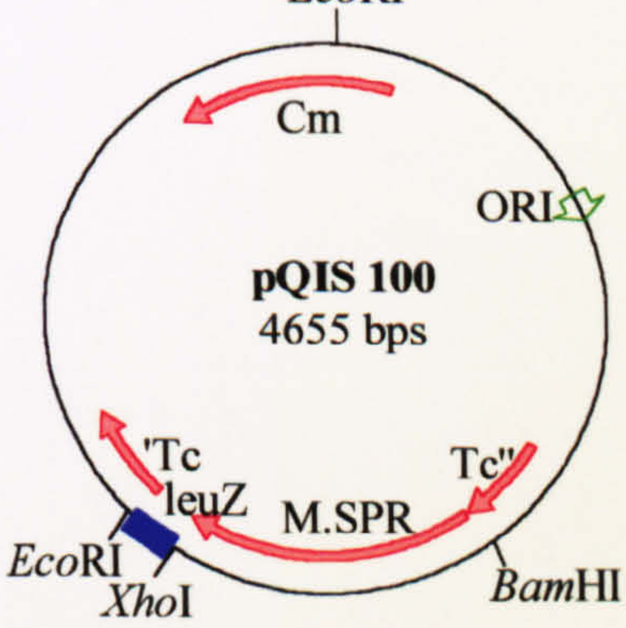
Insert

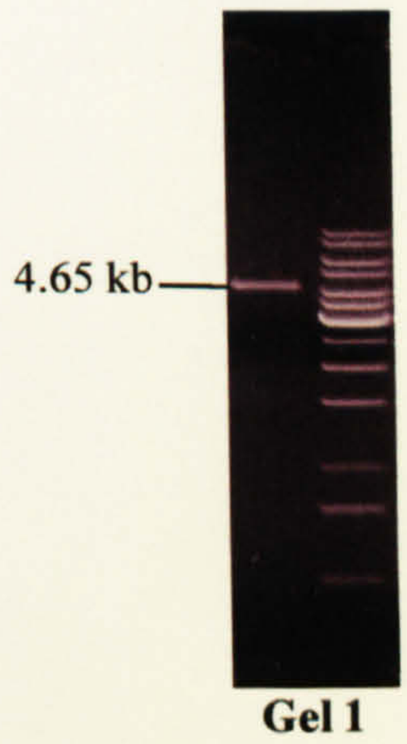
Cleavage with *Bam*HI and *Stu*I, 1009 bps fragment recovered from the gel

Vector

Cleavage with *Bam*HI and *Nru*I, 3646 bp fragment recovered from the gel.

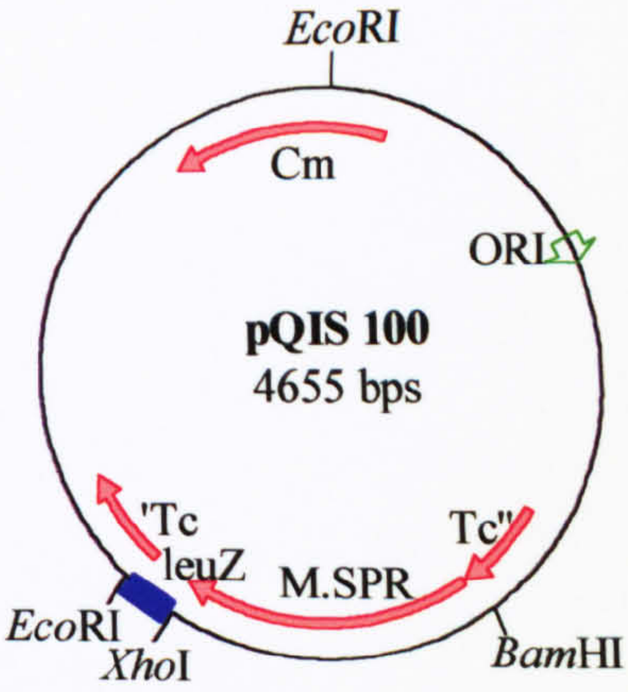
Ligation





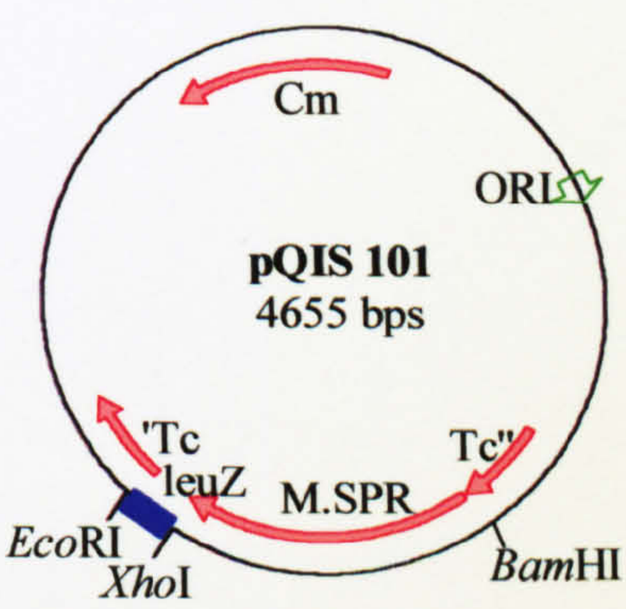
Gel 1

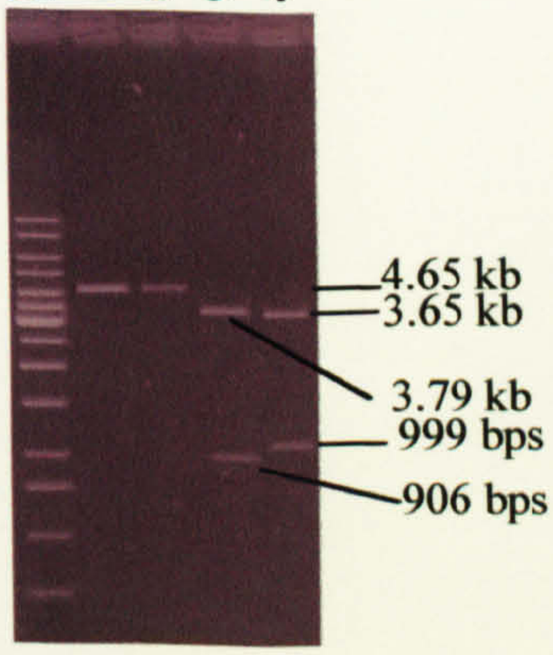
Figure 6.5: Schematic representation showing the construction of pQIS 100. This results in a construct which encodes motif I-VIII of M.SPRI fused with the leucine zipper region. **Gel 1:** pQIS 100 cut with *Kpn*I (for the verification of the construct). The M.SPRI gene contain a unique *Kpn*I site and therefore the plasmid pQIS 100 has been linearised. **M:** GeneRuler^{TM} 1 kb DNA ladder.



Gel 1

Site-directed silent mutagenesis to delete *Eco*RI site from Chloramphenicol gene.





Gel 2

Figure 6.6: Schematic representation showing the construction of pQIS 101. Gel 1: PCR product generated after site-directed mutagenesis. **Gel 2: lane 1** pQIS 101 digested with *Eco*RI. (Linear DNA). **Lane 2:** pQIS 101 digested with *Kpn*I (Linear DNA). **Lane 3:** pQIS 101 cut with *Bam*HI and *Xho*I. (Fragment sizes 3749 and 906 bps). **Lane 4:** pQIS 101 cut with *Bam*HI and *Eco*RI. (Fragment sizes 3656 and 999 bps). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

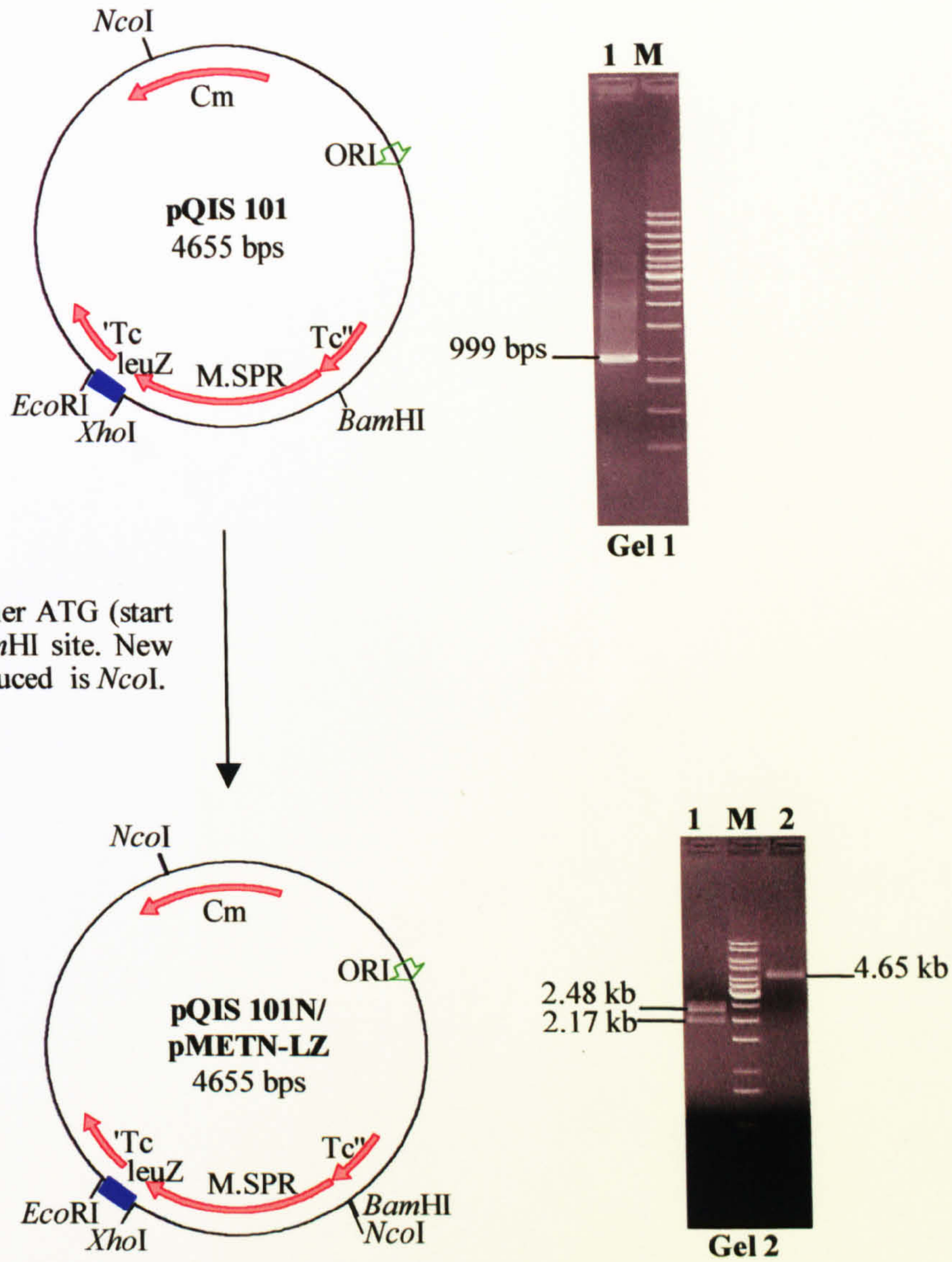


Figure 6.7: Schematic representation showing the construction of pMETN-LZ (pQIS 101N) (with corrected start codon). Gel 1: Lane 1 shows the generation of 999 bp PCR product. (Using primers NIS 77 and NIS 73). **Gel 2: Lane 1:** pQIS 101N cut with *Nco*I. (Fragment sizes 2174 and 2481) (pQIS 101N contains a corrected start codon and a new *Nco*I site). **Lane 2:** pQIS 101 cut with *Nco*I. (Linear DNA). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

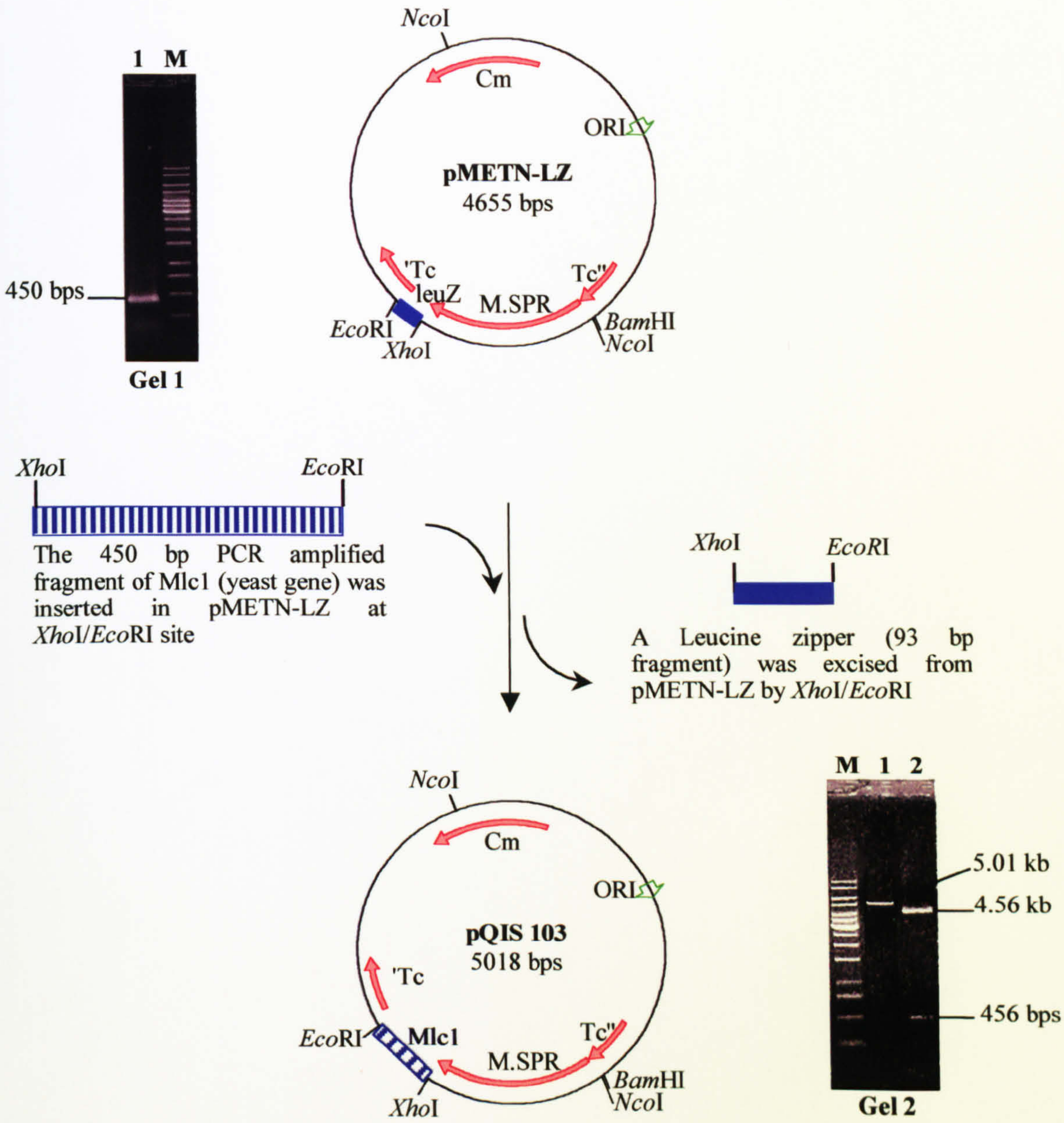
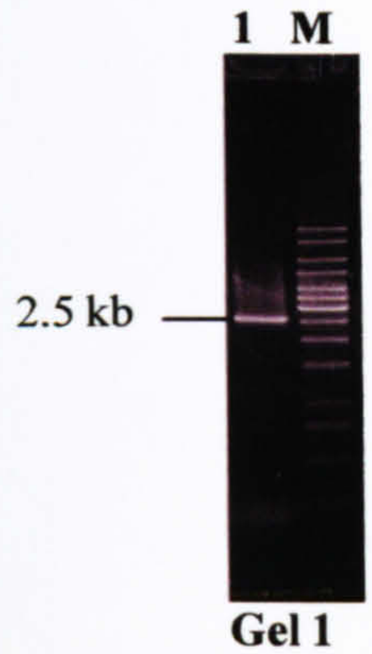
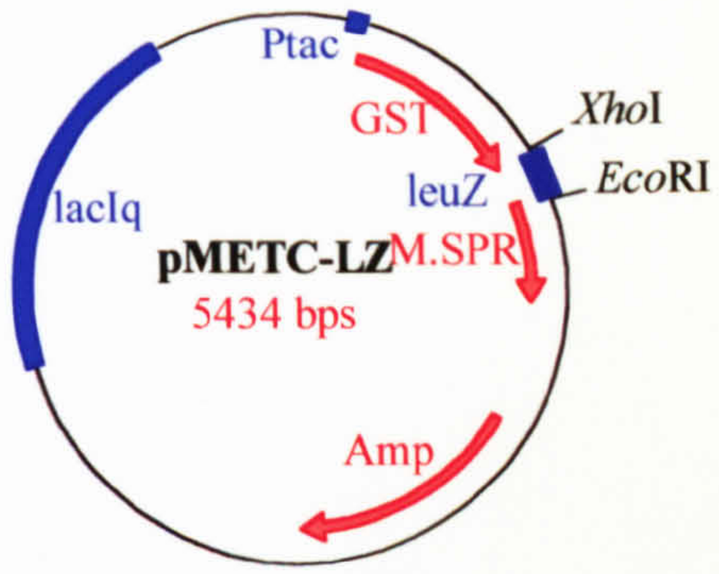


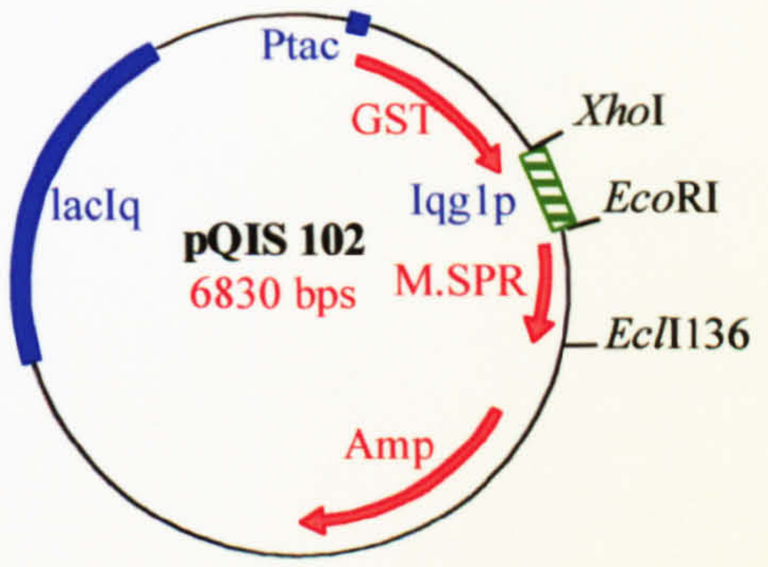
Figure 6.10: A Schematic representation showing the construction of pQIS 103. This plasmid encodes a protein composed of the Mlc1 gene from yeast fused to motif I-VIII from M.SPRI. **Gel 1:** shows the 450 bps Mlc1 fragment generated by PCR. **Gel 2: Lane 1:** pQIS 103 cut with *Stu*I. (linear DNA, 5018 bps). **Lane 2:** pQIS 103 cut with *Xho*I and *Eco*RI. (Fragment sizes 4562 and 456 bps). **M:** GeneRuler^{TM} 1 kb DNA ladder in both gels.

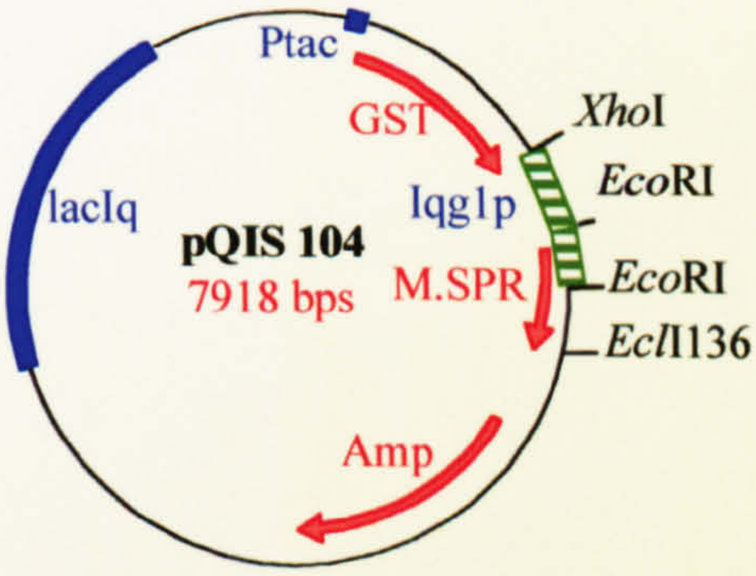




The 2.5 kb PCR amplified fragment of Iqg1p yeast gene was inserted in pMETC-LZ at *Xho*I/*Eco*RI site in two steps

A Leucine zipper (93 bp) was excised from pMETC-LZ by *Xho*I/*Eco*RI





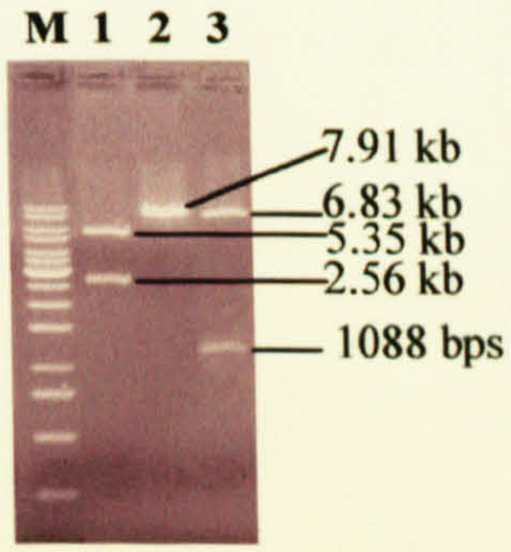
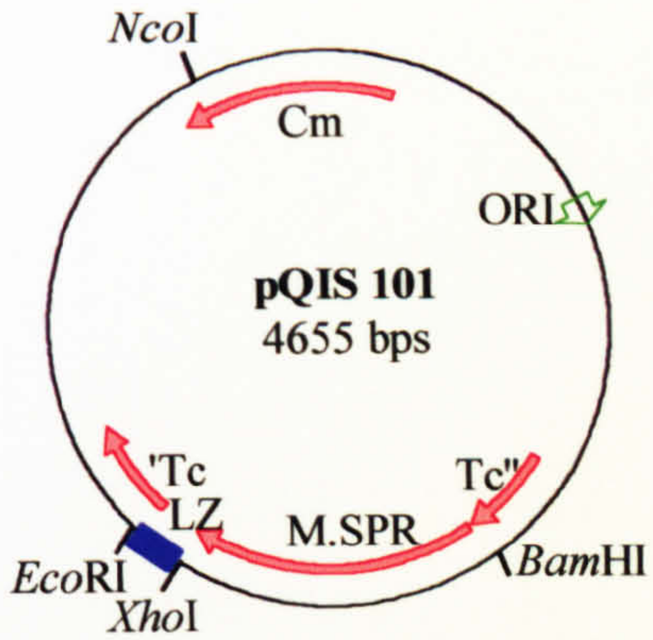


Figure 6.11: A schematic representation showing the construction of pQIS 118. This plasmid encodes a protein composed of the Iqg1p1 gene from yeast fused to motif IX-X from M.SPRI. **Gel 1: Lane 1** shows the 2.5 kb Iqg1p fragment generated by PCR. **Gel 2: Lane 1:** pQIS 104 cut with *Bam*HI and *Sal*I. (Fragment sizes are 5353 and 2565 bps). **Lane 2:** pQIS 104 cut with *Xho*I. (Linear DNA 7.9 kb size). **Lane 3:** pQIS 104 cut with *Eco*RI. (Fragment sizes are 6830 and 1088 bps). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

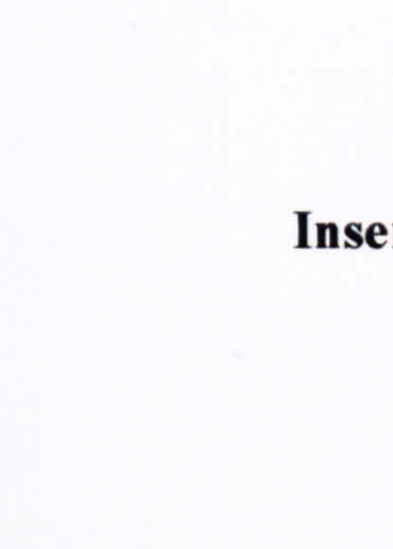
85 bp Oligonucleotide duplex NIS 71/NIS 72 inserted in pMETN-LZ at *Xho*I/*Eco*RI site



Cleavage with *Xho*I and *Eco*RI, 4562 bp fragment recovered from the gel.

Insert **Vector**

Ligation



Gel 1

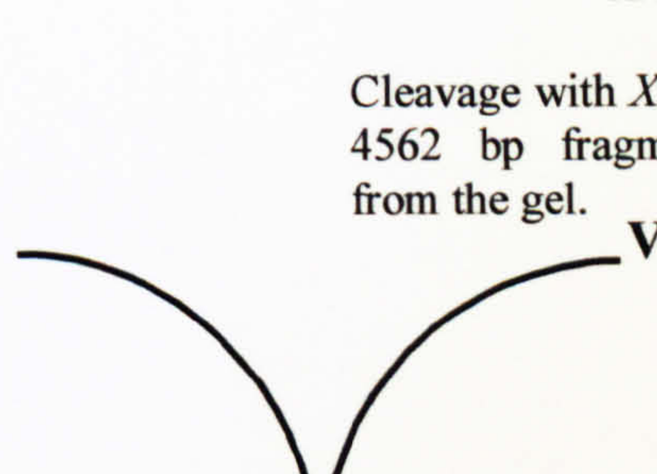
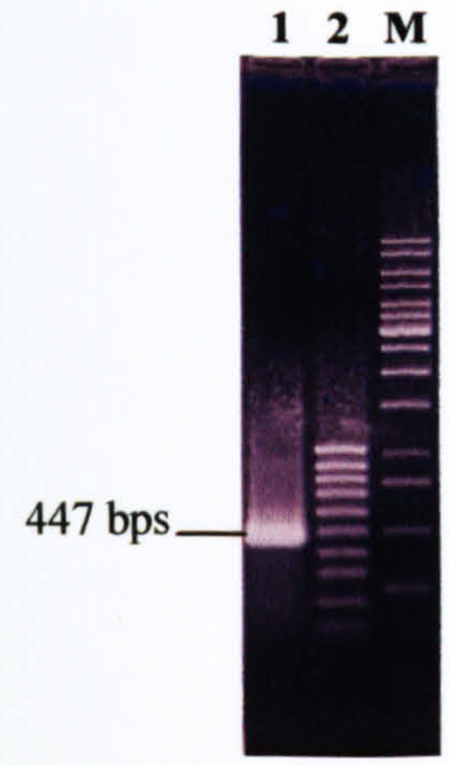
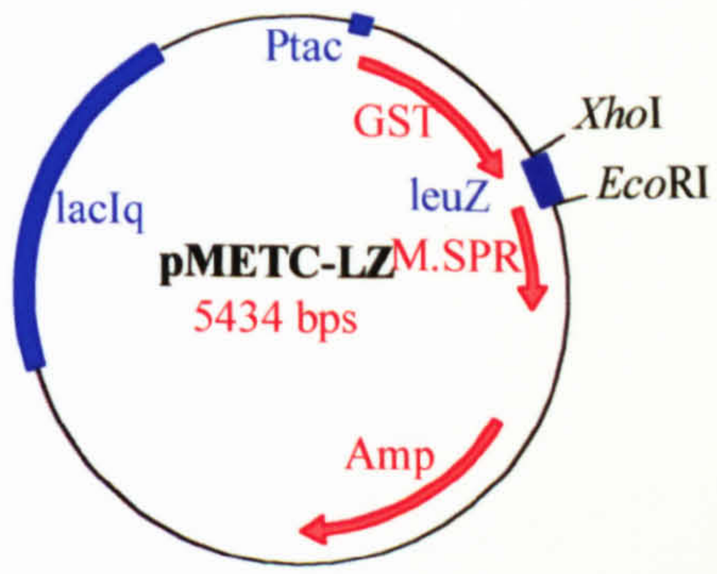


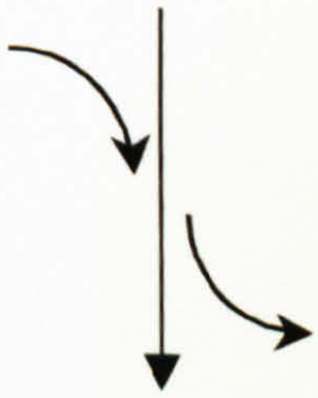
Figure 6.14: Schematic representation showing the construction of pQIS 113. This plasmid contains motif I-VIII of M.SPR fused with the Calmodulin Binding Peptide sequence. The oligonucleotide contain a *Nco*I site and there is an existing *Nco*I site in the plasmid. **Gel 1:** Upon restriction the plasmid generated two fragments (3149 + 1497 bps) which confirmed the identity of the new construct. **M:** GeneRuler™ 1 kb DNA ladder.



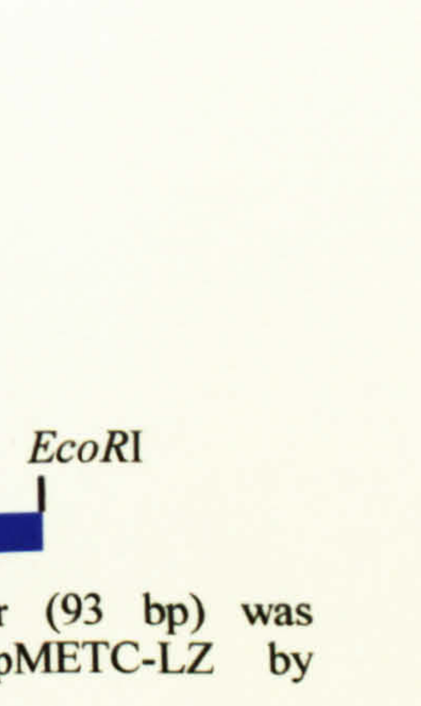
Gel 1



A 447 bp PCR amplified fragment of Calmodulin gene was inserted in pMETC-LZ at *Xho*I/*Eco*RI site

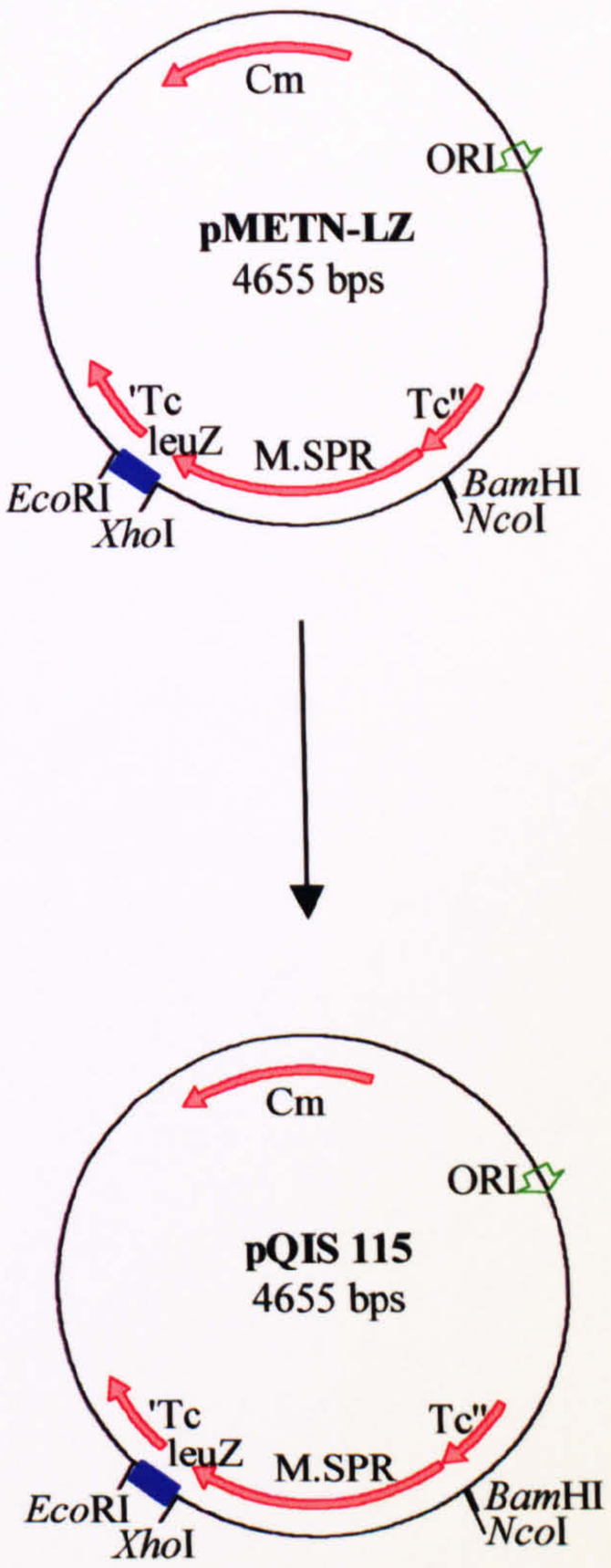


A leucine zipper (93 bp) was excised from pMETC-LZ by *Xho*I/*Eco*RI



Gel 2

Figure 6.15: Schematic representation showing the construction of pQIS 114R, which encodes a calmodulin gene fused with motif IX-X from M.SPRI. Gel 1 (1.5 % agarose): Lane 1 447 bps fragment generated by PCR. **Lane 2:** 100 bp DNA ladder. **Gel 2: Lane 1** shows the restriction digestion of pQIS 114R by *Xho*I and *Eco*RI which generated a 450 bp fragment and confirmed the insertion of calmodulin gene and the formation of new construct. **M:** GeneRuler™ 1 kb DNA ladder in both gels.



Gel 1

Figure 6.16: Schematic representation showing the construction of pQIS 115. This plasmid contains the gene for a C/EBP leucine zipper with a Asn 5 Lys mutation). **Gel 1** shows the restriction digestion of pQIS 115 with *Cla*I. The resulting fragments 3375 and 1280 bps confirmed the identity of the new construct. **M:** GeneRuler™ 1 kb DNA ladder.

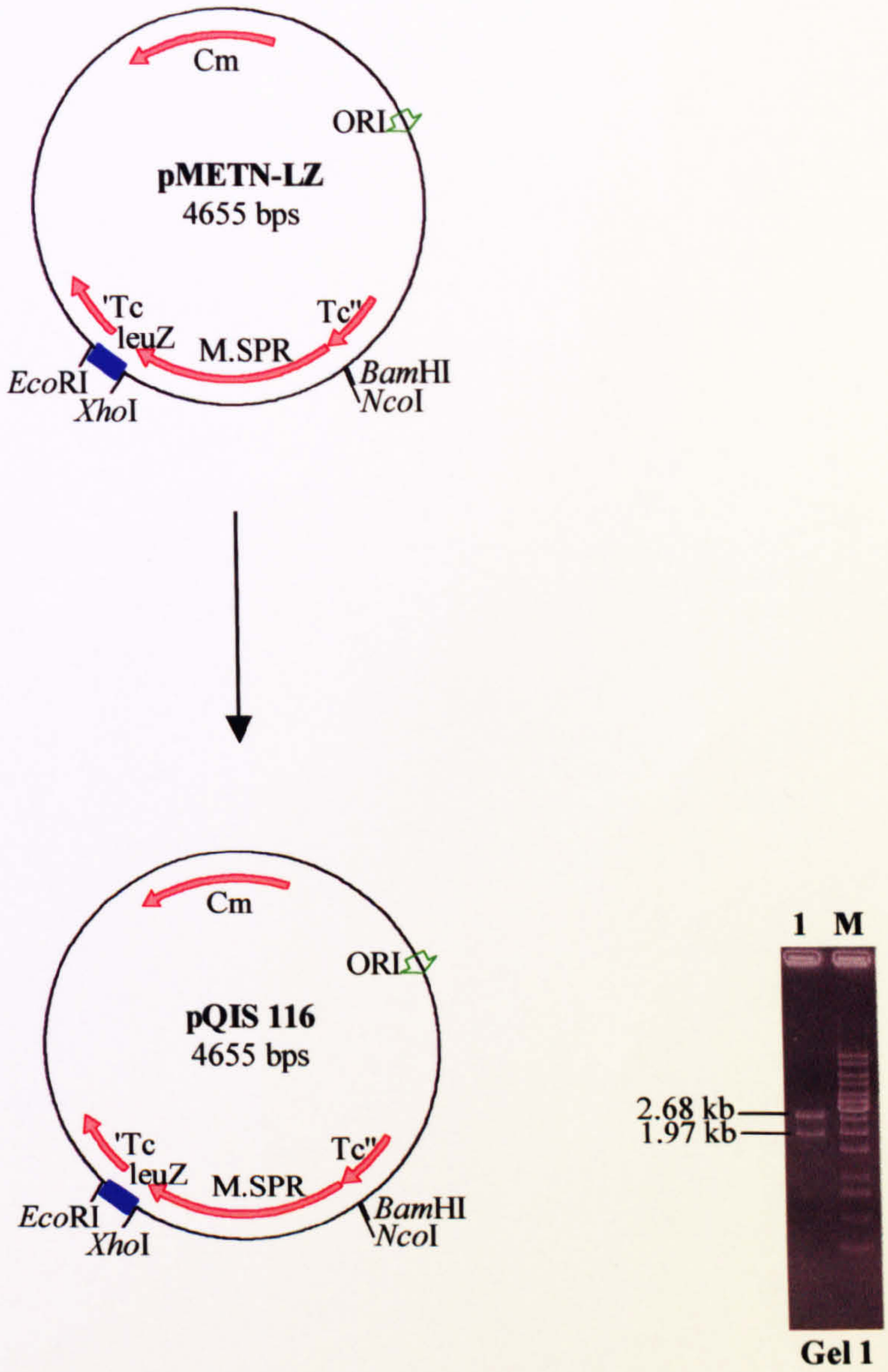
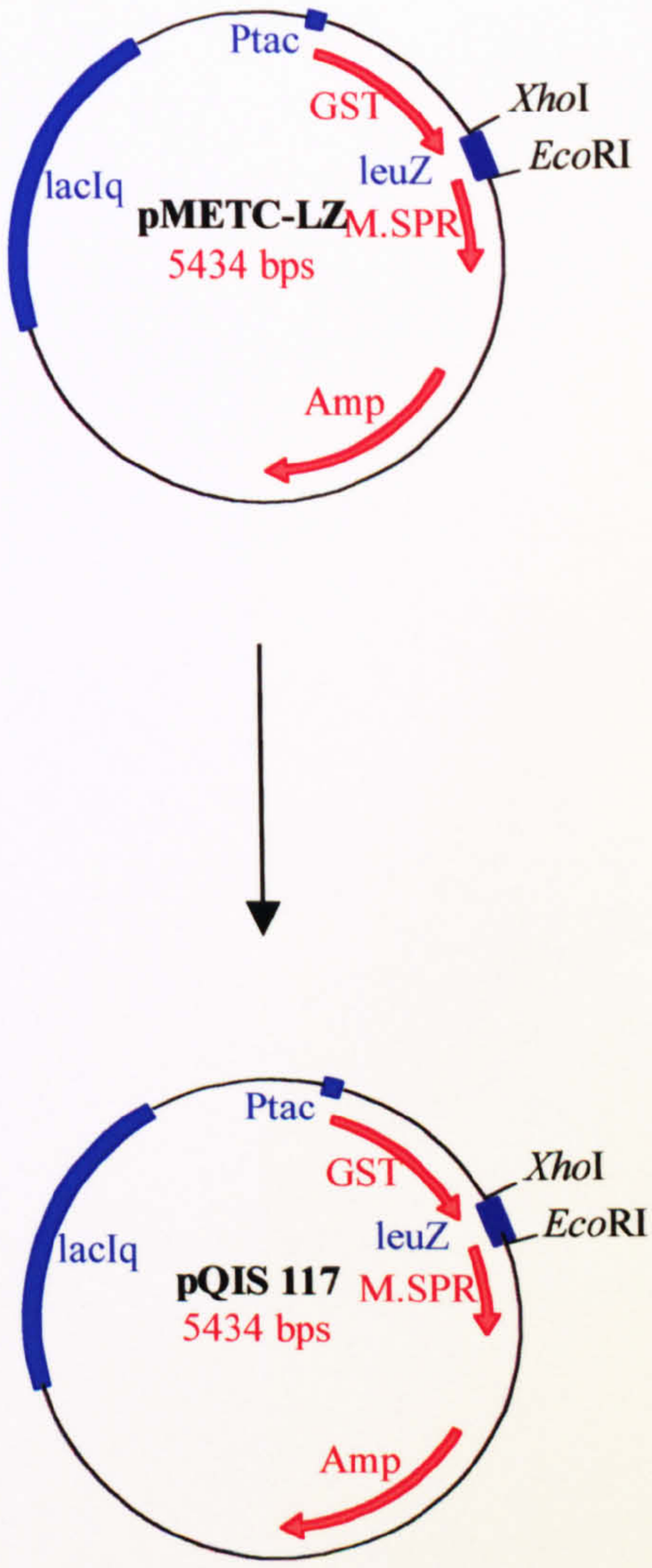
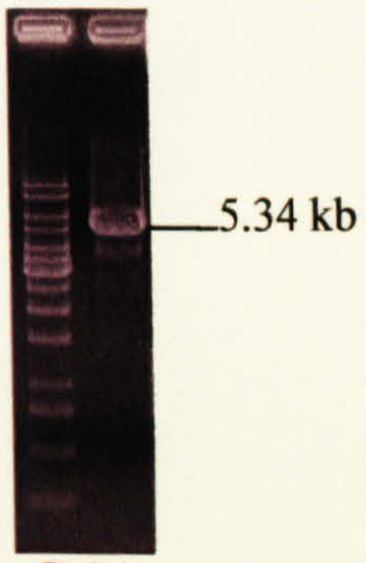


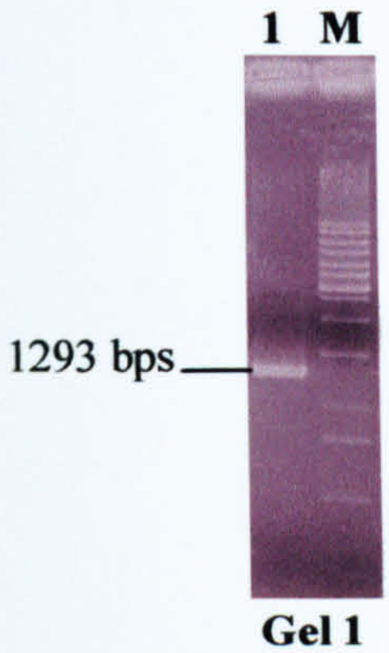
Figure 6.17: Schematic representation showing the construction of pQIS 116. This plasmid contains the gene for a C/EBP leucine zipper with a Leu 8 Pro mutation). The plasmid DNA was cut with restriction enzyme *Sac*II to confirm the identity of newly introduced mutation. **Gel 1** shows the restriction digestion of pQIS 116 with *Sac*II. The resulting fragments 2684 and 1971 bps confirmed the identity of the new construct. **M:** GeneRuler™ 1 kb DNA ladder.



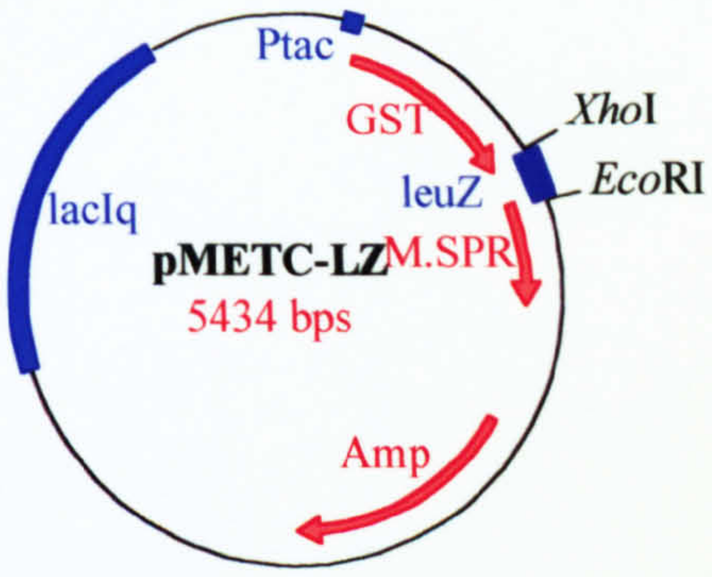


Gel 1

Figure 6.18: Schematic representation showing the construction of pQIS 117. This plasmid contains the gene for a C/EBP leucine zipper with a Leu 8 Pro mutation). **Gel 1** shows the digestion of pQIS 117 with *Sac*II. Linear DNA fragment of 5.4 kb confirmed the identity of newly introduced mutation and formation of new construct. **M:** GeneRuler™ 1 kb DNA ladder.

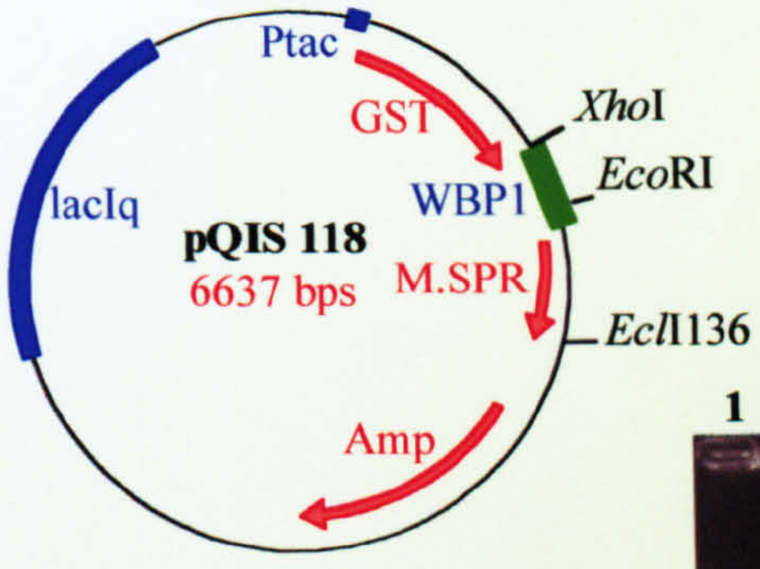


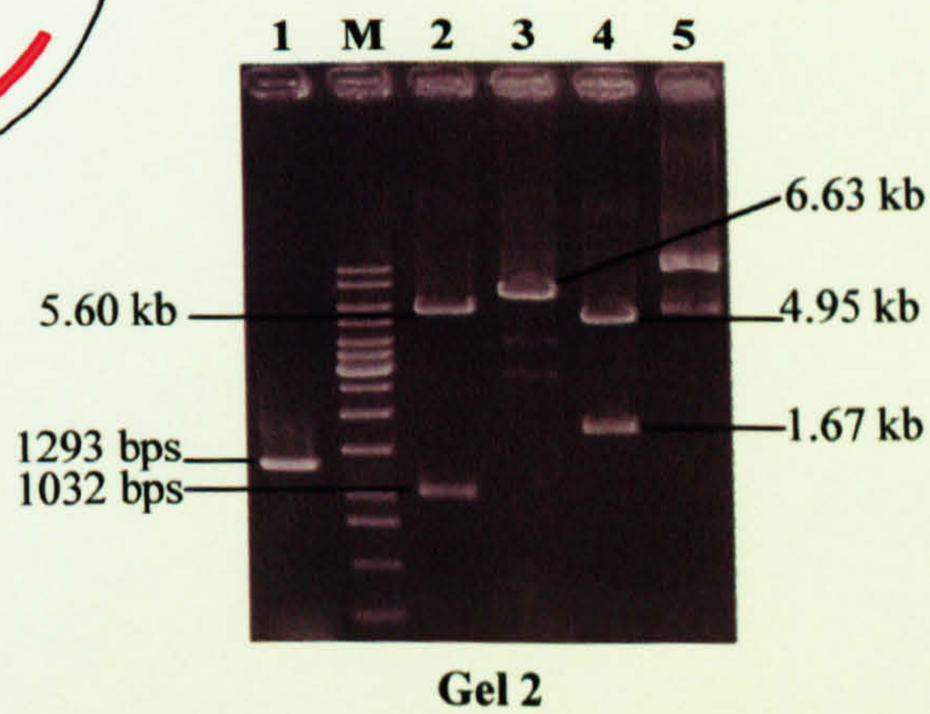
Gel 1



1293 bp PCR amplified fragment of WBP1 yeast gene was inserted in pMETC-LZ at *Xho*I/*Eco*RI site

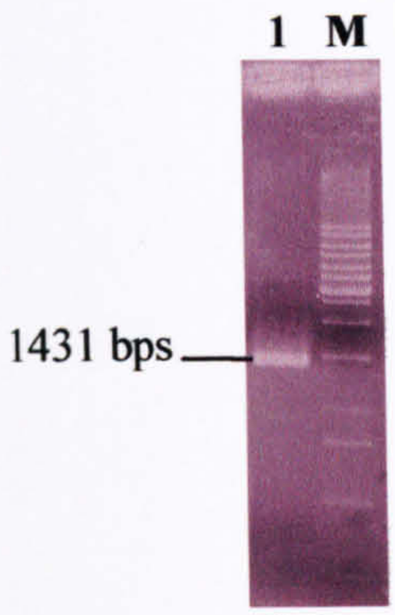
Leucine zipper 93 bp fragment was excised from pMETC-LZ at *Xho*I/*Eco*RI



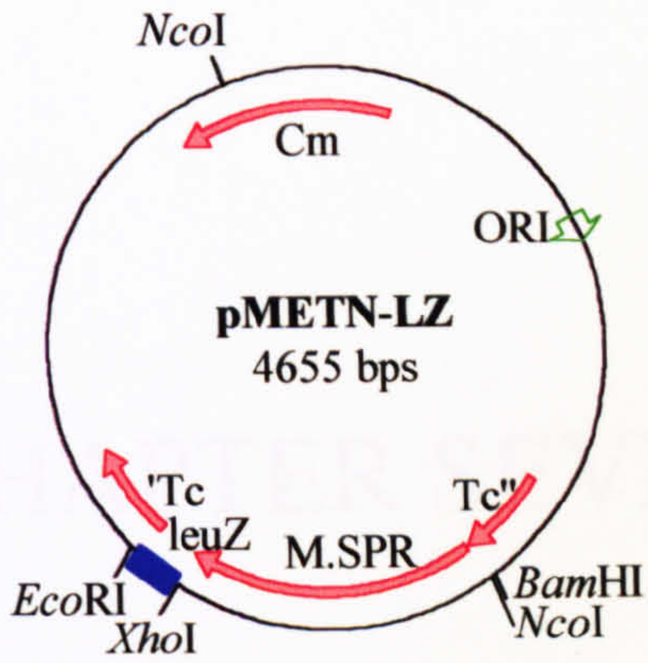


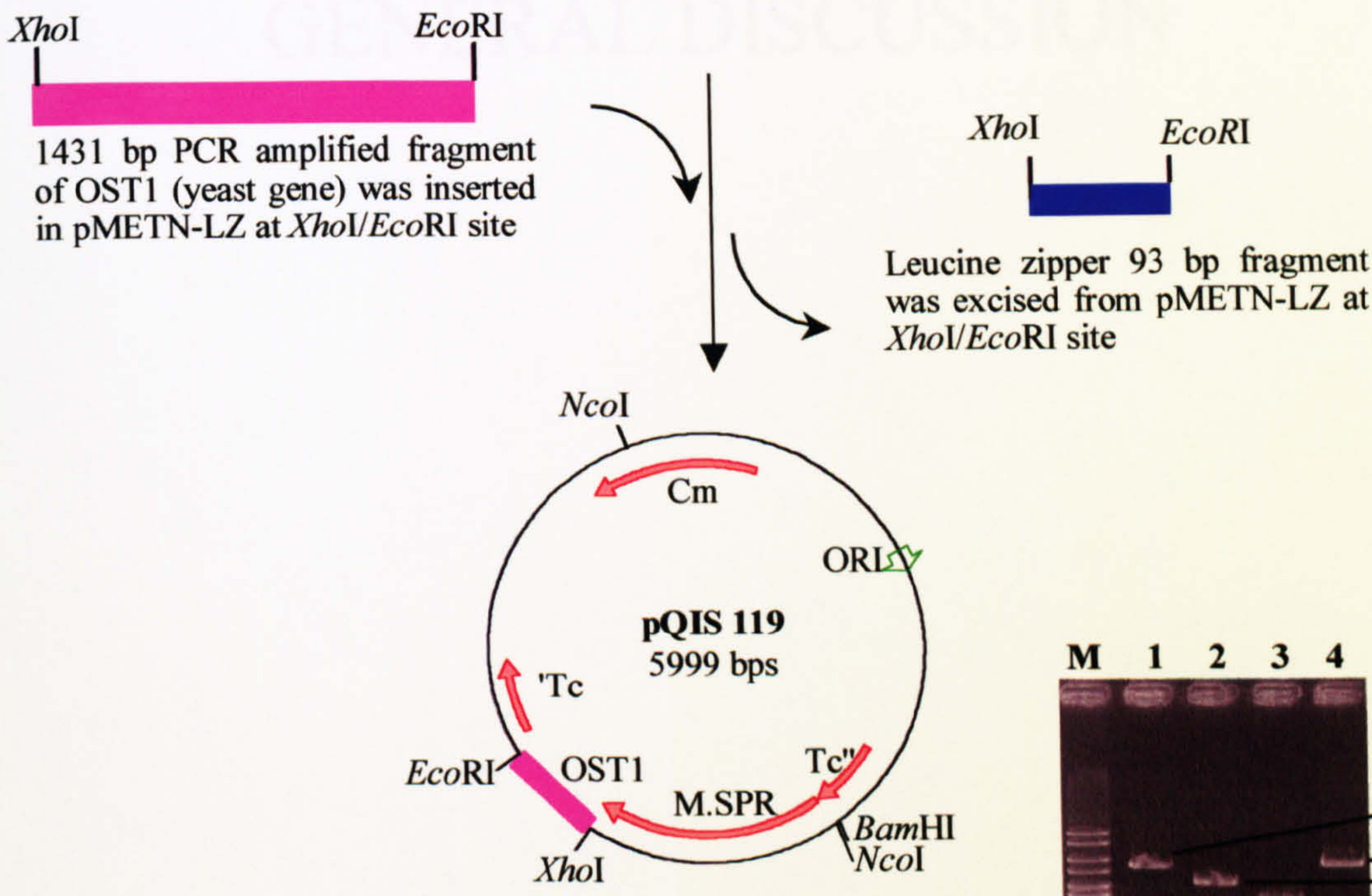
Gel 2

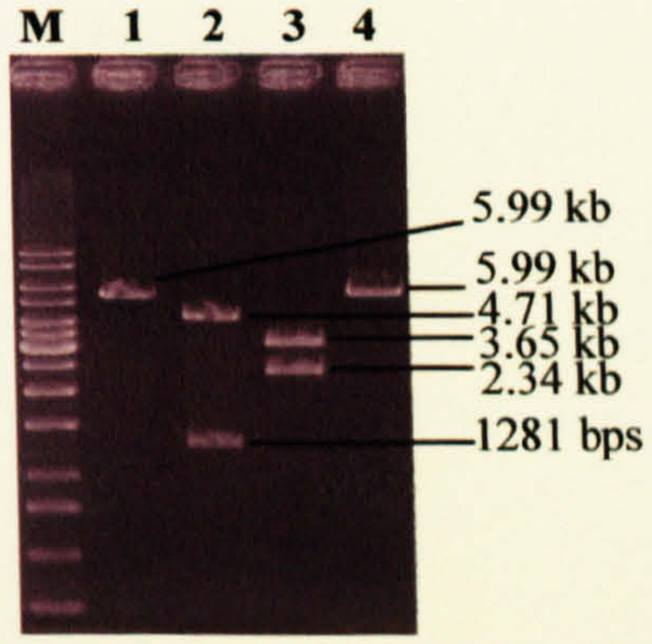
Figure 6.19: Schematic representation showing the construction of pQIS 118. This plasmid contains a yeast gene which encodes WBP1 and fused with the sequence of motif IX-X of M.SPRI. **Gel 1: Lane 1** shows the 1293 bps WBP1 fragment generated by PCR. **Gel 2: Lane 1:** 1293 bp PCR product as control DNA. **Lane 2:** pQIS 118 cut with *Xho*I and *Eco*RI (Fragment sizes 5605 and 1032 bps). (A smaller insert size due to internal *Eco*RI site. Compare lane 1 and lane 2). **Lane 3:** pQIS 118 cut with *Bam*HI. Linear DNA. (*Bam*HI site is present only in WBP1 gene). **Lane 4:** pQIS 118 cut with *Xho*I and *Ecl*136. (Fragment sizes 4958 and 1679 bps). **Lane 5:** pQIS 118 cut with *Xba*I. (Uncut DNA, blocked by ovelapping dam methylation). **M:** GeneRuler™ 1 kb DNA ladder in both gels.



Gel 1







Gel 2

Figure 6.20: Schematic representation showing the construction of pQIS 119. This plasmid contains a yeast gene which encodes OST1 fused with the sequence of motif I-VIII of M.SPRI. **Gel 1:** Lane 1 shows the 1431 bps OST1 fragment generated by PCR. **Gel 2: Lane 1:** pQIS 119 cut with *Mun*I yielded linear DNA. (*Mun*I site is present only in OST1 gene). **Lane 2:** pQIS 119 cut with *Kpn*I and *Mun*I (Fragment sizes 4718 and 1281 bps). **Lane 3:** pQIS 119 cut with *Bam*HI and *Eco*RI. (Fragment sizes 3656 and 2343 bps). **Lane 4:** pQIS 119 cut with *Stu*I yielded linear DNA. (*Stu*I site is present only in OST1 gene). **M:** GeneRuler™ 1 kb DNA ladder in both gels.

CHAPTER SEVEN

GENERAL DISCUSSION

CHAPTER SEVEN

GENERAL DISCUSSIONS AND FUTURE WORK

7.1: GENERAL DISCUSSION

C5-DNA Mtases are found in a range of organisms and catalyse the transfer of a methyl group from the co-factor SAM to the C5 position of a specific cytosine within DNA. Sequence comparisons of multispecific phage-encoded C5-Mtases has led to the identification of contiguous overlapping domains that are responsible for the recognition of specific DNA sites. These regions of the polypeptide chain are termed target-recognition domains (TRDs). The cytosine-specific Mtase M.SPRI from *Bacillus subtilis* bacteriophage contains three TRDs which methylate DNA sites recognised by the restriction enzymes *Eco*RII, *Msp*I and *Hae*III (Trautner *et al.*, 1988). M.SPRI has a structure consisting of 10 well-conserved motifs surrounding the variable region containing the TRDs.

Crystallographic studies on two mono-specific bacterial Mtases have shown that general enzymatic functions such as cofactor (SAM) binding and catalysis are provided by a large "catalytic" domain formed by the N-terminal part of the enzyme, which includes the conserved motifs I-VIII. The second smaller "recognition" domain of the Mtases is composed of the central variable region containing the target-recognising domain(s) (TRDs) and part of the C-terminal motifs IX and X (Klimašauskas *et al.*, 1994; Reinisch *et al.*, 1995). Motif X forms one of the bridges between the two domains. In the monospecific Mtases, studies of hybrids between two Mtases of different specificities, have demonstrated that there may be some direct structural interactions between the variable region and motif IX. The integrity of the small domain, provides a suitable interface between the variable, TRD-containing region and the conserved motifs IX and X, and appears to be crucial for enzymatic activity (Mi and Roberts, 1992).

The experiments presented in this thesis describe the introduction and analysis of mutations into two C5-DNA Mtases, M.SPR and M.*Hha*I. M.*Hha*I was chosen owing to our knowledge of its structure. The main aim of this mutational study was to dissect the internal relationship of the small domain comprising variable region and motif IX.

In the first phase of the experimental work (chapter 3) three individual point mutations were engineered into the sequence of M.SPRI in motif IX, in order to mimic motif IX of mono-specific Mtases. The protein encoded by the Phe 386 Ala mutant of M.SPRI gave partial activity. Analysis of the kinetics of covalent complex formation (Pinerbasi *et al.*, 1998) has shown that such mutants exhibit a delay in the initial phase of covalent complex formation, but after prolonged incubation, the extent of covalent complex formation appeared very similar to the wild type. The altered kinetic properties observed may be explained by altered conformational properties of the mutant or a modification of one or more chemical steps. While the M.SPRI Phe 386 Ala mutant features a decrease in methylation activity, a similar trend of reduced activity was observed when a similar mutation was carried out in M.*Hha*I i.e. Ala 280 Phe mutant. The crystallographic data from M.*Hha*I (Cheng *et al.*, 1993) revealed the networks of interactions of surrounding amino acids. From such analysis it became apparent that by altering the Ala 280 Phe, its neighbouring amino acids were perturbed (Figure 3.26 and 3.27) and as a result M.*Hha*I exhibited reduced active. Using a rational approach for mutagenising M.*Hha*I, a series of subsequent amino acids substitution were introduced systematically in order to simulate the properties of motif IX and X of multispecific Mtase. Finally a triple mutant of M.*Hha*I was generated that had regained the enzymatic activity, which was initially lost.

With this active multispecific like M.*Hha*I mutant in hand it was tempting to generate a chimaeric Mtase combining a TRD from M.SPRI. Insertion mutagenesis experiments in the triple mutant of M.*Hha*I were carried out in the variable region of M.*Hha*I by inserting the *Msp*I specific TRD from M.SPRI. In addition, a number of DNA fragments and ORFs were inserted from the variable region and TRDs of multispecific M.SPRI in the triple mutant of M.*Hha*I, in such a way that the reading frame of M.*Hha*I remained intact (chapter 4). In another series of experiment, chimaeras were constructed by the exchange of motifs between wild type M.SPRI and M.*Hha*I. Unfortunately none of these constructs encoded an active enzyme (chapter 5).

During this exercise certain other mutations were also engineered in the sequence of wild type of M.SPRI and M.*Hha*I for the introduction of various restriction sites at different locations. The sequence of MSPRI and M.*Hha*I was aligned and the mutants are summarised in Figure 7.1. In M.*Hha*I, the activity of the double mutant

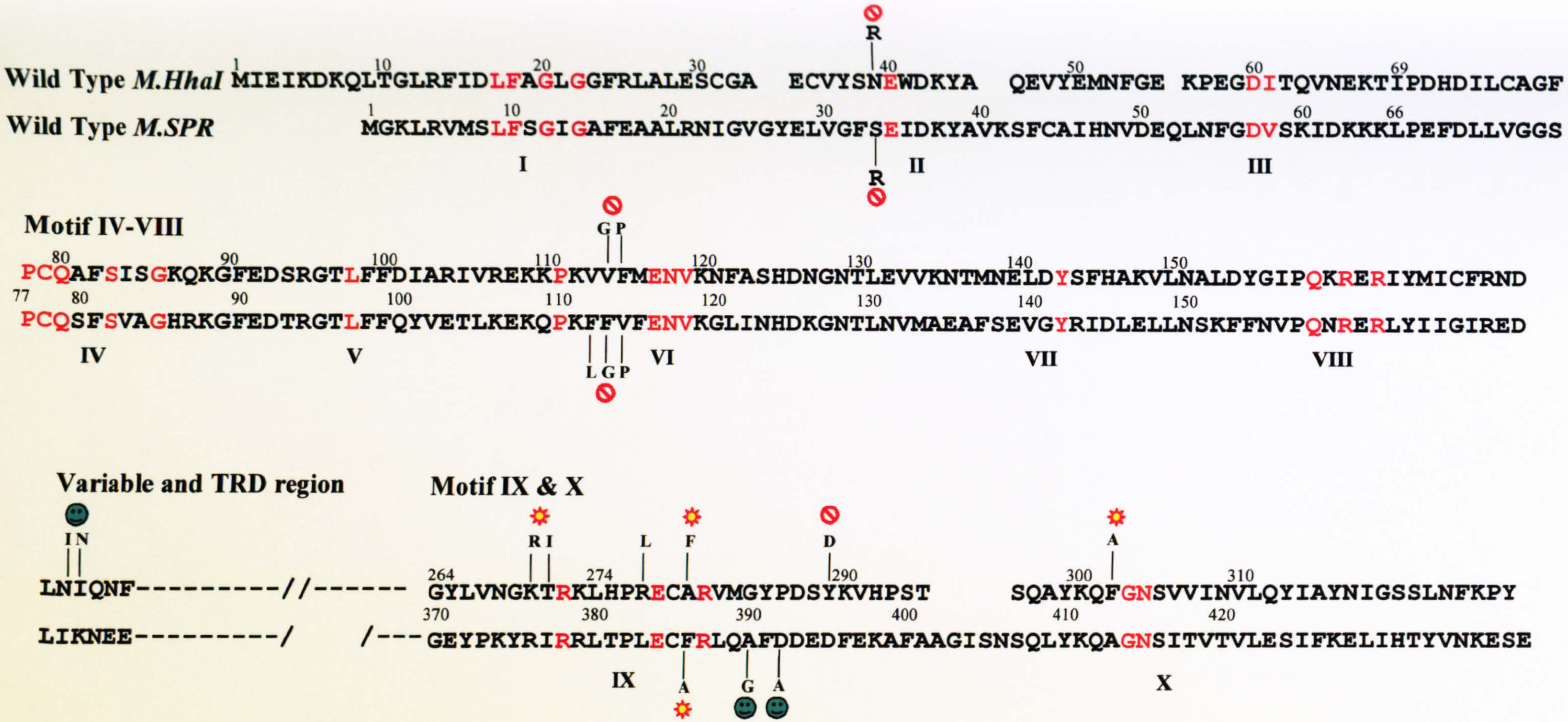


Figure 7.1: Mutants and sequence alignment of M.*Hha*I & M.SPRI. Motifs are numbered with mutations shown above and below at relevant position in both sequences. Conserved consensus sequences are shown in red. Point mutations for either one, two or three codon changes are designated with symbols as follows. Mutations shown in red "NO" symbol eliminate the Mtase activity. Mutations that do not interfere with the activity of enzyme are shown with a "smiley face" symbol. The mutations indicated by a "yellow sun" symbol reduce methylation potential of the encoded protein. Multiple mutations for codon changes at different locations and their methylation potential are described in the relevant text. Most of the TRD region is omitted and is indicated by a dashed line between F180 and G264 of M.*Hha*I and E177 and G370 of M.SPRI.

NI176/177IN was undisturbed. However, the activity of the enzyme encoded by the double mutant KT270/271RI was considerably reduced. The mutation that was introduced into the variable region, illustrated the high flexibility in this region of M.*Hha*I: the latter mutant had reduced activity. Here I should point out the level of reduced activity of different mutants observed in this study. Although protein expression and purification was not carried out for all mutants (as it was not the aim of the study), *in vivo* methylation protection assay revealed that there are different levels of reduced methylation potential which is apparent from close observation of restriction profiles on gels. Generally, point mutations in motif IX and X render M.*Hha*I partially active with the exception of one that makes it inactive (Figure 7.1). In this study M.SPRI mutations in motif IX are tolerated with one exception that makes it partial active (Figure 7.1). However data suggests that generally mutations in this region are not tolerated and produce class I mutants, in which the general methylation capacity of the Mtases is reduced or destroyed (Wilke *et al.*, 1988). Trautner *et al.*, (1996) also indicated that amino acids beyond the C-terminus of the ensemble of TRDs, although not necessary for individual TRD activity, may very well modulate the overall methylation potential of these enzymes. This clearly shows that motif IX and X have more important role in the structure of mono-specific M.*Hha*I.

The basis of losing the methylation potential

The methylation capacity of an enzyme can be disrupted by a single point mutation. This is however, dependent on the position of mutation and its interaction with either the catalytic region or the target recognition domain. Residues that are remote from the active site in the primary structure of a protein may play an important role (Oue *et al.*, 1999).

Mutational analysis of C5 Mtases has revealed that a spectrum of activity exists ranging from fully active to inactive. In between, the term partially active is defined by plasmids that show some resistance to restriction. For any given mutant that exhibits 'partial' activity it is possible that the enzyme is fully active, but rapidly degrades *in vivo,* or that the enzyme turns over at a reduced rate owing to some catalytic or substrate binding defect. A partially active mutant is defined here by the characteristic restriction pattern obtained when the plasmid encoding the mutant is exposed to the cognate restriction enzyme. For example wild type M.SPRI when

expressed in a plasmid, will be refractory to digestion with *Msp*I and *Hae*III, an inactive mutant will fail to methylate the plasmid and digestion with either *Msp*I or *Hae*III leads to complete cleavage at all sites. A partially active mutant either leads to a cleavage at a limited number of sites or due to the reduced methylation potential of mutant protein. It is difficult to determine the reason for partial protection without a detailed analysis of the kinetic properties of the purified mutant enzymes.

In this study it is also observed that the mono-specific Mtase loses catalytic activity due to insertion of large sequences, conflicting with the flexible nature of multi-specific Mtases as observed by Walter *et al.*, (1992). One possible explanation is that the mere enlargement of TRDs of the variable region by insertion of the *Msp*I specific TRD would exceed a limit of TRD tolerated by the variable region and would hence interfere with the proper function of the protein. Alternatively when the *Msp*I specific TRD was inserted, it might have impaired the activity of the neighbouring *Hha*I TRD, as observed in the chimera encoded by pQIS23 and pQIS 24. In another experiment when three TRDs of M.SPRI were inserted, once again the enlargement of the TRD of the variable region takes place. This shows that although the construct encoded by pQIS 18 mimics the multispecific Mtase M.SPRI it is still unable to tolerate extra polypeptide in the variable region.

Two-hybrid system based on an engineered form of M.SPRI

An emerging field for the analysis of biological systems is the study of the complete protein complement of the genome, the 'proteome'. There are several complementary tools available for proteome analysis including 2D protein electrophoresis and mass spectrometry. In addition, two-hybrid technology has also been shown to be helpful in defining protein-protein interactions. These technologies provide a wealth of information that is useful in discovery-based science. However, there are some key limitations of these approaches and new technology is required to be able to fully integrate proteomic information with information obtained about DNA sequence, mRNA profiles and metabolite concentrations into effective models of biological systems.

The principle of two-hybrid technology is based on the interaction of two chimaeric ("hybrid") components (Fields and Song, Nature, 1989). In their version of the system, hybrid 1, "the bait", is composed of a protein fused to a DNA-binding domain that specifically binds to the promoter of a reporter gene. Hybrid 2, "the

prey", is composed of a polypeptide fused to an "activation domain" which recruits the transcription machinery of the yeast cell. If a functional, transcription complex is reconstituted which drives the expression of a reporter gene, a specific gene whose expression (direct two-hybrid methods), or inhibition (reverse two-hybrid methods), reveals the existence of a protein-protein interaction. Using molecular tools and high-throughput screening procedures based on two-hybrid interaction systems, it is possible to identify protein-protein interactions to build protein interaction maps and thereby derive dedicated bioinformatic databases.

Bacterial two-hybrid system

A bacterial two-hybrid system would offer significant advantages over eukaryotic alternatives owing to rapid cell growth rates, efficient expression of heterologous proteins and higher permeability of the cells to small molecules as potential inhibitors of protein-protein interactions. New systems that utilise bacteria for two-hybrid analyses are emerging (Karimova *et al.*, 1998). This genetic test is based on the reconstitution, in an *Escherichia coli cya* strain, of a signal transduction pathway that takes advantage of the positive control exerted by cAMP. Two putative interacting proteins are genetically fused to two complementary fragments, T25 and T18, which constitute the catalytic domain of *Bordetella pertussis* adenylate cyclase. Association of the two-hybrid proteins results in functional complementation between T25 and T18 fragments and leads to cAMP synthesis. Cyclic AMP then triggers transcriptional activation of catabolic operons, such as lactose or maltose that yields a characteristic phenotype.

Benefits of bacterial two-hybrid system

Despite the widespread use of the yeast two-hybrid system, there are reasons why similar genetic assays in *E. coli* would be useful. One advantage of an *E. coli* system is that the speed with which large numbers of interactions can be tested could be increased, a commercially important consideration. *E. coli* grows faster than yeast and, with present methods, can be transformed with higher efficiency, allowing better coverage in library based screens. In addition, since the cloning steps in yeast often involve passing libraries through *E. coli* hosts, using *E. coli would* remove a step from each cycle in a high throughput-screening programme. The smaller genome complexity and greater evolutionary distance from higher eukaryotes mean

that *E. coli* systems should not generate as many false positives and negatives due to interactions between endogenous proteins and eukaryotic baits and preys as *S. cerevisiae* systems do. With respect to drug screening, another potential advantage is that the cell envelope of *E. coli* seems to be somewhat more permeable to small molecules than the cell wall of yeast. Finally, yeast two hybrid systems require nuclear localization of the hybrid proteins and the presence of localization signals for other cellular compartments could prevent bait or prey from reaching the nucleus, a problem that should not occur in *E. coli.*

Leucine zipper proteins

One protein motif whose function can be measured effectively using the Two-Hybrid system is the "leucine zipper". The leucine zipper is a structure that is believed to mediate the function of several eukaryotic gene regulatory proteins. The zipper consists of a periodic repetition of leucine residues at approximately every seventh position, and regions containing them typically span 8 turns of alpha helix (Hurst 1994). Initially it was thought that the Leu residues interdigitated and hence they were named "zippers". Now it is known that the leucine side chains that extend from one helix interact with those from a similar helix, hence facilitating dimerisation in the form of a coiled-coil (Nelson and Cox 2000).

A rapid functional test has been described in this study for the evaluation of protein-protein interactions utilising bacteria instead of yeast. The system is based on the construction of an artificial heterodimeric Mtase M.SPRI, and its *in vivo* assembly through dimerisation as explained in chapter six. By using an *mcr+ E. coli* strain that loses its plasmid-conferred antibiotic resistance in the presence of an active M.SPRI, it is possible to positively identify interacting protein "bait-prey" pairs.

A superficial search through the SWISSPROT protein database reveals at least 22 human genes that are listed as containing "putative" leucine zippers (Table 7.1). There are about 71 human leucine zippers that are "verified" in SWISSPROT, and there are about 9000 human proteins of all types listed in SWISSPROT to date. Therefore out of the total human proteome there may be hundreds of "controversial" leucine zippers that we might be able to verify using out leucine zipper *E. coli* system.

In those cases where the two-hybrid data are borderline, mathematical analysis helped resolve the results after the plasmids had been digested with *Hae*III and

analysed on DHPLC. Therefore, this biochemical analysis that includes the use of the DHPLC may facilitate the measurement of the strength of interaction of the bait and prey proteins.

Table 7.1: List and amino acid sequence of potential leucine zipper domains found in human protein through Swiss-Prot sequence database.

Human Protein	Primary Accession Number	"Putative" leucine zippers
Leucine zipper		LxxxxxxLxxxxxxLxxxxxxLxxxxxxL
ATF4	P18848	LTGECKELEKKNEALKERADSLAKEIQYL
BIK	Q13323	LLALLLLLALLLPLLSGGLHLL
HTF4	Q99081	LYSRDTGLPGCQSSLLRQDLGL
ITF2	P15884	LHSLQSRIEDRLERLDDAIHVL
M3K9	P80192	LRAKEKELRTWEEELTRAALQQKNQ and LRRREQELAEREIDILERELNI
M3KA	Q02779	IQHMFDDLRTKEKELRSREEELLR and LRRREQELAEREMDIVERELHL
MAFG	O15525	LQQEVEKLASENASMKLELDALRSKYEAL
MYCL	P12524	LVGAEKRMATEKRQLRCRQQQLQKRIAYL
MYCN	P04198	LQAEEHQEKEKLQARQQQLLLL
MYC	P01106	LISEEDLLRKRREQLKHKLEQL
NEFA	P80303	LKEYENIIALQENELKKKADELQKQKEELQRQHDQL
NEMO	Q9Y6K9	LAEKKELLQEQLEQLQREYSKL
NPT2	Q06495	LAVGLILLAGSLVLLCTCLILL
NUCB1	Q02818	LRRFEEELAAREAELNAKAQRLSQETEALGRSQGRLEAKKREL
RPB6	P41584	LEGETDPLLIAMKELKARKIPIIIRRYLPDGSYEDWGVDELIITD
SRE1	P36956	LQHSNQKLKQENLSLRTAVHKS
SRE2	Q12772	LQQVNHKLRQENMVLKLANQKN
TFEB	P19484	LENHSRRLEMTNKQLWLRIQEL
USF1	P22415	LQLDNDVLRQQVEDLKNKNLLL
USF2	Q15853	LQMDNELLRQQIEELKNENALL

Pitfalls and potential limitations

One disadvantage of this bacterial two-hybrid system is that the interaction of protein "bait-prey" pairs results in the death of the *E. coli* strain used for the test; in other words this is a "reverse" two-hybrid system. This means that it is not possible to recover the DNA that codes for a successful "prey" for identification. This is particularly important when trying to "fish" for novel "prey" genes from genomic libraries.

Therefore an *E. coli* based system that allows the recovery of "prey" genes in the same way as conventional yeast two-hybrid systems, is needed to complement the reverse M.SPRI system.

7.2: FUTURE WORK

I believe that the bacterial two-hybrid project needs refinement. There is a need to establish those conditions under which well defined interactions (both strong and weak) takes place. In addition a positive selection for homodimerization is desirable. A model has been proposed for making the reverse two-hybrid system a positive one and is summarised in Figure 7.2; this involves obtaining a temperature sensitive mutation in the Mtase. It has been possible to generate a temperature sensitive mutant of M.SPRI using site directed mutagenesis (Trautner, personal communication). In the same manner a temperature sensitive *mcr* system could be generated. A temperature sensitive M.SPRI or *mcr* system of *E. coli* could be used to create a positive bacterial two-hybrid system. Temperature sensitive enzyme is known to be active at 30°C and inactive at 42°C. When colonies from master plate are replica plated and grown at two different temperatures, Mtase being inactive at 42°C, all transformants will grow. However at 30°C only the inactive Mtase-carrying transformants will develop colonies. Therefore missing colonies in 30°C plate are the hits: they carry the interacting partners (proteins), which ultimately bring the two halves of Mtase (N- and C-terminal of M.SPRI) close enough to express Mtase activity. The missing colonies can be amplified from 42°C plate. These hits therefore carry clones with interacting gene products. However, at the moment this system may be adopted as a general tool to identify and characterise various types of known interactions. There is also a need to determine whether this system could be used to identify interacting proteins among an excess of non-interacting ones e.g. screening of libraries etc.

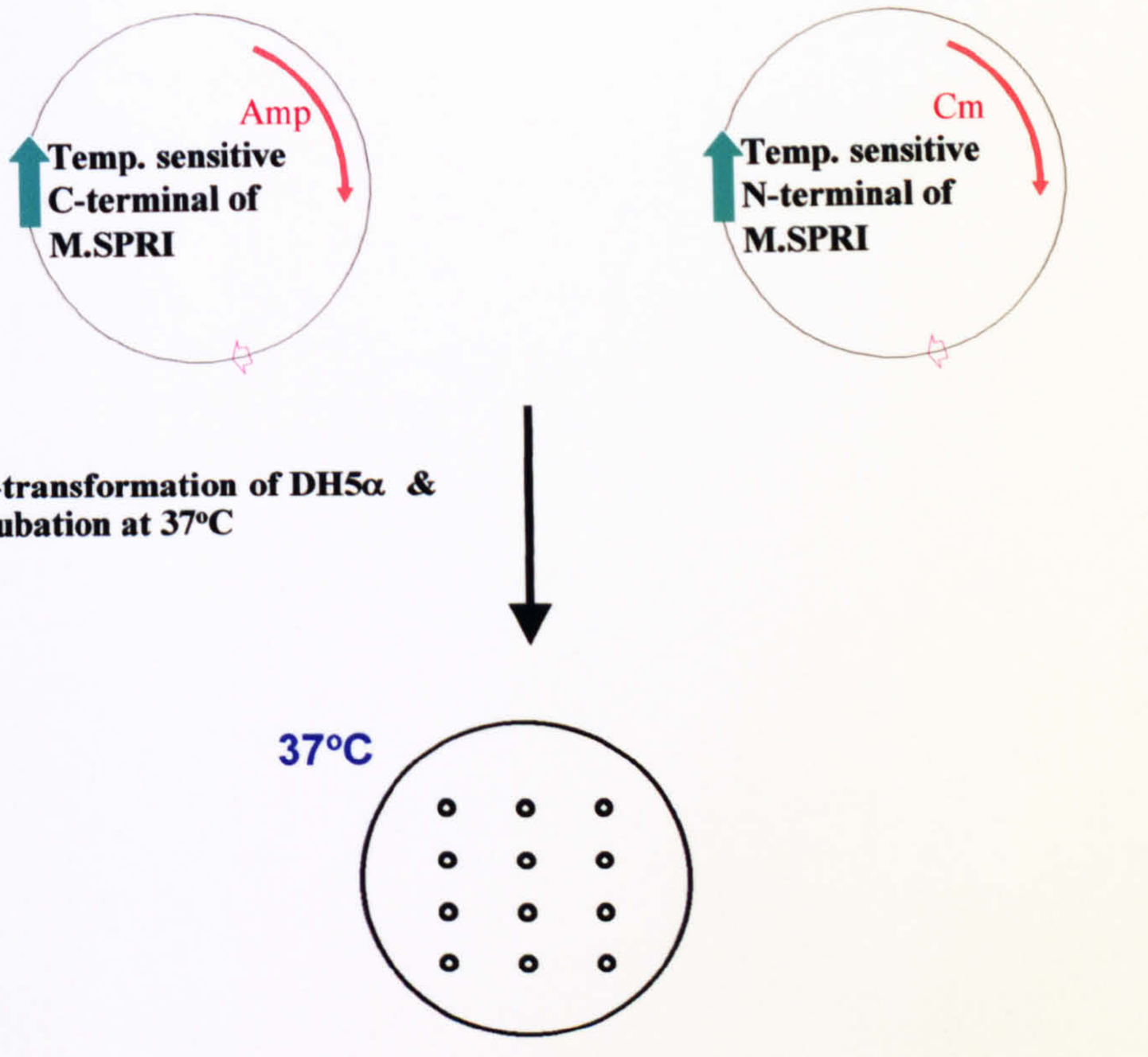


Figure 7.2: A proposed model for the generation of a temperature sensitive M.SPRI, and its use in a two-hybrid experiment. It is possible to generate temperature-sensitive mutants of the M.SPRI using site-directed mutagenesis (Trautner, personal communication). A temperature sensitive M.SPRI could be used to create a two-hybrid system as shown schematically above. Temperature sensitive enzyme M.SPRI is known to be active at 30°C and inactive at 42°C. When colonies from master plate will be allowed to grow at two different temperatures, Mtase being inactive at 42°C, all transformants will grow. However at 30°C only the inactive Mtase carrying transformants will develop colonies. Therefore, missing colonies in 30°C plate are the hits: they carry the interacting partners (proteins) which ultimately bring the two halves of Mtase (N- and C-terminal of M.SPRI) close enough to express Mtase activity. The missing colonies can be amplified from 42°C plate. These hits therefore carry clones with interacting gene products.

REFERENCES

REFERENCES

Adams, R. L.P. (1990) DNA methylation. The effect of minor bases on DNA-protein interactions. *Biochemical Journal* **265**, 309-320.

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

Alberts, B. (1998) The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* **92**, 291-294.

Amir, R. E., Van den Veyver, I. B., Wan, M. Tran, C. Q., Francke, U. and Zoghbi, H. Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics* **23**, 185-188.

Arber, W. (1979) Promotion and limitation of genetic exchange. *Science* **205**, 361-365.

Balganesh, T. S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K., and Trautner, T. A. (1987) Construction and use of chimeric SPR/ϕ3T DNA methyltransferase in the definition of sequence recognising enzyme regions. *The EMBO Journal* **6**, 3543-3549.

Bartel, P. L. and Fields, S. (1997) The yeast two-hybrid system, Oxford University Press Inc, USA; ISBM 0195109384.

Beary, T. P., Braymer, H. D. and Achberger, E. C. (1997) Evidence of participation of $McrB_S$ in McrBC restriction in *Escherichia coli* K-12. *Journal of Bacteriology* **179**, 7768-7775.

Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T. S. and Trautner, T. A. (1987) Organisation of multispecific DNA methyltransferase encoded by temperate *Bacillus subtilis* Phages. *The EMBO Journal* **6**, 1137-1142.

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

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) Cloning and sequencing of a cDNA-encoding DNA methyltransferase of mouse cells - the carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology* **203**, 971-983.

Bickle, T. A. and Kruger, D. H. (1993) Biology of DNA restriction. *Microbiological Reviews* **57**, 434-450.

Blattner, F. R., Plunkett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F.,

Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y. (1997) The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462.

Blundell, T. L. and Srinivasan, N. (1996) Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. *Proc. Natl. Acad. Sci. USA*. **93**, 14243-14248.

Bornberg-Bauer E., Rivals, E. and Vingron, M. (1998) Computational approaches to identify leucine zippers. *Nucleic Acids Research* **26**, 2740-2746.

Boyne, J. R., Yosuf, H. M., Bieganowski, P. Brenner, C. and Price, C. (2000) Yeast myosin light chain, Mlc1p, interacts with both IQGAP and class II myosin to effect cytokinesis. *Journal of Cell Science* **113**, 4533-4543.

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 255-260.

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

Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G. G. and Wells, R. D. (1987) Cloning, sequencing, *in vivo* promoter mapping and expression in *E. coli* of the gene for the *Hha*I methyltransferase. *The Journal of Biological Chemistry* **262**, 4770-4777.

Chang, A. C. Y. and Cohen, S. N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicle derived from the P15A cryptic miniplasmid. *Journal of Bacteriology* **134**, 1141-1156.

Chen, L., Macmillan, A. M., Chang, W., Ezaznikpay, 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.

Chen, L., Macmillan, A. M., Verdine, G. L. (1993) Mutational separation of DNA binding from catalysis in a DNA (cytosine-5)-methyltransferase. *Journal of American Chemical Society* **115**, 5318-5319.

Cheng, X. D. (1995) Structure and function of DNA methyltransferases. *Annual Review of Biophysics and Biomolecular Structure* **24**, 293-318.

Cheng, X. and Blumenthal, R. M. (1996) Finding a basis for flipping bases. *Structure* **4**, 639-645.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R. J. (1993) Crystal structure of the *Hha*I DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**, 299-307.

Colas, P. and Brent, R. (1998) The impact of two-hybrid and related methods on biotechnology. *Trends in Biotechnology* **16**, 355-363.

Colot, V. and Rossingol, J. (1999) Eukaryotic DNA methylation as an evolutionary device. *BioEssays* **21**, 402-411.

Coombs, G. S. and Corey, D. R. (1998) Chapter 4; Site-directed mutagenesis and Protein engineering. Edited by Angeletti, R. H., *Protein analysis and design.* Published by Academic Press, USA. p 260.

Corcoran, E. E. and Means, A. R. (2001) Defining Ca^2+/Calmodulin-dependent protein kinase cascades in transcriptional regulation. *The Journal of Biological Chemistry* **276**, 2975-2978.

Davis, J., Thompson, D. and Beckler, G. S. (1998) Large-scale dialysis reactions using *E. coli* S30 extract systems. *Promega Notes, magazine number* **56**, p 14.

De Bruijn, F. J., Lupski, J. R. and Weinstock, G. M. (1998) *Bacterial Genomes, Physical structure and Analysis.* (Chapter 8, Raleigh, E. A. and Brooks J. E.) Published by Chapman & Hall, Page 78-92.

Drees, B. L. (1999) Progress and variations in two-hybrid and three-hybrid technologies. *Current Opinion in Chemical Biology* **3**, 64-70.

Feil, R., Charlton, J., Bird, A. P., Walter, J. and Reik, W. (1994) Methylation analysis on individual chromosomes: Improved protocol for bisulphate genomic sequencing. *Nucleic Acids Research* **21**, 695-696.

Field, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246.

Finnegan, E. J., Peacock, W. J. and Dennis, E. S. (2000) DNA methylation, a key regulator of plant development and other processes. *Current opinion in Genetics and Development* **10**, 217-223.

Frangioni, J. V. and Neel, B. G. (1993) Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* **210**, 179-187.

Friedman, S. (1985) The irreversible binding of azacytosine-containing DNA fragments to bacterial DNA (cytosine-5) methyltransferases. *J. Biological Chemistry* **260**, 5698-5705.

Friedman, S. and Ansari, N. (1992) Binding of the *Eco*RII methyltransferase to 5-fluorocytosine-containing DNA. Isolation of a bound peptide. *Nucleic Acids Research* **20**, 3241-3248.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molly, P. L. and Paul, C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA*. **89**, 1827-1831.

Fu, J., Ren, M., and Kreibich, G. (1997) Interactions among subunits of the oligosaccharyltransferase complex. *The Journal of Biological Chemistry*. **272**, 29687-29692.

Gentz, R., Rauscher III, F. J., Abate, C. Curran, T. (1989) Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243**, 1695-1699.

Gietz, R. D., Triggs, R. B., Robbins, A., Graham, K. C. and Woods, R. A. (1997) Identification of proteins that interact with a protein of interest: Applications of the yeast two-hybrid system. *Journal Molecular and Cellular Biochemistry* **172**, 67-79.

Gilbert, S. F. (1994) Developmental Biology, Published by Sinauer Associates, inc. Sunderland, Massachusettes, USA. (4th edition) Page 423.

Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. and Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase M.*Taq*I in complex with DNA and a cofactor analog. *Nature Structural Biology* **8**, 121-125.

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

Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997) Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acid Research* **25**, 2702-2715.

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

Hanahan, D. (1985) Techniques for transformation of E. Coli, in DNA Cloning Vol I, A Practical Approach, Glover, D. M. (ed) pp 109-130, IRL Press, Oxford, UK.

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K. Weemaes, C. M. R. and Gartler, S. M. (1999) The DNMT3B DNA methyltransferase

gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. USA*. **96**, 14412-14417.

Heinz, D. W., Baase, W. A., Zhang, X. J., Blaber, M., Dahlquist, F. W. and Matthews, B. W. (1994) Accommodation of amino acid insertions in an alpha-helix of T4 lysozyme, structural and thermodynamic analysis. *Journal of Molecular Biology* **236**, 869-886.

Heitman, J. (1993) On the origins, Structure and function of Restriction and modification enzymes. Genetic engineering, Vol. 15, pp 57-108, Plenum Press New York.

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

Hodgkin, J. (2000) Genome sequencing: A view of Mount *Drosophila*. *Nature* **404**, 442–443.

Hsieh, C. L. (2000) Dynamics of DNA methylation pattern. *Current opinion in Genetics and Development*. **10**, 224-228

Hu, J. C., Kornacker, M. G., and Hochschild, A. (2000) *Escherichia coli* One- and Two-hybrid systems for the analysis and identification of protein-protein interactions. *Methods* **20**, 80-94.

International Human Genome Sequencing Consortium*. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-922 (*See a partial list of contributors and groups for this project at http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v409/n6822/full/409860a0_fs.html).

Johnston, K., Clements, A., Venkataramani, R. N., Trievel, R. C. and Marmorstein, R. (2000) Co expression of proteins in bacteria using T7-based expression plasmids: Expression of heteromeric cell-cycle and transcriptional regulatory complexes. *Protein expression and purification* **20**, 435-443.

Kan, M. S. (1999) Mutational analysis of M.*Hha*I to mimic M.*Spo*I from *Schizosaccharomyces pombe* and MascI from *Ascobolus immersus*. PhD. Thesis, University of Sheffield.

Karimova, G. Pidoux, J. Ullmann, A. and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. USA*. **95**, 5752-5756.

Karlsson, R. and Stahlberg, R. (1995) Surface plasmon resonance detection and multispot sensing for direct monitoring of interactions involving low-molecular weight analytes and for determination of low affinities. *Analytical Biochemistry* **228**, 274-280.

Karreman, C., and Waard, A. (1990) *Agmenellum quadruplicatum* M.*Aqu*I, a novel modification methylase. *Journal of Bacteriology* **172**, 266-272.

Kelleher, D. J. and Gilmore, R. (1994) The *Saccharomyces cerevisiae* oligosaccharyltransferase is a protein complex composed of Wbp1p, Swp1p, and four additional polypeptides. *The Journal of Biological Chemistry*. **269**, 12908-12917.

Kellehher, J. E. and Raleigh, E. A. (1991) A novel activity in *Escherichia coli* K-12 that directs restriction of DNA modified at CG dinucleotides. *Journal of Bacteriology* **173**, 5220-5223.

Kerppola, T. K. and Curran, T. (1991) Transcription factor interactions: basics on zippers. *Current Opinion in structural Biology* **1**, 71-79.

Klimašauskas, S., Kumar, S., Roberts, R. J. and Cheng, X. (1994) *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369.

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

Kong, H., Lin, L. F., Porter, N., Stickel, S., Byrd, D., Posfai, J. and Roberts, R. J. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acid Research* **28**, 3216-3223.

Kouzarides, T. and Ziff, E. (1988) The role of the leucine zipper in the fos-jun interaction. *Nature* **336**, 646-651.

Kühnlein, U. and Arber, W. (1972) Host specificity of DNA produced by *E. coli* XV. The role of nucleotide methylation in invitro B-specific modification. *Journal of Molecular Biology* **63**, 9-19.

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

Kumar, S., Cheng, X., Klimašauskas, S., Sha Mi, Posfai, J., Roberts, R. J., and Wilson, G. G., (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Research* **22**, 1-10.

Labahn, J., Granzin, J., Schlukebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I., and Saenger, W. (1994) 3-dimensional structure of the adenine- specific DNA methyltransferase M..*Taq*I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. USA*. **91**, 10957-10961.

Lacks, S. and Greenberg, B. (1977) Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *Journal of Molecular Biology* **114**, 153-168.

Laird P.W. and Jaenisch R., (1996) The role of DNA methylation in cancer genetics and epigenetics. *Ann. Rev. Genet.* **30**, 441-464.

Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988) The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764.

Lange, C., Noyer-Weidner, M., Trautner, T. A., Weiner, M. and Zahler, S. A. (1991) M.H21, a multispecific 5C-DNA methyltransferase encoded by *Bacillus amyloliquefaciens* phage H2. *Gene* **100**, 213-218.

Lange, C., Wild, C., and Trautner, T. A., (1996) Identification of a subdomain within DNA-(cytosine-C5)-methyltransferases responsible for the recognition of the 5' part of their DNA target. *The EMBO Journal* **15**, 1443-1450.

Lauster, R., Trautner, T. A. and Noyer-Weidner, M. (1989) Cytosine-specific type II DNA methyltransferase. A conserved enzyme core with variable target-recognising domains. *Journal of Molecular Biology* **206**, 305-312.

Leonhardt, H. and Bestor, T. H. (1993). *DNA Methylation: Molecular Biology and Biological Significance* (Jost, J. P. and Saluz, H. P., Eds.) First edition. Birkhauser Verlag Basel, Switzerland. 109-119.

Lee, K., Kam, K., and Shaw, P., (1995) A bacterial methyltransferase M.*Eco*HK31I requires two proteins for *in vitro* methylation. *Nucleic Acids Research* **23**, 103-108.

Li, E, Bestor, T. H. and Jaenisch R. (1992) Targeted mutation of the DNA methyltransferase gene result in embryonic lethality. *Cell* **69**, 915-926.

Luria, S.E. and Human, M. L. (1952) A nonhereditary, host induced variation of bacterial viruses. *Journal of Bacteriology* **64**, 557-569.

Margot, J. B., Aguirre-Arteta, A. M., Di Giacco, B. V., Pradhan, S., Roberts, R. J., Cardoso, M. C. and Leonhardt, H. (2000) Structure and function of the mouse DNA methyltransferase gene: *Dnmt1* shows a tripartite structure. *Journal of Molecular Biology* **297**, 293-300.

Matin, M. M. (2000) Analysis and application of mutant C5-DNA methyltransferase. PhD. Thesis, University of Sheffield.

Matthews, B. W. (1987) Genetic and structural analysis of the protein stability problem. *Biochemistry* **26**, 6885-6888.

Maxam, A. M. and Gilbert, W. (1980) Sequencing end-labelled DNA with bases-specific chemical cleavages. *Methods in Enzymology* **65**, 499-560.

Meador, W. E., Means, A. R., Quiocho, F. A. (1992) Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science* **257**, 1251.

Mi, S. and Roberts, R. J. (1992) How M.*Msp*I and M.*Hpa*II decide which base to methylate. *Nucleic Acids Research* **20**, 4811-4816.

Modrich, P (1997) Strand-specific mismatch repair in mammalian cells. *The Journal of Biological Chemistry* **272**, 24727-24730.

Mostoslavsky, R. and Bergman, Y. (1997) DNA methylation: regulation of gene expression and role in the immune system. *Biochimica et Biophysica Acta* **1333**, F29-F50.

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.

Naito, Y., Naito. T. and Kobayashi, I. (1998) Selfish restriction modification genes: resistance of resident R/M plasmid to displacement by an incompatible plasmid mediated by host killing. *Biological Chemistry* **379**, 429-436.

Nelson, D. L. and Cox, M. M. (2000) Lehninger Principles of Biochemistry. (Chapter 28, Regulation of gene expression) 3rd Ed. Worth Publishers, 41 Madison Avenue, New York, NY10010 p:1084.

Nieba, Lars., Nieba-Axmann, S. E., Persson, A., Hämäläinen, M., Edebratt, F., Hansson, A., Lidholm, J., Magnusson, K., Karlsson, Å. F., and Plückthun, A. (1997) BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip, *Analytical Biochemistry* **252**, 217-228.

Noyer-Weidner, and Trautner, T. A. (1993) DNA methylation: Molecular Biology and Biological Significance (Jost, J. P. and Saluz, H. P., Eds.) First edition, Birkhauser Verlag Basel, Switzerland. 39-108.

Noyer-Weidner, M., Diaz, R. and Reiners, L. (1986) Cytosine-specific DNA modification interferes with plasmid establishment in *Escherichia coli* K 12: involvement of rglB. *Molecular and General Genetics* **205**, 469-475.

O'Gara, M., Horton, J. R., Roberts, R. J. and Cheng, X. (1998) Structure of *Hha*I methyltransferase complexed with substrates containing mismatches at the target base. *Nature Structural Biology* **5**, 872-877.

O'Neill, 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 loss 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.C., Matsuda, A., and Santi, D.V. (1988) 5-Fluorocytosine in DNA is a mechanism-based inhibitor of *Hha*I methylase. *Biochemistry* **27**, 5204-5210.

Oue, S., Okamoto, A., Yano, T., and Kagamiyama, H. (1999) Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *The Journal of Biological Chemistry.* **274**, 2344-2349.

Ohmori, H., Tomizawa, J. I. And Maxam, A. M. (1978) Detection of 5-methylcytosine in DNA sequences. *Nucleic Acids Research* **5**, 1479-1485.

Panne, D., Raleigh, E. A. and Bickle, T. A. (1998) McrB$_S$, a modulator peptide for McrBC activity. *The EMBO Journal* **17**, 5477-5483.

Phizicky, E. M. and Fields, S. (1995) Protein-protein interactions: Methods for detection and analysis. *Microbiological Reviews* **59**, 94-123.

Posfai, G., Baldauf, F., Erdei, S., Posfai, J., Venetianer, P., and Kiss, A. (1984) Structure of the gene coding for the sequence-specific DNA-methyltransferase of the *B. subtilis* phage SPR. *Nucleic Acids Research* **12**, 9039-9049.

Posfai, G., Kim, S. C., Szilak, L., Kovacs, A. and Venetianer, P., (1991) Complementation by detached parts of GGCC-specific DNA methyltransferase. *Nucleic Acids Research* **19**, 4843-4847.

Pósfai, J., Bhagwat, A. S., Posfai, G. and Roberts, R. J. (1989) Predictive motifs derived from cytosine methyltransferase. *Nucleic Acids Research* **17**, 2421-2435.

Posfai, J., Bhagwat, A. S. and Roberts, R. J. (1988) Sequence motifs specific for cytosine methyltransferases *Gene* **74**, 261-265.

Raleigh, E. A. (1987) Restriction and modification in vivo by *Escherichia coli* K-12. *Methods in Enzymology* **152**, 130-141.

Raleigh, E. A. (1992) Organisation and function of the mcrBC genes of *Escherichia coli* K-12. *Molecular Microbiology* **6**, 1079-1086.

Raleigh, E. A., Benner, J., Bloom, F., Braymer, H. D., DeCruz, E., Dharmalingam, K., Heitman, J., Noyer, W. M., Piekarowicz, A., and Kretz, P. L. (1991) Nomenclature relating to restriction of modified DNA in *Escherichia coli. J. Bacteriol.* **173**, 2707-2709.

Raleigh, E. A. and Wilson, G. (1986) *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. USA*. **83**, 9070-9074.

Raleigh, E. A., Trimarchi, R. and Ravel, H. (1989) Genetic and physical mapping of the mcrA (rglA) and mcrB (rglB) loci of *Escherichia coli* K-12. *Genetics* **122**, 279-296.

Razin, A. (1998) CpG methylation, chromatin structure and gene silencing-a three-way connection. *The EMBO Journal* **17**, 4905-4908.

Razin, A., Cedar, H. and Riggs, A. D., eds (1984) DNA methylation; Biochemistry and biological significance. Published by Springer-Verlag New York Inc.

Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1995) The crystal structure of *Hae*III methyltransferase covalently complexed to DNA. *Cell* **82**, 143-153.

Roberts, R. J. (1995) On base flipping. *Cell* **82**, 9-12.

Roberts, R. J. and Cheng, X. (1998) Base flipping. *Annu. Rev. Biochem.* **67**, 181-198.

Roberts, R. J. and Macelis, D. (1999) REBASE-restriction enzymes and methylases. *Nucleic Acids Research* **27**, 312-313.

Robertson, K. and Wolffe, A. (2000) DNA methylation in health and disease. *Nature Genetics*, **1**, 11-19.

Robertson, K. D and Jones, P. A. (2000) DNA methylation: past, present and future directions. *Carcinogenesis* **21**, 461-467.

Ross, T. K., Achberger, E. C. and Braymer, H. D. (1987) Characterisation of the *Escherichia coli* modified cytosine restriction (mcrB) gene. *Gene* **61**, 277-289.

Russell, D. W. and Zinder, N. D. (1987) Hemi-methylation prevents DNA replication in *E. coli*. *Cell* **50**, 1071-1079.

Sambrook, J., Fritsch, E. R. & Maniatis, T. (1989) *Molecular cloning: A Laboratory Manual*. Second Edition. Cold Spring Harbor Press, N.Y.

Santi, D. V., Garrett, C. E. and Barr, P. J. (1983) On the mechanism of inhibition of DNA- cytosine methyltransferase by cytosine analogs. *Cell* **33**, 9-10

Santi, D. V., Norment, A. and Garrett, C. E. (1984) Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc. Natl. Acad. Sci. USA* **81**, 6993-6997.

Schuermann, M., Neuberg, M., Hunter, J. B., Jenuwein, T., Ryseck, R., Bravo, R. and Müller, R. (1989) The leucine repeat motif in fos protein mediates complex formation with Jun/AP-1 and is required for transformation. *Cell* **56**, 507-516.

Schumann, J., Walter, J., Willert, J., Wild, C., Koch, D. and Trautner, T. A. (1996) M.BssHII, a multispecific cytosine-C5-DNA-methyltransferase with unusual target recognising properties. *Journal of Molecular Biology* **257**, 949-959.

Sethmann, S., Ceglowski, P., Willert, J., Iwanicka-Nowicka, R., Trautner, T. A., and Walter, J. (1999) M.(ϕ)*Bss*HII, a novel cytosine-C5-DNA-methyltransferase with target-recognising domains at separated locations of the enzyme. *The EMBO Journal* **18**, 3502-3508.

Shannon, K. B. and Li, R. (2000) A myosin light chain mediates the localization of the budding yeast IQGAP-like protein during contractile ring formation. *Current Biology* **10**, 727-730.

Sheikhnejad, G., Brank, A., Christman, J. K., Goddard, A., Alvarez, E., Ford Jr.H., Marquez, V. E., Marasco, C. J., Surfin, J. R., O'Gara, M. and Cheng, X. (1999) Mechanism of inhibition of DNA (Cytosine C5)-Methyltransferases by oligodeoxyribonucleotides containing 5,6-Dihydro-5-azacytosine. *Journal of Molecular Biology* **285**, 2021-2034.

Smith, D. B. and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione s-transferase. *Gene* **67**, 31-40.

Smith, S. S., Kaplan, B. E., Sower, 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

Stryer, L. (1995) Biochemistry. 4th Edition. Published by W. H. Freeman and Company, New York. Page 222.

The *C. elegans* Sequencing Consortium*. (1998) Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* **282**, 2012-2018 (*See http://www.sanger.ac.uk/Projects/C_elegans/ for a list of authors).

Tran-Betcke, A., Behrens, B., Noyer-Weidner, M., Trautner, T. A. (1986) DNA methyltransferase gene of *Bacillus subtillus* phages: comparison of their nucleotide sequences. *Gene* **42**, 89-96.

Trautner, T. A., Balganesh, T. S., Pawlek, B. (1988) Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Research* **16**, 6649-6658.

Trautner, T.A., Balganesh, T., Wilke, K., Noyer-Weidner, M., Rauhut, E., Lauster, R., Behrens, B., Pawlek, B. (1988) Organization of target-recognizing domains in the multispecific DNA (cytosine-5) methyltransferases of *Bacillus subtilis* phages SPR and Phi-3T. *Gene* **74**, 267.

Trautner, T. A., Pawlek, B., Behrens, B., and Willert, J. (1996) Exact size and organisation of DNA target-recognising domains of multispecific DNA-(cytosine-C5)-methyltransferases. *The EMBO Journal* **15**, 1434-1442.

Trautner, T. A., Pawlek, B., Gunthert, U., Canosi, U., Jentsch, S. and Freund, M. (1980) Restriction and modification in *Bacillus subtilis*: identification of a gene in the temperate phage SPβ coding for a *Bsu*R specific modification methyltransferase. *Mol. Gen. Genet.* **180**, 361-367.

Uetz, P., Giot, L., Cagney, G. Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan,V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Rothberg, J. M. and Fields, S. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623 – 627.

Vilkaitis, G. Dong, A. Weinhold, E. Cheng, X. Klimasauskas, S. (2000) Functional roles of the conserved threonine-250 in the target recognition domain of *Hha*I DNA methyltransferase. *The Journal of Biological Chemistry* **275**, 38722-38730.

Waite-Rees, P. A., Keating, C. J., Moran, L.S., Slatko, B. E., Hornstra, L. J. and Benner, J. S. (1991) Characterisation and expression of the *Escherichia coli* Mrr restriction system. *Journal of Bacteriology* **173**, 5207-5219.

Walter, J., Trautner, T. A., and Noyer-Weidner, M. (1992) High plasticity of multispecific DNA methyltransferases in the region carrying DNA target recognising enzyme modules. *The EMBO Journal* **11**, 4445-4450.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., and Trautner, T. A., (1988) Sequential order of target-recognising domains in multispecific DNA-methyltransferases. *The EMBO Journal* **7**, 2601-2609.

Wilson, G. G. (1991) Organisation of restriction-modification systems. *Nucleic Acids Research* **10**, 2539-2566.

Wilson, G. G. and Murray, N. E. (1991) Restriction and modification systems. *Ann. Rev. Genet.* **25**, 585-627.

Wolffe, A. P. (1998) Packaging principle: How DNA methylation and histone acetylation control the transcriptional activity of chromatin. *The Journal of Experimental Zoology* **282**, 239-244.

Woodcock, D. M., Adams, J. K., Allan, R. G. and Cooper, I. A. (1983) Effect of several inhibitors of enzymatic DNA methylation on the *in vivo* methylation of different classes of DNA sequences in a cultured human cell line. *Nucleic Acids Research* **11**, 489-499.

Xiong, Z. and Laird, P. W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Research* **25**, 2532-2534.

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J. and Pequignot, E. V. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187-191.

Xu, S. Y., Xiao, J-P, Posfai, J. Maunus, R. and Benner II, J. (1997) Cloning of the *Bss*HII restriction-modification system in *Escherichia coli*: *Bss*HII methyltransferase contains circularly permuted cytosine-5 methyltransferase motifs. *Nucleic Acids Research* **25**, 3991-3994.

Zeng, X., Herndon, A. M. and Hu, J. C. (1997) Buried asparagines determine the dimerization specificities of leucine zipper mutants. *Proc. Natl. Acad. Sci. USA.* **94**, 3673-3678.

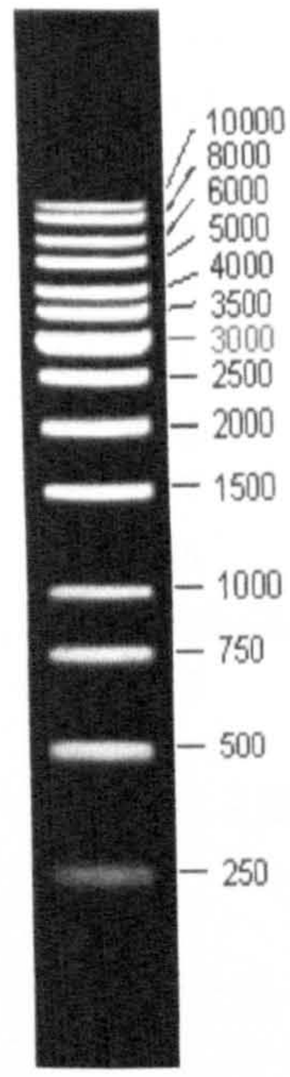
Zubay, G. (1973) *In-vitro* synthesis of protein in microbial systems. *Ann. Rev. Genet.* 7, 267.

APPENDICES

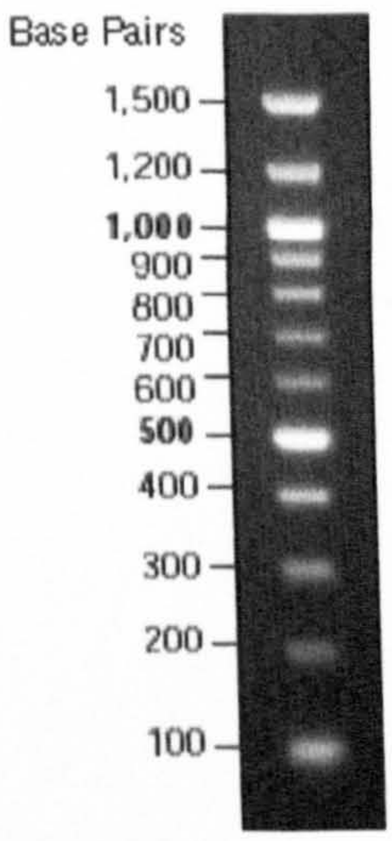
Appendix I

GeneRuler™ 1 kb and 100 bp DNA LADDER

Fragment	Base Pairs	DNA
1	10,000	20 ng
2	8,000	24 ng
4	6,000	30 ng
5	5,000	30 ng
6	4,000	40 ng
7	3,500	51 ng
8	3,000	154.5 ng
9	2,500	33.5 ng
10	2,000	35 ng
11	1,500	24.5 ng
12	1,000	10 ng
12	750	16.5 ng
13	500	14.5 ng
14	250	16.5 ng



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

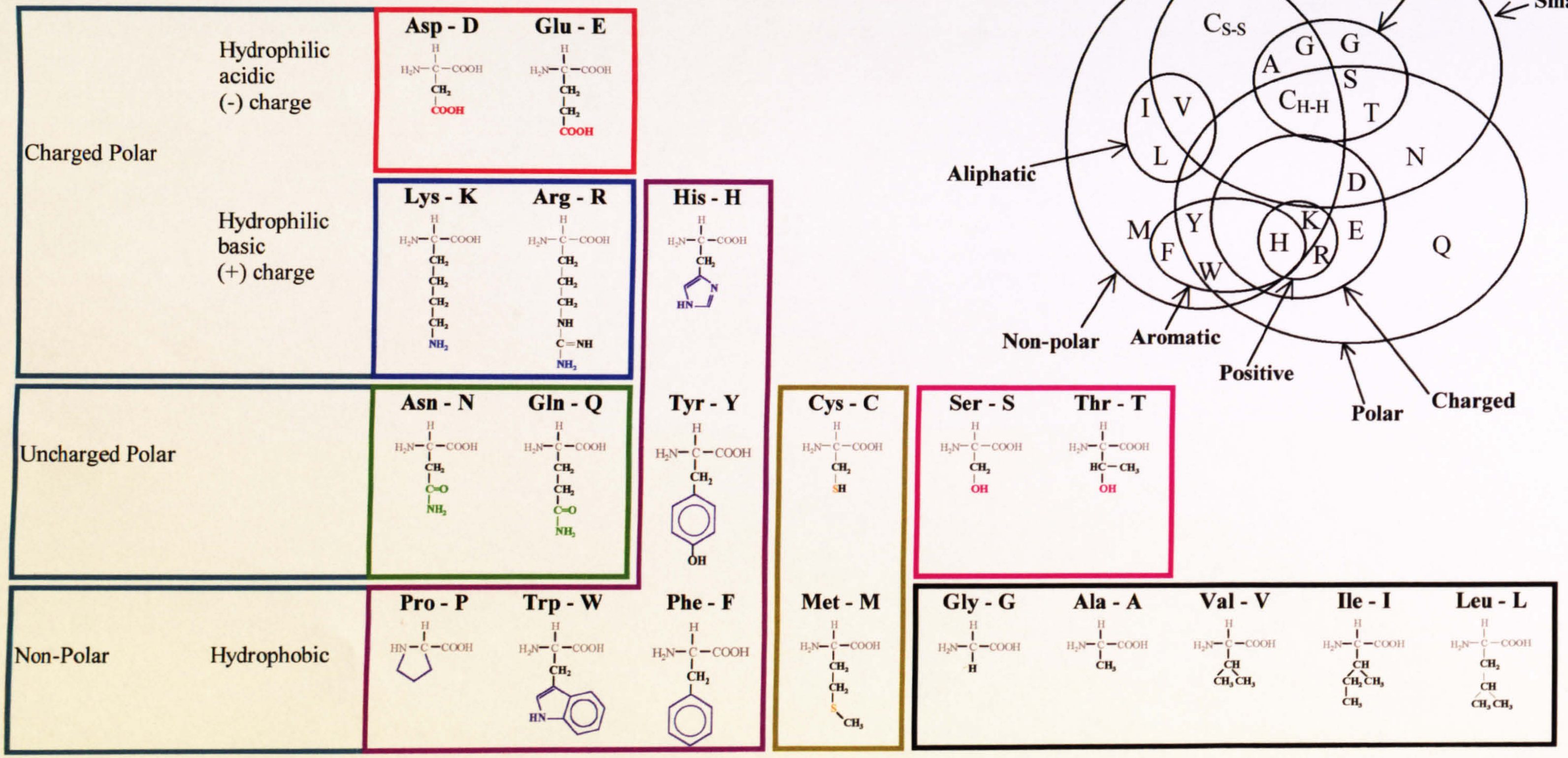


Electrophoretic molecular weight standards for DNA fragments separated on agarose.
(a) GeneRuler™ 1kb DNA Ladder (MBI) (14 fragments) with 0.5µg loading on 1.0% agarose gel and the (b) 100 bp ladder (MBI) (12 fragments) analysed on 1.8.0% agarose gel. The approximate DNA mass has also been tabulated above for each fragment. (Data adapted from MBI Fermentas).

Appendix II
PROTEIN BROAD RANGE MARKER

Fragment	Protein	Source	Calculated MW (Da)
1	Myosin	rabbit muscle	212,000
2	MBP-β-galactosidase	*E. coli*	158,194
3	b-galactosidase	*E. coli*	116,351
4	Phosphorylase b	rabbit muscle	97,184
5	Serum albumin	bovine	66,409
6	Glutamic dehydrogenase	bovine liver	55,561
7	MBP2	*E. coli*	42,710
8	Lactose dehydrogenase	porcine muscle	36,487
9	Triose phosphate isomerase	rabbit muscle	26,625
10	Trypsin inhibitor	soybean	(20,040-20,167)
11	Lysozyme	chicken egg white	14,313
12	Aprotinin	bovine lung	6,517
13	Insulin A, B chain	bovine pancreas	(2,340-3,400)

It is a mixture of purified proteins with known amino acid sequences. They are resolved to 13 sharp bands when analysed by SDS-PAGE and stained with Coomassie Blue R-250. Protein bands numbered 1-10 are visible in this gel. Two bands (BSA, MW 66.4 kDa and triphosphate isomerase, MW 26.6 kDA) are at double intensity to serve as reference point. MBP is abbreviation for maltose binding protein (NEB, 1999 Catalogue).



Structure of the 20 amino acids commonly found in proteins, arranged in groups with similar chemical properties of the side chains. (□) carboxylic group; (□) amino group; (□) amide group (□) ring/phenolic group; (□) thiol group; (□) hydroxylic group (hydrophilic polar); and (□) aliphatic group.

Appendix IV

Sequences of proteins used for testing dimer formation in two hybrid experiments

Plasmids	Sequence of test protein
pQIS 84 (pSPR-LZ) No of amino acids: 33 Molecular weight: 4047.5	**LE**LTSDNDRLRKRVEQLSRELDTLRGIFRQL**EF**
Plasmids carrying C-terminal region sequence of M.SPR (motif IX-X)	
pQIS 98 or pMETC (LZ) No of amino acids: 33 Molecular weight: 4047.5	**LE**LTSDNDRLRKRVEQLSRELDTLRGIFRQL**EF**
pQIS 114R or pMETC (Calmodulin) No of amino acids: 151 Molecular weight: 17098.9	**LE**MADQLTEEEIAEFKEAFSLFDKDGDGTITTKELG TVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEF LTMMARKMKDTDSEEIREAFRVFDKDGNGYISAAE LRHVMTNLGEKLTDEEVDMIREADIDGDGQVNYEE FVQMMTAK**EF**
pMETC (N-terminal fragmentof IqgI) yeast No of amino acids 861 Molecular weight: 98744.6	**LE**GSMTAYSGSPSKPGNNNSYLNRYVENLGTNVTP PLRPQSSSKINSSLNIASPSHLKTKTSASNSSATILSK KVESSVSKLKPSLPNKLVGKYTVDLSNYSKIELRYY EFLCRVSEVKIWIEAVIEEALPSEIELCVGDSLRNGV FLAKLTQRINPDLTTVIFPAGDKLQFKHTQNINAFFG LVEHVGVPDSFRFELQDLYNKKNIPQVFETLHILISM INKKWPGKTPALTNVSGQISFTKEEIAACKKAWPRI RDFKSLGTNINTAPASPEEPKEKRSGLIKDFNKFERP NIPVEEILITPRKNITDANCSDFSNTPSPYNEAPKMSN LDVVVEKRKFTPIEPSLLGPTPSLEYSPIKNKSLSYYS PTISKYLTYDTEFYTRRSRAREEDLNYYQTFKYSPS HYSPMRRERMTEEQFLEKVVQLQNICRGVNTRFNL YIQKRLLNLFEQDILRFQACLRGNKFRVLSSMYLPIR RAKIDVPHVEAIQSRIKGSRIRYKYDKLKFTLSRFSC TVELLQAYCRSKLLKTTVNTKLNDIEISHYPLTKLQ SYMRASYVRKKVMSLNTKLNDERESIMKFSAIIRG NVVRCSEDAILSAVHDVHKENISKLQSLIRGIFTRSC LASIIYSLGKENCNIIQLSACIRGNAVRHKVQSLFAP ENNLSETVHDLQGLVRGILVRYTLDLVDDIVEYNN LALFQAFSRGALVRESLDQKSSFYKRNVRSVIMIQS WIRKSLQRSAYLELLDCPNPSLWAVKKFVHLLNGT ATIEEVQNQLESCQASLDSENMKKERLLKSIRQQLN INGVLDKFGLLKDKDHELGISDSTIPKSKYQKYEKL FYMLQVDPSYWKLLYLKEPEVD**EF**
pQIS 117 or pMETC (LZ-Leu 8 Pro) No of amino acids: 33 Molecular weight: 4031.5	**LE**LTSDNDRPRKRVEQLSRELDTLRGIFRQL**EF**

pQIS 118 or pMETC (WBP1) No of amino acids: 435 Molecular weight: 49966.3	**LE**MRTDWNFFFCILLQAIFVVGTQTSRTLVLYDQST EPLEEYSVYLKDLEQRNYKLEYLDINSTSTTVDLYD KEQRLFDNIIVFPTKGGKNLARQIPVKQLIKFFENEG NILCMSSPGAVPNTIRLFLNELGIYPSPKGHVIRDYF SPSSEELVVSSNHLLNKYVYNARKSEDFVFGESSAA LLENEQIVPILNAPRTSFTESKGKCNSWTSGSQGFLV VGFQNLNNARLVWIGSSDFLKNKNQDSNQEFAKEL LKWTFNEKSVIKSVHAVHSHADGTSYDEEPYKIKD KVIYSVGFSEWNGEEWLPHIADDIQFELRQVDPYYR LTLSPSGNDSETQYYTTGEFILPDRHGVFTFLTDYR KIGLSFTTDKDVKAIRHLANDEYPRSWEISNSWVYI SAICGVIVAWIFFVVSFVTTSSVGKKLETFKKTNQF **RG**	
Plasmids carrying N-terminal region sequence of M.SPR (motif I-VIII)	
pQIS 101N or pMETN (LZ) No of amino acids: 33 Molecular weight: 4047.5	**LE**LTSDNDRLRKRVEQLSRELDTLRGIFRQL**EF**
pMETN (MLC1 yeast gene) No of amino acids: 153 Molecular weight: 16963	**LE**MSATRANKDIFTLFDKKGQGAIAKDSLGDYLRA IGYNPTNQLVQDIINADSSLRDASSLTLDQITGLIEV NEKELDATTKAKTEDFVKAFQVFDKESTGKVSVGD LRYMLTGLGEKLTDAEVDELLKGVEVDSNGEIDYK KFIEDVLRQ**EF**
pQIS 113 or pMETN (Myosin Light Chain Kinase peptide (MLCK) for Calmodulin) No of amino acids: 30 Molecular weight: 3553.2	**LE**KYMARRKWQKTGHAVRAIGRLSSMAM**EF**
pQIS 115 or pMETN (LZ-Asn 6 Lys) No of amino acids: 33 Molecular weight: 4047.5	**LE**LTSDKDRLRKRVEQLSRELDTLRGIFRQL**EF**
pQIS 116 or pMETN (LZ-Leu 8 Pro) No of amino acids: 33 Molecular weight: 4061.6	**LE**LTSDNDRPRKRVEQLSRELDTLRGIFRQL**EF**
pQIS 119 or pMETN (OST1) No of amino acids: 480 Molecular weight: 54590.4	**LE**MRQVWFSWIVGLFLCFFNVSSAAQYEPPATWEN VDYKRTIDVSNAYISETIEITIKNIASEPATEYFTAFES GIFSKVSFFSAYFTNEATFLNSQLLANSTTAPGDDGE SEIRYGIIQFPNAISPQEEVSLVIKSFYNTVGIPYPEHV GMSEEQHLLWETNRLPLSAYDTKKASFTLIGSSSFE EYHPPNDESLLGKANGNSFEFGPWEDIPRFSSNETL AIVYSHNAPLNQVVNLRRDIWLSHWASTIQFEEYYE LTNKAAKLSKGFSRLELMKQIQTQNMRQTHFVTVL DMLLPEGATDHYFTDLVGLVSTSHAERDHFFIRPRF PIFGGWNYNFTVGWTNKLSDFLHVSSGSDEKFVASI PILNGPPDTVYDNVELSVFLPEGAEIFDIDSPVPFTN VSIETQKSYFDLNKGHVKLTFSYRNLISQVANGQVL IKYDYPKSSFFKKPLSIACYIFTALMGVFVLKTLNM NVTN**EF**

Appendix V

The expected DNA fragments sizes (shown in base pairs) from two-hybrids plasmids. The sizes shown in black are generated from pMETN-oriented plasmid and the one shown in red are originated from pMETC plasmid. The combinations of plasmids shown in series A to F are same as explained in text.

A	B	C	D	E	F
pQIS 101n/ pQIS 98 (W/T LZ)	**pQIS 113 (CaM peptide)/ pQIS 98 (W/T LZ)**	**pQIS 115/ pQIS 98 (W/T LZ)**	**pQIS 113 (CaM peptide)/ pQIS 114 (CaM)**	**pQIS 116 (L 8 P)/ pQIS 117 (L 8 P)**	**pQIS 119 (Ost1)/ pQIS 118 (WBP1)**
8,	8,	8,	8,	8,	8,
11, 18	11, 18	11, 18	11, 18	11, 18	11, 18
	25		25		
43, 45	43, 45	43, 45	43, 45	43, 45	43, 45
57	57	57	57	57	56, 57
	72		72		
80, 87	80, 87	80, 87	80, 87	80, 87	80, 87
98	98	98	98	98	98
102	102	102	102	102	101, 102
117	117	117	117	117	117
	123	123	123	123	123
135	135	135	135	135	130, 135, 139
155	155	155	155	155	155
169	169	169	169	169	162, 169
173	173	173	173	173	173, 177
184, 185	184, 185	184, 185	184, 185	184, 185	184, 185
200, 203	200, 203	200, 203	200, 203	200, 203	200, 203
213	213	213	213	213	213, 213
222	222	222	222	222	222
255	255	255	255	255	255
267	267	267	267	267	267
287	287	287	287	287	287
					308
326	326	326	326	326	326
337	337	337	337	337	337
346	346	346	346	346	346
355	355	355	355	355	355
369	369	369	369	369	369
392, 395	392, 395	392, 395	392, 395	392, 395	392, 395
434	434	434	434	434	434
458	458	458	458	458	458
484	484	484	484	484	484
			539		
563	563	563	563	563	563
587	587	587	587	587	587
653	653	653	653	653	653
	662		662		
768		768		768	
					1045

Appendix VI

Sequence of Oligonucleotides used in this study

Name	Sequence 5'--- 3'	Brief Description
NIS 7	TATGTTAATAAAGAATCTGAA**CATCATCATCATCATCA** **T**TAA**TCTAGA**TC	Top oligo (linker) for histidine tag with new *XbaI* site before *XhoI*
NIS 8	TAGA**TCTAGA**TTA**ATGATGATGATGATGATG**TTCAGAT TCTTTATTAACA	Bottom oligo (linker) for histidine tag with new *XbaI* site before *XhoI*
NIS 9	CCGTTAGAG**TGCGCA**AGGCTACAGGC	Top oligo for site directed mutagenesis of F 386 A in M.SPRI
NIS 10	GCCTGTAGCCT**TGCGCA**CTCTAACGG	Bottom oligo for site directed mutagenesis of F 386 A in M.SPRI
NIS 11	GCTTTAGGCTACAG**GGT**TTTGATGACGAAG	Top oligo for site directed mutagenesis of A 390 G in M.SPRI
NIS 12	CTTCGTCATCAAA**ACC**CTGTAGCCTAAAGC	Bottom oligo for site directed mutagenesis of A 390 G in M.SPRI
NIS 13	GGCTACAGGCTTTT**CCG**GACGAAGATTTTG	Top oligo for site directed mutagenesis of D 392 P in M.SPRI
NIS 14	CAAAATCTTCGTC**CGG**AAAAGCCTGTAGCC	Bottom oligo for site directed mutagenesis of D 392 P in M.SPRI
NIS 15	GGAAATTACAC**CCTAGG**GAGTGT**TTT**AGAGTAATGGGC TAC	Top oligo for site directed mutagenesis of A 280 F in *M.Hha*I (StyI site)
NIS 16	GTAGCCCATTACTCT**AAA**ACACTC**CCTAGG**GTGTAATT TCC	Bottom oligo for site directed mutagenesis of A 280 F in *M.Hha*I (StyI site)
NIS 17	CGGAAATTACACCC**TCTAGA**GTGTTTTAGA	Top oligo for site directed mutagenesis of R 277 L in *M.Hha*I (*Xba*I site)
NIS 18	TCTAAAACAC**TCTAGA**GGGTGTAATTTCCG	Bottom oligo for site directed mutagenesis of R 277 L in *M.Hha*I (*Xba*I site)
NIS 19	GTTATAAAGTCCACCC**GTCTAC**CAGCCAAGCATATAAA CAA**GC**TGGTAACTCAGTTGTTATC	Top oligo for site directed mutagenesis of F 302 A in *M.Hha*I (*Acc*I site)
NIS 20	GATAACAACTGAGTTACCAGCTTGTTTATATGCTTGGC TG**GTAGAC**GGGTGGACTTTATAAC	Bottom oligo for site directed mutagenesis of F 302 A in *M.Hha*I (*Acc*I site)
NIS 21	GGCTACC**CAGATTCTG**A**C**AAAGTCCACCCGTC	Top oligo for site directed mutagenesis of Y 289 D in *M.Hha*I (*Alw*NI site)
NIS 22	GACGGGTGGACTTTGT**CAGAATCTG**GGTAGCC	Bottom oligo for site directed mutagenesis of Y 289 D in *M.Hha*I (*Alw*NI site)
NIS 23	TTCAAACCATATGCTCG**T**GG**G**ATTCCCCCTAAGTGAG	Top oligo to delete EcoRI & XhoI site at the end of *M.Hha*I (in pQIS26/27)
NIS 24	CTCACTTAAGGGGAATCCCACGAGCATATGGTTTGAA	Complementary to NIS 23 to delete *Eco*RI & *Xho*I site at the end of *M.HhaI* (in pQIS

		26/27)
NIS 25	**TCGAG**CGTCGACCTGCAGCAATTGGAATTCGGAAGACA C	Linker top oligo designed to fit in pQIS 28/29 @ *Xho*I/*Bbs*I to facilitate the subcloning of TRDs of SPR in *M.HhaI*.
NIS 26	**TTCCGTGTCTTC**CGAATTCCAATTGCTGCAGGTCGACG C	Linker bottom oligo designed to fit in pQIS 28/29 @ *Xho*I/*Bbs*I to facilitate the subcloning of TRDs of SPR in *M.Hha*I.
NIS 27	CCAAAATACAGAATTCGGA**GAATT**CGGAAATTACACCC TCTAG	Primer designed to introduce another *Eco*RI site 9 bp before existing *Eco*RI site.
NIS 28	CTAGAGGGTGTAATTTCC**GAATTC**TCCGAATTCTGTAT TTTGG	Bottom oligo to NIS 27 (*Eco*RI site)
NIS 31	AGAAAGGATATA**CTCGAG**AAAGGGAAACAGAG	Top oligo for for site directed mutagenesis of Q 189 E in M.SPRI (*Xho*I site)
NIS 32	CTCTGTTTCCCTTT**CTCGAG**TATATCCTTTCT	Bottom oligo for site directed mutagenesis of Q 189 E in M.SPRI (*Xho*I site)
NIS 33	GATTTGTTTTCGCAATGAT**TTAATTAA**TCAAAATTTCC AATTTCC	Top oligo for site directed mutagenesis of NI 176/177 IN in *M.Hha*I (*Pac*I site)
NIS 34	GGAAATTGGAAATTTTGA**TTAATTAA**ATCATTGCGAAA ACAAATC	Bottom oligo for site directed mutagenesis of NI 176/177 IN in *M.Hha*I (*Pac*I site)
NIS 35	CAAAGCAACCAAAGTT**GGGCCC**TTTTGAAAATGTTAAA G	Top oligo for for site directed mutagenesis of FFV 112/113/114 LGP in M.SPRI (*Apa*I site)
NIS 36	CTTTAACATTTTCAAAA**GGGCCC**AACTTTGGTTGCTTT G	Bottom oligo for site directed mutagenesis of FFV 112/113/114 LGP in M.SPRI (*Apa*I site)
NIS 37	AAAAAACCTAAAGT**GGGCCC**TATGGAAAATGTG	Top oligo for site directed mutagenesis of VF 116/117 GP in *M.Hha*I (*Apa*I site)
NIS 38	CACATTTTCCATA**GGGCCC**ACTTTAGGTTTTTT	Bottom oligo for site directed mutagenesis of VF 116/117 GP in *M.Hha*I (*Apa*I site)
NIS 41	GAAAAAAAACCTAAAGTGGTCTTTTTTGAAAATGTTAA AGGG	Top oligo for site directed mutagenesis of GP 116/117 VF in chimaeric construct (deleted *Apa*I site)
NIS 42	CCCTTTAACATTTTCAAAAAAGACCACTTTAGGTTTTT TTTC	Bottom oligo for site directed mutagenesis of GP 116/117 VF in chimaeric construct (deleted *Apa*I site)
NIS 43	TATTTAGTAAACGGGA**GAATTC**GGAAATTACACCCTAG	Top oligo for site directed mutagenesis of KT 270/271 RI in *M.Hha*I (*Eco*RI site)
NIS 44	CTAGGGTGTAATTTCC**GAATTC**TCCCGTTTACTAAATA	Bottom oligo for site directed mutagenesis of KT 270/271 RI in *M.Hha*I (*Eco*RI site)
NIS 47	GAAGATACAAGAGGTACCTTGTTTTTTCAATACG	Top oligo for site directed mutagenesis in M.SPRI (Silent, *Kpn*I site)
NIS 48	CGTATTGAAAAAACAAGGTACCTCTTGTATCTTC	Bottom oligo for site directed

		mutagenesis in M.SPRI (Silent, *Kpn*I site)
NIS 51	GCTGAGTGCGTTTATTC**TCGCGA**ATGGGATAAATATGC	Top oligo for site directed mutagenesis of N 39 R in *M.Hha*I (*Nru*I site)
NIS 52	GCATATTTATCCCAT**TCGCGA**GAATAAACGCACTCAGC	Bottom oligo for site directed mutagenesis of N 39 R in *M.Hha*I (*Nru*I site)
NIS 53	GAGCTGGTTGGTTT**TCGCGA**GATTGATAAATATGCC	Top oligo for site directed mutagenesis of S 33 R in M.SPRI (*Nru*I site)
NIS 54	GGCATATTTATCAATC**TCGCGA**AAACCAACCAGCTC	Bottom oligo for site directed mutagenesis of S 33 R in M.SPRI (*Nru*I site)
NIS 55	GCTGAGTGCGTTTATTC**TAACGA**GATTGATAAATATGCCGTC	Top oligo for site directed mutagenesis of R 39 N in chimaeric construct (deleted *Nru*I site)
NIS 56	GACGGCATATTTATCAATC**TCGTTA**GAATAAACGCACTCAGC	Bottom oligo for site directed mutagenesis of R 39 N in chimaeric construct (deleted *Nru*I site)
NIS 57	GAGCTGGTTGGTTT**TAGCGA**ATGGGATAAATATGCAC	Top oligo for site directed mutagenesis of R 33 S in chimaeric construct (deleted *Nru*I site)
NIS 58	GTGCATATTTATCCCAT**TCGCTA**AAACCAACCAGCTC	Bottom oligo for site directed mutagenesis of R 39 S in chimaeric construct (deleted *Nru*I site)
NIS 59	TCGAGTTGACTAGTGACAATGACCGCCTGCGCAAACGCGTGGAACAGCTGAGTCGCGAAC	Leucine zipper coding sequence, top oligonucleotide
NIS 60	TCAGCTGTTCCACGCGTTTGCGCAGGCGGTCATTGTCACTAGTCAAC	Leucine zipper coding sequence, bottom oligonucleotide
NIS 63	CTTTCATTGCCATACGAAATTCCGGATGAGCATTC	Top oligo to delete extra *Eco*RI site from pQIS 100
NIS 64	GAATGCTCATCCGGAATTTCGTATGGCAATGAAAG	Bottom oligo to delete extra *Eco*RI site from pQIS 100
NIS 71	TCGAGAAATATATGGCCAGAAGAAAATGGCAGAAAACAGGCCATGCTGTCCGAGCAATAGGAAGACTGTCATCCATGGCAATGG	Calmodulin binding peptide coding sequence, top oligo
NIS 72	GGACAGCATGGCCTGTTTTCTGCCATTTTCTTCTGGCCATATATTTC	Calmodulin binding peptide coding sequence, bottom oligo
NIS 73	GCGC**GAATTC**CAGCTGGCGGAAGATGC	Reverse primer from *Eco*RI site in leucine zipper-coding reggion
NIS 74	GCGC**CTCGAG**ATGGC**TGATCA**ACTGACAGAAGAG	Forward primer for amplification of Calmodulin gene (*Xho*I site)
NIS 76	GCGC**GAATTC**CTTCGCTGTCATCATCTGTACAAAC	Reverse primer for the amplification of Calmodulin gene (*Eco*RI site)
NIS 77	GCGC**GGATCCATGG**GTAAACTACGTGTA**ATG**	Forward primer to correct the start codon of M.SPRI in pACYC 184
NIS 80	GTTGACTAGTGACAAAG**ATCGAT**TGCGCAAACGCGTGG	Top oligo for site directed mutagenesis N 5 K in leucine zipper ORF (*Cla*I site)

NIS 81	CCACGCGTTTGCGCA**ATCGAT**CTTTGTCACTAGTCAAC	Bottom oligo for site directed mutagenesis N 5 K in leucine zipper ORF (*Cla*I site)
NIS 82	AGTGACAATGACCGC**CCGCGG**AAACGCGTGGAACAG	Top oligo for site directed mutagenesis L 8 P in leucine zipper ORF (*Sac*II site)
NIS 83	CTGTTCCACGCGTTT**CCGCGG**GCGGTCATTGTCACT	Bottom oligo for site directed mutagenesis L 8 P in leucine zipper ORF (*Sac*II site)
NIS 84	GCGC**CTCGAG**ATGAGGCAGGTTTGGTTCTC	Forward primer for amplification of yeast OST1 gene (*Xho*I site)
NIS 85	GCGC**GAATTC**TCAGTTAGTTACGTTCATGT	Reverse primer for the amplification of yeast OST1 gene (*Eco*RI site)
NIS 86	GCGC**CTCGAG**ATGCGGACCGATTGGAATTT	Forward primer for amplification of yeast WBP1 gene (*Xho*I site)
NIS 87	GCGC**CAATTG**GTTTGTTTTTTTGAATGTTT	Reverse primer for the amplification of yeast WBP1 gene (*Eco*RI site)
QY 1	GGCC**CTCGAG**ATGTCAGCCACCAGAGCCAATAAAG	Forward primer to amplify Mlc1 gene sequence from yeast genomic DNA
QY 2	GGCC**GAATTC**TCATTGTCTCAAAACATCTCGATGAACT TC	Reverse primer to amplify Mlc1 gene sequence from yeast genomic DNA
QY 3	GGCC**CTCGAG**GGATCCATGACAGCATATTCAGGCTCT	Forward primer to amplify Iqg1 gene sequence from yeast genomic DNA
QY 4	GGCC**GAATTC**GTCGACTTCTGGCTCTTTCAAATACAGC	Forward primer to amplify Iqg1gene sequence from yeast genomic DNA
IP 1	AC**GAATTC**C**CATATG**ATTGAAATAAAAGAT	Oligo at the beginning of *HhaI* with *EcoRI* and *NdeI* site
IP 4	AGGGGCTATGCCCCTTTC**GTCGACATATG**GTTTGAAAT TTAA	Oligo at the end of *HhaI* with *SalI* and *NdeI* site
ABF 27	GTGTTGGCCGGTGTCGGGGCG	Reverse primer for sequencing SPR C-terminal region
ABF19	GCGC**GGATCC**TTGGGTAAACT**CCGG**GTAATG	Forward primer of *M.SPR* containing *BamHI* and *MspI* sites used for sequencing

Note: Oligonucleotides are abbreviated as oligo and the restriction sites are shown red and bold where appropriate.

Appendix VII

List of 52 prokaryotic methyltransferases

Swiss-prot/ TrEMBL ID	Name	Organism
P05102	M.*Hha*I	*Haemophilus haemolyticus*
P11876	DCM OR MEC	*Escherichia coli*
Q57983	MJ0563	*Methanococcus jannaschii*
Q58600	MJ1200	*Methanococcus jannaschii*
P31974	M.*Alu*I	*Arthrobacter luteus*
P34882	M.*Aqu*I α subunit	*Synechococcus* sp
P34905	M.*Bbv*I	*Bacillus brevis*
P13906	M.*Bspr*I	*Bacillus sphaericus*
P10283	M.*Bep*I	*Brevibacterium epidermidis*
P25262	M.*Hgi*BI	*Herpetosiphon aurantiacus*
P43420	M.*Bsp*6I	*Bacillus* sp Plasmid pXH13
P19888	M.*Ban*I	*Bacillus aneurinolyticus*
P09389	M.SpβI	Bacteriophage SP-beta
P17044	M.*Bsu*FI	*Bacillus subtilis*
P05795	M.ϕ3TI	*Bacillus subtilis*, and Bacteriophage phi-3T
P06530	M.*Bsu*RI	*Bacillus subtilis*
P09915	M.ρ11SI	Bacteriophage rho-11s
P00476	M.SprI	Bacteriophage SPR
P36216	M.CviJI	Chlorella virus IL-3A (CV-IL3A)
P25263	M.*Hgi*CI	*Herpetosiphon aurantiacus*
P25264	M.*Hgi*CII	*Herpetosiphon aurantiacus*
P05302	M.*Dde*I	*Desulfovibrio desulfuricans*
P24600	M.*Hgi*DI	*Herpetosiphon aurantiacus*
P25265	M.*Hgi*DII	*Herpetosiphon aurantiacus*
P50185	M.*Dsa*V	*Dactylococcopsis salina*
P45000	HI1041	*Haemophilus influenzae*
P25266	M.*Hgi*EI	*Herpetosiphon aurantiacus*
P05101	M.*Eco*RII	*Escherichia coli*
P50196	M.*Eco*47II	*Escherichia coli*
P34906	M.*Fnu*DI	*Fusobacterium nucleatum*
P25282	M.*Hga*I-1	*Haemophilus gallinarum*
P25283	M.*Hga*I-2	*Haemophilus gallinarum*
P15446	M.*Hpa*II	*Haemophilus parainfluenzae*
P20589	M.*Hae*III	*Haemophilus aegyptius*
P50192	M.*Hph*I(C)	*Haemophilus parahaemolyticus*
P29567	M.*Mth*TI	*Methanobacterium thermoformicicum*
P11408	M.*Msp*I	*Moraxella* sp
P50188	M.*Nae*I	*Nocardia aerocolonigenes*
P50182	M.*Nla*IV	*Neisseria lactamica*
P08455	M.*Ngo*PII	*Neisseria gonorrhoeae*
P31033	M.*Ngo*MI	*Neisseria gonorrhoeae*
P24581	M.*Nla*X	*Neisseria lactamica*
P34879	M.*Sso*II	*Shigella sonnei*
P16668	M.*Sau*3AI	*Staphylococcus aureus*
P23737	M.*Sau*96I	*Staphylococcus aureus*
P34877	M.*Scr*FIA	*Lactococcus lactis (subsp. cremoris)*
P34878	M.*Scr*FIB	*Lactococcus lactis (subsp. cremoris)*
P09795	M.*Sin*I	*Salmonella infantis*
P15840	M.*Sss*I	*Spiroplasma* sp. (strain MQ-1)
P52311	M.*Xor*II	*Xanthomonas oryzae*
S61228	M.*Bss*HII	*Bacillus stearothermophilus*
JQ1019	M.H2I	Bacteriophage H2

Appendix VIII

List of 16 eukaryotic methyltransferases

SWISS-prot/ TrEMBL ID	Gene & Organism
P05102	*Hha*I methyltransferase
Q27746	Cytosine-specific methyltransferase from *Paracentrotus lividus* (common sea urchin)
Q48867	Cytosine-specific methyltransferase from *Daucus carota* (carrot)
O49889	Cytosine-specific methyltransferase from *Lycopersicon esculentum* (tomato)
O65343	Cytosine-specific methyltransferase from *Zea mays* (maize)
O23273	Cytosine-specific methyltransferase from *Arabidopsis thaliana* (mouse-ear cress)
O48866	Cytosine-specific methyltransferase from *Daucus carota* (carrot)
Q9Z33)	Cytosine-specific methyltransferase from *Rattus rattus* (black rat)
Q92072	Cytosine-specific methyltransferase from *Gallus gallus* (chicken)
P79922	Cytosine-specific methyltransferase from *Xenopus laevis* (african clawed frog)
Q9YHD9	Cytosine-specific methyltransferase from *Brachydanio rerio* (zebrafish)
P13864	Cytosine-5-methyltransferase (M.*Mmu*I) from *Mus musculus* (mouse)
P40999	DNA Methyltransferase PMT1 (M.*Spo*mI) *Schizosaccharomyces pombe* (fission yeast)
P34881	Cytosine-5-methyltransferase from *Arabidopsis thaliana* (mouse-ear cress)
P26358	Cytosine-5-methyltransferase M.HsaI (DNMT1) from *Homo sapiens* (human)
O14717	Putative DNA methyltransferase DNMT2 from *Homo sapiens* (human)
Q9W6K7	*De novo* DNA methyltransferase 3 DNMT3 (Fragment) from *Brachydanio rerio* (zebrafish)