Cloning of Methyltransferase cDNAs from *Pisum sativum*

Michele Cummings

Thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Department of Biochemistry and Molecular Biology,

University of Glasgow,

October 1995.

Acknowledgements

I would like to thank the following:

Dr. R.L.P. Adams for his supervision of this work and for his consistently friendly and helpful advice.

The Science and Engineering Research Council for financial support.

Members of lab C30, past and present, who have created such a friendly atmosphere in which to work.

Special thanks are also due to Michael, for listening with (sometimes very well) feigned interest to lengthy accounts of my project, for walking me to and from the lab at odd hours in the evening, and for just being there.

Finally, I would like to thank my Mum and Dad for all their support and encouragement over the years.

Abbreviations

The abbreviations used in this thesis are in agreement with the recommendations of the editors of the Biochemical Journal (*Biochem. J.* (1994) **297**, 1–15) with the following additions:

5-mC	5-methylcytosine
AdoMet	S-adenosyl-L-methionine
AdoHcy	S-adenosyl-L-homocysteine
CTAB	cetyltrimethylamonium bromide
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulphoxide
DNase	deoxyribonuclease
DTT	dithiothreitol
IPTG	isopropyl-β-D-galactopyranoside
MOPS	morpholinopropane-sulphonic acid
MTase	DNA methyltransferase
nt	nucleotide
PEG	polyethylene glycol
RNase	ribonuclease
SSC	saline sodium citrate
SSPE	saline sodium phosphate/EDTA
TEMED	N, N, N´ N´-tetramethylethylenediamine
TMACl	tetramethylammonium chloride
TNS	triisopropyl-napthalenesulphonic acid
Tris	tris(hydroxymethyl)aminoethane
TWEEN-20	polyethylene sorbitan monolaurate
UTR	untranslated region
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Contents

List of Figures

Summary

A number of approaches were adopted to clone pea MTase cDNAs. The only successful approach, however, was PCR amplification of pea MTase-specific segments using degenerate primers, that were designed mainly according to the conserved motifs IV and VI of mouse MTase. A downstream antisense primer based on motif VI was used to prime first-strand cDNA synthesis from pea total RNA. This was then amplified using the same primer in combination with a sense primer designed according to motif IV. Several of the PCR products which were amplified were cloned into pBluescript and then sequenced. Only one type of PCR product was found to be MTase-specific upon translation of the sequences. On comparison with the database, this PCR product was found to share the highest homology with *Arabidopsis* MTase, followed by mouse and human MTases. No other MTase-specific PCR product was identified out of a total of 18 independent clones that were sequenced. Only one MTase mRNA (of 5.0 kb in length) was identified upon hybridisation of the cloned pea MTase-specific PCR product to a northern blot of pea poly(A)$^+$ RNA.

The 3′ end of the pea MTase cDNA (corresponding to most of the catalytic domain) was cloned using a 3′ RACE strategy and *Pfu* polymerase. Seven independent clones were mapped, and three were sequenced. The three clones were found to be absolutely identical in sequence, except for at the extreme 3′ end, where the sites used for polyadenylation were found to be different for each clone. The nucleotide sequence of the 3′ RACE products was also found to match that of the original MTase-specific PCR products in the region between motifs IV and VI. Again, the translated sequence of the 3′ RACE product shared the highest homology with *Arabidopsis* MTase. No evidence for alternative splicing in the 3′ end of the pea MTase transcript was found.

Developmental changes in the steady state levels of pea MTase mRNA were investigated by hybridisation of one of the cloned PCR products to a northern blot of

total RNA extracted from different tissues from pea seedlings, taken at different stages of development. The strength of the hybridisation signal was found to parallel the changes in MTase specific activity (Yesufu *et al.*, 1991), being highest in rapidly dividing tissues .

To determine whether a family of related pea MTase genes exists in pea, the 3′ RACE product was hybridised at moderate stringency to a Southern blot of pea genomic DNA digested with various restriction enzymes insensitive to methylation at CpG and CpNpG sites. The results were consistent with the presence of only one type of MTase gene in pea.

Chapter 1 Introduction

1.1 DNA methylation

The presence of methylated bases in DNA is an almost universal feature of both prokaryotes and eukaryotes. Prokaryotic DNA contains the modified bases N^6-methyladenine, 5-methylcytosine (5-mC) and N^4-methylcytosine whereas the only methylated base in eukaryotes is 5-methlylcytosine with very few exceptions (Burdon and Adams, 1980).

The most widely distributed kind of methylation is that associated with the restriction modification systems of bacteria. These systems serve to protect the bacteria from phage infection; the chromosomal DNA is methylated in a sequence specific manner by a methyltransferase, protecting it from digestion by the 'cognate' restriction endonuclease which will cleave unprotected phage DNA (reviewed by Noyer-Weidner and Trautner, 1993). It has been proposed that in the course of evolution, the prokaryotic methyltransferase system has been adapted by higher eukaryotes for the purpose of regulating the structure and expression of the genome (Bestor, 1990).

5-methylcytosine is readily converted to thymine by oxidative deamination (Coulondre *et al.*, 1978). Failure of the mismatch repair system (Wiebauer *et al.*, 1993) results in C-G to T-A mutations. Thus 5-methylcytosines represent mutational hot spots in the genome. This would seem a very powerful argument for the supposed advantages conferred by DNA methylation in higher eukaryotes (whose genomes are extensively methylated), which would outweigh the mutational load of methylated cytosines.

1.2 Methods for studying DNA methylation

The presence of 5-mC was first detected in DNA by Hotchkiss (1948), at a time when the structure and function of DNA was still a mystery. Since then remarkable progress has been made in the methodologies used to study nucleic acids in general,

as well as the modified bases in DNA. Some of the latter will be discussed in the following sections.

1.2.1 Methods to study 5-methylcytosine distribution in DNA

To determine the total 5-mC content of DNA, the DNA is first hydrolysed by chemical or enzymatic means. Enzymatic hydrolysis involves little 5-mC deamination (Ford *et al.*, 1980), and leads to the production of mononucleotides with 5′ phosphates (e.g. in the case of DNase I) or with 3′ phosphates (e.g. with micrococcal nuclease). The cleavage products are then fractionated by thin-layer chromatography (TLC), or by high performance liquid chromatography (HPLC) (reviewed by Adams and Burdon, 1985). An improved method was developed whereby nicks were introduced into the DNA using DNase I and the nucleotide 3′ to the nicked base was labelled using [α-^{32}P]dNTP. On hydrolysis to the 3′ monophosphates and separation via TLC, this modified nearest neighbour analysis indicates what proportion of each of the four CpN sequences are methylated (Pollack *et al.*, 1984).

Much information about patterns of DNA methylation has been gained by the use of bacterial restriction enzymes. There are several isoschizomeric pairs which have the same recognition sequence but have differential sensitivity to the presence of methylated bases within that sequence. For example, *Hpa*II will only cleave the target sequence (CCGG) if the internal cytosine is not methylated on either strand whereas *Msp*I will cleave regardless. Comparison of the digestion patterns therefore reveals the methylation status of DNA (reviewed by Adams and Burdon, 1985; Saluz and Jost, 1993). The sensitivity of this method has been improved by using PCR in such a way that only undigested (i.e. methylated) DNA will be amplified (Singer-Sam *et al.*, 1990).

However, using methylation-sensitive restriction enzymes limits the detection of 5-mCs to those contained within the recognition sequences. This is not a problem

with genomic sequencing. The original method as devised by Church and Gilbert (1984) involves digestion of total genomic DNA with restriction enzymes which provides fragments of defined length containing the target sequence. The DNA is then subject to chemical sequencing reactions (Maxam and Gilbert sequencing) and the cleavage products are then resolved on a sequencing gel and transferred onto a nylon filter. Southern hybridisation using the appropriate probe reveals the sequencing ladder of the target sequence. Methylated cytosines are represented as gaps in the ladder as hydrazine does not react with 5-mC in the C specific reaction. Several improvements have been made on this procedure, such as linear PCR amplification of the cleavage products with labelled primers (Saluz and Jost, 1989). A new method for genomic sequencing is the bisulphite method (Frommer *et al.*, 1992) in which bisulphite is used to specifically deaminate cytosine (but not 5-mC) to uracil. The target sequence is then amplified by PCR and subject to dideoxy (Sanger) sequencing. Upon sequencing the uracil residues become detectable as thymine residues whereas 5-mC is detected as cytosine

1.2.2 The use of methylase inhibitors to study the effects of DNA methylation *in vivo*

Inhibition of DNA methylation is widely used as a means to study the effect of DNA methylation on biological processes. Several drugs that inhibit methylation both *in vivo* and *in vitro* have been identified (reviewed by Adams and Burdon, 1985). The most widely used of these are the base analogues 5-azacytidine and 5-azadeoxycytidine. On incorporation into DNA, 5-azacytosine brings about an inhibition of DNA methylation (Jones and Taylor, 1980). The triazine ring is unable to accept a methyl group, but this alone is not enough to account for the extent of inhibition. Moreover, heavily substituted DNA can inhibit the methylation of unsubstituted DNA *in vitro* (Adams *et al.*, 1984). This suggests that DNA methyltransferase binds irreversibly to DNA containing 5-azacytosine. This

'demethylating' ability of 5-azacytidine has been exploited in many experiments investigating the effects of DNA methylation (mainly with cultured animal cells), some of which are mentioned later. However, 5-azacytidine is toxic at concentrations over 1 μM and leads to a marked change in the metabolism and development of the exposed cells (Jones, 1984). Therefore the significance of the data obtained by these experiments must be assessed critically (Bird, 1992). Recently, more doubt has been cast on the traditional interpretations of these experiments by the work of Juttermann *et al.* (1994), who showed that methylation-deficient mutant mouse embryonic stem (ES) cells and embryos (Li *et al.*, 1992) were significantly more resistant to the toxic effects of 5-azadeoxycytidine than their wild-type counterparts. They suggested that the covalent trapping of methyltransferase rather than the subsequent demethylation was responsible for the cytotoxicity of the base analogue.

1.3 Distribution of 5-methylcytosine in higher eukaryotes

Eukaryotic organisms differ widely in their content of 5-mC. The traditional organisms used in molecular biology, whose genomes are most well characterised such as *Saccharomyces cerevisiae* (Proffit *et al.*, 1984), *Drosophila melanogaster* (Pollack *et al.*, 1984) and *Caenorhabditis elegans* (Simpson *et al.*, 1986) have virtually no methylated cytosines. The DNA of most species of fungi either lack or have very low amounts of 5-mC (Magill and Magill, 1990). However, vertebrates have between 4 and 6% of their cytosines methylated (Ehrlich *et al.*, 1982; Gamma-Sosa *et al.*, 1983), and the levels of 5-mC in higher plants can reach up to 30% (Adams and Burdon, 1985). In vertebrate DNA, 5-mC is found in the CpG dinucleotide whereas in higher plants both CpG and CpNpG sites can be methylated (Gruenbaum *et al.*, 1981).

The extent of methylation has a rough correlation with genome size. In simple eukaryotes such as some fungi (Rothnie *et al.*, 1991) and in some invertebrates such as echinoderms (Bird *et al.*, 1979), 5-mC appears to be confined to satellite

sequences. It has been postulated that methylation of highly repetitive elements within the genome provides a mechanism for sequence diversification which suppresses recombination, helping to maintain genome stability (Selker 1990; Kricker *et al.*, 1992). Even in a higher vertebrate such as the mouse, 50% of 5-mC is found in repetitive sequences (Miller *et al,* 1974). However, methylation in higher plants and vertebrates is not confined to repetitive DNA sequences and is involved in the control of gene expression and development in these organisms. This will be discussed in later sections.

The generation of methylated DNA is mainly a postreplicative process in prokaryotes (and in eukaryotes exclusively so). The class of enzymes responsible for this phenomenon are called DNA methyltransferases (MTases).

1.4 Prokaryotic DNA methyltransferases

As mentioned earlier, the most widely distributed kind of methylation is that which is associated with the restriction modification systems of bacteria. There are three classes of MTases involved in these systems: Types I, II, and III MTases and all require S-adenosyl-L-methionine (AdoMet) as the methyl-donating cofactor.

Types I and III are multi-subunit enzymes in which the restriction and modification activities reside in different subunits and, in type I enzymes, yet another subunit is responsible for determining target sequence specificity. All members of types I and III are N^6-methyladenine MTases. The type II MTases are separate proteins from their cognate restriction endonucleases and members of this class can be 5-methylcytosine, N^4-methylcytosine or N^6-methyladenine MTases. The remainder of this section will be devoted to the prokaryotic type II 5-methylcytosine (cytosine-C5) MTases, as they are evolutionarily linked to the eukaryotic MTases.

1.4.1 Type II cytosine-C5 methyltransferases

This class of MTases recognise palindromic sequences (type IIS MTases recognise non palindromic sequences and usually two different MTases are responsible for methylating cytosines on either strand). There are both mono and multispecific type II cytosine-C5 MTases (the multispecific MTases are encoded by certain *Bacillus subtilis* phages). The type II enzymes methylate the target cytosines on either strand of the recognition sequence in a sequential manner and show no preference for either unmethylated or hemimethylated substrates.

The type II and type IIS cytosine-C5 MTases (reviewed by Noyer-Weidner and Trautner, 1993; Kumar *et al.*, 1994) exhibit the highest degree of sequence similarity at the amino acid level. Sequence alignment showed highly conserved motifs interspersed with less conserved regions (Posfai *et al.*,1989; Lauster *et al.*, 1989). Posfai and co-workers identified ten separate conserved regions, I to X, whose order is absolutely invariant. The addition of further cytosine-C5 MTases to this list revealed motifs I, IV, VI, IX and X to be the most strongly conserved.

Regions VIII and IX are separated by a variable region which shows the most diversity both in length and sequence composition. Complementation was observed between separate N-terminal and C-terminal segments overlapping the variable region. Furthermore, this complementation could be achieved with heterologous combinations of two closely related MTases, M.*Bsp*RI and M.*Bsu*RI which methylate the same target sequence (Posfai *et al.*, 1991). This suggests that the segments on either side of the variable region represent independent functional entities, and indeed in the unusual case of M.*Aqu*I, the two 'domains' do exist as separate subunits (Karreman and de Waard, 1990). The target recognising domain (TRD) was originally assigned to the variable region purely because this was the region which varied between enzymes of different target specificities which otherwise had particularly high levels of sequence similarity, though there is no shared homology between the variable regions and other characterised DNA binding motifs

(Buhk *et al.*, 1984). Direct evidence was provided by the construction of chimaeric multispecific MTases in which the composite parts were derived from MTases with different specificities (Trautner *et al.*, 1988). Hybrid swap experiments with the monospecific MTases established that not only did the variable region determine sequence specificity, but also the choice of the specific base to be methylated within the target sequence (Klimasauskas *et al.*, 1991; Mi and Roberts, 1992). In addition, site-directed mutagenesis was used to map TRDs within the variable region of multispecific MTases (Wilke *et al.*, 1988). The data are consistent with the idea that the TRDs act in a modular fashion to confer target specificity and in general, target specificity is independent of the conserved regions. Sequence comparisons between TRDs (which are about 40 amino acids in length) suggested a weak consensus sequence: T(V/I/L)XXXXXXG(V/L) (Lauster *et al.*, 1989). It should be stressed however, that *a priori* predictions of the target sequence from the amino acid sequence of the variable region are not possible.

Conserved region I, consensus (F/G)XGXG forms part of the AdoMet binding site. This motif is found in MTases (DNA, RNA and protein) which utilise AdoMet as a methyl donor (Ingrosso *et al.*, 1989).

Region IV is responsible for the transfer of the methyl group onto the C5 position of the cytosine ring, and contains the highly conserved motif PCXXXS. The details of catalysis will be discussed later.

X-ray crystallography of M.*Hha*I complexed with AdoMet has revealed the tertiary structure to consist of a large N-terminal and small C-terminal domain linked by a flexible hinge region, the variable domain making up the hinge region and most of the small domain. All the conserved motifs are located in the large domain, with the exception of motif IX, which is mostly in the small domain. Motif IX is probably necessary for correct folding of the small domain. As expected, the conserved motifs form the core of the protein and constitute most of the structures that surround the active site cleft (Cheng *et al.*, 1993).

1.4.2 Mechanism of methyl transfer

Studies on the kinetics of M.*Hha*I DNA methylation led to the proposal by Wu and Santi (1987), that methylation proceeded through nucleophilic attack of the C6 position of the cytosine ring by the cysteine thiol group of the highly conserved PCXXXS motif mentioned earlier. This leads to the formation of a transient covalent MTase-DNA complex via a thioether bond. (Fig 1.1). Evidence for this mechanism came from the observation that the the exchange of the conserved cysteine residue for other amino acids by site-directed mutagenesis led to destruction of enzyme activity (Wilke *et al.*, 1988, Wyszynski *et al.*, 1992). Direct evidence came from covalent trapping of M.*Hae*III onto an analogue of the target cytosine, 5-fluoro-2′-deoxycytidine (FdC), which was contained in the oligonucleotide substrate. Exhaustive proteolysis of the DNA-M.*Hae*III complex followed by peptide sequencing revealed the covalent attachment to the DNA to be via the conserved cysteine residue (Chen *et al.*, 1991).

A profound insight into the mechanism of methyl transfer came from the determination of the co-crystal structure of a DNA MTase bound covalently to DNA (Klimasauskas *et al.*, 1994; see Fig. 1.2). Klimasauskas and co-workers had crystalised a chemically trapped intermediate between M.*Hha*I and a self complementary duplex 13-mer oligonucleotide. The enzyme is linked through a thioether bond to the target FdC residue on one strand and the cofactor product of methyl transfer, S-adenosyl-L-homocysteine (AdoHcy), remains bound to the active site. The 3D structure of this crystal revealed the DNA to be located in the cleft between the two domains predicted by Cheng *et al.* (1993). The major groove faces the variable region in the small domain, and the minor groove faces the catalytic site in the large domain. The DNA is in the B conformation except for a disrupted G-C base pair that contains the target cytosine. This cytosine has swung completely out of the double helix and is positioned in the active site, which itself has undergone a large conformational change. Though the DNA is contacted in both the major and minor

grooves, almost all the base-specific interactions occur in the major groove, via two Gly rich loops (loops I and II) from the small domain which correspond to the putative TRD predicted by Lauster *et al.*, (1989). This mechanism represents a hitherto unknown mode of protein-DNA interaction.

The mechanistic roles of the other conserved amino acids apart from the catalytic cysteinemay be inferred from the crystal structure. For example, the amide carbonyl preceding the absolutely conserved proline of the Pro-Cys dipeptide forms a hydrogen bond with the exocyclic amine of the target FdC, and the proline carbonyl appears to hydrogen bond with the asparagine residue of the absolutely conserved ENV sequence of motif VI. The target FdC is stabilised by interactions with Phe79 of motif IV, Glu119 of region VI, and Arg165 of region VIII. The latter two amino acids are absolutely conserved in all the cytosine-C5 MTases. The conserved Glu residue may determine specificity for cytosine since it hydrogen bonds to the N^3 and N^4 of cytosine, and plays a part in catalysis. Surprisingly, no obvious candidate enzymic base for the abstraction of the hydrogen has been revealed by this structure. The unpaired guanine residue is stabilised by hydrogen bonds with Gln237 (from loop I of the small domain), maintaining the stacking of the helix. The Gln is in turn stabilised by Ser87 from region IV. However, neither of these amino acids appear to be well conserved. AdoMet binds to a pocket in the large domain facing the cleft. Residues from motifs I–V and from motif X contribute to this pocket, which means that the cofactor is in very close proximity to the catalytic site.

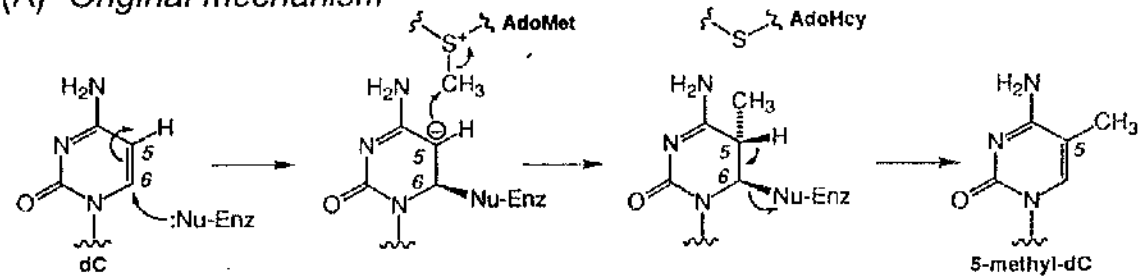
Figure 1.1 Catalytic mechanism of methyl transfer (adapted from Bestor and Verdine, 1994)

A. As originally proposed by Wu and Santi (1987), the catalytic mechanism of cytosine-C5 MTases by analogy to thymidylate synthase which also contains a Pro-Cys dipeptide, involves nucleophilic attack on C6 of the target cytosine ring by the conserved cysteine of motif IV (Enz-Nu: = the cysteine thiol). This results in the formation of a transient covalent MTase/DNA complex via a thioether bond and the conversion of C5 to a carbanion which then, through nucleophilic attack on the methyl group of AdoMet, induces the transfer of the methyl group to C5. Elimination of the hydrogen at C5 and the enzyme at C6 results in the restoration of the C5-C6 double bond to produce 5-methylcytosine.

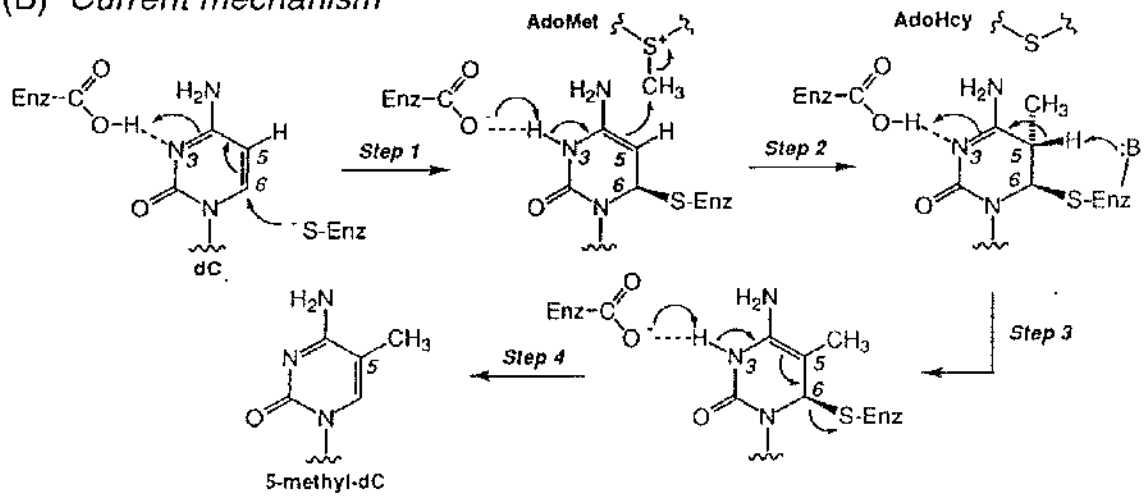
B. The revised mechanism, as proposed by Erlanson *et al.*, 1993, who proposed more favourable enamine intermediates. This mechanism implicitly requires strand separation for the proposed attack trajectories. In the co-crystal structure of M.*Hha*I (Klimasauskas *et al.*, 1994), Enz-S is Cys 81 (of motif IV) Enz-COOH is Glu 119 (of motif VI), and Enz-B: is unknown.

C. With the cytosine analogue, FdC, elimination of the fluorine atom at C5 is energetically unfavourable. This results in the covalent trapping of the enzyme to C6 of the methylated FdC via a thioether bond (Chen *et al.*, 1991).

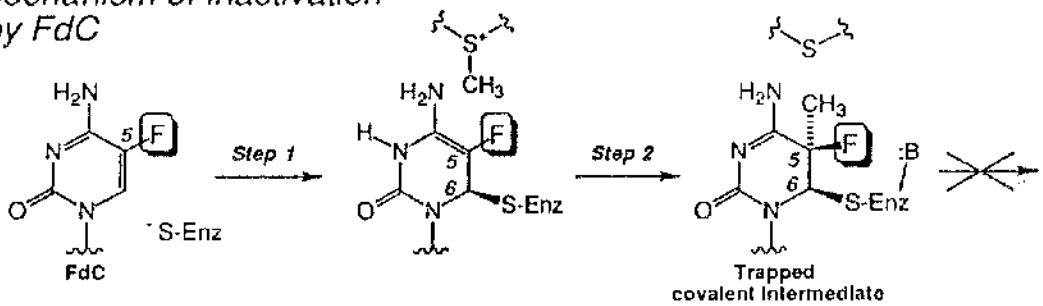
(A) *Original mechanism*



(B) *Current mechanism*



(C) *Mechanism of inactivation by FdC*



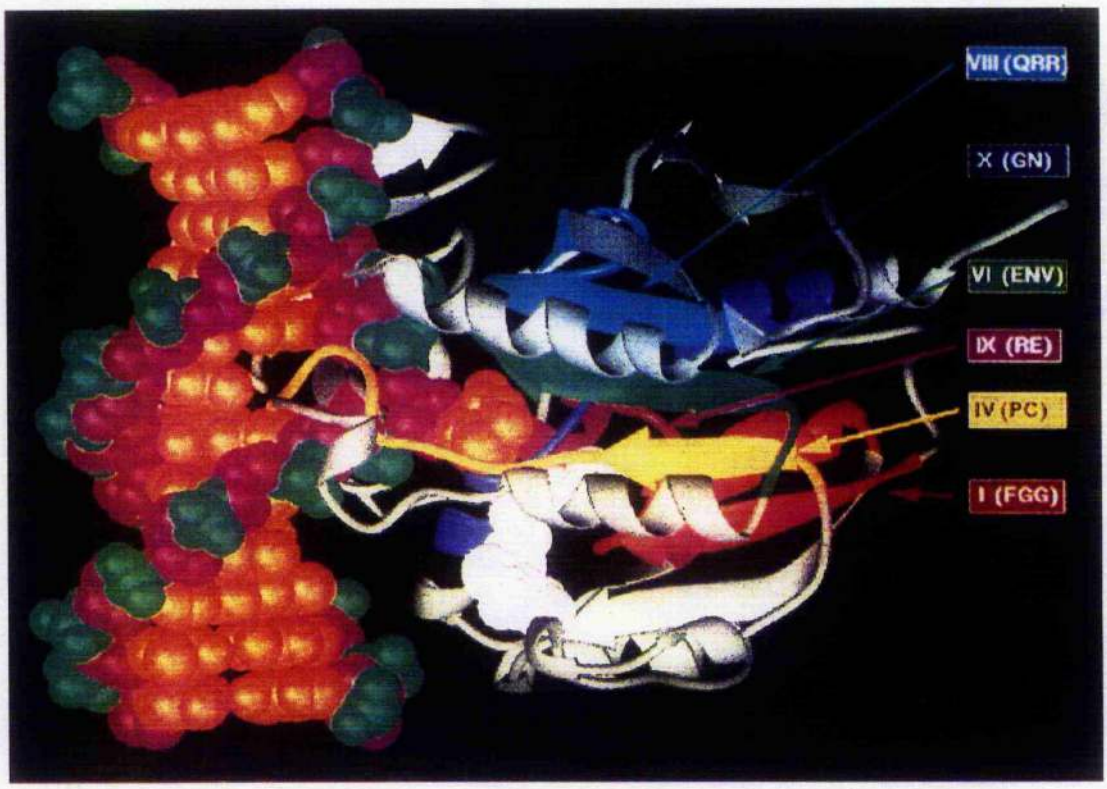


Figure 1·2
The structure of M.*Hha*I in a ternary complexed with a substrate duplex DNA oligonucleotide and the end product of the reaction AdoHcy. The motifs are shown in colour: AdoHcy is in white, the DNA bases are in orange, the deoxyribose is in purple and the phosphates are green.

Roberts R.J. (1993), Les Prix Nobel, p156

1.5 Eukaryotic DNA methyltransferases

Unlike their prokaryotic counterparts, which do not discriminate between unmethylated and hemimethylated targets, the eukaryotic cytosine-C5 MTases preferentially methylate hemimethylated substrates (Gruenbaum *et al.*, 1982). This reflects the different role that MTases play in eukaryotes.

1.5.1 Maintenance and *de novo* methylation

Patterns of DNA methylation are inherited and maintained over many cell generations (Wigler *et al.*, 1981) by a process known as maintenance methylation. The semi-conservative nature of DNA replication results in the parental strand remaining methylated while the newly synthesised daughter strand is unmethylated. In an early postreplicative step (Burdon and Adams, 1969) this hemimethylated DNA is recognised by DNA MTase which, using the parental strand as a template, symmetrically imposes a methyl group on the newly synthesised strand. This mechanism was first proposed by Holliday and Pugh (1975) and Riggs (1975). However, this model may be an over-simplification as methylation patterns of non-symmetrically methylated sequences in transgenic plants (Meyer *et al.*, 1994) and at origins of replication in cultured hamster kidney cells (Tasheva and Roufa, 1994) appear to be maintained following replication. Also, certain CpG dinucleotides appear to be maintained in a hemimethylated state in avian liver cells (Jost and Saluz, 1993). A model has been suggested to explain these phenomena, in which the DNA MTase recognises a hemimethylated region of DNA (rather than an individual CpG site), that is associated with proteins characteristic of inactive chromatin. In contrast, active chromatin lacks these proteins and is replicated in early S phase (when the level of DNA MTase is limiting). The maintenance of methylation patterns could therefore be seen as a consequence of recognition of regions of chromatin structure, rather than individual CpG sites (Adams, 1995).

New patterns of methylation are established during embryogenesis by *de novo* methylation of previously unmethylated DNA (reviewed by Adams *et al.*, 1993). However, little is known of how sequence-specific methylation patterns are established in the first place, as the sequence-specificity of purified mammalian MTase does not extend much past the CpG dinucleotide (Gruenbaum *et al.*, 1981b; Bestor and Ingram, 1983; Hubrich-Khuner*et al.*, 1989). It remains possible that there are different specific *de novo* MTases expressed in a tissue specific and stage specific manner, though there is no evidence for the presence of a gene family in mammals. However, a longer MTase mRNA (6.2.kb as opposed to the 5.2 kb mRNA found in somatic cells) has been identified in pachytene spermatocytes, a cell type that is known to be active in *de novo* methylation (Trasler *et al.*, 1992). The size of the MTase enzyme in these cells, at 180–200 KDa, is similar to that of somatic tissues. Until this alternative form of MTase has been cloned, one cannot ascribe the high rates of *de novo* methylation to a different form of MTase enzyme in these cells. It has been suggested that the pattern of *de novo* methylation is dependent on other factors beside the specificity of DNA MTase itself. An important development in this field has been the discovery that the presence of Sp1 sites in close proximity to CpG islands protects these islands from *de novo* methylation *in vivo*, though the mechanism of this protection is not clear (Brandeis *et al.*, 1994).

It has been shown that *de novo* and maintenance activities reside in the same protein and, in the case of the mammalian enzymes, the preference for hemimethylated sites is 30–40 fold higher (Bestor and Ingram, 1983; Bolden *et al.*, 1984).

1.5.2 Structure of mammalian DNA methyltransferases

The first of the eukaryotic MTase cDNA to be cloned was from murine erythroleukaemia (MEL) cells (Bestor *et al.*, 1988). The cDNA sequence revealed an open reading frame corresponding to a protein of about 170 KDa, which is the size of

the protein yielded when the cloned cDNA is expressed in COS cells (Czank *et al.*, 1991). However, the size of the protein yielded when the cDNA is expressed in *E. coli* is 190 KDa (Tollefsbol and Hutchison, 1995). Purified MEL MTase has an apparent molecular weight of ~190 KDa, as does MTase purified from a variety of cell lines. However, MTase from oocytes and early embryos is 170 kDa (Carlson *et al.*, 1992). This observation suggests that DNA MTase normally undergoes post-translational modification, the nature which is as yet unknown. Alternatively, the smaller enzyme may arise from the use of an alternative translation start site, or from *in vivo* N-terminal proteolysis. However, the specificity of both forms of the enzyme is the same *in vitro*, and the significance of this smaller form of MTase is obscure (see section 1.8).

The predicted amino acid sequence of the cDNA reveals a 1000 amino acid N-terminal domain linked to a C-terminal domain of about 500 amino acids that is closely related to the prokaryotic type II cytosine-C5 MTases discussed earlier. In the original paper, only conserved regions IX and X were apparent due to the presence of one or more frame-shifts. After sequence corrections, all of the conserved motifs (with the exception of motifs III and V), were identified (Bestor, personal communication). The variable domain (between motifs VIII and IX) is much longer than most prokaryotic monospecific MTases, and is in fact the longest known of the monospecific enzymes. Figure 1.3 is a diagram of mouse MTase, and shows various features of the N-terminal domain, and the position of the catatalytic centre (motif IV) in the C-terminal domain.

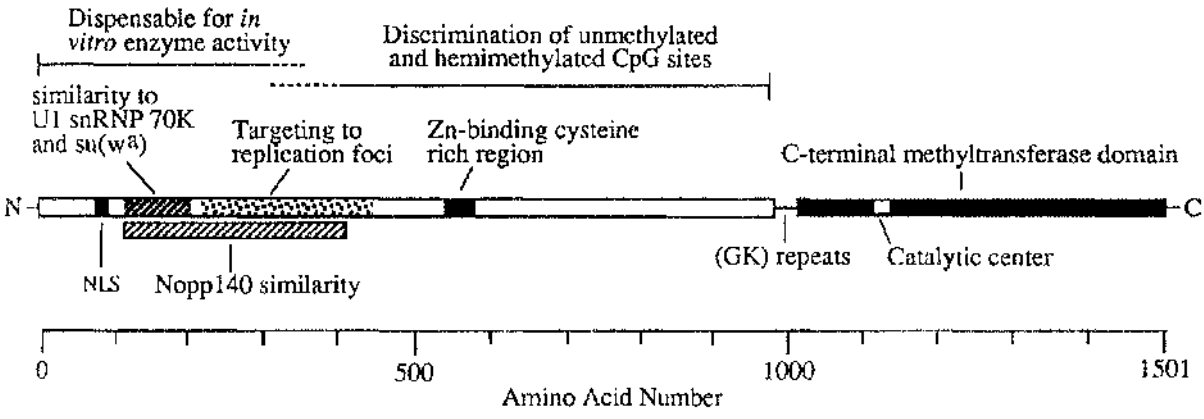
The mechanism of methyl transfer (Fig 1.1) is presumably the same for the eukaryotic DNA MTases because of the presence of the conserved motifs in all the MTases cloned to date. Smith *et al.* (1992) demonstrated the formation of a high molecular weight complex when a ^{32}P labelled double-stranded oligonucleotide substrate containing FdCpG with ^{m}CpG on the opposite strand was incubated with partially purified human MTase. The apparent molecular weight of this complex was consistent with that of the oligonucleotide plus the catalytic unit present in

Figure 1.3 Structure of mouse MTase (adapted from Bestor and Verdine, 1994)

Mouse MTase consists of a C-terminal (catalytic) domain linked to an N-terminal domain by a series of Lys/Gly repeats.

The C-terminal domain is related to the prokaryotic type II cytosine-C5 MTases, and contains all the conserved motifs except for III and V. (The catalytic centre corresponds to motif IV.)

The N-terminal contains a nuclear localisation signal (NLS) and a region that has been implicated in the localisation of the enzyme to the replication factories in the nucleus. This region shares significant homology with the nucleolar protein, Nopp140. Adjacent to this region is a region homologous to an Arg- and Ser-rich motif found in many splicing complex proteins. There is also a Cys-rich region which binds to Zn and may be involved in the discrimination between unmethylated and hemimethylated DNA.



homogeneous preparations of human MTase. This complex was only formed in the presence of AdoMet, which suggests that complex formation is dependent on methyl transfer, as with the prokaryotic MTases.

The C-terminal catalytic domain of mouse MTase is joined to the N-terminal domain by a run of 13 alternating glycyl and lysyl residues. It has been proposed that during the evolution of eukaryotic MTases the 'prokaryotic' catalytic domain has become fused at the genetic level to an unrelated protein, and that the site of this fusion is the Lys-Gly repeats (Bestor, 1990). The first 200 amino acids of mouse MTase are very rich in charged and polar amino acids and the first ~300 amino acids are very susceptible to proteolytic cleavage during enzyme purification. Deletion of these sequences does not affect *de novo* or maintenance methylation *in vitro* (Bestor and Ingram, 1985).

Residues 207 to 455 have been shown to target mouse MTase to replication foci in S phase nuclei. Thus replication and methylation are closely coordinated, which is to be expected for faithful inheritance of methylation patterns to occur. The nature of this association is not known and indeed, the structural organisation of replication foci and associated proteins is poorly understood at present. Incidentally, this work revealed a functional nuclear localisation signal at residues 72–92 (Leonhardt *et al.*, 1992). Surprisingly, the N-terminal domain was found to contain a region similar to an Arg- and Ser-rich motif which may be involved in the assembly of splicing complexes. This region is overlapped by a region related to Nopp140 (a protein which is thought to shuttle between the nucleolus and the nuclear membrane via some kind of nuclear filament). The region of Nopp140 similarity represents most of the replication foci-targeting domain. As MTase is never associated with splicing complexes or the nucleolus, it has been suggested that the similarities between these proteins represent divergent forms of a sequence which interact with different classes of hypothetical motile nuclear filaments, which direct the proteins to specific nuclear domains (Bestor and Verdine, 1994).

In the centre of the N-terminal domain is a cluster of 8 cysteinyl residues

which has been shown to bind zinc ions and which may be part of a DNA binding motif (Bestor, 1992). In addition the motif S/TPXX, where X tends to be a basic amino acid, occurs ten times in the N-terminal domain; these motifs are frequently found in regulatory proteins and bind to the minor groove of DNA, the narrower groove of AT rich DNA being the preferred binding site (Churchill and Suzuki, 1989).

The N-terminal domain is involved in the discrimination between unmethylated and hemimethylated DNA. Adams *et al.* (1983) observed an increased rate of *de novo* methylation following the treatment of a crude DNA MTase preparation with trypsin and concluded that the enzyme must contain a protease-sensitive domain that makes contact with the C5 methyl group of hemimethylated sites. Protease V8 was shown to cleave DNA MTase between the N- and C-terminal domains. Cleavage caused a large stimulation of *de novo* methylation, without significant alteration in maintenance methylation activity. This demonstrates that the N-terminal domain suppresses the *de novo* activity of the C-terminal domain (Bestor, 1992). However, since the 3D structure of M.*Hhal* MTase discussed earlier reveals that the DNA is contacted in the major groove by regions from the variable domain, there appears to be little room for contacts with a component of the N-terminal domain of the eukaryotic MTases. It has been suggested that it is methylation dependent structural alterations in DNA, rather than the direct contacts with the major groove which are responsible for the discrimination between unmethylated and hemimethylated substrates, and that the zinc binding domain is responsible for sensing the methylation status of DNA (Bestor, 1992).

The central importance of DNA MTases in mammalian development has been demonstrated by the targeted mutation of mouse MTase. Embryonic stem cells (ES cells) homozygous for this mutation showed a 3-fold drop in 5-mC levels, yet were ostensibly normal in other respects. However, when this mutation was introduced to the germ line, the resultant homozygous embryos were developmentally retarded and died mid-gestation. Histological analysis of the embryos showed wide-spread cell

death and reduced proliferation. The underlying cause of cell death is not known, though a likely possibility is inappropriate gene expression (Li *et al.*, 1992).

1.5.3 Plant DNA methyltransferases

As mentioned in section 1.2, the DNA of higher plants contains 5-mC in both CpG and CpNpG sequences, and plants must therefore have a MTase or MTases specific for both targets. The CpNpG motif is also symmetrical, which would allow clonal inheritance of ^{m}CpNpG through maintenance methylation.

DNA MTases have been partially purified and characterised from a number of plant species. Purification of MTase from wheat embryo revealed two polypeptides of 50 and 30 KDa (Theiss *et al.*, 1987), whereas MTase purified from cultured rice cells had an apparent molecular weight of 54 KDa (Giordano *et al.*, 1991). Initial work in our laboratory on the purification of MTase from pea seedlings showed a single polypeptide of 160 KDa. This protein was very unstable and rapidly broke down in storage into smaller polypeptides which were still active (Yesufu *et al.*, 1991). It is this instability which may have accounted for the lower molecular weights quoted for the other plant MTases. The MTases isolated from the different plant sources all differed in their preferred substrates and levels of maintenance over *de novo* methylation. Both the wheat and the pea enzyme would methylate cytosines in the CNG trinucleotide but it was not clear whether CG sequences were also methylated. The study of plant MTases is complicated by the fact that there are two different specificities involved which may or may not reside in the same enzyme. The picture is further complicated by proteolytic breakdown which may alter the maintenance/*de novo* activities and, of course, the molecular weight determinations. The substrate specificity of the partially purified pea MTase was investigated further by using a variety of methylated and hemimethylated synthetic oligonucleotides as substrates. A strong preference for hemimethylated substrates (relative to the mammalian MTases) was observed using crude extracts. Further purification revealed

two peaks of activity with different relative activities for di- and tri-nucleotide substrates, and the two activities had different heat stabilities. This was the first indication that the two different specificities did reside in two separate MTases in plants (Houlston *et al.*, 1993). The two different activities in pea were subsequently purified by Pradhan and Adams (1995). They identified an enzyme of 140 KDa (the putative CpG MTase) which methylates only CG and CI dinucleotide sequences, and an enzyme of 110 KDa (the CpNpG MTase) which methylates only CAG and CTG sequences but not CCG or CGG sequences. Both the enzymes had maintenance and *de novo* activity. No evidence for the presence of a third enzyme activity specific for the outer cytosine of the CCG motif was found in any of the fractions (from genomic sequencing of oligonucleotides methylated *in vitro*), although the CG enzyme did show weak affinity for the outer cytosine in the CCG motif, when it was replaced by the analogue 2-pyrimidinone-1-β-D-2′-deoxyribofuranoside (d^4HC). (This analogue reacts with the enzyme to form a covalent adduct (Taylor *et al.*, 1993).) Although Gruenbaum *et al.* (1991) had presented data suggesting the presence of ^{m}CpCpG sequences in plant DNA, this has never been demonstrated directly in endogenous plant genes.

Two putative MTases have been cloned in *Arabidopsis thaliana*. The first of these, a cDNA clone, encodes a protein of ~150 KDa which has an N-terminal domain linked to a catalytic C-terminal domain. The C-terminal domain shares 50% homology with the mouse enzyme, and contains all the conserved motifs discussed earlier (section 1.5.2). There is also sequence homology in the variable region, except that that of *Arabidopsis* is shorter by about 40 amino acids. Sequence homology is much less in the N-terminal domain (24%), and is limited to short stretches. One of these is the sequence that targets the MTase to replication foci in mouse. The other is the motif S/TPXX (see section 1.5.2) which occurs five times in the *Arabidopsis* sequence. Unlike mouse MTase, the N-terminal domain of this putative MTase does not contain a zinc binding motif (Finnegan and Dennis, 1993). The evidence suggests that this cDNA encodes a DNA MTase capable of maintenance

(and possibly *de novo*) methylation, but it is not possible to predict whether this is a CpG or CpNpG MTase. The authors also found evidence for the existence of a family of related genes in *Arabidopsis*.

The other putative *Arabidopsis* DNA MTase clone was a genomic clone (Scheidt *et al.*, 1994a). The protein predicted lacks an N-terminal domain and, in that sense, is much more like the prokaryotic type II cytosine-C5 MTases. It was proposed that this MTase, because it lacked the modulating function and replication foci targeting sequence provided by the N-terminal domain, was a *de novo* MTase. The problem however, is that it also lacks a nuclear localisation signal which would be necessary for the enzyme to methylate genomic DNA. This putative MTase had very little sequence homology with the mouse catalytic domain, or that of the other *Arabidopsis* clone, even in the conserved motifs (with the exception of motif I). This paper, however, has since been retracted (Scheidt *et al.*, 1994b) because of falsification of sequence data of motif I by the first author. Subsequent cloning and sequencing of the cDNA corresponding to this genomic clone has revealed only "motif VI" (i.e. the sequence ENV) to be encoded (Nebendahl and Baumlein, 1995). The function of the protein encoded by this cDNA is unknown, but it is extremely unlikely to be a MTase.

1.5.4 Eukaryotic DNA methyltransferase homologues

Two different eukaryotic genes with homology to DNA MTases have been cloned recently. The first, *All-1*, was cloned from mouse and is related to the trithorax gene of *Drosophila*, which control the expression of homeobox genes. The predicted *All-1* protein is not a eukaryotic MTase homologue *per se*, but it does share the unusual Cys-rich zinc binding motif of the N-terminal domain of mammalian MTases. It also contains the AT rich minor groove binding motifs which are present in the N-terminal domains of both the mammalian and *Arabidopsis* MTases. Interestingly, the trithorax protein of *Drosophila* lacks this zinc-binding motif. As there is no 5-mC in

Drosophila DNA, it has been suggested that *All-1* codes for a DNA binding protein which may have the ability to distinguish methylated DNA (Ma *et al.*, 1993; Bestor and Verdine, 1994). As mentioned in section 1.5.2, the N-terminal domain of mouse MTase (which includes the Zn binding motif) inhibits *de novo* methylation, and must therefore be able to sense the methylation status of DNA. This lends support to the proposed role of this zinc-binding motif.

The second gene, *pmt1*$^+$, was cloned from the fission yeast *Schizosaccharomyces pombe*. The cDNA encodes a protein which contains all the conserved motifs of the type II cytosine-C5 MTases. However, there appears to have been an insertion of a serine residue before the catalytic cysteine residue of motif IV. The expressed protein is not able to methylate DNA *in vitro*; which is consistent with the fact that *S. pombe*, like other yeasts, has no 5-mC in its DNA. So far, the specificity of this enzyme has not been identified but it has been proposed to be an RNA methyltransferase (Wilkinson *et al.*, 1995). However, deletion of the serine residue resulted in the restoration of DNA MTase activity, specific for the sequence CC(A/T)GG (which is also the sequence modified by M.*Eco*RII) (Pinabarsi *et al.*, personal communication).

The study of DNA MTase homologues and pseudo DNA MTases may help towards understanding the evolution and functional significance of eukaryotic DNA MTases.

1.6 CpG islands

As mentioned earlier, 5-mC is only found in the dinucleotide CpG in vertebrate DNA. However, CpG sequences are under-represented in the vertebrate genome, occurring at only one fifth the expected frequency. Although an excision repair mechanism which operates at 90% efficiency exists to repair G/T mismatches back to GC (Wiebauer and Jiricny, 1989; Neddermann and Jiricny, 1993), the replication of the unrepaired product of 5-mC deamination has led to the depletion of CpG

sequences during vertebrate evolution. That this is the cause of CpG depletion is corroborated by the fact that TpG and CpA sequences (which would be expected to arise after replication of a TpG mismatch) are correspondingly over-represented in higher eukaryotic DNA (Sved and Bird, 1990).

However, there are localised regions of vertebrate genomes which have a G+C content of over 60% (in the case of warm-blooded vertebrates), and contain CpG sequences at the expected frequency. These regions contribute about 1–2% of the vertebrate genome and are known as CpG islands (Bird, 1986).These islands contain many *Hpa*II sites close together (*Hpa*II will cleave the sequence CCGG provided the second cytosine is unmethylated), and on digestion give rise to *Hpa*II tiny fragments (HTFs). These islands are further characterised by the fact that they are unmethylated in the germ line and in all somatic tissues except in the inactive X-chromosome in female mammals (section 1.8.1) and in imprinted genes (section 1.8.2). According to this model, it is their unmethylated status in the germ line which has led to their preservation throughout vertebrate evolution.

CpG islands are about 0.5–3 kb long and are mostly associated with the 5′ end of housekeeping genes; often covering the first few exons as well as upstream regions. However, some CpG islands have been found in the 3′ end of genes; for example the glucose-6-phosphate dehydrogenase gene (Wolf *et al.*, 1984). All housekeeping genes sequenced to date have been found to contain a CpG island. On the other hand, only about 40% of tissue specific genes or genes of limited expression have been found to be associated with CpG islands, and there seems to be no bias in the position of the island towards the 5′ end of these genes (Larsen *et al.*, 1992). It has been suggested that tissue specific genes that contain CpG islands represent a transitionary stage in the adaptation of a housekeeping gene to a tissue specific function (Edwards, 1990).

Abnormal methylation of a CpG island can have dramatic medical implications. For example, Fragile X syndrome, which is the most common inherited form of mental retardation, is linked to the abnormal methylation and unstable length

of a CpG island which maps to the breakage point of the fragile X (Bell *et al.*, 1991). It is known that the affected gene (*FMR 1*) is transcriptionally silent in affected individuals (Pieretti *et al.*, 1991), so it may be that CpG island methylation itself is responsible for the phenotype.

The involvement of *de novo* methylation of CpG islands associated with certain tumour suppressor genes, for example the retinoblastoma gene (Sakai *et al.*, 1991), has been implicated in tumourigenesis in some cancers. *De novo* methylation of CpG islands has also been observed in cell lines (Antiquera *et al.*, 1990). This *de novo* methylation has occurred in genes non-essential for growth in culture or in genes whose expression prevents immortalisation (Jones *et al.*, 1990), and is presumably a random event. In many cases, these 'epimutations' (Holliday, 1987) can be reversed by treatment with 5-azacytidine. However, as these islands are never methylated in normal cells in the organism, this phenomenon is of limited value as a model for understanding methylation-mediated gene inactivation.

To summarise: CpG islands could have evolved to 'mark out' promoter sequences and simplify their interaction with the transcriptional apparatus (Antequera *et al.*,1989). This is corroborated by the fact that only organisms with very large genomes have CpG islands (Antequera and Bird, 1993).

The question that remains however, is how do CpG islands remain methylation free. Since they become methylated on the inactive X-chromosome (Xi), it cannot be that they are not substrates for methyltransferase, and yet it has been shown that CpG islands remain methylation free once introduced into transgenic mice, regardless of their expression (Kolsto *et al.*, 1986). The somatic methylation pattern of an adult is established at implantation (Kafri *et al.*,1992), the methylation pattern from the gametes having been erased in the morula and early blastula (Monk *et al.*, 1987). At this stage of embryogenesis there must be a mechanism of distinguishing CpG island from non CpG island sequences so that CpG islands are protected from *de novo* methylation and/or are actively demethylated. The evidence suggests there is a combination of the two. Frank *et al.* (1991) demonstrated that an

in vitro methylated hamster adenine phosphoribosyltransferase (APRT) gene became demethylated at its CpG island in transgenic mice, whereas the 3′ end remained methylated. Transient transfection studies with the same construct showed this demethylation to be rapid, which implies an active enzymatic process, and to be a feature of embryonic cells. Such a demethylating activity (specific for hemimethylated CpG sites) has been identified in nuclear extracts of chick embryos and operates by an excision repair mechanism (Jost, 1993). Brandeis and co-workers found that the hamster APRT CpG island was protected against *de novo* methylation upon transfection into cultured ES cells (somatic cells are much less efficient at *de novo* methylation; Jahner and Jaenisch, 1985). Crucially, they also found that the presence of upstream Sp1 elements was necessary for both the prevention of *de novo* island methylation and the active demethylation of *in vitro* methylated islands. Moreover, the insertion of an Sp1 consensus element about 10–50 bp upstream of a *Hpa*II site in the non island containing β-globin gene induced substantially demethylation of that site upon transfection into ES cells (Brandeis *et al.*, 1994). As Sp1 sites are frequently associated with CpG islands, this could be the general mechanism by which CpG islands remain methylation free. Though Sp1 elements are involved with transcription, promoter activity is not essential for demethylation. Tissue specific genes containing CpG islands remain unmethylated at these islands in tissues where the genes are not expressed.

1.7 DNA methylation and transcriptional activity

It is now well established that DNA methylation inhibits the expression of many tissue specific and housekeeping genes (reviewed by Doerfler, 1983; Razin and Cedar, 1991). However, to date there is no clear cut answer as to the mechanism of transcriptional activation, or indeed whether a causal relationship exists at all *in vivo*. or whether methylation serves to maintain established states of inactivation. In some instances DNA methylation appears to act directly, affecting the binding of

transcription factors to promoter sequences. In others, DNA methylation may inhibit transcription indirectly by altering chromatin structure. These alternative mechanisms are discussed below.

1.7.1 DNA methylation affects the binding of transcription factors

The physical presence of a methyl group in the major groove of DNA can prevent transcription factor binding. This provides a mechanism whereby methylation of specific sites in promoter and enhancer elements can inhibit transcription. Binding can be inhibited by eliminating specific protein-DNA contacts or by altering local DNA geometry (Molloy and Watt, 1990).

A number of transcription factors whose binding is inhibited by methylation of their target sequences have been identified (reviewed by Tate and Bird, 1993). These include MLTF (Molloy and Watt, 1990), c-Myc/Myn (Prendergast *et al.*, 1991), the cAMP responsive element (CRE) binding protein (Iguchi-Ariga and Schaffner, 1989), the activating protein 2 (AP2) (Comb and Goodman, 1990; Hermann and Doerfler, 1991) and NF-κB (Bednarik *et al.*, 1991). In most of these cases, methylation was found to be associated with lack of factor binding (in gel retardation experiments and *in vivo* footprinting), and with transcriptional inactivity (transient expression of reporter gene constructs and *in vitro* transcription assays).

However, in the case of the rat tyrosine aminotransferase (TAT) gene, though lack of binding of ubiquitous transcription factors to upstream elements *in vivo* was found to be associated with methylation of these sites, treatment with 5-azacytidine did not induce factor binding. In this case, the nucleosomal organisation of the potential binding sites suggested that chromatin structure was the dominant determinant of transcription factor binding (Weih *et al.*, 1991). Other transcription factors such as Sp1 (Harrington *et al.*,1988; Bryans *et al.*, 1992) or CTF (Ben Hattar *et al.*, 1989) are insensitive to methylation of their binding sites.

Clearly additional, more general mechanisms are involved in methylation-mediated gene repression.

1.7.2 Methylcytosine binding proteins

The alternative mechanism by which methylation can inhibit gene expression is through the binding of proteins specific for methylated DNA which in turn prevent the binding of transcription factors. Several of these methylated DNA binding proteins have been identified in vertebrates (Meehan *et al.*, 1989; Lewis *et al.*, 1992; Jost and Hoffsteenge, 1992) and in higher plants (Zhang *et al.*, 1989; Ehrlich, 1992; Ehrlich *et al.*, 1992) and are referred to as methylated DNA binding proteins (MDBPs), methylated cytosine binding proteins (MeCPs) or DBP-m in the case of the plant protein.

MDBP-2 was first identified as a down-regulating transcription factor for the chicken vitellogenin gene (Pawlak *et al.*, 1991). The purified protein requires a minimum 30 base pairs and the presence of one symmetrically methylated cytosine for binding *in vitro*. Interestingly, sequence analysis revealed MDBP-2 to be a member of the histone H1 family (Jost and Hoffsteenge, 1992).

MeCP-1 binds to a minimum of 12 symmetrically methylated CpGs (^{m}CpGs) and can repress the transcription of methylated promoters *in vitro* and in transfected cells. The extent of repression depends upon the density of ^{m}CpGs and on the promoter strength (Boyes and Bird, 1991; Boyes and Bird, 1992).

MeCP-2, a much more abundant protein, will bind to sequences containing only one symmetrically methylated CpG and is associated with heterochromatin (Lewis *et al.*, 1992). However, MeCP-2 does not preferentially inhibit transcription of methylated genes *in vivo* (Meehan *et al.*, 1992). MeCP-2 may be involved in the genome-wide protection of ^{m}CpGs against nucleases observed by Antiquera *et al.* (1989).

The current model hypothesised by Bird and co-workers is that MeCP-1

competes with transcription factors for binding to methylated DNA. Binding of MeCP-1 then guides the DNA into a heterochromatic structure which is stabilised by association with MeCP-2 (Meehan *et al.*, 1992). This model can only be tested however, once purified MeCP-1 becomes available. Nevertheless, the central importance of MeCP-2 in mammalian development has been demonstrated by knockout experiments. Mice lacking both MeCP-2 alleles die very early on in embryogenesis and show developmental abnormalities similar to those exhibited by the hypomethylation mutant embryos described in section.1.5.2 (Bird, personal communication). This finding is in agreement with the proposed role of MeCP-2 as a mediator of the effects of DNA methylation on the developmental regulation of gene expression. It would be interesting to see if plants have analogous MeCPs which bind to ^{m}CpGs and even ^{m}CpNpGs and which may play a similar role.

1.7.3 DNA methylation and chromatin structure

Transcriptonally active euchromatin is more susceptible to nuclease digestion than inactive heterochromatin as a result of its more open structure. It is considered that an open chromatin structure is necessary for initiation of transcription (Adams and Workman, 1993). The first indication that methylated DNA was in a specific chromatin structure came from the work of Razin and Cedar (1977). Mouse nuclei were digested with micrococcal nuclease and the solubilised chromatin analysed for 5-mC content. After extensive digestion, 50% of the chromatin remained unsolubilised and contained 70% of the total 5-mC. Subsequently, Solage and Cedar (1978) showed that 5-mC is released more slowly than thymidine, and does not start to appear until 15% of the DNA had been solubilised. Thus 5-mC is organised in chromatin structures that are relatively resistant to nuclease digestion. Keshet *et al.* (1986) provided direct evidence that methylation of DNA leads to the formation of a closed chromatin structure by transfecting mouse L cells with methylated constructs which were integrated into the genome. The methylated constructs adopted a DNase I

resistant structure characteristic of inactive chromatin. Nuclease insensitivity has been observed for a variety of methylated genes (Antequera *et al.*, 1989; Levine *et al.*, 1991) and it has been reported that a closed chromatin structure is associated with transcriptional inactivity. Furthermore Buschhausen *et al.* (1987) demonstrated that transcriptional inhibition of a methylated thymidine kinase (TK) construct occurred 8 h after injection into *Xenopus* oocytes, corresponding with the assembly of the transfected constructs into chromatin. If the construct was assembled into chromatin prior to injection, inhibition was immediate.

Since CpG islands usually contain the promoters of housekeeping genes (section 1.6) which are by definition transcriptionally active, CpG islands should represent open chromatin. This has shown to be the case with two X-linked genes, hypoxanthine phosphoribosyltransferase *(hprt)* and glucose 6-phosphate dehydrogenase *(g6pd)*, whose GpG islands are hypersensitive to nucleases such as *Msp*I, DNase I and S1 nuclease on the Xa (Kerem *et al.*, 1983). Analysis of CpG island chromatin revealed it to be much less abundant in H1 (about 10%) compared to bulk chromatin. In addition, histones H3 and H4 were found to be highly acetylated compared to those of bulk chromatin (Tazi and Bird, 1990). Hebbes *et al.* (1988) had previously demonstrated enrichment of actively transcribed regions upon immunoprecipitation with an anti-acetylated H4 antibody. On digestion of CpG island chromatin fractions with micrococcal nuclease, a sub-nucleosomal sized fraction was identified. Southern hybridisation with *c-fos* revealed that a region of the promoter of this gene belonged to this fraction, and must therefore be lacking a nucleosome (Tazi and Bird, 1990). This region had previously been identified as a DNase I hypersensitive site. In a detailed study by Pfeifer and co-workers, the CpG island of the human X-linked phosphoglycerate kinase *(pgk)* gene on the Xa was found to be completely unmethylated, free of nucleosomes and showed footprints for putative transcription factors on the Xa, whereas the same sequence on the Xi was shown to be wrapped around two nucleosomes and had no evidence for transcription factor binding (Pfeifer *et al.*, 1990a; Pfeifer *et al.*, 1990b; Pfeifer and Riggs, 1991; Pfeifer

et al., 1991; Pfeifer *et al.*, 1992).

Heterochromatin can be constitutive or facultative. Constitutive heterochromatin often contains sequences that are highly repetitive, whereas facultative heterochromatin arises from inactivation of previously euchromatic regions. The best example of facultative heterochromatin is the inactive X chromosome (Xi) in female mammals. Both the mouse major satellite DNA and the Xi are extensively methylated, heterochromatinised and replicate late in S phase. Treatment with 5-azacytidine brings about earlier replication both of genes on the Xi (Jablonka *et al.*, 1985) and of the mouse major satellite DNA (Selig *et al.*, 1988). The temporal separation of replication of hypermethylated heterochromatinised DNA could result in a difference in cellular environment and thus could be involved in maintenance of differential chromatin structure and methylation levels.

In an attempt to understand the higher order structures responsible for nuclease resistance of methylated DNA, Ball *et al.* (1983) fractionated nucleosomes on agarose gels following micrococcal nuclease treatment. They found that the histone H1 associated nucleosome fraction was significantly enriched for 5-mC compared to nucleosomes lacking H1 and concluded that association of H1 with nucleosomes contributed to nuclease insensitivity. However, some *in vitro* binding studies have not shown preferential binding of H1 to methylated DNA but that binding results in a change in conformation, rendering the DNA insensitive to *Msp*I digestion. This does not happen with unmethylated DNA (Higurashi and Cole 1991). Johnson *et al.*(1995) also showed no preferential binding of H1 to methylated DNA, but that H1 preferentially inhibits *in vitro* transcription of methylated templates. In contrast, other workers have found that H1 does preferentially bind methylated DNA (Levine *et al.*, 1993). This is supported by the work of Jost and Hoffsteenge (1992) on H1 related MDBP-2 (see previous section) who also showed that H1 is a more efficient inhibitor of transcription from a methylated template than an unmethylated template.

Kass *et al.*, (1993) transfected *in vitro* patch-methylated plasmid constructs

containing the CAT reporter gene into cultured cells, and found that the methylated patch was able to initiate the spreading of chromatin and subsequent transcriptional inactivation of the reporter gene, even when the methylated patch did not include the reporter gene, and that the degree of inhibition was dependent on the size of the patch. This spreading effect has also been observed by C. Johnson, working in our laboratory, on chromatin assembled *in vitro* on the same constructs.

The association of other proteins with methylated DNA was implicated by the work of Antequera *et al.* (1989) who observed protection from digestion with *Msp*I and *Pth*I of bulk chromatin. These enzymes have a CpG in their recognition sites and will cut irrespective of its methylation. In contrast, enzymes which do not have a CpG in their recognition sequence cleave bulk chromatin extensively. The interpretation of these findings was that there was an associated protein specific for mCpG. Two such proteins have been identified (MeCP-1 and -2), and have been proposed to be the mediators of the effects of DNA methylation on transcription and chromatin structure (see section 1.7.2), though *in vitro* studies have shown DNA methylation to have an effect on transcription when nucleosomes are assembled in the absence of these proteins (Englander *et al.*, 1993).

1.8 DNA methylation and development

It has been proposed that DNA methylation is an important epigenetic mechanism involved in the developmental program (reviewed by Holliday, 1987; Cedar and Razin, 1990). That DNA methylation is absolutely essential for mammalian development has been demonstrated by Li *et al.*(1992) (see section 1.5.2). Monk *et al.* (1987) compared the genome-wide methylation levels between germ cells and embryos of mice by digestion with *Msp*I and *Hpa*II. They observed that oocyte DNA is undermethylated relative to sperm DNA and that the level of methylation in the 8-cell embryo was the average of the maternal and paternal genomes. This inherited pattern of methylation is lost between the 8-cell and blastocyst stages, resulting in

undermethylation of both extraembryonic and embryonic tissues (Howlett and Reik, 1991; Kafri *et al.*, 1992). Paradoxically, the level of nuclear MTase is at its highest at the 8-cell stage, although it does not appear to be localised to replication foci. The MTase present at this stage of development is the smaller form which is inherited from the oocyte (see section 1.5.1). However, since it has the same *in vitro* activity as somatic MTase, an interaction with other factors which may modify its activity in the embryo implicated (Carlson *et al.*, 1992).

It is possible that demethylation is necessary in order to establish somatic methylation patterns. *De novo* methylation and X-chromosome inactivation occur around the time of implantation. Jahner and Jaenisch (1984) demonstrated that implantation embryos were capable of *de novo* methylation but this ability did not extend into postimplantation stages. *De novo* methylation occurs just before gastrulation and leads to the hypermethylation of non-island DNA whereas island DNA remains unmethylated (Kafri *et al.*, 1992). It is possible that CpG islands are both actively demethylated and protected from *de novo* methylation by virtue of the fact that they contain multiple Sp1 elements (Brandeis *et al.*, 1994), as discussed in section 1.6. The second step involves the establishment of tissue specific and gene specific methylation patterns, and occurs in late embryonic development. A well documented example is the muscle-specific demethylation of the α-actin gene. Yisraeli *et al.* (1986) transfected *in vitro* methylated α-actin constructs into cultured cells, and found that the promoter was specifically demethylated and actively transcribed in myoblasts but not in fibroblasts. Thus muscle cells have the ability to recognise and activate by demethylation this muscle-specific gene. The demethylation was found to be active and dependent on the presence of cis-acting sequences (Paroush *et al.*, 1990). The establishment of tissue-specific methylation patterns appears to develop by a wave of demethylation followed by methylation in early embryogenesis,which is then followed by specific demethylation in later stages of development. Methylation may therefore have a role to play in the coordination of gene expression in a tissue-specific and temporally specific manner during

development. Once established, somatic methylation patterns can then be clonally inherited through many cell divisions through maintenance methylation.

1.8.1 X-Chromosome inactivation

X-chromosome inactivation is a mechanism of gene dosage compensation in female mammals. One of the X-chromosomes (which is randomly assigned in the case of the eutherian mammals) becomes inactivated in each cell of the embryo, around the time of *de novo* methylation. The inactivation is characterised by the heterochromatinisation of the inactive X chromosome (Xi), which is recognisable in interphase nuclei as the densely staining Barr body. The Xi is further characterised by the fact that its CpG islands are hypermethylated. A number of studies on the CpG islands of X-linked genes such as *g6pd* (Wolf *et al.*, 1984; Wolf and Migeon 1985; Toniolo *et al.*, 1988) and *hprt* (Lock *et al.*, 1986; Sasaki *et al.*, 1992) have shown that the CpG islands are hypermethylated and nuclease insensitive on the Xi, and that treatment of the Xi with 5-azacytidine could bring about nuclease sensitivity. Inactivation is thought to occur in three stages: initiation of inactivation, the spreading of the inactivation along the length of the chromosome and the maintenance of the inactive state of the chromosome. However, most of the data so far give evidence for the involvement of DNA methylation in the maintenance of the inactive X chromosome. The role of DNA methylation as the initial signal for the spreading of heterochromatinisation is disputed by Lock *et al.* (1987), who found that heterochromatinisation of the Xi preceded the methylation of the CpG island of *hprt* by several days. Heterochromatinisation is initiated at a region known as the X chromosome inactivation centre (known as XIC in humans) which maps to Xq13 (Brown *et al.*, 1991a). A gene (*XIST*) has been identified which maps to the XIC and, in contrast to other X-linked genes, is unmethylated and transcriptionally active only on the Xi (Brown *et al.*, 1991b). The *XIST* cDNA is 17 kb long, includes several tandem repeats and does not appear to code for a protein, which suggests that

the RNA has some structural function. Evidence for this was provided by *in situ* hybridisation experiments that showed *XIST* RNA to be localised in the Barr body (Brown *et al.*, 1992). Thus, the association of *XIST* RNA may initiate X-chromosome inactivation through some as yet unknown mechanism which leads to heterochromatinisation, a state that is subsequently maintained by DNA methylation. However, it could be merely the fact that the gene is transcriptionally active, and therefore in an open chromatin structure, that prevents methylation (see section 1.5.1), and the spread of heterochromatinisation throughout the X chromosome. The work of Kass *et al.*, (1993) indicated that methylated DNA is able to initiate the spreading of chromatin into unmethylated regions (see previous section), and this at least points to a possible mechanism for the spreading of heterochromatinisation from the XIC.

1.8.2 Imprinting

During development, functional differences of maternal and paternal chromosomes in terms of differential gene expression are conferred by genomic imprinting. This may prevent either the female or male genome becoming totipotent by itself so that both germ lines are are required for normal mammalian development (reviewed by Surani *et al.*, 1990). Although little is known of the molecular basis of the imprint, it has been suggested that the mechanism must fulfil the following criteria. Firstly, the primary imprint must be laid down during gametogenesis as this is the only time at which male and female genomes are separated and can therefore be differentially imprinted. Second, this imprint must be clonally inherited which means that the imprint must be re-imposed following DNA replication. Thirdly, re-programming must be possible, so that for example a maternally derived gene can be given a male-specific imprint in the male germline. Finally, the imprint must affect the transcription of the gene, so that parental-specific expression of the locus is achieved. Methylation can in theory fulfil all the above criteria.

Methylation has been shown to be involved in the imprinting of X chromosomes in female eutherian mammals (in extra-embryonic tissues only), where the paternally-derived X chromosome is preferentially inactivated (Takagi and Sasaki, 1975). The site of the primary imprint appears to be the *XIST* gene as expected. Recently, the murine equivalent of the *XIST* gene (*Xist*) has been shown to be unmethylated in sperm at CpG sites at its 5´ end. In contrast, the same sites on the maternal allele are initially methylated on the oocyte, and remain modified in the early embryo; the differential parental methylation patterns persisting to the blastocyst stage when the paternal X chromosome inactivation occurs. In the male germ line demethylation of these sites occurs during spermatogenesis, an event that precedes reactivation of *Xist* in the testis. This presents a clear cut example of a potential methylation imprinting signal that is reprogrammable and gamete derived (Zuccotti and Monk, 1995; Ariel *et al.*, 1995).

Most of the data concerning methylation and autosomal genomic imprinting comes from studying the methylation patterns of transgenes in mice. These transgenes are differentially methylated in a parental origin dependant manner and the methylation pattern can be switched upon passage through the germ-line of the opposite sex (Sapienza *et al.*, 1987; Reik *et al.*, 1987).

With the isolation of endogenous imprinted genes from mouse, it became possible to directly tackle the molecular mechanism of genomic imprinting. For example, the insulin-like growth factor II receptor gene (*Igf2r*), is expressed only on the maternally inherited chromosome (Barlow *et al.*, 1991), whereas the insulin-like growth factor II gene (*Igf2*) gene is only expressed on the paternal chromosome (DeChiara *et al.*, 1991). Biallelic expression of the *IGF2* gene in humans is associated with Beckwith-Weidemann syndrome (a congenital overgrowth syndrome). The CpG island associated with the promoter of the *Igf2r* gene was found to be methylated only on the inactive paternal allele (Stoger *et al.*, 1993). However, the promoter-associated CpG island of the *Igf2* gene is unmethylated and DNase I hypersensitive on both the maternal and paternal alleles. A link has been

suggested between the reciprocally imprinted *H19* gene, which is situated 90 kb away from the *Igf2* gene. Li *et al.* (1993) followed the transcriptional activity of the three genes in methylation-deficient mutant embryos (Li *et al.*, 1992). Though they found that the normally inactive paternal allele of *H19* was active, the normally active alleles of the other two genes were inactive.

It appears that in the case of *Igf2*, paternal methylation of the *H19* gene forms the primary imprint, as targeted deletion of *H19* results in biallelic expression of *Igf2* in mice that have inherited the mutation from their mother, but not their father (Leighton *et al.*, 1995). A pattern of differential methylation of the 5′ end of *H19* was found to be established in male and female gametes, a subset of which is maintained in the preimplantation embryo, and it has been suggested that paternal specific methylation of the far 5′ end of *H19* forms the primary imprint (Tremblay *et al.*, 1995). *H19*, like the *XIST* gene, codes for an untranslatable RNA. An enhancer competition model has been suggested (Leighton *et al.*, 1995), whereby *H19* and *Igf2* share enhancer elements and inactivation of the (perhaps stronger) *H19* promoter by imprinting results in the enhancer driving *Igf2* expression.

1.9 DNA methylation in higher plants

Though the study of the DNA methylation in higher plants is a relatively young field, there are data accumulating that support the notion that it is involved in the regulation of gene expression (reviewed by Finnegan *et al.*, 1993).

Patterns of methylation in higher plants are far from random. For example unique sequences which comprise ~38% of the cotton genome, contain only 4% of the total 5-mC (Guseinov *et al.*, 1975). This is consistent with the proposed role of methylation in the suppression of recombination of repetitive sequences (see section 1.3). Furthermore, even though 80% of cytosines in CpG dinucleotides are modified, a number of plants have been shown to have a HTF fraction in their DNA (see section 1.6), such as maize tobacco and wheat (Antequera and Bird, 1988) and

also tomato (Messeguer et al, 1991), suggesting that plants also have CpG islands. Unmethylated CpG islands containing DNase I hypersensitive sites have been found to be associated with the 5′ end of several genes such as the alcohol dehydrogenase (*Adh* 1) and the sucrose synthase (*Sh* 1) genes of maize (Antequera and Bird, 1988). In addition, the CpG island of *Adh* 1 remained methylation free at the CpG and CpNpG sites tested, even in leaf tissue where it is not expressed (Nick *et al.*, 1986). Antequera and Bird (1988) postulated that CpG islands have arisen independently in plants and vertebrates because they do not share a common ancestor that has CpG islands. Like the situation in vertebrates, there is under-representation of CpG sequences (though to a lesser extent) and a corresponding over-representation of TpG and CpA sequences in the DNA of higher plants (Gardiner-Garden *et al.*, 1992). This is consistent with the proposed mutagenicity of 5-mC. No evidence was found for CpNpG depletion however, and it has been proposed that a mechanism to correct for deamination of 5-mC evolved prior to the occurrence of methylation in the CpNpG trinucleotide (McClelland, 1983). Belanger and Hepburn (1990) suggested that this was unlikely since their phylogenetic studies with regard to methylation patterns indicated that CpNpG methylation is at least as old as CpG methylation in vascular plants. It is now thought that plants must have a much more efficient mismatch repair system, particularly with regard to the CpNpG methylating system (Belanger and Hepburn, 1990; Gardiner-Garden *et al.*, 1992). The alternative hypothesis is that cell lineages that contribute to germ cells in plants may have lower levels of 5-mC (Gardiner-Garden *et al.*, 1992).

1.9.1 DNA methylation and transcriptional regulation in plants

Direct evidence for the inhibitory effect of cytosine methylation on transcription in plants comes from experiments which follow the transient expression of an *in vitro* hemimethylated construct, consisting of the cauliflower mosaic virus *35S* promoter linked to a reporter gene, upon transfection into protoplasts of either tobacco or

petunia (Hershkovitz *et al.*, 1989; Hershkovitz *et al.*, 1990; Weber and Graessman 1989). These experiments indicate that to inhibit gene expression the level of methylation must exceed a certain threshold and/or cover specific sites. However, methylation was not restricted to the promoter so the effect of methylation of critical sites within the promoter could not be studied. The work of Hershkovitz *et al.* (1990) indicated that CpG hemimethylation alone was sufficient to suppress transcription, suggesting that CpNpG methylation could have an unrelated function, but this might not reflect the true situation *in vivo*. The methylation status of the integrated construct in stably transformed tobacco callus was examined; the *in vitro* hemimethylated construct was shown to retain methylation only at CpG and CpNpG sites and in general, hypermethylation of the construct was associated with its transcriptional repression. This effect could be reversed by treatment with 5-azacytidine (Weber *et al.*, 1990).

A number of plant proteins with differential binding affinities for methylated and unmethylated DNA have been identified. Most of the proteins identified so far bind DNA in a sequence-specific manner, and are probably methylation-sensitive transcription factors. For example, CG-1 is a DNA binding factor found in tobacco nuclear extracts which binds a sequence containing a CACGTG motif in the promoter of *Antirrhinum* chalcone synthase. Binding was inhibited by *in vitro* methylation of the target sequence (Staiger *et al.*, 1989). Other sequence-specific factors have been identified in nuclear extracts of pea, wheat, cauliflower and soybean which bind to the CRE consensus (see section 1.7.1), which is also found in the promoters of several plant genes. Binding of the factors was inhibited by the methylation of a single cytosine in the element (Inamdar *et al.*, 1991). The cDNA of one such protein (VBP-1) was cloned from a broad bean expression library using the CRE consensus as a probe (Ehrlich *et al.*, 1992). A sequence unspecific methyl-binding protein, DBP-m has also been identified by gel-shift assays on pea nuclear extracts. This protein binds to as little as one methyl cytosine, (Zhang *et al.*, 1989), has an apparent molecular weight of 50 kDa (Ehrlich, 1993) and may be analogous to mammalian

MeCP-2. As yet the effect of DBP-m binding on transcription has not been tested. However, DBP-m is not distinguishable from possible pea DNA MTases or their proteolytic breakdown products using this method.

1.9.1.1 DNA methylation of genes regulated in a tissue specific or inducible manner

A number of plant genes whose expression correlates with their extent of methylation have been identified including seed storage protein genes of maize, barley and bean (Bianchi and Viotta, 1988; Sorensen, 1992; Riggs and Chrispeels, 1990), photosynthetic genes in the bundle sheath and mesophyll cells of the C4 plant, maize (Ngernprasirtsiri *et al.*, 1989; Langdale *et al.*, 1991), ripening specific genes of tomato (Hadfield, *et al.*, 1993) and rRNA genes in pea (Watson *et al.*, 1987). Moreover, for the M1 zein gene family, evidence was found for this hypomethylation to be independent of transcription factor binding (Bianchi and Viotta, 1988). Riggs and Chrispeels (1990) have demonstrated using restriction digestion that in the cotyledons of *Phaseolus vulgaris*, the expression of seed storage proteins encoded by the lec 1 and lec 2 is correlated with tissue specific demethylation of sites upstream of lec 1, though no cotyledon specific nuclear proteins bind to these sites (Riggs *et al.*, 1989). Demethylation of a single *Pvu*II site 3.3 kb upstream of the phosphoenolpyruvate carboxylase (PEPCase) gene was found to correlate with induction of its transcription in mesophyll cells during greening. The cytosine involved is within a CpTpG trinucleotide, suggesting a role for CpNpG methylation in transcriptional regulation, though as yet there is no known function for this site (Langdale *et al.*, 1991). Sorensen (1992) analysed in detail (by genomic sequencing) the promoters of B-hordein genes. They identified 7 CpG sites which were unmethylated in endosperm where the genes are transcribed, but methylated in leaf tissue. No methylation was observed at CpNpG sites within the promoter in either endosperm or leaf, so the role of CpNpG methylation in transcriptional inhibition

remains controversial. No direct proof has been shown for active demethylation to be taking place in any of these instances, and it is not known at present whether an active demethylating system exists in plants. As with vertebrates, a number of tissue specific or inducible genes have been found in plants which do not show a correlation between methylation and gene expression.

Plants differ from animals in that they can respond developmentally to a wide range of environmental stimuli. One such response in which methylation is implicated is the promotion of early flowering by cold (vernalisation) in *Arabidopsis* which can be mimicked by treatment of imbibed seeds with 5-azacytidine. Both cold treated and 5-azacytidine treated plants show decreased levels of methylation. A candidate gene for the control of early flowering which shows a decrease in methylation upon cold treatment is kaurenoic hydroxylase which is involved in giberellic acid biosynthesis (Burn *et al.*, 1993).

To investigate the importance of DNA methylation in plant development, *Arabidopsis* hypomethylation mutants were isolated, identifying the locus *DDM1* to be important in cytosine methylation (Vongs *et al.*, 1993). These mutants had a 70% reduction in 5-mC levels yet showed no gross developmental abnormalities, unlike those observed in mutant mouse embryos (see section 1.5.2). This could be explained by the fact that plants are much more developmentally plastic than vertebrates, with cell differentiation continuing throughout the life of the plant and terminal differentiation being rare (reviewed by Goldberg *et al.*, 1989). However, more detailed analysis of the mutants did reveal altered leaf morphology and delayed flowering; perhaps a greater reduction in 5-mC would lead to a more severe phenotype. Nuclear extracts of the mutants showed the same methyltransferase activity as those of wild type plants, and the levels of AdoMet were the same. It was therefore suggested that the hypomethylation phenotype stems from alteration of the regulation of MTase, the mechanism of which may be investigated once the *DDM1* gene has been cloned (Kakutani *et al.*, 1995).

1.9.1.2 Methylation and inactivation of transposons and foreign transgenes

There is a correlation between loss of activity (i.e. ability to transpose) and methylation in all the transposable elements that have been studied in detail. For example, inactivation of the autonomous transposable element Activator (*Ac*) of maize is correlated with methylation of cytosines within several *Hpa*II sites.at the 5′ end of the transposase gene. Moreover, the methylation is specific to the element and does not extend into flanking plant DNA (Schwartz and Dennis, 1986). The mechanism of this selectivity is unclear as yet, and may be a consequence of high element copy number, or the disruption of chromatin structure caused by transposition. It is possible that plants have evolved a mechanism for silencing transposable elements to maintain genome stability.

The variability observed in the expression of transgenes between independent transformants has been ascribed to transgene copy number and positional effects, but these factors alone are not sufficient to explain transgene silencing. The site of integration has been shown to be important in transgene silencing in some cases; integration of the maize A1 gene into a highly repetitive (heterochromatic) site in *Petunia hybrida* resulted in hypermethylation and complete silencing of the transgene, whereas integration into a hypomethylated (euchromatic) region resulted in high expression of the transgene (Prols and Meyer, 1992). A number of other studies have shown a correlation between methylation of transgenes introduced via *Ag robacterium tumifaciens* T-DNA and their transcriptional repression (Peerebolte *et al.*, 1986; John and Amasino, 1989, Klaas *et al.*, 1989; Hobbs *et al.*, 1990). In most of these cases, multiple copies of the T-DNA were integrated. Treatment with 5-azacytidine was shown to reactivate silenced transgenes. Linn *et al.* (1990) introduced the maize A1 gene under the control of the CaMV *35S* promoter into the white flowering mutant *Petunia hybrida*. The plants produced had a range of flower colours from white through variegated to brick red. The majority of the white

flowered plants had multiple copies of the transgene whereas plants with brick red flowers usually had one intact copy. Two *Hpa*II sites within the *35S* promoter were found to be methylated in the white flowered plants indicating that methylation was responsible for transgene silencing. The work of Matzke *et al.* (1989) and Matzke and Matzke (1990;1991) showed suppression of genes encoded by an integrated T-DNA in transgenic tobacco, when a second T-DNA carrying different selectable markers was also integrated. This suppression was dependent on promoter methylation, and could be a consequence of sequence homology of the T-DNA and promoter regions of the two different transgenes. Because only 15% of the double transformants exhibited this phenomenon, the proximity of the transgenes to each other was thought to be an important factor. In addition, the integration of transgenes is sometimes associated with the inactivation of the endogenous gene (a phenomenon called co-suppression) (Napoli *et al.*, 1990; van der Krol *et al.*, 1990).

In conclusion, plants can recognise and inactivate introduced DNA and perhaps the mechanisms evolved to protect plant genomes from the effects of multiple insertions of transposable elements or T-DNA, also inactivate artificially introduced transgenes. The possibility that methylation of transgenes and integrated viral DNA in animals represents an 'immune' response to foreign DNA has been discussed (Bestor, 1990; Doerfler, 1991). However, it has been suggested that this is an unlikely explanation for all the results with plant transgenes, because inactivation can depend on the site of integration, and sometimes requires the insertion of an additional transgene copy (Matzke *et al.*, 1994).

Several models have been proposed to explain the silencing of homologous genes in plants. Matzke *et al.*, (1994) have suggested a pairing mechanism in which the inactive (hypermethylated) state of one transgene is imposed onto its active (hypomethylated) partner. This mechanism may also be responsible for the inactivation of endogenous allelic A1 genes in petunia, a phenomenon known as paramutation (Meyer *et al.*, 1993). An analogy has been drawn between the silencing of transgenes and homologous gene silencing in pre-meiotic *Neurospora crassa* (a

multicellular fungus) described by Selker *et al.* (1987). This process, known as repeat-induced point mutation (RIP), involves the detection of duplicated segments, perhaps through a pairing mechanism, which are then riddled with C to T point mutations (possibly by active deamination of methylated cytosines). The segments become extensively methylated and are rendered permanently inactive. Both closely linked and unlinked copies can be 'ripped', but the latter are ripped less frequently. Ripping could represent a mechanism to select against gene duplications in *N. crassa*. (Selker, 1990; see section 1.3). Another feature of genes silenced by ripping is that they are densely methylated, in both symmetrical and non-symmetrical sequences (Selker *et al.*, 1993). Methylation in non-symmetrical sequences has also been reported in transgenic *Petunia hybrida*, as mentioned in section 1.5.1. However, silencing of transgenes in plants is not associated with the introduction of point mutations, and may be more analogous to the silencing via methylation alone of homologous genes in another fungus, *Ascobolus immersus* (a phenomenon known as methylation induced pre-meiotically (MIP)), or the 'quelling' phenomenon (Romano and Macino, 1992) of *N. crassa*. Quelling involves the reversible silencing of an endogenous gene which occurs in mitotic cells after introduction of a homologous transgene, a process that does not involve transcription of the transgene. However, it may be the imposition of the chromatin structure itself that is responsible for both ectopic and allelic silencing, which would be analogous to dominant position-effect variagation in *Drosophila* (Henikoff and Dreesen, 1989).

Another model that has been suggested to be responsible for transgene inactivation and co-suppression is RNA-directed methylation. It has been shown that methylation of integrated viroid cDNA in transgenic tobacco occurs only after autonomous viroid RNA-RNA replication has taken place in these plants. Moreover, the methylation is restricted to viroid sequences only, which suggests that specific interaction between the transcripts and the integrated cDNA is able to direct methylation of the cDNA (Wasseneger *et al.*, 1994). A similar model has been suggested to account for X-chromosome inactivation by the *XIST* RNA in mammals

(see section 1.8). Thus, the ability to recognise and methyltate unusual DNA structures may be a general feature of eukaryotic MTases.

1.10 Summary

Despite many years of research, the role of DNA methylation in higher eukaryotes is not fully understood. In general DNA methylation is correlated with gene silencing in both vertebrates and higher plants, though it remains to be established whether methylation initiates the inactivation of genes or whether it is a secondary event which acts to maintain transcriptional inactivation. The discovery that normal levels of methyltransferase and the methylated binding protein, MeCP-2 are crucial for mammalian development, at least points at a general mechanism whereby methylation inhibits transcription, though there may be specific instances where the binding of transcription factors is affected directly.

Only one methyltransferase activity, specific for CpG sequences has been identified in mammals, and both the mouse and the human MTase cDNAs have been cloned. Both of these clones share sequence homology in their C-terminal domains with the prokaryotic type II cytosine-C5 MTases. Plants on the other hand, methylate cytosines in both CpG and CpNpG sequences, and evidence has been found to suggest that two different enzymes are responsible for the different specificities.

1.11 Project aims

Our laboratory has been working for several years on methylase from *Pisum sativum* (pea) (Yesufu *et al.*, 1991; Houlston *et al.*, 1993; Pradhan and Adams, 1995). This organism is a good model system for the study of the enzymology of DNA methylation in higher plants, as high activities of methylase can be extracted. This is in contrast to *Arabidopsis*, where levels of methylation are low (~4%), and from which very low levels of methylase activity have been extracted (Kakutani *et al.*,

1995). Indeed, purification of methylase from *Arabidopsis* has never been attempted. At the beginning of this project, no plant methyltransferases had yet been cloned, and it was not known whether two seperate enzymes were responsible for CpG and CpNpG methylation in plants. The aim of this project was to clone a MTase cDNA from pea and to investigate whether two different MTase mRNAs exist, which could possibly code for enzymes with different target specificities. The data obtained could then be related to the enzymological work to give a meaningful picture of DNA methylation in higher plants.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 List of suppliers

Unless otherwise stated, all chemicals were Analar grade, supplied by BDH Chemicals, Poole, Dorset, or Fisons Scientific, Loughborough, Leics. All radiochemicals were purchased from Amersham International plc. Growth media were supplied by Difco Laboratories, Detroit, USA. Pea seeds (var. Feltham First) were obtained from Brooker Seeds Ltd, Sleaford, Lincs. Where chemicals, kits or equipment were obtained from other sources, this is indicated in the text. A list of suppliers of special reagents is given below:

Aldrich Chemicals Co., Gillingham, Dorset

Amersham International plc., Amersham, Bucks.

Bochringer Corporation (London) Ltd., Lewes, E. Sussex

Fuji Photofilm Co., Japan

Gibco/BRL (Bethesda Research Laboratories) Ltd., Paisley

IBI Ltd., Cambridge

Kodak Ltd., Kirby, Liverpool

Pharmacia Ltd., Milton Keynes

Promega Ltd., Southampton

Sigma chemicals Co., Poole, Dorset

2.1.2 Bacterial growth and selection media stocks

All media and solutions used for the growth of microorganisms were sterilised by autoclaving for 20 minutes at 15 psi, or if heat labile, were filter sterilised.

LB medium (per litre)	: 10 g bacto-tryptone
(adjusted to pH 7.0 with 5 N	5 g bacto-yeast extract
NaOH)	10 g NaCl
LB plates (per litre)	: 1 litre LB
	15 g bacto-agar
LB top agar/agarose	: 1 litre LB
	7 g agar/agarose
2 x YT medium (per litre)	: 16 g bacto-tryptone
(adjusted to pH 7.0 with 5 N	10 g bacto-yeast extract
NaOH)	5 g NaCl
SOB medium (per litre)	: 20 g bacto-tryptone
(adjusted to pH 7.0 with 5 N	5 g bacto-yeast extract
NaOH)	5 g NaCl
	10 ml KCl solution (250 mM)
Add just before use:	5 ml sterile $MgCl_2$ solution (2 M)
ampicillin stock	: 50 mg/ml in sterile water
(filter sterilised and stored frozen)	
tetracycline stock	: 5 mg/ml in ethanol
(stored at –20°C)	

kanamycin stock : 10 mg/ml in sterile water

(filter sterilised and stored frozen)

IPTG stock : 20% (w/v) in sterile water

(filter sterilised and stored frozen)

X-Gal : 20 % (w/v) in dimethylformamide

(stored at –20°C)

2.1.3 Buffers

PBS mix : 8:1:1 mix of solutions A, B and C (each was sterilised by autoclaving)

Solution A : 170 mM NaCl
0.5 mM KCl
10 mM Na_2HPO_4
2 mM KH_2PO_4

Solution B : 7 mM $CaCl_2.2H_2O$

Solution C : 5 mM $MgCl_2.6H_2O$

10 x TBE (pH 8.3) : 0.89 M Tris
0.89 M boric acid
0.025 M EDTA

50 x TAE (pH 8.0) : 2 M Tris
1 M acetic acid
50 mM EDTA

20 x SSC	: 3 M NaCl 0.3 M tri-sodium citrate Adjusted to pH 7.0 with HCl
20 x SSPE	: 3.6 M NaCl 0.2 M sodium phosphate, pH 7.5 0.02 M EDTA
100 x Denhardt's solution (filter sterilised and stored frozen)	: 2% (w/v) Ficoll (type 100, Pharmacia) 2% (w/v) polyvinylpyrrolidone (PVP) 2% (w/v) BSA (fraction V, Sigma)
TE (sterilise by autoclaving)	: 10 mM Tris-HCl 1 mM EDTA Adjusted to pH 8.0 with HCl
Kirby reagent	: 6% (w/v) 4-amino-salicyclic acid 10 mM Tris-HCl pH 7.6 50 mM KCl 1% (w/v) TNS (Tri-isopropyl napthalene sulphonic acid (sodium salt)
10 x MOPS buffer (pH 7.0) (made in DEPC-treated water)	: 0.2 M 3-[N-Morpholino]-propane sulphonic acid 0.05 M sodium acetate 0.01 M EDTA

Solution I (for plasmid preps)		: 50 mM glucose
(filter sterilise)		2 mM Tris-HCl, pH 8.0
		10 mM EDTA
Solution II	"	: 0.2 M NaOH
		1% SDS
Solution III	"	: 60 ml 5 M potassium acetate
(make in sterile distilled water)		11.5 ml acetic acid
		28.5 ml water

2.1.4 Bacterial strains and storage

Bacterial strains were stored in the short term (for up to one month) at 4°C on tightly sealed agar plates. Longer term storage was in 50% (v/v) glycerol at –70°C.The following strains were used in this study.

a) Strain used as host to plasmid pBluescript II and for helper phage mediated single stranded rescue: *E. coli* XL1-Blue.

Phenotype: r_k^- lacZΔM15 Rec$^-$

Genotype: *supE*44 *hsdR*17 *recA*1 *endA*1 *gyrA*46 *thi relA*1 *lac*$^-$

F´[*proAB*$^+$ *lacI*^{q} *lacZ*ΔM15 Tn*10*(*tet*^{r})] (Bullock *et al.*, 1987)

b) Strain used as host for bacteriophage λgt11: *E. coli* Y1090*hsdR*

Phenotype: r_k^-Lac$^-$ Lon$^-$

Genotype: *hsdR supF* Δ*Lac*U169 Δ*Lon araD*139 *rpsL*

*trpC*22::Tn*10*(*tet*^{r}) pMC9 (Young and Davis, 1983; Jendrisak *et al.*, 1987)

2.1.5 DNA vectors, recombinants and oligonucleotides used in this study

DNA vectors, recombinants and oligonucleotides were stored in TE (see section 2.1.3) at –20°C.

2.1.5.1 λgt11 pea cDNA library "5′-stretch" (Clontech)

mRNA source : *Pisum sativum* var. Alaska, pea buds of etiolated 5 day old seedlings.

Random hexamers and oligo dT were used to prime first strand cDNA synthesis in the construction of this library.

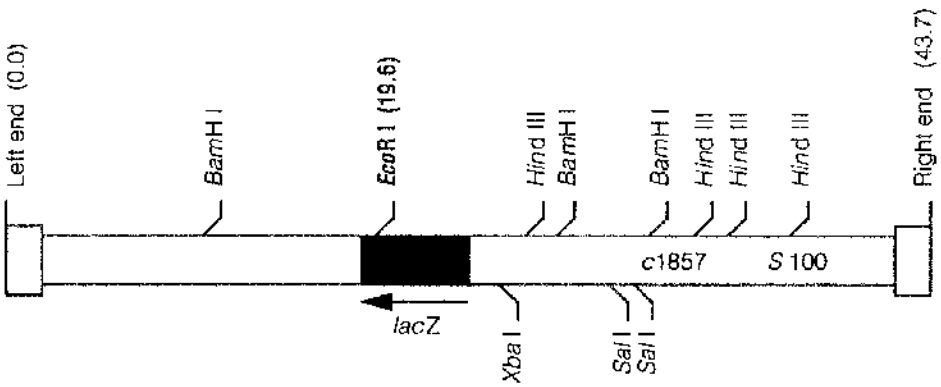
Titre of library : 3 x 10^{10} pfu/ml at time of purchase.

No. of independent clones : 1.8 x 10^{6}.

Average insert size: 1.7 kb (range: 0.8 to 4.0 kb), as estimated by PCR using Clontech gt11 primers.

λgt11 is a vector commonly used for immunological screening. DNA fragments (up to 7.2 kb) are cloned into the unique *Eco* RI site located in *lacZ*, allowing expression of a fusion protein if the cloned sequence is in frame with *lacZ*. A *sup F* host is used for propagation and screening. In a non suppressing host, the Sam100 mutation prevents cell lysis and accumulation of the desired fusion protein. The bacteriophage carries *red* and *gam* and can thus be grown in a *recA*⁻host. The temperature sensitive repressor allows control of bacteriophage replication and production of fusion proteins. Therefore, λgt11 libraries can also be screened with DNA probes. A restriction map of λgt11 is shown in figure 2.2.1.

Figure 2.1.1 Map of λgt11



Notes: The direction of transcription of the *lac*Z gene is indicated by the arrow. Also shown are the temperature-sensitive *c*I repressor, *c*I857, and the S amber mutation, *S*100.

2.1.5.2 pBluescript II KS +/– (Stratagene)

This vector is used for subcloning λgt11 inserts of interest and PCR products obtained during this study. The F′ origin allows single stranded rescue by helper phage. See figure 2.2.2.

Bluescript® II KS +/- phagemid

The Bluescript® II KS (+/-) phagemid is a 2961 basepair phagemid derived from pUC19. The KS designation indicates the polylinker is oriented such that *lacZ* transcription proceeds from Kpn I to Sac I.

f1 (+) origin: (3-459 bp) f1 filamentous phage origin of replication allowing recovery of the sense strand of the *lacZ* gene when a host strain containing the Bluescript II phagemid is co-infected with helper phage.

f1 (-) origin: (3-459 bp) f1 filamentous phage origin of replication allowing recovery of the antisense strand of the *lacZ* gene when a host strain containing the Bluescript II phagemid is co-infected with helper phage.

ColE1 origin: (1032-1972 bp) Plasmid origin of replication used in the absence of helper phage.

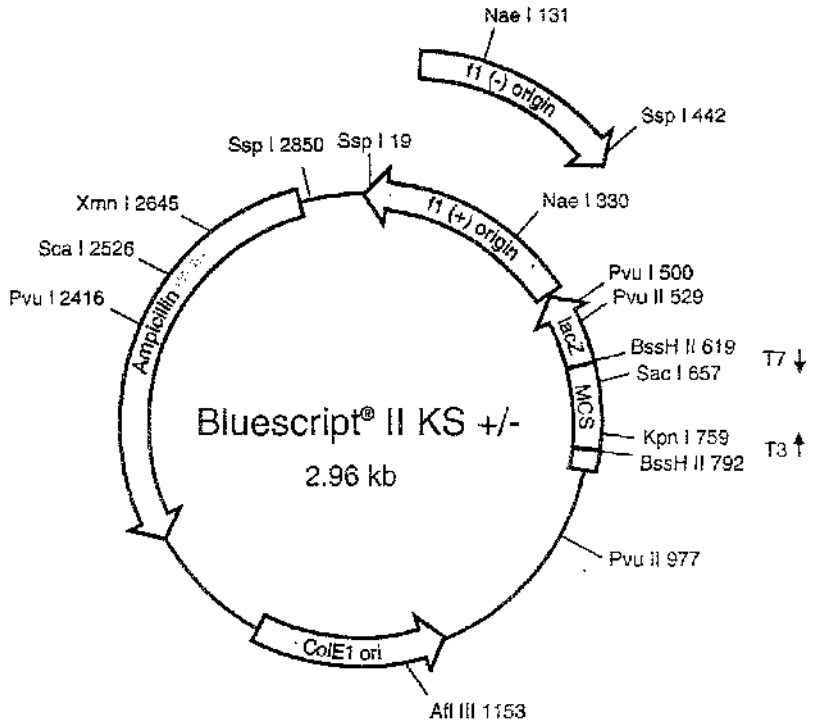
LacZ: (lac promoter: 816-938 bp) This portion of the *lacZ* gene provides α-complementation for blue/white color selection of recombinant phagemids. An inducible lac promoter upstream from the *lacZ* gene permits fusion protein expression with the β-galactosidase gene product.

MCS: (657-759 bp) Multiple cloning site flanked by T3 and T7 promoters; please see sequence below.

Ampicillin: (1975-2832 bp) Ampicillin resistance gene for antibiotic selection of the phagemid vector.

Please Note: The upper strand is designated the (+) strand and the lower strand is designated the (-) strand

GenBank® # X52327 (KS+), X52329 (KS-)



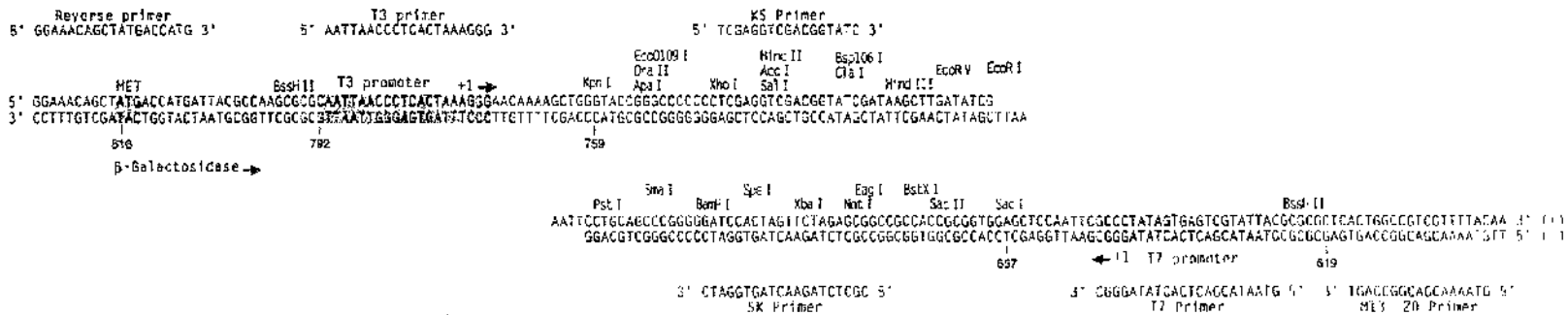


Figure 2.1.2 Map of pBluescript II KS +/-

2.1.5.3 Synthetic oligonucleotides

Synthetic oligonucleotides (for use as primers in the polymerase chain reaction) were made in the Department of Biochemistry by Dr. V. Math, using an Applied Biosystems 381A DNA synthesizer for the phosphite-triester method (Atkinson and Smith, 1984).

Oligonucleotides for use as sequencing primers were obtained from Pharmacia.

2.2 Methods

2.2.1 Growth of pea seedlings

Pea seeds (*Pisum sativum* var. Feltham First) were swollen in water for 4–6 h and then seeded into a tray (21 cm x 33 cm) under compost. The seedlings were grown at 26°C under continuous white light illumination (which was provided by a fluorescent light situated approximately 30 cm above the tray), and kept moist by watering. Unless otherwise stated, the portion selected for RNA and genomic DNA isolation and MTase partial purification was the apical shoot tips from 4–5 day old seedlings. The collected material was stored at –70°C.

2.2.2 Antibody work

2.2.2.1 SDS-PAGE of partially purified pea methylase

Partially purified pea methylase fractions were a kind gift from Claire Houlston (for enzyme preparation protocols see Houlston *et al.*, 1993). SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to the method of Laemmli (1970). An 8.75% resolving gel was prepared using 10 ml resolving gel buffer (1.5 M Tris-HCl (pH 8.8), 0.4% SDS), 11.7 ml 30% acrylamide stock (30% acrylamide,

0.8% NN´-methylenebisacrylamide), 18 ml distilled water, 200 µl 10% ammonium persulphate and 50 µl TEMED which was added last. The solution was mixed and poured into a pair of clamped gel plates. A thin layer of 1% SDS was applied to the gel to allow it to set. A 6% stacking gel was prepared using 2.5 ml stacking gel buffer (0.5 M Tris-HCl (pH 6.8), 0.4% SDS), 2.0 ml 30% acrylamide stock, 5.5 ml distilled water, 150 µl ammonium persulphate and 5 µl TEMED which again was added last. The layer of SDS was removed and the stacking gel mixed and poured on top of the resolving gel. The well forming comb was inserted immediately and removed as soon as the gel had polymerised. Samples were incubated with 0.5 volumes SDS-PAGE sample buffer (0.1875 M Tris-HCl (pH 6.8), 30% glycerol, 6% SDS, 15% mercaptoethanol and 0.1% bromophenol blue) for 5 minutes in a boiling water bath and subject to electrophoresis at 30–40 mA (< 200 V). The electrode buffer was Tris/glycine buffer (25 mM Tris, 191 mM glycine; pH 8.3) and protein molecular weight standards were used (myosin 200 kDa, phosphorylase B 97.4 kDa, bovine serum albumin 69 kDa, ovalbumin 46 kDa, carbonic anhydrase 30 kDa, trypsin inhibitor 21.5 kDa and lysozyme 14.3 kDa). After electrophoresis the gel was western blotted or washed in 50% methanol prior to silver staining.

2.2.2.2 Silver staining of protein gels

The gel was fixed in 50% methanol overnight before treatment with solution C. This solution consisted of solution A (0.8 g of $AgNO_3$ in 4 ml distilled water) added dropwise, whilst vigorously stirring to solution B (21 ml of 0.36% NaOH with 1.4 ml 14.8 M NH_4OH) until the saturation point was reached. Solution C was made up to 100 ml with distilled water and used immediately to wash the gel for 10–15 minutes with gentle agitation at room temperature. The gel was then washed for 1–2 hours in distilled water, again with gentle agitation. The gel was then washed with solution D (2.5 ml of 1% citric acid and 0.25 ml of 38% formaldehyde), made up to 500 ml with distilled water and used immediately. The bands developed within 10–

15 minutes and development was stopped by washing in distilled water and then in 50% methanol. The gel was photographed using a Polaroid C-U5 camera or imager.

2.2.2.3 Transfer of proteins to nitrocellulose: western blotting

A sheet of nitrocellulose impregnated nylon membrane (Hybond C^+ extra), cut to the same size as the gel, was soaked in transfer buffer (25 mM Tris-HCl (pH 7.2), 0.19 M glycine, 0.02% SDS, 20% methanol) and placed on top of the gel which had been placed on a sheet of Whatman 3MM paper pre-soaked in transfer buffer. Care was taken to ensure there were no bubbles between the gel and the membrane. A second sheet of pre-soaked Whatman 3MM paper was placed on top of the membrane and everything was sandwiched between two sponge pads soaked in transfer buffer and placed in a perspex cassette. The cassette was placed in a BioRad transblot tank filled with transfer buffer so that the nitrocellulose sheet was facing the cathode. The electrophoretic blotting was carried out at 40–60 mA overnight.

2.2.2.4 Enhanced chemiluminescence (ECL)

Enhanced chemiluminescence was carried out according to the manufacturer's protocol. The electrophoretic blot was washed in blotting buffer (20 mM Tris-HCl (pH 7.2), 0.15 M NaCl), supplemented with 3% non-fat dried milk and 0.5 mg/ml NaN_3 for 1 hour with continuous shaking at room temperature to block non-specific binding to the membrane. The membrane was rinsed five times for 5 minutes with blotting buffer containing 0.1% TWEEN 20. The blot was then incubated for 1 hour with blotting buffer supplemented with 0.5 mg/ml NaN_3, 5% (v/v) heat inactivated normal goat serum plus the appropriate dilution of the primary antibody (usually 1/100). The blot was washed five times as before, then incubated in blotting buffer containing 0.1% TWEEN 20 and 0.1% of the secondary antibody (anti-rabbit Ig linked to horseradish peroxidase). The membrane was again washed five times as

before and blotted dry on Whatman 3MM paper. The chemiluminescence was detected after the addition of 10 ml of a 50:50 mixture of solutions A+B supplied by Amersham. The membrane was left for exactly one minute in this mixture, then wrapped in cling film and exposed to hyperfilm-ECL (Amersham) for 10 seconds and the film developed. The exposure time could be varied as required.

2.2.3 Extraction and precipitation of DNA

2.2.3.1 Phenol/chloroform extraction

Phenol/chloroform stock was made by mixing equal amounts of phenol and chloroform then equilibrating the mixture by extracting several times with TE until the pH was 8.0. 8-Hydroxyquinoline was added to 0.1% as an antioxidant and to assist in distinguishing the phenol/chloroform from the aqueous phase in extractions. The equilibrated mixture was stored under TE at 4°C in the dark. Extraction with phenol/chloroform was performed as follows: an approximately equal volume of phenol/chloroform was added to the DNA solution to be extracted, mixed and centrifuged for 10 minutes; the upper aqueous layer was transferred to a fresh tube and the extraction repeated as necessary.

2.2.3.2 Ethanol and isopropanol precipitation

Ethanol used was of the absolute alcohol 100 grade and was stored at –20°C. A salt solution was added to the DNA solution to the appropriate concentration (0.1 volume of 5 M ammonium acetate or 0.1 M volume of 3 M sodium acetate) and mixed with 2–3 volumes of cold ethanol. The samples were placed at –20°C for 20 minutes then centrifuged for 15 min at 12,000 rpm in a microfuge at 4°C. The precipitated DNA was washed with 70% ethanol at –20°C and dried at room temperature under vacuum. DNA was usually stored at –20°C in TE. When the DNA sample was such

that the addition of 2–3 volumes ethanol was impractical, precipitation was carried out by the addition of an equal volume of isopropanol to the DNA + salt solution.

2.2.3.3 Estimation of nucleic acid concentration

Spectrophotometric estimation of nucleic acid concentration was carried out as follows: A thousand-fold dilution of the purified nucleic acid solution was made in a volume of 1 ml. The absorbance at 260 nm and 280 nm was measured using a spectrophotometer (Cecil instruments, model CE 2020). One OD_{260} unit corresponds to approximately 50 µg/ml double-stranded DNA, 40 µg/ml for single stranded DNA and RNA, and ~25µg/ml for single stranded oligonucleotides. The OD_{260}/OD_{280} ratio provides an estimate of the purity of the nucleic acid. Pure solutions of DNA and RNA have OD ratios of 1.8 and 2.0 respectively.

In cases where the concentration of DNA was too low to be estimated spectrophotometrically, the DNA solution was compared against a series of DNA concentration standards which were made by serial dilution of λ DNA (Gibco BRL). Aliquots (1 µl) of each standard (plus 1 µl of the unknown) were pipetted onto the surface of a 1.5% agarose gel containing 0.5 µg/ml ethidium bromide which had been poured into a 90 mm petri dish. The degree of fluorescence under u.v. light of the unknown was compared to that of the standards in order to estimate the DNA concentration of the unknown.

2.2.4 DNA digestion with restriction endonucleases

Restriction digests were routinely carried out in 1.5 ml Eppendorf tubes in the presence of the appropriate buffer. One unit of restriction enzyme activity is defined as the amount of enzyme required to digest 1 µg of λ DNA to completion in one hour at the correct incubation temperature and appropriate buffer. In practice a several-fold excess was usually added to digest the DNA. A typical reaction mixture contained 1

μg of DNA, the correct enzyme buffer (usually supplied by the manufacturer) and 5 units of the restriction enzyme in a final volume of 10 μl. The mixture was incubated for at least 1 hour at the temperature recommended by the manufacturer, which was usually 37°C. The completion of the digest was monitored by electrophoresis of an aliquot through an agarose gel. After digestion, restriction enzymes were heat inactivated (65°C for 10 minutes) or removed by phenol/chloroform extraction if appropriate. Digests of genomic DNA were incubated overnight, using 8 units of restriction enzyme per μg of DNA. For multiple restriction digests, the digestion which required the lowest ionic strength was carried out first, then the ionic conditions were adjusted, the second enzyme added and the incubation continued.

2.2.5 DNA electrophoresis in agarose gels

Agarose gels were routinely used to separate DNA fragments in the size range 0.1–20 kb. They were prepared by dissolving agarose (BRL ultra PURE) in 1 x TAE or 1 x TBE buffer (section 2.1.3) to the desired concentration (0.7–2%). Ethidium bromide was added to 0.5 μg/ml and the gel poured into a horizontal mould (BRL model H5 or model 200). Wells with a suitable capacity were formed by insertion of a comb into the gelling agarose. After submerging the gel in 1 x TAE containing 0.5 μg/ml ethidium bromide, one tenth-volume loading buffer (30% Ficoll type 400 (Pharmacia), 0.50% bromophenol blue) was added to the DNA samples and the mixture loaded into the wells (genomic digests were heated to 65°C for 5 min than quenched on ice before loading). Electrophoresis was carried out at 1–5 V/cm (measured as the distance between electrodes) until the bromophenol blue had migrated the appropriate distance along the gel. The gel was then rinsed in distilled water. DNA fragments were visualised by the use of a long wavelength u.v. transilluminator (U.V. Products Inc.) and photographed with a U.V.P. imager. The distance moved by a DNA fragment during gel electrophoresis is directly proportional to the log of its molecular weight or size in base pairs. The size of

unknown fragments could therefore be determined by comparison with DNA marker fragments of known size which were electrophoresed on the same gel.

2.2.6 Recovery of DNA from low melting point agarose

Low melting point agarose (BRL ultra PURE) gels were prepared and used in the same way as standard agarose gels (using 1 x TAE as the buffer), except that they were run in the cold room (at 4°C). The gel was placed over a u.v. transilluminator and the desired band was excised from the gel with scalpel and transferred to an Eppendorf tube. Three volumes of TE buffer were added to the agarose slice and the sample was placed in a waterbath for 8 minutes to melt the agarose. The sample was extracted twice with phenol saturated with TE buffer (pH 8), and three times with ether saturated with distilled water to remove the agarose. The DNA was then ethanol precipitated.

A quicker, more efficient method for isolating DNA from agarose in the base pair range 300 to 15 kb is to use the Geneclean II kit. The DNA band containing up to 5 µg DNA was excised from low melting point agarose as above. Three volumes of NaI (6 M) were added and the tube was incubated at 50°C to dissolve the agarose. Glassmilk suspension (5 µl) was added and the mixture was incubated on ice for 5 minutes to allow binding of the DNA to the silica matrix, whilst mixing every 1–2 minutes. The silica matrix with the bound DNA was pelleted by centrifugation in a microfuge for 5 seconds. The supernatant was removed by aspiration and the pellet was resuspended in ice cold New Wash (200 µl). The matrix was again pelleted by centrifugation for 5 seconds and the supernatant removed. This washing step was repeated twice. The DNA was eluted by resuspending the matrix in 10–20 µl TE buffer, pelleting the matrix and recovering the TE buffer containing the DNA.

2.2.7 Subcloning DNA fragments into pBluescript

Insert DNA fragments and pBluescript vector DNA were digested with the appropriate enzyme(s) and the extent of digestion was checked by electrophoresis. The DNA was then phenol/chloroform extracted and ethanol precipitated. Insert DNAs were usually purified from low melting point agarose as described in section 2.2.6.

In cases where only one enzyme was used to cut the vector, religation of the vector to itself was prevented by treating the cut vector with alkaline phosphatase to remove 5´ phosphates: the reaction mix (50 µl), containing 5 µg digested vector DNA, 1 unit of calf intestinal alkaline phosphatase (Northumbria Biologicals), 10 mM Tris-HCl (pH 8.3), 1 mM $ZnCl_2$, 1 mM $MgCl_2$ was incubated at 37°C for 30 minutes. At the end of the incubation, the mixture was heated to 75°C for 10 minutes in the presence of 5 mM EDTA. The reaction was cooled to room temperature, phenol/chloroform extracted and ethanol precipitated. The DNA was resuspended in sterile distilled water at a concentration of 100 ng/µl.

Typical ligation conditions used contained: vector DNA (50 ng), an equimolar quantity of insert DNA (if possible), 20 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 5 mM DTT, 0.5 mM ATP and 1 Weiss unit of T4 DNA ligase in a final volume of 10 µl. The ligation mixture was incubated at 15°C overnight. A greater concentration of ligase (5 Weiss units in a volume of 10 µl) was used to ligate DNA molecules with blunt ends.

2.2.8 Transformation of bacterial cells with plasmid DNA

Frozen competent *E.coli* cells (strain XL1 Blue, section 2.1.4), were prepared according to the method developed by Hanahan (1983): a single colony of XL1 Blue taken from a fresh LB plate was used to inoculate 100 ml SOB medium. The cells were grown with vigorous shaking at 37°C for 2–3 hours until the OD_{600} of the

culture reached 0.3. The culture was cooled on ice and the cells harvested by centrifugation at 3000 rpm in a Beckman benchtop centrifuge (model CS 6R). The medium was decanted from the cell pellets and the tubes inverted to allow the last of the fluid to drain away. The cells were resuspended in 40 ml ice-cold FSB buffer (10 mM potassium acetate, pH 7.5, 45 mM $MnCl_2.4H_2O$, 10 mM $CaCl_2.2H_2O$, 100 mM hexamminecobalt chloride, 10% glycerol). The cells were recovered by centrifugation at 3000 rpm for 10 minutes at 4°C and the fluid allowed to drain away. The cells were resuspended by gentle vortexing in 8 ml ice-cold FSB. DMSO (280 µl) was added to the bacterial suspension which was mixed gently by swirling and the suspension was incubated on ice for 15 minutes. An additional 280 µl aliquot of DMSO was then added to the suspension and was mixed by swirling and the suspension was returned to the ice-bath. Working quickly, 200 µl aliquots of the bacterial suspension were dispensed into chilled Eppendorf tubes which were then snap frozen by immersing them in liquid nitrogen. The tubes were then transferred to the –70°C freezer for storage.

Transformation of competent cells was carried out as follows: an Eppendorf tube containing frozen competent cells was thawed in the hand and placed on ice for 10 minutes. A 5–10 µl aliquot of ligation mix containing 5–50 ng ligated DNA was added to the competent cells and the mixture was incubated for 30 minutes on ice. The cells were heat shocked at 42°C for 1 minute then placed back on ice. A mixture containing 40 µl IPTG (2% w/v), 40 µl X-gal (2% w/v in dimethylformamide) and 40 µl sterile, distilled water was spread out onto an LB plate containing ampicillin (100 µg/ml) and tetracycline (7.5 µg/ml). After the liquid had absorbed, the transformed cells were spread onto the plate which was then inverted and incubated at 37°C overnight. Typical transformation efficiencies achieved by this method (as determined by transformation with 1 ng control supercoiled plasmid) are in the range of 10^6 transformants/µg. Colonies transformed with recombinant pBluescript were colourless in the presence of IPTG and X-gal because they lacked a functional β-galactosidase. Colonies transformed with non-recombinant pBluescript are blue in

the presence of IPTG and X-gal.

2.2.9 Preparation of plasmid DNA

2.2.9.1 Small scale preparation of plasmid DNA

This alkaline lysis method is a modification of the method of Birnboim and Doly (1975). A 3 ml overnight culture was prepared from a single colony of transformed bacteria in LB supplemented with the appropriate antibiotic (100 µg/ml ampicillin for pBluescript and recombinants derived from this vector). A 1.5 ml aliquot of the overnight culture was transferred to an Eppendorf tube and the bacteria harvested by centrifugation for 20 seconds at 12,000 rpm in a microfuge. The remainder of the culture was set aside for the preparation of a glycerol stock if necessary (see section 2.1.4). The bacterial pellet was resuspended in 100 µl ice cold solution I. Freshly prepared alkaline lysis solution (solution II) (200 µl) was then added and the tube inverted several times until cell lysis had occurred. Neutralising solution (solution III) (150 µl) was added and the tube vortexed and put on ice for 3–5 minutes; (see section 2.1.3 for the composition of solutions I, II and III). The tube was then centrifuged at 12,000 rpm for 5 minutes at 4°C. The supernatant was extracted with phenol/chloroform (section 2.2.3.1); then with chloroform. The aqueous layer was ethanol precipitated (section 2.2.3.2). The DNA pellet was resuspended in 10–50 µl TE containing 20 µg/ml DNase-free pancreatic RNase (Boehringer). DNA prepared using this method is suitable for general use but is unsuitable for sequencing.

2.2.9.2 Small scale preparation of plasmid DNA using Promega Magic Miniprep columns

This system is a modification of the alkaline lysis miniprep method described in the previous section. The supernatant obtained after the neutralisation step was mixed

with 1 ml purification resin. The mixture was then applied onto a Promega Magic Miniprep column fitted to a vacuum manifold. The vacuum was applied to draw the liquid through the column, which was then washed with 2 ml column wash solution (100 mM NaCl, 10 mM Tris-HCl pH 7.5, 2.5 mM EDTA, 50% ethanol), whilst maintaining the vacuum. The column was then fitted in an open Eppendorf tube and dried by centrifuging for 20 seconds at 12,000 rpm in a microfuge. The plasmid DNA was eluted by applying 50 µl TE buffer to the column (fitted in a fresh Eppendorf tube) and centrifuging for 20 seconds at 12,000 rpm in a microfuge. Plasmid DNA prepared using this system is free of salt and macromolecular contaminants and was found to be suitable for double stranded sequencing using the Applied Biosystems automated sequencer (see section 2.2.21).

2.2.9.3 Large scale preparation of plasmid DNA

This protocol is essentially a scale-up of the alkaline lysis method described in section 2.2.9.1. A 200 ml overnight culture was prepared and the bacteria pelleted by centrifugation at 3000 rpm (Beckman GS-6R benchtop centrifuge) for 10 minutes at 4°C. The bacteria were resuspended in 10 ml solution I. Freshly prepared solution II (10 ml) was added to the bacterial suspension and the mixture incubated on ice for 5 minutes. Solution III (7.5 ml) was then added and the mixture was incubated on ice for 10 minutes; (see section 2.1.3 for the composition of solutions I, II and III). The cell debris was pelleted by centrifugation at 10,000 rpm in a JA 20 rotor at 4°C. The supernatant was then filtered through Whatman 3MM filter paper. Isopropanol (0.6 x volume) was added to the filtrate which was then incubated for 15 minutes at room temperature. The precipitated nucleic acids were recovered by centrifugation at 3000 rpm in a Beckman benchtop centrifuge at 4°C. The pellet was resuspended in 2 ml TE buffer. LiCl (5.0 M in 50 mM Tris-HCl, pH 8.0 (2.0 ml) was added and the mixture was incubated for 10 minutes on ice to precipitate the RNA. The RNA was removed by centrifugation for 5 minutes at 3000 rpm in a Beckman benchtop

centrifuge at 4°C. The DNA in the supernatant was precipitated with 2 volumes of ethanol for 15 minutes on ice and was then pelleted by centrifugation for 15 minutes at 3000 rpm in a Beckman benchtop centrifuge at 4°C. The pellet was dissolved in 500 µl TE buffer and the DNA solution was transferred to an Eppendorf tube. DNase-free RNase was added to a concentration of 0.6 µg/µl and the DNA was incubated for 1 hour at 37°C. Two phenol/chloroform extractions were carried out (section 2.2.3.1), followed by one chloroform extraction, and the DNA was precipitated with two volumes of ethanol. After recovery by centrifugation at 12,000 rpm for 15 minutes at 4°C (in a microfuge), the DNA pellet was washed with 70% ethanol and resuspended in 100 µl TE buffer. Typically, this method yields 500 µg to 1 mg plasmid DNA (mostly supercoiled), which is of suitable quality for all applications described in this thesis.

2.2.9.4 Preparation of single stranded plasmid DNA

Single stranded rescue of pBluescript DNA, using helper phage (M13 KO7), was performed using the method of Blondel and Thillet (1990) and was as follows: a single freshly grown colony, harbouring the phagemid, was picked from an LB agar plate containing 100 µg/ml ampicillin and 12.5 µg/ml tetracycline. The colony was then suspended in 20 µl 2 x YT medium. Helper phage was added (5 µl) from a stock with a titre of at least 10^{10} pfu/ml. The mixture was incubated for 15 min at room temperature, before adding 500 µl 2 x YT supplemented with ampicillin (100 µg/ml) and tetracycline (12.5 µg/ml). The mixture was then grown for 1 h at 37°C with shaking. The bacterial suspension (120 µl) was added to 3 ml of LB medium in a falcon tube, supplemented with ampicillin (100 µg/ml), tetracycline (12.5 µg/ml) and kanamycin (75 µg/ml). The mixture was incubated with shaking for 18–20 h at 37°C. The overnight culture (1.4 ml) was microfuged (7 min, 7000 rpm) to remove bacterial debris. The supernatant (1.2 ml) was removed and 200 µl of a solution containing 20% PEG 6000 and 2.5 M NaCl was added. The mixture was incubated

for 30 min at room temperature. The precipitated phage particles were spun down (5 min, 12000 rpm) in a microfuge and the phage pellet dissolved in 100 µl TE buffer, extracted twice with 50 µl phenol and once with 100 µl chloroform. The DNA was then ethanol precipitated and resuspended in 30 µl TE buffer. A typical concentration obtained was 0.2 µg/µl.

2.2.10 Plating out the lambda phage library for screening with DNA probes

The cDNA library used in this study was in λgt11. The host for this vector is *E.coli* Y1090*hsdR* (see section 2.1.4). Plating bacteria were prepared as follows: Sterile LB medium (20 µl), supplemented with 0.2% maltose, was inoculated with a single colony of Y1090 taken from a freshly grown plate. The culture was grown overnight with moderate shaking at 37°C. The bacteria were pelleted by centrifugation at 4000 rpm (Beckman GS-6R benchtop centrifuge) for 10 minutes at room temperature and the cells were resuspended by vortexing in 20 ml sterile 10 mM $MgSO_4$. Plating bacteria prepared in this way can be stored at 4°C for up to 3 weeks.

To titre the phage library, a serial dilution was made of the library stock.in SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5, 0.01% gelatin). A 100 µl aliquot of each dilution was mixed with 100 µl plating bacteria. The tubes were incubated for 20 minutes at 37°C to allow the phage particles to infect the bacteria. Molten top agar at 47°C (2.5 ml) was added to each tube, the mixture vortexed briefly and poured onto pre-warmed 90 mm LB plates. Once the agar had hardened, the plates were inverted and incubated overnight at 37°C.

For library screening, the phage were plated out using the appropriate dilution to give approximately 30,000–75,000 plaques per 150 mm LB plate. Molten top agarose (10 ml) was used in this case to prevent the top layer peeling off during plaque lifts. Plaques which gave a positive signal upon hybridisation of the plaque lifts were isolated as follows: An agar plug (approximately 5 mm in diameter) was

removed around the area of the plaque with the snipped-off end of a 1 ml disposable pipette tip. The agar plug was placed in 1 ml SM buffer containing a drop of chloroform (the chloroform had been extracted with sodium bicarbonate to remove photolysis products). The plug was left to elute for 1 hour with occasional agitation and the phage were plated out at a density of 100 to 300 plaques per 90 mm plate so that well isolated plaques which hybridised to the probe upon secondary screening could be picked.

2.2.11 Small scale preparation of lambda DNA

Either liquid lysate or plate lysates can be used as starting material. For this study, liquid lysates were used and prepared as follows: A single, well isolated plaque was picked from a fresh plate and transferred to an Eppendorf containing 100 µl SM buffer. This was placed at 4°C overnight. A fresh culture of *E.coli* Y1090 *hsD* was started by inoculating a single colony into 5 ml LB (supplemented with 0.2% maltose and 10 mM $MgSO_4$), and shaking at 37°C overnight. The overnight culture (500 µl) was added to 20 µl of the phage plug eluate and the mixture incubated at 37°C for 20 minutes. The infected culture was transferred to a 250 ml Erlenmeyer flask containing 100 ml pre-warmed LB (to 37°C) supplemented with 10 mM $MgSO_4$. The flask was incubated at 37°C with shaking until lysis occurred, at which point 500 µl chloroform was added. If after 7 hours lysis had not occurred, 500 µl chloroform was added and shaking was continued for a further 15 minutes. The lysate was centrifuged at 8000 rpm (Beckman JA-20 rotor) for 10 minutes at 4°C to remove cellular debris. A 10 ml aliquot of the lysate was transferred to a Corex tube. Nuclease mixture (40 µl) (0.25 mg/ml RNase A, 0.25 mg/ml DNase I, 150 mM NaCl, 50% glycerol) was added and the tube incubated at 37°C for 15 minutes. Phage Precipitant (4 ml) (33% PEG 8000, 3.3 M NaCl) was then added to the lysate, the solution mixed gently and placed on ice for 30 minutes. The tube was centrifuged at 10,000 rpm (Beckman JA-20 rotor) for 10 minutes at 4°C. The supernatant was

decanted and the pellet resuspended in 500 µl Phage Buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 10 mM $MgSO_4$). The resuspended phage were transferred to an Eppendorf tube which was centrifuged for 10 seconds at 12,000 rpm to remove any insoluble particles. Purification Resin (1 ml) was added to the supernatant. The resin/lysate mixture was applied onto a Magic Lambda Prep Column (which was attached to a vacuum manifold, Promega), and the vacuum applied. The column was then washed with 2 ml 80% isopropanol whilst maintaining the vacuum. The column was fitted in an open Eppendorf tube and dried by centrifuging for 20 seconds at 12,000 rpm in a microfuge. The lambda DNA was eluted by applying 100 µl TE buffer pre-heated to 80°C to the column and centrifuging the column (fitted in a fresh Eppendorf tube) for 20 seconds at 12,000 rpm.

2.2.12. Preparation of genomic DNA

2.2.12.1 Preparation of pea or mouse genomic DNA (method 1)

Genomic DNA was extracted from pea nuclei, which were prepared using the method described by Yesufu *et al.* (1991), were kindly provided by Claire Houlston according to the method described by Yesufu *et al.*, 1991). Nuclei from 30 g pea seedlings were suspended in 1 ml phosphate buffered saline (see section 2.1.3.) and 1 ml SDS mix (2% SDS, 4 mM EDTA, butanol 10%). They were then incubated with proteinase K (50 µg/ml) for 1 hour at 37 °C. An equal volume of phenol/chloroform was added and the two phases were mixed by gentle inversion over a period of 10 minutes. The mixture was then centrifuged at 3000 rpm in a Beckman bench top centrifuge, model TJ-6. The upper aqueous layer was extracted with phenol/chloroform in the same way until there was no interface (1 or 2 further extractions). This was followed by a chloroform extraction. The DNA was precipitated with two volumes of ethanol at –20°C. The precipitated DNA was spooled out of solution and washed twice with 70% ethanol and left to air dry before

being dissolved in 250 µl 50 mM KCl (at 4 °C overnight). Mouse genomic DNA was prepared using this method on mouse ascites nuclei which were kindly provided by Dr. R.L.P. Adams.

2.2.12.2 Preparation of pea genomic DNA (method 2)

This method (adapted from that of Rogers and Bendich, 1988) employed the detergent CTAB (cetyltrimethylammonium bromide), and yields plant genomic DNA which is relatively free of contaminating polysaccharides, which can inhibit digestion by restriction enzymes. Genomic DNA was made from 5 day old pea shoot tips. The harvested material (~5–10 g) was ground under liquid nitrogen with a pestle and mortar. An approximately equal volume of 2 x CTAB buffer (2% CTAB (w/v), 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, 1% (w/v) polyvinylpyrrolidone (PVP 40,000), pre-heated to 65°C, was mixed with the ground plant tissue. One volume of chloroform/isoamyl alcohol (24:1) was then added, and mixed gently to form an emulsion. The mixture was then centrifuged at 10,000 rpm in a Beckman JA-20 rotor for 10 min, and the upper aqueous phase collected. A 1/10 volume of 10% CTAB buffer was then added (10% w/v CTAB, 0.7 M NaCl) and mixed gently. The chloroform/isoamyl alcohol extraction was repeated and an equal volume of CTAB precipitation solution (1% CTAB, 50 mM Tris-HCl (pH 8.0), 10 mM EDTA) was then added and mixed gently. The DNA/CTAB precipitate was dissolved in high salt TE buffer (TE with 1.0 M NaCl) then precipitated with 2 volumes of ethanol. The precipitated DNA was spooled out, and resuspended in high salt TE buffer before purification by CsCl equilibrium density centrifugation as follows:

The volume of the DNA solution was measured precisely and 1 g CsCl was added per ml plus 0.8 ml ethidium bromide (10 mg/ml). The final density of the solution was 1.55 g/ml. The solution was centrifuged at 12,000 rpm (Beckman JA-20 rotor) for 5 min. The clear red solution was removed from the precipitated protein complexed with ethidium bromide and transferred to a 5.1 ml polypropylene

ultracentrifuge tube, and the tube heat sealed. The tube was centrifuged overnight at 45,000 rpm in a VTi 65 rotor (Beckman), at 22°C. After centrifugation, the banded DNA (visualised under u.v.) was removed by piercing the top of the tube with a hypodermic needle to let in air, then piercing just underneath the DNA band with an 18-guage hypodermic needle attached to a 1 ml syringe. The ethidium bromide was removed from the DNA by extracting several times with an equal volume of water-saturated butan-1-ol (no centrifugation required) until the aqueous phase was colourless. The CsCl was then removed by dialysis against TE buffer overnight.

2.2.13 RNA isolation and handling

Where practical, all solutions and plasticware used for the extraction and handling of RNA were treated with DEPC (diethylpyrocarbonate) as follows: DEPC was added to solutions (except those containing Tris) at a concentration of 0.1% (v/v). The solutions were left overnight and then autoclaved at 15 psi for 20 min. Tris solutions were prepared by dissolving Tris in DEPC-treated water. Disposable plasticware was soaked in 0.1% DEPC overnight before autoclaving. Glassware was washed thoroughly and baked at 180°C for 2 h.

2.2.13.1 Preparation of total RNA from pea seedlings

The Kirby method was used for total RNA isolation, and was as follows: the material (1 g of pea seedlings, stored at –70°C) was ground under liquid nitrogen with a pestle and mortar.and the ground tissue transferred to a sterile falcon tube containing 10 ml phenol (equilibrated to pH 7.6 with 10 mM TE) plus 10 ml Kirby reagent (section 2.1.3). The two phases were mixed by vortexing and the tube was centrifuged at 3000 rpm (Beckman GS-6R bench top centrifuge) for 10 minutes. The aqueous layer was removed and was re-extracted with 5 ml phenol reagent mixed with 5 ml chloroform. This was repeated until there was a clear interface between the

phases (usually twice). The RNA was precipitated with an equal volume of isopropanol at 4°C and pelleted by centrifugation at 3000 rpm (Beckman GS-6R benchtop centrifuge) at 4°C. The pellet was then washed 3–4 times with 2.5 M sodium acetate (pH 5.2) to remove contaminating polysaccharides, then washed with 80% ethanol made with DEPC-treated water and resuspended in 200 µl DEPC-treated water.

2.2.13.2 Isolation of poly(A)$^+$ RNA

Poly(A)$^+$ RNA was isolated from total RNA (2.5.13) by oligo dT cellulose chromatography according to the method described by Sambrook *et al.*(1989). The oligo dT cellulose (1 g) was suspended in 0.1 N NaOH. A column (bed volume = 1 ml) was poured into a DEPC-treated 2.5 ml syringe plugged with glass wool. The column was rinsed with sterile, RNase-free 1 x column-loading buffer (20 mM Tris-HCl (PH 7.6), 0.5 M NaCl, 1 mM EDTA (pH 8.0), 0.1% sodium lauryl sarcosinate) until the pH of the effluent was less than 8.0. Total RNA (10 mg) (prepared from ~15 g pea shoots) was diluted to 2 ml with DEPC-treated water. The solution was then heated to 65°C for 5 minutes, cooled quickly to room temperature (in a water bath) before adding an equal volume of 2 x column-loading buffer. The total RNA sample was then applied to the column and the eluate collected. The column was then rinsed with 1 ml 1 x column loading buffer and the eluate collected. The eluates were then pooled, heated to 65°C and cooled to room temperature as before, then re-applied to the column. The eluate was collected in 1 ml fractions. The column was then washed with 10 ml 1 x column loading buffer, followed by 10 ml 1 x column loading buffer (containing 0.1 M NaCl). Throughout, 1 ml samples were collected. The OD_{260} values were measured for each fraction. The elution was stopped when the OD_{260} of the fractions approximated to zero. The poly(A)$^+$ RNA was then eluted with 4 ml sterile RNase-free elution buffer (10 mM Tris-HCl, pH 7.8, 1 mM EDTA (pH 8.0), 0.05% SDS). The OD_{260} of each fraction was then measured in a quartz

cuvette which had been soaked in concentrated HCl:methanol (1:1) then washed extensively with DEPC-treated water. The peak fractions were pooled, ethanol precipitated and resuspended in DEPC-treated water. An aliquot was checked by denaturing gel electrophoresis (section 2.2.13.3). It was always found that the chromatography step had to be repeated to get rid of residual ribosomal RNA.

2.2.13.3 RNA electrophoresis in denaturing agarose gels

The method used was that suggested by Sambrook *et al.*(1989). A 1% agarose gel with formaldehyde was prepared as follows: the appropriate amount of agarose was dissolved in water and the solution cooled down to 60°C. Formaldehyde and 10 x MOPS buffer (section 2.1.3) were added to concentrations of 2.5 M and 1 x respectively. The gel was cast in the same apparatus used for DNA agarose gel electrophoresis which had been pre-treated with 3% H_2O_2 and rinsed with DEPC treated water to inactivate any RNase. The samples were prepared by mixing the following: RNA (up to 30 µg) (5.5 µl), 10 x MOPS buffer (1.0 µl), formaldehyde (3.5 µl), and deionised formamide (10 µl). The samples were incubated for 15 minutes at 65°C and then chilled on ice. To each sample was added 2 µl formaldehyde gel-loading buffer (50% glycerol, 1 mM EDTA, 0.25% bromophenol blue). The gel was pre-run at 5 V/cm with 1 x MOPS buffer as the electrophoresis buffer. The samples were then loaded and the gel was run at 4 V/cm. RNA size markers were also run to enable size estimation of unknown messages. The quality of the RNA and the distance moved by the size markers were checked by staining the gel with ethidium bromide (0.5 µg/ml in 0.1 M ammonium acetate) for 30 minutes and the RNA visualised on the u.v. transilluminator. A better picture was obtained if the gel was left to destain overnight in distilled water at 4°C.

2.2.14 Nucleic acid transfer to membranes

2.2.14.1 DNA transfer from agarose gels to membranes (Southern blotting)

This method is basically that of Southern (1975). Positively charged nylon membranes (Hybond N^+, Amersham international plc) were used throughout this study. DNA can be fixed to these membranes with alkali, and thus the blotting and fixing steps can be combined if alkali is used as the blotting buffer.

After electrophoresis (section 2.2.5) the agarose gel was immersed in 0.25 M HCl for 15 minutes at room temperature with gentle agitation. This "depurination" step was only necessary to increase the efficiency of transfer of DNA fragments of 10 kb and over. The gel was rinsed in distilled water and an alkali blot was set up as follows: NaOH (0.4 M) was used to fill two reservoirs connected over a solid support by two wicks of Whatman 3 MM paper and the gel was placed on the bridge. Any of the 3 MM paper surface not covered by the gel was covered with parafilm, to prevent side capillary transfer. A nylon filter cut to the size of the gel was placed on top of the gel and three sheets of Whatman 3 MM paper cut to size and pre-wet with 0.4 M NaOH were placed on top of the filter. An excess stack of absorbent paper was further placed on top and the whole system compressed using a 1 kg weight on top of a glass plate. Transfer was allowed to proceed for 2–3 hours. The filter was then rinsed in 2 x SSC (section 2.1.3), wrapped in SaranWrap and stored at 4°C, ready for hybridisation.

DNA was also transferred from agarose gels to nylon membranes by vacuum blotting using the Hybaid vacuum blotter (model VA). Again 0.4 M NaOH was used as blotting buffer. A sheet of Whatman 3 MM paper pre-wet with blotting buffer, was placed on top of the solid, porous support of the blotting apparatus and the nylon filter placed on top. A rubber gasket was placed over the filter so that the filter was framed. The gasket was held in place by a sealed plastic frame which would

accommodate the blotting buffer. A little blotting buffer was poured over the filter and the gel positioned over the window in the gasket. A vacuum was applied beneath the porous support, which held the gel in place. The gel was then immersed in blotting buffer and vacuum blotting continued for 20–30 minutes, after which the filter was rinsed in 2 x SSC.

2.2.14.2 Transfer of DNA from phage plates to membranes (plaque lifts)

Phage libraries were plated out as described in section 2.2.10. A circular 150 mm Hybond N+ membrane (Amersham International plc) was placed on each plate. Each filter was left in place for 1 minute whilst piercing with a syringe needle through to the agar in various asymmetric positions in order to orientate the autoradiograph to the plate later on. The filter was then transferred (DNA side up) to a pad of Whatman 3 MM paper soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) and left for 7 minutes. The filter was transferred next to a pad of Whatman 3 MM paper soaked in neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 10 mM EDTA) and left for 3 minutes. This was repeated with a fresh pad soaked in the same solution. The filter was then rinsed briefly in 2 x SSC and transferred to a 3 MM pad soaked in fixing solution (0.4 M NaOH). Filters were left to fix for 20 minutes before rinsing in 2 x SSC, after which they were ready for hybridisation. Duplicate plaque lifts were taken from each plate in order to eliminate spurious hybridisation signals upon autoradiography.

2.2.14.3 Transfer of RNA from agarose gels to membranes (northern blotting)

After electrophoresis (section 2.13.3), the gel is rinsed in distilled water and is capillary blotted overnight on to a Hybond N+ filter (Amersham) using 20 x SSC as

the blotting buffer. After transfer, the filter is washed briefly with 2 x SSC and the RNA fixed to the membrane by placing the membrane RNA side down on to a u.v. transilluminator and exposing to u.v. for 90 seconds.

2.2.15 Radiolabelling of DNA probes

2.2.15.1 Random oligonucleotide-primed synthesis from DNA fragments

The method used was based on that developed by Feinberg and Vogelstein (1983 and 1984). The Amersham Megaprime kit (RPN 1606) was used, which employs synthetic random nonamers as primers instead of the original hexamer primers. The desired DNA fragment was resolved on a low melting point agarose gel as described in section 2.2.5. Enough DNA was loaded on the gel so that there would be 250 ng of the desired fragment. The gel slice obtained was weighed and 3 volumes of sterile distilled water were added. The slice was melted at 100°C for 5 minutes, the volume of agarose measured and aliquots which contained 25 ng were dispensed into Eppendorf tubes which were stored at –20°C. For one labelling reaction, an Eppendorf tube was thawed and the agarose melted by boiling for 30 seconds. Primer solution (5 µl) was added plus sterile water to bring the final volume to 50 µl. The primer/DNA mixture was boiled for five minutes, cooled to room temperature and 10 µl labelling buffer (dATP, dGTP, dTTP in Tris-HCl, pH 7.5, 2-mercaptoethanol and $MgCl_2$), 50 µCi of [α-^{32}P]dCTP (Amersham, 3000 Ci/mmol), and 2 units of Klenow fragment (cloned) were added. The mixture was incubated for 30 minutes at 37°C, then the reaction stopped by adding 4 µl EDTA (0.25 M). The percentage incorporation (typically 50–80%) of the ^{32}P labelled dCTP was measured using Whatman DE81 paper with 0.3 M ammonium formate (pH 8) as the solvent for vertical paper chromatography. Unincorporated ^{32}P labelled dCTP migrates along with the solvent front whereas that incorporated into DNA stays at the origin. The

percentage incorporation can therefore be estimated by Cerenkov counting in a scintillation counter. The specific activity of the DNA labelled by this method is typically > 10^9 dpm/μg DNA.

2.2.15.2 End-labelling of oligonucleotides

Single stranded synthetic oligonucleotides were end-labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP as follows: 10 pmoles of oligonucleotide were labelled with 10 pmoles [γ-^{32}P]ATP (Amersham, 5000 Ci/mmol) in a reaction buffer containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 1 mM EDTA, 0.1 mM spermidine and 10 units T4 polynucleotide kinase (NBL), in a final volume of 10 μl. The reaction was carried out for 1 hour at 37°C, after which the polynucleotide kinase was inactivated by heating to 68°C for 10 minutes. The extent of incorporation was estimated and the specific activity calculated as described in the previous section.

2.2.15.3 Removal of unincorporated nucleotides using Sephadex spin columns

To separate out unincorporated nucleotides from DNA fragments and synthetic oligonucleotides of 14 bases and over, ready-made Sephadex spin columns (Chromaspin-10, Clontech) were used. A column was resuspended by inverting several times, and the matrix buffer (100 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA) was allowed to drain out by gravity. The column was then centrifuged twice for 3 minutes at 1200 rpm in a Beckman GS-6R benchtop centrifuge to remove excess buffer and to pack the column. After stopping the labelling reaction by inactivating the relevant enzyme, 1 μg fragmented salmon sperm DNA was added to the reaction mix to act as a carrier. The mix was loaded onto the centre of the prepared column, which was then centrifuged at 1200 rpm for 5 minutes. Over 95% of unincorporated nucleotides are removed by columns of this size exclusion range,

whereas only ~30% of 14-mer oligonucleotides are removed and more than 90% of DNA fragments of 26 bp and over are recovered.

2.2.16 Hybridisation of blotted DNA and RNA

2.2.16.1 Hybridisation using DNA probes labelled by random priming

The following conditions (as recommended by Sambrook *et al.*, 1989) were used for the hybridisation of DNA fragments (labelled by random priming, section 2.2.15.1) to denatured DNA or RNA immobilised on positively charged nylon membranes. The nylon membrane blotted with DNA or RNA was placed in a polythene bag containing 5–10 ml hybridisation buffer (5 x SSPE (section 2.1.3), 0.5% (w/v) SDS, 100 µg/ml denatured fragmented salmon sperm DNA and 5 x Denhardt's solution (section 2.1.3). The bag was sealed and the membrane pre-hybridised for 1 hour at 65 °C with constant agitation. The appropriate DNA probe was denatured (if double-stranded) by boiling for 5 minutes and then cooling on ice for 5 minutes. The probe was then added to the membrane plus prehybridisation buffer and the bag re-sealed. Hybridisation was allowed to take place for at least 12 hours at 65°C with constant agitation. (Alternatively, hybridisation was carried out in a Hybaid hybridisation oven in bottles which held 5 ml hybridisation solution). After hybridisation, the following washes were carried out in order to remove non-specifically hybridised probe from the filter: twice with 2 x SSC, 0.1% SDS at room temperature for 10–30 minutes, once with 1 x SSC, 0.1% SDS at 65°C for 15–30 minutes and once with 0.1 x SSC, 0.1% SDS at 65°C for 10–30 minutes (this wash was only carried out when the probe was completely homologous). After washing, the filter was wrapped in SaranWrap and exposed to X-ray film (Fuji) using an intensifying screen (Dupont, Cronex-lightning plus) for a suitable period at –70°C. Alternatively, autoradiography was carried out using a phosphoimager (FujiX BAS

1000). This method is faster and more sensitive than conventional autoradiography. The blot was exposed to a phosphoimager plate at room temperature for an appropriate length of time. The plate was then scanned and the digitalised image viewed on a Macintosh computer terminal using the MacBAS package.

Nylon membranes were stripped for subsequent re-screening by pouring over boiling 0.5% SDS which was then allowed to cool down to room temperature. Adaptations to this basic protocol were as follows:

If hybridisation needed to be carried out at lower stringency, the prehybridisation, hybridisation and post hybridisation washes were carried out at a lower temperature which was empirically determined. In some cases it was appropriate to include deionised formamide in the hybridisation solutions. The presence of formamide has the same effect on stringency as does raising the temperature, and thus the stringency can be varied in parallel hybridisations by varying the percentage formamide (up to 50%) whilst keeping the temperature constant.

Northern blots of total RNA and Southern blots of genomic DNA have low target concentrations, thus the hybridisation needs to be pushed by having the probe at high concentration; i.e. by keeping the volume to a minimum. If, on the other hand, there is a high concentration of target DNA on the filter (for example plaque lifts), hybridisation can be carried out in volumes of ~50 ml in a plastic box.

2.2.16.2 Screening the cDNA library with end-labelled degenerate primers

The method used is that described by Sambrook *et al.*(1989). Pools of short degenerate oligonucleotides whose design was based on conserved amino acid motifs were end labelled as described in section 2.2.15.2 to a specific activity of >1 x 10^9 dpm/μg. Hybridisation was carried out in the presence of 3.0 M tetramethylammonium chloride (TMACl). The presence of salts of this type serves to

ensure that all species of the degenerate pool of oligonucleotides have an equal annealing temperature, regardless of their (G+C) content, so that there will be no bias in target hybridisation. Filters were pre-hybridised for 2–6 hours in oligonucleotide prehybridisation solution (6 x SSC), 10 mM sodium phosphate (pH 6.8), 1 mM EDTA, 0.5% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, and 0.1% nonfat dried milk at the appropriate hybridisation temperature. The pre-hybridisation solution was replaced with 25 ml hybridisation solution (3.0 M TMACl, 10 mM sodium phosphate (pH 6.8), 1 mM EDTA, 0.5% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.1% nonfat dried milk, and the filters hybridised in a sealed plastic bag at the same temperature for the appropriate length of time (based on the level of probe degeneracy). The filters were washed firstly in 6 x SSC, 0.5% SDS, 10 mM sodium phosphate (pH 6.8) for 10 minutes at room temperature. This was repeated twice. Next, the filters were washed with hybridisation solution without salmon sperm DNA or nonfat milk for 5 minutes at room temperature. Finally, the filters were washed for 5 minutes with the previous solution at the temperature used for hybridisation. The filters were then wrapped in SaranWrap and autoradiographed.

(A stock solution of 6 M TMACl was prepared, activated charcoal was added to 10% (w/v) and the solution stirred for 30 minutes. The solution was then filtered through Whatman no. 1 paper and then through a Nalge 0.45-micron filter. The concentration of TMACl was then accurately measured by measuring the refractive index of the solution).

2.2.17 The polymerase chain reaction (PCR)

The polymerase chain reaction is a powerful method for the *in vitro* amplification of specific DNA fragments. The basic PCR methodology is as follows:

Standard PCR amplification reactions were set up in a total volume of 100 µl. The reaction mix contained: 10 µl of Promega 10 x *Taq* DNA polymerase buffer (500

mM KCl, 100 mM Tris-HCl, pH 9.0, 100 mM NaCl, 0.1 mM EDTA, 5 mM DTT, 1.0% Triton X-100), 0.2 mM each dATP/dGTP/dCTP/dTTP, 2 mM $MgCl_2$, 2.5 units *Taq* DNA polymerase (Promega), 20 pmol each of the two oligonucleotide primers used and a suitable amount of target DNA, depending upon its complexity (2 ng plasmid, ~30–50 ng cDNA). The optimum magnesium concentration can vary for different primer pairs between 0.5 mM and 2.5 mM over the total dNTP concentration (Innes and Gelfand, 1990). The samples were overlaid with 100 µl mineral oil to prevent evaporation during heating. Amplification was carried out using a programmable DNA thermal cycler (Perkin Elmer Cetus (model 480) or Appligene Crocodile II). The exact conditions used for amplification varied with the nature and length of the target sequence and primers. The stages involved in the PCR were an initial denaturation of 94°C, followed by repeated cycles of denaturation (94°C), annealing of primers (50–60°C), extension from primer (72°C). The cycles were followed by a final extension time at 72°C. Finally, the samples were cooled to 4°C to stop the reaction. PCR amplified DNA products were analysed by agarose gel electrophoresis (section 2.2.5).

2.2.18 PCR amplification of first strand cDNA using degenerate oligonucleotides as primers

2.2.18.1 First strand cDNA synthesis

First strand cDNA was prepared from total RNA obtained from 4–5 day old pea seedlings (section 2.2.13.1) using a Pharmacia first strand cDNA synthesis kit which uses Moloney Murine Leukemia Virus (M-MuLV) reverse transcriptase. Prior to first strand synthesis, the RNA was treated with RNase free DNase 1 to remove any contaminating genomic DNA; Total RNA (200 µg) was incubated in 30 mM Tris-HCl (pH 8.0), 10 mM NaCl, 10 mM $CaCl_2$, 6 mM $MgCl_2$ and 1 unit of RNase free DNase I (Promega), in a total volume of 50 µl, for 1 hour at 37°C. The RNA was

then ethanol precipitated.

Reverse transcriptions using the Pharmacia kit were carried out as follows: 5 µg RNA in 8 µl was heated to 65°C for 10 minutes then chilled on ice. Bulk first strand cDNA reaction mix (5 µl), containing cloned, FPLC*pure* MuLV Reverse Transcriptase, RNAguard, RNase/DNase-free BSA, dATP, dCTP, dGTP, dTTP in aqueous buffer was added to the heat denatured RNA as well as 1 µl DTT and 1 µl *Not*I-d(T)$_{18}$ primer (5′-[AAC TGG AAG AAT TCG GGG CCG CAG GAA T_{18}]-3′), 0.2 µg/µl (13 pmol/µl) or 0.4 µg (50 pmol) of the downstream antisense primer. The mixture was then incubated for 1 hour at 37°C.

2.2.18.2 PCR amplification of first strand cDNA

After first-strand synthesis was complete, the *Not*I-dT$_{18}$ primer was removed by applying the sample to a Chromaspin 100 column (Clontech). (For column use see section 2.2.15.3). The RNA/cDNA was then ethanol precipitated, resuspended in 10 µl sterile water, 3 µl of which was used in each PCR reaction. If the downstream antisense primer was used to prime first-strand cDNA, the chromaspin step was omitted and the cDNA was ethanol precipitated using ammonium acetate to remove unincorporated dNTPs, then resuspended in 10 µl sterile water. A 3 µl aliquot of this was used in each PCR reaction.

The PCR reaction mix was as detailed in section 2.2.17, with 100 pmol (~0.8 µg) of each degenerate primer, whose design was based on conserved amino acid motifs. The primers also included an *Eco*RI site at their 5′ end to facilitate cloning of PCR products. The PCR temperature profile differed from the basic profile listed in section 2.2.17. There was an initial denaturation step of 94°C for 3 min, followed by 3 x [94°C for 1 min, 32–37°C for 1 min, 72°C for 1 min]. This was followed by 30–35 x [94°C for 1 min, 46°C for 1 min, 72°C for 1 min] and then a final extension at 72°C for 2 min, after which the samples were cooled down to 4°C. This protocol was based on that described by Compton (1990). The PCR products were then

ethanol precipitated and analysed by agarose gel electrophoresis. Repeat PCR amplification (from the second cycling stage) was also carried out on 1 µl aliquots taken from the original PCR samples after amplification in order to visualise very faint bands.

2.2.19 Rapid amplification of cDNA 3′ ends (3′ RACE)

A 3′ RACE kit (purchased from Gibco BRL, Cat. No. 8373SA) was used for all 3′ RACE experiments. This kit contained all the reagents required for reverse transcription and subsequent PCR amplification, with the exception of the gene-specific primer and thermostable polymerase.

2.2.19.1 First stand cDNA synthesis using poly(A)$^+$ RNA

The protocol as recommended by the manufacturers was as follows: poly(A)$^+$ RNA (0.5 µg) from 5 day old pea seedlings (prepared as in section 2.2.13.2) was diluted to 13 µl with DEPC-treated water in a sterile DEPC-treated Eppendorf. To this was added 1 µl Adapter Primer (AP) supplied with the kit (sequence 5′-[GGC CAC GCG TCG ACT AGT AC$(T)_{17}$] -3′. The mixture was incubated at 65°C for 10 min and chilled for 2 min on ice. The tube was then centrifuged briefly in a microfuge (12,000 rpm) before adding the following: 2 µl 10 x synthesis buffer (200 mM Tris-HCl (pH 8.4), 500 mM KCl, 25 mM $MgCl_2$, 1 mg/ml BSA), 1 µl dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP) and 2 µl DTT (0.1 M). The mixture was then equilibrated to 42°C for 2 min. 1 µl SuperScript reverse transcriptase (200 U/µl) was then added and the mixture was incubated at 42°C for 30 min. After reverse transcription, the RNA was removed using *E.coli* RNase H, supplied with the kit (this step has been shown in some cases to improve the efficiency of PCR); the tube was placed on ice and 1 µl RNase H (2 U/µl) was added. The mixture was then incubated at 42°C for 10 min. The sample was then ready for PCR amplification.

The following modifications were made to the reverse transcription reaction to improve the yield of full length product and reduce the level of truncated artefacts: the amount of AP was reduced from 10 pmol to 1.2 pmol, and the incubation temperature for the reverse transcription was increased to 45°C.

2.2.19.2 PCR amplification using *Taq* polymerase

Amplification was carried out on 2 µl samples neat and 2 µl samples of 10-fold dilution of the first strand synthesis reaction mentioned above. The PCR reaction mix was assembled as follows using the kit reagents mentioned above: 5 µl 10 x synthesis buffer, 1 µl dNTP mix, 39.5 µl sterile water, 1 µl gene-specific primer (10 µM) (5´-[AAC TGC AGA ATT GGC CTC TAA]-3´), plus either 1 µl Adapter Primer (10 µM) (mentioned above) or 1 µl Universal Amplification Primer (10 µM) which was also supplied with the kit (5´-[CUA CUA CUA CUA GGC CAC GCG TCG ACT AGT AC]-3´), and 0.5 µl (2.5 U) *Taq* polymerase.

The following PCR parameters were found to be optimum: [94°C for 5 min, 75°C for 5 min] x 1, [60°C for 2 min, 72°C for 40 min] x 1, [94°C for 1 min, 60°C for 1 min, 72°C for 3 min] x 30, [72°C for 15 min] x 1. Cycling was carried out using the Appligene Crocodile II thermocycler using a temperature probe inside a 0.5 ml Eppendorf tube. Also a "Hot Start" was employed, whereby the *Taq* polymerase then the mineral oil (pre-heated to 75°C) were added to the samples once the temperature had reached 75°C.

2.2.19.3 PCR amplification using *Pfu* polymerase

First-strand cDNA (1 µl), prepared using the modified protocol (section 2.2.19.1), was used as template. PCR was carried out using 10 pmol each of the gene-specific primer and the Adapter Primer (both described above), using 2.5 U native *Pfu*

polymerase (Stratagene), in the presence of 20 mM Tris-HCl (pH 8.2), 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 15 mM $MgCl_2$, 0.1% Triton X-100 and 100 µM each dNTP. (final volume = 50 µl). The following cycling parameters were found to work well: [95°C for 5 min, 75°C for 5 min] x 1, [58°C for 10 min, 75°C for 80 min] x 1, [95°C for 1 min, 60°C for 5 min and 75°C for 10 min] x 30, [75°C for 60 min] x 1. Again, a hot start was performed as described in the previous section.

2.2.20 Manual DNA sequencing

Sequencing of both double and single-stranded plasmid templates was carried out using the dideoxy chain termination method (Sanger *et al.*, 1980). A Sequenase Version 2.0 kit (United States Biochemical Corp.) was used for all manual sequencing reactions. Sequenase 2.0 is a genetic variant of bacteriophage T7 DNA polymerase created by *in vitro* genetic manipulation. in which the 3´–5´ exonuclease activity of the wild type enzyme has been completely removed. Its other features include high processivity and high speed. The primers used for sequencing were the M13 Reverse or –20 Universal primers (see Fig. 1.4.2), which were purchased from Pharmacia. Alternatively, unique primers designed according to known sequence data from the insert were used; these were also purchased from Pharmacia.

2.2.20.1 Annealing and sequencing reactions

For single stranded templates (prepared as in section 2.2.9.4) primer annealing was carried out as follows: 1 pmol primer and 1 µg template were mixed with 2 µl 5 x Sequenase reaction buffer (200 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl), in a final volume of 10 µl. The mixture was heated to 65°C for 2 min and allowed to cool to < 30°C over a period of 30 min. The tube was then placed on ice and the labelling reaction carried out as follows: to the annealed primer + template were added 2 µl dGTP labelling mix (7.5 µM dGTP, 7.5 µM dCTP, 7.5 µM dTTP)

(this is usually diluted five-fold), 5–10 μCi [α-^{35}S]dATP (600 Ci/mmol) and 2 μl Sequenase enzyme (1.5 U/μl). The labelling reaction was carried out for 2–5 minutes at 18°C.

After this time, four 3.5 μl samples were taken from the labelling mix and transferred to four Eppendorf tubes containing 2.5 μl of each of the four ddNTP termination mixes (pre-incubating at 38°C). ddGTP termination mix contains 80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddGTP, 50 mM NaCl. The other termination mixes are identical, except that they contain their appropriate ddNTP at 8 μM). The termination reactions were incubated at 38°C for 5 minutes before adding 4 μl stop solution (95% formamide, 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF). For reading beyond 300 nucleotides, the labelling mix was used undiluted, more [α-^{35}S]dATP (20 μCi) added, and the labelling time increased to 5 min.

For double stranded templates, alkali denaturation and primer annealing were carried out on 5 μg plasmid (from a maxiprep, section 2.2.9.3) according to the method described by Stephen *et al.* (1990). The plasmid solution was made up to 9 μl with TE. NaOH (2 M, 1 μl) was added and the mixture incubated for 15 minutes at 37°C to denature the plasmid. Sequencing primer (10 pmol) was then added and the mixture vortexed briefly. The alkali was neutralised by adding 3 μl sodium acetate (3 M, pH 5.2) and the DNA precipitated with ethanol (75 μl). Following centrifugation at 12,000 rpm for 10 minutes in a microfuge, the pellet was washed with 100 μl 70% ethanol, dried and resuspended in 8 μl water. To this was added 2 ml of the 5 x Sequenase reaction buffer mentioned above. The labelling and termination reactions were then carried out as described for single-stranded sequencing.

2.2.20.2 Gel electrophoresis and autoradiography

Polyacrylamide-urea denaturing gels were prepared as follows: a 40% stock solution

of acrylamide (38% acrylamide, 2% N´N´-methylene bisacrylamide) was prepared of which 15 ml was added to 46 g urea (BRL ultra PURE) along with 20 ml 5 x TBE (section 2.1.3) and the mixture made up to 100 ml with distilled water. To the mixture 400 µl fresh 10% ammonium persulphate and 70 µl TEMED were added and the gel poured between two clean glass plates, which were separated by spacers of 2 mm in width. The notched gel plate was treated with dimethyldichlorosilane to prevent the gel from adhering to both plates. The gel was left for at least 2 hours to polymerise fully then transferred to the sequencing apparatus (BRL model S2), the reservoirs of which were filled with 1 x TBE buffer. A sharkstooth comb was then fitted so that the points indented (but did not pierce) the surface of the gel. The gel was pre-run for 30 minutes at 30 mA and 40 W. The sequencing reaction tubes were heated to 75–80°C prior to loading 2–3 µl of each reaction mix; the wells were flushed with buffer before loading. Electrophoresis was carried out at 30 mA and 40 W. Two separate loadings were necessary to maximise the length of sequence that could be read: the first loading was subjected to electrophoresis for 3–4 hours and the second loading for 1.5–2 hours. After electrophoresis and removal of the notched gel plate the gel was fixed by soaking in a 2 litre bath of 5% (v/v) acetic acid and 10% (v/v) methanol for 15 min to remove the urea. The gel was then drained for a few minutes, transferred onto a sheet of Whatman 3MM paper, covered in cling film and dried under vacuum on a gel drier (model 1125B Bio-Rad) for 1 hour at 80°C. After drying, the gel was exposed directly onto Fuji X-ray film overnight at room temperature. Longer exposures were used as necessary. Following autoradiography it was usually possible to read 200–300 bases.

2.2.21 Automated DNA sequencing

Automated sequencing was carried out on double stranded plasmid template using the Applied Biosystems *Taq* DyeDeoxy Terminator Cycle Sequencing Kit. The method employs random dideoxy chain termination, and each of the ddNTPs is tagged with a

different fluorescent dye, so that the terminated products can be distinguished by the wavelength emitted when exposed to u.v. All four termination reactions can therefore be carried out in the same tube. Terminated DNA fragments were resolved, and their incorporated fluorescent dideoxy terminators detected using the Applied Biosystems Automated Sequencer (model 373A). Sequencing primers were as used for manual sequencing (section 2.20.1).

2.2.21.1 Double stranded sequencing using the *Taq* Dyedeoxy Terminator Cycle Sequencing Kit.

Plasmid template was prepared as described in section 2.2.9.2. Sequencing primer (~0.3 pmol) was added to 1–2 µg template DNA and the volume made up to 10.5 µl with sterile water in a 500 µl PCR tube. Bulk sequencing reaction mix was made up according to requirement. For example, for 20 sequencing reactions, the following reagents were mixed: 80 µl 5 x Terminator Ammonium Cycle Sequencing (TACS) buffer (400 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM $(NH_4)_2SO_4$, pH 9.0), 20 µl DyeDeoxy A Terminator and 20 µl of each of the other DyeDeoxy Terminators, 20 µl dNTP mix (750 µM dITP, 150 µM dATP, 150 µM dTTP, 150 µM dCTP), and 10 µl Ampli*Taq* DNA Polymerase (8 U/µl). A 9.5 µl aliquot of this mixture was dispensed into each template/primer solution and mixed. The mixtures were then overlaid with 40 µl mineral oil and the tubes placed in a Perkin Elmer Cetus thermal cycler (model 480), pre-heated to 96°C. Immediately after placing the tubes in the thermal cycler, thermal cycling was begun as follows: 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes, and this cycle was repeated 25 times before cooling to 4°C.

The extension products were purified away from the unincorporated DyeDeoxy Terminators by phenol/chloroform extraction: sterile water (80 µl) was added to each sample, and the sample was removed from the mineral oil and transferred to a fresh 1.5 ml Eppendorf tube. A 100 µl aliquot of

phenol/water/chloroform (68:18:14, Applied Biosystems) was added and the mixture vortexed and centrifuged (12,000 rpm in a microfuge) for 5 minutes. The lower organic phase was discarded and the aqueous phase was re-extracted with phenol/water/chloroform (100 µl). The extension products were then ethanol precipitated by adding 10 µl 3 M sodium acetate (pH 5.2) plus 250 µl absolute ethanol and incubating at –70°C for 15 minutes. After centrifugation (in a microfuge) at 12,000 rpm for 15 minutes, the pellet was washed with 70% ethanol at –20°C and dried under vacuum. The DNA pellet was then resuspended in 4 µl of a mixture containing 5 µl deionised formamide plus 1 µl 50 mM EDTA, pH 8.0. When the gel was ready for loading, the samples were heated at 90°C for 2 minutes to denature, then transferred onto ice.

2.2.21.2 Gel electrophoresis and detection of extension products using the Automated Sequencer

A 6% polyacrylamide-urea gel was cast as follows: the u.v. transparent gel plates, spacers and comb (Applied Biosystems) were washed with Alkonox detergent and dried with non dye-containing tissues. The gel was cast using pre-made acrylamide solutions (National Diagnostics). Monomer solution (48 ml) was mixed with buffer reagent (12 ml). To this was added .48 ml 10% ammonium persulphate. The mixture was then poured and allowed to polymerise for 2 hours, after which the tape was removed and the outside of the gel plates cleaned thoroughly with Alkonox. The gel was then fitted into the ABI sequencing apparatus and the Plate Check program run to check that the plates were clean in the region read by the laser. The reservoirs were filled with 1 x TBE, and the sharkstooth comb (24 well) fitted. The gel was then pre-run at 512 V for 5 min. The sample data were stored on a new sample sheet on the Macintosh, and the settings for automatic data collection and analysis were selected. Odd numbered samples were loaded first (after flushing the wells) and the samples electrophoresed at 512 V for 5 min. Even numbered samples were then loaded and

electrophoresis continued overnight. Printouts of analysed data were obtained, which show the profile of peaks for each fluorescently-tagged base together with the sequence. Up to 400 bases could be read without ambiguities on the automated sequencer.

2.2.22 Analysis of sequence data

Sequence data were stored using the Wisconsin GCG package (accessed via UNIX) and various programs such as MAP, BESTFIT, GAP and PILEUP were used in sequence analysis and comparison with known methylase sequences. Also GenEmbl databank searches such as FASTA and WORD were carried out with unknown sequences to establish which of the sequences on the database best matched the sequence in question, and whether these matches were significant.

Chapter 3 Results

3.1 Introduction

A number of different approaches were tried in order to achieve the aims of this project. However, the only approach that proved successful was the use of degenerate primers based on conserved amino acid motifs. These were used to amplify pea MTase-specific cDNA fragments by PCR (section 3.6). The 3′ end of pea MTase cDNA was then cloned using a 3′ RACE strategy (section 3.7). Each of the different approaches will be discussed in the following sections.

3.2 Screening a pea cDNA library for MTase clones using a mouse MTase cDNA clone

The aim of these experiments was to establish whether suitable hybridisation conditions could be found for library screening using probes derived from the mouse MTase cDNA clone (Bestor *et al.*, 1988), which was kindly provided by Timothy Bestor.

3.2.1 Probe selection

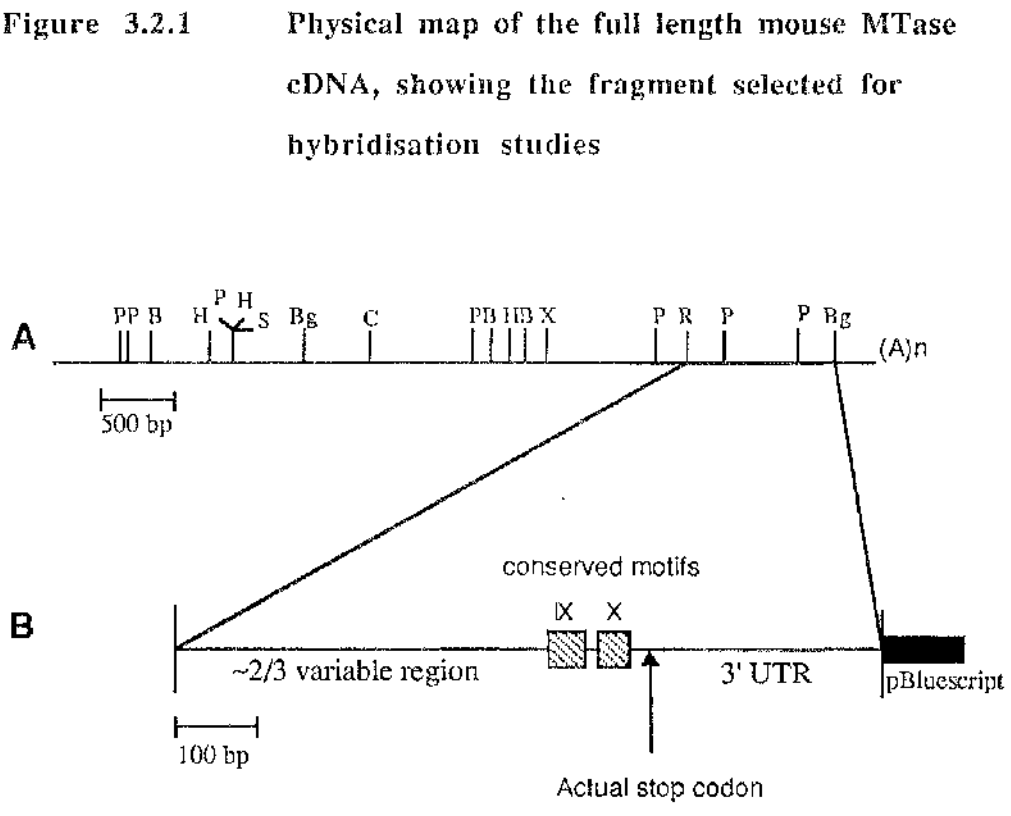
The mouse MTase cDNA clone (pMG) represents the entire cDNA except for the first 20 nucleotides from the 5′ UTR and the last 50 nucleotides from the 3′ end. The cDNA is cloned into pBluescript SK+; the 5′ end is cloned into an *Eco*RI site and the 3′ end is cloned into a *Bgl*II/*Bam*HI hybrid site. In the construction of the clone, the *Sma*I site of the vector has been duplicated so that the insert can be excised by digestion with *Sma*I. The restriction map of the published full length cDNA is shown in figure 3.2.1.

At the time of this work, the mouse MTase nucleotide sequence had not yet been revised, so we were relying on the published nucleotide and amino acid sequence data for the selection of a suitable probe. According to the published MTase

amino acid sequence, only two motifs had been identified that were conserved between mouse MTase and the prokaryotic MTases (motifs IX and X). The region of the cDNA specifying these motifs is contained within the ~1 kb *Eco*RI/*Bgl*II fragment shown in figure 3.2.1. At the time, it was believed that this fragment contained only coding sequences. Later analysis of the revised amino acid sequence data on the GenEmbl database revealed that the published position of the stop codon (lying ~30 nucleotides downstream of the *Bgl*II site) was wrong. The stop codon actually occurs ~30 nucleotides downstream of the region specifying motif X (see Fig. 3.2.1). Therefore a significant proportion of the probe used consisted of 3′ UTR sequences. Had this been known at the time, an alternative segment of the mouse MTase cDNA would have been used as a probe.

Isolation of the 1 kb fragment was achieved firstly by digesting clone pMG with *Sma*I, which excises the insert in one piece (~5.0 kb). The insert was then purified from LMP agarose by phenol/chloroform extraction and digested with *Eco*RI to give two fragments of ~4 kb and ~1 kb. The latter fragment was used as a probe in the hybridisation experiments described in this section.

Figure 3.2.1 **Physical map of the full length mouse MTase cDNA, showing the fragment selected for hybridisation studies**



Enzyme key : B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; H, *Hin*cII; P, *Pst*I; R, *Eco*RI; S, *Sac*I; X, *Xho*I

A. The published restriction map of the full length mouse MTase cDNA (adapted from Bestor *et al.*, 1988).

B. The probe used for the experiments described in this section corresponds to the region between *Eco*RI and *Bgl*II at the 3′ end of the cDNA and is approximately 1 kb in length. The 3′ *Bgl*II site is the one used to clone the 3′ end of the cDNA into pBluescript II SK+ in the clone pMG (kind gift from Timothy Bestor). The probe sequence specifies approximately two thirds of the variable region, the conserved motifs (IX and X) and contains most of the 3′ UTR. The position of the stop codon (after revision of the sequence) is indicated on the diagram.

3.2.2 Hybridisation: theoretical considerations

The theoretical aspects of hybridisation are reviewed by Anderson and Young (1985) and Sambrook *et al.* (1989).

The reversible melting temperature (T_m) of perfectly matched DNA/DNA hybrids of length greater than 100 nucleotides is given by the following empirical equation (Bolton and McCarthy, 1962):

$T_m = 81.5°C + 16.6[\log Na^+] + 0.41(\%G+C) - 0.63(\%\text{ formamide}) - 600/l$

where l = length of hybrid

This equation applies only to Na^+ concentrations of between 0.01 and 0.4 M; the second parameter can be omitted for Na^+ concentrations > 0.4 M. Since hybridisation is carried out using 5 x SSPE or SSC (= 1 M Na^+), the equation simplifies to:

$T_m = 81.5°C + 0.41\ (\%G+C) - 0.63(\%\text{ formamide}) - 600/l$

And, since the average length of a labelled probe obtained from random nonamer-primed synthesis is 150 nucleotides and the average G+C content of eukaryotic DNA is 40%, the T_m of perfectly-matched hybrids works out to be 94°C in the absence of formamide. Hybridisations are routinely carried out at 20–25°C below the theoretical T_m (the makers of Hybond N^+ recommend 65°C). At this temperature the rate of hybridisation of perfectly-matched hybrids is at a maximum. The final post-hybridisation wash for completely homologous probes is as stringent as possible and is routinely 0.1 x SSC, 0.1% SDS, 0% formamide at 65°C.

When hybridisation is carried out between related sequences, the T_m decreases by ~1°C for every 1% mismatch (Bonner *et al.*, 1973). If the two sequences share 75% identity, the T_m will therefore be 69°C in the absence of formamide, and the appropriate hybridisation temperature will be between 44 and

49°C. Homologies have been detected with as much as 35% mismatch at 50°C below the T_m of the perfectly-matched hybrid (Howley *et al.*, 1979). Of course, the T_m between mismatched hybrids will depend upon the distribution of the mismatches; empirical determination of appropriate hybridisation and washing conditions is usually necessary.

3.2.3 Screening pea genomic blots with part of the mouse MTase cDNA

Because mammals and plants are distantly related eukaryotes, we were expecting low identities to exist at the nucleotide level between the mouse and pea MTase cDNAs; identities which may be too low for library screening. The fragment of the mouse MTase cDNA clone shown in Fig. 3.2.1 was selected for use as a probe because it was hoped to contain regions of sufficient identity with the pea MTase cDNA for library screening to be possible.

In an attempt to establish appropriate hybridisation conditions, a Southern blot of a series of *Eco*RI digested samples of pea genomic DNA (10 µg) was prepared. The filter was probed cut into strips and probed with the 1 kb fragment of pMG (Fig. 3.2.1). The stringency of hybridisation and post-hybridisation washes was varied between each strip. The autoradiograph is shown in figure 3.2.2. It is clear that very low stringencies are necessary for any cross-hybridisation between the probe and pea genomic DNA. A signal was obtained at the lowest stringency (strip 1), but little detectable signal was obtained when the stringency of hybridisation was increased to 10% formamide at 42°C (which should allow up to ~25% mismatch between hybrids). It is therefore highly likely that there is 30% or more mismatch between the region of the mouse MTase cDNA used as a probe and pea MTase sequences. The problems with probing at low stringency are that there are high levels of background hybridisation, which in this case has not been removed even by prolonged washing at stringencies which should allow no less than 37% mismatch

between hybrids (in the case of strip 1). From the gel photograph, it appears that at least some of the hybridising bands in the pea genomic digest correspond to satellite sequences, and the strength of signal is largely a consequence of the amount of DNA present. Whether any of the hybridising bands correspond to MTase specific sequences is unclear. The control Southern blot (of mouse genomic DNA digested with *Eco*RI) shows that the sensitivity is sufficient to detect the homologous target, although non-specific hybridisation to satellite bands does not appear to have been removed completely by washing at moderate stringency.

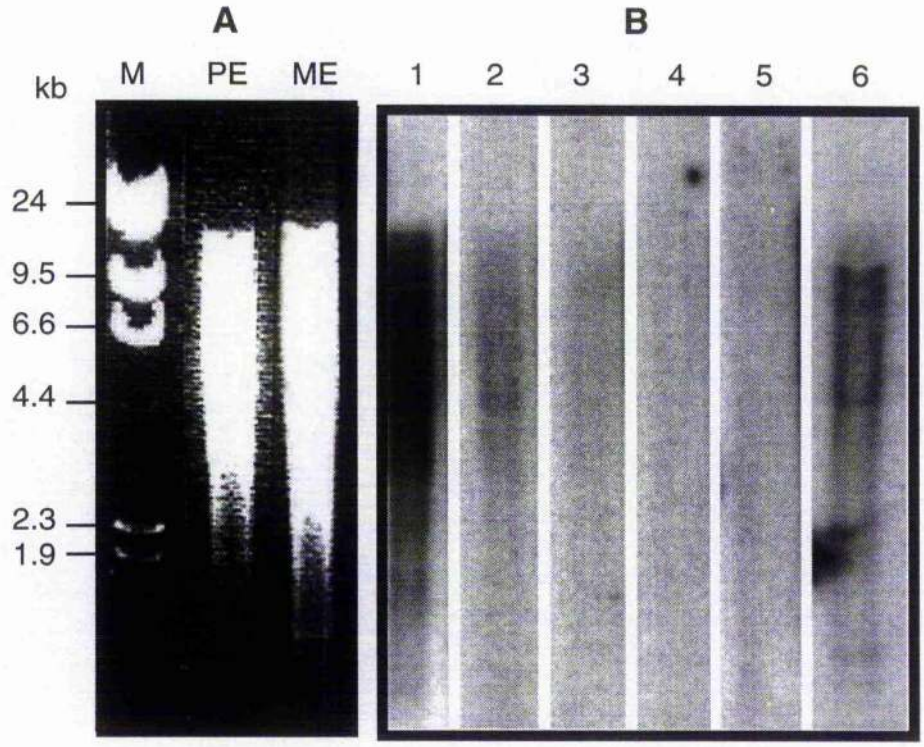
A trial screening of a portion of the pea cDNA library in λgt11 (section 2.1.5.1) was undertaken to assess whether it was feasible to screen the library with the cDNA clone. 35,000 plaques were plated on a 150 mm plate as described in section 2.2.10. Plaque lifts were performed in duplicate as described in 2.2.14.2, and the filters were probed with the *Eco*RI fragment of pMG at 42°C in the absence of formamide. Initially, the final post-hybridisation wash was for 1 hour at 50°C in 2 x SSC; however, the levels of background hybridisation were high. Background hybridisation was still high after washing at 55°C in 2 x SSC, and was only significantly reduced when the stringency of washing was increased to 1 x SSC at 55°C, which was estimated to be too high for specific hybridisation.

Figure 3.2.2 Screening pea genomic Southern blots with a fragment of the mouse MTase cDNA

Six 10 μg samples of pea genomic DNA plus a 10 μg sample of mouse genomic DNA (both prepared according to section 2.2.12.1) were digested with *Eco*RI then resolved on a 0.8% agarose TBE gel (section 2.2.5), together with marker DNA (λ DNA digested with *Hin*dIII).

A. A picture of part of the gel showing 10 μg *Eco*RI digested pea genomic DNA (lane PE) and 10 μg *Eco*RI digested mouse genomic DNA (lane ME) and marker (lane M).

B. The gel was alkali blotted onto a Hybond N^+ filter (section 2.2.14.1), and the filter cut into strips. Each strip was probed with the 1 kb fragment of the mouse MTase clone (pMG), which is shown in Fig. 3.2.1. (For details of probe labelling and hybridisation see sections 2.2.15.1 and 2.2.16.1). The filter strips were all hybridised at 42°C in 5 x SSPE, but with different concentrations of formamide, as shown in the table. Post-hybridisation washes were carried out as described in section 2.2.16.1, except the stringency of the final post-hybridisation wash was reduced. The final washes (2 x 30 min), were at the temperatures and salt concentrations (x SSC) shown below the figure.



strip no.	% formamide	wash temperature	x SSC
1	0	50°C	2
2	0	55°C	2
3	10	60°C	2
4	10	65°C	2
5	20	65°C	1
6	50	65°C	1

Strips 1–5 *Eco*RI digested pea genomic DNA (10 µg)

Strip 6 *Eco*RI digested mouse genomic DNA (10 µg)

3.3.4 Discussion

It was decided on the basis of these results not to attempt library screening using this probe. If the mRNA would have been expected to be relatively highly represented in the library, this approach perhaps would have been valid. However, screening at low stringencies risks the isolation of false positives, especially if the genuine mRNA is poorly represented in the library. This is well illustrated by the work of Scheidt *et al.* (1994a) who isolated a number of *Arabidopsis* genomic clones by screening at low stringency with the same region of the mouse MTase cDNA as was used in the experiments described here. None of these clones, however, turned out to be convincing MTase candidates (the reasons for this will be described in more detail in the following section). When the nucleotide sequence of the analogous part of the pea MTase cDNA which was eventually cloned (see section 3.7) was compared to the region of the mouse MTase cDNA used as a probe in this experiment, it is clear that the approach described in this section had very little chance of success. When the sequences were aligned using GAP, the identity was 47% over the entire length of the sequences; the program BESTFIT gave an alignment over ~270 bp (covering the conserved motifs IX and X) of 64% identity. So, even at best, the level of identity is at the very limit required for heterologous screening, and this is confined to less than one third of the probe length.

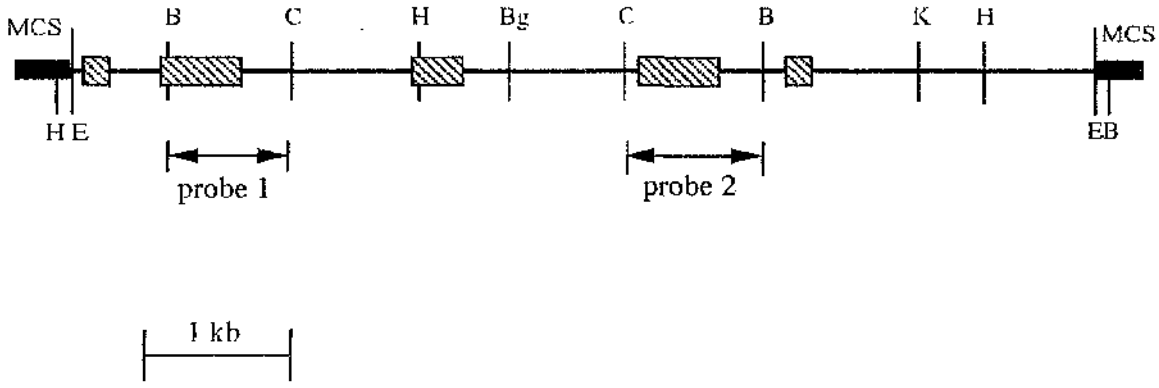
3.3 Screening a pea cDNA library for MTase clones using a putative *Arabidopsis* MTase genomic clone

This clone (kindly provided by Adolf Graessman) had been obtained by screening an *Arabidopsis* genomic library at low stringency with a 1 kb *Eco*RI fragment of the mouse MTase cDNA. (The same *Eco*RI fragment was used as a probe in the experiments described in the previous section (see Fig. 3.2.1).) The putative *Arabidopsis* MTase genomic clone provided (pGr) consisted of a 7 kb *Eco*RI fragment which contained conserved motifs (figure 3.3.1). There was no information about the sequences of the conserved motifs because this work had not yet been published. However, it was assumed that this clone must be reasonably homologous to the mouse cDNA for it to have been isolated by library screening (see section 3.2.2). It was therefore expected that the pea MTase cDNA would be closely related to regions of this genomic clone, which would make this clone a suitable probe for library screening.

3.3.1 Probe selection

It was decided on the basis of the map sent with the clone, that the best probes to use corresponded to the 1 kb and 0.9 kb fragments obtained upon double digestion of pGr with *Bam*HI and *Cla*I. This was because they contained minimum amounts of intronic sequences. Fig 3.3.1 shows a copy of the map sent with the clone, and the location of the probes used. Since the 1 kb and 0.9 kb fragments could not be resolved, they were used together as probes.

Figure 3.3.1 **Linear map of a putative methylase genomic clone from *Arabidopsis* (pGr) showing fragments used as probes**



This is a copy of the restriction map provided with the clone. This clone consists of a 7 kb *Eco*RI fragment, cloned into pBluescript II SK +, of the putative *Arabidopsis* MTase genomic clone (Clone 11). Clone 11 was isolated by heterologous screening with a fragment of mouse MTase cDNA as probe (Scheidt *et al.*, 1994a).

The restriction map is shown in the 5′ to 3′ direction of the coding strand. The positions of proposed exons (which contain putative conserved motifs) are indicated by boxes. A key to the restriction sites is as follows: E = *Eco*RI, H = *Hin*dIII, B = *Bam*HI, C = *Cla*I, Bg = *Bgl*II, K = *Kpn*I.

The two fragments obtained upon digestion with *Bam*HI and *Cla*I, are indicated below the map. These two fragments were used together as probes for all experiments described in this section.

3.3.2 Establishment of appropriate conditions for library screening

The theoretical aspects of hybridisation are discussed in section 3.2.2. It was expected that the level of nucleotide sequence identity between the *Arabidopsis* MTase and pea MTase within the conserved regions would be high. If it is assumed that there is about 25% identity at the nucleotide level between the coding sequences of the *Arabidopsis* genomic clone and the pea MTase cDNA, the appropriate temperature for hybridisation is about 50°C in the absence of formamide, and appropriate washing conditions are 1 x SSC at ~60°C. This assumption has since turned out to be quite valid for the *Arabidopsis* cDNA clone isolated by Finnegan and Dennis (1993), which shares ~75% identity at the nucleotide level in the C-terminal domain with pea MTase.

With this in mind, a trial hybridisation experiment was carried out with Southern blots of pea genomic DNA. A series of 20 µg samples of pea genomic DNA (prepared according to section 2.2.12.1), were digested with *Hin*dIII (section 2.2.4). The probe used consisted of two fragments of approximately 1 kb in length each (see Fig. 3.3.1). These are treated as a single sequence of 2 kb in length in the following calculations. The haploid genome size of *Pisum sativum* (a diploid plant) is 5.0 pg or 4.8 x 10^9 bp (Rogers and Bendich, 1991). Thus, a single copy sequence of 2 kb in length would represent 1000/4.8 x 10^9 of the genome, or ~8 pg in a 20 µg genomic digest. To test the sensitivity of the hybridisation, a control lane which consisted of undigested pGr plasmid (50 pg) was also included in the hybridisation experiment. Since the length of the insert plus vector is ~10.7 kb, the amount of the 2 kb fragment present would be ~10 pg. The limit of sensitivity was also tested by including a lane with 5 pg undigested pGr, which should contain ~1 pg of the 2 kb sequence. The pea genomic digests plus the control pGr samples were resolved on a 0.8% agarose TBE gel (section 2.2.5), along with marker DNA (*Hin*dIII digest of λ DNA). The gel was alkali blotted onto a Hybond N^+ filter according to section

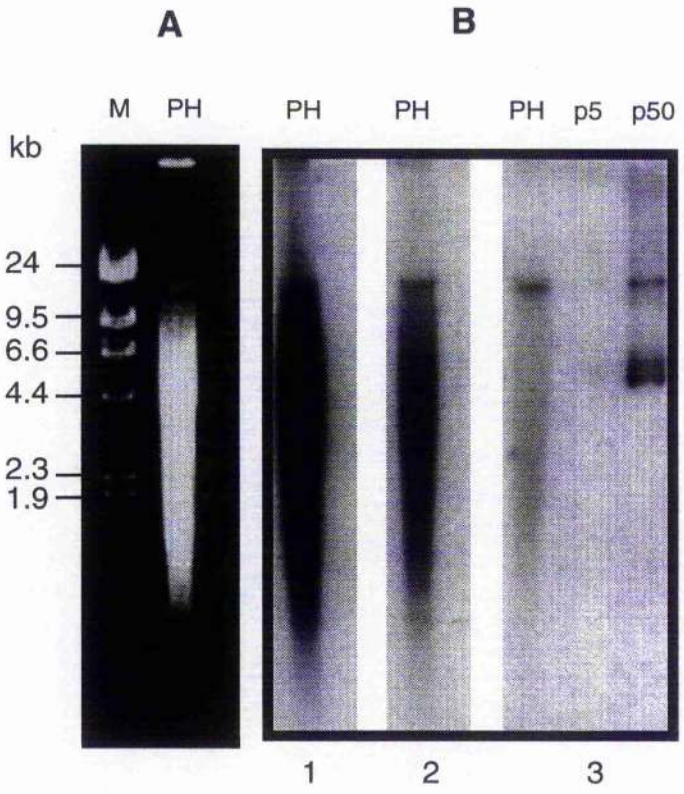
2.2.14.1). The filter was then hybridised at 50°C in the absence of formamide according to section 2.2.16.1 with the pGr probes (Fig. 3.2.1), which were labelled to a specific activity of > 10^9 dpm/μg (section 2.2.15.1). The filter was then washed twice in 2 x SSC, 0.1% SDS at room temperature. After this, the filter was cut into strips and the stringency of subsequent washes was varied between each strip. The strips were then autoradiographed for 2 days at -70°C with an intensifying screen. The autoradiograph and a picture of part of the agarose gel are shown in figure 3.3.2.

As can be seen from Fig. 3.3.2, the sensitivity of probing is sufficient to give a signal with 50 pg pGr (i.e.~10 pg target sequence). There is also a very faint signal with 5 pg pGr. Because of reduced stringency of hybridisation and washing, individual bands are difficult to discern in the pea genomic digests due to background hybridisation. A high molecular weight band hybridised even at the highest stringency used. However, this was judged to be background hybridisation to a satellite band which is is visible on the gel. Apart from this band, the signal from the pea genomic digest on filter 3 was very weak, so the stringency of washing was judged to have been too high. In filter 2, there appears to be sufficient signal for specific hybridisation to the probe to have taken place. Individual bands can be observed within the smear. It is not clear whether these bands are due to the specific hybridisation to target sequence, or whether this is background hybridisation to satellite sequences. However, when the results of this experiment are compared to the results of probing pea genomic digests with part of the mouse MTase cDNA (see section 3.2), it is clear that a positive signal is obtained at much higher stringencies of hybridization with the pGr probe.

Since the intensity of the signal is much reduced in filter 3, it was decided to screen the library using a hybridisation temperature of 50°C with washes corresponding to those used for filter 2. This should enable detection of cDNA clones which share 75% identity (or more) with regions of the probe.

Figure 3.3.2 Hybridisation of the pGr probe to Southern blots of pea genomic DNA

Samples of pea genomic DNA (20 µg) were digested with *Hin*dIII and resolved on a 0.8% agarose TBE gel (A). The gel was alkali blotted to a Hybond N^+ filter and the filter cut into strips (1–3) and probed with pGr *Bam*HI/*Cla*I fragments (Fig. 3.3.1). The fragments (25 ng) were labelled by random priming (section 2.2.15.1). Hybridisation was carried out at 50°C in 5 ml hybridisation buffer (section 2.2.16.1), containing 5 x SSPE and no formamide. The strips were washed twice for 30 min in 2 x SSC, 0.1% SDS at room temperature and individual strips were subject to additional washes of different stringencies. The autoradiograph is shown in B.



Strip	1. Final wash	=1 x SSC, 0.1% SDS		at 55°C;	2 x 30 minute washes
"	2.	"	=	"	at 60°C	"
"	3.	"	=	"	at 65°C	"

Lane M	Marker (10 µg of *Hin*dIII digested λ DNA)
Lanes PH	20 µg of pea genomic DNA digested with *Hin*dIII
Lane P50	50 pg of pGr (undigested) diluted in carrier (λ DNA, 2 ng/µl)
Lane P5	5 pg of pGr "

3.3.3 Screening a cDNA library with *Bam*HI/*Cla*I fragments of pGr

The library used in this experiment was the pea cDNA library in λgt11 (section 2.1.5.1). Although the tissue used to construct this library (5-day old pea shoot tips), contains a relatively high specific activity for MTase, the actual amount of MTase protein is low (Yesufu *et al.*, 1991). Therefore it can be assumed that pea MTase mRNA(s) represent low abundance mRNAs. The number of cDNA clones needed to be screened to ensure that low-abundance types of mRNA (< 10 copies per cell) are represented in a clone bank is given by the following equation:

$$N = \ln (1 - P)/\ln (1 - 1/n)$$

where N = the number of clones required, P = the desired probability (usually 0.99), and $1/n$ = the fraction of the total mRNA population which the low-abundance mRNA types represent. Because the population of low-abundance mRNAs is not known for most plant cells, one can assume that in the mRNA population of interest there are about 10,000 unique low-abundance mRNAs, and that these represent about 30% of the total mRNA population; thus, $1/n$ = 1/33,000 (Slightom and Quemada, 1991). Therefore, the number of cDNA clones required to be screened to have a 99% probability of obtaining a single clone of a particular low abundance mRNA is ~140,000. It was decided to screen three times as many clones to ensure a high probability of obtaining a MTase clone.

The library was titred and plated out onto six 150 mm plates at a density of 70,000 plaques per plate (2.2.10). Duplicate plaque lifts were carried out and the DNA fixed to the filters as described in section 2.2.14.2. The filters were hybridised overnight at 50°C in 100 ml of hybridisation solution (section 2.2.16.1) with 50 ng of the probe (Fig. 3.3.1), which was labelled to a specific activity of > 10^9 dpm/µg. The filters were washed in large volumes to ensure minimum background. The post-

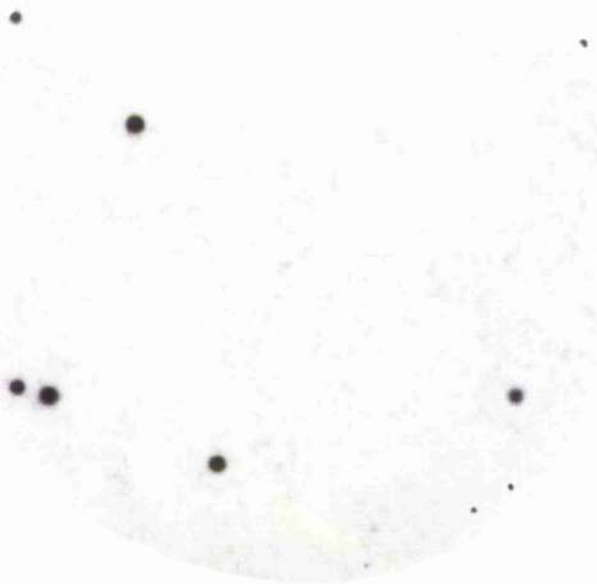
hybridisation washes were as described in section 2.2.16.1, except that the final post-hybridisation washes were at 60°C in 1 x SSC, 0.1% SDS (2 x 30 min washes).

From the initial screening, one strong positive signal and five weaker positives were identified upon autoradiography of the filters. All of the positives hybridised on duplicate filters. Agar plugs corresponding to the approximate position of the hybridising plaques were removed from the plates and placed into 1 ml aliquots of SM buffer (section 2.2.10). The eluted phage were then plated out at a density of 100–300 plaques per 90 mm plate, except for clone λ5.1, which was plated out at a density of 1000 plaques on a 150 mm plate, because the agar plug taken was larger. Duplicate plaque lifts were taken from the plates and were screened with the same probe. The stringency of hybridisation of the secondary screening was increased to 55°C, and the final post-hybridisation wash was with 1 x SSC at 60°C. Of the 6 clones, 3 still gave a strong signal upon secondary screening. These were called λ5.1, λ1.1 and λ6.2. Clone λ5.1 gave the strongest signal in the secondary screening. An autoradiograph of the secondary screening of clone λ5.1 is shown in figure 3.3.3.

Figure 3.3.3 An example of positive clones obtained upon secondary screening with pGr

The positive clones which were obtained upon primary screening of the library with the *Bam*HI/*Cla*I fragments of pGr (Fig. 3.3.1), were picked and plated out at a density of 100–300 plaques per 90 mm plate according to section 2.2.10. (clone λ5.1 was plated out at a density of ~1000 plaques on a 150 mm plate, because a larger plug was picked).

Plaque lifts were performed in duplicate as described in section 2.2.14.2. The filters were then hybridised in the absence of formamide and at 55°C (section 2.2.16.1) with the same probe as was used for library screening. (The probe was labelled by random priming, section 2.2.2.15.1.) The final post-hybridisation was was in 1 x SSC at 60°C. Of the six clones identified in the primary screening, three still hybridised strongly after secondary screening. An autoradiograph of the secondary screen of clone λ5.1 (which was the most strongly hybridising clone) is shown below. All the positives hybridised on duplicate filters. (The faint spots at the edge of the filter are radioactive ink).

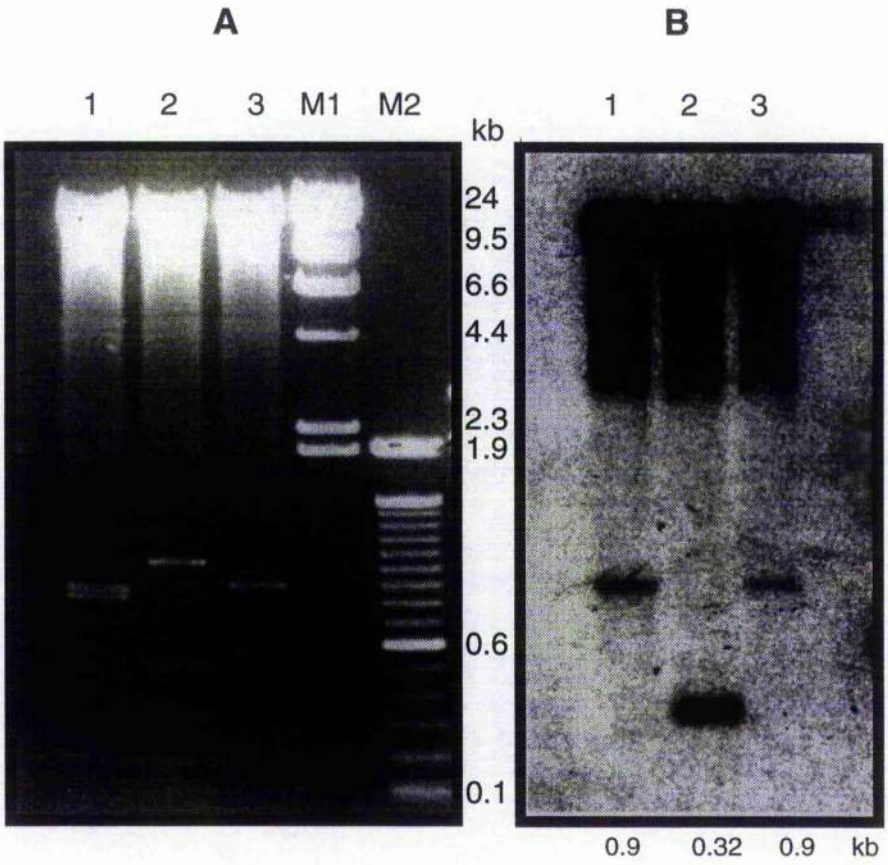


Well isolated positive plaques were picked and plated out for tertiary screening, which was carried out under the same conditions as used for secondary screening. One of the duplicate filters was hybridised at 60°C and washed at 1 x SSC at 65°C. Clone λ5.1 still gave a strong signal at this stringency, whereas the strength of the signal from the other two clones was reduced. Well isolated positive plaques were picked and used to make λ DNA. Small scale preparations of λ DNA were made from liquid lysates using the Promega Magic Lambda Prep columns (section 2.2.11). Aliquots containing 10 μg λ DNA were digested with *Eco*RI to excise the inserts. Aliquots of each digest (1 μg) were resolved on a 1% TAE gel, which was then blotted and probed with the same probe and using the same conditions as in the secondary screening. A picture of the agarose gel and the autoradiograph is shown in figure 3.3.4. Each λ clone gave rise to more than one insert upon digestion with *Eco*RI. These inserts are either independently cloned fragments or are due to the presence of *Eco*RI sites within one cloned fragment. Only one insert-derived fragment from each clone was found to hybridise to the probe. Although this does not prove that the other non-hybridising fragments are not contiguous to the hybridising fragments, it was decided to concentrate on the hybridising fragments for further analysis. The signal obtained with the hybridising fragments from clones λ1.1 and λ6.2 was weaker than that obtained for clone λ5.1, despite the fact that these fragments are approximately three times the size of the fragment from clone λ5.1. This could either be because the degree of overlap between the probe and these fragments is less for these fragments, or because the probe is less homologous in these regions.

Figure 3.3.4 Analysis of the positive λ clones obtained upon library screening

Single, well isolated plaques which were obtained upon tertiary screening were used to prepare λ DNA according to section 2.2.11. The DNA was digested with *Eco*RI to excise the inserts. An aliquot of each digest (1 μg) was resolved on a 1% agarose TAE gel (section 2.2.5). The gel was alkali blotted onto a Hybond N+ filter (section 2.2.14.1). The filter was hybridised in 5 ml hybridisation buffer with the same probe and using the same conditions as described in Fig. 3.3.3. The stringency of the post-hybridisation washes was also the same as in Fig. 3.3.3.

The agarose gel is shown in A, and an autoradiograph of the filter is shown in B. One insert-derived fragment from each λ clone was shown to hybridise to the probe. The sizes of the hybridising fragments were ~900 bp for clones λ1.1 and λ 6.2, and ~320 bp for clone λ5.1.



Lane 1. *Eco*RI digest of clone λ1.1

2. " λ5.1

3. " λ6.2

4. Marker 1 (*Hin*dIII digest of λDNA)

5. Marker 2 (100 bp ladder)

The remainder of the *Eco*RI digests of the λ clones described above were resolved on a LMP agarose gel. Each of the insert fragments were excised from the gel. The DNA was purified from the gel slices by phenol extraction (section 2.2.6), and the fragments cloned into the *Eco*RI site of pBluescript II KS+, as described in section 2.2.7. The plasmid clones were numbered as follows: for example, the insert-derived fragments of clone λ1.1 were designated λ1.1a and λ1.1b (λ1.1a being the largest). Accordingly, the plasmid clones were called p1.1a and p1.1b respectively. Each plasmid clone was digested with *Eco*RI and resolved on a 1% TAE gel. To ensure that the correct fragments (i.e. the hybridising fragments) were used for subsequent analyses, the gel was blotted and probed with the same probe used in the previous experiments.

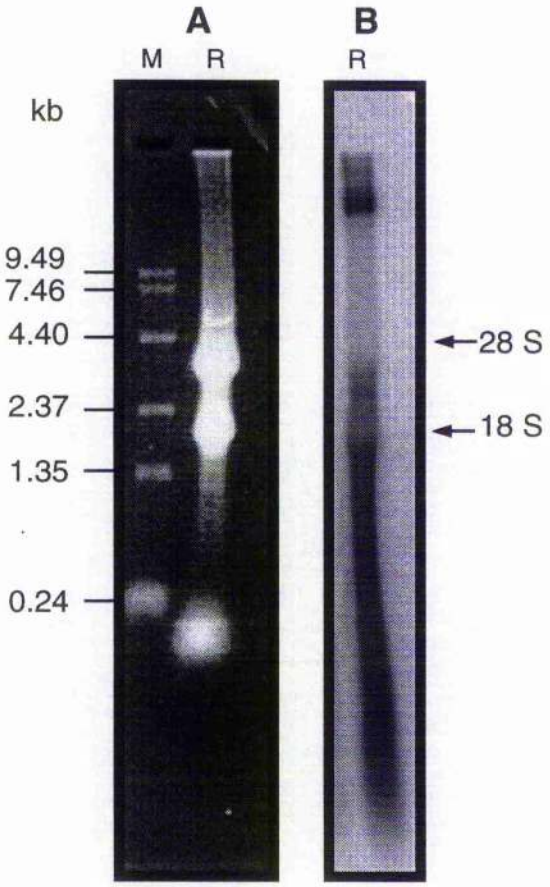
3.3.4 Northern analysis with the positive clones obtained from library screening with pGr

To test whether the three fragments which hybridised with the pGr probe derived from the same mRNA, northern blots of 25 μg total RNA (prepared from 4 day old pea shoot tips according to section 2.2.13.1 were probed at high stringency with the inserts from each of the plasmid clones (p1.1a, p5.1c and p6.2a). Only one clone (p1.1a) gave a signal; this is shown in Fig. 3.3.5. A single band of very high molecular weight (> 10 kb) hybridised. This would represent an unusually large mRNA, and it was therefore hypothesised that this fragment corresponded to an intronic sequence from an abundantly expressed gene, and that the band which hybridised represented hnRNA. Though hnRNA is unstable, the steady state levels of the unspliced version of a highly expressed mRNA could be sufficient to be detected in a northern blot, and to be represented in a cDNA library. It can also be seen in Fig. 3.5.5 that there is a smear of hybridised material extending from about 3 kb downwards. This is not due to non-specific hybridisation, since the blot was washed at high stringency, and the ribosomal RNA bands do not hybridise. Neither

does it appear to be due to general RNA degradation, as the hybridising band appears undegraded. The smear could therefore represent degradation of the intron once it has been spliced out of the mRNA.

The other clones (p5.1c and p6.2a) were next hybridised to northern blots of 5 µg poly(A)$^+$ RNA, which was prepared from 4 day old pea shoot tips according to section 2.2.13.2. As controls, samples of the *Eco*RI digested plasmid clones were diluted so that they contained 0.1 pg, 1 pg 10 pg of the insert. The samples were resolved on a 1% agarose gel, which was capillary blotted and fixed in the same manner as the northern blots (section 2.2.14.3). The filter was cut in half and the appropriate half was included in each hybridisation. After hybridisation, the filters were exposed for 5 days on a phosphoimager plate. Still, neither clone p5.1c or p6.2a gave a positive signal on the northern blot, even though the probes were able to detect 0.1 pg of the target sequence. These fragments could be derived from mRNAs which are expressed in the tissue used in construction of the cDNA library (i.e etiolated 5 day old pea shoot tips), and not in the tissue used for northern analysis (4-day old pea shoot tips grown under continuous illumination).

Figure 3.5.5 **Results of northern analysis with the subcloned λ fragments**



Samples of total RNA (25 μg) from 4 day old pea shoot tips (prepared according to section 2.2.13.1) were resolved by denaturing agarose gel electrophoresis (section 2.2.13.3), along with RNA size-markers (Gibco BRL). A portion of the gel containing the markers and one of the total RNA samples was removed and stained with ethidium bromide (A). The remainder of the gel was then blotted onto a Hybond N^+ filter (2.2.14.3), which was then cut into strips. Each strip was probed at high stringency (section 2.2.16.2) with the insert from each plasmid clone (p1.1a, p5.1c, p6.2a; section 3.3.3). Only one of these clones (clone p1.1a) gave a positive signal on the northern blot (B). The positions of the 28 S and 18 S ribosomal subunits are indicated by arrows.

Lane M RNA size-markers (Gibco BRL)

Lanes R pea total RNA (25 μg)

3.3.5 Sequence analysis of the subcloned λ fragments

The plasmid clones selected for sequencing were p1.5c and p1.1a. Although from the results of northern analysis, neither appeared to derive from a MTase mRNA, it was necessary to confirm this in the case of clone p1.1a, and since clone p5.1c gave a strong signal when probed with part of the insert of pGr, it was of interest to sequence this clone.

High quality plasmid DNA was prepared for each of the clones as described in section 2.2.9.2. Automated sequencing (section 2.2.21) was carried out using M13 Universal (–20) and Reverse primers (Fig. 2.2.2). The whole of the insert from clone p5.1c was sequenced, whereas ~300 bp from either end of clone p1.1a were sequenced. Each of the clones was found to contain stop codons in each frame. For clone p1.1a, this is consistent with the hypothesis that the fragment corresponds to an intronic sequence.

A FASTA search of both the ends of p1.1a (and their reverse complements) was carried out. However, no discernible pattern of matches could be identified with this clone. This was repeated with both strands of clone p5.1c; again, no discernible pattern of matches were obtained. A FASTA search of the SwissProt database was then carried out with the translated sequence of clone 5.1.c. The translation in frame a was used because this contained only one stop codon; whereas the other five frames contained at least five stop codons. The best match was obtained with mouse MTase, and was over a short sequence of amino acids close to the C-terminus. However, it was later discovered that this sequence in the mouse MTase cDNA actually belonged to the 3′ untranslated region, because of the incorrect assignment of the stop codon (see section 3.1). The sequences of clones p5.1c and p1.1a were then compared to the sequence of the *Arabidopsis* putative MTase genomic clone (which was used for library screening), when it became available on the database.

3.3.6 Sequence comparison of clones p1.1a and p5.1c with the putative *Arabidopsis* MTase genomic clone

The nucleotide sequence of both strands of clone 5.1a were compared with the *Arabidopsis* genomic clone, using the program BESTFIT. The best match was identified with the strand which was translated and used in the FASTA SwissProt database search described in the previous section. The region of this match corresponded to the first *Bam*HI/*Cla*I fragment of the pGr clone (probe 2; Fig. 3.3.1). This match is shown in figure 3.3.6, and covers a repeating GGTTT motif. This motif was also identified in the mouse MTase cDNA sequence (also shown in Fig. 3.3.6), and corresponds to the amino acid sequence which was previously found to match in the SwissProt database search with p5.1c. The published map of the *Arabidopsis* MTase gene (clone 11) (Scheidt *et al.*, 1994a) is very large (~12,000 kb), and shows the conserved motifs to be dispersed throughout the gene in a complicated arrangement of introns and exons (which is unusual for *Arabidopsis*). The gene itself was shown to hybridise with a mRNA of 2.2 kb, and so the putative MTase was proposed to lack an N-terminal domain. Only two of the conserved motifs were shown in the paper (motifs I and IV), and I was unable to unambiguously identify the others in the appropriate regions. Furthermore, there were several discrepancies between the published map and the map provided with the subcloned *Eco*RI fragment of clone 11 (pGr), with respect to the arrangement of the exons and the restriction sites. As mentioned in the introduction (section 1.5.3), this paper was retracted (Scheidt *et al.*, 1994b) because of falsification of sequence data for motif I by the first author. Aside from motif IV, the other conserved motifs must have been identified on the basis of very relaxed criteria. Sequence comparison between this genomic clone and the *Eco*RI fragment of the mouse MTase cDNA (see Fig. 3.2.1), which was used for genomic library screening (Scheidt *et al.*, 1994) revealed very little homology, except for within the repeating GGTTT motif (see Fig. 3.3.6). This genomic clone has probably been selected on the basis of this short

region of identity alone. This sequence is absent from the human MTase cDNA (Yen *et al.*, 1992), the *Arabidopsis* MTase cDNA (Finnegan *et al.*, 1993) and the pea MTase cDNA which was eventually cloned (section 3.7).

A 1.8 kb cDNA has been cloned which corresponds to the putative *Arabidopsis* MTase genomic clone. This cDNA corresponds to only part of the proposed gene (extending approximately from the *Eco*RI site to the first *Hin*dIII site of pGr in the 5′ to 3′ direction), and overlaps the region corresponding to probe 1 but not probe 2 (Fig. 3.3.1). There are also sequence discrepancies between the cDNA and the gene, such that sequences present in the cDNA are not present in the gene (Nebendahl and Baumlein, 1995). The function of this cDNA is unknown; it does contain the sequence ENV (of motif VI), but this hardly makes it a MTase cDNA. The predicted protein is very rich in lysine and glutamic acid residues. Like the MTase mRNA, this mRNA is expressed in young seedlings, but unlike MTase, the expression levels are much higher in the stem than in the shoot tip (A. Nebendahl, personal communication). Neither clone p5.1c or p1.1a have any significant homology with this cDNA; it has been shown that p5.1c corresponds to a region of the genomic clone that lies outside the region specifying this mRNA (probe 2; Fig. 3.3.1), whose function is unknown. Clone p1.1a does not seem to be particularly homologous in the regions sequenced to probe 2 either. However, it appears that the sequence data from the genomic clone is not to be trusted.

A FASTA search was carried out with the repeating motif of clone p5.1c and the best score was obtained with the putative *Arabidopsis* genomic clone and a high score was also obtained with the mouse MTase cDNA. High scores were also identified with the 3′ UTRs of various eukaryotic cDNAs which is not surprising since they are generally GT rich. Notably, an *Arabidopsis* cDNA clone obtained using expressed sequence tagging (EST), whose function is unknown, also contains this repeat. It is not clear whether clone p5.1c corresponds to an mRNA or not, since no band was shown to hybridise with pea poly(A)$^+$ RNA. However, it could correspond to an mRNA specific to the material used for construction of the cDNA

library (i.e. shoot tips of etiolated pea seedlings).

3.3.7 Summary

No MTase cDNA clones were obtained from library screening with fragments of pGr. This is not surprising, considering that this genomic clone almost certainly does not encode a MTase, and has probably been cloned on the basis of a repeating GGTTT motif, which is present in the 3′ UTR of the mouse MTase cDNA. This same sequence was also responsible for the selection of p5.1c described in this section. This sequence occurs in the 3′ UTRs of some plant and mammalian mRNAs, but the significance of this is unclear.

It was also shown that this sequence is capable of forming stable hybrids at relatively high stringency (hybridisation at 60°C; final wash at 65°C, 1 x SSC). This illustrates the importance of not including 3′ UTR sequences in a probe, because of the unpredictable nature of cross-hybridisation with sequences within this region.

Figure 3.3.6 **Sequence homology between p5.1c and the putative *Arabidopsis* MTase genomic clone and the mouse MTase cDNA**

The sequences were aligned using the programs, BESTFIT and GAP; homology was shown to be restricted to the repeating motif GGTTT. This was also shown to be the case when the putative *Arabidopsis* MTase genomic clone was compared with the mouse MTase cDNA. Therefore, this genomic clone was probably selected on the basis of this region of homology alone. The BESTFIT alignments are shown below.

A. Clone 5.1.c x putative *Arabidopsis* MTase genomic clone

```
  47 TCAGTTTCTCGGTTTGGTTTGGTTTGGTTTGGTTTGAACTT 87
     | | ||||  ||||||||||||:||||||||||||||   ||
9161 TGAATTTCCTGGTTTGGTTTGGTTTGGTTTGGTTTGAGTTT 9201
```

B. Clone 5.1c x mouse MTase cDNA

```
  51 TTTCTCCGTTTGGTTTGGTTTGGTTTGGTTTGAACTTTATCC 92
     ||| |  ||||||||||||||||||||||||      ||||||
4907 TTTTTTTGTTTGGTTTGGTTTGGTTTGGTTTTTTTTCTTATCC 4948
```

C. *Arabidopsis* putative MTase genomic clone x mouse MTase cDNA

```
9165 TTTCCTGGTTTGGTTTGGTTTGGTTTGGTTT 9195
     |||  : |||||||||||||||||||||||||
4907 TTTTTTTGTTTGGTTTGGTTTGGTTTCGTTT 4937
```

3.4 Screening pea DNA methyltransferases with antibodies

The aim of these experiments was to assess the potential use of polyclonal antibodies (already present in the lab) for screening a pea expression library (section 2.1.5.1) for MTase clones. The antibodies used (antibodies 156 and 157) had been raised against synthetic peptides corresponding to conserved sequences derived from a consideration of the mouse MTase cDNA clone (Bestor *et al.*, 1988). The revised sequence (section.1.5.2) had been referred to in the design of these peptides. Figure 3.4.1 shows the peptide sequences used, which correspond to the catalytic motif IV and to motif X (sections 1.4.1 and 1.4.2).

3.4.1 Antibody detection procedure

Samples of partially purified pea methylase were kindly provided by Claire Houlston. Both crude nuclear extracts (E1) and the peak fraction from the Q-Sepharose column (Yesufu *et al.*, 1991, Houlston *et al.*, 1993) were resolved by SDS PAGE (section 2.2.2.1) in duplicate. The duplicate halves of the gel were either silver stained (section 2.2.2.2) or western blotted (section 2.2.2.3) and probed by either antibody 156 or 157 according to the method described in section 2.2.2.4. As a control, samples of mouse methylase purified from ascites nuclear extract (Adams *et al.*, 1986), which was kindly provided by Roger Adams) were also run on the gels to test the effectiveness of the antibodies in reacting to the peptide sequence when it is within a protein. The results of screening with antibody 156 are shown in figure 3.4.2.

Figure 3.4.1 Sequence of the synthetic peptides to which antibodies 156 and 157 had been raised

a) The mouse MTase peptide sequences corresponding to parts of motifs IV and IX, used for raising antibodies 156 and 157 respectively.

b) The prokaryotic consensus (Posfai *et al.*, 1989), showing alternative amino acids at each position present in the 13 prokaryotic type II cytosine-C5 MTases which were known about at the time (the alternative amino acids are placed in order of frequency of occurrence where possible).

c) The amino acids which remain absolutely conserved to date with the addition of many more prokaryotic sequences to the list (Kumar *et al.*, 1994).

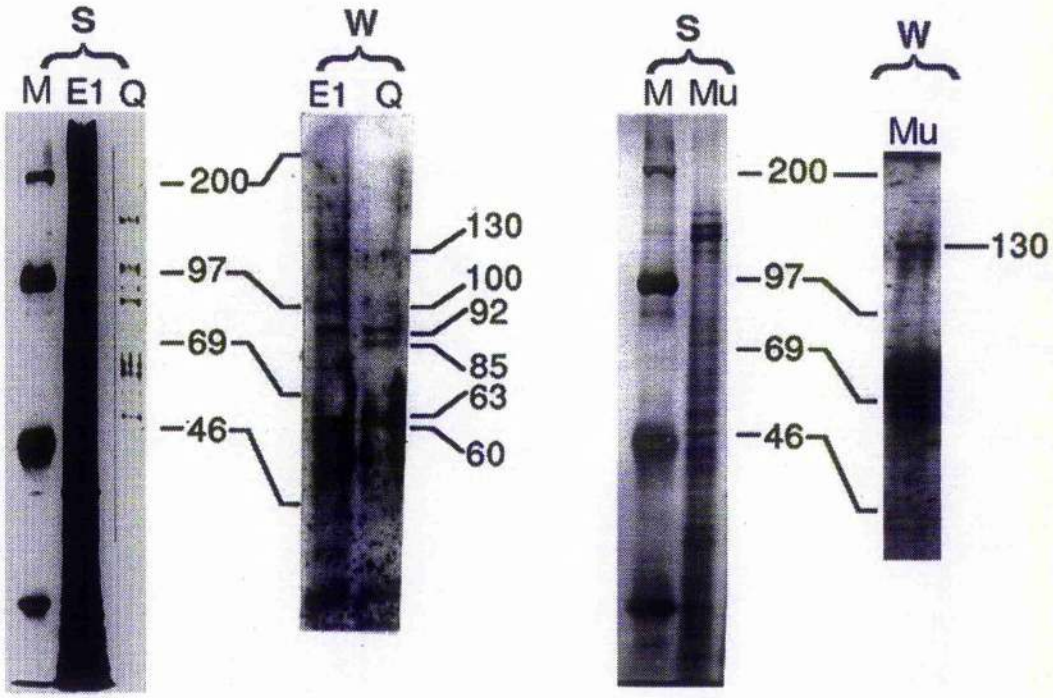
As can be seen from the diagram, there is better agreement between the mouse peptide sequence corresponding to motif IV and the original prokaryotic consensus, than there is for the peptide sequence corresponding to motif X. This still applies to the present date (data not shown).

156 peptide sequence	**157 peptide sequence**	
L C G G P P C Q G F S	D R H R Q V G N A V	*a*
L C.G G F P C Q G F S	Q Q Y K Q I G N S V	*b*
I L A . P . . P P W .	D L . R M A . . A I	
V I S . S . . . A . .	K A . Q E F	
. V . . . . . S . .	S G . . . L	
. M . . . . . T . .	. M . . . M	
. T	. K	
	. S	
. . . . . P C . . . S	. . . . . G N .	*c*

Figure 3.4.2 Antibody cross-reactivity of pea DNA methylase extracts

The cross-reaction profiles of antibody 156 against partially purified pea methylase fractions and partially purified mouse methylase, resolved by SDS PAGE. The silver-stained gels are marked S and the western blots are marked W. The molecular weights of cross-reacting bands are expressed in kDa.

Lanes M	Protein molecular weight standards
Lanes E1	Crude salt extract of pea nuclei (Yesufu *et al.*, 1991; Houlston *et al.*, 1993)
Lanes Q	E1 that has been treated with heparin Sepharose then fractionated on a Q-Sepharose column. Pooled fractions from the methylase activity peak were used (Houlston *et al.*, 1993).
Lanes MU	Partially purified mouse methylase (91.15 AS supernatant; Adams *et al.*, 1986)


S
M E1 Q
W
E1 Q
200
97
69
46
130
100
92
85
63
60
S
M Mu
W
Mu
200
97
69
46
130

3.4.2 Discussion

As can be seen in figure 3.4.2, a number of bands in both crude nuclear extracts and the Q-Sepharose fractionated activity peaks cross-reacted with antibody 156.

Antibody 156 reacted with partially purified mouse methylase. The most strongly reacting band was a 130 kDa band (the native MTase protein has a molecular weight of 190 kDa and smaller species are due to N-terminal proteolysis).

Bands of 130 kDa and 100 kDa present in E1 of pea methylase cross-react with antibodies 156. The most strongly reacting bands were a band of 92 kDa in E1 and in Q, and a band of 84 kDa in Q. These two bands appear to correlate with methylase activity along the Q-Sepharose elution profile (C. E. Houlston, MSc. thesis) but this assignment is entirely speculative due to the sheer number of bands present. Other bands of lower molecular weight cross-react with antibody 156 in both the partially purified mouse methylase sample and the E1 and Q fractions of pea methylase. Antibody 157 cross-reacted with many more bands in both partially purified mouse methylase and E1 and Q fractions of pea methylase, which indicates that this antibody is not particularly specific (results not shown).

Taken together, these results are hard to interpret. This work has been problematic because of the difficulties associated with plant DNA methyltransferase purification, the pea methylase being much more difficult to extract than the mouse enzyme. Though high specific activities are obtained, the amount of pea methylase protein relative to other contaminating proteins is much less in comparison to the mouse methylase extracts. In addition, the pea methylase (like other plant methyltransferases) is much more susceptible to N-terminal proteolysis during extraction than the mammalian MTases, and so the methylase protein cannot be assigned unambigously to any particular band.

Both antibodies reacted with the 130 kDa band of partially purified mouse MTase. However, the strength of the signal was not particularly strong, relative to that of smaller contaminating bands in this fraction, which indicates that the

antibodies have neither a high titre nor are particularly specific (this is especially the case with antibody 157).

It was decided on the basis of these results not to go ahead with the expression library screening. This was firstly because of the non-specific cross-reactivities of the antibodies which would obviously be problematic in library screening. Secondly, the signal obtained for the mouse methylase was not particularly high, relative to the amount of protein present. Because high titre antibodies are recommended for use with the library by the manufacturers, antibodies 156 and 157 were not considered suitable.

The subsequent cloning of the 3′ end of a pea MTase cDNA (section 3.7), has shown that the predicted amino acid sequence of pea MTase differs by two amino acids from each of the synthetic peptide sequences corresponding to mouse motifs IV and VI. The effect this may have had on the reactivity of the antibodies 156 and 157 to pea MTase proteins is unknown.

3.5 Screening a pea cDNA library for MTase clones using degenerate oligonucleotides

The aim of these experiments was to isolate MTase specific clones from the pea cDNA library (in λgt11, section 2.1.5.1) by virtue of their hybridisation to two separate pools of degenerate oligonucleotides designed according to conserved amino acid motif IV of mouse MTase (section 1.5.2).

3.5.1 Oligonucleotide design

There are two different approaches to library screening using degenerate oligonucleotides, as described by Sambrook *et al.*, 1989. The first is to use pools of short oligonucleotides whose sequences are highly degenerate, and the second is to use pools of longer oligonucleotides of lesser degeneracy (or "guessmers"). The former was the method of choice because less amino acid sequence data is required and also because all possible combinations of codon use can be represented. When longer oligonucleotides are used (> 20 nucleotides), the degeneracy has to be kept to a minimum because the effect of mismatches on the melting temperature becomes progressively less significant, and thus each member of the degenerate pool has a greater range of sequences it can hybridise with. It is possible to minimise degeneracy through consultation of codon preference tables; however this does not guarantee that the correct codons have been chosen.

At the time this work was carried out, the only eukaryotic MTase cDNA that had been cloned was that of mouse (Bestor *et al.*, 1988). We had also by then access to the revised amino acid sequence (Bestor, personal communication), which contained most of the conserved amino acid motifs (section 1.5.2). The MTase motif selected for oligonucleotide design was motif IV, since this is the catalytic motif and, hopefully, the best conserved between mammals and plants. In designing the oligonucleotides, it was not possible to allow alternative amino acids present in

prokaryotic MTases, as this would have made the degeneracy too high and would result in an unacceptably high number of false positives. The design of the oligonucleotides (oligos 1015 and 1017) and the corresponding amino acid sequence is shown in figure 3.5.1. Oligo 1017 is a 17-mer of degeneracy 256 which was intended for preliminary screening of the library. The reasoning behind oligonucleotide design was as follows:

The complexity (C) of a typical mammalian cDNA library is ~1 x 10^7 bp and the probability (P) of a sequence of length L occurring by chance is given by the equation:

$$P = (x)^L \times 2C$$

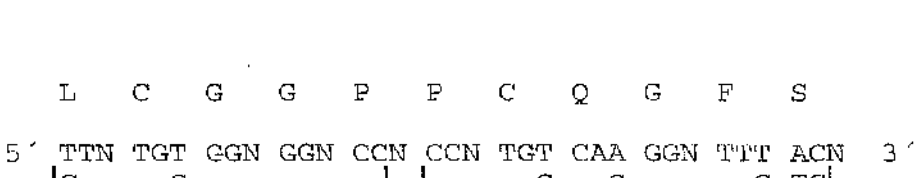
where x is the chance of a particular nucleotide occurring at any position.

If the distribution of bases is random, the value of x is 0.25 and therefore the probability of a particular 17-mer occurring by chance in a mammalian cDNA library is ~1/900. However, the distribution of bases in coding sequences is non-random and empirical testing reveals that x is closer to 0.283 (Lathe, 1985). This is because of the redundancy of the genetic code that gives a better than 46% chance of correct selection of the third position in the codon, without consideration of codon utilisation tables (this figure takes into account the variability at all positions for Leu, Ser and Arg codons). This brings the value of P up to 1/104 for a 17-mer.

The 17-mer designed for library screening (oligo 1017) has a degeneracy of 256, which brings the value of P up to 2.4 for a mammalian cDNA library. It was assumed that the complexity of the pea cDNA library (made from 5 day old pea shoot tips) would be roughly similar. For example, a human fibroblast cell contains 12,000 different mRNAs (Sambrook *et al.*, 1989), and the number of different mRNAs in tobacco cells has been quoted by Slightom and Quemada (1991) to be 25,000, about 6000 of which are unique to each organ (leaf, stem, root etc.). Therefore, in theory, only five fortuitous matches were expected at the most. However, the number of positive clones obtained upon screening with oligo 1017 cannot be predicted statistically; if there happens to be a chance match with a well represented cDNA

clone, the number of positives will be much greater. It was therefore intended to re-screen positives (obtained upon screening with oligo 1017) with a second degenerate oligonucleotide (oligo 1015). This oligo is shorter (a 14-mer) and has a degeneracy of 256, and so the P value works out to be ~100. Thus oligo 1015 is only suitable for screening a subset of clones identified by screening with oligo 1017.

Figure 3.5.1 Design of degenerate oligonucleotide pools for library screening



The figure shows the predicted amino acid sequence of motif IV of mouse MTase (Bestor *et al.*, 1988), after revision of sequence data (Bestor, personal communication). Underneath is the nucleotide sequence coding for this motif, showing all positions of degeneracy. The regions corresponding to degenerate oligonucleotides 1015 and 1017 are indicated below.

3.5.2 Library screening with oligo 1017

Because individual members of a pool of degenerate oligonucleotides will vary widely in their (G+C) content, so will their melting temperatures. For this reason, hybridisation was carried out in the presence of the quaternary alkylammonium salt, tetramethylammonium chloride (TMACl), instead of sodium chloride. (Jacobs *et al.*, 1988). In the presence of TMACl, the T_m of a hybrid is independent of base composition, and is dependent primarily on its length. The use of TMACl therefore allows hybridisation to be carried out at temperatures which discriminate against mismatched hybrids. The recommended hybridisation temperature (at ~5°C below the T_i (or irreversible melting temperature) for a 17-mer is 48–50°C in 3 M TMACl (Sambrook *et al.*, 1989). Hybridisation using degenerate oligonucleotides also requires high concentrations of probe (0.1–1.0 pmol/ml) and long hybridisation times (proportional to the degeneracy of the probe). This is because the effective concentration of the probe is inversely proportional to its degeneracy.

The degree of representation of MTase clones in the λgt11 library is difficult to estimate. The library was constructed from poly(A)$^+$ mRNA from etiolated 5-day old pea shoot apices. 5 day old shoot apices show peak MTase specific activity (Yesufu *et al.*, 1991). However, the levels of presumptive MTase protein are low, which is presumably reflected in the steady-state levels of MTase mRNA. In addition, the library is amplified, which will lead to even less representation of rare mRNAs. However, as a preliminary experiment, it was decided to screen ~200,000 clones with oligo 1017.

The library was plated out onto 7 x 150 mm plates (section 2.2.10), at a density of 30,000 plaques per plate. Plaque lifts were performed in duplicate as described in section 2.2.14.2. Oligo 1017 was end-labelled to high specific activity (~10^9 dpm/μg), as described in section 2.2.15.2. The filters were hybridised at 49°C in 25 ml hybridisation buffer in a sealed plastic bag. The hybridisation buffer contained 3.0 M TMACl and 20 pmol end-labelled oligo 1017; the other components

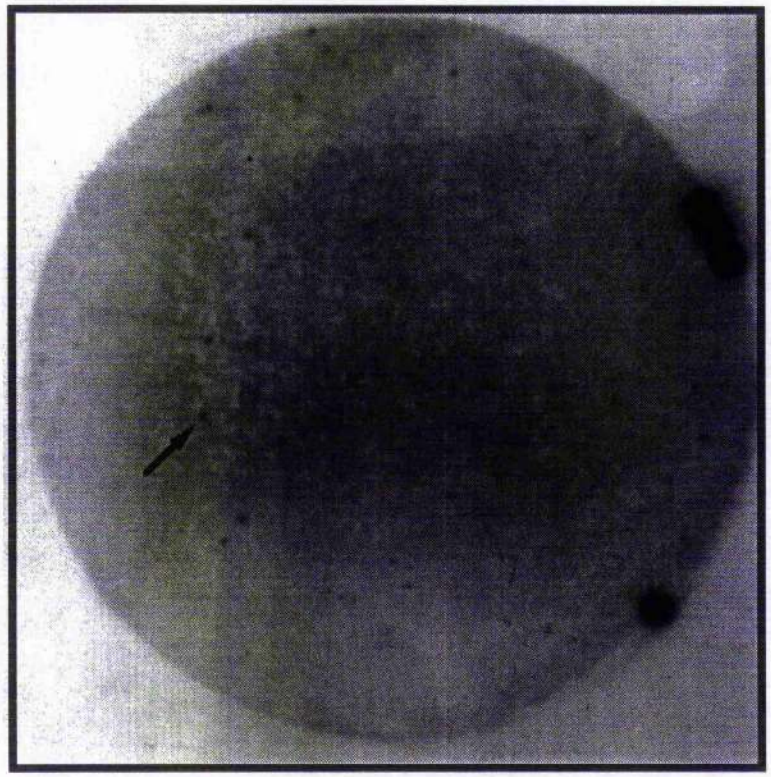
of the prehybridisation and hybridisation buffers are listed in section 2.2.16.2. Because of the degeneracy of oligo 1017, hybridisation was carried out for 6 days. Post hybridisation washes were carried out as described in section 2.2.16.2. The final wash was carried out in a solution containing 3.0 M TMACl at 49°C.

3.5.3 Results of screening with oligo 1017

An autoradiograph of one of the filters is shown in figure 3.5.2. The faint positives obtained also hybridised on the duplicate filter. The positives are faint because the signal is expected to be 1/256 (or less) the signal obtained with a unique oligo. There is also a low level of non-specific hybridisation.

Library screening with oligo 1017 yielded many more positives than anticipated and an average of ~50 clones per 150 mm plate (i.e. 30,000 clones) hybridised. This is probably because of a chance match between one or more highly represented cDNA clones. In retrospect, it would have been better to have used a longer degenerate oligonucleotide as the concomitant increase in degeneracy is offset by the increase in length. Oligos of up to 19 nucleotides in length show good discrimination between perfectly matched and mismatched hybrids (Sambrook *et al.*, 1989).

Figure 3.5.2 **Results of screening the pea cDNA library with oligo 1017**



The figure shows an autoradiograph of one of the filters screened with the degenerate oligonucleotide, 1017 (Fig. 3.5.1). 200,000 clones from the pea cDNA library in λgt11 were plated out at a density of 30,000 per 150 mm plate (section 2.2.10). Plaque lifts were performed using Hybond N^+ filters as described in section 2.2.14.1, and the library screened with end-labelled oligo 1017 (Fig. 3.5.1) in hybridisation buffer containing 3.0 M TMACl (section 2.2.16.2)

Approximately 50 positive clones per plate were identified, one of which is indicated with an arrow. The large spots on the edge of the filter are radioactive ink.

3.5.4 Cross-screening of the library with oligo 1015

The anticipated number of clones expected upon hybridisation with oligo 1015 were much greater than with oligo 1017. However, the filters used in the previous experiment were stripped by washing with 0.5% SDS at 55°C for 1 h, autoradiographed to check whether all the probe had been removed, and then screened with oligo 1015. Hybridisation was carried out using the same hybridisation protocol as described for oligo 1017 but with the following modifications: 5 x SSPE (section 2.1.3) was used instead of 3.0 M TMACl, because the reduction in the range of melting temperatures when TMACl is used instead of sodium salts is much less for 14-mers (Jacobs *et al.*, 1988). Hybridisation was carried out at 40°C; the temperature of hybridisation was selected to be 2°C less than the calculated T_m of the most AT rich species, as recommended by Sambrook *et al.* (1989); the calculation $T_m = 4(G+C) + 2(A+T)$ °C holds well for oligos less than 18 nucleotides in length. The final post-hybridisation wash was carried out at 40°C in hybridisation buffer without the blocking agents for 5 min. However, high levels of background hybridisation were obtained, which made detection of individual positive clones very difficult (results not shown). It was anticipated from the outset of these experiments that empirical testing of appropriate hybridisation temperatures was necessary for oligo 1015. This is possible when only a small subset of clones are to be screened, but is not really feasible with the large numbers of positive clones which hybridised to oligo 1017.

3.5.5 Discussion

At the time, this seemed a perfectly reasonable approach to take to obtain cDNA clones of pea MTase; oligo 1017 was the minimum length recommended for library screening by Lathe (1985). Indeed, Mason and Williams.(1985) recommend the use of 17-mers and salt-containing hybridisation solutions for library screening. Because

hybridisation temperatures in the presence of sodium ions have to be appropriate for the most AT rich oligo species, this allows a considerable proportion of shorter hybrids to form and effectively reduces the probe length (Wood *et al.*, 1985). As discussed earlier, decreasing the probe length increases the number of false positives.

The number of positives obtained with the 17-mer described in this section was much greater than expected. In retrospect, it would have been better to increase the probe length by synthesizing partially overlapping oligonucleotides corresponding to motif IV, or to use additional conserved motifs for oligonucleotide design.

With the benefit of hindsight, it is clear that this approach would have met with little success, even if the probe length had been increased. First of all, the actual predicted amino acid sequence of the pea MTase cDNA that was eventually cloned deviates from the mouse MTase motif IV in the region corresponding to oligo 1015 (see Fig. 3.7.8). Ironically, the amino acid sequences of pea and mouse MTases are identical in the region of motif IV (sequence GMNRFN) that shows much less homology with the prokaryotic consensus than does the region corresponding to oligo 1015. Secondly, it was shown later that the region corresponding to motif IV is, for some reason, very much under-represented in the library (<1 in 3 million; section 3.6.13). Therefore even if a longer degenerate oligonucleotide corresponding to motif IV had been used, the chances of success would have still been very small.

The technique of PCR amplification using degenerate primers offers several advantages over the approach described above. Firstly, a greater degree of degeneracy is allowed as two separate degenerate oligonucleotides are annealed in the same reaction which greatly increases the effective specificity. This in turn gives a better guarantee of success. Secondly, the technique is much less time consuming in the identification of true positives. This approach is described in the next section.

3.6 RT-PCR of methylase-specific DNA fragments from pea RNA using degenerate primers

The aim of these experiments was to specifically amplify a segment of pea methylase cDNA using primers designed taking into account the conserved amino acid motifs (sections 1.4 and 1.5). This is a very powerful technique for cloning new members of gene classes, one or more of which have already been characterised. The experiment was designed using the recommendations of Compton (1990). First strand cDNA was used in preference to genomic DNA, because the pea genome is large and therefore it is possible that there are introns between the conserved motifs.

3.6.1 Degenerate primer design

At the start of these experiments, the only eukaryotic MTases to have been cloned were mammalian (i.e. murine (Bestor *et al.*, 1988) and human (Yen *et al.*, 1992)). These two MTases are, not surprisingly, very closely related. No examples of cloned plant MTases existed to refer to for primer design. It was considered highly likely that plant MTases would be more closely related to the mammalian enzymes than to any of the prokaryotic type II cytosine-C5 MTases. Nevertheless, the motifs exhibiting the greatest sequence conservation at the amino acid level between the prokaryotic MTases (Posfai *et al.*, 1989) were selected for primer design. The other criteria were that highly conserved amino acids had to be uninterrupted as far as possible by less conserved amino acids, and that the expected product should not be too large (smaller products amplify more efficiently during PCR).

Of the conserved regions, the best candidates were motif IV (containing the catalytic Pro-Cys) and motif VI. Figure 3.6.1 shows the amino acid sequence of the regions of motifs IV and VI used for primer design. Initially, two primers for each motif were designed, and the design of each was based on the amino acid sequence of the mammalian motifs, with some allowance for the variability in the amino acid

sequence seen in the prokaryotic type II cytosine-C5 MTases. Oligos IVa and IVb were sense primers designed according to motif IV, and oligos VIa and VIb were antisense primers designed according to motif VI. The 'b' set of primers were more degenerate than the 'a' set.

The oligos were synthesized to contain the 'neutral' base inosine at positions of absolute degeneracy (after Knoth *et al.*, 1988). Inosine forms stable base pairs with all four bases, and is the best neutral base available for positions of four-fold degeneracy (Martin *et al.*, 1985). The use of inosines at positions of degeneracy has the advantage that there are fewer different species within the degenerate pool, so that less of the primer needs to be added to the PCR reaction. N.B. the degrees of degeneracy illustrated in Fig. 3.6.1 are not the "true" degeneracy, and are calculated as if I = N. The degeneracy figures are meant as a guide to the likelihood of chance annealing to non-MTase sequences.

The primers were also synthesized to contain an *Eco*RI site plus a GG clamp (Scharf, 1990) at their 5′ ends in order to facilitate cloning of PCR products. The presence of the 5′ *Eco*RI site has the added advantage of increasing the stability of the priming duplex after the first round of amplification, which increases the overall efficiency of the reaction (Mack and Sninsky, 1988).

Each oligo was checked by denaturing gel electrophoresis on a 20% polyacrylamide gel containing 6 M urea to ensure the population contained only full length species.

Figure 3.6.1 **The design of degenerate oligonucleotides based on conserved motifs IV and VI**

a) The figure shows regions of motifs IV and VI selected for primer design. Underneath the mammalian sequence (murine and human) is the consensus of 13 prokaryotic type II cytosine-C5 MTases (Posfai *et al.*, 1989). The number to the right of each amino acid indicates the degree of representation amongst the 13 prokaryotic MTases.

b) All degenerate oligonucleotides synthesized were 22-mers, had inosine at positions of complete degeneracy, and included *Eco*RI sites at their 5′ ends. The 'degeneracy' of each oligonucleotide is indicated on the right hand side, and is calculated as if I = N.

Oligo IVa was synthesized according to the mammalian sequence, whereas oligo IVb was made to allow for possible sequence deviations in pea MTase and was completely degenerate at the second codon (it was not possible to allow for the presence of proline because that would have made the degeneracy too high).

Oligo VIa was synthesized according to the mammalian sequence, except that both codons for lysine and only one codon for arginine (the one used in mouse motif IV) were allowed at the 4th position (from the right hand side). This was because lysine is the most well represented amino acid at that position amongst the prokaryotic enzymes. Oligo VIb was synthesized to be completely degenerate at this position to allow for the presence of any amino acid. Neither of the oligos allowed for the possible presence of glycine at the last codon position due to degeneracy constraints.

Oligo VIc was synthesized to allow for all possible arginine and lysine codons. Also included were the codons for asparagine and threonine at the last codon position. This was because the recently cloned *Arabidopsis* MTase was found (surprisingly) to code for a threonine instead of asparagine at this position (Jean Finnegan, personal communication).

a) Conserved motifs

	Motif IV				**Motif VI**					
G	P	P	C	Q	E	N	V	R	N	mammalian
G13	F8 P3 S2	P13	C13	Q10 P3	E13	N13	V13	K10 R1 L1 P1	G10 N3	prokaryotic

b) Primer sequences **'Degeneracy'**

IVa (oligo 1334) (sense)

	*Eco*RI	**G**	**P**	**P**	**C**	**Q**		
5´	GGGAATTC	GGI	CCI	CCI	TGT C	CA	3´	64

IVb (oligo 1333) (sense)

	*Eco*RI	**G**	**X**	**P**	**C**	**Q**		
5´	GGGAATTC	GGI	III	CCI	TGT C	CA	3´	2048

VIa (oligo 1332) (antisense)

	E	**N**	**V**	**K(R)**	**N**	*Eco*RI		
3´	CTT C	TTA G	CAI	TTT CC	TT	CTTAAGGG	5´	64

VIb (oligo 1331) (antisense)

	E	**N**	**V**	**X**	**N**	*Eco*RI		
3´	CTT C	TTA G	CAI	III	TT	CTTAAGGG	5´	1024

VIc (oligo 1394) (antisense)

	E	**N**	**V**	**(R/K)**	**(N/T)**	*Eco*RI		
3´	CTT C	TTA G	CAI	TTI GC	TT G	CTTAAGGG	5´	512

3.6.2 Reverse transcription of pea RNA

Total RNA was prepared from 4–5 day old pea seedlings using the Kirby method (section 2.2.13.1). The RNA was checked for possible degradation prior to reverse transcription by denaturing gel electrophoresis (section 2.2.13.3). Reverse transcription was carried out according to section 2.2.18.1, using MuLV reverse transcriptase.

In initial experiments, first strand synthesis was primed using an oligo dT primer, which was removed (by gel filtration on a chromaspin 100 column) prior to PCR. The rationale behind the use of oligo dT-primed cDNA synthesis was to eliminate the possibility of amplifying ribosomal RNA cDNA fragments. This is a particular problem when highly degenerate primers are used (Compton, 1990). In any case, the use of oligo dT primed first strand cDNA should not affect the efficiency of amplifying methylase-specific fragments providing the reverse transcription step is efficient.

3.6.3 PCR amplification of first strand cDNA

The PCR cycling parameters (section 2.2.13.2) were based on the recommendations of Compton (1990). The cycling is begun by performing 1–5 cycles with a low, non-stringent annealing temperature of 37°C. After this stage, the PCR products containing the 5′ extensions (in this case *Eco*RI sites) serve as targets for the second stage of amplification, which can thus be carried out with a higher annealing temperature. The number of cycles recommended for the second stage was 25–40 cycles. PCR was carried out using various combinations of the degenerate primers (Fig 3.6.2 *a*). A number of controls were included for the PCR step, and were as follows:

1. RNA (5 μg) to check for possible amplification of traces of genomic DNA.

2. Negative control (no target), to eliminate possibility of carry-over.

3. PCR reactions with only one primer included, to eliminate non MTase-specific PCR products.

3.6.4 Results of PCR amplification of oligo dT-primed first strand cDNA

The initial set of experiments involved the use of primers IVa, IVb, VIa and VIb. The expected size of authentic pea MTase PCR product was calculated as follows: The exact size of the PCR product which would be obtained from mouse cDNA is 159 bp (including the *Eco*RI 5′ ends); this corresponds to 48 amino acids. The analogous segment in the prokaryotic enzymes (Posfai *et al.*, 1989) is typically 46–48 amino acids in length. Two species have segments of 56 amino acids and one (Phi3T) has a segment of 80 amino acids. Therefore the size of the authentic PCR product from pea cDNA should be ~160 bp if Phi3T MTase is ignored.

Experiment 1 used three cycles with an annealing step at 37°C, followed by 30 cycles with an annealing step of 46°C. The annealing temperature used for the second stage of amplification was lower than recommended by Compton (1990). The reasoning behind this was as follows: the reversible melting temperature (T_m) of oligonucleotide hybrids is estimated using the rule of thumb calculation (Thein and Wallace, 1986) where $T_m = 4(G+C) + 2(A+T)$ °C. The effect of inosine is hard to predict, as the order of stability of I-N pairs is I-C > I-A > I-T ≈ G-T, with I-C pairs being slightly less stable than A-T pairs. Also, the stability of base-pairing depends on local sequence contexts (Martin *et al.*, 1985). Oligos VIa and VIb are more AT rich than IVa and IVb, so the annealing temperatures used were based on a consideration of the T_ms of oligos VIa and VIb. A conservative estimate for the T_m of the most AT rich species, counting inosines as zero, is 30°C and 24°C for oligos VIa and VIb respectively, not including the *Eco*RI sites. The inclusion of the *Eco*RI sites in the calculation brings the T_m of oligo VIa and VIb up to 40°C and 46°C

respectively. However, it has been reported that up to 20% mismatch between degenerate primers and their templates is tolerated in the PCR reaction (Lee and Caskey, 1990), and so it was hoped that the primers would anneal at temperatures higher than the calculated T_m. For this reason, a high annealing temperature (relative to the calculated T_m) was used in the first three rounds of amplification, which should maximise the specificity of the PCR. In subsequent rounds, a less stringent annealing temperature (equal to the calculated T_m of VIa) was used, which should maximise the overall efficiency of amplification. Repeat PCRs (using the second stage amplification step only) were carried out on 1 µl (out of 100 µl) samples of the first PCR reaction. Figure 3.6.2 shows the results of experiment 1. However, even after repeat PCR, no product of the expected size was obtained.

It was thought that perhaps the initial annealing temperature may be too stringent for some members of VIa and VIb. This parameter was therefore reduced to 32°C in experiment 2, whilst keeping the other parameters the same as in Experiment 1. Again repeat PCR was carried out on 1 µl samples of the initial reaction. The results of experiment 2 are shown in figure 3.6.3.

Figure 3.6.2 PCR amplification of pea first-strand cDNA (experiment 1)

3.6.2 A

Cycling parameters:

Stage 1 [94°C, 3 min] x 1; [94°C 1 min, 37°C 1 min, 72°C 1 min] x 3

Stage 2 [94°C 1 min, 46°C 1 min, 72°C 1 min] x 30; [72°C 2 min] x 1

Targets:

Lanes 1–8: Oligo dT-primed first-strand cDNA (section 2.2.18.1)

Lane 9: RNase-free DNase I treated pea total RNA (2 µg)

Lane 10: No target

Primer combinations for both 3.6.2 A and 3.6.2 B:

Lane:	1	2	3	4	5
	IVa+VIa	IVa+VIb	IVb+VIa	IVb+VIb	IVa
	6	7	8	9	10
	IVb	VIa	VIb	IVa +VIa	IVa +VIa

Each PCR sample was ethanol precipitated and resuspended in 10 µl TE buffer. Half of each sample was then resolved on 2% agarose TAE gels along with 0.5 µg 100 base-pair ladder marker (lanes M).

3.6.2 B (repeat PCR)

Cycling parameters:

[94°C 3 min] x 1;[94°C 1 min, 46°C 1 min, 72°C 1 min] x 30; [72°C 2 min] x 1

Targets:

1 µl (out of 100 µl) of the corresponding PCR reactions from 3.6.2 A

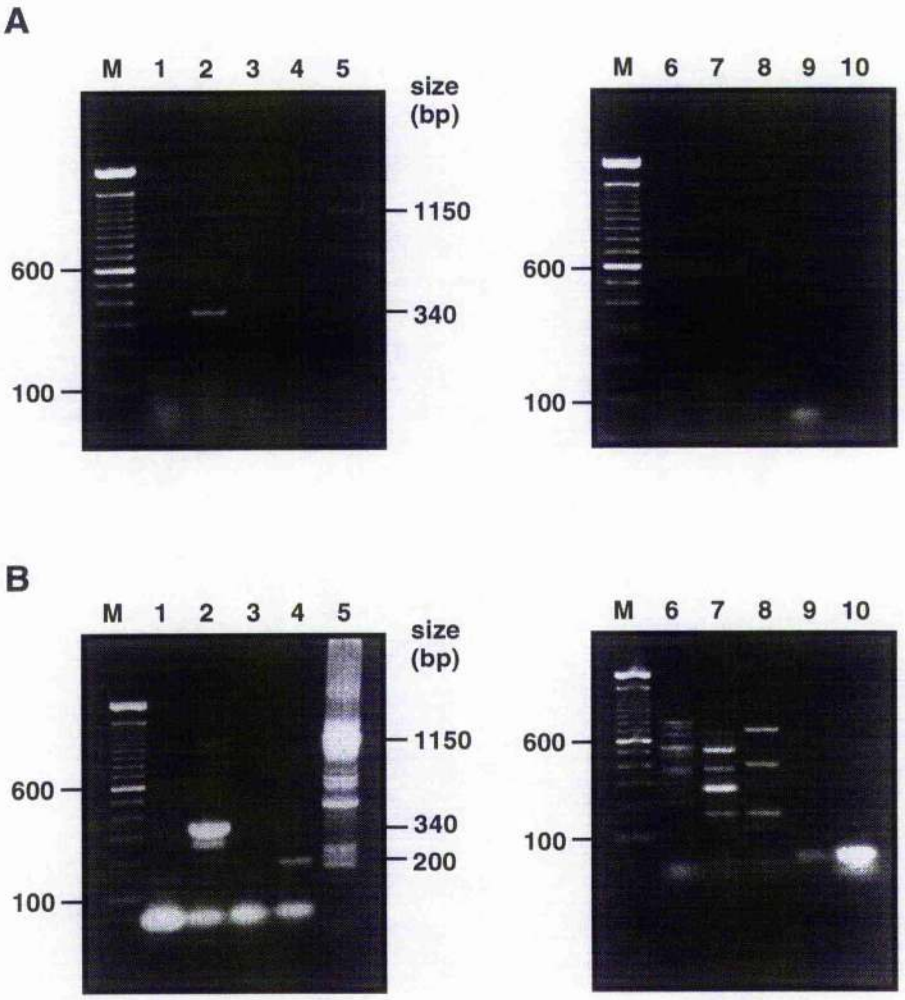

A
M 1 2 3 4 5
size (bp)
1150
600
340
100
M 6 7 8 9 10
600
100
B
M 1 2 3 4 5
size (bp)
1150
600
340
200
100
M 6 7 8 9 10
600
100

Figure 3.6.3 PCR amplification of pea first strand cDNA (experiment 2)

3.6.3 A

Cycling parameters:

Stage 1: [94°C 3 min] x 1;[94°C 1 min, 32°C 1 min, 72°C 1 min] x 3

Stage 2: [94°C 1 min, 46°C 1 min, 72°C 1 min] x 30; [72°C 2 min] x 1

Targets: Same as in Fig. 3.6.2 A

3.6.3 B (repeat PCR)

Cycling parameters:

[94°C 3 min] x 1;[94°C 1 min, 46°C 1 min, 72°C 1 min] x 30; [72°C 1 min] x 1

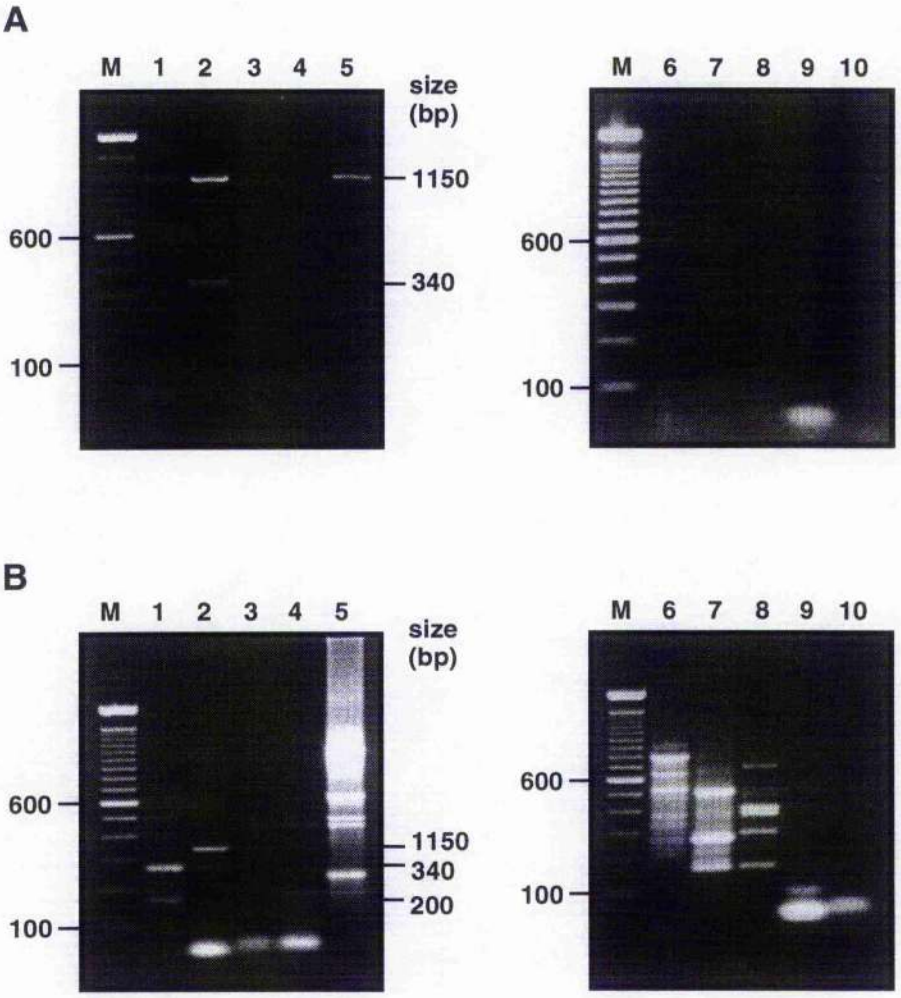
Targets:

1 µl (out of 100 µl) of the corresponding reactions from 3.6.3 A

Primer combinations:

Same as in figure 3.6.2 for each lane

Again, each PCR reaction was ethanol precipitated and resuspended in 10 µl TE buffer. Half of each sample was electrophoresed on a 2% agarose TAE gel, along with marker DNA (100 bp ladder; lanes M).


A
M 1 2 3 4 5
size (bp)
1150
340
600
100
M 6 7 8 9 10
600
100
B
M 1 2 3 4 5
size (bp)
600
100
1150
340
200
M 6 7 8 9 10
600
100

As can be seen in figures 3.6.2 and 3.6.3, there appears to be no contaminating genomic DNA in the total RNA preparation because there was no amplification in either the first or repeat PCR steps. The main PCR products obtained in the first round of PCR are a band of ~340 bp and one of ~1.15 kb. The 340 bp band is too large to be a MTase-specific PCR product. This band was amplified by IVa in combination with either VIa and VIb, even when the initial annealing temperature is 37°C. The fact that this band is amplified by both VIa and VIb suggests that the annealing temperatures are not too high for the inosine-rich oligo VIb to work. However, this band was not amplified with primer combinations that included oligo IVb, even when the initial annealing temperature is reduced to 30°C. This is surprising, considering that even though oligo IVb contains one more inosine than oligo VIb, it is relatively GC rich and this should raise the T_m of the hybrids. Perhaps the failure of IVb to amplify the 340 bp band is a function of its high effective degeneracy (see Fig. 3.6.1). Similarly, IVb does not amplify the 1.15 kb band which is amplified by IVa.

The band of 1.15 kb appears in all lanes where oligo IVa was used. There are two possible explanations for the amplification of this fragment. The first is that it is a IVa specific band; the second is that the fragment is a MTase 3´ RACE product amplified by IVa and residual amounts of the oligo dT primer, not removed by the chromaspin column (theoretically, even a 1% residue of the oligo dT primer (~0.05 pmol per PCR reaction)) would be capable of generating a 1.15 kb band in the range of 30 ng). However, the first explanation appears more likely because in the second stage of PCR using 1 µl out of 100 µl amplified product as target, highly amplified material appears in the region of ~1.15 kb when oligo IVa alone is used as a primer, and the amounts of residual oligo dT would be negligible.

In experiment 2, a similar profile of fragments to that obtained from experiment 1 was obtained in the first PCR step. However, the profile of bands obtained in the repeat PCR step is significantly different. In particular, oligos IVa and VIa had amplified fragments of estimated size 160, 270 and 520 bp. The first

matches exactly the size of the product expected from mouse MTase cDNA. A worrying feature of this result however, was that the main bands obtained from amplification by VIa alone match this profile closely. Nevertheless, a repeat PCR amplification was carried out on the gel-purified 160 bp band using oligos IVa and VIa. Disappointingly, only a smear was visible after agarose gel electrophoresis.

It was decided next to test whether the degenerate primers could amplify the correct segment from a positive control target. This was not done concurrently with the previous experiments as a precaution against contamination of the test reactions by the positive control. The results of these experiments are summarised in the following section.

3.6.5 Amplification of a plasmid target using the degenerate primers

The positive control used in these experiments was the mouse MTase cDNA cloned into pBluescript (pMG), (see section 3.2). The amount of target used in each case was 2 ng plasmid, and the parameters used were the same as in experiment 1. Amplification of the expected fragment (160 bp in length) from the control target was achieved using all combinations of IV and VI primer pairs, and the intensity of the band was approximately equal in each case (results not shown). The yield of authentic product was low, however, but was improved when the number of cycles in the second step was increased to 35. The intensity of the 160 bp band amplified from plasmid targets was always low in comparison to the intensities of smaller products (< 100 bp). These smaller products were either due to chance annealing to regions other than the intended target within the plasmid template or were primer-dimer artefacts. Nevertheless, these experiments did show that all the degenerate oligonucleotides were capable of amplifying the authentic product.

3.6.6 PCR amplification of oligo VI-primed first strand cDNA

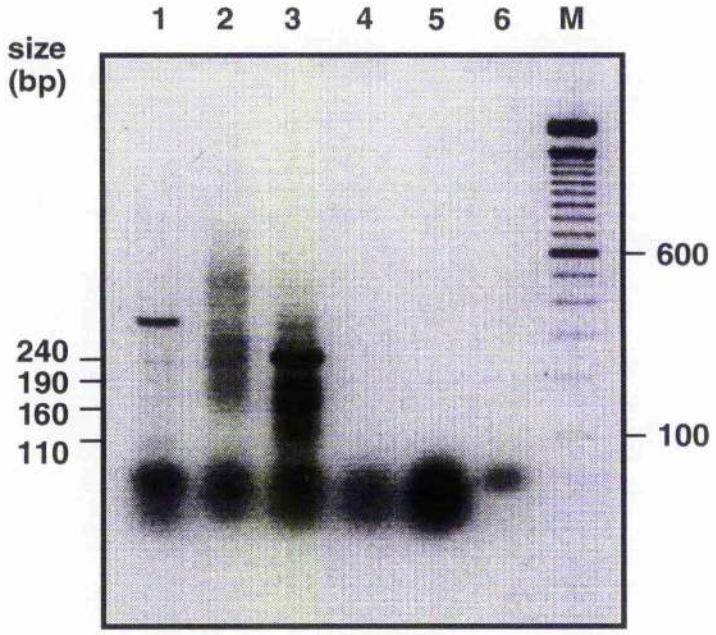
On the basis of previous experiments, it was concluded that there was nothing basically wrong with the PCR conditions and parameters used for pea cDNA amplification. The fact that a cDNA fragment of over 1 kb has been amplified (figures 3.6.2 and 3.6.3) suggests that first-strand synthesis should have been efficient enough to generate a long enough MTase cDNA to contain the primer annealing sites. Another factor which may be affecting the yield of authentic product could be the presence of secondary structure in MTase mRNA, which would prematurely terminate first-strand synthesis. Though MuLV reverse transcriptase was recommended in the literature, it is more sensitive to secondary structure than AMV reverse transcriptase, presumably because reverse transcription is carried out at a higher temperature (42°C) with the latter enzyme. To circumvent the problem of secondary structure, a downstream degenerate primer was used to prime reverse transcription. Though there was a possibility of priming from, and subsequent amplification of ribosomal RNA, it was hoped that sufficient amounts of authentic product would be amplified. By the time these experiments were undertaken, I had access to the predicted amino acid sequence of motifs IV and VI of the cloned *Arabidopsis* MTase cDNA (Jean Finnegan, personal communication). *Arabidopsis* motif IV was identical to mouse motif IV in amino acid sequence. *Arabidopsis* motif VI, however, differed from mouse motif VI by one amino acid (the asparagine residue was replaced by a threonine residue). Oligo VIc was therefore designed to allow for either asparagine or threonine to be present in motif VI of pea MTase, (see Fig. 3.6.1). Oligos IVa and VIc were used in the following experiment.

Reverse transcription was carried out as described in section 2.2.18.1, using oligo VIc (50 pmol) instead of the oligo dT primer to prime synthesis. PCR was carried out using the same parameters as in experiment 1 (Fig. 3.6.2) except that 35 cycles were used in the second stage of PCR, and a repeat PCR reaction was not

carried out. As a control, oligo dT-primed cDNA was also used as target. The results of this experiment are shown in figure 3.6.4.

As can be seen from the figure (shown in negative for clarity), four fragments of estimated molecular sizes 240, 190, 160 and 110 bp are amplified strongly from oligo VIc-primed first strand cDNA (the expected size of MTase-specific product is 160 bp) The fragments amplified from oligo dT-primed pea cDNA (with the oligo dT primer removed prior to PCR amplification) are similar to those obtained in previous experiments. (A control in which the oligo-dT primer was not removed was also included to test whether oligo dT removal by chromaspin 100 and recovery of target really was efficient.) Together the results show that it is extremely likely that secondary structure of RNA was the cause of the failure to amplify a pea MTase-specific fragment.

Figure 3.6.4 **Amplification of oligo VIc-primed first strand cDNA**



The primers used for each PCR reaction were oligos IVa and VIc. The PCR parameters were as described in section 3.6.6. The reactions were then ethanol precipitated and resuspended in 10 µl TE buffer. Half of each sample was resolved on a 2% agarose TAE gel along with 0.5 µg marker (100 base-pair ladder). The gel photograph is shown in negative.

Targets:

Lane 1	Oligo dT-primed first strand cDNA in which the oligo dT had been removed by a chromaspin 100 column
Lane 2	Oligo dT-primed first-strand cDNA, chromaspin step omitted
Lane 3	Oligo VIc-primed first strand cDNA
Lane 4	Non-DNase I-treated pea total RNA (5 µg)
Lane 5	DNase I-treated pea total RNA (5 µg)
Lane 6	Negative control – no target

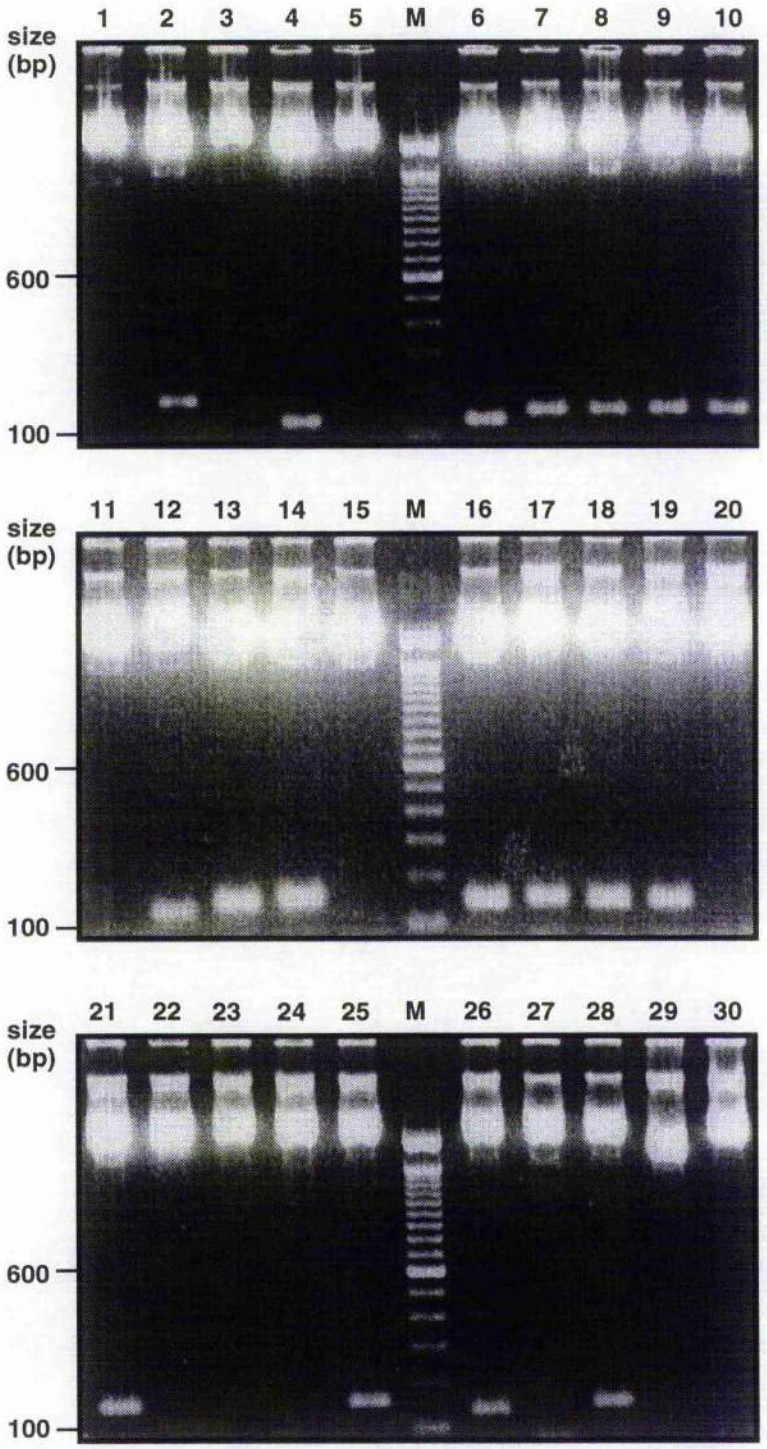
3.6.7 Cloning of PCR fragments

Firstly, a repeat of the PCR experiment was carried out in duplicate, together with a control in which only oligo VIc was added. The samples were resolved on a 2% LMP agarose gel. The profile of bands obtained with IVa and VIc amplification was exactly the same as in the previous experiment. The profile of bands obtained in the control lane was different, but because of the number of bands it was not clear which of the IVa + VIc amplified fragments could be due to amplification by VIc alone (results not shown). The bands of 160 and 190 bp were excised from the gel and the DNA recovered by phenol extraction (section 2.2.6).

The DNA samples were then digested with *Eco*RI, phenol/chloroform extracted, ethanol precipitated and resuspended in sterile water. The DNA fragments were then cloned into the *Eco*RI site of pBluescript II KS+. Ligations were carried out as described in section 2.2.7 with 2 ng insert DNA (as estimated against DNA concentration standards section 2.2.3.3). 5 µl of each ligation mix was used to transform competent *E.coli* XL1 Blue cells (section 2.2.8). Pale blue colonies were selected for overnight culture (the insertion of small fragments of DNA can lead to incomplete disruption of the Lac Z gene). Plasmid minipreps were made from each overnight culture (section 2.2.9.1). *Eco*RI digested samples of the minipreps are shown in figure 3.6.5. For some reason, fewer transformants were obtained from the ligation of the 190 bp fragment. Only one recombinant plasmid contained an insert approximating to this size (lane 2). The majority of inserts appeared to be the same size (~150 bp) which would result from the cleavage of a 160 bp PCR fragment by *Eco*RI. In addition, some of the recombinants contained smaller inserts (~130 bp). For future reference, the recombinant plasmids are numbered as they are in figure 3.6.5.

Figure 3.6.5 ***Eco*RI digests of recombinant plasmid minipreps containing cloned PCR products**

5 µg of each plasmid was digested with *Eco*RI and resolved on a 2% agarose gel along with marker DNA (100 bp ladder). Lanes 1–5 are transformants obtained from the ligated 190 bp fragment whereas lanes 6–29 are transformants obtained from the ligated 160 bp fragment. Lane 30 is a control (from a blue colony).


1 2 3 4 5 M 6 7 8 9 10
size (bp)
600
100
11 12 13 14 15 M 16 17 18 19 20
size (bp)
600
100
21 22 23 24 25 M 26 27 28 29 30
size (bp)
600
100

3.6.8 Sequencing of recombinant plasmids

Each recombinant plasmid described in the previous section was sequenced on the ABI automated sequencer (section 2.2.21). Double stranded plasmid templates for sequencing were prepared from the overnight cultures using the Promega Magic Miniprep columns (section 2.2.9.2). Some of the recombinant plasmids were sequenced in both directions (using the M13 universal and reverse sequencing primers (Pharmacia). Due to space (and financial) constraints, most were sequenced using the reverse primer only.

The major species of cloned PCR product (clones 7, 8, 9, 10, 13, 16, 17, 18, 19, 21, 28) were identified as pea MTase-specific fragments upon sequence analysis and translation. Figure 3.6.6 shows an alignment of the sequences with each other. Only clone 7 was sequenced in both directions, and the sequences were found to match exactly (results not shown). The clones that did not belong to this group were clones 6 and 12 (which were identical to each other), and clones 2, 4 ,14, 25 and 26, which were unique. These non-MTase clones will be discussed later.

3.6.9 Sequence analysis of cloned pea MTase-specific PCR products

As can be seen in figure 3.6.6, all the PCR products were ligated in the antisense orientation in pBluescript II KS (see also Fig. 2.2.2), although the orientation should theoretically be random. This could be because of toxicity to the host cells of the β-galactosidase fusion protein following induction by IPTG. Alternatively, perhaps insertion in the sense orientation leads to recombination in the host; indeed all the cloned PCR products (whether they belong to this group or not) have been cloned in the same orientation.

Figure 3.6.6 Sequence alignment of cloned pea MTase specific PCR fragments

The figure shows the reverse complement of clones 7–28 which are MTase-specific PCR fragments cloned into the *Eco*RI site of pBluescript II KS+. Automated sequencing was performed on double-stranded plasmid templates using M13 reverse sequencing primer. As can be seen, all inserts have been cloned in the same (antisense) orientation.

Underneath the alignment, the sequences of degenerate primers oligo VIc and IVa.(omitting *Eco*RI sites) are depicted, showing the alternative bases (oligo VIc is shown in antisense).

Also shown is the sequence of part of the pea DNA MTase 3′ RACE product which was cloned later (see section 3.7). Where the sequences of the PCR products differ from that of the cloned 3′ RACE product is indicated in bold type.

clone

```
          1                                                  50
  7       GGCCCGCCGT GCCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
  8       GGCCCGCCGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
  9       GCNCCGCCGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
 10       GGNCCGCCGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
 13       GGNCCGCCGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
 16       GGNNCCCCGT GCCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
 17       GCCCCGCCGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
 18       GGCCCGCCGT GTCAGGGTAT CTCTGGGATG AATAGATTTA ACACAAGCAC
 19       GGNCCGCTGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC
 21       GGNCCGCCGT GTCAGGGTTT CTCTGCGATG AATAGATTTA ACACAAGCAC
 28       GGNCCGCCGT GTCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC

RACE      GGGCCTCCAT GCCAGGGTTT CTCTGGGATG AATAGATTTA ACACAAGCAC

Oligo IVa GGICCICCIT GCCA
                     T

          51                                                 100
  7       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
  8       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
  9       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 10       TTGGAGTGAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 13       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 16       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 17       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 18       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCNTA TCCTTTGCTG
 19       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 21       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTTA TCCTTTGCTG
 28       TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTAA TCCTTTGCTG

RACE      TTGGAGTAAA GTCCAGTGTG AGATGATATT AGCGTTCTAA TCCTTTGCTG


          101                                          143
  7       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AGAACGTCAG AAA
  8       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AGAACGTCAG AAA
  9       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AGAACGTCAA GAA
 10       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AAAACGTCAA GAA
 13       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AGAATGTCAA AAA
 16       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AGAACGTCAA AAA
 17       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AAAACGTCAA GAA
 18       ATTATTNCCG NCCGAGNTAC TTCCTNCTGG AGAACGTCAG AAA
 19       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AGAACGTCAG GAA
 21       ATTATTTCCG GCCGAGGTAT TTCCTCTTGG AAAACGTCAA AAA
 28       ATTATTTCGG GCCGAGGTAT TTCCTCTTGG AGAACGTCAA AAA

RACE      ATTATTTCGG GCCGAGGTAT TTCCTCTTGG AGAATGTGAG GAA

Oligo VIc (antisense of)                  G AGAACGTIAG IAA
                                            A T   CA  C
```

Comparison of the sequences with each other and with the authentic sequence (which was later obtained by 3′ RACE using *Pfu* polymerase; section 3.7) shows that there were few *Taq* misincorporations (excluding the primer sequences). The exception was clone 18 which appeared to contain several point mutations, but this may be a feature of poor sequencing in this instance. Apart from clone 18, only clone 10 had any point mutations (just one at position 58). Analysis of the sequences corresponding to the degenerate primers showed significant deviation at positions of degeneracy from the authentic sequence. When the reverse primer is used for sequencing, the sequence of oligo VIc itself, and the sequence of the strand opposite to oligo IVa is read. Analysis of the sequence corresponding to oligo IVa shows a marked preference for G on the coding strand, at positions where where there are inosines in oligo IVa. This shows that there has been a bias towards the incorporation of C residues opposite I residues during PCR with the degenerate primers, especially towards the 3′ end of oligo IVa. This is probably because I-C pairs are the most stable (Martin *et al.*,1985). There also appears to be a tolerance of mismatches at positions where two alternative nucleotides are possible in the primer sequences. For example, with oligo VIc up to 3 mismatches are tolerated (= 14% of a 22-mer). As mentioned earlier, the studies of Lee and Caskey (1990) have shown a tolerance of up to 20% mismatch, though the type of mismatch was not specified. All of the sequence deviations at positions of two-fold degeneracy in this experiment have arisen from G-T mismatches between the primers and the template cDNA. G-T pairs are relatively stable, and are in fact somewhat more stable than G-I pairs (Sambrook *et al.*, 1989).

3.6.10 Comparison of MTase PCR products with sequences in the database

The reverse complement of clone 7 (figure 3.6.6) was translated in three forward frames, revealing the first frame to be an ORF as expected. A FASTA search of the

SwissProt database was then carried out with the predicted amino acid sequence. The highest score was obtained with *Arabidopsis* MTase (Finnegan and Dennis, 1993), followed by human MTase (Yen *et al.*, 1992) and mouse MTase (Bestor *et al.*, 1988). The alignments are shown in figure 3.6.7 together with the two most closely matching prokaryotic MTases (M.*Sin*I and M.*Aqu*I). The best alignment (not surprisingly) is to *Arabidopsis* MTase. The two plant MTases are slightly more divergent from each other than the two mammalian MTases (whose sequences are identical in this region). The eukaryotic MTases appear to form a closely related group which are distinct from the prokaryotic MTases.

3.6.11 Sequence analysis of the other cloned PCR products

The other cloned PCR fragments (referred to in section 3.6.8) all derived from amplification by oligos IVa and VIc. Because oligo VIc was used to prime first-strand synthesis, the sense strand reads in the IVa to VIc direction. It was first necessary to investigate the possibility that any of these PCR products have been amplified from a different pea MTase mRNA (since there are two different plant MTases with CG and CNG substrate specificities (section 1.5.3). For this to be a possibility, the ORF of a particular PCR product must begin as GPPC and end as ENV(R/K) in the same frame. In addition, the absolutely conserved serine residue from motif IV must be present i.e. PCXXXS (section 1.4). As mentioned earlier, each of these PCR products had been cloned in the same orientation as the MTase-specific PCR products. Therefore, the reverse complement of each of these cloned PCR products was translated in three forward frames. Three of the clones (14, 25 and 26) did contain a serine at the appropriate position; however both clones 14 and 25 required a frame shift to read ENV(R/K). Clone 26 did have the sequence ENVK in frame with GPPCQ, and therefore a FASTA search was carried out with the predicted amino acid sequence (from frame a). The translations in three frames of each of clones 14, 25 and 26 are shown in figure 3.6.8, together with the best

Figure 3.6.7 **Predicted amino acid sequence of cloned pea MTase-specific PCR product; homology to other MTases**

The sequence of the cloned PCR product (clone 7, Fig. 3.6.6) was translated in the first reading frame. A FASTA search was then carried out on the SwissProt database with the predicted amino acid sequence, and the four best matches are shown in the figure.

The best matches are obtained with the eukaryotic MTases (*Arabidopsis* MTase and the mouse and human MTases). Of the mammalian sequences, only the match with mouse MTase is shown, as it is identical to the human sequence in this region. The alignments with the two most closely related prokaryotic MTases between regions IV and VI, M.*Sin*I and M.*Aqu*I are also shown. Also indicated is the sequence specificity of each MTase; the target cytosine is underlined.

***Arabidopsis* MTase** Target specificity: CG or CNG

```
SCORES          Init1:   264 Initn:   264 Opt:   264
          95.7% identity in 47 aa overlap

               10        20        30        40
PCR7.p GPPCQGFSGMNRFNTSTWSKVQCEMILAFLSFADYFRPRYFLLENVR
       ||||||||||||||| |:|||||||||||||||||||||||||||||
mtdm_a GPPCQGFSGMNRFNQSSWSKVQCEMILAFLSFADYFRPRYFLLENVR
         1200      1210      1220      1230      1240
```

Mouse MTase Target specificity: CG

```
SCORES          Init1:   202 Initn:   202 Opt:   202
          68.1% identity in 47 aa overlap

               10        20        30        40
PCR7.p GPPCQGFSGMNRFNTSTWSKVQCEMILAFLSFADYFRPRYFLLENVR
       ||||||||||||||::|:|| : :::::||.:.||:| |:|||||||
mtdm_m GPPCQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVR
                1120      1130      1140      1150
```

M.*Sin*I Target specificity: GGWCC

```
SCORES          Init1:    46 Initn:    75 Opt:    79
          34.0% identity in 47 aa overlap

               10        20        30        40
PCR7.p GPPCQGFSGMNRFNTSTWSKVQCEMILAFLSFADYFRPRYFLLENVR
       |||.|:||: ::    : :: : :::: :|: |  :||:|:::|||
mtsi_s GPPCQAFSTAGK--RLGLEDERGNVFIKYLDVALDIRPKYIVIENVR
              160       170       180       190
```

M.*Aqu*I (α subunit) Target specificity: CYCGRG

```
SCORES          Init1:    45 Initn:    65 Opt:    68
          31.9% identity in 47 aa overlap

               10        20        30        40
PCR7.p GPPCQGFSGMNRFNTSTWSKVQCEMILAFLSFADYFRPRYFLLENVR
       ||.|||:.|  ::    : :: :  ::|:||: :    |::|::|||:
mtaa_s GPPCQSFSLAGK--RMGMDDPRGMLVLEFLRVVREALPKCFVMENVK
       80        90        100       110       120
```

sequence match obtained with a MTase from the FASTA search with clone 26 (translated in the first frame). It can be seen from the figure that though the FASTA search did identify a match between clone 26 and a MTase sequence, the match is restricted to the region of motif IV, which is by definition dictated by the degenerate primer sequence. This match was therefore not considered to be significant.

The possibility that the frame shifts in clones 14 and 25 were due to sequencing errors was also considered. FASTA searches were therefore carried out with the sequences translated in all three frames. Again, no significant alignments with MTase sequences were identified (data not shown).

Only one type of MTase-specific product has been amplified by oligos IVa and VIc has been identified. This does not of course completely exclude the possibility that another MTase mRNA exists in pea; it is possible that the sequence of an alternative MTase mRNA could be different enough to prevent amplification by these primer pairs.

Figure 3.6.8 Sequence analysis of other cloned PCR products

a) Each of the cloned PCR products (Fig. 3.6.5) that did not belong to the MTase group (Fig. 3.6.6) were analysed to see if they represented alternative MTase PCR products. The reverse complement of each sequence was translated in three forward frames. Those which coded for the sequence PCXXXS in the first frame (clones 14, 25 and 26) are shown below. Clones 14 and 25 contain stop codons in the first frame and are therefore not MTase-specific segments.

b) A FASTA search with the amino acid sequence predicted by frame a of clone 26. The best match obtained with a MTase is shown, but was not considered significant.

a) Translation of clones 14, 25 and 26

Clone 14

```
          ggncencegtgtcaccgaagtaccaacagaagaaacaaccgaaaaaccaacagaaacaac
        1 ---------+---------+---------+---------+---------+---------+ 60
a         G P P C H R S T N R R N N R K T N R N N  -
b          ? ? R V T E V P T E E T T E K P T E T T -
c           ? ? V S P K Y Q Q K K Q P K N Q Q K Q L -

          tgcagcaggagaagctgtagatgctcccgaaacccaagatccacttgaagttgagaacaa
       61 ---------+---------+---------+---------+---------+---------+ 120
a         C S R R S C R C S R N P R S T * S * E Q  -
b          A A G E A V D A P E T Q D P L E V E N K -
c           Q Q E K L * M L P K P K I H L K L R T R -

          ggaagtggtaacagagaacgtcaaaaa
      121 ---------+---------+------- 147
a         G S G N R E R Q K   -
b          E V V T E N V K    -
c           K W * Q R T S K   -
```

Clone 25

```
          ggncegccgtgccatgcatgcagctatcgatagacaaaaaaatcatggtatgcactttcg
        1 ---------+---------+---------+---------+---------+---------+ 60
a         G P P C H A C S Y R * T K K S W Y A L S  -
b          ? R R A M H A V I D R Q K N H G M H F R -
c           ? A V P C M Q L S I D K K I M V C T F V -

          tgtattagctaaagccttacgtttgtctggtggagatcatattcacgctggtactgtagt
       61 ---------+---------+---------+---------+---------+---------+ 120
a         C I S * S L T F V W W R S Y S R W Y C S  -
b          V L A K A L R L S G G D H I H A G T V V -
c           Y * L K P Y V C L V E I I P T L V L * * -

          aggtaaacttgaaggcgaacgtcaagaa
      121 ---------+---------+-------- 148
a         R * T * R R T S R   -
b          G K L E G E R Q E  -
c           V N L K E N V K   -
```

Clone 26

```
        gggccgccgtgtcaaacgtcatccgtcatccacattgcaatgactctggacgctaccnac
      1 ---------+---------+--------  .  --------+---------+---------+ 60
a       G  P  P  C  Q  T  S  S  V  I  H  I  A  M  T  L  D  A  T  Y  --
b        G  R  R  V  K  R  H  P  S  S  T  L  Q  *  L  W  T  L  P  T  -
c         A  A  V  S  N  V  I  R  H  P  H  C  N  D  S  G  R  Y  L  P -

        ctccgtggctccattgccggtgttttctctgtcttgcaacacgcttcgtgtccagagaac
     61 ---------+---------+---------+---------+---------+---------+ 120
a       L  R  G  S  I  A  G  V  F  S  V  L  Q  H  A  S  C  P  E  N  -
b        S  V  A  P  L  P  V  F  S  L  S  C  N  T  L  R  V  Q  R  T -
c         P  W  L  H  C  R  C  F  L  C  L  A  T  R  F  V  S  R  E  R -

        gtcaaaaa
    121 -------- 128
a       V  K     -
b        S  K    --
c         Q  K   -
```

b) FASTA search with clone 26 amino acid sequence (frame a)

Clone 26 x M.*Hgi*dI

```
SCORES            Init1:     33 Initn:     45 Opt:     33
          100.0% identity in 5 aa overlap

                                                    10        20        30
PCR26.                                          GPPCQTSSVIHIAMTLDATYLRGSTAGVFS
                                                |||||:
mtd1_h FTHPIHELDLAQIDAAVSLIKTHSPELIIGGPPCQDFSSAGKRDEGLGRANLTLDFAKIV
               50        60        70        80        90        100

                40
PCR26. VLQHASCPENVK

mtd1_h LAIQPAWVIMENVERARLSKIHQQACSMLGDEGYSLAQVVLDASLCGVPQLRKRTFVIGH
              110       120       130       140       150       160
```

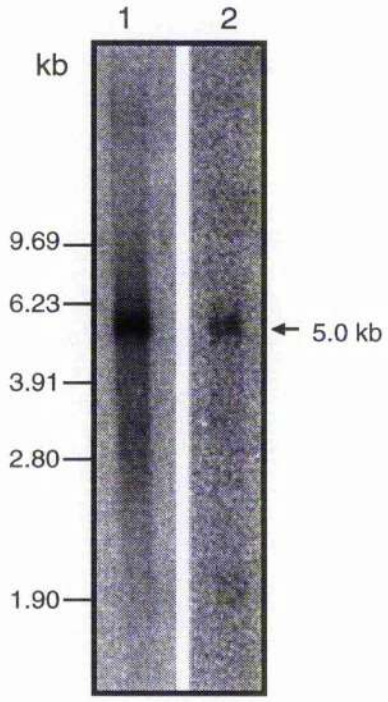
3.6.12 Hybridisation of a cloned pea MTase-specific PCR product to pea RNA

The insert from clone 7 (Fig. 3.6.6) was radiolabelled by random priming and hybridised to a northern of poly(A)$^+$ RNA (prepared from 4–5 day old pea shoots). Hybridisation and washes were carried out at high stringency (2.2.16.1), but the last wash was omitted because the probe length is less than 150 bp.

For comparison, a duplicate northern blot was probed at 65°C with a fragment of a pea cDNA clone obtained independently by S. Pradhan, working in our laboratory. (This cDNA clone was isolated partly on the basis of N-terminal sequencing of purified pea CWG MTase). The final, high stringency wash was also omitted for this blot, to see if there was a family of related MTase mRNAs were present.

Figure 3.6.9 shows the results of hybridisation. A single band of 5.0 kb was found to hybridise with each probe. The size of the mouse MTase mRNA is 5.2 kb (Bestor *et al.*, 1988), and that of the *Arabidopsis* MTase mRNA is 4.7 kb (Finnegan and Dennis, 1993). The results suggest (but do not prove) that the MTase-specific PCR product and the cDNA clone isolated by S. Pradhan originate from the same mRNA. The results also suggest that there is only one type of MTase mRNA (unless there is another MTase mRNA that is different enough not to have hybridised under these conditions).

Figure 3.6.9 **Hybridisation of a cloned pea MTase-specific PCR product to a pea northern**



Poly(A)+ RNA was prepared from 4–5 day old pea shoot tips according to section (2.2.13.2), and 4 µg samples were resolved by electrophoresis on a denaturing agarose gel (section 2.2.13.3), along with RNA size-markers (Promega). The gel was blotted onto a Hybond N+ membrane (section 2.2.14.3), and the membrane was then cut into strips. One strip (strip 2) was probed with the insert from the cloned MTase-specific PCR product (clone 7; Fig. 3.6.6), and the other with a fragment of a pea MTase cDNA clone isolated by S. Pradhan.

Both probes were radiolabelled by random priming (section 2.2.15.1) and hybridised at 65°C (section 2.2.16.1). Post-hybridisation washes were as recommended by the manufacturers for homologous probes, except that the final high stringency wash was omitted. A single band of 5.0 kb was found to hybridise in each case.

3.6.13 Screening a pea cDNA library with the cloned MTase-specific PCR product

At this stage, we did not know whether the cloned MTase-specific PCR product and the MTase cDNA clone isolated by S. Pradhan (mentioned in the previous section) corresponded to the same MTase mRNA, because we did not have overlapping sequence data. To obtain a larger cDNA fragment the pea cDNA library in λgt11 (section 2.1.5.1) was probed with the insert of clone 7 (Fig.3.6.6). The library was plated out (section 2.2.10) on 12 plates at a density of 75,000 plaques/150 mm plate (= 10^6 plaques). Duplicate plaque lifts were performed as described in section 2.2.14.2. The insert from clone 7 was radiolabelled by random priming (section 2.2.15.1) to a specific activity of >1 x 10^9 dpm/μg. Hybridisation was carried out at 65°C in 50 ml hybridisation solution (section 2.2.16.1) with 50 ng probe. Post hybridisation washes were carried out as follows: 500 ml 2 x SSC, 0.1% SDS at room temp (2 x 30 min); 250 ml 1 x SSC, 0.1% SDS, at 65°C (2 x 15 min). The final, most stringent wash was omitted because of the shortness of the probe. However, upon autoradiography of the filters (for 1 week), no positive signals were obtained on duplicate filters. The library screening was repeated on 2 x 10^6 plaques. As a control, samples of clone 7 plasmid were dot blotted onto a Hybond N^+ membrane at concentrations of 1 ng, 100 pg and 10 pg homologous target (i.e. insert).The filter was fixed in the same manner as the plaque lifts and hybridised and washed together with the plaque lift filters. The limit of detection was found to be 10 pg target DNA after 1 week's exposure. This sensitivity should be enough to detect homologous λ recombinants on the plaque lift filters. However, no positive plaques were identified. It was therefore concluded that the sequence corresponding to regions IV–VI of pea MTase must be very poorly represented in this library. This conclusion was also reached independently by S. Pradhan. The reason for this under-representation is unclear. The results from this section suggest that there may be some secondary structure in the MTase mRNA, which would inhibit oligo dT-

primed synthesis. However, oligo dT- and random hexamer-primed first strand cDNA synthesis was used in the construction of this library. Another factor could be toxicity of this particular region to the host through low level basal expression, as it has been shown in this section that the PCR product spanning motifs IV and VI has been cloned in the antisense orientation only (though induction with IPTG was used in this instance). However if this has occurred, it would only affect cDNA fragments cloned in the sense orientation and in frame with the β-galactosidase gene of the vector. Perhaps there has been an initial under-representation due to a combination of factors, which has been exaggerated through library amplification.

3.6.14 Summary

PCR amplification from pea cDNA using the degenerate primers IVa and VIc yielded only one type of MTase-specific PCR product. This hybridises to a mRNA of 5.0 kb. An mRNA of the same size was also found to hybridise with the pea MTase cDNA clone isolated by Sriharsa Pradhan. This cDNA clone was isolated from the same pea cDNA library (in λgt11) by screening with a PCR product amplified by degenerate primers based on N-terminal protein sequence data (from the CWG MTase) (Pradhan and Adams, 1995). This cDNA clone was incomplete and did not extend into region IV (efforts to find overlapping cDNA clones extending beyond region IV in this library had failed). As both this clone and the cloned pea MTase-specific PCR product hybridised to the same-sized mRNA, it seemed likely that both originated from the same mRNA. To prove this however, (and to clone the 3′ end of the cDNA) a 3′ RACE strategy was adopted. Sequence data close to the 3′ end of the cDNA clone (kindly provided by Sriharsa Pradhan) was used to design a unique primer for 3′ RACE. The resultant 3′ RACE product should overlap with my PCR product and sequencing would reveal whether the two clones derive from the same MTase mRNA.

Most importantly, a 3′ RACE strategy is the quickest, most effective way of

identifying alternative splicing events in the 3´ end of the mRNA. This is a particularly pertinent question to answer because of the fact that there are both CG and CWG MTase activities in plants, and that the two substrate specificities reside in different proteins (Pradhan and Adams, 1995). It could be hypothesised that an alternative splicing event in the 3´ end of the pea MTase pre mRNA could give rise to two mRNAs, which differ in the region specifying the variable domain . The variable domain determines target specificity, at least in the prokaryotic MTases (section 1.4.1). This difference in size may not be large enough to discern on a northern blot.

The 3´ RACE experiments will be described in the next section.

3.7 Cloning of the 3′ end of a pea MTase cDNA using a RACE strategy

A unique primer (oligo 1838) was synthesized to correspond to part of the sense strand of the pea MTase cDNA clone isolated by Sriharsa Pradhan (section 3.6.13). The sequence of the primer and the predicted amino acid sequence are shown in figure 3.7.1. This particular sequence was selected as it contained a natural *Pst*I site, which would facilitate subsequent cloning of the RACE product. The location of this sequence corresponds to 40 or so amino acids downstream of motif II; as discussed earlier, none of the clones isolated from the library extended as far as motif IV.

3.7.1 Protocol for 3′ RACE

As it was suspected that there was inefficient oligo dT-primed reverse transcription using MuLV reverse transcriptase (3.6.6), it was thought best not to use this enzyme for the 3′ RACE experiment. A 3′ RACE kit was purchased from Gibco BRL (Cat. No. 8373SA). This kit utilises SuperScript reverse transcriptase (a cloned, genetically modified AMV RTase). This enzyme has been engineered to eliminate the RNase H activity (found in other reverse transcriptases, especially AMV RTase) which degrades the mRNA during the reverse transcription. This enzyme therefore results in greater full-length cDNA synthesis and a higher yield of first-strand cDNA. The higher temperature optimum for SuperScript may minimise problems associated with RNA secondary structure.

a) **Gene-specific primer (GSP, = oligo 1838) (sense)**

5´ **AACTGCAG AATTGGCCTCTAAGC** 3´

*Pst*I

Predicted amino acid sequence = AELASK as read from the 5th base (G)

b) **Adapter Primer (AP) (antisense)**

*Mlu*I *Spe*I

5´ **GGC CAC GCG TCG ACT AGT AC(T)$_{17}$** 3´

*Not*I half-site *Sal*I

c) **Universal Amplification Primer (UAP) (antisense)**

*Mlu*I *Spe*I

5´ **CUA CUA CUA CUA GGC CAC GCG TCG ACT AGT AC** 3´

UDG cloning region *Not*I half-site *Sal*I

3.7.1.1 3′ RACE using *Taq* polymerase

Reverse transcription was carried out as detailed in section 2.2.19.1, initially as recommended by the manufacturers of the 3′ RACE kit (i.e. using 10 pmol Adapter Primer at 42°C). PCR amplification was carried out (as described in section 2.2.19.2) using the gene-specific primer and either the Adapter primer or the Universal Amplification Primer (Fig. 3.7.1). The PCR parameters (using a "Hot Start") were based on those recommended by Frohman (1994) and were as follows:

1. The reactions (minus the *Taq* and mineral oil) were heated to 94°C for 5 min. This step is to denature the RNA/DNA hybrids in case the RNase H treatment was not effective. The samples were then cooled to 75°C for 5 min, during which time the *Taq* and mineral oil (which was pre-heated to 75°C) were added.

2. The primers were annealed at 60°C for 2 min followed by a long extension at 72°C (40 min). This is to ensure the target molecules are full length for subsequent amplification. A stringent annealing temperature was used to maximise specificity at this initial stage (the rule of thumb calculation of the reversible T_m for the gene-specific primer, ignoring the two 5′ adenosines, is 64°C). It was not certain how raising the annealing temperature above 60°C would affect the annealing of the Adapter Primer (which contains a stretch of 17 dTs at its 3′ end). However, the annealing temperature could be raised in subsequent experiments, should there be no specific amplification.

3. 30 cycles of the the following temperatures were carried out: [94°C for 1 min, 60°C for 1 min, 72°C for 3 min]. The long extension time should ensure that full-length species (~1.3 kb) can compete effectively with truncated products for amplification.

4. A final extension at 72°C for 15 min (to ensure all PCR products are "flushed").

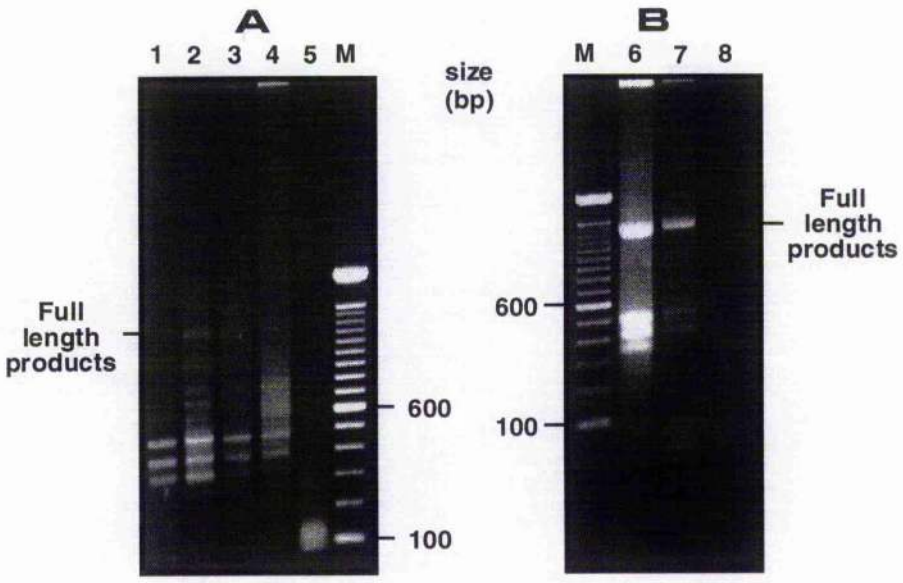
The results of amplification are shown in figure 3.7.2 A.

Figure 3.7.2 3′ RACE amplification of pea first strand cDNA using *Taq* polymerase

The figure shows the results of PCR amplification of reverse first strand cDNA as described in section 3.7.1.1 using various amounts of target and different primer combinations.

Figure 3.7.2 A shows the result of 3′ RACE using first strand cDNA which was reverse transcribed from pea poly(A)$^+$ RNA as suggested by the 3′ RACE kit instruction manual (2.2.19.1). Figure 3.7.2 B shows the result of amplification of cDNA synthesized using the modified protocol described in 3.7.1.1. Half of each PCR reaction was resolved on a 1% agarose TAE gel along with marker (100 bp ladder, lanes M).

The full length products (a doublet of ~1400 and 1320 bp) are indicated on figure 3.7.2 A. The resolution is not sufficient on figure 3.7.2 B to show the presence of a doublet (which was visible by eye). The results show that the yield of full length products has been vastly improved by modifying the reverse transcription reaction.



Target cDNA

Lanes 1,3 First strand cDNA (2 µl neat) prepared according to section 2.2.19.1 (following manufacturer's protocol)

2,4 As above (2 µl of a 10-fold dilution)

5 50 ng pea poly(A)$^+$ RNA (–ve control)

6 First-strand cDNA (2 µl neat) section 2.2.19.1, modified protocol

7 As above (2 µl of a 10-fold dilution)

8 50 ng pea poly(A)$^+$ RNA (–ve control)

Primers (see Fig. 3.7.1)

Lanes 1,2,5,6,7,8 GSP + UAP

3,4 GSP+AP

It can be seen from the figure that two bands of approximately 1400 and 1320 bp. However, these bands are very faint when compared to the ladder of smaller PCR products. It was suspected that these products represent truncated 3′ RACE products (the largest, at ~500 bp, is far too small to represent an alternatively spliced MTase mRNA).

Truncated 3′ RACE products arise because of internal priming by the adapter primer during reverse transcription to internal A rich regions of the mRNA. Because these truncated products are much shorter than the full-length species, they compete very effectively during amplification and become the predominant PCR products. The reverse transcription reaction used 10 pmol of Adapter Primer, which is in vast excess over the mRNA (~15:1 molar excess, assuming the average mRNA is 2 kb). Under the low stringency used for reverse transcription (37–42°C), a stretch of A residues as short as 6 nucleotides is sufficient to prime synthesis. Frohman (1994) recommended the use of 2.5 pmol adapter primer (for 1 µg mRNA).

On the basis of this information, it was decided to reduce the primer concentration to 1.2 pmol (for 0.5 µg mRNA). Also, the temperature of the reverse transcription was raised to 45°C (the maximum temperature recommended for use with SuperScript). PCR amplification was carried out using identical PCR conditions and parameters to the ones described above. UAP was selected as the downstream primer because it gave the most specific amplification when a 10-fold dilution of template was used (Fig. 3.7.2 A). The results of amplification are shown in figure 3.7.2 B. A vast improvement in efficiency of amplification of the full-length products has been achieved (the doublet is not resolved very well on this gel).

The next stage was to clone and sequence the full length 3′ RACE products. However, it was decided first to attempt amplification using an alternative thermostable polymerase to avoid the possibility of sequence errors due to *Taq* misincorporations (even though conditions had been used which minimise the number of misincorporations, i.e. low [dNTP]).

The thermostable polymerase selected was *Pfu* DNA polymerase

(Stratagene). This enzyme, like *Vent* polymerase (and unlike *Taq*) has a proof reading ability in the form of a 3′–5′ exonuclease activity. The proof reading DNA polymerases are believed to produce errors much less frequently than *Taq* polymerase (Eckert and Kunkel, 1991). In a study by Lundberg *et al.* (1991), the error rates of *Pfu* and *Taq* polymerase were compared. Twenty cycles of amplification were carried out and the concentration of each dNTP was 200 μM. The error rate (per nucleotide) for *Pfu* polymerase was calculated to be 1.6×10^{-6}, whereas that for *Taq* was 2.0×10^{-5}. If a 1 kb fragment is amplified for 20 effective cycles, 3.2% of *Pfu* generated PCR products and 40% of *Taq* generated products will contain mutations. Extrapolating this to the 3′ RACE experiment (i.e. amplification of a ~1.5 kb fragment using 30 cycles), 90% of *Taq* generated products will contain mutations whereas only 7.2% of *Pfu* generated products will contain mutations. Of course, this assumes that the error frequency is independent of cycle number, and that amplification is exponential throughout the 30 cycles (the latter assumption is probably not correct, i.e. the number of doublings will be less than the number of cycles). However, it is likely that the majority of *Pfu* amplified RACE products will be free of misincorporations.

3′ RACE using *Pfu* polymerase will be described in the next section.

3.7.1.2 3′ RACE using *Pfu* polymerase

3′ RACE was carried out using pea first-strand cDNA prepared using the modified protocol for reverse transcription (described in sections 2.2.19.1 and 3.7.1.1). Native *Pfu* polymerase (Stratagene) was used for PCR amplification under the buffer conditions described in section 2.2.19.3). 10 pmol each of the GSP and AP were used as primers (Fig. 3.7.1). AP had to be used instead of UAP as the downstream primer because *Pfu* polymerase will not copy from templates containing uracil residues. Thermal cycling (again using the Appligene Crocodile thermal cycler) was as follows:

1. The reactions were heated to 95°C for 5 min, then cooled to 75°C for 5 min, during which time the *Pfu* and mineral oil (which was pre-heated to 75°C) were added.

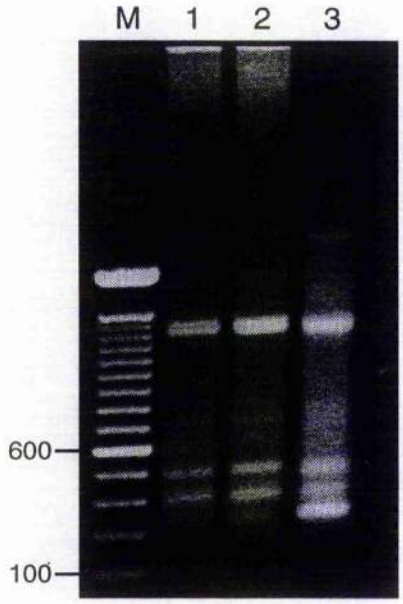
2. The initial annealing step was at 58°C for 10 min followed by an extension step at 75°C for 80 min. A longer annealing step at a lower temperature is used because the salt concentration in the 1 x *Pfu* buffer is less than that used for *Taq* (10 mM KCl as opposed to 50 mM KCl). Stratagene recommend that the annealing temperature should be 5–10°C lower than that used for *Taq*. It was decided (to keep the annealing conditions as stringent as possible for the initial round of amplification) that the annealing temperature should be 6°C lower than the calculated T_m of the GSP (= 64°C, minus the two 5′ adenosines). Stratagene also recommend annealing steps of 5 min, and this was doubled for the first round of amplification. Extension is carried out at 75°C (the optimum temperature for *Pfu*). The extension time was twice that used for *Taq* (3.7.1.1) because of the lower processivity of *Pfu*.

3. 30 cycles of the following were carried out: [95°C for 1 min, 60°C for 5 min, 75°C for 10 min]. A denaturing step of 95°C could be used because of the superior thermostability of *Pfu*. The annealing temperature could be decreased in subsequent experiments, should it prove to be too stringent.

4. A final extension of 75°C for 1 h.

The results of 3′ RACE using *Pfu* polymerase are shown in figure 3.7.3. A doublet of full length 3′ RACE products (of molecular sizes ~1400 and 1320 bp) can clearly be seen. For comparison, a sample of *Taq* amplified 3′ RACE product was loaded in parallel. Unfortunately, this sample has been somewhat overloaded so the doublet was not resolved on the photograph (though it was visible by eye), and the bands slightly displaced. However, since the pattern of truncated products was identical, I was confident that the same pattern of amplification was being achieved with *Pfu* and *Taq*. The next stage was to clone the full length *Pfu* 3′ RACE products in pBluescript for sequencing. (The term "full length" is used to distinguish the 3′ RACE products from truncated products which are an artifact of 3′ RACE).

Figure 3.7.3 **3′ RACE amplification of pea first strand cDNA using *Pfu* polymerase**



The figure shows the results of 3′ RACE amplification of first strand cDNA which was reverse transcribed from pea poly(A)+ RNA following the modified protocol described in section 3.7.1.2.

Lanes 1 and 2 are duplicate 3′ RACE reactions, amplified by *Pfu* polymerase (section 3.7.1.2), using GSP and AP as primers. Lane 3 is a sample of a 3′ RACE reaction amplified by *Taq* polymerase (section 3.7.1.1), using GSP and UAP as primers. Half of each PCR reaction was resolved on a 1% agarose TAE gel along with 0.5 μg 100 bp ladder marker (lane M).

Similar results are obtained with *Pfu* and *Taq*, though *Taq* amplifies more efficiently. The doublet consisting of the two full length products (apparent molecular sizes ~1400 and 1320 bp) are clearly visible in lanes 1 and 2.

3.7.2 Cloning the 3′ RACE products (amplified by *Pfu*) into pBluescript II KS+

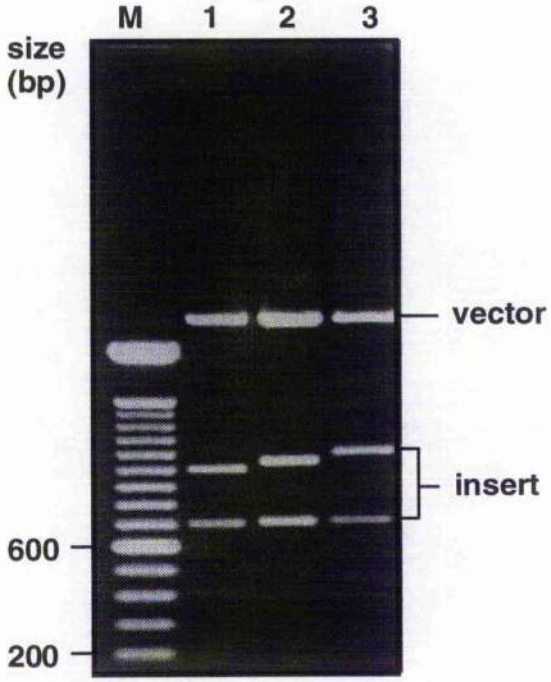
The other halves of the duplicate 3′ RACE reactions (shown in Fig.3.7.3) were resolved on a 1% LMP agarose TAE gel. The doublet consisting of the full length 3′ RACE products was excised from the gel (the doublet could not be separated) and purified using the Geneclean II kit (section 2.2.6). As mentioned earlier, the GSP was designed to incorporate a *Pst*I site at its 5′ end (Fig. 3.7.1) which occurs naturally in the cDNA, to facilitate cloning of the RACE products. The idea was to digest the 3′ RACE products with *Pst*I (which cuts the GSP) and *Sal*I (which cuts the AP), and ligate into pBluescript.

Aliquots of the purified 3′ RACE products were digested with either *Pst*I or *Sal*I and resolved by agarose gel electrophoresis. Unfortunately, *Pst*I did cleave both the 3′ RACE products internally and therefore could not be used as a cloning enzyme. However *Sal*I (which cleaves eukaryotic DNA relatively infrequently for a 6-cutter) did not cleave either RACE product internally.

The alternative cloning strategy was to cut the RACE products with *Sal*I alone, then ligate into *Sal*I + *Sma*I digested pBluescript II KS+ (*Sma*I gives a blunt end). Unlike PCR products amplified with *Taq*, which usually have protruding 3′ ends, *Pfu* amplified PCR products have blunt ends, thus simplifying blunt-ended cloning. This cloning strategy was also undertaken because ligation into the *Sma*I and *Sal*I sites of pBluescript II KS+ also directs the RACE products to be cloned in an antisense orientation. This was to ensure success in cloning, since problems with cloning MTase cDNA fragments in the sense orientation were anticipated (see section 3.6). The purified 3′ RACE products mentioned above were digested with *Sal*I, which was then heat inactivated at 70°C for 15 min. The DNA was then purified using the Geneclean II kit (section 2.2.6). The DNA was then ligated into pBluescript II KS+ which had been digested with *Sal*I and *Sma*I. Ligations were carried out as described in section 2.2.7, using an increased amount of ligase as

appropriate for blunt ended ligation. The ligation mix was used to transform competent *E.coli* strain XL1-Blue (section 2.2.8). A number of white colonies were selected for overnight culture and small-scale preparation of plasmid (section 2.2.9.1). The plasmids were digested with *Pvu*II (*Pvu*II has two sites in pBluescript II KS+, at positions 529 and 977, which flank the multiple cloning site). Only 7 recombinants contained inserts approximating to the size of the 3′ RACE products. Three of these recombinants (digested with *Pvu*II) are shown in figure 3.7.4. As can be seen in the figure, in each case the insert has been cleaved in two by *Pvu*II. The smaller fragment was the same size for each recombinant, whereas the size of the larger fragment varied between recombinants. Three types of recombinant plasmid were obtained as judged by size estimation of the larger insert-derived fragment. Representative clones from each group were selected for sequencing; these are the ones shown in Fig. 3.7.4. Three examples of clone Pfu 1 and Pfu 3 were identified in total (as judged by size estimation of the larger insert-derived fragment obtained upon digestion with *Pvu*II, whereas only one example of Pfu 2 was identified.

Figure 3.7.4 ***Pvu*II digests of 3′ RACE products cloned into pBluescript II KS+**



Samples of each recombinant plasmid (clones Pfu 1, 2 and 3) were digested with *Pvu*II and resolved on a 1% agarose TBE gel (section 2.2.5) along with marker DNA (100 bp ladder). The insert from each clone has been cleaved in two by *Pvu*II. The smallest (insert-derived) fragment in each case was estimated to be 710 bp. The largest insert-derived fragment varied in size between the different clones, and was estimated to be 1100, 1050 and 1010 bp for clones Pfu 3, 2 and 1 respectively.

Lane 3: clone Pfu 3 digested with *Pvu*II

2: clone Pfu 2 "

1: clone Pfu 1 "

M: marker (100 bp ladder)

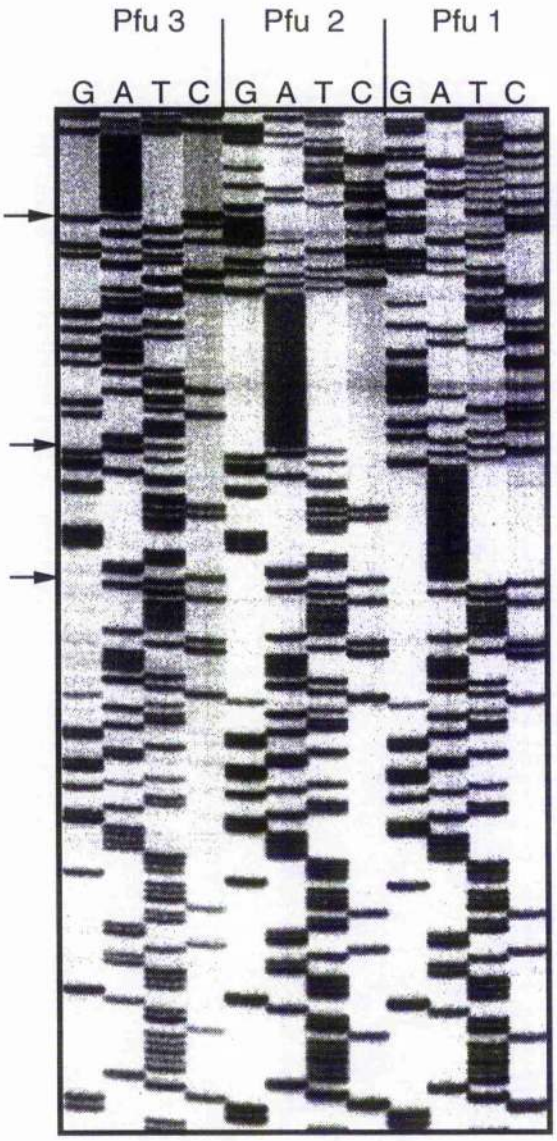
3.7.3 Sequencing the cloned 3′ RACE products

Sequencing of the cloned 3′ RACE products (clones Pfu 1, Pfu 2 and Pfu 3) was carried out manually using the Sequenase II kit (USB). Initially, M13 (–20) Universal and Reverse sequencing primers (Pharmacia) were used on single-stranded and double-stranded templates respectively, as described in section 2.2.20. Because the products were cloned in the antisense orientation in pBluescript II KS+, the antisense strand of the insert is the sequencing template when M13 Universal primer is used, and the sequence data generated corresponds to the sense strand in the 5′–3′ direction. The three clones were found to have identical sequence at the 5′ end but not at the 3′ end. It was therefore decided to continue sequencing manually on single-stranded template (plus strand) using a 'primer walking' strategy (the sequencing primers were synthesized by Pharmacia). A picture of part of a sequencing gel is shown in figure 3.7.5.

Each clone was found to be absolutely identical in sequence, except at the extreme 3′ end, where the variability in fragment size was found to be due to alternative polyadenylation sites in each of the three clones. Ideally, the clones should have been sequenced in both directions. However, as there were no problems with sequencing artefacts (such as band compressions), and the fact that the clones were identical for the most part, thus acting as internal controls, sequencing the opposite strand was deemed not to be necessary. Where the clones differed at their 3′ ends, sequencing of the opposite strand was carried out (i.e. with the Reverse primer).

The overlapping sequence data was assembled to form a contiguous sequence, which was then translated in 3 forward frames, starting from the natural *Pst*I site of the gene-specific primer (Fig.3.7.1*a*). Frame 3 was found to be the correct reading frame as expected (i.e. it consists of an ORF which codes for all the conserved motifs from IV–X). The sequence of the longest clone (translated in the third frame) is shown in figure 3.7.6. Also illustrated are the polyadenylation sites used in the shorter 3′ RACE clones, Pfu 1 and Pfu 2.

Figure 3.7.5 **An example of sequence data for the cloned 3′ RACE products**



Part of a sequencing gel showing the extreme 3′ end of the cloned 3′ RACE products (clones Pfu 1, 2 and 3) is shown. The arrows indicate the start of the poly(A) tails for each clone. The sequence of each clone is absolutely identical except for the extreme 3′ end, where differences in sequence are due to the use of different polyadenylation sites in each clone.

Figure 3.7.6 Sequence of the cloned 3′ RACE products

The figure shows the sequence of the longest pea MTase 3′ RACE clone (Pfu 3), along with the predicted sites of selected enzymes (six-cutters only).

The sequence of the gene-specific primer (Fig. 3.7.1), ignoring the first two adenosines, is in upper case and is underlined. This primer was used as the upstream primer for 3′ RACE.

The sequencing primers designed from the sense strand sequence were as follows:

SEQ 1 5′ CCTCTTGGAGAATGTGA 3′

SEQ 2 5′ ATGTCCAGTATGCTGCC 3′

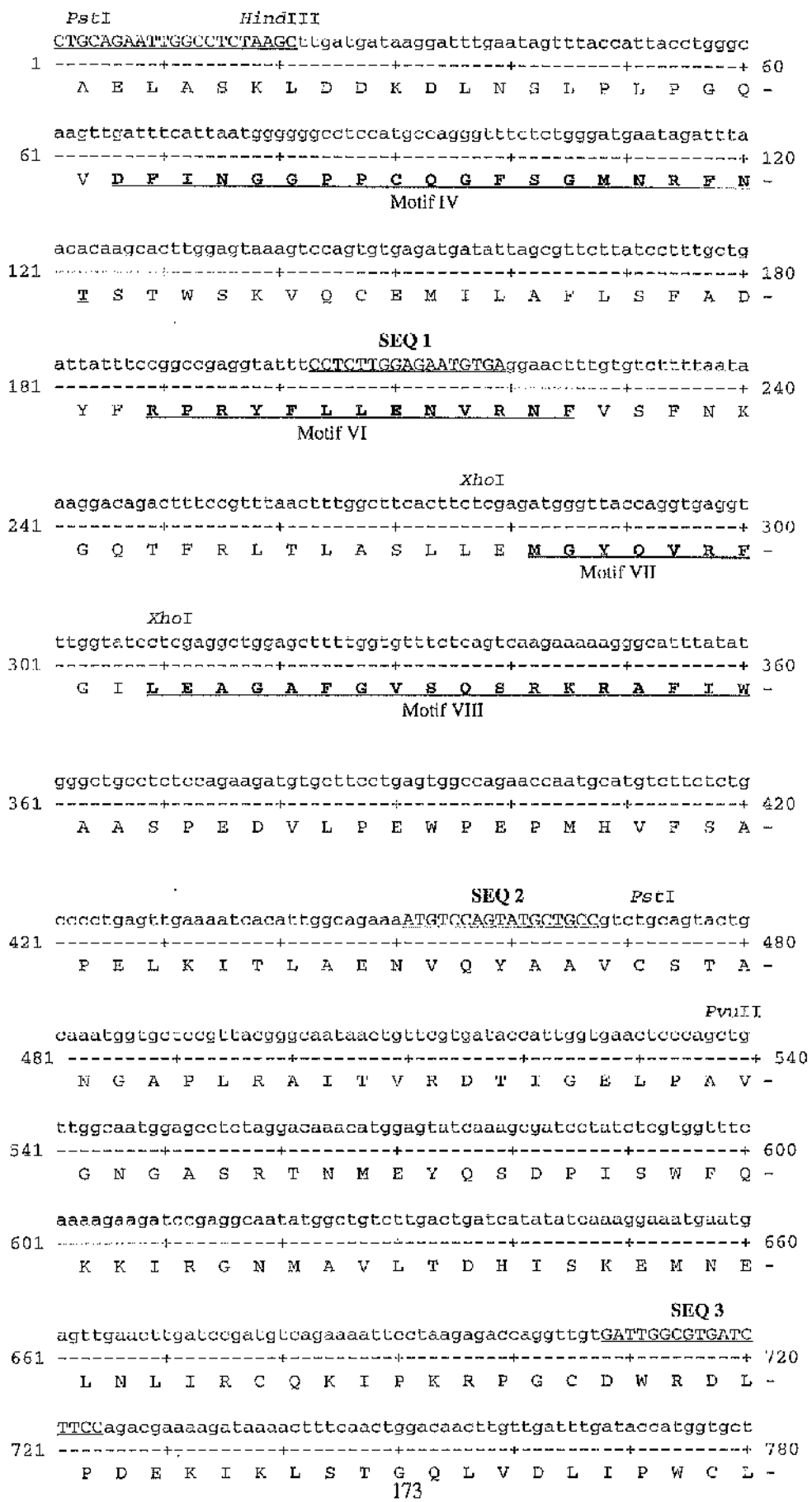
SEQ 3 5′ GATTGGCGTGATCTTCC 3′

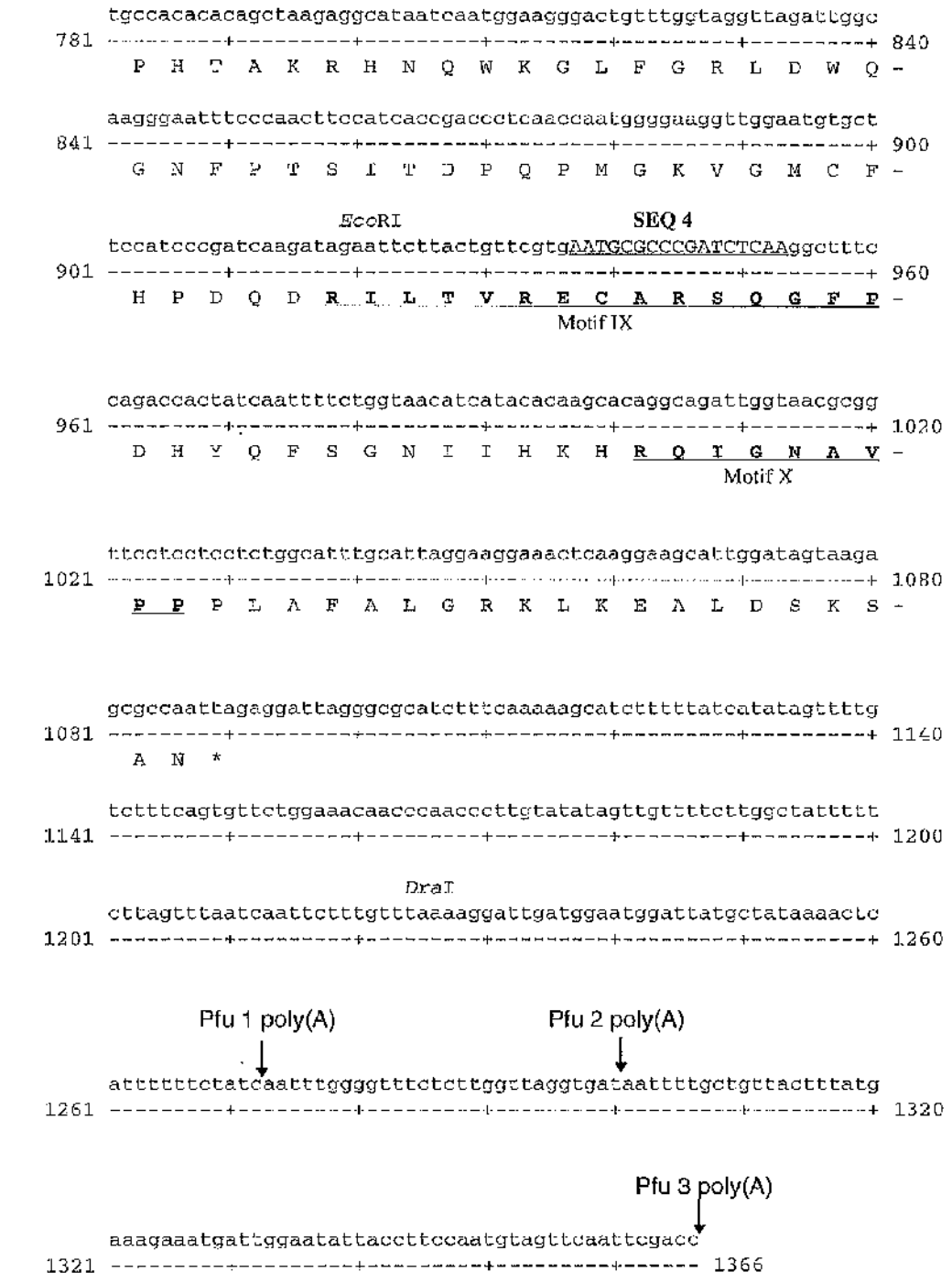
SEQ 4 5′ AATGCGCCCGATCTCAA 3′

These sequences are shown in upper case and are underlined in the figure

The translated sequence (frame 3 ORF) is shown below the nucleotide sequence. The conserved motifs identified (IV to X) are indicated in bold and are underlined. Motif V is absent, which is also the case for the other eukaryotic MTases cloned so far.

The polyadenylation sites of the two shorter clones (Pfu 1 and Pfu 2) are indicated by arrows.


PstI
HindIII
CTGCAGAATTGGCCTCTAAGCttgatgataaggatttgaatagtttaccattacctgggc
1
60
A E L A S K L D D K D L N S L P L P G Q -
aagttgatttcattaatgggggcctccatgccagggtttctctgggatgaatagattta
61
120
V D F I N G G P P C Q G F S G M N R F N -
Motif IV
acacaagcacttggagtaaagtccagtgtgagatgatattagcgttcttatcctttgctg
121
180
T S T W S K V Q C E M I L A F L S F A D -
SEQ 1
attatttccggccgaggtatttCCTCTTGGAGAATGTGAggaactttgtgtcttttaata
181
240
Y F R P R Y F L L E N V R N F V S F N K
Motif VI
XhoI
aaggacagactttccgtttaactttggcttcacttctcgagatgggttaccaggtgaggt
241
300
G Q T F R L T L A S L L E M G Y Q V R F -
Motif VII
XhoI
ttggtatcctcgaggctggagcttttggtgtttctcagtcaagaaaaagggcatttatat
301
360
G I L E A G A F G V S Q S R K R A F I W -
Motif VIII
gggctgcctctccagaagatgtgcttcctgagtggccagaaccaatgcatgtcttctctg
361
420
A A S P E D V L P E W P E P M H V F S A -
SEQ 2
PstI
ccccctgagttgaaaatcacattggcagaaaATGTCCAGTATGCTGCCgtctgcagtactg
421
480
P E L K I T L A E N V Q Y A A V C S T A -
PvuII
caaatggtgctccgttacgggcaataactgttcgtgataccattggtgaactcccagctg
481
540
N G A P L R A I T V R D T I G E L P A V -
ttggcaatggagcctctaggacaaacatggagtatcaaagcgatcctatctcgtggtttc
541
600
G N G A S R T N M E Y Q S D P I S W F Q -
aaaagaagatccgaggcaatatggctgtcttgactgatcatatatcaaaggaaatgaatg
601
660
K K I R G N M A V L T D H I S K E M N E -
SEQ 3
agttgaacttgatccgatgtcagaaaattcctaagagaccaggttgtGATTGGCGTGATC
661
720
L N L I R C Q K I P K R P G C D W R D L -
TTCCagacgaaaagataaaactttcaactggacaacttgttgatttgataccatggtgct
721
780
P D E K I K L S T G Q L V D L I P W C L -



Enzymes that do cut (six cutters):

*Dra*I *Eco*RI *Hind*III *Pst*I *Pvu*II *Xho*I

3.7.4 Discussion of the 3′ RACE experiment

A number of conclusions can be made about the results of the 3′ RACE experiment.

1. The sequence of all three cloned 3′ RACE products is identical to that of the pea MTase-specific PCR products amplified using degenerate primers (excluding the degenerate primer sequences), which were described in the previous section (see Fig. 3.6.6). Because the upstream gene-specific primer was designed on the basis of sequence data of a MTase clone isolated by Sriharsa Pradhan, this proves that both the degenerate PCR amplified fragments and the cDNA clone isolated on the basis of N-terminal sequencing of the CWG MTase originate from the same pea MTase mRNA. Only one mRNA hybridised on the northern blot (Fig. 3.6.11), the degree of resolution being insufficient to distinguish alternative polyadenylation sites.

2. As mentioned in section 3.7.3, the three 3′ RACE clones isolated were absolutely identical in sequence and differed only in the choice of polyadenylation site. Three classes of clones were identified on the basis of the size of the largest insert-derived fragment obtained upon digestion with *Pvu*II. The number of poly(A) tail derived adenosines was 25, 41 and 28 for clones Pfu 1, 2 and 3 respectively. In the reverse transcription step of the 3′ RACE reaction (where there is an excess of oligo dT primer), primers annealing nearest to the polyadenylation site will be preferentially reverse transcribed to form full length cDNA. This is reflected in the fact that both clones Pfu 1 and 3 contain relatively short stretches of the poly(A) tail. Clone Pfu 2 on the other hand has a slightly longer stretch of adenosines. This has led to this particular clone being distinguished from the members of the shortest class in terms of insert size. It is quite probable that other members of the shortest class have the same polyadenylation site as Pfu 2, but could not be distinguished in terms of insert size. It is also possible that other polyadenylation sites exist. The use of alternative polyadenylation sites has also been reported for *Arabidopsis* MTase mRNA (Finnegan and Dennis, 1993). The existence and significance of alternative polyadenylation sites will be discussed in more detail later.

The possibility that the other 4 clones which were not sequenced were alternatively spliced within the variable region was excluded by digesting samples of each clone with *Eco*RI. This enzyme contains one site within the insert, at the beginning of the region specifying motif IX (Fig. 3.7.6). The vector *Eco*RI site lies outside the *Sma*I site (Fig. 2.1.2), into which the 5′ end of the RACE products have been cloned. Digestion of all 7 clones yielded a single fragment, of ~930 bp, which is the size expected from the sequence data (Fig. 3.7.6). The size of this fragment appeared the same in each case, which shows that the variablity in size does not lie within the 5′ end (including the region specifying the TRD) of the cloned RACE products.

3. These experiments provided no evidence for alternative splicing events occurring in the 3′ end of pea MTase. To exclude the possibility that alternatively spliced forms have been 'missed' by the cloning procedure, a sample of the gel-purified uncloned 3′ RACE products was digested with *Eco*RI. However, only one band of apparent molecular weight ~920 bp was visible, which corresponds to that predicted by the sequence of the cloned 3′ RACE fragments. This proves that the heterogeneity giving rise to the doublet in the 3′ RACE product arises from alternative polyadenylation sites only. This however does not exclude the possibility that alternative splicing giving rise to very small changes in the size of this fragment might occur; this might be revealed by direct sequencing of the PCR products.

It was also possible that the region corresponding to the gene-specific primer is excluded in a differently spliced mRNA. To test this, sequencing primers 1 and 2 were used in an analogous 3′ RACE experiment (using *Taq* polymerase). The protocol had to be modified because of the shortness of the primers, i.e. the annealing temperature was decreased to 50°C. The reaction was found to be very inefficient and to visualise specifically amplified bands, the gel was blotted and probed with the insert of the cloned 3′ RACE product, Pfu 3. This was obtained by digesting clone Pfu 3 with *Bam*HI (which cuts outside the *Sma*I site into which the 5′ end of the insert has been cloned) and *Sal*I (which is the cloning site of the 3′ end

of the insert). Neither of these enzymes cleave the insert internally (see Fig. 3.7.6). The results are shown in figure 3.7.7. Apart from truncated products, two bands were found to be amplified with both SEQ 1 and SEQ 2 primers. The sizes were ~1200 and ~1100 for SEQ 1 and ~950 and ~850 bp for SEQ 2. The size of the band obtained with the control primer (i.e. the original primer used for 3´ RACE (oligo 1838; Fig. 3.7.1) was ~1300 on the autoradiograph (the doublet could not be resolved). These values agree with those expected from the relative positions of SEQ 1 and SEQ 2 and show no evidence for an alternatively spliced mRNA, assuming there would still be heterogeneity of polyadenylation sites in the alternatively spliced molecule.

3.7.5 Sequence comparisons between the pea MTase 3´ RACE product with other MTases

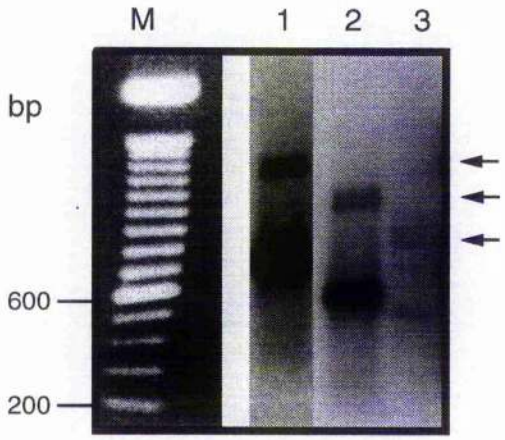
First of all, a FASTA search was carried out using the translated sequence of pea MTase (Fig. 3.7.6) with the SwissProt database on GCG. Not surprisingly, the best match was found with *Arabidopsis* MTase, followed by human and mouse MTases. Of the prokaryotic sequences which matched, only one sequence (M.*Ban*I) had an alignment which spanned the variable domain. Using the Pileup program in GCG, a multiple alignment was created between the C-terminal end of the eukaryotic MTases and selected prokaryotic type II C5-cytosine MTases. This is shown in figure 3.7.8. The prokaryotic enzymes selected for comparison were M.*Ban*I (sequence specificity = GGYRCC) because it gave the best alignment (out of the prokaryotic MTases) with the pea MTase sequence in the FASTA search, M.*Sss*I (sequence specificity = CG) because its recognition sequence is identical to that of the mammalian enzymes, and M.*Eco*RII (sequence specificity = CCWGG). The latter enzyme was chosen as an example of a bacterial CWG MTase (no bacterial MTases have been identified which recognise just the motif CWG). M.*Hha*I (sequence specificity = GCGC) was also included because its co-crystal structure has been solved (see section 1.4.2).

Figure 3.7.7 3′ RACE experiment using SEQ 1 and SEQ 2 primers

3′ RACE was carried out using Taq polymerse as detailed in section 3.7.1.1, using either SEQ 1 or SEQ 2 as the upstream gene-specific primers (Fig. 3.7.6) and UAP as the downstream primer (Fig. 3.7.1). The upstream gene-specific primer used in previous experiments (oligo 1838) (Fig. 3.7.1) was also used in a control PCR reaction. The PCR parameters were identical to those described in section 3.7.1.1, except that the annealing temperature was reduced to 50°C.

The PCR products wrer resolved on a 1% agarose TAE gel along with marker DNA (100 bp ladder). The gel was then blotted and probed at high stringency with the insert from clone Pfu 3 (see Fig. 3.7.6), which was obtained by digesting Pfu 3 with *Bam*HI and *Sal*I.

The amplified doublet was of the expected size in each case; the position of each doublet is indicated by an arrow.



Lane M marker DNA (100 bp ladder)

Upstream primer

Lane 1 oligo 1838

2 SEQ 1

3 SEQ 2

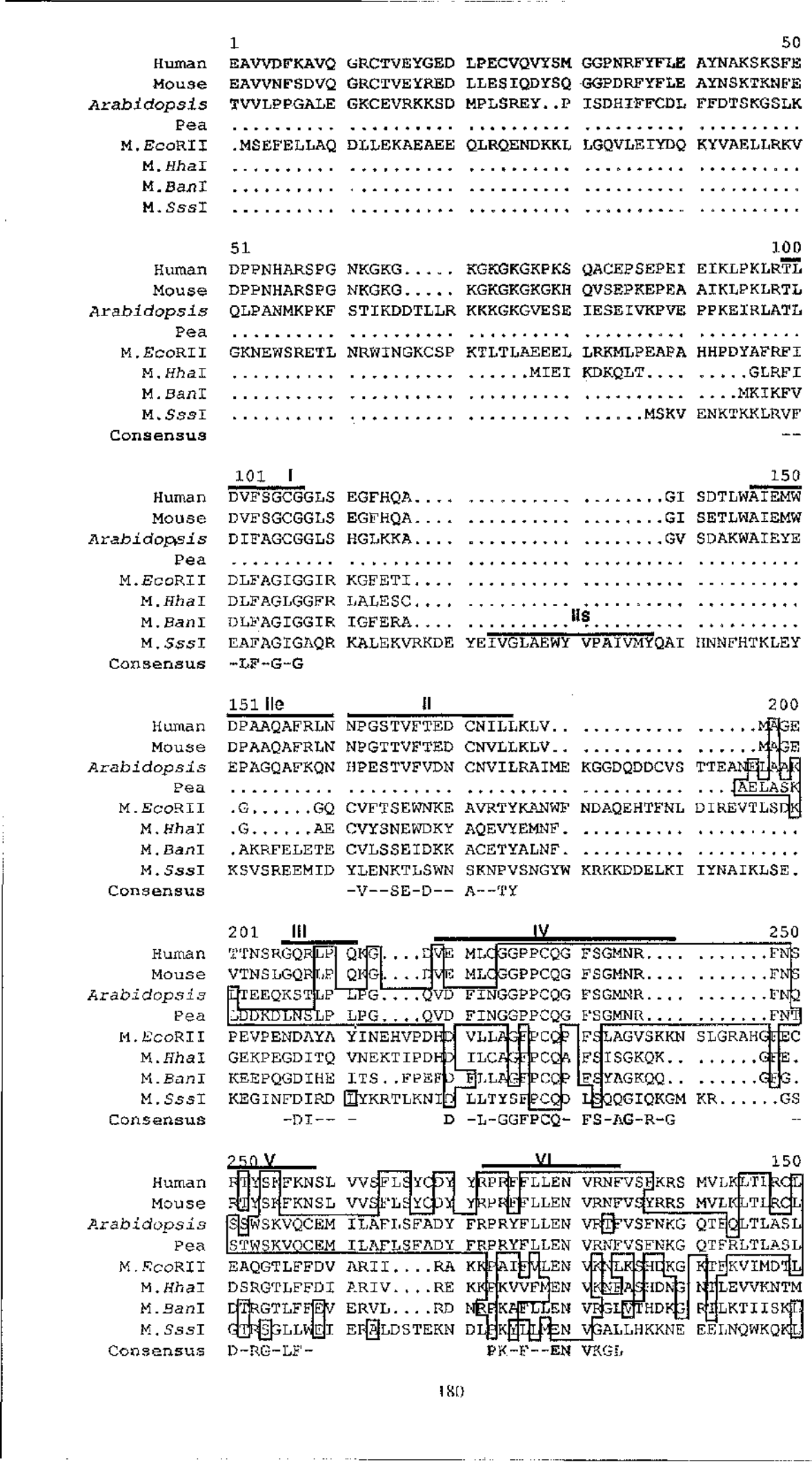
Figure 3.7.8 Sequence alignment between pea MTase and other MTases

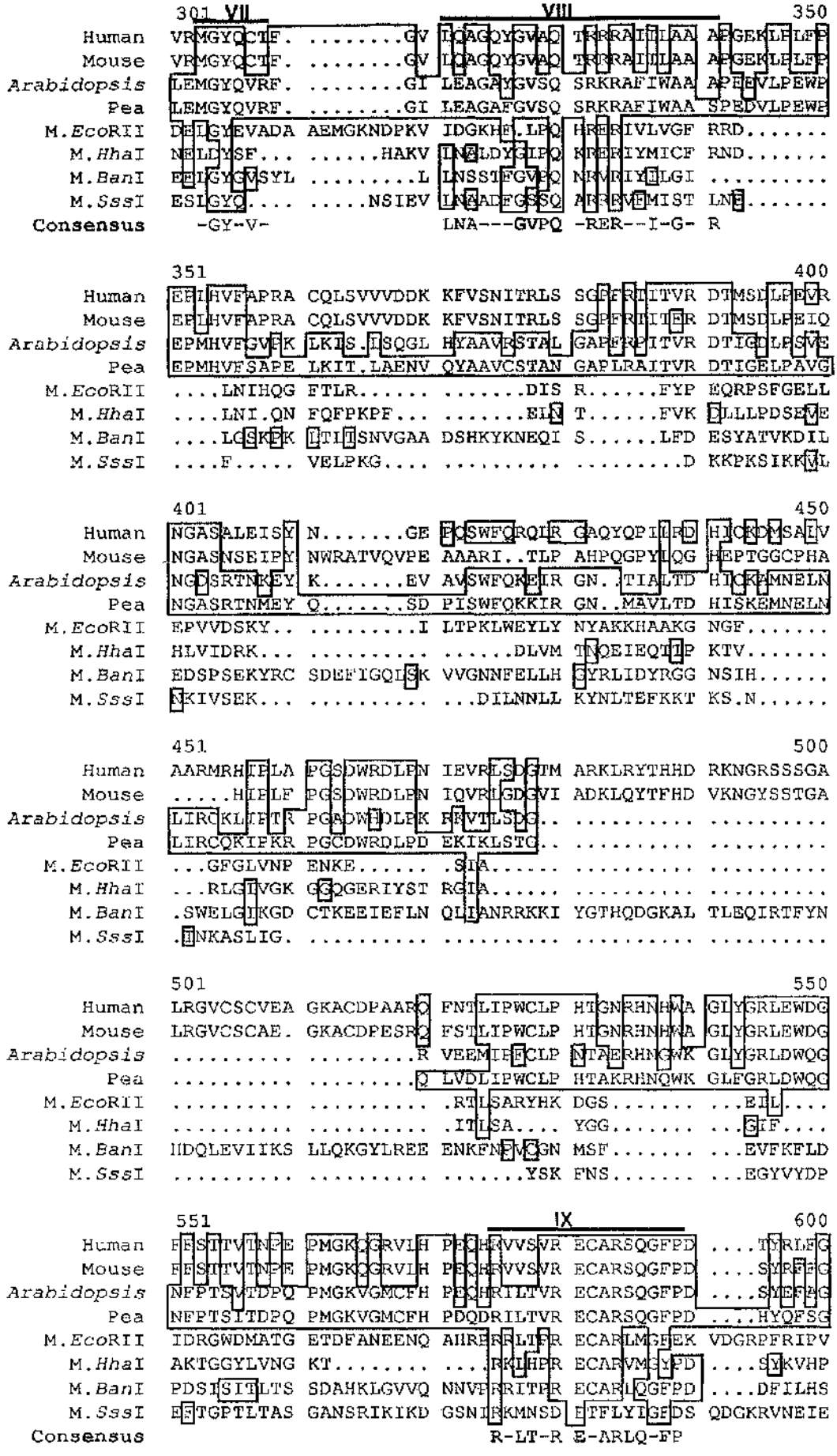
The predicted amino acid sequence of the cloned pea MTase 3´ RACE product was aligned with other MTases using the Pileup program of GCG (the default parameters were used, i.e. gap creation penalty = 3, gap extension penalty = 0.1). This program creates a multiple sequence alignment from a group of related sequences using progressive pairwise alignments. The MTases used for the Pileup and their sequence specificities (either known or inferred) are as follows:

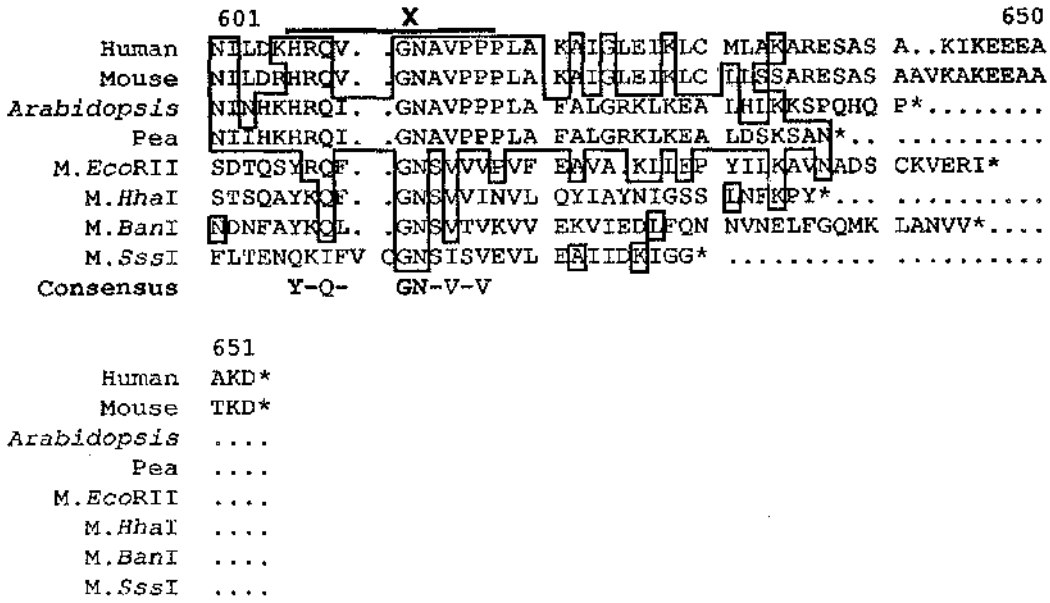
Pea MTase	^{m}CWG or ^{m}CG (?)
Mouse MTase	^{m}CG
Human MTase	^{m}CG
Arabidopsis MTase	^{m}CG or ^{m}CNG (?)
M.*Ban*I	GYR^{m}CC or GGYRC^{m}C (?)
M.*Eco*RII	C^{m}CWGG
M.*Sss*I	^{m}CG
M.*Hha*I	G^{m}CGC

The positions of the conserved motifs are indicated above the alignment, and the prokaryotic consensus shown below. The consensus identified by Posfai *et al.* (1989) is indicated under the motifs II, III, V and VII, although these have since been shown to be less well conserved. Under the remainder of the motifs is depicted the consensus identified by Wilson, 1992 (which was derived from comparisons between 30 prokaryotic MTases). Amino acids occurring in 50% or more of the MTases are indicated in plain text; those occurring in 75% or more of the MTases are indicated in bold. In the case of motif II, where the eukaryotic motifs fail to align with their prokaryotic counterparts, the eukaryotic motif II is indicated separately (IIe). Motif II of M.SssI is also indicated separately (IIs).

The boxed amino acids are absolutely identical to the pea MTase sequence.


1 50
Human EAVVDFKAVQ GRCTVEYGED LPECVQVYSM GGPNRFYFLE AYNAKSKSFE
Mouse EAVVNFSDVQ GRCTVEYRED LLESIQDYSQ GGPDRFYFLE AYNSKTKNFE
Arabidopsis TVVLPPGALE GKCEVRKKSD MPLSREY..P ISDHIFFCDL FFDTSKGSLK
Pea .....
M.EcoRII .MSEFELLAQ DLLEKAEAEE QLRQENDKKL LGQVLEIYDQ KYVAELLRKV
M.HhaI .....
M.BanI .....
M.SssI .....
51 100
Human DPPNHARSPG NKGKG..... KGKGKGKPKS QACEPSEPEI EIKLPKLRTL
Mouse DPPNHARSPG NKGKG..... KGKGKGKGKH QVSEPKEPEA AIKLPKLRTL
Arabidopsis QLPANMKPKF STIKDDTLLR KKKGKGVESE IESEIVKPVE PPKEIRLATL
Pea .....
M.EcoRII GKNEWSRETL NRWINGKCSP KTLTLAEEEL LRKMLPEAPA HHPDYAFRFI
M.HhaI .....MIEI KDKQLT..... .....GLRFI
M.BanI ....MKIKFV
M.SssI .....MSKV ENKTKKLRVF
Consensus --
101 I 150
Human DVFSGCGGLS EGFHQA..... ....GI SDTLWAIEMW
Mouse DVFSGCGGLS EGFHQA..... ....GI SETLWAIEMW
Arabidopsis DIFAGCGGLS HGLKKA..... ....GV SDAKWAIEYE
Pea .....
M.EcoRII DLFAGIGGIR KGFETI..... ....
M.HhaI DLFAGLGGFR LALESC..... ....
M.BanI DLFAGIGGIR IGFERA..... ....
IIs
M.SssI EAFAGIGAQR KALEKVRKDE YEIVGLAEWY VPAIVMYQAI HNNFHTKLEY
Consensus -LF-G-G
151 IIe II 200
Human DPAAQAFRLN NPGSTVFTED CNILLKLV.. ....MAGE
Mouse DPAAQAFRLN NPGTTVFTED CNVLLKLV.. ....MAGE
Arabidopsis EPAGQAFKQN HPESTVFVDN CNVILRAIME KGGDQDDCVS TTEANELAAK
Pea ....AELASK
M.EcoRII .G.....GQ CVFTSEWNKE AVRTYKANWF NDAQEHTFNL DIREVTLSDK
M.HhaI .G.....AE CVYSNEWDKY AQEVYEMNF. ....
M.BanI .AKRFELETE CVLSSEIDKK ACETYALNF. ....
M.SssI KSVSREEMID YLENKTLSWN SKNPVSNGYW KRKKDDELKI IYNAIKLSE.
Consensus -V--SE-D-- A--TY
201 III IV 250
Human TTNSRGQRLP QKG....DVE MLCGGPPCQG FSGMNR..... .....FNS
Mouse VTNSLGQRLP QKG....DVE MLCGGPPCQG FSGMNR..... .....FNS
Arabidopsis LTEEQKSTLP LPG....QVD FINGGPPCQG FSGMNR..... .....FNQ
Pea LDDKDLNSLP LPG....QVD FINGGPPCQG FSGMNR..... .....FNT
M.EcoRII PEVPENDAYA YINEHVPDHD VLLAGFPCQP FSLAGVSKKN SLGRAHGFEC
M.HhaI GEKPEGDITQ VNEKTIPDHD ILCAGFPCQA FSISGKQK.. ......GFE.
M.BanI KEEPQGDIHE ITS..FPEFD FLLAGFPCQP FSYAGKQQ.. ......GFG.
M.SssI KEGINFDIRD IYKRTLKNID LLTYSFPCQD LSQQGIQKGM KR......GS
Consensus -DI-- - D -L-GGFPCQ- FS-AG-R-G -
250 V VI 150
Human RTYSKFKNSL VVSFLSYCDY YRPRFFLLEN VRNFVSFKRS MVLKLTLRCL
Mouse RTYSKFKNSL VVSFLSYCDY YRPRFFLLEN VRNFVSYRRS MVLKLTLRCL
Arabidopsis SSWSKVQCEM ILAFLSFADY FRPRYFLLEN VRNFVSFNKG QTFQLTLASL
Pea STWSKVQCEM ILAFLSFADY FRPRYFLLEN VRNFVSFNKG QTFRLTLASL
M.EcoRII EAQGTLFFDV ARII....RA KKPAIEVLEN VKNLKSHDKG KTFKVIMDTL
M.HhaI DSRGTLFFDI ARIV....RE KKPKVVFMEN VKNFASHDNG NTLEVVKNTM
M.BanI DTRGTLFFEV ERVL....RD NRPKAFLLEN VRGLVTHDKG RTLKTIISKL
M.SssI GTRSGLLWEI ERALDSTEKN DLPKYLLMEN VGALLHKKNE EELNQWKQKL
Consensus D-RG-LF- PK-F--EN VKGL


VII
VIII
IX
Human
Mouse
Arabidopsis
Pea
M.EcoRII
M.HhaI
M.BanI
M.SssI
Consensus


601
X
650
Human NILDKHRQV. .GNAVPPPLA KAIGLEIKLC MLAKARESAS A..KIKEEEA
Mouse NILDKHRQV. .GNAVPPPLA KAIGLEIKLC MLSSARESAS AAVKAKEEAA
Arabidopsis NINHKHRQI. .GNAVPPPLA FALGRKLKEA LHIKKSPQHQ P*.....
Pea NIIHKHRQI. .GNAVPPPLA FALGRKLKEA LDSKSAN*.. .....
M.EcoRII SDTQSYRQF. .GNSVVVEVF EAVA.KILEP YIIKAVNADS CKVERI*
M.HhaI STSQAYKQF. .GNSMVINVL QYIAYNIGSS LNFKPY*... .....
M.BanI NDNFAYKQL. .GNSMTVKVV EKVIEDLFQN NVNELFGQMK LANVV*....
M.SssI FLTENQKIFV CGNSISVEVL EAIIDKIGG* ....
Consensus Y-Q- GN-V-V
651
Human AKD*
Mouse TKD*
Arabidopsis ....
Pea ....
M.EcoRII ....
M.HhaI ....
M.BanI ....
M.SssI ....

It is obvious from figure 3.7.8 that the eukaryotic MTases form a class of their own in terms of sequence conservation in the C-terminal catalytic domain (over the range studied). Using the program BESTFIT, The predicted pea MTase amino acid sequence was found to share 84% similarity and 77% identity with *Arabidopsis* MTase, and 72% similarity and 52% identity at the amino acid level with human MTase. As a group, the eukaryotic MTases share regions of identity with the prokaryotic enzymes only in the highly conserved motifs (i.e. I, IV, VI, VIII, IX and X), and also motif VII. Motifs III and V are absent in eukaryotic MTases. Motif II can be identified in the eukaryotic enzymes but it is dissimilar enough from the prokaryotic consensus not to align with prokaryotic motif II. However, The same is true of motif II of M.*Sss*I.

Comparison of the variable domain (between motifs VIII and IX) revealed very little sequence similarity between eukaryotic and prokaryotic MTases, regardless of sequence specificity. M.*Ban*I, the most closely related prokaryotic MTase to pea MTase identified in the FASTA search, had very little similarity in the variable domain; the reason the match spanned the variable domain is probably more a function of the similar size of the variable domain rather than its sequence.

M.*Sss*I is an unusual prokaryotic MTase in that it methylates cytosines processively, as it interacts with duplex DNA irrespective of the presence of CpG sequences (Renbaum and Razin, 1992). In this sense, as well as in its target specificity, it is similar to the mammalian MTases. However, its amino acid sequence is quite dissimilar to the C-terminal domain of the mammalian MTases. It also differs from the mammalian enzymes in that it does not prefer hemimethylated targets over unmethylated ones.

M.*Eco*RII is a member of a highly related group of prokaryotic MTases whose members methylate the inner cytosine of either CCWGG or CCNGG motifs. Members of this group share significant homology in the variable domain. Gopal *et al.* (1994) identified conserved segments in the variable domain in this class of enzyme, which interestingly were found to be similar both in sequence and in

predicted secondary structure to the region of M.*Hha*I which is known to contact the DNA bases and confers sequence specificity (Klimasauskas *et al.*, 1994). The fact that M.*Hha*I has a quite different sequence specificity (= G^{m}CGC) implies that MTases with differing specificities utilise a similar structural framework for sequence recognition. The variable domain from the plant MTases was compared (by eye) to that of the CCWGG/CCNGG group of enzymes. However, nothing resembling the consensus identified by Gopal *et al.*, (1994) could be found. Perhaps the apparent sequence similarity in the variable domain between M.*Hha*I and the CCWGG/CCNGG group of MTases reflects relative evolutionary closeness of these enzymes. Sequence comparison of this sort was found to be uninformative about the sequence specificity of the plant MTases. The eukaryotic "pseudo DNA MTase" encoded by the *pmt1*$^{+}$ gene from *S. pombe* (Wilkinson et al., 1995; Pinabarsi *et al.*, personal communication) was also compared to the eukaryotic MTases. As mentioned in section 1.5.4, the predicted amino acid sequence of this gene contains all the conserved motifs (I–IX) of the prokaryotic cytosine-C5 MTases, except that motif IV is altered. However, this enzyme bears little resemblance to the eukaryotic MTases (even less than some other prokaryotic MTases) and is therefore unlikely to be ancestrally related (comparison not shown).

Comparison of the variable domains between the mammalian and the plant MTases revealed a high degree of sequence similarity, albeit less than that of the conserved motifs. Interestingly, the mammalian enzymes contain an extra 40 amino acids inserted within a region which is otherwise highly conserved. Either this region has been lost by (dicotyledonous) plants or gained by mammals during evolution. The possibility that another MTase containing this sequence is present in pea as a result of alternative splicing has already been excluded. The fact that the surrounding amino acid sequences in the variable region are fairly well conserved between mammals and plants implies that this difference may have functional significance. It is possible that this difference in the variable region could be responsible for the different target specificities of the mammalian and plant MTases. It has been

hypothesised that the multispecific C5-cytosine MTases encoded by *Bacillus* phages are able to methylate different target sequences because of conformational flexibility (Noyer-Weidner and Trautner, 1993). As we only have evidence for the existence of one MTase mRNA in pea (see chapter 4), it is conceivable that the absence of this stretch of 40 amino acids allows two different conformations to be adopted in the variable region of the plant enzymes, which would give rise to either CG or CNG target specificities. Of course, the difference may merely represent the relative evolutionary closeness of the organisms from which the MTases were cloned, rather than the sequence specificity of the enzymes; the cloning of other plant and animal MTase cDNAs from more phylogenetically distant organisms would test the validity of this argument. Also, mammalian MTase has been shown to methylate CWG sequences *in vitro*, albeit to a small extent (Adams *et al.*, 1993), and it has recently been shown that mammalian cells are able to methylate CWG sites in transfected plasmid DNA (Clark *et al.*, 1995), so there may be other factors influencing target specificity of eukaryotic MTases.

This discussion has so far neglected the possible influence of the N-terminal domain on the target specificity of eukaryotic MTases. It has previously been demonstrated that the N-terminal domain of mouse MTase suppresses *de novo* methylation of CpG sites, as discussed in section 1.5.2. The N-terminal domain may also have an influence on sequence specificity in the plant MTases. As discussed in section 1.5.3, the CG and CWG specificities have been shown to reside in separate MTases in pea, even though there appears to be only one MTase mRNA. The different specificities could arise from differences in the N-terminal domain, which could be due to post-translational modification or the use of an alternative translational start site. One could envisage specific interactions between regions in the N-terminal domain and the variable domain that bring about the correct conformation in the variable domain to allow sequence recognition. However, the exact nature of the N-terminal involvement in target specificity will only be determined once the crystal structure of a eukaryotic MTase complexed with DNA has been determined.

3.7.6 Alternative polyadenylation sites in the pea MTase mRNA

Unlike animal mRNAs which usually have well defined 3′ ends, plants generally have heterogeneous 3′ ends which are spread over a region of up to 150 nucleotides in length (Dean *et al.*, 1986). Moreover, mammalian polyadenylation signals function inefficiently and inaccurately in plants (Hunt *et al.*, 1987), which suggests that 3′ end processing may be related but is not identical. In a study by Joshi (1987), less than half of the plant genes examined contained a perfect copy of the metazoan AAUAAA consensus upstream of the polyadenylation site. This is similar to the situation in yeast, where about 50% of genes were found to contain the full consensus (Hyman *et al.*, 1991). In plants, conservation of the AAUAAA motif is often limited to the first 3 nucleotides (Ellison and Messing, 1988), though single mutations even in the first 3 nt of the CaMV polyadenylation signal AAUAAA motif were shown to be tolerated well (Rothnie *et al.*, 1994). Plant 3′ end processing also differs from that of mammals in that specific sequences downstream of the polyadenylation site do not seem to be required. Mogen *et al.* (1992) identified (by linker scanning mutagenesis), near upstream elements (NUEs) and a far upstream element (FUE) which were essential for polyadenylation at the 3 main sites used in a pea *rbc*S gene. Each site had its own NUE within –40 nt, whereas deletion of the FUE (–60 to –120 away from site 2) abolished polyadenylation at all 3 sites (the use of a 4th site, 50 nt downstream from site 2, was greatly increased). The NUEs contain sequences related to AAUAAA, whereas the FUE contains a repeated UUGUA motif, as well as UG rich sequences (the individual elements of the FUE appear to be functionally redundant). Interestingly, sequences related to this UUGUA motif have been found to be essential for 3′ end formation in yeast (Zaret *et al.*, 1982) and are also found in the SV40 late polyadenylation signal (Schek *et al.*, 1992).

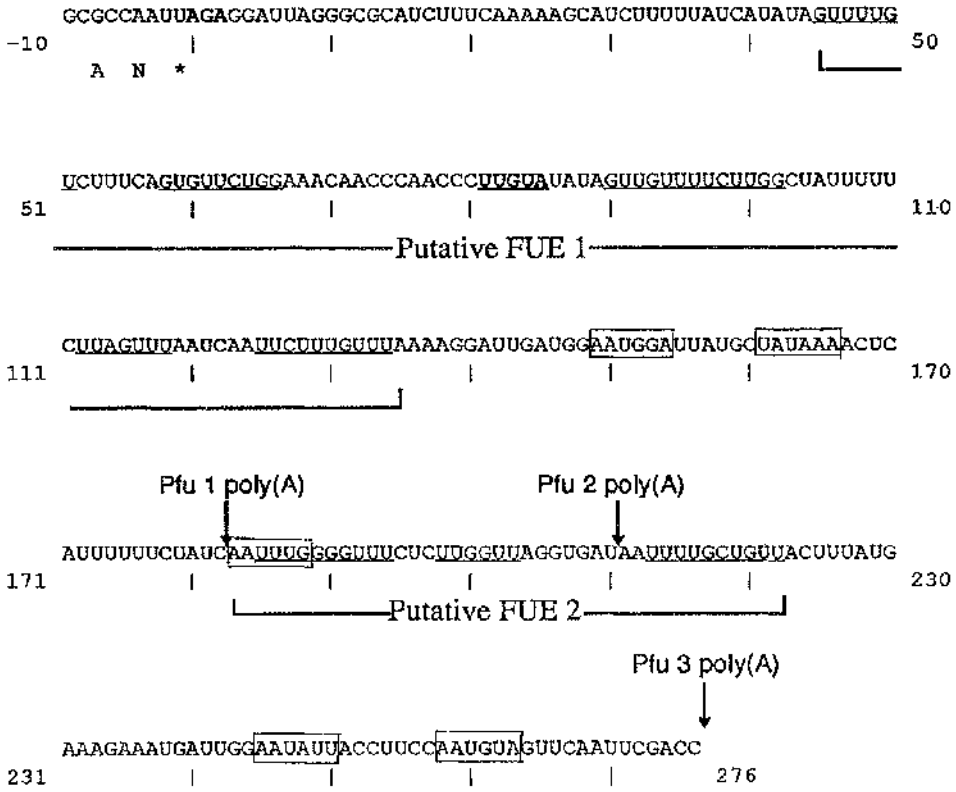
The sequences of the cloned pea MTase 3′ RACE products were analysed for putative polyadenylation signals (Fig. 3.7.9). Each polyadenylation site conforms to the consensus of cleavage sites observed by Joshi (1987), i.e. YA. It is likely that the

first adenosine of the poly(A) tail for each site has arisen from the gene and not polyadenylation. The sequence AAUAAA is absent, though several related sequences are present. Sequences which match this consensus as closely as possible, and which are within the correct distance (i.e. less than 40 nt upstream) from each of the three polyadenylation sites are highlighted. Of course, these merely represent putative NUEs. Also identified is a putative FUE which is approximately the correct distance from sites 1 and 2 (after Mogen *et al.*, 1992). This contains the motif UUGUA plus GU rich sequences. However, this motif is possibly too far away from site 3. A second putative FUE was also identified at approximately the correct distance upstream of site 3. This contains GU rich sequences, but does not contain a UUGUA motif. It is tempting to speculate that sites 1 and 2 are under the control of FUE 1 and site 3 is under the control of FUE 2. This might merely reflect the redundancy of plant polyadenylation signals, or could have some functional significance. Montoliu *et al.* (1990) identified two main polyadenylation sites used in the α-1 tubulin mRNA of maize which were approximately 70 bp apart. On analysis, the published sequence of the 3′ UTR appeared to have features in common with that of the pea MTase mRNAs. Two putative FUEs were identified, both of which contained GU rich elements. In addition, the upstream FUE contained a UUGUA motif whereas the downstream FUE did not. Interestingly, Montoliu and co-workers found that a higher proportion of the smaller species was present in actively dividing tissues, whereas the longer species was more evenly distributed amongst different tissues. They suggested that the smaller species may be more stable, giving rise to an increase in steady-state levels in rapidly dividing tissues. Though the levels of pea MTase mRNA appear to be tightly controlled at the transcriptional level (see next section), it would be interesting to test whether there is differential distribution of the differently polyadenylated forms of the MTase mRNA in a tissue or developmental stage-specific manner. This could be achieved by quantitative RT-PCR, using oligos specific for each type of MTase mRNA.

Figure 3.7.9 Putative polyadenylation signals in pea MTase mRNA

The cloned 3′ RACE products (Pfu 1, 2 and 3) were analysed for putative polyadenylation signals. The 3′ end of the longest clone (Pfu 3) is shown and the polyadenylation sites of clones Pfu 1 and Pfu 2 are indicated by arrows (cleavage probably occurs on the right hand side of the first adenosine of the poly(A) tail).

Possible near upstream elements (NUEs), which correspond to the canonical AAUAAA motif in the first 3 nucleotides are boxed. Two putative far upstream elements (FUEs) are also indicated. It is postulated that polyadenylation sites used in Pfu 1 and Pfu 2 are under the control of FUE 1, and that used in Pfu 3 is under the control of FUE 2. The individual components of the putative FUEs are indicated; the sequence UUGUA is in bold and underlined and UG rich sequences are just underlined.


GCGCCAAUUAGAGGAUUAGGGCGCAUCUUUCAAAAAGCAUCUUUUUAUCAUAUAGUUUUG
-10
50
A N *
UCUUUCAGUGUUCUGGAAACAACCCAACCCUUGUAUAUAGUUGUUUUCUUGGCUAUUUUU
51
110
Putative FUE 1
CUUAGUUUAAUCAAUUCUUUGUUUAAAAGGAUUGAUGGAAUGGAUUAUGCUAUAAAACUC
111
170
Pfu 1 poly(A)
Pfu 2 poly(A)
AUUUUUUCUAUCAAUUUGGGGUUUCUCUUGGUUAGGUGAUAAUUUUGCUGUUACUUUAUG
171
230
Putative FUE 2
Pfu 3 poly(A)
AAAGAAAUGAUUGGAAUAUUACCUUCCAAUGUAGUUCAAUUCGACC
231
276

3.8 Steady state levels of pea MTase mRNA in different tissues

The levels of MTase mRNA at different stages of seedling growth and in different structures was measured to see if the changes in MTase activity during seedling growth observed by Yesufu *et al.* (1991) are paralleled by changes in the levels of MTase mRNA. Total RNA was prepared from various structures (apex, stem, leaf and root) and at different stages of development (counted in days after imbibition). Sections of 1 cm in length were taken from the shoot tips (= shoot apex, which includes the immature leaves) and the radical (= root apex). The shoot sub-terminal portion was taken from the second cm down from the apical portion. The leaf portions were fully differentiated leaves from 7 day old seedlings. Because there was slight variation in the stage of growth between the seedlings, the seeding tray was divided into squares for each days sampling and all the seedlings from each square were used, so that the average stage of growth for a particular day was represented. A northern blot was prepared with 30 µg samples of each RNA preparation, and was probed at high stringency with the whole insert isolated from clone Pfu 3 (Fig. 3.7.6) by a *Sal*I/*Bam*HI double digest. As a control, 0.1 pg of the insert from Pfu 3 was run on an agarose TAE gel, which was blotted and fixed in the same manner as the northern. This blot was included in the hybridisation and exposed to the phosphoimager plate together with the northern. The results of this experiment are shown in figure 3.8.1. In each case, only one band of 5.0 kb hybridised with the probe. The intensities of the hybridised band in each lane were measured using MacBAS; comparison with the control target (0.1 pg Pfu 3 insert) enabled the number of transcripts/µg total RNA to be calculated.

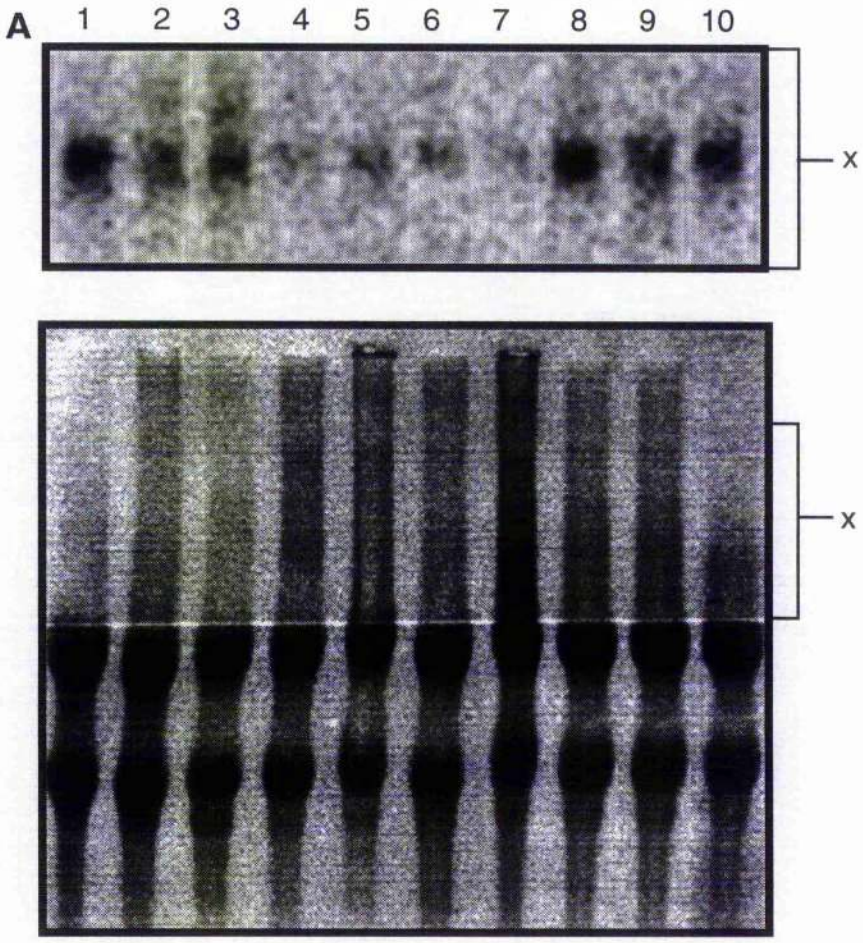
From the results it is clear that the concentrations of MTase mRNA are at their highest in actively dividing tissues, such as root and shoot apical meristems and immature leaves. This is largely consistent with the specific activities of pea MTase found by Yesufu *et al.* (1991), except that relatively low specific activities of MTase

were recovered from root tips. This however, is probably due to difficulties in enzyme extraction from this tissue, which may be because of the high polysaccharides levels in root tips. The MTase specific activities and MTase mRNA concentrations are lower in structures such as mature leaf and stem (the sub terminal portion, lane 4). This is reflected in the fact that growth in these structures is due to cell elongation rather than cell division. High specific activities of MTase are expected in rapidly dividing tissues in pea because the main substrate (hemimethylated DNA) is formed only in the S-phase of the cell cycle; DNA MTase extracted from 5–7 day old pea shoot tips exhibits a 20–60 fold preference for hemimethylated substrates over unmethylated substrates (Houlston *et al.*, 1993).

These results could indicate that pea MTase expression is regulated at the level of transcription. However, the possibility that the regulation is all or partly at the level of mRNA stability cannot be ruled out, and there may be additional regulation at the translational level. As discussed in section 3.7.6, there are at least 3 differently polyadenylated species of MTase mRNA in pea, which appear as a doublet upon gel electrophoresis of 3′ RACE products. There appear to be equal amounts of the longer and shorter species, though after 30 rounds of amplification the results will probably not be strictly quantitative. The 3′ RACE experiments were carried out on poly(A)$^+$ RNA from 5 day old shoot tips (the top two cm) and, as discussed earlier, there may be a differential distribution of differently polyadenylated mRNA species. Unfortunately, as these differences are small and the size of the mRNA is so large, the degree of resolution of the northern blot (Fig. 3.8.1) does not permit distinction between mRNA species. As mentioned in section 3.7.6, the only feasible way to investigate the distribution of the differently polyadenylated species would be a to use quantitative RT-PCR.

Figure 3.8.1 Variation in pea MTase mRNA levels with location and age

Total RNA was isolated (section 2.2.13.1) from the structures and at the times (in days after swelling, section 2.2.1) stated below the figure. 30 µg of each RNA sample was resolved by denaturing agarose gel electrophoresis (section 2.2.13.3). The gel was stained with ethidium bromide (section 2.2.13.3) to confirm equal loading. The gel was then blotted (section 2.2.14.3) and probed at high stringency (section 2.2.16.1) with the whole insert isolated from clone Pfu 3 (Fig.3.7.6) by *Sal*I/*Bam*HI double digestion. The blot was exposed for 4 days on a phosphoimager plate. The figure shows the autoradiograph and below it, the ethidium-stained gel (shown in negative). A band of 5.0 kb hybridises in each lane, but with varying intensity.



Material used for RNA isolation			Level of MTase message (transcripts/µg total RNA; x10⁴)
	Location	Age (days)	
Lane 1:	shoot apex	4	46
2:	"	5	28
3:	"	6	33
4:	shoot (sub-terminal)	6	14
5:	shoot apex	7	19
6:	"	9	13
7:	leaf	7	9
8:	radical apex	2	42
9:	"	3	31
10:	"	5	31

3.9 Is there a family of related MTase genes in pea?

3.9.1 Introduction

Finnegan and Dennis (1993) have reported evidence for the existence of a family of related MTase genes in *Arabidopsis*. Their evidence was as follows: firstly, they amplified a segment from *Arabidopsis* genomic DNA using degenerate primers corresponding to motifs IX and X of mouse MTase. Only one type of MTase-specific PCR product was identified, which was then used to isolate a clone from a genomic library. The isolated genomic clone contained a region which was identical in nucleotide sequence to the PCR product. An *Eco*RI fragment of the genomic clone was then used to screen a cDNA library. Four clones which overlapped the region between IX and X were isolated and, where sequenced, were found to be identical (except for in the 3′ UTR). However, the original PCR product shared only 86% identity at the nucleotide level with the equivalent region in the assembled cDNA sequence. When genomic blots were probed with the cloned PCR product (87 bp) and a 398 bp fragment of the cDNA clone encompassing regions IX and X, different patterns of hybridising fragments were obtained with the two probes. The bands that hybridised with the PCR probe also appeared when the 398 bp fragment was the probe, but only upon prolonged exposure, or when the stringency was reduced. This presents good evidence for the existence of a second gene which is related to the MTase gene in *Arabidopsis*. However, it has not yet been proven that the gene from which the PCR fragment was amplified is actually expressed (the authors do not mention any attempts at northern analysis with the PCR product, or at amplification of this fragment from cDNA).

Using degenerate primers corresponding to motifs IV and VI, only one type of MTase-specific cDNA PCR product was identified (section 3.6). However, other mRNAs may exist that are sufficiently different in this region for the degenerate

primers not to anneal. To investigate whether other genes related to the isolated MTase cDNA fragment exist in pea, I probed at moderate stringency a Southern blot of pea genomic DNA that had been restricted as described as described below. As a probe, I used the whole insert from clone Pfu 1 (Fig. 3.7.6). If a closely related MTase gene exists, it is likely to be homologous in all the conserved motifs, not just IX and X.

3.9.2 Selection of restriction enzymes

The restriction enzymes used in this experiment were selected on the basis of the following criteria: firstly, the enzymes should have no sites within the probe used. Secondly, the enzymes should not be sensitive to the methylation of cytosines within CG or CNG sites, as one might expect heterogeneity in patterns of methylation of MTase gene(s) in different tissues, which would distort the Southern blot data. The selected enzymes plus their recognition sequences are illustrated in figure 3.9.1.

As discussed in the introduction (section 1.5.3), there is little evidence for the presence of a MTase which methylates the outer cytosine of CCG motifs, or the presence of cytosines methylated at this this position in pea at least (Pradhan and Adams, 1995). Therefore, the only enzyme which may give ambiguous data is *Bgl*II, if the first base following a target sequence happens to be G.

Figure 3.9.1 Enzymes selected for Southern analysis

*Bgl*I	G^{m}CCNNNNNGGC
*Sac*I	GAG^{m}CTC
*Sph*I	GCATGC
*Xba*I	T^{m}CTAGA
*Bam*HI	GGAT^{m}CC
*Bgl*II	AGAT^{m}CT

The restriction sites are shown in the 5′–3′ direction. The enzymes will not cleave their target sequences if the cytosines are methylated (at the 5-position of the cytosine ring) at the positions shown.

These enzymes can be grouped into 4 categories in terms of sensitivity to the presence of 5-mC in their target sequences i.e.:

1. *Sph*I will cleave regardless of the methylation status of cytosines
2. *Sac*I, *Xba*I are sensitive to 5-mC which is not present either a CG or a CNG motif
3. *Bam*HI, *Bgl*I are sensitive to 5-mC which is potentially in a CCG motif (25% chance for each site)
4. *Bgl*II is sensitive to 5-mC which is potentially in a CWG motif (25% chance for each site)

3.9.3 Southern analysis of pea genomic digests

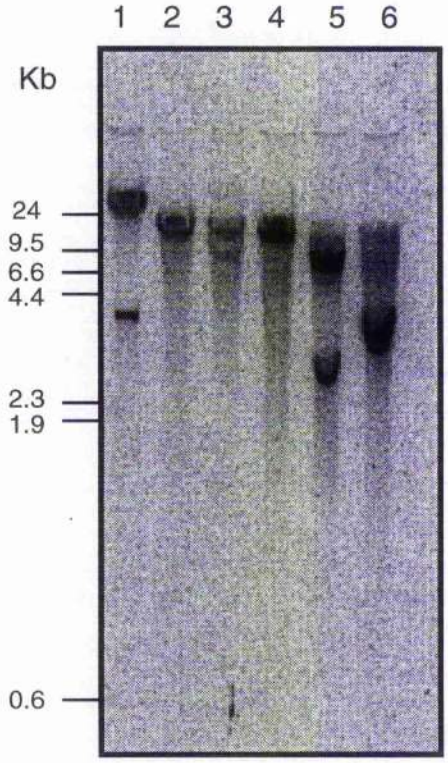
Pea genomic DNA was prepared using the CTAB method (section 2.2.12.2). Digests were carried out using each of the selected restriction enzymes on 40 μg samples of pea genomic DNA (a large amount of DNA was used in order to maximise the signal obtained from weakly hybridising fragments). The digests were resolved on a 0.7% agarose TBE gel and alkali blotted onto a Hybond N^+ filter. The blot was then probed with the whole insert from clone Pfu 1 (Fig. 3.7.6), which was isolated by *Bam*HI/*Sal*I digestion of clone Pfu 1.

Hybridisation was carried out under standard high stringency conditions (section 2.2.16.1), and the filter washed at moderate stringency (1 x SSC, 0.1% SDS at 65°C) for 30 min. Hybridisation under these conditions (65°C, 5 x SSPE, 0% formamide) is ~30°C below the T_m of a perfect hybrid (see section 3.2.2), as the G+C content of the probe is ~40%. The hybridisation used should, in theory, allow up to 10% and possibly 15% mismatch between hybrids, and washing at the stringency mentioned above allows hybrids with up to 17% mismatch to be retained on the filter. Hybridisation using these conditions should, therefore, allow closely related members of a gene family to be identified. Indeed, as mentioned earlier, Finnegan and Dennis found cross hybridisation to occur at high stringency. After hybridisation, the filter was exposed overnight on a phosphoimager plate; the autoradiograph is shown in figure 3.9.2, together with an estimate (against DNA size markers) of the sizes of hybridised bands obtained for each restriction enzyme.

Figure 3.9.2 Southern analysis of pea genomic DNA

40 μg samples of pea genomic DNA, digested with different restriction enzymes (section 2.2.4), were resolved by TBE agarose gel electrophoresis (section 2.2.5), along with marker DNA (*Hin*dIII digest of λ DNA). The gel was alkali blotted (section 2.2.14.1) onto a Hybond N+ filter, which was probed with the whole insert from Pfu 1 (Fig. 3.7.6), which was labelled by random priming (section 2.2.15.1). Hybridisation was carried out in the absence of formamide at 65°C (section 2.2.16.1), and the final post-hybridisation wash was of moderate stringency (1 x SSC, 0.1% SDS at 65°C) for 30 min. The blot was exposed overnight on a phosphoimager plate.

The distance migrated by the size markers is indicated on the left of the figure. The restriction enzymes used, and an estimate of the sizes of hybridising fragments (in kb) for each enzyme are indicated below the figure.



Lane	Enzyme	Fragment size(s) (kb)
1	*Bgl*I	> 24 + 3.6
2	*Sac*I	14
3	*Sph*I	14 + 7
4	*Xba*I	14
5	*Bam*HI	6 + 3
6	*Bgl*II	4

3.9.4 Discussion

The results of this experiment do not appear to indicate the presence of a MTase gene family in pea, at least with respect to the catalytic domain. Unfortunately, there is smearing of some of the bands due to overloading the DNA, and therefore the size estimations of some of the bands are very approximate. Nevertheless, digestion of pea genomic DNA with *Bgl*II clearly yielded only one hybridising fragment, as did digestion with *Sac*I and *Xba*I. The other restriction enzymes (*Bgl*I, *Sph*I and *Bam*HI) yielded two hybridising fragments. As there were no sites for these enzymes within the probe, the presence of two hybridising bands must be due to the presence of a single site for each enzyme within intronic sequences. Furthermore, all hybridising bands were still present after washing at high stringency (0.1 x SSC 0.1% SDS, 65°C). No other hybridising fragments were identified even after exposure for 1 week on the phosphoimager plate (results not shown).

These results contrast with the findings of Finnegan and Dennis (1993), and it therefore seems unlikely that a MTase gene family exists in pea. However, it could be possible that the region of homology exists in the region spanning motifs IX and X alone, and that the signal obtained with cross-hybridising fragments would be too low to be visible because this region represents a small proportion of the probe used. It would seem unlikely for another MTase gene to be closely related only in two of the conserved motifs, but the possibility could be investigated by probing with a smaller restriction fragment of the Pfu 1 insert, which includes , for example, regions IX and X.

Chapter 4 Discussion

4.1 The methyltransferase cDNA of *Pisum sativum*

The 3′ end of the pea MTase cDNA cloned in this project corresponds to an mRNA which is approximately 5.0 kb in length (section 3.6.12). The size of the mouse MTase mRNA is 5.0 kb (Bestor *et al.*, 1988), and the *Arabidopsis* MTase mRNA is 4.7 kb (Finnegan and Dennis, 1993). The steady-state levels of the pea MTase mRNA parallel the specific activities of MTase in pea seedlings from different structures and at different stages of development, being relatively high in rapidly dividing tissues (section 3.8). The control of MTase expression may be at the transcriptional level or at the level of mRNA stability, or it could be a combination of the two. There are at least three different polyadenylation sites utilised in the pea MTase mRNA (section 3.7.6) and a similar observation was made for the *Arabidopsis* MTase mRNA (Finnegan and Dennis, 1993). As discussed in section 3.7.6, this could represent an additional level of control (at the level of mRNA stability) on the steady-state levels of MTase message. This hypothesis, however, remains to be tested.

The predicted amino acid sequence of the cloned pea MTase 3′ RACE product shares 77% identity with *Arabidopsis* MTase and 52% identity with mouse MTase. As discussed in section 3.7.5, the eukaryotic MTases (i.e. from higher plants and mammals) form a closely related group with respect to their catalytic domains, whose members show a greater degree of homology at the amino acid level with each other than with any prokaryotic MTase. Therefore, these enzymes must be ancestrally related. The degree of conservation in the N-teminal domain is much less between plant and mammalian MTases; *Arabidopsis* MTase shares only 24% homology at the amino acid level with the murine enzyme. The predicted *Arabidopsis* MTase protein also lacks the Cys-rich zinc binding domain of the mammalian MTases, and has an acidic region which is absent in the mammalian enzymes.

4.2 Are the different sequence specificities encoded by distinct mRNA species in plants?

As discussed in the introduction, 5-mC is found in CG and CNG sequences in plants, whereas in vertebrates the vast majority of 5-mC is found in CG sequences (Gruenbaum *et al.*, 1981). It has recently been shown that two different specificities (for CG and CWG sequences) reside in different enzymes in pea (Pradhan and Adams, 1995). (There is no direct evidence, from genomic sequencing, for the methylation of the outer cytosine in the CCG motif in plants.). Intuitively, one would expect the two different MTases to be encoded by separate genes or by different mRNAs as a result of alternative splicing.

4.2.1 Are there two different MTase genes in plants?

As described in section 3.9, evidence was found for the existence of another gene which is related to the MTase gene in *Arabidopsis*, and whose cDNA has not yet been cloned (Finnegan and Dennis, 1993). Only the sequence spanning conserved motifs IX and X of this second gene was published. Comparison with the pea MTase clone in this region reveals it to be more homologous to the *Arabidopsis* MTase cDNA than to the other *Arabidopsis* gene in this region. It still remains to be proven that this second gene encodes the other conserved motifs, which would be required for it to be a *bone fide* DNA MTase gene, and it cannot be ruled out that this gene may have a function unrelated to DNA methylation. This is exemplified by the *Pmt1*$^+$ gene of the fission yeast, *S. pombe* (Wilkinson *et al.*, 1995; Pinarbarsi *et al.*, personal communication), which encodes a protein that contains all the conserved amino acid motifs present in prokaryotic type II C5-cytosine MTases (except for a mutated motif IV), yet is not a functional DNA MTase. However, the predicted amino acid sequence of the second gene in *Arabidopsis* identified by Finnegan and Dennis (1993) does resemble the eukaryotic MTases more than any prokaryotic

MTase (and more than the *Pmt1*$^+$ protein), and must therefore have diverged relatively recently. It is striking that no cDNA clone corresponding to this gene was identified upon library screening; this may be either because it is not expressed in the material used for library construction, or because the mRNA is expressed at low levels relative to the MTase mRNA that was cloned.

The work presented in this thesis gives evidence for the existence of only one type of MTase mRNA. Firstly, when degenerate primers corresponding to conserved motifs IV and VI were used to amplify pea cDNA, only one type of MTase-specific PCR product was identified, which comprised 11 clones out the 18 independent cloned PCR products that were sequenced (section 3.6). It is of course possible that the degenerate primers used (oligos IVa and VIc, which were designed according to the sequences GPPCQ and ENV(R/K)(N/T) respectively) would not be able to amplify the other MTase mRNA (if it exists), because of the effect of mismatches. Although up to three primer-template mismatches were found to be tolerated, they were all G-T mismatches (Fig. 3.6.6). Similarly, Finnegan and Dennis (1993) were able to amplify by PCR using degenerate primers, only one type of segment of *Arabidopsis* genomic DNA which was homologous to a MTase, and this segment derived from the second putative MTase gene (not the one corresponding to the MTase cDNA which was subsequently cloned). The amino acid sequences used for primer design were QGFPD and QVGNAV of mouse motifs IX and X respectively. The failure to amplify a segment of the gene which corresponded to the MTase cDNA (which contains the sequence QIGNAV) was attributed to the presence of a C-A primer-template mismatch 4 nucleotides from the 3′ end of the downstream primer. However, the degenerate primers used in this thesis agree much better with the prokaryotic consensus, especially in their 3′ ends (where the presence of mismatches is more deleterious), than do the primers used by Finnegan and Dennis (1993) (see Fig. 3.7.8). Therefore, one would expect these primers to amplify a segment from any cDNA from pea,which corresponds to a functional MTase mRNA that is closely related to the eukaryotic MTases. However, it is not possible to predict

with certainty the positions where mutations may occur outside the absolutely conserved amino acids, and the effect these may have on the efficiency of PCR amplification.

The second line of evidence supporting the existence of only one type of MTase mRNA came from ~~came from~~ hybridisation to northern blots of pea poly(A)$^+$ RNA (section 3.6.12), followed by washing at reduced stringency. Using the cloned pea MTase-specific PCR product as a probe (which would be expected to share a high level of identity with a related MTase mRNA), only one hybridising band (of 5.0 kb in length) was identified. The hybridisation conditions used should permit the detection of closely related mRNA species (see section 3.9). Of course, it could be possible that the band represents two closely related mRNA species which are superimposed; the degree of resolution does not permit the alternatively polyadenylated forms of the characterised mRNA which are known to exist (section 3.7.6) to be distinguished. A pea MTase cDNA was cloned independently by S. Pradan, working in our laboratory. This clone was isolated partly on the basis of N-terminal sequencing of the purified CWG MTase. The 3′ end of ~~the~~ this cDNA was cloned via 3′ RACE, using a unique primer designed according to part of the sequence of this clone (section 3.7). Seven independent clones were mapped and three were sequenced. The nucleotide sequence between motifs IV and VI of the cloned cloned 3′ RACE products was found to be identical with that of the MTase-specific PCR product described in section 3.6 (see Fig. 3.6.6), and therefore both are derived from the same MTase mRNA.

To establish whether a family of MTase genes exists in pea, Southern blots of pea genomic DNA were probed at moderate stringency with the insert from one of the cloned 3′ RACE products (section 3.9). This fragment encodes all the conserved motifs from IV to X, and therefore it was reasoned that if another, closely related gene exists (as appears to be the case in *Arabidopsis*), it would be detected by the probe. However, the results of Southern analysis were consistent with the presence of one type of MTase gene. Thus, the presence of more than one kind of MTase gene

does not appear to be an essential feature of higher plants.

4.2.2 Alternative splicing events in pea MTase mRNA

As discussed in section 1.4.1, the sequence specificity of prokaryotic MTases is defined by the variable domain, which lies between motifs VIII and IX. The obvious hypothesis was that a single MTase gene could give rise to mRNA species that are differently spliced in the region specifying the variable domain. As mentioned earlier, the 3′ end of the pea MTase cDNA (which contained the conserved motifs IV to X and therefore spanned the variable domain) was cloned via a 3′ RACE strategy (section 3.7). The amplified product (excluding the truncated 3′ RACE artefacts, which are too small to be full length 3′ RACE products) showed size heterogeneity which was visible as a doublet upon gel electrophoresis. The possibility that some of this size heterogeneity could be as a result of alternative splicing was investigated. However, the size heterogeneity observed was found to be due to the use of different polyadenylation sites in the mRNA (see section 3.7.4).

The analogy with the target recoginition domains of prokaryotic MTases is perhaps inappropriate for eukaryotic MTases, all of which that have been cloned so far code for a large N-terminal domain which is linked to the 'prokaryotic' catalytic domain. It has been shown that cleavage of the N-terminal domain of mouse MTase results in a stimulation of the rate of de novo methylation *in vitro*, whereas there is little effect on the rates of maintenance methylation (Bestor, 1992) (section 1.5.2). It was suggested by the author that proteolytic cleavage *in vivo* could be responsible for the ectopic methylation which is associated with some tumours and in aging animals As the N-terminal domain is capable of modifying the specificity for unmethylated CG sites in mouse MTase, perhaps the specificity for CG as opposed to CWG sites in plant MTases is also affected by the N-terminal domain; this is discussed below.

4.3 How can one MTase mRNA give rise to both CG- and CWG-specific MTases?

As mentioned earlier, the CG and CWG specificities do reside in different proteins in pea (Pradhan and Adams, 1995). The purified CWG MTase activity corresponds to a 110 kDa protein plus a 100 kDa protein, which is considered to be due to partial proteolysis of the former protein. The CG MTase activity corresponds to two proteins of 140 and 55 kDa. Again, it was considered that the 140 kDa protein represented the intact enzyme. Sequencing of the region specifying the N-terminal domain of the pea MTase cDNA which was cloned by S. Pradhan revealed the putative translational start-site to encode a protein of at least 180 kDa. The N-terminal domain also encodes a Cys-rich region which may bind zinc ions, as in mouse MTase. However, the homology between the plant and ~~and~~ mammalian MTases is much less in the N-terminal domain than the C-terminal domain. Interestingly, the Cys-rich sequence was found to be absent in the *Arabidopsis* MTase cDNA. The N-terminal sequence of the pea MTase cDNA also contains several regions which could act as signals for proteolytic cleavage *in vivo*, two of which would give rise to proteins which correspond to the 140 and 110 kDa MTases. These hypothetical cleavage sites would give rise to proteins which contain putative nuclear localisation signals near their N-terminus. This would explain how proteolytic processing in the cytoplasm could give rise to different MTase enzymes which are able to function in the nucleus.

In summary, the specificity of plant MTases could be dictated by *in vivo* proteolytic cleavage of a single translation product, which removes portions of the N-terminus which could modify the sequence specificity of the enzyme. The sequence specificity of eukaryotic MTases is more relaxed than their prokaryotic counterparts; it has been shown that mouse MTase methylates CWG sequences to a small extent *in vitro* (Adams *et al.*, 1993), although whether this is associated with

proteolytic cleavage at specific sites is not known. If the N-terminal domain is able to modulate the sequence specificity of plant MTases, what is the mechanism of this modulation? As discussed earlier (sections 1.5.2 and 3.7.5), the variable regions of the eukaryotic MTases that have been cloned are very large in comparison with those of the majority of mono-specific prokaryotic MTases. It is possible that the variable domains contain regions which interact in some way with the N-terminal domain in a way that dictates the sequence specificity. However, until the structure of a eukaryotic MTase interacting with DNA has been solved, the nature of the involvement of the N-terminal domain remains speculative.

The sequence specificity of eukaryotic MTases perhaps cannot be considered in isolation. As discussed in section 1.5.1, methylation at non-symmetrical sequences occurs in mammals (Tasheva and Roufa, 1994), in plant transgenes (Meyer *et al.*, 1994), and also in 'ripped' sequences in *Neurospora crassa* (Selker *et al.*,1993). It has been proposed that eukaryotic MTases are able to recognise and methylate particular regions of chromatin structure, perhaps via interaction with unknown accessory proteins, or perhaps as a consequence of accessibility to a MTase with relaxed specificity. In addition, recognition of DNA-RNA triple helixes by DNA MTases has also been suggested (Wassenegger *et al.*, 1994). The elucidation of these interactions represents a major challenge in this field.

References

Adams, C.C. and Workman, J.L. (1993) *Cell* **72**, 305–308

Adams, R.L.P., Burdon, R.H., McKinnon, K. and Rinaldi, A. (1983) *FEBS Lett.* **163**, 194–198

Adams, R.L.P., Davis, T., Fulton, J., Kirk, D., Qureshi, M. and Burdon, R.H. (1984) *Curr. Top. Microbiol. and Immunol.* **108**, 143–156

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

Adams, R.L.P., Gardiner, K., Rinaldi, A., Bryans, M., McGarvey, M. and Burdon, R.H. (1986) *Biochim. Biophys. Acta* **868**, 9–16

Adams, R.L.P., Lindsay, H., Reale, A., Seivwright, S., Kass, S., Cummings, M., Houlston, C. (1993) In: *DNA Methylation: Molecular Biology and Biological Significance*, J.-P. Jost and H.P. Saluz (ed.), pp. 120–144, Birkhauser Verlag, Basel

Adams, R.L.P.(1995) *BioEssays* **17**, 139–145

Anderson, M.L.M. and Young, B.D. (1985) In: *Nucleic acid hybridisation: A practical approach*, B.D. Hames and S.J. Higgins (ed.), pp. 47–111, IRL press, Oxford

Antequera, F. and Bird, A.P. (1988) *EMBO J.* **7**, 2295–2299

Antequera, F., Macleod, D. and Bird, A.P. (1989) *Cell* **58**, 509–517

Antequera, F., Boyes, J. and Bird, A.P. (1990) *Cell* **62**, 503–514

Antequera, F. and Bird, A.P. (1993) In: *DNA Methylation: Molecular Biology and Biological Significance*, J.-P. Jost and H.P. Saluz (ed.), pp. 169–185, Birkhauser Verlag, Basel

Ariel, M., Robinson, E., McCarrey, J.R. and Cedar, H. (1995) *Nature Genetics* **9**, 312–315

Atkinson, T. and Smith, M. (1984) In: *Oligonucleotide Synthesis, a Practical Approach*, pp. 35–81, M.J. Gait (ed.), I.R.L. Press

Ball, D.J., Gross, D.S. and Garrard, W.T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5490–5494

Barlow, D.P., Stoger, R., Hermann, B.G., Saito, K. and Schweifer, N. (1991) *Nature* **349**, 84–87

Bednarik, D.P., Duckett, C., Kim, S.U., Perez, V.L., Griffis, K., Guethner, P.C. and Folks, T.M. (1991) *New Biol.* **3**, 969–976

Belanger, F.C. and Hepburn, A.G. (1990) *J. Mol. Evol.* **30**, 26–35

Bell, M.V., Hirst, M.C., Nakahori, Y., MacKinnon, R.N., Roche, A., Flint, T.J., Jacobs, P.A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R.H., Winter, R., Pembrey, M., Thibodeau, S. and Davies, K.E. (1991) *Cell* **64**, 861–866

Ben Hattar, J., Beard, P. and Jiricny, J. (1989) *Nucl. Acids Res.* **17**, 10179–10190

Bestor, T.H. and Ingram, V.M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5559–5563

Bestor., T.H., and Ingram, V.M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2674–2678

Bestor, T.H., Laudano, A., Mattalanio, R. and Ingram, V. (1988) *J. Mol. Biol.* **203**, 971–983

Bestor, T. (1990) *Phil. Trans. R. Soc. Lond. B* **326**, 179–187

Bestor, T. (1992) *EMBO J.* **11**, 2611–2617

Bestor, T.H. and Verdine, G.L. (1994) *Curr. Opin. Cell Biol.* **6**, 380–389

Bianchi, M.W. and Viotti, A. (1988) *Plant. Mol. Biol.* **11**, 203–214

Bird, A.P., Taggart, M.H. and Smith, B.A. (1979) *Cell* **17**, 889–901

Bird, A.P. (1986) *Nature* **321**, 209–213

Bird, A. (1992) *Cell* **70**, 5–8

Blondel, A. and Thillet, J. (1990). *Nucl. Acids Res.* **19**, 181

Bolden, A., Ward, C., Siedlecki, J.A. and Weissbach, A. (1984) *J. Biol. Chem.* **259**, 12437–12443

Bolton, E.T. and McCarthy, B.J. (1962) *Proc. Natl. Acad. Sci. USA* **48**, 1390–1395

Bonner, T.I., Brenner, D.J., Neufield, B.R. and Britten, R.J. (1973) *J. Mol. Biol.* **81**, 123–135

Boyes, J. and Bird, A. (1991) *Cell* **64**, 1123–1134

Boyes, J. and Bird, A. (1992) *EMBO J.* **11**, 327–333

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and Cedar, H. (1994) *Nature* **371**, 435–438

Brinboim, C. and Doly, J. (1979) *Nucl. Acids Res.* **7**, 1513–1523

Brown, C.J., Lafreniere, R.G., Powers, V.E., Sebastio, G., Ballabio, A., Pettigrew, A.L., Ledbetter, D.H., Levy, E., Craig, I.W. and Willard, H.F. (1991a) *Nature* **349**, 82–84

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R. and Willard, F. (1991b) *Nature* **349**, 38–44

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J. and Willard, H.F. (1992) *Cell* **71**, 527–542

Bryans, M., Kass, S., Seivwright, C. and Adams, R.L.P. (1992) *FEBS Lett.* **309**, 97–102

Buhk, H.-J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Gunthert, U., Noyer-Weidener, M., Jentsch, S. and Trautner, A. (1994) *Gene* **29,** 51–61

Bullock, W.O., Fernandez, J.M., Short, J.M. (1987) *Biotechniques* **5**, 376

Burdon, R.H. and Adams, R.L.P. (1969) *Biochim. Biophys. Acta* **174**, 322–329

Burdon, R.H. and Adams, R.L.P. (1980) *TIBS* **5**, 294–297

Burn, J.E., Bagnall, P.J., Metzger, J.D., Dennis, E.S. and Peacock, W.J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 287–291

Buschhausen, G., Wittig, B., Graessman, M. and Graessman, A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1177–1181

Carlson, L.L., Page, A.W. and Bestor, T.H. (1992) *Genes Dev.* **6**, 2536–2541

Cedar, H. and Razin, A. (1990) *Biochim. et Biophys. Acta.* **1049**, 1–8

Chen, L., Macmillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry* **30**, 11018–11025

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W. and Roberts, R.J. (1993) *Cell* **74**, 299–307

Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995

Churchill, M. and Suzuki, M. (1989) *EMBO J.* **8**, 4189–4195

Clark, S.J., Harrison, J. and Frommer, M. (1995) *Nature Genetics* **10**, 20–27

Comb, M. and Goodman, H.M. (1990) *Nucl. Acids Res.* **18**, 3975–3982

Compton, T. (1990) In: *PCR Protocols; a Guide to Methods and Applications* (M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White (ed.), pp 39–45, Academic Press Ltd., London

Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) *Nature* **274**, 775–780

Czank, A., Hauselman, R., Page, A.W., Leonhardt, H., Bestor, T.H., Schaffner, W. and Hergersberg, M. (1991) *Gene* **109**, 259–263

Dean, C., Tamaki., S., Dunsmuir, P., Favreau, M. and Katayama, C. (1986) *Nucl. Acids Res.* **14**, 2229–2240

DeChiara, T.M., Robertson, E.J. and Efstratiadis, A. (1991) *Cell* **64**, 849–859

Doerfler, W. (1983) *Annu. Rev. Biochem.* **52**, 93–124

Doerfler, W. (1991) *Biol. Chem. Hoppe-Seyler* **372**, 557–564

Eckert, K.A. and Kunkel, T.A. (1991) In: *Polymerase Chain Reaction I: A Practical Approach*, M.J. McPherson, P. Quirke and G.R. Taylor (eds.) IRL Press, Oxford

Edwards, Y. H. (1990) *Phil. Trans. R. Soc. Lond.* B **326**, 207–215

Ehrlich, K.C. (1992) *Biochim. et Biophys. Acta* **1172**, 108–116

Ehrlich, K.C. (1993) *Preparative Biochem.* **23**, 423–438

Ehrlich, M., Gamma-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A. and Doerfler, W. (1983) *Anal. Biochem.* **135**, 165–171

Ellison, K. and Messing, J. (1988) In: *Plant Biotechnology*, S.D. Kung and C.J. Arntzen (ed.) pp 115–135, Stoneham. Butterworths. Mass.

Englander, E.W., Wolffe, A.P. and Howard, B.H. (1993) *J. Biol. Chem.* **268**, 19565–19573

Erlanson, D.A., Chen, L. and Verdine, G.L. (1993) *J. Am. Chem. Soc.* **115**, 12583–12584

Feinberg, A.R. and Vogelstein, (1983) *Anal. Biochem.* **132**, 6–13

Feinberg, A.R. and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–266

Finnegan, E.J., Brettel, R.I.S. and Dennis, E.S. (1993) In: *DNA Methylation: Molecular Biology and Biological Significance*, J.-P Jost and H.P Saluz (ed.), pp. 218–261, Birkhauser Verlag, Basel

Finnegan, E.J. and Dennis, E.S. (1993) *Nucl. Acids Res.* **21**, 2383–2388

Ford, J.P., Coca-Prados, M. and Hsu, M.-T. (1980) *J. Biol. Chem.* **225**, 7544–7547

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831

Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A. and Cedar, H. (1991) *Nature* **351**, 239–241

Frohman, M.A., (1994) in: *The Polymerase Chain Reaction*, K.B. Mullis, F. Ferre and R.A. Gibbs (ed.) pp. 14–37, Birkhauser, Boston

Gamma-Sosa, M.A., Midgett, R.M., Slagel, V.A., Githens, S., Kuo, K.C., Gehrke, C. and Ehrlich, M. (1983) *Biochim. Biophys. Acta* **740**, 212–219

Gardiner-Garden, M., Sved, J.A. and Frommer, M. (1992) J. Mol. Evol. **34**, 219–230

Giordano, M., Mattachini, M.E., Cella, R. and Pedrali-Noy, G. (1991) *Biochem. Biophys. Res. Commun.* **177**, 711–719

Goldberg, R.B., Barker, S.J. and Perez-Grau, L. (1989) *Cell* **56**, 149–160

Gopal, J., Yebra, M. and Bhagwat, A.S. (1994) *Nucl. Acids Res.* **22**, 4482–4488

Gruenbaum, Y., Cedar, H. and Razin, A. (1982) *Nature* **295**, 620–622

Gruenbaum , Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) *Nature* **292**, 860–862

Gruenbaum, Y., Stein, R., Cedar, H. and Razin, A. (1981)b *FEBS Lett.* **124**, 67–81

Guseinov, V.A., Kiryanov, G.I. and Vanyushin, B.F. (1975) *Mol. Biol. Rep.* **2**, 59–63

Hadfield, K.A., Dandekar, A.M. and Romani, R.J. (1993) *Plant Science* **92**, 13–18

Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580

Harrington, M.A., Jones, P.A., Imagawa, M., and Karin, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2066–2070

Hebbes, T.R., Thorne, A.W., and Crane-Robinson, C. (1988) *EMBO J.* **7**, 1395–1402

Hermann, R. and Doerfler, W. (1991) *FEBS Lett.* **291**, 191–195

Hershkovitz, M., Gruenbaum, Y., Zakai, N. and Loyter, A. (1989) *FEBS Lett.* **253**, 167–172

Hershkovitz, M., Gruenbaum, Y., Renbaum, P. and Loyter, A.(1990) *Gene* **94**, 189–193

Heinikoff, S. and Dreesden, T.D. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6704–6708

Higurashi, M. and Cole, R.D. (1991) *J. Biol. Chem.* **226**, 8619–8625

Hobbs, S.L.A., Kpodar, P. and DeLong, C.M.O. (1990) *Plant Mol. Biol.* **15**, 851–864

Holliday, R. and Pugh, J. (1975) *Science* **187**, 226–232

Holliday, R. (1987) *Science* **238**, 163–170

Hotchkiss, R.D. (1948) *J. Biol. Chem.* **168**, 315–332

Houlston, C.E., Cummings, M., Lindsay, H., Pradhan, S. and Adams, R.L.P. (1993) *Biochem J.* **293**, 617–624

Howlett, S.K. and Reik, W. (1991) *Development* **113**, 119–127

Howley, P.M., Israel, M.A., Law, M.F., Martin, M.A. (1979) *J. Biol. Chem.* **254**, 4876–4883

Hubrich-Khuner, K., Buhk, H.J., Wagner, H., Kroger, H. and Simon, D. (1989) *Biochem. Biophys. Res. Commun.* **160**, 1175–1182

Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F. and Chua N.-H. (1987) *Plant Mol. Biol.* **15**, 9627–9640

Hyman, L.E., Seiter, S.H., Whoriskey, J., Moore, C.L. (1991) *Mol. Cell. Biol.* **11**, 2004–2012

Iguchi-Ariga, S.M.M. and Schaffner, W. (1989). *Genes Dev.* **3**, 612–619.

Inamdar, N.M., Ehrlich, K.C., and Ehrlich, M. (1991) *Plant Mol. Biol.* **17**, 111–123

Ingrosso, D., Fowler, A.V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.* **264**, 20130–20139

Jablonka, E., Goiten, R., Marcus, M. and Cedar, H. (1985) *Chromosoma* **93**, 152–156

Jacobs, K.A. Rudersdorf, R.,. Neill, S.D, Dougherty, J.P., Brown, E.L. and Fritsch, E.F. (1988) *Nucl. Acids Res.* **16**, 4637–4641

Jahner, D. and Jaenisch, R. (1984) In: *DNA Methylation: Biochemistry and Biological Significance,* A. Razin, H. Cedar and A.D. Riggs (ed.), pp. 165–187,Springer-Verlag, New York

Jahner, D. and Jaenisch, R. (1985) *Mol. Cell. Biol.* **5**, 2212–2220

Jendrisak, J., Young, R.A., Ingel, J.D. (1987) *Methods Enzymol.* **152**, 359–371

John, M.C. and Amasino, R.M. (1989) *Mol. Cell. Biol.* **9**, 4298–4303

Johnson, C., Goddard, J.P. and Adams, R.L.P. *Biochem J.* **305**, 791–798

Jones, P.A. (1984) In: *DNA Methylation: Biochemistry and Biological Significance.* A. Razin, H. Cedar and A.D. Riggs (ed.), pp. 165–187, Springer-Verlag, New York

Jones, P.A. and Taylor, S.M. (1980) *Cell* **20**, 85–93

Joshi, C.P. (1987) *Nucl. Acids Res.* **15**, 9627–9640

Jost, J.-P. and Hofsteenge, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9499–9503

Jost, J.-P. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4684–4688

Jost, J.-P. and Saluz, H.-P. (1993) In: *DNA methylation: Molecular Biology and Biological Significance* J.-P. Jost and H.-P Saluz (ed.), pp. 425–451, Birkhauser Verlag, Basel

Juttermann, R., Li, E. and Jaenisch, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11791–11801

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H. and Razin, A. (1992) *Genes Dev.* **6**, 705–714

Kakutani, T., Jeddeloh, J.A. and Richards, E.J. (1995) *Nucl. Acids Res.* **23**, 130–137

Karreman, C. and de Waard, A. (1990) *J. Bacteriol.* **172**, 266–272

Kass, S.U., Goddard, J.P. and Adams, R.L.P. (1993) *Mol. Cell. Biol.* **13**, 7372–7379

Kerem, B.-S., Goiten, R., Richler, C., Marcus, M. and Cedar, H. (1983) **304**, 88–90

Keshet, L., Lieman-Hurwitz, J. and Cedar, H. (1986) *Cell* **44**, 535–543

Klaas, M., John, M.C., Crowell, D.N. and Amasino, R.M. (1989) *Plant Mol. Biol.* **12**, 413–423

Klimasauskas, S., Nelson, J.L. and Roberts, R.J. (1991) *Nucl. Acids Res.* **19**, 6183–6190

Knoth, K., Roberds, S., Poteet, C. and Tamkun, M. (1988) *Nucl. Acids Res.* **16**, 10932

Kolsto, A.B., Kollias, G., Giguere, V., Isobe, K.I., Prydz, H. and Grosveld, F. (1986) *Nucl. Acids. Res.* **14**, 9667–9678

Kricker, M.C., Drake, J.W. and Radman, M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1075–1079

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. and Wilson, G. (1994) *Nucl. Acids Res.* **22**, 1–10

Laemmli, U.K. (1970) *Nature* **227**, 680–685

Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12

Langdale, J.A. and Tailor, W.C. and Nelson, T. (1991) *Mol. Gen. Genet.* **225**, 49–55

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992) *Genomics* **13**, 1095–1107

Lauster, R., Trautner, T.A., and Noyer-Weidener, M. (1989) *J. Mol. Biol.* **206**, 313–321

Lee, C. C. and Caskey, C.T. (1990) In: *PCR Protocols; a Guide to Methods and Applications* (M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White (ed.), pp 46–53, Academic Press Ltd., London

Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A. and Tilghman, S.M. (1995) *Nature* **375**, 34–39

Leonhardt, H., Page, A.W., Weier, H.-U. and Bestor, T.H. (1992) *Cell* **71**, 865–873

Levine, A., Cantoni, G.L.and Razin, A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6515–6518

Levine, A., Yeivin, A., Ben-Asher, E., Aloni, Y and Razin, A. (1993) *J. Biol. Chem.* **268**, 21754–21759

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) *Cell* **69**, 905–914

Li, E., Bestor, T.H. and Jaenisch, R. (1992) *Cell* **69**, 915–926

Li, E., Beard, C. and Jaenisch, R. (1993) *Nature* **366**, 362–365

Linn, F., Heidmann, I., Saedler, H. and Meyer, P. (1990) *Mol. Gen. Genet.* **222**, 329–336

Lock, L.F., Melton, D.W., Caskey, C.T. and Martin, G.R. (1986) *Mol. Cell. Biol.* **6**, 914–924

Lock, L.F., Tagaki, N, and Martin, G. (1987) *Cell* **48**, 39–46

Lundberg, K.S., Shoemaker, D.D., Adams, M.W.W., Short, J.M., Sorge, J.A. and Mathur, E.J.(1991) *Gene* **108**, 1–6

Ma, Q., Alder, H., Nelson, K.K., Chatterjee, D., Gu, Y., Nakamura, T., Canaani, E., Croce, C.M., Siracusa, L.D. and Buchberg, A.M. (1993) *Proc Natl. Acad. Sci. USA* **90**, 6350–6354

Mack, D. and Sninsky (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6977–6981

Magill, J.M.and Magill, C.W. (1989) *Dev. Genet. (N.Y.)* **10**,63–69

Martin, F.H., Castro, M.M., Aboul-ela, F. and Tinoco, I. (1985) *Nucl. Acids Res.* **13**, 8927–8930

Mason, P.J. and Williams, J.G. (1985) In: *Nucleic acid hybridisation: A practical approach*, B.D. Hames and S.J. Higgins (eds.), pp. 113–137, IRL press, Oxford

Matzke, M.A., Primig, M., Trnovsky, J. and Matzke, A.J.M. (1989) *EMBO J.* **8**, 643–649

Matzke, M.A. and Matzke, A.J.M. (1990) *Devel. Genet.* **11**, 214–223

Matzke, M.A. and Matzke, A.J.M. (1991) *Plant Mol. Biol.*, **16**, 821–830

Matzke, M., Matzke, A.J.M. and Mittelsten Scheid, O. (1994) In: *Homologous recombination and gene silencing in plants* J. Paszkowski (ed.), pp. 271–307, Kulwer Academic Publishers, Dordrecht

McClelland, M. (1983) *J. Mol. Evol.* **19**, 346–354

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A. (1989) *Cell* **58**, 499–507

Meehan, R.R., Lewis, J.D. and Bird, A.P.(1992) *Nucl. Acids Res.* **20**, 5085–5092

Messeguer, R., Ganal, M.W., Steffens, J.C. and Tanskley, S.D. (1991) *Plant Mol. Biol.* **16**, 753–770

Meyer, P., Heidmann, I. and Neidenhof, I (1993) *Plant. J.* **4**, 89–100

Meyer, P., Niedenhof, I., and ten Lohuis, M. (1994) *EMBO J.* **13**, 2084–2088

Mi, S. and Roberts, R.J. (1993) *Nucl. Acids Res.* **21**, 2459–2464

Miller, O.J., Schnedl, W., Allen, J. and Erlanger, B.F. (1974) *Nature* **251**, 636–637

Mogen, B.D., MacDonald, M.H., Leggewie, G. and Hunt, A.G. (1992) *Mol. Cell. Biol.* **12**, 5406–5414

Molloy, P.L. and Watt, F. (1990) *Phil. Trans R. Soc. Lond.* B. **326**, 267–275

Monk, M., Boubelik, M. and Lehnert, S. (1987) *Development* **99**, 371–382

Monk, M., Adams, R.L.P. and Rinaldi, A. (1991) *Development* **112**, 189–192

Montoliu, L., Rigau, J. and Puigdomenech, P. (1990) *FEBS Lett.* **277**, 29–32

Napoli, C, Lemieux, C. and Jorgensen, R. (1990) *Plant Cell* **2**, 279–289

Nedderman, P. and Jiricny, J. (1993) *J. Biol. Chem.* **268**, 21218–21224

Ngernprasitsiri, J., Chollet, R., Kobayashi, H., Sugiama, T. and Akazawa, T. (1989) J. *Biol. Chem.* **264**, 8241–8248

Nick, H., Bowen, B., Ferl, R.J. and Gilbert, W. (1986) *Nature* **319**, 243–236

Noyer-Weidener, M. and Trautner, T.A. (1993) In: *DNA Methylation: Molecular Biology and Biological Significance*, J.-P. Jost and H.P. Saluz (ed.), pp. 39–108, Birkauser Verlag, Basel

Paroush, Z., Keshet, I., Yisraeli, J. and Cedar, H. (1990) *Cell* **63**, 1229–1237

Pawlak, A., Bryans, M. and Jost, J.-P. (1991) *Nucl. Acids Res.* **19**, 1029–1034

Peerbolte, R., Leenhouts, K., Hooykaas-van Slogteren, G.M.S., Wullems, G.J. and Schilperoot, R.A. (1986) *Plant Mol. Biol.* 7, 285–299

Pfeifer, G.P. and Riggs, A.D. (1991) *Genes Dev.* **5**, 1102–1113

Pfeifer, G.P., Steigerwald, S.D., Hansen, R.S., Gartler, S.M. and Riggs, A.D. (1990a) *Proc. Natl. Acad. Sci. USA* **87**, 8252–8256

Pfeifer, G.P., Tanguay, R.L., Steigerwald, S.D. and Riggs, A.D. (1990b) *Genes Dev.* **4**, 1277–1287

Pfeifer, G.P., Drouin, R., Riggs, A.D. and Holmquist, G.P. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1374–1378

Pfeifer, G.P., Drouin, R., Riggs, A.D.and Holmquist, G.P. (1992) *Mol. Cell. Biol.* **12**, 1798–1804

Pieretti, M., Zhang, F., Fu, Y., Warren, S., Oostra, B., Caskey, C.T. and Nelson, D. (1991) *Cell* **66**, 817–822

Pinabarsi, E., Elliot, J. and Hornby, D.P. (personal communication)

Pollack, Y., Kasir, J., Shemer, T., Metzger, S.and Szyf, M. (1984) *Nucl. Acids Res.* **12**, 4811–4824

Posfai, G., Kim, S.C., Szilac, L., Kovaks, A. and Venetianer, P. (1991) *Nucl.Acids Res.* **19**, 4843–4847

Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 2421–2435

Pradhan, S. and Adams, R.L.P. (1995) *Plant Journal* 7, 471–481

Prendergast, G.C, Lawe, D. and Ziff, E.B. (1991) *Cell* **65**, 395–407

Proffitt, J.H., Davie, J.R., Swinton, D. and Hattman, S. (1984) *Mol. Cell. Biol.* **4**, 985–988

Prols, F. and Meyer, P. (1992) *Plant J.* **2**, 465–475

Razin, A. and Cedar, H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2725–2728

Razin, A. and Cedar, H. (1991) *Microbiol. Rev.* **55**, 451–458

Reik, W., Collick, A., Norris, M.L., Barton, S.C. and Surani, M.A. (1987) *Nature* **328**, 248–251

Renbaum, P. and Razin, A. (1992) *FEBS Lett.* **313**, 243–247

Renbaum, P. and Razin, A. (1995) *Gene* **157**, 177–179

Riggs, A.D. (1975) *Cytogenet. Cell Genet.* **14**, 9–25

Riggs, C.D., Voeller, T.A., and Chrispeels, M.J. (1989) *Plant Cell.* **1**, 609–621

Riggs, C.D. and Chrispeels, M.J. (1990) *Plant Mol. Biol.* **14**, 629–632

Rogers, S.O. and Bendich, A.J. (1988) In: *Plant Molecular Biology Manual* , S.B Gelvin, R.A. Schilperoort and D.P.S. Verma (eds.), Kulwer Academic Publishers, Dordrecht. **A6**, 1–10

Romano, N. and Macino, G. (1992). *Mol. Microbiol.* **6**, 3343–3353

Rothnie, H.M., McCurrach, K.J., Glover, L.A. and Hardman, N. (1991) *Nucl.Acids Res.* **19**, 279–286

Rothnie, H.M., Reid, J. and Hohn, T. (1994) *EMBO J.* **13**, 2200–2210

Sakai , T., Toguchida, J., Ohtani, N., Yandell, D, Ranaport, J. and Dryja, T. (1991) *Am. J. Hum. Genet.* **48**, 880–888

Saluz, H.P. and Jost, J.-P. (1989) *Anal. Biochem.* **176**, 201–208

Saluz, H.P. and Jost, J.-P. (1993) In: *DNA Methylation: Molecular Biology and Biological Significance*, J.-P. Jost and H.P. Saluz (ed.), Birkhauser Verlag, Basel. pp 11–26

Sambrook, J. Fritsch, E.F. and Maniatis, T. (1989): *Molecular Cloning. A Laboratory Manual* (second edition), Cold Spring Harbor Laboratory Press

Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* **143**, 161–178

Sapienza, C., Peterson, A.C., Rossant, J. and Balling, R. (1987) *Nature* **328**, 251–254

Sasaki, T., Hansen, R.S. and Gartler, S.M. (1992) *Mol. Cell. Biol.* **12**, 3819–3826

Scharf, S.J. (1990) In: *PCR Protocols; a Guide to Methods and Applications* (M.A. Innis, D.H. Gelfand, J.J. Sniskey and T.J. White (ed.), pp 84–91, Academic Press Ltd., London.

Schek, N., Cooke, C.and Alwine, J.C. (1992) *Mol. Cell. Biol.* **12**, 5386–5393

Scheidt, G., Weber, H., Graessmann, M. and Graessmann A. (1994a) *Nucl. Acids Res.* **22**, 953–958

Scheidt, G., Weber, H., Graessmann, M. and Graessmann A. (1994b) *Nucl. Acids Res.* **22**, 1538

Schwartz, D. and Dennis, E. (1986) *Mol. Gen. Genet.* **205**, 476–482

Selig, S., Ariel, M. and Cedar, H. (1988) *EMBO J.* **7**, 419–426

Selker, E.U. (1990) *A. Rev. Genetics* **24**, 579–613

Selker, E.U., Camberari, E.B., Jensen, B.C.and Haack, K.R. (1987) *Cell* **51**, 741–752

Selker, E.U. , Fritz, D.Y. and Singer, M.J. (1993) *Science* **262**, 1724–1728

Simpson, V.J., Johnson, T.E. and Hammen, R.F. (1986) *Nucl. Acids Res.* **14**, 6711–6719

Singer-Sam, J., Grant, M., LeBon, J.M., Okuyama, K., Chapman, V., Monk, M. and Riggs, A.D. (1990) *Mol. Cell. Biol.* **10**, 4987–4989

Skadsen, R.W. and Knauer, N.S. (1995). *FEBS Lett.* **361**, 220–224

Slightom, J.L. and Quemada, H.D. (1991) In: *Plant Molecular Biology Manual* , S.B Gelvin, R.A. Schilperoort and D.P.S. Verma (eds.), Kulwer Academic Publishers, Dordrecht. **A7**, 1–52

Smith, S.S., Kaplan, B.E., Sowers, L.C. and Newman, E.M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4744–4748

Solage, A. and Cedar, H. (1978) *Biochemistry* **17**, 2934–2838

Sorensen, M.B. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4119–4123

Southern, E.M. (1975) *J. Mol. Biol.* **98**, 503–507

Staiger, D., Kauleen, H. and Schell, J. (1989) *Proc. Natl. Acad. Sci.* USA **86**, 6930–6934

Stephen, D. Jones, C. and Schofield, J.P. (1990) *Nucl. Acids Res.* **18**, 7463–7464

Stoger, R., Kubicka, P., Liu, C.-G., Kafri, T., Razin, A., Cedar, H. and Barlow, D.P. (1993) *Cell* **73**, 61–71

Surani, M.A., Allen, M.D., Barton, S.C., Fundele, R., Howlett, S.K., Norris, M.L. and Reik, W. (1990) *Phil. Trans. R. Soc. Lond.* B **326**, 313–327

Sved, J. and Bird, A.P. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4692–4696

Szynter, L.A., Slatko, B., Moran, L., O'Donnell, K.H. and Okada, S. (1989) *Nucl. Acids Res.*, **15**, 1846–1866

Takagi, N. and Saski, M. (1975) *Nature* **256**, 640–642

Tasheva, E. and Roufa, D.J. (1994) *Mol. Cell. Biol.* **14**, 5636–5644

Tate, P.H. and Bird, A.P. (1993) *Curr. Opin. Gen. Dev.* **3**, 219–225

Taylor, C., Ford, K., Connolly, B.A. and Hornby, D.P. (1993) *Biochem. J.* **291**, 493–504

Tazi, J. and Bird, A.P. (1990) *Cell* **60**, 909–920

Thein, S.L. and Wallace, R.B. (1986). In: *Human genetic diseases: a practical approach*, K.E.Davies (ed.), pp 33–50. IRL Press, Herndon, Virginia.

Theiss, G., Schleicher, R., Schimpff-Weiland, G. and Follman, H. (1987) *Eur. J. Biochem.* **167**, 89–96

Tollefsbol, T.O. and Hutchison, C.A. (1995) J. *Biol. Chem.* **270**, 1–8

Toniolo, D., Martini, G., Migeon, B.R. and Dono, R. (1988) *EMBO J.* **7**, 401–406

Trasler, J.M., Alcivar, A.A., Hake, L. E., Bestor, T.and Hecht, N.B. (1992) *Nucl. Acids Res.* **20**, 2541–2545

Trautner, T.A., Balganesh, T.S. and Pawlek, B. (1988) *Nucl. Acids Res.* **16**, 6649–6657

Tremblay, K.D., Saam, J.R., Ingram, R.S., Tilghman, S.M. and Bartolomei, M.S. (1995) *Nature Genetics* **9**, 407–413

Triboli, C, Tamanini, F., Patrosso, C., Milanesi, L., Villa, A., Pergolizzi, R., Maestrini, E., Rivella, S., Bione, S., Mancini, M., Vezzoni, P. and Toniolo, D. (1992) *Nucl. Acids. Res.* **20**, 727–733

Van der Krol, A.R., Mur, L.A., Beld, M., Mol., J.N.M. and Stuitje, A.R. (1990) *Plant Cell* **2**, 291–299

Vongs, A., Kakutani, T., Martienssen, R.A. and Richards, E.J. (1993) *Science* **260**, 1926–1928

Watson, J.C., Kaufman, L.S. and Thompson, W.F. (1987) *J. Mol. Biol.* **193**, 15–26

Weber, H. and Graessman, A. (1989) *FEBS Lett.* **253**, 163–166

Weber, H., Zeichmann, C. and Graessman, A. (1990) *EMBO J.* **9**, 4409–4415

Weibauer, K. and Jiricny, J. (1989) *Nature* **339**, 234–236

Weibauer, K., Nedderman, P.and Jiricny, J. (1993) In: *DNA Methylation: Molecular Biology and Biological significance*, J.P. Jost and H.P. Saluz (ed.), pp. 510–522, Birkhauser Verlag, Basel

Weih, F., Nitsch, D., Reik, A., Schultz, G. and Becker, P.B. (1991) *EMBO J.* **10**, 2559–2567

Wigler, M., Levy, D., and Perucho, M. (1981) Cell 24, 33–40

Wilke K., Rauhut, E, Noyer-Weidener, M., Lauster, R., Pawlek, B., Behrens, B. and Trautner, T.A. (1988) *EMBO J.* 7, 2601–2609

Wilkinson, C.R.M., Bartlett, R., Nurse, P. and Bird, A.P. (1995) *Nucl. Acids Res.* **23**, 203–210

Wilson, G.G. (1992) *Methods in Enzymology* **216**, 260–279

Wolf, S.F. and Migeon, B.R. (1985) *Nature* **314**, 467–469

Wolf, S.F., Dintzis, S., Toniolo, D., Persico, G., Lunnen, K.D., Axelman, J. and Migeon, B.R. (1984) *Nucl. Acids Res.* **12**, 9333–9348

Wood, W.I. , Gitschier, J., Laskey, L.A. and Lawn, R.M. (1985) *Proc. Natl. Acad.Sci. USA* **82**, 1585–1588

Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.* **262**, 4776–4786

Wyszynski, M.W., Gabbara, S. and Bhagwat, A.S. (1992) Nucl. Acids Res. **20**, 319–326

Yen, C. R.-W., Vertino, P.M., Nelkin, B.D., Yu, J.J., El-Deiry, W., Cumarswamay, A., Lennon, G.G., Trask, B.J., Celano, P. and Baylin, S.B. (1992) *Nucl. Acids Res.* **20**, 2287–2291

Yesufu, H.M.I., Hanley, A., Rinaldi, A. and Adams, R.L.P. (1991) *Biochem. J.* **273**, 469–475

Yisraeli, J., Adelstein, R.S., Melloul, D, Nudel, U., Yaffe, D. and Cedar, H. (1986) *Cell* **46**, 409–416

Young, R.A. and Davis, R.W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198

Zaret, K.S.and Sherman, F. (1982) *Cell* **28**, 563–573

Zhang, D., Ehrlich, K.C., Supakar, P.C. and Ehrlich, M. (1989) *Mol. Cell. Biol.* **9**, 1351–1356

Zhang, X.-Y., Jabrane-Ferrat, N., Asiedu, C.K., Samac, S., Peterlin, B.M. and Ehrlich, M. (1993) *Mol. Cell. Biol.* **13**, 6810–6818

Zuccotti, M. and Monk M. (1995) *Nature Genetics* **9**, 316–320

Wassenegger, M., Heimes, S., Riedel, L. and Sanger, H.L. (1994) *Cell* **76**, 567–576