

Chapter Six

Multiple cytosine methyltransferase genes in *Arabidopsis thaliana*

6.1 Introduction

Cytosine-5 methyltransferases catalyse the transfer of methyl groups from the cofactor S-adenosyl methionine (S-Adomet) to the C5 position of cytosine in DNA, and are found in bacteria, fungi, plants and animals. The methyltransferase enzymes of all these organisms contain highly conserved amino acid motifs which are important in the methylation process. In prokaryotic methyltransferases, ten conserved motifs are present, and six of these motifs are highly conserved (I, IV, VI, VIII, IX and X; Posfai *et al.*, 1989; Kumar *et al.*, 1994). These six motifs, plus two others, are retained in the carboxy-terminal domains of all eukaryotic methyltransferases so far described (Bestor *et al.*, 1988; Yen *et al.*, 1992; Finnegan and Dennis, 1993; Tajima *et al.*, 1995; Aniello *et al.*, 1996; Kimura *et al.*, 1996; Malagnac *et al.*, 1997; Pradhan *et al.*, 1998; Bernacchia *et al.*, 1998a, 1998b; Yoder and Bestor, 1998; Malagnac *et al.*, 1999). This structural conservation indicates that the methylation mechanism of prokaryotic methyltransferases (reviewed in Kumar *et al.*, 1994; Bestor and Verdine, 1994) may be conserved in eukaryotic methyltransferases. The conserved motifs in methyltransferase enzymes are separated by more variable regions, one of which, between motifs VIII and IX, determines the target sequence for methylation (Balganesh *et al.*, 1987; Klimasauskas *et al.*, 1991; Mi and Roberts, 1992), and is therefore known as the target recognition domain

(TRD, Lauster *et al.*, 1989). The TRD is highly variable among the prokaryotic methyltransferases, mirroring their differing target specificities.

A feature of some eukaryotic methyltransferases, not shared by the prokaryotic enzymes, is a large and relatively poorly conserved amino-terminal domain, thought to target the enzyme to hemimethylated DNA. The mouse Dnmt1 enzyme, which possesses an amino-terminal domain, has a 30-50 fold preference for hemi-methylated DNA over unmethylated DNA (Bestor and Ingram, 1983). Cleavage of the amino-terminal domain activates methylation of unmethylated DNA (Bestor, 1992). Methyltransferases which possess the large amino-terminal domain may be involved in maintaining methylation patterns by targeting sites which are hemi-methylated after DNA replication (Holliday and Pugh, 1975; Riggs, 1975; Wigler *et al.*, 1981). Four eukaryotic methyltransferases lack the amino-terminal domain: murine and human Dnmt2 (Yoder and Bestor, 1998; Okano *et al.*, 1998b), murine Dnmt3a and Dnmt3b (Okano *et al.*, 1998a), and *Ascobolus* Masc1 (Malagnac *et al.*, 1997). While Dnmt2 is not essential for *de novo* or maintenance methylation (Okano *et al.*, 1998b), Dnmt3a, Dnmt3b and Masc1 are thought to be involved in *de novo* methylation of cytosines (Okano *et al.*, 1998a; Malagnac *et al.*, 1997, 1999).

The methyltransferase genes isolated from mouse appear to fall into two functional classes, with Dnmt1 acting as a maintenance methyltransferase (Bestor and Ingram, 1983; Bestor, 1992) and Dnmt3a and Dnmt3b acting as *de novo* (and possibly also maintenance) methyltransferases (Okano *et al.*, 1998a). Dnmt2 may yet prove to function

in *de novo* methylation, possibly being limited to cytosines in specific sequences (Okano *et al.*, 1998b). Human DNA sequences corresponding to the four mouse methyltransferase genes have been isolated (Yen *et al.*, 1992; Yoder and Bestor, 1998; Xie *et al.*, 1998; NCI-CGAP, 1999, <http://www.ncbi.nlm.nih.gov/ncicgap>). *Ascobolus* also contains methyltransferase genes in more than one functional class. *Masc1* is essential for the *de novo* methylation that occurs during sexual reproduction (methylation induced premeiotically, MIP) (Malagnac *et al.*, 1997). *Masc2*, which possesses a large amino-terminal domain thought to target the enzyme to hemi-methylated DNA (Chernov *et al.*, 1997; Goyon, 1998), is not essential for maintenance methylation, MIP, or vegetative *de novo* methylation (Malagnac *et al.*, 1999), though it does have *in vitro* methyltransferase activity (Chernov *et al.*, 1997). Evidently *Ascobolus* possesses at least one other methyltransferase gene active in maintenance and vegetative *de novo* methylation.

While two carrot methyltransferase genes have been identified (Bernacchia *et al.*, 1998a), they are 85 % identical, indicating that they arose from a recent duplication event and may have closely related or identical functions. However, there is evidence for the existence of different functional classes of methyltransferases in plants. In vertebrates, methylation of symmetrical sites occurs mainly in CG dinucleotides, while in plants, cytosines in both CG dinucleotides and CNG trinucleotides are methylated (Gruenbaum *et al.*, 1981; Clark *et al.*, 1995). Separate methyltransferase enzymes exist for methylation of cytosines in CG or CNG sites in pea, and possibly in other plants: two pea methyltransferase enzymes have been purified, one of which methylates cytosines in CG

dinucleotides, while the other acts on C^A/T_TG sequences (Pradhan and Adams, 1995). These two enzymes may be separately encoded, or be produced by proteolytic processing of a single translation product (Adams *et al.*, 1996). A single methyltransferase gene, *MET1*, has been isolated from Arabidopsis, and Southern analysis of Arabidopsis DNA using this gene indicates the presence of a small family of related genes (Finnegan and Dennis, 1993). Arabidopsis plants expressing an antisense transcript of *MET1* have reduced levels of cytosine methylation (Finnegan *et al.*, 1996; Ronemus *et al.*, 1996), particularly in CG and CCG sites, indicating that MET1 may preferentially methylate cytosines in CG dinucleotides (Finnegan *et al.*, 1996), and that another methyltransferase gene may be responsible for methylation of cytosines in CNG trinucleotides. Treatment of tobacco with three different demethylating agents showed that 5-azacytidine inhibited methylation within CG and CCG sites, but not within C^A/T_TG, while ethionine and dihydroxypropyladenine inhibited methylation within CNG sites (where N is any nucleotide) but not within CG sites (Kovarik *et al.*, 1994), again suggesting the presence of two methyltransferase enzymes, with differing specificities and differing sensitivity to inhibitors.

Results reported in previous chapters indicated that, while DNA demethylation has an important role in the early flowering response to cold treatment, demethylation does not wholly substitute for the cold treatment. Plants transformed with the *MET1* antisense construct flower significantly earlier than unvernallized wildtype plants, but, with the exception of one transgenic family, #39, do not flower as early as vernalized wildtype plants. Also, plants from all transgenic families, including #39, retain a proportion of the

wildtype vernalization response. Moreover, *METI* antisense transgenics retain the ability to reset this proportion of the vernalization response in their progeny (Finnegan *et al.*, 1998b), indicating that part of the vernalization response is not mediated by METI-catalyzed methylation. This suggests that the vernalization response is mediated by more than one mechanism, one of which involves demethylation of sites methylated by the METI methyltransferase. It is possible that other methyltransferases, not affected by the *METI* antisense construct, are involved in the vernalization response. The characterization of a second *Arabidopsis* methyltransferase gene, related to *METI* (Finnegan and Dennis, 1993), is described here.

6.2 Materials and Methods

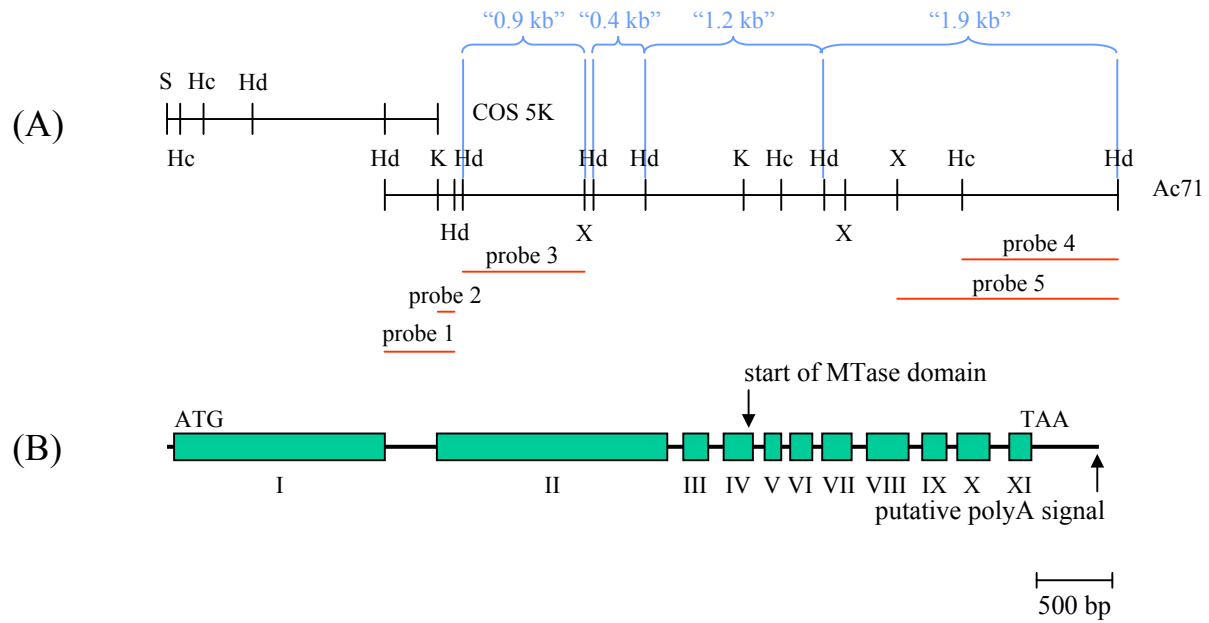
6.2.1 Isolation of *METII* genomic clones

The Ac71 clone was isolated from an *Arabidopsis thaliana* (ecotype Columbia) genomic library (gift of Calgene Australia) by hybridization with a 400 bp *HincII-EcoRI* fragment of the genomic clone Ag42 containing a short *METI*-like sequence (Finnegan and Dennis, 1993; Genger *et al.*, 1999). A 400 bp *HindIII* fragment from the 5' end of the methyltransferase-like sequence within Ac71 (probe 3, Figure 6.1) was used to screen a cosmid library containing *A. thaliana csr-1* DNA (background ecotype Columbia; Keith Davis, Arabidopsis Stock Centre, Ohio State University). A cosmid clone, COS5K, containing the 5' end of the methyltransferase-like ORF, was isolated from this screen. The library screening protocol is described in Section 2.5.1.

Figure 6.1: Structure of the *METII* coding region

(A) Restriction map of genomic clones Ac71 and COS5K spanning *METII* open reading frame. Restriction sites are represented as follows: S: *Sph*I; Hc: *Hinc*II; Hd: *Hind*III; K: *Kpn*I; X: *Xho*I. Sequencing extended from the 5' end of the *Sph*I-*Kpn*I fragment of COS5K to the putative polyadenylation signal in the 3' *Hinc*II-*Hind*III fragment of Ac71. The location of probes used in isolation of the 5' regions of the *METII* gene are shown, as are the four *Hind*III-*Hind*III fragments originally isolated from Ac71 (1.9, 1.2, 0.9 and 0.4kb).

(B) Structure of the *METII* gene, showing exons I to XI (shaded boxes) and 10 introns (black lines). The 5'-most intron, identified by E. J. Finnegan, is not shown. The 3' untranslated region shown is terminated by a putative polyadenylation signal. Putative start and stop codons in the open reading frame are shown.



6.2.2 Restriction and hybridization analysis of Ac71 and COS5K

The Ac71 clone was digested with *EcoRI*, *HindIII* and *BamHI*, transferred to Hybond-N+ nylon membrane and hybridized with ³²P-labelled *METI* cDNA clones using the hybridization protocol described in Section 2.5.2. The *METI* clones were Yc8 and a 1.2 kb *BamHI* fragment of Yc2, which does not contain the 3' end of the gene (Finnegan and Dennis, 1993). The Yc8 clone contains approximately 500 bp of the amino-terminal domain, and the entire methyltransferase domain, while the sequence of the 1.2 kb *BamHI* fragment of Yc2 lies within the amino-terminal domain (Finnegan and Dennis, 1993). *HindIII* fragments hybridizing to either probe were cloned into pUC119. A 5 kb *KpnI* fragment from COS5K, hybridizing to the 400 bp *HindIII* fragment used to isolate COS5K, was cloned into pJKKmf(-).

6.2.3 DNA sequencing and sequence analysis

DNA sequencing methods are described in Section 2.5.3. Nucleotide and amino acid comparisons were done using GCG sequence analysis package version 8.1.

6.2.4 PCR to identify *METI* intron positions

Primers for PCR were designed to flank positions in the *METI* cDNA homologous to the putative positions of intron-exon junctions in *METII*. Sequences of each primer pair are shown in Table 2.1, and the PCR conditions are given in Section 2.6.2.

6.2.5 Protein modeling

The structure of the predicted METI and METII methyltransferase domains was modelled using the protein structure of M.HhaI (Cheng *et al.*, 1993). The Swiss-Model protein modelling server, which used the ProMod protein modelling tool, was used (Peitsch and Jongeneel, 1993; Peitsch, 1995; Peitsch, 1996). The protein structures generated by Swiss-Model were viewed using RasMol version 2.0.

6.2.6 Expression studies

(a) Screening cDNA libraries by hybridization with METI and Ac71 probes

The expression libraries screened are described in detail in Section 2.7.2. Briefly, library 1 contained cDNAs from tissue cultured roots, 7 day old etiolated seedlings, rosettes, stems, flowers and siliques from plants of different ages and light regimes. Library 2 contained cDNAs from above ground tissue of plants from young seedlings through to flowering plants. Library 3 contained cDNAs from flower buds, and library 4 contained cDNAs from apices of vernalized and unvernallized plants.

173 000 plaques from library 1 and 144 000 plaques from library 2 were screened using the 1.0 kb *HincII-HindIII* fragment of Ac71 as a ³²P radiolabelled probe (probe 4, Figure 6.1). 240000 plaques each from libraries 1 and 2, and 120000 plaques each from libraries

3 and 4, were screened using a fragment of the *METI* cDNA clone, containing the active site motif, motif IV, as a ^{32}P radiolabelled probe. Details of the method are given in Section 2.6.2.

(b) Screening cDNA libraries with Ac71 PCR primers

Primers flanking *METII* intron III (Figure 6.2) were used to screen libraries 1 and 2. The method is described in Section 2.6.3. A total of 240000 phage from library 1 and 288000 phage from library 2 were screened.

(c) RT-PCR

Total RNA was isolated from *Arabidopsis* ecotype C24 as described by Dolferus *et al.* (1994). A range of tissues and developmental stages were studied; tissues included roots from plants grown on MS medium; rosette leaves from 6 week old plants grown on MS medium; flower buds (including inflorescence meristems), open flowers and green siliques from soil grown plants, while the different ages were whole seedlings at the 4 leaf stage grown on MS media, and above-ground tissue of 15, 21, 25 and 30 day old soil grown plants. The RT-PCR protocols for both *METII* and *METI* are described in Section 2.6.4.

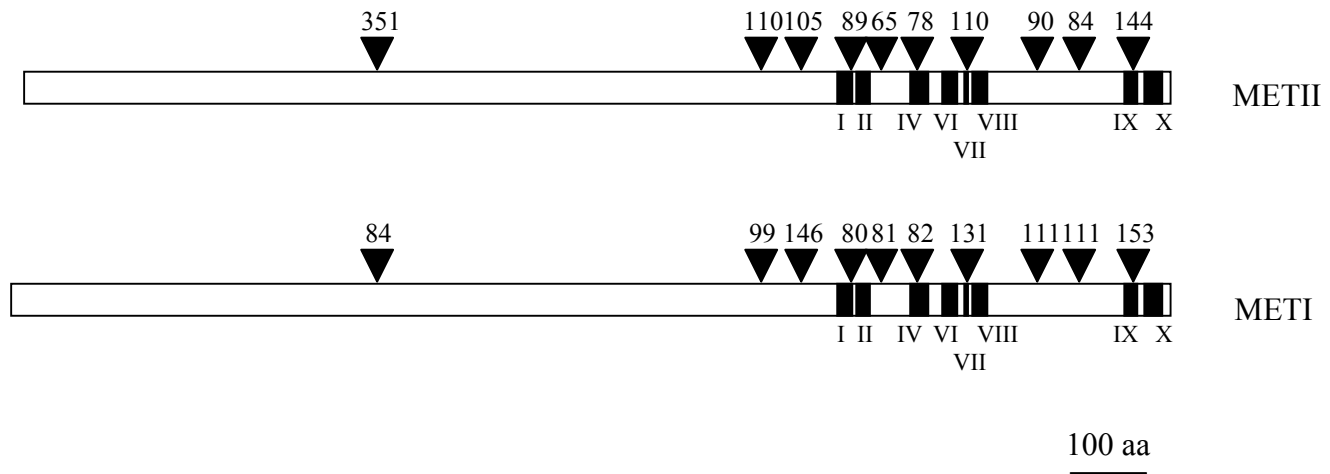


Figure 6.2: Structure of *METII* and *METI* genes, showing the identical positions of introns II to XI (▼); the 5'-most intron, I, identified by E. J. Finnegan, is not shown. The length of each intron in bp is shown above its position. Conserved methyltransferase motifs are shown as shaded boxes.

6.2.7 Mapping the chromosomal location of *METII*

RFLP data for mapping was obtained using recombinant inbred (RI) lines, from a Columbia-Landsberg *erecta* cross, and RFLP data for these lines, both provided by Caroline Dean (Norwich). A polymorphism between Columbia and Landsberg *erecta* was identified for *METII* by Southern analysis, digesting with *Bgl*III and using the 1.6 kb *Xho*I-*Hind*III fragment at the 3' end of Ac71 as a probe (probe 5, Figure 6.1). This polymorphism was scored for 53 recombinant inbred lines, and the data was analysed using MapMaker Macintosh Version 2.0.

6.3 Results

6.3.1 Sequence analysis of a methyltransferase-like ORF

Methyltransferase-like regions of Ac71 were detected by digesting with restriction enzymes including *Hind*III and hybridizing with either the Yc8 or Yc2 cDNA clones of *METI* (Finnegan and Dennis, 1993). Four *Hind*III fragments, approximately 1.9 kb, 1.2 kb, 0.9 kb and 0.4 kb (Figure 6.1), which hybridized to the Yc8 clone of *METI* (Figure 6.3), were cloned. The sequence of these subclones was homologous to *METI*, but the subclones did not contain sequence homologous to the 5' end of *METI*.

The 5' end of the *METI*-like gene was isolated by identification of DNA fragments from Ac71, and then from a cosmid library, which hybridized to the most 5' *METI*-like

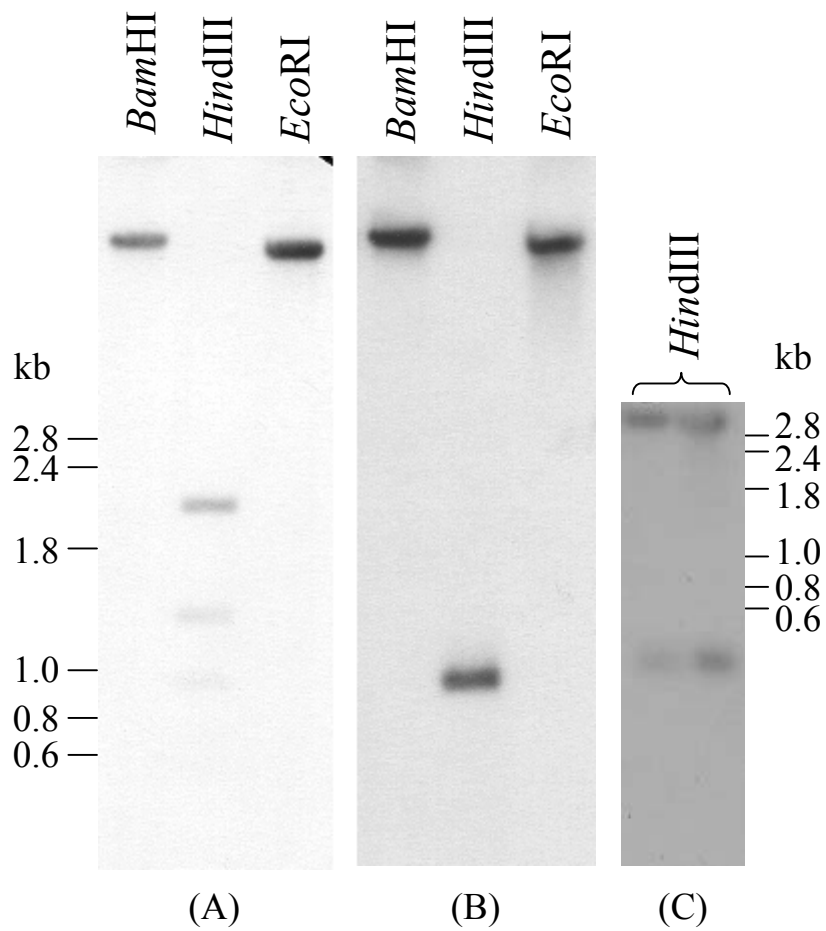


Figure 6.3: Southern analysis of Ac71 clone, showing fragments which hybridize to regions of the *METII* cDNA clone, described in Section 6.2.2. Three *Hind*III fragments, approximately 1.9, 1.2 and 0.9 kb in length, hybridize to the Yc8 *METI* cDNA clone (A), and the 0.9 kb fragment also hybridizes to the 1.2 kb *Bam*HI fragment of Yc2 (B). A fourth *Hind*III fragment, approximately 0.4 kb long, which hybridizes to Yc8, but not to Yc2, was detected after cloning of Ac71 *Hind*III fragments (C). The 3 kb fragment (C) is the linearised vector.

sequences. A *KpnI* fragment of approximately 2 kb was identified from Ac71 by hybridization with the 0.9 kb *HindIII* fragment (probe 1, Figure 6.1). A 120 bp *KpnI*-*HindIII* fragment at the 5' end of this fragment (probe 2, Figure 6.1) was isolated and used to probe *HindIII*-digested Ac71 (Figure 6.1). A 400 bp *HindIII* fragment extending 5' (probe 3, Figure 6.1) was isolated and used to screen a cosmid library (Keith Davis, Arabidopsis Stock Centre, Ohio State University) containing *A. thaliana csr-1* DNA (Columbia background; Haughn and Somerville, 1986). A cosmid clone overlapping Ac71, and extending at least 5 kb upstream (COS5K), was identified from this screen. Subclones from the 3' end of COS5K were sequenced and shown to contain *METI*-like regions corresponding to the 5' end of *METI*. Sequencing identified two potential translation start sites within a *SphI*-*HindIII* subclone of COS5K (Figure 6.1). Both ATG codons are inframe, and occur close together, 47 bp and 53 bp from the *SphI* site respectively. However, the sequence flanking the second ATG conforms most closely to the consensus translation start (Kozak, 1984).

6.3.2 Arabidopsis has a second DNA methyltransferase gene

The length of the assembled transcribed region from Ac71 and COS5K, extending from the putative translation start site to a putative poly-adenylation site downstream of the stop codon, is 5901 bp. The sequence of this gene, designated *METII* (AF138283), is 81% identical with the coding region of *METI*. Nucleotide sequence similarity between *METII* and *METI* decreases slightly towards the 5' end of the genes. In the first 500 bp of coding sequence, the two genes are 76% identical. Through approximately the next 3200

bp of coding sequence the two genes are 80-81% identical, and identity rises to 83% in the final 2065 bp.

The position of introns in *METII* was predicted on the basis of gaps in the alignment of the *METII* genomic sequence and the *METI* cDNA sequence. Eleven exons were identified, separated by introns ranging in size from 65 to 350 bp (Figure 6.2). For all 10 putative introns there is good agreement with the Arabidopsis intron splice site consensus sequence (Brown *et al.*, 1996). The predicted splice junctions for introns III and IV were confirmed by sequencing the RT-PCR product (Figure 6.4).

PCR amplification, cloning and sequencing of regions of the *METI* genomic clone homologous to intron insertion sites in *METII* revealed that the intron locations predicted for *METII* are conserved in *METI*, and intron size is similar between the two genes (Figure 6.2). As Figure 6.2 shows, four of the ten introns occur within sequence coding for conserved motifs.

The cDNA sequence assembled from the *METII* genomic sequence is 4551 bp long, plus 123 bp of 3' untranslated sequence terminated by a putative polyadenylation signal (AATAAA). Thus the predicted protein length is 1517 aa, slightly shorter than the 1534 aa predicted *METI* protein (assuming the second transcription start site, 53 bp upstream of the *SphI* restriction site, to be the correct one). The structure of the predicted *METII* protein is similar to that of *METI* (Finnegan and Dennis, 1993) and other eukaryotic methyltransferases (Bestor and Verdine, 1994; Tajima *et al.*, 1995; Aniello *et al.*, 1996),

Figure 6.4: Comparison of the sequence of the *METIII* RT-PCR product, spanning the predicted position of introns III and IV, with the genomic sequence in this region. Introns III and IV have been spliced out of the RT-PCR product as predicted; intron splice sites are shown in bold. Priming sites used to amplify the RT-PCR product are underlined. Although the 3' primer was inadvertently designed to span intron IV, a comparison of the PCR product amplified from genomic DNA with the RT-PCR product indicated that introns III and IV are spliced as shown.

consisting of a carboxy-terminal domain with homology to bacterial methyltransferases (methyltransferase domain), 463 aa in length, and a 1049 aa amino-terminal domain. The two domains are separated by the sequence KKGKG, as in METI (Finnegan and Dennis, 1993); in other eukaryotic methyltransferases the two domains are also separated by lysine-glycine repeats. The *METI* and *METII* methyltransferase domains are 79% identical, and the amino-terminal domains are 68% identical.

6.3.3 Methyltransferase domain

Alignment of the METI and METII amino acid sequences from the beginning of motif I to the end of the peptide demonstrates that throughout the catalytic region of the methyltransferase domain, the two proteins are identical in length and structure. The methyltransferase domain of METII retains the 8 motifs seen in all eukaryotic methyltransferases of thi0s class. These show high similarity to the METI motifs (Figure 6.5), with amino acid identity ranging from 70 - 100%. Invariant residues in prokaryote methyltransferases are indicated by capitals in the consensus sequence; all but one of these invariant residues are conserved in METII, as in METI. The target recognition domain, between motifs VIII and IX, is 72% identical between METI and METII.

The structure of the METI and METII methyltransferase domains was modelled on the protein structure of *M.HhaI*, which was solved by X-ray crystallography (Cheng *et al.*, 1993; Klimasauskas *et al.*, 1994; Reinisch *et al.*, 1995). Tertiary structure was well conserved between the bacterial and plant enzymes in the regions of the conserved

Figure 6.5: Comparison of the conserved methyltransferase motifs in *METI* and *METII*.

Amino acid identity between the two methyltransferases is shown. The consensus sequence (Cons) of a range of prokaryote methyltransferases, based on consensus sequences given by Posfai *et al.* (1989), is shown below the plant sequences; absolutely conserved residues are shown in upper case.

MOTIF I

METI LDIFAGCGGLSHGLKKAGV 84 % identity
METH LDIFAGCGGHSHGLENAGV

Cons mdlFaG-ggf--a----G-
isf s mai g
l c l s

MOTIF II

METI AKWAIEYEYPAGQAFKQN 89 % identity
METH TKWAIEYEYPAGHAFKQN

Cons -i--se-d--a--sy--i
v nd n i vf n
l i f ti

MOTIF IV

METI VDFINGGPPCQGFSGMNRFNQSSW 87.5 % identity
METH VDFINGGPPCQGFSGMNRFSHGWS

Cons -d-l-sGfPCp-fS--G-----e-
n i g s q w d
v a p s

MOTIF VI

METI RPRYFLLENVRTFVSFNKGQ 70 % identity
METH RPKYFLLENVKKFVTYNKGR

Cons -p-----ENV-g-----g-
t n n
k

MOTIF VII

METI MGYQV 100 % identity
METH MGYQV

Cons -gY-i
d f
v

... continued on next page

MOTIF VIII

METI LEAGAYGVSQSRKRAFIWAA 80 % identity
METH LEAGTYGVSQPRKRVIWAA

Cons -ns--f_{nv}-Q-R-R---ig-
ia ygi va
dd hfl ec

MOTIF IX

METI HRILTVRECARSQGFPD 88 % identity
METH DRIITVRECARSQGFPD

Cons -R-ls--E--rlq-fd-
mt avm yp
fh vi e

MOTIF X

METI GNINHKHRQIGNAVPPPLAFALGR 87.5 % identity
METH GTTKHKHRQIGNAVPPPLAFALGR

Cons -s----Ykq-GNsi-v-v----f-
k re av i a a
r qm p l g

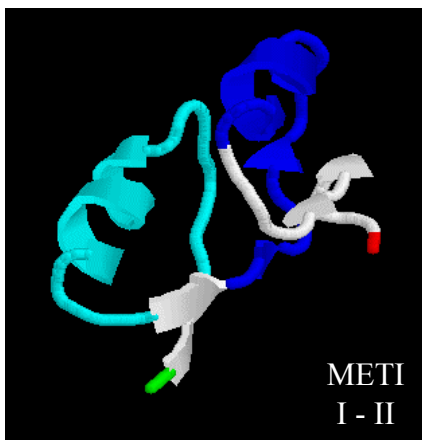
motifs. Three stretches of peptide sequence with good conservation of structure were seen; the first encompassed motifs I and II, the second motifs IV, VI, VII and VIII, and the third motifs IX and X. No conservation of structure was seen in the TRD (Figure 6.6).

The methyltransferase domains of ten plant methyltransferases, and of the mouse and human Dnmt1 (DNMT1) genes, were compared using PileUp analysis (Figure 6.7). With the exception of the Arabidopsis chromo-methyltransferases, *CMT1* and *CMT2* (Henikoff and Comai, 1988; Genger *et al.*, 1999), all plant methyltransferases showed high amino acid similarity throughout the methyltransferase domain, both within the conserved motifs and the TRD. In the TRD, *CMT1* and *CMT2* had two deletions of 8 and 12 aa, and one 36 aa insertion, relative to the other plant methyltransferases. This insertion occurred at the same point as a 40 aa insertion in the mouse and human TRDs, relative to the other plant methyltransferases; however there was little similarity between the *CMT* and *Dnmt1* sequences within these insertions. The methyltransferase domains of both *CMT* genes contain a chromodomain (Koonin *et al.*, 1995; Cavalli and Paro, 1998), between motifs II and IV, which has no homology with any of the other genes in the analysis.

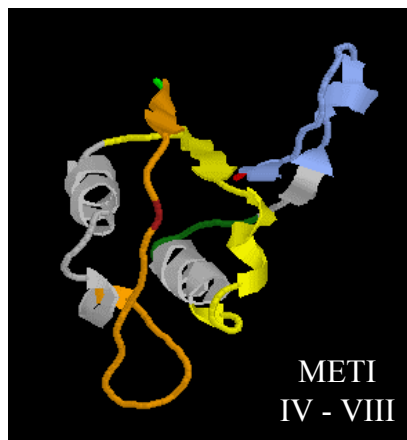
6.3.4 Amino-terminal domain

Homology between METI and METII in the amino-terminal domain is lower than in the methyltransferase domain: alignment of the two peptides introduces small gaps throughout the amino-terminal domain, and amino acid identity is 68%. Like METI

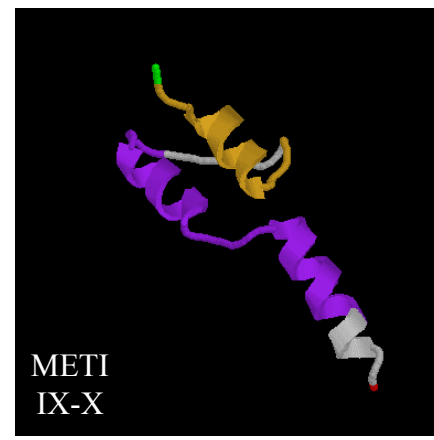
Figure 6.6: Structures of *METI* and *METII* predicted proteins, modelled on the structure of M.HhaI. Output of SwissModel modified with CHARMM. (A) Region containing motifs I (cyan) and II (blue). (B) Region containing motifs IV (orange), VI (yellow), VII (dark green) and VIII (light blue); the cysteine residue within the active site in motif IV is shown in burgundy. (C) Region containing motifs IX (brown) and X (purple). For each peptide region shown, the amino terminal residue is shown in green, and the carboxy terminal residue is shown in red.



(A)



(B)



(C)

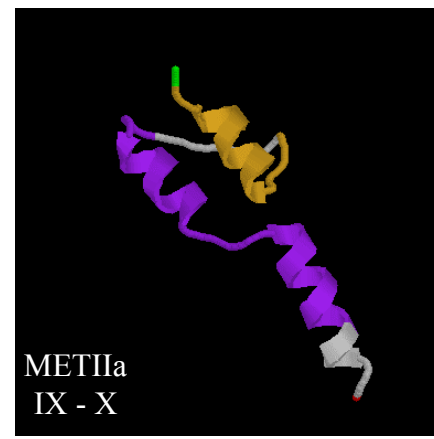
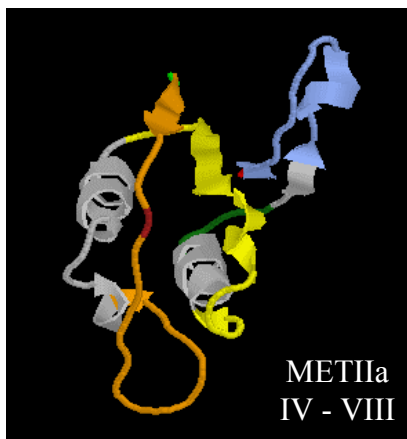
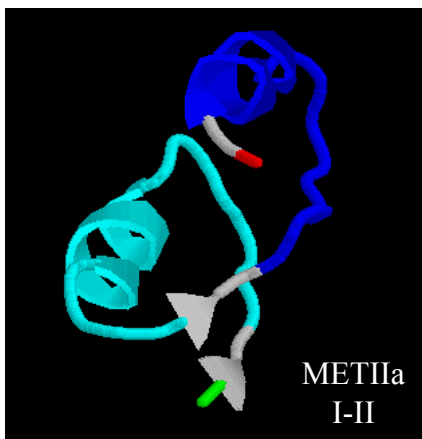


Figure 6.7: Alignment of methyltransferase domains of twelve eukaryotic cytosine methyltransferases: METI, METIIa, METIIb, the carrot methyltransferases CMET5 and CMET21 (Bernacchia *et al.*, 1998a), the tomato methyltransferase TMET (Bernacchia *et al.*, 1998b), the pea methyltransferase PMET (Pradhan *et al.*, 1998), the maize methyltransferase ZMET (GenBank accession AF063403), the Arabidopsis chromomethyltransferases CMT1 and CMT2 (Henikoff and Comai, 1998; Genger *et al.*, 1999), the mouse methyltransferase Dnmt1 (Bestor, 1988) and the human methyltransferase DNMT1 (Yen *et al.*, 1992). For CMT1 and CMT2, the sequence used began 3 amino acids before conserved motif I. The 57 amino acid chromodomain was deleted from both CMT peptides to allow correct alignment of the methyltransferase motifs; the position of this deletion is shown by ▼. For the other enzymes, all residues after the lysine-glycine repeats which separate the amino terminal and methyltransferase domains were included. The degree of shading indicates the degree of conservation, with absolutely conserved residues white on a black background. The positions of the eight conserved motifs are shown.

```

CMT1      : ~~~~~~KFLLDLYSGCGAMSTGFCMGASISGVKLIITKWSVDINKFACDSLKLNHPETE-V
CMT2      : ~~~~~~LPVLDLYSGCGGMSLGLSLGAKISGVDVVTKWAVDQNTAACKSLKLNHPNTQVV
DNMT1     : ~~~~~~LRITLDVFSGCGGLSEGL----HQAGIS-DTLWAIEMWDPAAQAFRLNPNPGSTVF
Dnmt1     : ~~~~~~LRITLDVFSGCGGLSEGL----HQAGIS-ETLWAIEMWDPAAQAFRLNPNPGTTFV
CMET21    : ~~~~~~LATLDFAGCGGLSEGL----QQSGVC-RTKWAIEYEEEPAGDAFKLNHPDITMF
CMET5     : ~~~~~~LATLDFAGCGGLSEGL----QKSGVC-TTKWAIEYEEAAGDAFKLNHPESLMF
TMET      : ~~~~~~LSTLDFAGCGGLSEGL----QHSGVT-DTNWAIYEYEPAGDAFRLNHPKTKVF
PMET      : ~~~~~~LATLDFAGCGGLSEGL----HKSGAS-STKWAIEYEEEPAGNAFKLNHPALVF
ZMET      : ~~~~~~LATLDFAGCGGLSEGL----QQAGVS-FTKWAIEYEEEPAGEAFKLNHPAIVF
METIIa    : ~~~~~~LATLDFAGCGGLSHGL----ENAGVS-TTKWAIEYEEEPAGHAFKLNHPAIVF
METIIb    : ~~~~~~LTLDFAGCGGLSHGL----EKAGVS-NTKWAIEYEEEPAGHAFKLNHPAIVF
METI      : VESEIESEIVKPEPPKEIRLATLDFAGCGGLSHGL----KKAGVS-DAKWAIEYEEEPAGQAFKLNHPSTVF

```

I

II

```

CMT1      : RNEAAEDFLALLKEWKRICEK--FSLVSSTEPVESISELEDEEVEENDDIDEASTGAELFVIDGFKSHL LPLPG
CMT2      : RNDAAAGDFLQLLKEWDKLCRYVFNNDQRTDTLRSVNSTKFTSGSSSSSDDDS-----DVTSGFKSKI LPLPG
DNMT1     : TEDC-----NILLKLV-----AGETTNSRGOR-----LPQKG
Dnmt1     : TEDC-----NVLKLV-----AGEVTNSLGOR-----LPQKG
CMET21    : INNC-----NVILKAIMDKSGDADDCISTPEAAELAAKLSEE-----ELKNLPLPG
CMET5     : INNC-----NVILKAIMDKTGADDCISTPEAAELAAKLSEE-----EIKNLPLPG
TMET      : IHNC-----NVILRAVMQKCGSDDCISTPEASELAAAMDES-----ELNSLPLPG
PMET      : INNC-----NVILRAIMEKCGDIDECISTAEAAELASKLDDK-----DLNSLPLPG
ZMET      : VDNC-----NVILKAIMDKCGDTDDCVSTSEAAEQAAKLPEV-----NINNLPLPG
METIIa    : VDNC-----NVILRAIMEKCGDVDDCVSTVEAAELAAKLLEN-----QKSTLPLPG
METIIb    : VDNC-----NVILRAIMEKCGDVDDCVSTVEAAELVAKLDEN-----QKSTLPLPG
METI      : VDNC-----NVILRAIMEKGGDQDDCVSTTEANELAAKLTEE-----QKSTLPLPG

```

```

CMT1      : TVYTVCGGPPCQGISGYNRRNNEAPLEDQKNQOLLVFLDIIIFLKPNYVLMENVVDLIRFSKGFRLARHAVASL
CMT2      : RVGVICGGPPCQGISGYNRRHRNVDSPLNDERNQOIVFMDIVEYLYKPSYVLMENVVDILRMDKGSGLGRYALSRL
DNMT1     : DVELCGGPPCQGFSGMNR--NSRTYSKFKNSLVVSVLSYCDYRPRFFLLENVRNFVSKRSMVLKLTLRCL
Dnmt1     : DVELCGGPPCQGFSGMNR--NSRTYSKFKNSLVVSVLSYCDYRPRFFLLENVRNFVSKRSMVLKLTLRCL
CMET21    : QVDFINGGPPCQGFSGMNR--NQSSWSKVQCEMILAFLSFADYRPRKYFLENVRNFVSNKQOTFRLATIASL
CMET5     : QVDFINGGPPCQGFSGMNR--NQSSWSKVQCEMILAFLSFADYRPRKYFLENVRNFVSNKQOTFRLATIASL
TMET      : QVDFINGGPPCQGFSGMNR--NQSTWSKVQCEMILAFLSFADYRPRKYFLENVRNFVSNKQOTFRLATIASL
PMET      : QVDFINGGPPCQGFSGMNR--NSTTWSKVQCEMILAFLSFADYRPRKYFLENVRNFVSNKQOTFRLATIASL
ZMET      : EVEFINGGPPCQGFSGMNR--NQSPWSKVQCEMILAFLSFAEYRPRKYFLENVRNFVSNKQOTFRLATIASL
METIIa    : QVDFINGGPPCQGFSGMNR--SHGWSKVQCEMILAFLSFADYRPRKYFLENVKKFVTVNKGRTFQLTMASL
METIIb    : QADFISGGPPCQGFSGMNR--SDGWSKVQCEMILAFLSFADYRPRKYFLENVKKFVTVNKGRTFQLTMASL
METI      : QVDFINGGPPCQGFSGMNR--NQSSWSKVQCEMILAFLSFADYRPRKYFLENVRNFVSNKQOTFRLATIASL

```

IV

VI

```

CMT1      : VANNYQTFGLGMAAGSYGLPQLRNRVFLWAAQPSKLP-PYPLPTHEVAKKFNTPKFKDLQVGRIQMEFLKLD
CMT2      : VNMRYQARLGMITAGCYGLSOFRRSRVFMGAVPNKNLP-PFPLPTHDVIVRYGLPLEFERNVVAEAGQPRKLE
DNMT1     : VRMGYQCTFGVLOAGQYGVAQTRRRRAILAAAPGKELPL-FPEPLH-VFAPRACQLSVVDDKKFVSNITRLSS
Dnmt1     : VRMGYQCTFGVLOAGQYGVAQTRRRRAILAAAPGKELPL-FPEPLH-VFAPRACQLSVVDDKKFVSNITRLSS
CMET21    : LEMGYQVRFGILEAGAFGVPOSRRKRAFIWAASPEETLPGSWPEPMH-VFAAPELKVA-LPGNKHYAAVRSTQAG
CMET5     : LDMGYQVRFGILEAGAYGVPOSRRKRAFIWAASPEETLPEWPEPMH-VFAAPELKIA-LPENKYAAVRSTQTG
TMET      : LEMGYQVRFGILEAGAYGVPOSRRKRAFIWAGSPEEVLP-EWPEPMH-VFAVPELKIA-LSETSYAAVRSTASG
PMET      : LEMGYQVRFGILEAGAFGVSQSRKRAFIWAASPEDVLP-EWPEPMH-VFASAPELKIT-LAENVQYAAVCSTANG
ZMET      : LEMGYQVRFGILEAGAFGVAQSRKRAFIWAAAPGEMLP-DWPEPMH-VFASPELKIT-LPDGQYAAARSTAGG
METIIa    : LEMGYQVRFGILEAGTYGVSQPRKRVIWAASPEEVLP-EWPEPMH-VFDNPGSKIS-LPRGLRYDAGCNTKFG
METIIb    : LEIGYQVRFGILEAGTYGVSQPRKRVIWAASPEEVLP-EWPEPMH-VFDNPGSKIS-LPRGLHYDTRNTKFG
METI      : LEMGYQVRFGILEAGAYGVSQSRKRAFIWAAAPPEEVLP-EWPEPMH-VFVGVPKLKIS-LSQGLHYAAVRSTALG

```

VII VIII

```

CMT1      : ---NALLIADAISDLPPVTNYVANDVMDYNDAAPKTEFENFI-SLKRSETLLPACGGDPTRRLFQPLVLGDD
CMT2      : ---KALVLKDAISDLPHVSNDEDREKLPY-ESLPKTDFORYIRSTRKRTGSAIDNC-NKRTMLLHDHRPFHINED
DNMT1     : GPFRTITVRDITMSDLPEVRNGASALEISYNGE-POSWFQRLR-----GAQYQPILRDHICKDMSAL
Dnmt1     : GPFRTITVRDITMSDLPEIONGASNSEIPYNGE-PLSWFQRLR-----GSHYQPILRDHICKDMSPL
CMET21    : APFRAITVRDITIGDLPVITNGASKTTLEYRCD-PISWFQKNIR-----AN--MMVLTDHISKEMNEL
CMET5     : APFRSITVRDITIGDLPVITNGASRTSIEYQMD-PISWFQKKIR-----AN--MMVLTDHISKEMNEL
TMET      : APFRSLITVRDITIGDLPVITNGASKTCIEYQGD-PVSWFQKKIR-----GS--SITLSDHISKEMNEL
PMET      : APLRAITVRDITIGELPAVGNNGASRTNMEYQSD-PISWFQKKIR-----GN--MAVLTDHISKEMNEL
ZMET      : APFRAITVRDITIGDLPKVGNGASKLTLEYGGE-PVSWFQKKIR-----GS--MMVLTDHISKEMNEL
METIIa    : APFRSITVRDITIGDLPVENGESKINKEYGTT-PASWFQKKIR-----GN--MSVLTDHICKGLNEL
METIIb    : APFRSITVRDITIGDLPVENGESKINKEYRTT-PVSWFQKKIR-----GN--MSVLTDHICKGLNEL
METI      : APFRPITVRDITIGDLPVENGDSRTNKEYKEV-AVSWFQKEIR-----GN--TIALTDHICKAMNEL

```

```

CMT1      : DLBRVSYIPKQKGANRDMPE-----GVLVHNNKAEINPRFR-AKLKSGKNVVPAYAISFIKGSKPKP-----
CMT2      : DYARVCQIPKRRGANFRDLP-----GLIVRNNTVCRDPSMEPVILPSGKPLVP--GVFTFQQKSKRP-----
DNMT1     : VAARMRHIPLAPGSDWRDLPNIEVRLSDGTMARKLRYTHHDKRNGRSSSGALRGVCSCVEAGKACDPAARQFNT
Dnmt1     : VAARMPHIPLFPGSDWRDLPNIQVRLGDGVIADLQYTFHDVKNGYSSTGALRGVCSCAE-GKACDPESRQFST
CMET21    : NLIRCQRIPKRRGADWHDLPEEKVKLSTGQLVD-----RIPRPGADWDLPEEKVKLSSGQLVD-----
CMET5     : NLIRCQRIPKRRGADWQDLPEEKVKLSSGQLVD-----RIPRPGADWDLPEEKVKLSSGQLVD-----
TMET      : NLIRCQRIPKRRGADWRDLEDEKVKLSNGQLVD-----RIPRPGADWDLPEEKVKLSSGQLVD-----
PMET      : NLIRCQKI PKRPGCDWRDLPEEKVKLSTGQLVD-----RIPRPGADWDLPEEKVKLSSGQLVD-----
ZMET      : NLIRCQHI PKRPGCDWHDLPEEKVKLSNGOMAD-----RIPRPGADWDLPEEKVKLSSGQLVD-----
METIIa    : NLIRCCKI PKRPGADWRDLPDENVTLNGLVEK-----RIPRPGADWDLPEEKVKLSSGQLVD-----
METIIb    : NLIRCCKI PKRPGADWRDLPDENVTLNGLVEK-----RIPRPGADWDLPEEKVKLSSGQLVD-----
METI      : NLIRCKLIPTRPGADWHDLPRKRVTLSNGRVEE-----RIPRPGADWDLPEEKVKLSSGQLVD-----

```

```

CMT1      : -----FGRLLWGDEIVNTVVTRAEAPHNQC--VIHPMQNRVLSVRENARLQGFPCYKLCGT
CMT2      : -----FARLWWDDETVPVTLVPTCHSQA--LLHPEQDRVLTIRE SARLQGFDPYFQFCGT
DNMT1     : LIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPTYRFLGN
Dnmt1     : LIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPSYRFFGN
CMET21    : LIPWCLPNTAKRHNQWKGLFGRLDWEGNFPTSITDPQPMGKVGMCFHDPQDRIVTVRECARSQGFPSYQFYGN
CMET5     : LIPWCLPNTAKRHNQWKGLFGRLDWEGSFPTSITDPQPMGKVGMCFHDPQHRIVTVRECARSQGFPSYQFYGN
TMET      : LIPWCLPNTAKRHNQWKGLFGRLDWQGNFPTSITDPQPMGKVGMCFHDPQDRIVTVRECARSQGFPSYQFAGN
PMET      : LIPWCLPHTAKRHNQWKGLFGRLDWQGNFPTSITDPQPMGKVGMCFHDPQDRILTVRECARSQGFDPHYQFSGN
ZMET      : LIPWCLPNTAKRHNQWKGLYGRLDWEGNFPTSITDPQPMGKVGMCFHDPQDRITVRECARSQGFPSYEFAGN
METIIa    : LRPLALSKTAKNHNEWKGLYGRLDWQGNLPISITDPQPMGKVGMCFHPEQDRITVRECARSQGFPSYEFSGT
METIIb    : LRPLALSKTAKNHNEWKGLYGRLDWQGNLPISITDPQPMGKVGMCFHPEQDRITVRECARSQGFPSYEFSGT
METI      : MIPFCLPNTAERHNGWKGLYGRLDWQGNFPTSITDPQPMGKVGMCFHPEQHRILTVRECARSQGFPSYEFAGN

```

IX

```

CMT1      : IKEYIQVGNVAVPVGVALGYAFGMA SQGLTDDEPVIKLPFKYPECMQAKDQ~~~~~
CMT2      : IKERYCQIGNAVAVSVSRALGYS LGMAFRGLARDEHLIKLPQNFSSHSTYPQLQETIPH
DNMT1     : ILDKHRQVGNVAVPPPLAKAIGLEIKLCMLAKARESASA--KIKEEEAAKD~~~~~
Dnmt1     : ILDRHRQVGNVAVPPPLAKAIGLEIKLCLLSSARESASAAVKAKEEAATKD~~~~~
CMET21    : ILHKHRQIGNAVPPPLAYALGRKLKEALESK GSM~~~~~
CMET5     : ILHKHQIGNAVPPPLAYALGMKLKEALESK GCM~~~~~
TMET      : ILHKHRQIGNAVPPPLAYALGRKLKEAVESKNRLT~~~~~
PMET      : I IHKHRQIGNAVPPPLAFALGRKLKEALDSKSAN~~~~~
ZMET      : IQNKHRQIGNAVPPPLAYALGRKLKEAVDKRQEASAGVPAP~~~~~
METIIa    : TKHKHRQIGNAVPPPLAFALGRKFKKEALYLKSSLQHQS~~~~~
METIIb    : TKHKHRQIGNAVPPPLAFALGRKLKEALYLKSSLQHQS~~~~~
METI      : INHKHRQIGNAVPPPLAFALGRKLKEALHLKKSPOHQP~~~~~

```

X

(Finnegan and Dennis, 1993), METII has low homology to the mouse Dnmt1 methyltransferase in the amino-terminal domain (27% identity). PileUp analysis was performed for the amino-terminal domains of eight plant methyltransferases. Three of these are Arabidopsis cytosine methyltransferases: METI, METII and a third methyltransferase, METIIb, identified by database searches (Genger *et al.*, 1999). A fourth methyltransferase-like gene, *METIII*, was not included in the analysis, as the predicted peptide for this gene truncates after 64 aa (Genger *et al.*, 1999). *METII* and *METIIb* are highly similar, as will be discussed below; thus *METII* will be designated *METIIa* from here on. The remaining five plant methyltransferases included in the PileUp analysis are the two carrot methyltransferases CMET5 and CMET21 (Bernacchia *et al.*, 1998a), tomato methyltransferase (TMET; Bernacchia *et al.*, 1998b), pea methyltransferase (PMET, Pradhan *et al.*, 1998) and maize methyltransferase (ZMET, AFO63403). All these enzymes are similar in this domain (Figure 6.8). Initial PileUp analysis revealed that CMT1 and CMT2 (Henikoff and Comai, 1998; Genger *et al.*, 1999) showed no similarity to the other plant methyltransferases in their amino-terminal domains, and therefore they were omitted from the final analysis.

A number of features are conserved in the amino-terminal domains of the eight plant methyltransferases analyzed. The motif S/TPXX, where X is usually a basic amino acid, binds DNA in the minor groove (Churchill and Suzuki, 1989). The S/TPXX motif occurs four times in the METIIA amino-terminal domain. Two of these motifs have at least one basic residue in the X position (SPER, TPKK), and the position and sequence of these

Figure 6.8: Alignment of amino-terminal domains of eight plant cytosine methyltransferases: METI, METIIa, METIIb, the carrot methyltransferases CMET5 and CMET21 (Bernacchia *et al.*, 1998a), the tomato methyltransferase TMET (Bernacchia *et al.*, 1998b), the pea methyltransferase PMET (Pradhan *et al.*, 1998), and the maize methyltransferase ZMET (GenBank accession AF063403). All residues before the glycine-lycine repeats that separate the amino terminal and methyltransferase domains were included. The first 211 residues of CMET21 were not included in the sequence comparison, as they were found to have little similarity to the other sequences. The degree of shading indicates the degree of conservation, with absolutely conserved residues white on a black background.

```

METIIa : ~~~~~METKAGKQKRR---SVDSDDDVSKEERR-PKRAAAC--TN
METIIb : ~~~~~METKVGKQKRR---SVDSDDDVSKEERR-PKRAAAC--RN
METI : ~~~~~MVENGAKAAKRRKRRPLPEIQEVEDVPRTRR-PRRAAAC--TS
CMET21 : AGSRKMPKRAADCADDKE-KAVKISK-KVSSAKAKMDQALDEEEVAVQEAHVSKSSEVSAGSRKMPKRAACA--D
CMET5 : ~~~~~MGSSAVBDAPALD-AGLETKKNKRNADCDSEKTABSGQ-KKQRAHALKSSETPVGSRKMPKRAACA--D
TMET : MASQPNSSESVLELPNND-KSGH-KKNKRKQDSVSKRKASATGKKEKKAQA-VSETIEEPTAGRKRPKRAAAC--D
PMET : ~~~~~MGSASLLNPSDSSSLPGGKDSTSKEEPVSNTEGEVMAGGKQKRRS--LSESEQAPPTRKVPKRSASAASKN
ZMET : ~~~~~MQSKATKEGRGIHRKQQAGEWISGYNRRGASWSRKS DGHVTRKRPRRSAAACS--D

```

```

METIIa : FKEKSI R I S D K S E T V E A K K E Q I L A E I V A I O L T S S L E S N D D P R P N R R L T D F V L H D S E G V P Q P V E M L E L G D I F I E G V
METIIb : FKEK P L R I S D K S E T V E A K K E Q N V E E I V A I O L T S S L E S N D D P R P N R R L T D F V L H N S D G V P Q P V E M L E L G D I F L E G V
METI : FKEK S I R V C E K S A T I E V K K Q Q I V E E F L A L R L T - A L E T D V E D R P T R R L N D F V L F D S D G V P Q P L E M L E I H D I F V S G A
CMET21 : FKEK V V Q I S K K A S I I E T K K D R C V D E E E M A V R L T A G - Q - - E D G R P C R R L T D F I L H N S D G V Q P F E M L E V D D L F I S G L
CMET5 : FKEK S I Q I S K K S S I I E T K K D R S V D E E E V A V R L T A G - Q - - E D G R P C R R L T D F I F H N S D G I P Q A F E M L E V D D L Y I S G L
TMET : FKEK S V H L S K K S S V I E T K K D H C V D E E D V A I R L T A G L Q - - E S O R P C R R L T D F V F H N S E G I P Q P F G M S E V D D L F I S G L
PMET : L K E K S F S I S D K S C L V E T K K D Q V A E G E L L A V R M T A G - Q - - E D D R P N R R L T D F I L H D E S G A A Q A L E M L E I K D L F I T G L
ZMET : FKEK S I R L S E K K S V M V K K N R M E E E V D A V N L T K - - L G P E D P P P C R K L I D F I L H D A E G N P Q P F E M S E I D D F I T A L

```

```

METIIa : V L P - - L G D E K K E E K G V R F Q S F G R V E N W N I S G Y E D G S P V I W I S T A L A D Y D C R K P S K K Y K K L Y D Y F F E K A C A C V E V F K
METIIb : V L P - - L G D D K N E E K G V R F Q S F G R V E N W N I S G Y E D G S P G I W I S T A L A D Y D C R K P A S K Y K K I Y D Y F F E K A C A C V E V F K
METI : I L P S D V C T D K E K E K G V R C T S F G R V E H W S I S G Y E D G S P V I W I S T E L A D Y D C R K P A A S Y R K V Y D Y F Y E K A R A S V A V Y K
CMET21 : I L P L E E S S Q K - E D C S I R C E G F G R I E D W A I S G Y E D G V P I I W V S T V A D Y D C V K P S A A Y K K H Y E H F F A K A T A C I E V Y K
CMET5 : I L P L E D S S Q K - E A C S I K C E G F G R I E N W A L S G Y E E G V P T I I W V S T V A D Y D C V K P S A S Y K K H Y E H L F A K A T A C V E V Y K
TMET : I L P L E D S L D K V K A K G I R C E G F G R I E W A I S G Y E D G T P V I W I S T E T A D Y D C L K P S G S Y K K F Y D H F L A K A T A C V E V Y K
PMET : I L P L E G N A D K K K E Q V R C H G F G R I E S W D I S G Y E D G S P V I W I S T E T A D Y D C Q K P A G T Y K K Y Y D L F F E K A R A C L E V Y K
ZMET : I M P M D D D L E K E R E R G V R C E G F G R I E D W N I S G Y D E G T P V I W V S T V A D Y E C V K P S T N Y K S Y F D H F Y E K A Q V C V E V F K

```

... continued on next page

```
METIIa : SLSK----NPDTSLDELLAAVSRSM-SGSKIIFSSGGAIQEFVISQGEFIYNQLAGLDETAKNHETCFVFNRLVLSL
METIIb : SLSK----NPDTSLDELLAAVSRSM-SGSKIIFSSGGAIQEFVISQGEFIYNQLAGLDETAKNHETCFVENSVLVSL
METI : KLSKSSGGDPDILGELLELLAAVVRSMSSGSKYFSSGAAIIDFVISQGDFIYNQLAGLDETAKKHESYVEIPVLVAL
CMET21 : KLSKSSGGNPDLSEFDELLAGVVRAM-NGMKCFSRGVSIKDFIISQGEFIYNQLVGLDETS-KDDQQFLELPVLVAL
CMET5 : KLSKSSGGNPDLSEFDELLAGVVRGL-SGMKCFSRVSVIKDFIISQGDFIYNQLVGLDETSKKDQQFLELPVLIAL
TMET : KLSKSSGGNPDLSEFDELLAGVVRAM-TGIKCFSGGVSIRDFVITOGGFIYKELIGLDDTSKKTDQLFVELPVLASL
PMET : KLAKSSGGDPDISLDELLAGMARS-SGSKYFSGTASLKEFIISQGDFIYKOLIGLDTMLKANDKGFEDIPALIAL
ZMET : KLAKSVGGNPNQGLDELLASVVRSTNAMKGYSGTMS-KDLVISIGEFVYNQLVGLDETSNNDDKEFATLPVLVLSL
```

```
METIIa : RDHESNKIHKAL-----SN-VALR----IDESKVVTSDHLDVGA-EDEDVKYAKLIQEEFYRKSMERSRNKRSST
METIIb : RDHESSKIHKAL-----SN-VALR----IDESQLVKS DHLDVGA-EAEDVRYAKLIQEEFYRISMERSRNKRSST
METI : RE-KSSKIDKPLQERPNPSNGVRRIKEVSQVAESEAALSDQLVDGT--DDDRRYAILLODEENRKSMMQQPR--KNSS
CMET21 : RDESSRHVNDFQERIGCTNGTLL---KIRDNEHQ-----KNSVTEEGEDKKMARLLOEFFFWKSMKQKKGQGRSV
CMET5 : RESSKHGDPDSIGKVASTNGTLLTIGPKIKDGENK-----KESATEEDEGVKVARLLOEFFFWNNSMKQKKGGRGSSST
TMET : RDESSKHETLAQPETISSNGLRITGPKAGNGGDKIVESGLANGPAP-EDEDLKLAKLHHEEYWCSLKQKDRNTSS
PMET : RDESKKQAHFANTQVRPSNATLRITGSGIVDEEKK----NQMDSVDEEDEDKTLARLLODEEYWKSNRQRKNSRSSS
ZMET : RDQCRSRVELTKLPSNFSNTSLKIL-----KDESEDETAEDDDDAKLARLLOEFEWKMMKKQGRGRGTP
```

```
METIIa : TSGGSSRFYIKISEDEIADDDYPLPSYYKNTKEETDELVLFEAGYEV-DTRDLP CRTLHNWTLYNSDSRMISLEVL
METIIb : TS-ASNKFYIKINEHEIANDYPLPSYYKNTKEETDELLLFEPGYEV-DTRDLP CRTLHNWALYNSDSRMISLEVL
METI : SGSASNMFYIKINEDEIANDYPLPSYYKTSEETDELILLYDASYEV-QSEHLPHRMLHNWALYNSDLRFISLELLP
CMET21 : A---STKYIKINEDEIANDYPLPAYYKTANQETDEYIIFDGGADACYTDDLPRSMMLHNWALYNSDRLISLELLP
CMET5 : S---SNKYIKINEDEIANDYPLPAYYKTANQETDEYIIFDGGADACYTDDLPRSMMLHNWALYNSDRLISLELLP
TMET : S---SSKTYIKINEDEIASDYPLPAYYKTSNEETDEYIVFDGSGVETYHIDELPRSMMLHNWALYNSDRLISLELLP
PMET : S---SNKTYIKINEDEIANDYPLPAYYKTSLOETDEFIVFDNDCDIYDTEDEPSRSMMLHNWALYNSDRLISLELLP
ZMET : S---QKNVYIKISEAEIANDYPLPAYYKPFSEEMDEYI-FDSD-DSIFSDDVPRILNNTLWALYNSDRLISLELIP
```

... continued on next page


```

METIIa : MRPCAEIDVTVFGSGVVAEDDGSGFCLD---DSESSTSTQSNHDGMNIFLSQIKEWMIEFGA-EMIFVTLRTDMA
METIIb : MRPCAEIDVTVFGSGVVAEDDGSGFCLD---DSESSTSTQSNVHDGMNIFLSQIKEWMIEFGA-EMIFVTLRTDMA
METI : MKQCDDIDVNIFGSGVVTDDNGSWISLN---DPS--GSQSHDPGMCIFLSQIKEWMIEFGSDDISISIRTDVA
CMET21 : MKPCAEIDVTIFGSGVMTEDDGSGFNLETDTSHSSSSGSGTANVDGIPIYLSAIKEWMIEFGS-SMVFISIRTDMA
CMET5 : MKGCADIDVTIFGSGVMTEDDGTGFNLDGDTSQSSSAGLGTANVDGIPIYLSAIKEWMIEFGS-SMVFISIRTDMA
TMET : MKACADIDVTIFGSGVMTADDGSGFNFDTDANHSSSGGSRSAEIDGMPIYLSAIKEWMIEFGS-SMIFISIRTDMA
PMET : MKPCAEMDVTIFGSGTMTSDDGSGFNLDTEAGQSSVA-SGAQDTDGIPIYLSAIKEWMIEFGS-SMVFISIRTDLA
ZMET : MKSGAENDVVIFGSGFMRDDDGSCCSTAESVKSSSSSKADQLDAGPIYLSPIKEWIEFGG-SMICITIRTDVA

```

```

METIIa : WYRLGKPSKOYAPWFGTVMKTVRVGISIFNMLRESRVAKLSYANVIKRLCGLENDKAYISSKLLDVERYVVVHG
METIIb : WYRLGKPSKOYAPWFETVMKTVRVVAISIFNMLRESRVAKLSYANVIKRLCGLENDKAYISSKLLDVERYVVVHG
METI : WYRLGKPSKLYAPWWKPVLKTARVGISILTFLRVESRVARLSFADVTKRLSGLQANDKAYISSDPLAVERYLVVHG
CMET21 : WYRLGKPSKOYAPWWEPVLKTARVAISIITILMEQARVSRLSFMDVIKRVSEFEKGHPAYISSIPAVVERYIVVHG
CMET5 : WYRLGKPSKOYASWWEPVLKTARVAISIITILLKEQARVSRLSFMDVIKRVSEFEKGHPAYISSVPAAVERYVVVHG
TMET : WYRLGKPLKOYAPWWEPVIKTARLAVSIITILLKEQNRVARLSFGEVIKRVSEFKKDHPAYISSNVDAVERYVVVHG
PMET : GIGLKPSKOYTPWYDTVLKTARIAISIITILLKEQSRVSRLSFPDVIKKVSEYTQDNKSYISSDPLAVERYIVVHG
ZMET : WYKLROPLKOYAPWCEPVLKTARLAVSIITILLKEQSRASKLSFADVIRKVAEFDKGNPAFISSNITLVERYIVVHG

```

```

METIIa : QIILQLFEEYPDKDIKRCPFVTSLASKMQDIHHTKWIIKKKKILQKGKNLNPRAGIAPVVSRMKAMOATTTRLVN
METIIb : QIILQLFEEYPDKDIKRCPFVTGLASKMQDIHHTKWIIKRKKKILQKGKNLNPRAGLAHVVTRMKPMOATTTRLVN
METI : QIILQLFAVYPDDNVKRCPFVVGLASKLEDRHHTKWIIK-KKISLKELNLNPRAGMAPVASKRKAMOATTTRLVN
CMET21 : QIILQFLEFPDEKIKSAFVAGLTKMERHHTKWLVKKKKILQRDEPNLNPRAAIAPVVSKRKAMOATTTRLIN
CMET5 : QIILQFLEFPDEKIKSAFVIGLTNKMEERHHTKWLNKKKKLLQRDEPNLNPRAALAPVDSKRKAMOATTTRLIN
TMET : QIILQFSEFPDVSIRNCAFAVGLSRKMEERHHTKWVIKKKKVMQRLEQNLNPRASMAPSV-KRKAMOATTTRLIN
PMET : QIILQLFAEFPDDKIRKSPFVTGLMNKMEERHHTKWLVKKKKLSPKSEPNLNPRAAMAPVVSKRKAMOATATKLIN
ZMET : QIILQFADFPDETIRRSAFVSGLLLKMEQRHTKLVMKKTQVMRGE-NLNPSAAMGP-ASRKKAMRATTTRLIN

```

... continued on next page

```

METIIa : RIWGEFYSIYSPEVPS EAINAENV EEELE-----EVEEEDENEEDDPENELEAVEIQNSP--TPKKIKGISEDM
METIIb : RIWGEFYSIYSPEVPS EAIH--EV EEEIE-----EDEEEDENEEDDI E---EAVEVQKSH--TPKKSRGNSEDM
METI   : RIWGEFYSNYS PEDPLOATAA ENGE DEVEEEGGNG EEEVEEENGLEDTVPEPVEVQKPH--TPKKIRGSSGKR
CMET21 : AIWGEFYSNYS PEDMKEGITSDEK EEE-----AEEQEEIDDEEEDDEKETLVALEKTPPTPTSTPRKSKSNSKLK
CMET5  : RIWGEFYSNYS PEDMKEGITG EDK EEE-----PEEQEEI---EEEEKETLTALEKTPPTPTSTPRKTKSIPKVK
TMET   : RIWGEYYSNYS PEVSK E VADCEVKDD EEE-----PDEQEE--NEEDDVPERNLDP EKAHTPSS TRRHIKSRSDSK
PMET   : RIWGEYYSNHL PESK EGTAI EKDDDE-----AEEQEE--NEDEDA EETV-LLEETLKPRIVSKQIKAFSDDG
ZMET   : RIWSDYVAHHF PEDSK EGDGN ETK EID-----DEQEE--NEDEDA EDEGQIEENISKTPPS TRSRKLLSQTC

```

```

METIIa : EIKWDGEILGKTSAGE PLYGRA FVGGDVV VGS AVILEVDDQDDTQLICFV--EFMFESSNHSKMLHGKLLQRGSE
METIIb : EIKWNGEILGETSDG E PLYGRALVGGETVAVGSAVILEVDDPDETPAIYFV--EFMFESSDQCKMLHGKLLQRGSE
METI   : EIKWDGESL G KTSAGE PLYQOALVGGEMVAVGGAVTLEVDDPDEMPAIYFV--EYMFESTDHCKMLHGRFLQRGSM
CMET21 : D VSWNGKPAVKRSS GEMLYKQATLHGNMIAVGGAVLTDDASCLNLP AIYFV--EYMFESSDG-KMIHGRLLQRGSE
CMET5  : DIRWN RKS VGETLSGEALYKQAIYVGT E IAVGGAVLV DDESA-QLPAIYFV--EYMFETLNGIKMLHGRMLQQGSL
TMET   : EINWDGESIGKTASGEOLF KKARVHGHE IAVGDSVLVEHDEPDELGCIYFV--EYMF EKLDGSKMLHGKMMQRGSD
PMET   : EVRWE G VPERKTSSGLPLYKQAIT H GSCFCGN-ICVSRKLMNQMSFLIYITLNICLNPKNGEKMFHGRMMQHGCH
ZMET   : EIRWEGETS GKTLSGE TLYKCA YVREL R I PVGGTVALEDDSGDTV--ICFV--EYMFQKVDGSKM VHGRILOKGSQ

```

```

METIIa : TVLGM AANERELFLINECLTVOLKDIKGTVSL EIRSRP WGHQYRKENIDVKLDRARAEERKTNGLPTDYYCKSLY
METIIb : TVIGT AANERELFLINECLTVHLKDIKGTVSLDIRSRP WGHQYRKENLVVDKLD RARAEERKANGLPTEYYCKSLY
METI   : TVLGN AANERELFLINECM T TOLKDIKGVASF EIRSRP WGHQYRKKNITADKLDWARALERKVKDLPTEYYCKSLY
CMET21 : TVLGN TANE QELFLINECM FEFLMDVKMPVIVEIRSRP WGHQHRKINANADKIDKARAV ERKKNGLETEYYCKSLY
CMET5  : TILGN TANE CEVFLINDC MDFELADVK-KAVEIRSRP WGHQYRKVNANADKIYRAGVE ERKKNGLETEYYCKSLY
TMET   : TVLGN AANEREVFLINECMNLQLGDVKESIAVNI RMP WGHQHR--NTNADKLETAKAEDRKRKGLPTEFYCKSFY
PMET   : TVLGN AA SEREVFLINECRDLGLQDVKQINVASIRKTPWGHQHRKASNAAGKIDRE RADERK KKKGLPTEYYCKSLY
ZMET   : TILGN AANEREVFLINDC L EFKLDDIKELV MVDIQRSP WGHK YRKENSEADKVEQVKAERK KKKGQPMVYFCKSLY

```

... continued on next page

```

METIIa : SPERGGFFSLPRNDMGLGSGFCSSCKIRENEEERSKTKLNDSKTFGLSNGIEYHNGDFVYVLPNYITKDGLKKG-S
METIIb : SPERGGFFSLPRNDIGLGS GFCSSCKIKEEEEERSKTKLNI SKTGVFSNGIEY YNGDFVYVLPNYITKDGLKKGTS
METI : SPERGGFFSLPLSDIGRSSGFCSTCKIREDEEKRSTIKLNVSKTFGFINGIEYSVEDFVYVNPDSI--GGLKEGS-
CMET21 : WPERGAFFSLPVNCMGLGSGICSSCSANKDHTEKEKFSVSSCKTSFVYKGT EYSVHDFLYVSPDQFATERVGOE--
CMET5 : CPDKGAFFSLPLNSMGLGSGICSSCKLDKDLTEKEK FVHSDKTSFVFNGT EYSIHDFLYVSPQOFSTERVGN E--
TMET : RPEKGAFFRLPFDKMG LCNGLCYSCETQOTDQEKESFKFDMSKSSFVYLGTEYSVDDFVYVSPDHFTAERGNG--
PMET : WPERGAFFSLPFDTLGLGSGVCHSCNIQ EADKAKEIFKVNSSKSSFVLDGTEYS LNDYVYVSPFEFE-EKLEQG--
ZMET : WPEKGAFFALSRDKMGLGSGICSSCDNI EPDSDELKI ---FSKTSFVYRKVTYNVNEFLYIRPDFFAED-----E

```

```

METIIa : RRTTLKCGRNVGLKAFVVCQLLDVIVLEESRKASKASFOVKLTRFYRPEDISEEKAYASDIQELYYSQDTYILPPE
METIIb : RRTTLKCGRNVGLKAFVVCQLLDVIVLEESRKASNASFOVKLTRFYRPEDISEEKAYASDIQELYYSHDTYILPPE
METI : -KTSFKSGRNI GLRAYVVCOLLEIVPKESRKADLGSFDVKVRRFYRPEDVSAEKAYASDIQELYFSQDTVVLPPG
CMET21 : ---TFKGGRNVGLKAF AICOLLEVVVPKKPOQADDSSTE VKVRRFYRPEDISDEKAYCSDIREVYYS EETHLLVE
CMET5 : ---TFKGGRNVGLKAY AICOLLEIIVPKAPKQAEPHSTEIKVRRFYRPEDISDEKAYCSDIREVYYS EETHIDAE
TMET : ---TFKAGRNVGLMAYVVCOLLEIVGPKGSKQAKVDS TNVKVRRFFRPEDISSDKAYSSDI REIYYS EDIHTVPE
PMET : ---THKSGRNVGLKAFVVCVLEI IAKKETKQAEIKSTELKVRRFFRPEDVSS EKAYCSDVQEVYFSDETYTISVQ
ZMET : DRATFKAGRNVGLKPYAVCQLLS IPEGAGSKKLNPA SANI SARRFYRPDDISSAKAYASDI REVYYS EDVIDVPE

```

```

METIIa : AIOGKCEVRKKS DMPLCREYPI LDHIFFC EYFYDSSTGYLKQFPANMKLKFSTIK--DETLLE
METIIb : AIOGKCEVRKKN DMPLCREYPI LDHIFFC EYFYDSSTGYLKQFPANMKLKFSTIK--DETLLE
METI : AIEGKCEVRKKS DMPLSREYPI SDHIFFC DLFDDTSKGS LKQLPANMKPKFSTIK--DDTLR~
CMET21 : AIEGRCEVRKKS DLPTCDAPTI YEHV FYCEYLYDPHKGSLKQLPSNIKLRYSTVKAYDSSLR~
CMET5 : TVEGRCEVRKKN DLPSCDAPTI FDHVF FCEYLYDPAKGS LKQLPPNIKLRYSAVKGAHVSSLR~
TMET : IIEGKCEVRKKY DISSEDVPAMFDHIFFC EYLYDPLNGSLKQLPAQINLILSKIKLDDATSR~~
PMET : SVEGKCEVRKKI DIPEGSAPGAFHN VFFCELLYDPATGSLKQLPSHIKVKYSSGPTADNAAR~~
ZMET : MIEGKCEVRKKN DLASSDLVMEHVFFCELLYDRASGALKQLPPNVRF-MSMVQRTSAL~~~

```

motifs is conserved perfectly in METI and partially in the other plant methyltransferases (Figure 6.8). The S/TPXX motif does not occur in the amino-terminal domains of the CMT enzymes. The S/TPXX motif occurs eleven times in the amino-terminal domain of the mouse enzyme, and in six cases there is at least one basic residue in the X position. Sequences resembling nuclear localization signals (reviewed by Garcia-Bustos *et al.*, 1991) are found throughout the METIIa amino-terminal domain, especially towards the 5' end.

The zinc binding domain seen in the amino-termini of the mouse (and human) Dnmt1 (DNMT1) (Bestor *et al.*, 1988; Yen *et al.*, 1992) is likely to bind DNA, perhaps being involved in targeting hemimethylated DNA (Bestor, 1992). This domain is lacking from METIIa and all other plant methyltransferases analyzed; in both METI and METIIa there is a deletion at that point. In METIIa, an acidic region is found near this deletion, containing 18 glutamic acid residues and 3 aspartic acid residues in 31 aa (residues 639 to 669). METI and the other plant methyltransferases analysed also contain this acidic region (Figure 6.8). A short basic region occurs upstream; in METIIa this region contains 8 lysine residues and 2 histidine residues in 18 aa (residues 573 to 590). The acidic region is not found in the mouse Dnmt1 or human DNMT1, though a short basic region occurs in both enzymes (residues 718 to 730 in Dnmt1).

Dnmt1 is targeted to replication foci during S phase by a 249 aa region in the amino-terminal domain (Leonhardt *et al.*, 1992). Dot plot analyses indicate that a stretch of 150 aa within this region shows homology in the amino-terminal domain of both METIIa (Figure 6.9) and METI (Finnegan and Dennis, 1993).

6.3.7 Chromosomal location of *METIIa*

Caroline Dean (John Innes Institute, Norwich) provided data for 68 markers covering all 5 Arabidopsis chromosomes, scored on a set of RI lines from a cross between Columbia and Landsberg *erecta*. An RFLP for *METIIa* was scored on 53 of these RI lines (Figure 6.10). *METIIa* was linked to a group of markers on chromosome IV (Figure 6.11).

6.3.8 Expression pattern of *METI* and *METIIa*

(a) Screening cDNA libraries

(i) With *METIIa* 3' probe

A 1.0 kb *HincII-HindIII* fragment of Ac71 was used to screen libraries 1 and 2. This probe was chosen because it contains the 3' end of the *METIIa* ORF (Figure 6.1), and thus would hybridize to incomplete as well as full length *METIIa* cDNAs. This region is 84 % identical to *METI* over 372 bp. Two positive clones were isolated from each library, and designated ZL1 and ZL3 (library 1 clones), and YES6 and YES12 (library 2

Figure 6.9: Dotplot comparison of the mouse Dnmt1 (horizontal axis) and Arabidopsis METIIa (vertical axis) amino terminal domains. Residues 1 to 1111 from Dnmt1 were compared to residues 1 to 1049 from METIIa; for both peptides, this includes all residues up to the glycine-lysine repeats separating the amino terminal and methyltransferase domains. The dotted lines indicate the region of the Dnmt1 amino terminal that targets the enzyme to the replication fork. The window size is 50 residues and the stringency of the match is 30.

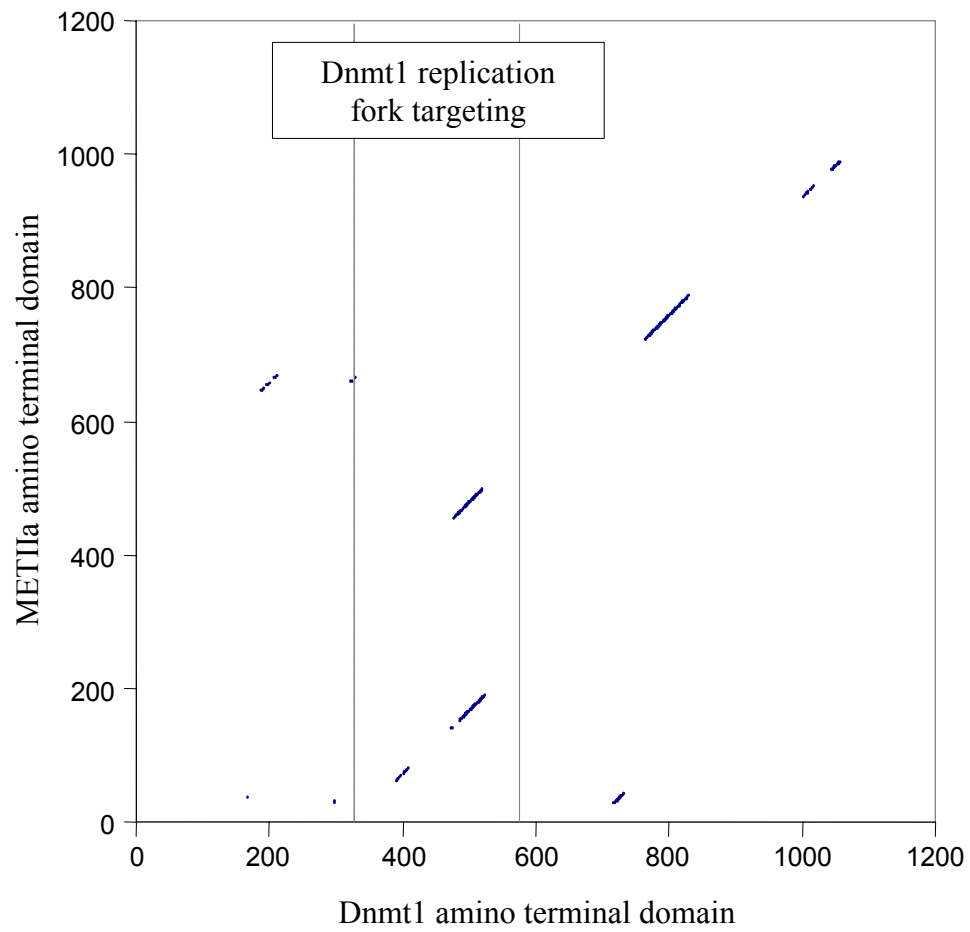
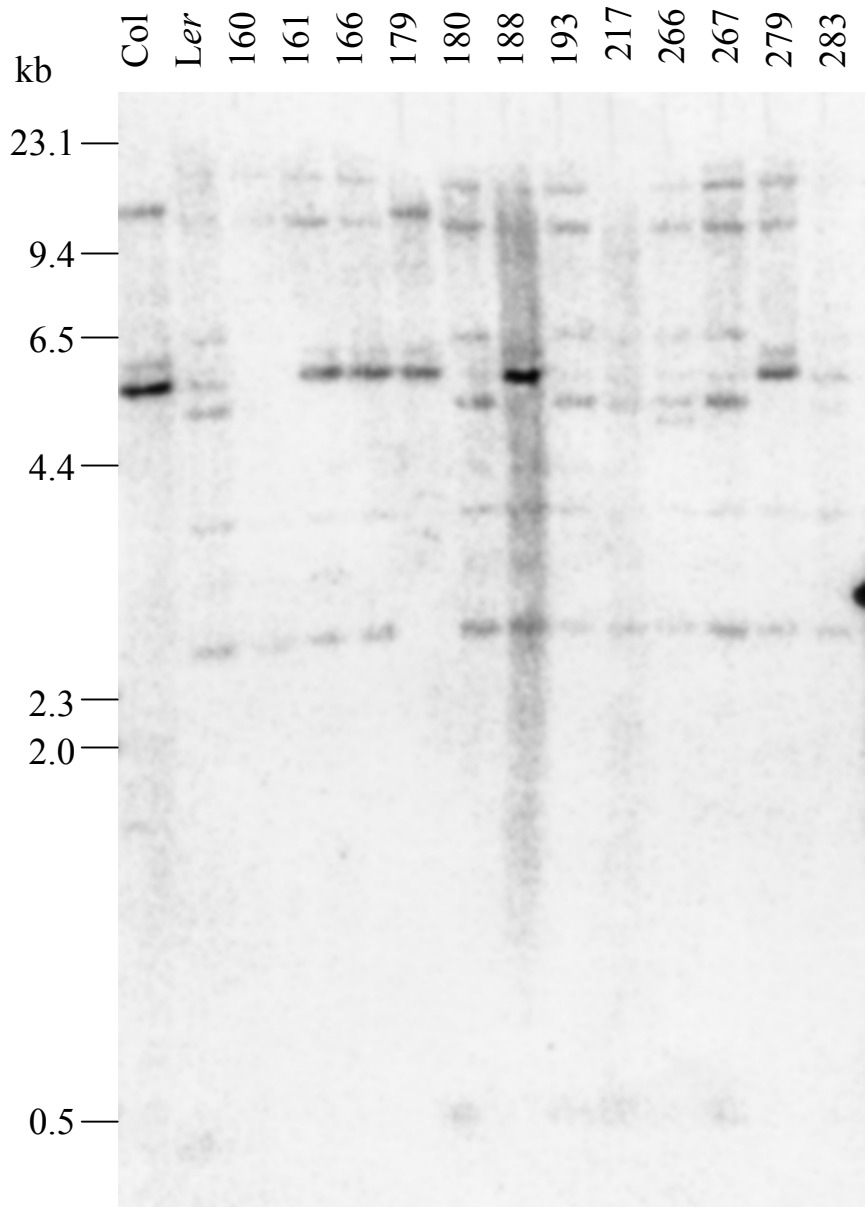


Figure 6.10: Detection of an RFLP for Ac71. Arabidopsis genomic DNA was digested with *Bgl*II and hybridized with the 1.6 kb *Xho*I-*Hind*III fragment at the 3' end of Ac71 (probe 5, Figure 6.1). The first two lanes contain Columbia (Col) and Landsberg *erecta* (*Ler*) DNA, and show a polymorphism between these two ecotypes. The remaining twelve lanes contain DNA of a subset of 53 recombinant inbred (RI) lines from a Columbia-Landsberg *erecta* cross which were scored for this polymorphism. Lines 180, 193, 217, 266 and 267 were scored as possessing the *Ler* pattern at this locus, while lines 161, 166, 179, 188 and 279 were scored for the Col pattern. Southern analysis of lines 160 and 283 was repeated; no clear pattern was seen for line 160, but line 283 showed the Col pattern.



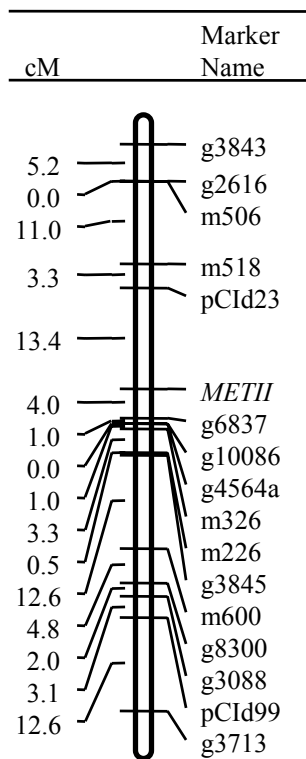


Figure 6.11: Linkage map of Arabidopsis chromosome 4, showing position of *METII* relative to 16 markers scored on the same set of recombinant inbred lines (data provided by Caroline Dean, John Innes Institute, Norwich). The map was generated using MapMaker Macintosh Version 2.0.

clones). Sequencing showed that none of these clones corresponded to any region of the *METIIa* ORF, but for ZL1 and ZL3 there was good sequence homology with a region 3' of the *METIIa* ORF. These clones were partially sequenced, and found to originate from an ORF 3' of *METIIa*, which is transcribed in the opposite direction to *METIIa* (Figure 6.12). A short region sequenced from the YES6 clone also has high homology with this region (Figure 6.12). The polyA tails of the ZL1 and ZL3 clones start at different points, with the ZL1 transcript extending 20 bp further than ZL3; possibly this ORF has multiple polyadenylation sites. Preliminary sequence of the ends of 2 YES12 subclones, totalling approximately 1.3 kb, showed that YES12 was identical to *METI*.

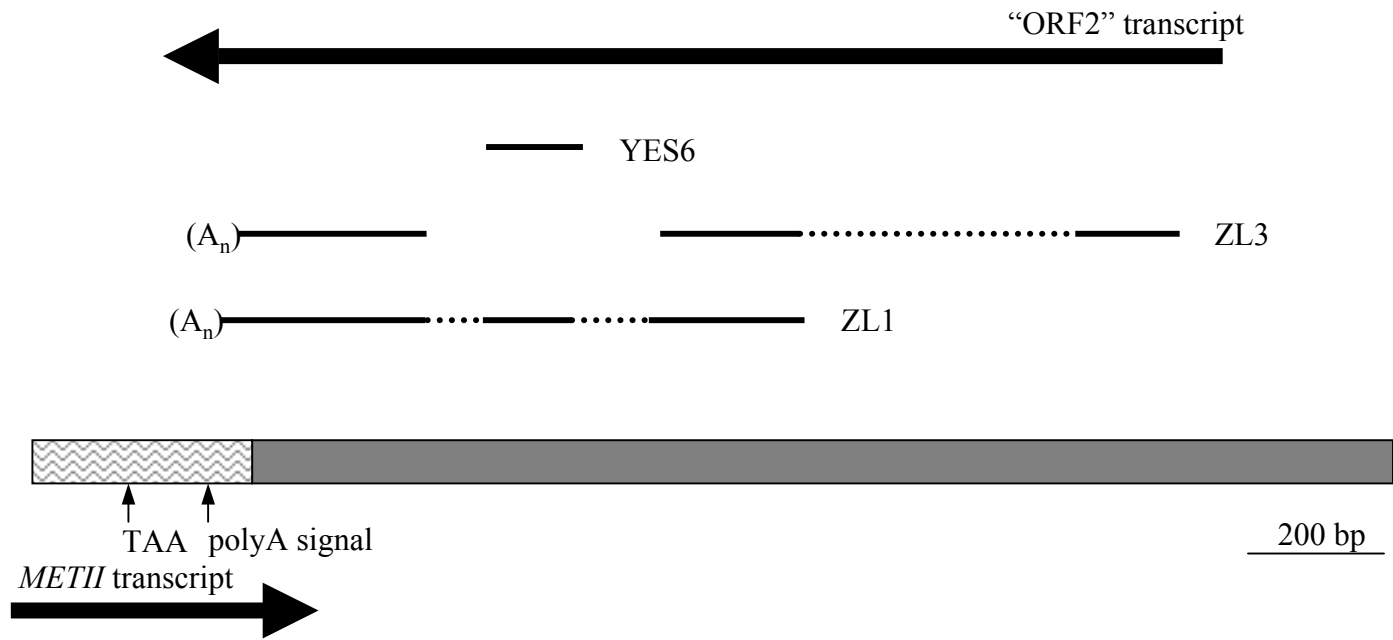
(ii) With *METI* active site probe

All four libraries were screened extensively with a region of *METI* containing motif IV, but no positive clones were obtained from libraries 3 and 4. Libraries 1 and 2 each yielded 4 positive clones. Sequencing revealed that all these clones were identical to *METI*.

(iii) With Ac71-specific primers

Libraries 1 and 2 were screened using primers that flank *METIIa* intron III (Figure 6.2). No product could be amplified from the *METI* cDNA clone Y8, which contains sequence corresponding to the amplified region of *METII* (Figure 6.13). Positive controls were the 1.2 kb *HindIII*-*HindIII* clone and intact Ac71 phage. To determine the sensitivity of the

Figure 6.12: DNA sequence downstream of the *METIIa* gene, containing a second open reading frame (ORF2). The genomic sequence is represented by a thick bar; the lightly shaded region of this bar represents the 3' end of the *METIIa* gene, which was sequenced as described in this chapter. The remainder of the genomic sequence shown here was derived from the sequence of a region of Arabidopsis chromosome IV (Genbank accession AL021711). Regions of the cDNA clones YES6, ZL1 and ZL3 which were sequenced are shown by solid lines; this sequence is shown aligned to the genomic sequence.



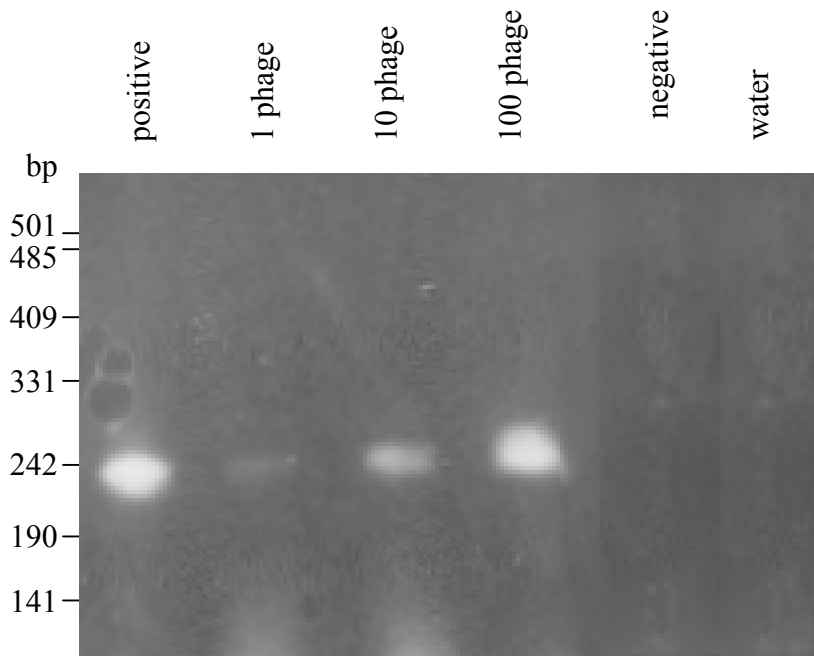


Figure 6.13: Products of PCR with primers specific to *METIIa*, separated on a 4% polyacrylamide gel. The 1.2 kb *HindIII-HindIII* fragment of Ac71 (Figure 6.1) was used as the positive control, while the Yc8 cDNA clone of *METI* was used as the negative control. Products of PCR using serial dilutions of the Ac71 phage in a flax cDNA library as template are shown; inclusion of only a single phage in the reaction yields a faint product.

screen, PCR reactions were performed using serial dilutions of the Ac71 phage in a flax cDNA library (provided by Dr Jean Finnegan) as template. When the Ac71 phage template was diluted so that only a single phage was added to the PCR reaction, a faint visible product was obtained. Despite extensive screening (Section 6.2.6(b)), no positive signals were obtained from either library.

(b) RT-PCR

Expression of both *METI* and *METIIa* was assayed by RT-PCR in vegetative and floral tissues, and in whole plants or above ground tissues of varying ages (Figure 6.14, Figure 6.15). Expression of both genes was detected in all tissues assayed, and in all cases *METI* transcript levels greatly exceeded *METIIa* levels. A rough estimate of the ratio of *METI* and *METIIa* transcripts was made on the basis of the amount of product from RT-PCR: on average, *METI* expression was 10000 fold higher than *METIIa* expression, though in siliques, the ratio dropped to approximately 5000 fold. Expression of *METIIa* in plants expressing an antisense transcript of *METI* (Finnegan *et al.*, 1996) was assayed by RT-PCR. The *METIIa* gene is expressed at wildtype levels in these plants (Figure 6.14, Figure 6.15).

Figure 6.14: RT-PCR products for *METI* and *METII*.

(A) and (B) Products of *METI* RT-PCR reactions on RNA from different plant tissues (A) and whole plants of different ages (B). The most abundant product is the expected 205 bp band. A larger band which corresponds to the 450 bp band expected from genomic DNA appears in a number of lanes, and is likely to be a product of genomic DNA contaminants in the original RNA sample. Tissues showing the highest expression levels are leaves, flower buds and flowers; expression is slightly lower in siliques, and much lower in roots. Plant age also affects expression, increasing from a very low level in young seedlings to peak in 21 and 25 day old plants; expression in 30 day old plants is very low. The positive control is the Yc8 cDNA clone of *METI*, which includes the methyltransferase domain sequence (Finnegan and Dennis, 1993). The negative control is a 1.2 kb *HindIII* fragment of Ac71 (Figure 6.1).

(C) and (D) Products of *METII* RT-PCR reactions on RNA from different plant tissues (C) and whole plants of different ages (D). The expected 493 bp product is seen in the positive control lane, Ac71, and a band of the expected cDNA product size, 277 bp, appears in lanes containing RT-PCR reactions. Expression is fairly consistent throughout plant tissues, though, as for *METI*, lower expression is seen in roots. As seen for *METI*, expression of *METII* increases from a very low level in seedlings to peak in 21 and 25 day old plants, and is undetectable in 30 day old plants. The positive control is a 1.2 kb *HindIII* fragment of Ac71 (Figure 6.1). The negative control is the Yc8 cDNA clone of *METI*, which includes the methyltransferase domain sequence (Finnegan and Dennis, 1993).

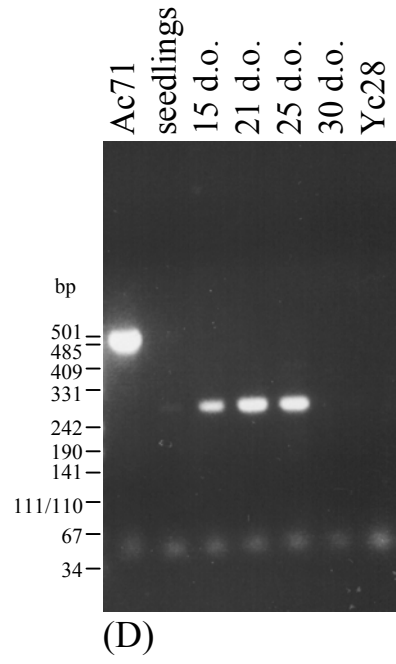
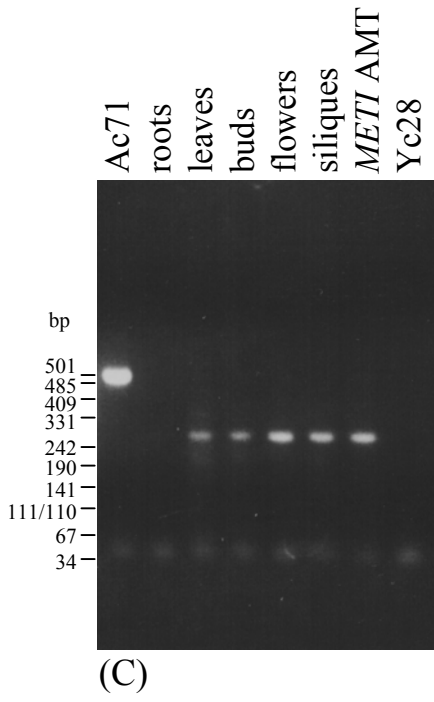
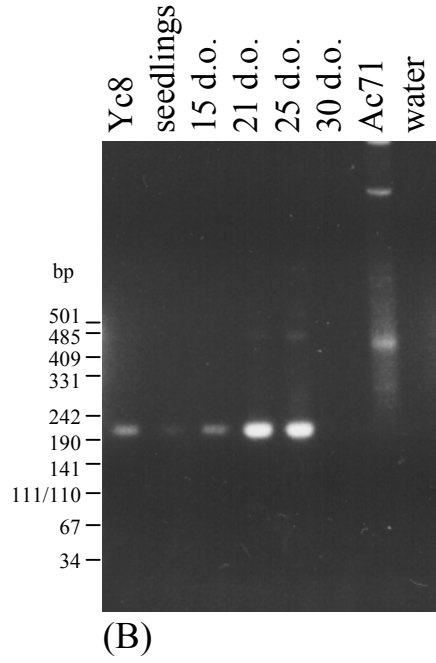
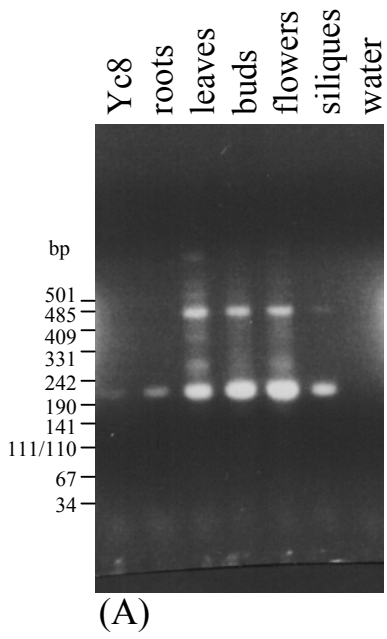
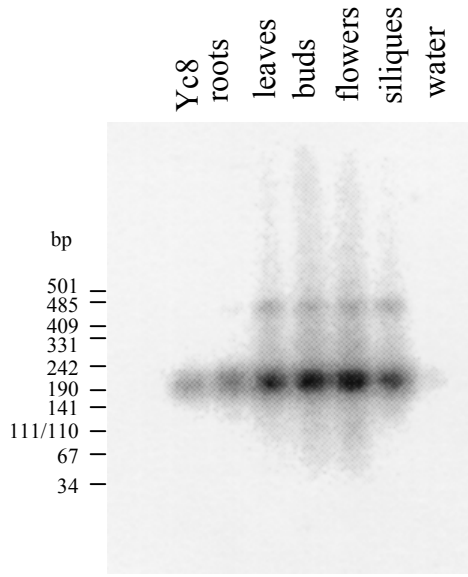


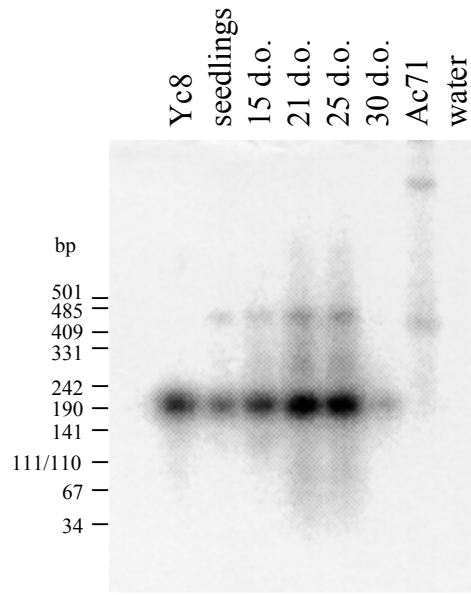
Figure 6.15: Southern analysis of *METI* and *METII* RT-PCR gels shown in Figure 6.14.

(A) and (B) The probe used was the 205 bp product amplified from the Yc8 cDNA clone of *METI* using the same primers as those used in the RT-PCR reaction.

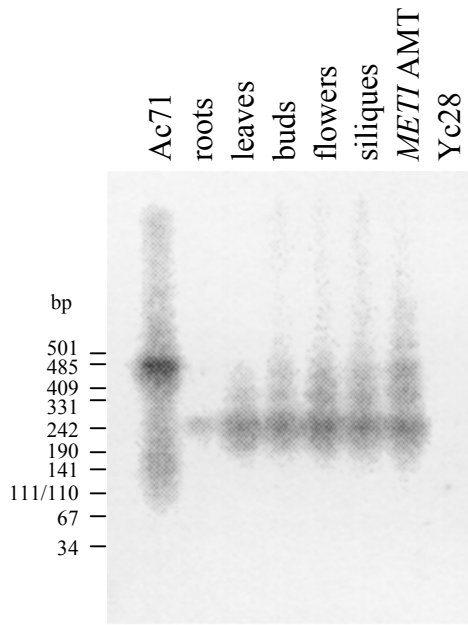
(C) and (D) The filters were probed using the 493 bp product amplified from the Ac71 genomic clone with the same primers as those used in the RT-PCR reaction. A faint band of the same size (493 bp) as that seen in the genomic control lane is seen in most lanes, indicating contamination of the original RNA preparations with genomic DNA.



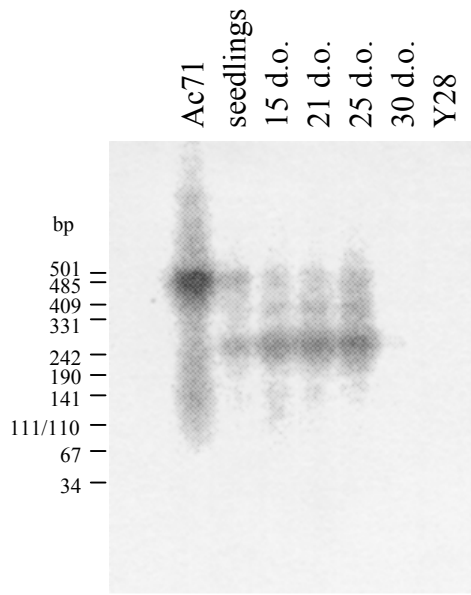
(A)



(B)



(C)



(D)

(c) Screening expressed sequence tag databases

The Genbank EST database was screened using the *METIIa* genomic sequence to search for Arabidopsis ESTs. Three Arabidopsis ESTs were identified with homology to *METIIa*; all three were identical to *METI* sequence.

6.4 Discussion

The *METI* and *METIIa* genes have a common structure, of which the most notable aspect is the perfect conservation of the position of the ten introns within the coding sequence. Four of these introns split conserved motifs I, IV, VII and IX, and two split the TRD. An eleventh intron has been identified for both genes, 11 nucleotides upstream of the start codon. The sequence of this intron is 72% identical between *METI* and *METIIa* (Genger *et al.*, 1999). Throughout the coding region, sequence identity is high at both the nucleotide and amino acid level, and the *METI* predicted protein is only slightly longer than the *METIIa* predicted protein. The amino acid sequence of the methyltransferase domain is better conserved than that of the amino-terminal domain; the similarity between the nucleotide sequences does not reflect this, being fairly uniform throughout. This indicates that base changes in the methyltransferase domain have been restricted to the third position of the codons more so than in the amino-terminal domain. This restriction is likely to be due to the need to retain a functional methyltransferase protein.

There is good reason to believe that *METI* is a functional cytosine methyltransferase. Not only does *METI* show sequence and structural similarity to other cytosine methyltransferases, but also, antisense expression of *METI* in *Arabidopsis* causes cytosine demethylation (Finnegan *et al.*, 1996). The high degree of similarity between *METI* and *METIIa* in both gene structure and sequence, and especially the presence of the eight motifs characteristic of cytosine methyltransferases, suggest that *METIIa* (AF138283) is also a cytosine methyltransferase. Protein structure modelling of *METI* and *METIIa*, based on the structure of the bacterial cytosine methyltransferase *M.HhaI* (Cheng *et al.*, 1993; Klimasauskas *et al.*, 1994), suggested that the tertiary structures of both proteins are identical to that of *M.HhaI* in and around the conserved motifs. The crystal structures of *M.HhaI* and *M.HaeIII* reveal that both these enzymes gain access to the target cytosine by rotating it out of the DNA helix (Cheng *et al.*, 1993; Klimasauskas *et al.*, 1994; Reinisch *et al.*, 1995). The predicted conservation of tertiary structure in *METI* and *METIIa* may indicate conservation of the methylation mechanism of prokaryote methyltransferases in these plant enzymes.

Finnegan and Dennis (1993) suggested the existence of a small family of cytosine methyltransferases in *Arabidopsis*, based on the multiple bands seen in low stringency Southern analysis using the *METI* cDNA as a probe. *METIIa* is one of these methyltransferases; recently, two more genes of this family were identified by searching *Arabidopsis* sequence databases (Genger *et al.*, 1999). Both these genes are more similar to *METIIa* than *METI*, and both, like *METIIa*, are located on chromosome IV. The predicted peptide for one of these genes, *METIII* (AL049656), truncates after 64 aa due to

a single base change creating a stop codon, and a single base deletion causes a frameshift at around amino 243 (Genger *et al.*, 1999). The other gene, *METIib* (AC005359) appears to encode an intact methyltransferase, and is over 90% identical to *METIia* (Genger *et al.*, 1999). It appears that *METIia*, *METIib* and *METIII* result from two recent duplication events, while the ancestor of these three genes and *METI* probably originated from an earlier duplication followed by movement to chromosome V.

METI, *METIia* and *METIib* all possess a large amino-terminal domain with similarity to the amino-terminal domain of Dnmt1, which confers a preference for hemi-methylated substrates (Bestor, 1992); thus these methyltransferases are likely to function in maintenance methylation. Expression of an antisense construct against *METI* causes demethylation of cytosines in CG and CCG but had a lesser effect on methylation in CAG and CTG sequences, suggesting that *METI* is a CG maintenance methyltransferase (Finnegan *et al.*, 1996). Ronemus *et al.* (1996) found that *METI* is expressed ubiquitously, but at higher levels in meristematic tissue. The two *METI*-like carrot methyltransferases are also expressed at high levels in meristematic tissue (Bernacchia *et al.*, 1998a). These observations are consistent with the hypothesis that *METI* is a maintenance methyltransferase, as this activity is required in dividing cells after DNA replication.

METIia and *METIib* are identical in the region used to assay for *METIia* expression (Genger *et al.*, 1999), so the expression observed could be derived from either gene. Assays specific to each gene are needed in order to determine their relative expression

levels and patterns. In most tissues, the expression level of the *METII* genes is roughly 10000 times lower than that seen for *METI*; this probably indicates that *METI* is the predominant methyltransferase of this class. The expression of the *METII* genes is not disrupted in *METI* antisense transgenics. This could suggest that the *METII* genes do not replicate the function of *METI*; however, it is also possible that the low level of expression of these genes prevents them from substituting for *METI*. Similarity between the *METI* and *METIIa/METIIb* genes is lower in the target recognition domain (TRD) than in the conserved motifs, so it is possible that *METIIa* and *METIIb* methylate cytosines in a sequence context other than the CG or CCG sequence targeted by *METI*, such as CNG. Alternatively, the conserved residues in the TRD may be those important in sequence recognition.

Why is the level of expression of the *METII* genes so low, compared to that of *METI*? Levels of cytosine methylation in CG and CNG sequences appear to be roughly similar in plant DNA (Gruenbaum *et al.*, 1981), so if the *METII* genes target CNG sequences, they should be expressed as highly as *METI*. It is possible that *METIIa* and *METIIb* methylate cytosines in very specific sequence contexts at low frequency, and therefore are not required at high levels. No phenotype has yet been found for *METIIa* antisense plants (K.A. Kovac, personal communication), which should also be disrupted in *METIIb* function. This may indicate that these genes are not essential; however, further analysis of *METIIa* antisense plants is in progress, to confirm the activity of the *METIIa* antisense construct.

Alignment of the methyltransferase domains of plant and mammalian methyltransferases revealed a high degree of similarity among all these enzymes throughout the whole domain (Figure 6.7). The alignment also revealed consistent differences between the plant and animal methyltransferases. Of these, the most significant is a 40-41 residue deletion (relative to mammalian proteins) in the TRD of the plant methyltransferases, which may indicate differences in the target sequences of plant and mammalian methyltransferases. The similarity seen in the TRD of eukaryote methyltransferases may reflect the fact that most cytosine methylation in eukaryote genomes occurs within CG or CNG sites (Gruenbaum *et al.*, 1981; Clark *et al.*, 1995). It has recently been shown that cytosine methylation also occurs at high frequency within non-symmetrical sites in plant DNA (Meyer *et al.*, 1994; Oakeley and Jost, 1996, Wang *et al.*, 1996), so an alternative explanation for this low variability is that the methyltransferase genes so far isolated target only cytosines within CG or CNG sites.

Certain characteristics of the amino-terminal domain are conserved between plants and mammals, indicating that this domain may fulfill similar roles in the mammalian and plant enzymes. These characteristics include regions which may target the replication fork, nuclear localization signals and DNA binding motifs (Leonhardt *et al.*, 1992; Finnegan and Dennis, 1993; Bernacchia *et al.*, 1998a, 1998b; Pradhan *et al.*, 1998). The zinc binding domain found in the amino-terminal domains of the mammalian enzymes is also likely to bind DNA (Bestor, 1992). This domain is not found in METI, METIIa, METIIb, or any other plant methyltransferase gene so far described. An acidic region occurs at a similar point in all plant methyltransferases except the chromo-

methyltransferases (Henikoff and Comai, 1998). The function of this acidic region is unknown. *In vitro* expression and assays of methyltransferase activity of the *MET1* class of enzymes, whether from Arabidopsis or other plants, may allow the function of the acidic region to be determined, by deleting it from the expression construct and observing the effect on methyltransferase activity and/or specificity.

The chromo-methyltransferases represent a second class of Arabidopsis cytosine methyltransferases. The two genes so far identified from this class, *CMT1* and *CMT2*, contain chromodomains and have little similarity to the *MET1* class of methyltransferases, except within the conserved motifs (Henikoff and Comai, 1998; Genger *et al.*, 1999). Chromodomains target proteins to heterochromatin (Paro and Harte, 1996), so it is likely that genes from this class methylate cytosines within heterochromatin, or have a role in the condensing of DNA into heterochromatin.

It is apparent from the above discussion that much further study will be needed to ascertain the functions of the multiple cytosine methyltransferase genes already isolated. Arabidopsis may contain even more cytosine methyltransferase genes, as Southern analysis indicates the presence of one more *MET1*-like sequence and possibly many more *CMT* class sequences (Finnegan and Dennis, 1993; Henikoff and Comai, 1998; Genger *et al.*, 1999). Other classes of methyltransferase with similarity to the *de novo* methyltransferases from mouse (*Dnmt3*) and *Ascobolus* (*Masc1*) may also be found. Such methyltransferases could be responsible for silencing repeated sequence in the genome, such as transposable elements, or for epi-mutations such as the *clark kent*

mutations, caused by non-sequence specific hypermethylation of the *SUPERMAN* gene (Jacobsen and Meyerowitz, 1997). Analysis of the phenotypes of knockout mutants or antisense transformants for these genes, more detailed analysis of expression patterns, and *in vitro* assays of protein function will be essential to gain an understanding of how methylation patterns are set, changed and maintained, as will a better understanding of the processes, whether active or passive, of cytosine demethylation.