The EMBO Journal vol.13 no.9 pp.2084–2088, 1994

Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*

Peter Meyer[1], Ingrid Niedenhof and Michael ten Lohuis

Max-Delbrück-Laboratorium in der MPG, Carl-von-Linné Weg 10, D-50829 Cologne, Germany
[1]Corresponding author

Communicated by H.Saedler

A considerable proportion of cytosine residues in plants are methylated at carbon 5. According to a well-accepted rule, cytosine methylation is confined to symmetrical sequences such as CpG and CpNpG, which provide the signal for faithful transmission of symmetrical methylation patterns by maintenance methylase. Using a genomic sequencing technique, we have analysed cytosine methylation patterns within a hypermethylated and a hypomethylated state of a transgene in *Petunia hybrida*. Examination of a part of the transgene promoter revealed that in both states m^5C residues located within non-symmetrical sequences could be detected. Non-symmetrical C residues in the two states were methylated at frequencies of 5.9 and 31.9%, respectively. Methylation appeared to be distributed heterogeneously, but some DNA regions were more intensively methylated than others. Our results show that at least in a transgene, a heterogeneous methylation pattern, which does not depend on symmetry of target sequences, can be established and conserved.

Key words: DNA methylation/gene silencing/genomic sequencing/*Petunia hybrida*

Introduction

DNA methylation has been associated with several cellular processes, including gene expression (Doerfler, 1983), chromatin organization (Bird, 1986), transposition (Schwarz and Dennis, 1986) and the inheritance of epigenetic states (Holliday, 1987). It is still a matter of debate whether DNA methylation is involved in these processes as a regulatory or a responsive mechanism. The function and control of DNA methylation in eukaryotes are still unclear. Animal, plant and fungal genomes differ in m^5C content and distribution, which might indicate distinct or multiple functions for DNA methylation in different eukaryotic systems.

In mammals the importance of DNA methylation for embryonic development has been convincingly demonstrated. A mutation of the murine DNA methyltransferase that led to a 3-fold reduction in levels of genomic m^5C residues caused abnormal development and lethality of embryos (Li *et al.*, 1992). Further characterization of mutant mice deficient in DNA methyltransferase activity showed that a normal level of DNA methylation is required for controlling differential expression of the paternal and maternal alleles of imprinted genes (Li *et al.*, 1993). In contrast DNA methylation mutants in *Arabidopsis thaliana* do not exhibit adverse effects on plant development or viability. Although in the DDM1 mutant m^5C levels are reduced >70%, the mutant develops normally and exhibits no striking morphological phenotype (Vongs *et al.*, 1993). A methylation-deficient mutant in the filamentous fungus *Neurospora crassa* is also viable, but frequently creates duplications of chromosomes or chromosomal parts (Foss *et al.*, 1993) suggesting a role for DNA methylation in the control of chromosomal behaviour.

In animals, 2–8% of total cytosines are methylated with CpG dinucleotides being the preferred targets for DNA methylation (Doerfler, 1983). In plants, up to 30% of cytosines are methylated (Adams and Burdon, 1985). Methylated C residues are located within CpG or CpNpG sequences, respectively (Gruenbaum *et al.*, 1981). It has been proposed that methylation of symmetrical sequences provides a signal for a maintenance methylase that methylates C residues in a newly synthesized strand, if the opposite strand carries a m^5C residue in the complementary sequence. The suggestion that symmetrical methylation patterns are transmitted in a semi-conservative fashion (Holliday and Pugh, 1975; Riggs, 1975) has obtained substantial support from the fact that hemimethylated DNA is the preferred substrate for maintenance methylase (Gruenbaum *et al.*, 1982).

Filamentous fungi contain a very low content of ~1.5% m^5C residues (Russell *et al.*, 1987). In *N.crassa* (Selker, 1990) and *Ascobolus immersus* (Rhounim *et al.*, 1992), however, methylation of C residues that do not belong to symmetrical sequences has been observed. Genomic sequencing revealed that non-symmetrical DNA methylation in *N.crassa* was present in a gene exposed to repeat-induced point mutation (Selker *et al.*, 1993).

We analysed part of the promoter region in two epigenetic variants of a transgenic *Petunia* line by genomic sequencing. The two variants carry differentially methylated states of an A1 transgene that is expressed in one variant, but inactive in the other. We observe non-symmetrical cytosine methylation at different intensities in the two states. Our data show that cytosine methylation of a plant transgene is not confined to CpG and CpNpG sequences.

Results

Experimental design

For genomic sequence analysis, we used two lines that both derived from the same transgenic petunia plant, Rl01–17 (Pröls and Meyer, 1992). This transformant contains one copy of the maize A1 gene, encoding dihydroflavanol reductase (DFR), an enzyme that induces a new pigmentation pathway in *Petunia hybrida* (Meyer *et al.*, 1987). A1 activity

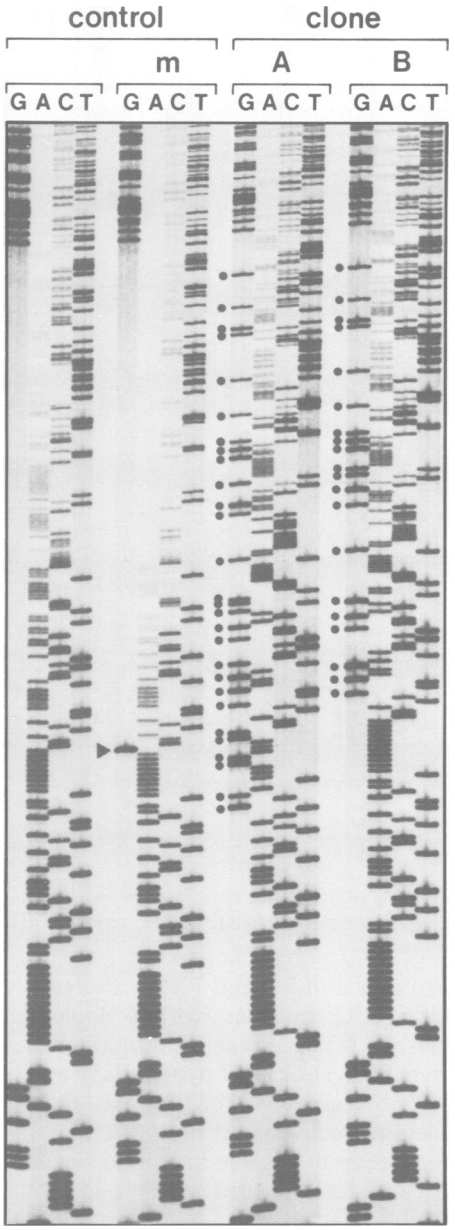

Fig. 1. Examples of sequences of the lower strand of the 35S promoter region between positions −276 and −44, relative to the transcription start site. For the bisulfite-treated DNA, the sequence of the complementary strand is shown. 'control' shows the sequence of unmethylated plasmid and plasmid that had been methylated *in vitro* with *Hae*III methylase. The arrow indicates the G residue that marks the methylated complementary C residue of the GGCC recognition sequence of *Hae*III. All other C residues of the control DNA have been converted into T residues due to bisulfite treatment and PCR amplification, as indicated by the empty G lane between the two internal primers. Clones A and B show the sequence of two clones from bisulfite-treated DNA '17 white'. Methylated C residues are represented by signals in the G lane that are indicated by dots. In the 'control m' lane the A block at the bottom of the gel lacks two nucleotides, which is due to a difference in primer sequences (see Materials and methods).

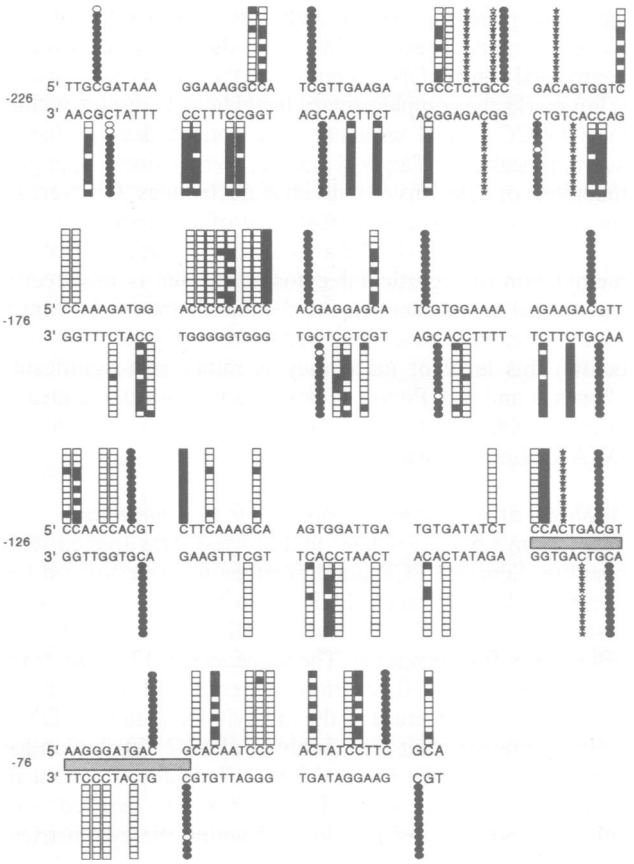

Fig. 2. Distribution of methylated C residues in the −226 to −44 region of the 35S promoter in line '17 white'. Data from 12 clones, representing individual cells have been compiled. The hatched box refers to the 21 bp element that contains the two TGACG binding motifs for the activation sequence factor. Methylated C residues are labelled with filled symbols. Circles mark CpG residues, asterisks indicate CpNpG residues. Other C residues are marked by boxes.

can be detected by a brick red pigmentation of the flower. Line '17 red' displays intensive pigmentation, while in line '17 white', a heterozygous derivative of line 17-W (Meyer *et al.*, 1993), no pigmentation is observed, due to a lack of A1 transcription. The A1 construct is driven by the 35S promoter of cauliflower mosaic virus, a strong constitutive promoter frequently used for plant transgenes. For an analysis of DNA methylation we selected the region between positions −226 and −44, relative to the transcription start site. The −90 to −46 A1 domain (Benfey *et al.*, 1990), which is required for maximal expression of the promoter, harbours a 21 bp element containing two TGACG motifs that bind the activation sequence factor (ASF-1) (Lam *et al.*, 1989). The −226 to −90 region contains domains B1, B2 and B3 with multiple *cis*-elements that confer synergistic interactions (Benfey *et al.*, 1990).

For analysis of DNA methylation patterns, we used a genomic sequencing protocol that is based upon bisulfite modification and subsequent PCR amplification of single-stranded DNA (Frommer *et al.*, 1992). Bisulfite treatment converts unmethylated C residues into U residues that appear as T residues in the PCR product. All m^5C residues remain unaltered. As a control for complete conversion of unmethylated C residues into U residues DNA of plasmid p35A1 containing the 35S promoter was used. Both unmethylated plasmid DNA or DNA that had been methylated *in vitro* by *Hae*III methylase were treated in a slightly modified bisulfite reaction (see Materials and methods for details). Representative sequences for the control reactions are shown in Figure 1. For the unmethylated control DNA (control) an empty G lane is detectable between

the two primers, which indicates conversion of all complementary C residues into T residues. For the *in vitro*-methylated control DNA (control m) the only G residue that is left marks the complementary methylated C residue within the GG^{m}CC *Hae*III sequence. We noticed that, probably due to mistakes of Taq polymerase, some clones contained mutations or deletions of individual nucleotides. On average such errors occurred at a frequency of approximately one in 200 nucleotides, which leaves a 0.5% probability that the methylation of a particular cytosine residue is incorrectly determined by this method. For the interpretation of the high methylation frequencies observed in the transgene DNA (see below) this level of inaccuracy is rather non-significant. Clones A and B in Figure 1 show examples of the sequence of clones obtained after PCR amplification of bisulfite-treated DNA of line '17 white'.

Analysis of m^{5}C distribution in the two variants

DNA from lines '17 white' and '17 red' was treated with bisulfite. Separate PCR amplifications were performed for upper and lower strands. PCR products were isolated after restriction with *Eco*RI and *Hin*dIII and cloned into pBlueScript II (Stratagene). The sequences of 12 clones from each of the four PCR reactions are compiled in Figures 2 and 3. Each line represents the methylation pattern of DNA from a different plant cell. Analysis of m^{5}C distribution for the transcriptionally inactive line '17 white' is shown in Figure 2, where methylated C residues are labelled with filled symbols. C residues located within the symmetrical sequences (circles and asterisks) are almost completely methylated: 97.1% of all CpG dinucleotides and 90.5% of all CpNpG trinucleotides contain a m^{5}C residue. In addition, C methylation in non-symmetrical sequences (filled boxes) is also detectable. If we do not consider CpN dinucleotides that are part of a symmetrical CpNpG trinucleotide, m^{5}C residues occur in 44.3, 38.7 and 17.3% of the CpT, CpA and CpC dinucleotides, respectively. Although certain nucleotide compositions are more frequently methylated than others, we cannot detect a strict sequence specificity for non-symmetrical methylation. The lowest methylation frequency of 7.1% is observed in the first C residue of CpCpC trinucleotides, while CpTpT trinucleotides are methylated with the highest frequency of 56.1%. Nevertheless local frequencies for methylation of individual CpTpT trinucleotides differ between 0% (lower strand at position −109) and 100% (lower strand at position −132) (Figure 2).

The highest density of m^{5}C residues in non-symmetrical sequences is observed in the lower strand of the −226 to −177 region. Although the 12 different clones contain some common, preferentially methylated sites, individual clones vary in methylation patterns at certain sites. This implies that m^{5}C residues are not at exactly defined sequences, but rather within a defined region. An example of this heterogeneously clustered distribution of m^{5}C residues is the CTCC sequence in the lower strand at position −149 (Figure 2). In 11 of the 12 clones analysed at least one C residue within the CTCC sequence is methylated, but methylation can occur in each of the three C residues.

When compared with the inactive 35S promoter in line '17 white', cells of line '17 red' contain a significantly lower degree of cytosine methylation within the promoter region (Figure 3). Methylation of CpG and CpNpG residues is

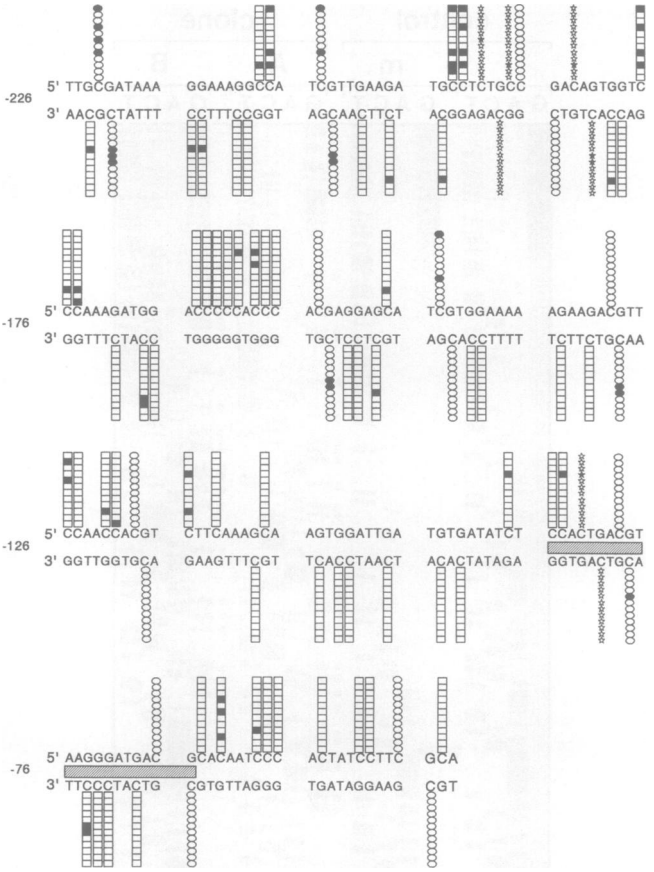

Fig. 3. Distribution of methylated C residues in the −226 to −44 region of the 35S promoter in line '17 red'. C residues are marked as described in Figure 2.

decreased to a total of 7.5 and 7.1%, respectively. Again we observe m^{5}C residues within non-symmetrical sequences. If we do not consider CpN dinucleotides that are part of a symmetrical CpNpG trinucleotide, m^{5}C residues occur in 6.3, 5.7 and 5.7% of the CpT, CpA and CpC dinucleotides, respectively. Again distribution of the few m^{5}C residues seems to be more intensive in certain regions, but not all clones carry methylated C residues at identical positions.

In conclusion, our data show that both the hypermethylated and the hypomethylated allele contain symmetrical as well as non-symmetrical m^{5}C residues within the 35S promoter region.

Discussion

Detection of non-symmetrical m^{5}C residues

Comparing a hypermethylated and a hypomethylated state of a transgene, we detect a significant portion of methylated cytosines within non-symmetrical sequences. This result is contrary to the accepted mechanism of base modification that confines cytosine methylation in plants to CpG dinucleotides or CpNpG trinucleotides (Gruenbaum *et al.*, 1981).

In the hypomethylated state of the A1 transgene, both symmetrical and non-symmetrical methylation occur at comparably low levels, with 7.5% CpG dinucleotides, 7.1% CpNpG trinucleotides and 5.9% C residues at non-symmetrical positions methylated. In the hypermethylated state of the Al transgene, symmetrical methylation is very efficient with 97.1% methylated CpG dinucleotides and

90.5% methylated CpNpG trinucleotides. However, 31.9% of non-symmetrical C residues are also methylated. Non-symmetrical methylation does not exhibit a completely random distribution. CpT and CpA dinucleotides are methylated at a higher frequency than CpC dinucleotides and methylation frequencies of different trinucleotides vary between 7.1 and 56.1%. Non-symmetrical methylation is clustered in certain regions, but individual clones do not always carry m^5C residues at identical positions.

Substrate specificities of DNA methyltransferases

It is unclear whether symmetrical and non-symmetrical methylation patterns are established by the same methyltransferase. While in mouse a single methyltransferase gene has been identified, a small multigene family with homology to cytosine-5 methylase has been detected in *A.thaliana* (Finnegan and Dennis, 1993). As it has not been shown that any of these methylase-like sequences actually encode methyltransferases, it remains unclear whether plants contain more than one methyltransferase or whether methyltransferase has different substrate specificities. It is interesting that the distribution frequencies of non-symmetrical m^5C residues in the hypermethylated A1 transgene match with data from a nearest-neighbour analysis of plant DNA. This study had shown that 19% of m^5C residues were located in CpA and CpT dinucleotides, respectively, but only 7% in CpC dinucleotides (Gruenbaum *et al.*, 1981). We observe a similar ratio with 39 and 44% methylation in non-symmetrical CpA and CpT dinucleotides, respectively, but only 17% in CpC dinucleotides. These data could support the assumption that methyltransferase recognizes CpT and CpA dinucleotides more efficiently than CpC dinucleotides. Methylation of CpC dinucleotides in plants suggests a different substrate specificity to that of mammalian methyltransferase, which has been shown to modify ~99% CpG, 14% CpA, 6% CpT, but 0% CpC dinucleotides *in vitro* (Simon *et al.*, 1983).

Conservation of non-symmetrical methylation patterns

The high level of cytosine methylation at symmetrical sequences detectable in the hypermethylated variant supports the model that symmetrical methylation patterns are efficiently transmitted in a semi-conservative fashion (Holliday and Pugh, 1975; Riggs, 1975). In contrast non-symmetrical methylation is less efficient and its perpetuation cannot be mediated by the methylation pattern of the non-replicated DNA strand. We do not detect any defined patterns within the DNA sequence that might mediate maintenance of non-symmetrical methylation. The clustered arrangement of m^5C residues and higher methylation frequencies at certain positions suggest that some regions are more susceptible to methylation than others. The signal for methylation might therefore be coded by structural elements and not by sequence. In this context some common features between non-symmetrical m^5C patterns and methylation products of RNA methyltransferases should be noticed. It has been demonstrated that a tRNA cytosine-5 methyltransferase recognizes the secondary structure of transcripts (Sakamoto and Okata, 1985) which results in clustering of methylated residues. Clustering of m^5A residues was also observed at three positions of Rous sarcoma virus RNA (Kane and Beemon, 1985). Modifications were heterogeneous since different RNA molecules appeared to be methylated differently. It may be considered that methylation of non-symmetrical sites results from an activity of methyltransferases that are guided by a particular secondary structure or chromatin conformation.

Non-symmetrical DNA methylation in other genes

So far it is unclear whether non-symmetrical methylation is confined to transgenes or whether it is also found in endogenous plant genes. In the few cases where genomic sequencing analysis has been performed, no non-symmetrical methylation patterns could be detected. In a *Hae*III fragment of a CG-rich satellite DNA from *Scilla siberica* methylation was confined to CpG and CpNpG sequences only (Deumling, 1981). In 1400 bp of 5′-flanking sequences of the maize *Adh* gene methylation was detected at CpG, CpApG and CpTpG sites (Nick *et al.*, 1986). Only one CpG residue was found methylated in 150 bp of the subterminal region of the maize Ac elements that had been transferred into tobacco. Analysing 200 bp of the chromosomal integration region into which the A1 transgene of line 17 had been inserted, we could not detect any cytosine methylation (data not shown). So far there is no indication that non-symmetrical methylation patterns are present in endogenous plant genes, but detection will be difficult, if methylation occurs at low frequencies, in certain species or in a limited number of DNA regions.

Non-symmetrical cytosine methylation has also been observed in filamentous fungi. The repeat-induced point mutation (RIP) mechanism in *N.crassa* mutates linked and unlinked duplicated sequences, that also become methylated. Genomic sequencing of a *Neurospora* gene exposed to RIP revealed cytosine methylation of non-symmetrical sequences at frequencies between 80 and 100%. Similar to our data methylation patterns were variable and clustered in particular regions. DNA methylation is accompanied with certain homology-dependent gene silencing phenomena in plants, such as *trans*-inactivation (Matzke *et al.*, 1989) or paramutation (Meyer *et al.*, 1993). For future analysis it should be interesting to examine whether non-symmetrical methylation patterns have a specific function for gene silencing effects in *N.crassa* and plants.

Materials and methods

Bisulfite treatment

Bisulfite treatment was performed as described by Frommer *et al.* (1992) with the following modifications: 2 μg of genomic DNA, restricted with *Hin*dIII, was used for bisulfite treatment. An additional denaturation step was performed, incubating DNA for 5 min at 95°C. The DNA was kept under nitrogen in 1.5 M sodium bisulfite−0.5 mM hydrochinon, pH 5.5 for 20 h at 50°C. After dialysis the samples were incubated in 0.1 M NH_4OH for 30 min at 37°C (Wang *et al.*, 1980). To avoid amplification of incompletely denatured fragments, samples were treated with *Mnl*I (Biolabs) before PCR. In the two control reactions 20 pg of linearized plasmid p35A1 were mixed with 2 μg salmon sperm DNA. Plasmid DNA was tested either unmethylated or after *in vitro* methylation with *Hae*III methylase (Biolabs).

PCR

PCR was performed in a volume of 50 μl Taq buffer (Perkin Elmer), containing 50 ng bisulfite-treated DNA, 6.25 μl of primer mixture A, 2 μl of a 2 mM dNTP solution, 2.5 mM $MgCl_2$ and 1 unit Ampli Taq (Perkin Elmer). For each strand two external (primer mixture A) and two internal primers (primer mixture B) were mixed at a concentration of 2 nmol/ml.For the top strand a 28mer (5′-GGAAATTTTTTTGGATTTTATTGTTTAG) and a 28mer (5′-AAAAATCCCCRRATACCCTRTCCTCTCC; R = G or A) were used as external primers and a 41mer (5′-GGGAATTCTT-

ATTTTATTGTGAAGA TAGTGGAAAAGGAAGG) and a 39mer (5'-GGAAGCTTAAAATAAACTTCCTTATATAAAAAAAAAATC) as internal primers. For the bottom strand a 34mer (5'-CTTTATTATAAAAATAATAAAAAAAAAAAAATAAC) and a 30mer (5'-TTAAATGAAATGAATTTTTTTATATAGAGG) were used as external primers and a 40mer (5'-AGTGAATTCAAAAAAAAAAATAACTCCTACAAATACCATC) and a 38mer (5'-GCCAAGCTTGAATTTTTTTATATAGAGGAAGGGTTTTG) as internal primers. As documented in Figure 1 (control m) we occasionally observe a lack of one or two A residues in the A-block of the 40mer, which is probably due to a contamination of the primer preparation. 33 cycles were performed with intervals of 1 min at 94°C, 2 min at 48°C and 2 min at 74°C. One microlitre of the first PCR was used in a second reaction that contained 6.25 μl of primer mixture B. This second PCR was performed under the same conditions as described above. After restriction with *Eco*RI and *Hin*dIII the amplified fragment was purified from an agarose gel and cloned into pBlueScript II SK$^-$ (Stratagene).

Acknowledgements

We would like to thank Dr M.Frommer for communication of the bisulfite sequencing technique before its publication and Dr T.Bestor for pointing our attention to the similarity between non-symmetrical methylation and RNA methylation patterns. This work was supported by Bundesministerium für Forschung und Technologie MDL 0316150A. M.t.L. is a research fellow of the Max-Planck-Society.

References

Adams,R.L.P. and Burdon,R.H. (1985) *Molecular biology of DNA methylation*. Springer Verlag, New York.
Benfey,P.N., Ren,L. and Chua,N.-H. (1990) *EMBO J.*, **9**, 1685–1696.
Bird,A.P. (1986) *Nature*, **321**, 209.
Deumling,B. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 338–342.
Doerfler,W. (1983) *Annu. Rev. Biochem.*, **52**, 93–124.
Finnegan,E.J. and Dennis,E.S. (1993) *Nucleic Acids Res.*, **21**, 2383–2388.
Foss,H.M., Roberts,C.J., Claeys,K.M. and Selker,E.U. (1993) *Science*, **262**, 1737–1741.
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.
Gruenbaum,Y., Naveh-Many,T., Cedar,H. and Razin,A. (1981) *Nature*, **292**, 860–862.
Gruenbaum,Y., Cedar,H. and Razin,A. (1982) *Nature*, **295**, 620–622.
Holliday,R. (1987) *Science*, **238**, 163–170.
Holliday,R. and Pugh,J.E. (1975) *Science*, **187**, 226–232.
Kane,S.E. and Beemon,K. (1985) *Mol. Cell. Biol.*, **5**, 2298–2306.
Lam,E., Benfey,P.N., Gilmartin,P.M., Fang,R.-X. and Chua,N.-H. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7890–7894.
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.
Matzke,M.A., Priming,M., Trnovsky,J. and Matzke,A.J.M. (1989) *EMBO J.*, **8**, 643–649.
Meyer,P., Heidmann,I., Forkmann,G. and Saedler,H. (1987) *Nature*, **330**, 677–678.
Meyer,P., Heidmann,I. and Niedenhof,I. (1993) *Plant J.*, **4**, 89–100.
Nick,H., Bowen,B., Ferl,R.J. and Gilbert,W. (1986) *Nature*, **319**, 243–246.
Pröls,F. and Meyer,P. (1992) *Plant J.*, **2**, 465–475.
Rhounim,L., Rossignol,J.L. and Faugeron,G. (1992) *EMBO J.*, **11**, 4451–4457.
Riggs,A.D. (1975) *Cytogenet. Cell Genet.*, **14**, 9–25.
Russel,P.J., Rodland,K.D., Rachlin,E.M. and McCloskey,J.A. (1987) *J. Bacteriol.*, **169**, 2902–2905.
Sakamoto,K. and Okata,N. (1985) *Nucleic Acids Res.*, **13**, 7195–7206.
Schwarz,D. and Dennis,E. (1986) *Mol. Gen. Genet.*, **205**, 476–482.
Selker,E.U. (1990) *Annu. Rev. Genet.*, **24**, 579–613.
Selker,E.U., Fritz,D.Y. and Singer,M.J. (1993) *Science*, **262**, 1724–1728.
Simon,D., Stuhlmann,H., Jahner,D., Wagner,H., Werner,E. and Jaenisch,R. (1983) *Nature*, **304**, 275–277.
Vongs,A., Kakutani,T., Martienssen,R.A. and Richards,E.J. (1993) *Science*, **260**, 1926–1928.
Wang,R.Y.-H., Gehrke,C.W. and Ehrlich,M. (1980) *Adv. Virus Res.*, **20**, 4777–4790.

Received on December 16, 1993; revised on February 2, 1994