

Heavy *de novo* methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation

Thierry Pélissier, Sabine Thalmeir, Dirk Kempe, Heinz-L. Sängler and Michael Wassenegger*

Max-Planck-Institut für Biochemie, Abteilung Viroidforschung, D-82152 Martinsried, Germany

Received December 31, 1998; Revised and Accepted February 12, 1999

ABSTRACT

Previous analysis of potato spindle tuber viroid (PSTVd) RNA-infected tobacco plants has suggested that an RNA–DNA interaction could trigger *de novo* methylation of PSTVd transgene sequences. Using the genomic sequencing technique, the methylation pattern associated with the RNA-directed DNA methylation process has been characterized. Three different PSTVd transgene constructs all showed a similar pattern of methylation. Most of the cytosines at symmetrical as well as non-symmetrical positions appeared to be methylated in both DNA strands of the viroid sequences. Heavy methylation was mostly restricted to the viroid cDNA sequences. Flanking DNA regions immediately adjacent to the viroid cDNA displayed a lower but significant level of cytosine methylation. The observation that the heavy methylation was essentially co-extensive with the length of the PSTVd cDNA sequences provided evidence that a direct RNA–DNA interaction can act as a strong and highly specific signal for *de novo* DNA methylation. These data also confirmed that *de novo* methylation was not limited to canonical CpG and CpNpG sites, but can also involve all the cytosine residues located in the genomic region where the RNA–DNA interaction takes place.

INTRODUCTION

The most common eukaryotic DNA modification is methylation of cytosine at position 5 (m^5C). In plants and animals, DNA methylation has been identified as a powerful mechanism to regulate gene expression (reviewed in 1,2) and is thought to play an essential role in a number of cellular processes, such as developmental control, genomic imprinting, control of parasitic elements and gene silencing (for reviews see 3–7).

In animals, C residues at CpG dinucleotides are the preferred targets for DNA methylation, while methylation at both CpG and CpNpG sequences is common in plants (8). The symmetry of the CpG and CpNpG sites was proposed to be important for stable maintenance of methylation patterns throughout DNA replication cycles. After replication, a maintenance methyltransferase

(MTase) could rapidly methylate C residues in the newly synthesized strand, if the parental strand contained an m^5C in the complementary sequence (9–11). In this semi-conservative model, the methylation pattern at non-symmetrical sequences is not efficiently maintained and should be lost after several cell divisions. However, cytosine methylation of non-symmetrical sequences was recently reported in mammals (12,13), in fungi (14,15) and in plants (16–20) and could contribute to the regulation of gene expression (21). Therefore, non-symmetrical methylation patterns have to be maintained by a mechanism different to that proposed in the semi-conservative model or they have to be established *de novo* after each DNA replication cycle.

Little is known about the molecular mechanisms that target DNA sequences for *de novo* methylation. It is also not clear if the processes involved in *de novo* methylation of symmetrical sequences are different from those taking place in the *de novo* methylation of non-symmetrical sequences. Because of the difficulty in analyzing cells where *de novo* methylation is initiated, the frequent appearance of symmetrical methylation patterns may simply reflect that only these patterns are efficiently maintained.

A large part of the information about *de novo* methylation in higher eukaryotes comes from the characterization of transgenic plants. In plants, heavy *de novo* methylation and silencing of multiple transgene copies integrated at the same locus have been proposed to occur through a DNA–DNA pairing process (22–24). The best evidence for such a mechanism is based on the methylation induced premeiotically (MIP) phenomenon that was discovered in the fungus *Ascobolus immersus* (25). In this filamentous fungus, tandemly duplicated sequences are very efficiently detected and methylated while variable methylation efficiencies are observed for duplicated sequences at ectopic positions. The observations that methylation is co-extensive with duplication and that either both duplicated sequences are affected, or neither of the two is, suggested that this premeiotic process required a direct interaction between the duplicated sequences (25,26). The methylation of an unmethylated sequence homologous to a methylated locus has been observed in different plant species (23,24,27). To account for this *de novo* methylation, involvement of a DNA pairing-dependent process termed ‘epigene conversion’ was proposed (24,28,29). In epigene conversion, strand interactions between methylated and unmethylated homologs produces hemimethylated intermediates which are the preferred substrates

*To whom correspondence should be addressed. Tel: +49 89 8578 2580; Fax: +49 89 8578 2937; Email: wasseneg@biochem.mpg.de

of maintenance MTase (11). Other possible mechanisms for *de novo* methylation propose either that the MTase has affinity for certain types of structural features in DNA (28,30) or a specific detection of sequences such as transgenes because of their different isochore content (i.e. GC content) compared with the locus of integration (31). Recognition of specific CpG or CpNpG sites localized in a particular sequence context by the DNA MTase, in conjunction with *cis*- and/or *trans*-acting regulatory factors, was also considered as a mechanism to initiate *de novo* methylation (for a review see 32).

Recent findings in our laboratory suggested that an RNA–DNA interaction served as a signal that triggered *de novo* DNA methylation (33). Viroid cDNA copies integrated into the tobacco genome were methylated after autonomous viroid RNA–RNA replication had taken place in these plants. However, only fragmentary information was obtained on the RNA-directed methylation of the transgenes. Only a few C residues, all of which were located in CpG and CpNpG sequences, were analyzed for methylation by Southern blot analysis using methylation-sensitive endonucleases.

In this paper, our interest was to characterize the RNA-directed DNA methylation (RdDM) process and the pattern of transgene methylation. RdDM was examined by the bisulphite method, which allows the detection of all methylated C residues within a genomic DNA region of interest (34). Here, we report that most if not all the cytosines located within the genome-integrated viroid sequences were methylated whatever their sequence context. This heavy methylation pattern was almost entirely restricted to the viroid sequences, indicating that the RNA-directed process was highly specific and led to *de novo* methylation of all the cytosines located in a putative RNA–DNA triplex region. A lower degree of cytosine methylation was observed in the DNA regions immediately flanking the viroid sequences and may partly result from the spreading of methylation from the heavily methylated viroid sequences into the adjacent regions.

MATERIALS AND METHODS

Plant material

Transgenic SRI-4(-), SRI-3(+), SRI-SB2(+) and SRI-SB2(+)/I tobacco plants containing genomically integrated PSTVd constructs were previously described by Wassenecker *et al.* (33).

Plant DNA isolation and Southern blotting

Plant nuclear DNA was isolated from 10–20 g of leaf material following the procedure of Bedbrook (35). Restriction analysis was performed with 15 µg of genomic DNA and 75–100 U of the appropriate endonuclease in the recommended buffer (Boehringer Mannheim). Digested DNA was electrophoresed in 0.8% agarose gels and transferred onto positively charged Qiabran membrane (Qiagen).

Southern hybridization was carried out according to the method of Amasino (36) using a PSTVd-specific DNA probe labeled with [α -³²P]dCTP by a random-primed DNA labeling reaction (Boehringer Mannheim).

Bisulphite conversion of DNA

Several series of bisulphite treatments of genomic DNA were performed using two methods, essentially as described by Clark *et al.* (37) or Olek *et al.* (38). Prior to the treatment, DNA from

SRI-4(-)-6, SRI-3(+)-9, SRI-SB2(+)-4 and SRI-SB2(+)/I-4 plants was digested with *SspI*, phenol/chloroform extracted, ethanol precipitated and resuspended in water.

'Classical' treatment according to Clark *et al.* (37). Digested DNA (1 µg) was denatured in 0.3 M NaOH for 20 min at 37°C in a total volume of 70 µl. Denatured DNA was mixed with 400 µl of a freshly prepared 2 M sodium metabisulphite (Merck)/0.6 mM hydroquinone (Sigma), pH 5.0, solution (1.7 M/0.5 mM, final concentrations). The reaction mixture was incubated under mineral oil, in a thermal cycler (Appligene) for 18 h at 55°C with a 30 s denaturation step at 95°C every 2 h. DNA was then purified via a desalting column step (Promega Wizard DNA Clean-Up System) and the eluted DNA was incubated in 0.3 M NaOH for 15 min at 37°C. After neutralization with ammonium acetate to a final concentration of 3 M, the DNA was ethanol precipitated and resuspended in 50 µl of water (20 ng/µl).

Bisulphite treatment of agarose beads according to Olek et al. (38). Digested DNA (1 µg) was boiled for 10 min, quickly chilled on ice and subsequently incubated in 0.3 M NaOH for 15 min at 50°C. Denatured DNA was mixed with 2 vol of 2% LMP agarose dissolved in water to give a final volume of 180 µl. This mixture was directly pipetted into chilled mineral oil to form beads of 15 µl (100 ng of denatured DNA/bead). The beads were incubated in 200 µl of 2.5 M sodium metabisulphite (Merck)/1.25 mM hydroquinone (Sigma), pH 5.0, for 4 h at 50°C in the dark and the protocol of Olek *et al.* (38) was then strictly followed.

To demonstrate the efficiency of bisulphite treatment and the sensitivity of the transgene sequences to bisulphite conversion, the pPCV702SM transformation vectors that carry the PSTVd-4(-), PSTVd-3(+) and PSTVd-SB2(+) transgene constructs, which had been introduced into the SRI tobacco genome, were used as controls. Conversion controls were performed on 135 pg of *SspI* digested plasmid DNAs, mixed with 1 µg of *SspI* digested genomic DNA isolated from SRI-SB2(+)-4, SRI-4(-)-6 or SRI-3(+)-9 plants (i.e. ~50 plasmid copies equivalent genome).

PCR amplification, cloning and sequencing

PCR was performed on 5 µl of converted DNA or on a single bead using a Crocodile II apparatus (Appligene). The amplifications were in a reaction volume of 100 µl containing 0.5 µl each primer, 0.2 mM each dNTP and 2.5 U of *TaKaRa LA Taq* in the recommended buffer. Samples were processed as follows: 1 cycle of 95°C for 3 min with annealing for 1 min at the T_{mp} (temperature corresponding to the lowest T_m of the primers used); 35 cycles of 72°C for 30 s, 95°C for 30 s, T_{mp} for 45 s; 1 cycle of 72°C for 5 min.

The following primers were used. Upper strand of p35S–PSTVd junction: PSTVd-4(-) construct, p540BiD (GTY-AYTTTATTGTGAAGATAGTG) and p180BiUS (TTTTCACCCTTCCTTTCTTC); PSTVd-3(+) construct, p710US (AAGYA-AGTGGATTGATGTG) and p280BiR (ARCTTCARTTRTTTC-ACC); PSTVd-SB2(+) construct, p710US (AAGYAAGTGG-ATTGATGTG) and p140BiR (CRCTRRRCACTCCCCAC). Lower strand of p35S–PSTVd junction: PSTVd-4(-) construct, p580BiD (CTCCTACAAATRCCATCATT) and p100BiLS (GAAAYTTGGAGYGAAYTGG); PSTVd-3(+)/SB2(+) constructs, p700BiLS (ACRTTCCAACCACRTCTTC) and p320LS (TTYAGTTGTTTTYYAYYGGGTAG). Upper strand of PSTVd–pAnos junction: PSTVd-4(-) construct, p320LS (TTY-

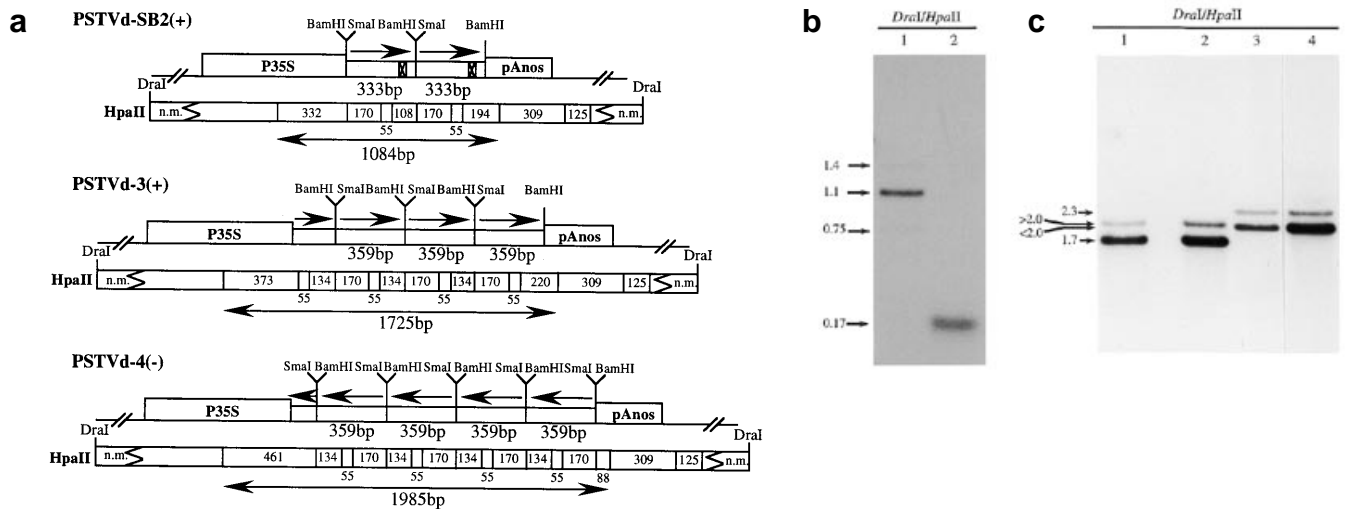


Figure 1. Evidence for methylation of the genome-integrated PSTVd-specific cDNAs by Southern blot analysis. **(a)** Physical map of the *DraI/HpaII* restriction pattern of the T-DNAs carrying the PSTVd-SB2(+), PSTVd-3(+) and PSTVd-4(-) transgene sequences (n.m. = not mapped). The PSTVd-SB2(+) construct contains two *SmaI*-*Bam*HI cDNA units (U) of a 26 bp PSTVd deletion mutant, which are arranged in tandem repeat and in plus orientation. The position of the 26 bp deletion is indicated (X). The PSTVd-3(+) construct contains 3.6 U of plus-orientated PSTVd cDNA and PSTVd-4(-) contains 4.4 U of minus-orientated PSTVd cDNA. **(b)** Southern blot hybridization of *DraI/HpaII* digested genomic DNA from PSTVd-free SRI-SB2(+)-4 (lane 2) and PSTVd-infected SRI-SB2(+)-I-4 (lane 1) plants with ³²P-labeled PSTVd cDNA. **(c)** Southern blot performed as above with SRI-3(+)-1 and -9 (lanes 1 and 2) and SRI-4(-)-6 and -8 (lanes 3 and 4) plant genomic DNA. The sizes of hybridized DNA fragments in kb are indicated. The low signal intensity on the original autoradiograph may render the 1.4 and 0.75 kb bands (b) hardly visible in print.

AGTTGTTTTYYAYYGGGTAG) and pNOS1020BiUS (CTCTAATCATAAAAACCCATCTC); PSTVd-3(+)/SB2(+) constructs, p10BiUS (AYTYGTGGTTYTGTGGT) and pNOS1020BiUS (CTCTAATCATAAAAACCCATCTC). Lower strand of PSTVd-pAnos junction: PSTVd-4(-) construct, p140BiR (CRCTRRRC-ACTCCAC) and p940LS (AGAAATTATATGATAATYATY-GYAAG) or pNOS1120BiLS (GGYAATTTYGATYTAGTAA-YATAGATG); PSTVd-3(+)/SB2(+) constructs, p30BiLS (TTCAC-ACCTRACCTCT) and pNOS1050BiLS (GTATTAAATGTAT-AATTGYGGG).

PCR products were excised from the gel and subcloned into the PGEM-Teasy vector (Promega). Sequencing was performed using an automated sequencer (ALFexpress; Pharmacia Biotech) and the Cy5 AutoRead Sequencing Kit (Pharmacia Biotech).

RESULTS

Plant material

Transgenic tobacco lines SRI-4(-), SRI-3(+) and SRI-SB2(+) carrying the PSTVd-4(-), PSTVd-3(+) and PSTVd-SB2(+) transgene constructs, respectively (Fig. 1a), have been previously described (33). SRI-4(-) and SRI-3(+) plants contained viroid (PSTVd) replication initiation (VRI)-competent cDNA constructs while the SRI-SB2(+) plants were transformed with a dimeric VRI-incompetent PSTVd cDNA construct. For PSTVd infection, viroid-free SRI-SB2(+)-4 tobacco cuttings were mechanically inoculated (39) and infected plants were named SRI-SB2(+)-I-4. Replication of PSTVd RNA in SRI-4(-), SRI-3(+) and SRI-SB2(+)-I-4 plants or its absence in the control SRI-SB2(+)-4 plant were monitored by northern blot analysis (data not shown).

PSTVd replication is associated with methylation of viroid cDNA sequences

To determine whether PSTVd cDNA sequences were specifically methylated in PSTVd-infected plants, genomic DNA from different lines was analyzed using the methylation-sensitive *HpaII* endonuclease (restriction site CCGG, with digestion blocked by methylation at either C residue). The hybridization patterns obtained with a PSTVd-specific probe on *DraI/HpaII* restricted genomic DNAs are shown in Figure 1b and c. For the viroid-free SRI-SB2(+)-4 plant DNA, a hybridized band of 170 bp was detected (Fig. 1b, lane 2). This corresponds to the predicted size for the non-methylated, completely restricted DNA (Fig. 1a). The 108 and 55 bp PSTVd cDNA-specific fragments, as well as the 194 bp fragment overlapping the 3' junction of the viroid cDNA, displayed signal intensities which were close to or at the Southern blot detection limit. The hybridization pattern observed for the PSTVd-infected SRI-SB2(+)-I-4 plant DNA revealed a major band of ~1.1 kb, indicating that the *HpaII* restriction sites located within the PSTVd cDNA sequence are protected from digestion by extensive methylation (Fig. 1b, lane 1). The minor ~1.4 kb band may reflect incomplete digestion of the *HpaII* site located within the pAnos region. The detection of a further ~750 bp faint band (Fig. 1b, lane 1) indicated that the *HpaII* site at the 5' border of the PSTVd transgene sequence was not fully methylated in the SRI-SB2(+)-I-4 plant (see Fig. 1a).

The hybridization patterns observed for *DraI/HpaII* digested DNAs of SRI-3(+)-1 and -9 (Fig. 1c, lanes 1 and 2) and SRI-4(-)-6 and -8 (Fig. 1c, lanes 3 and 4) transformants revealed a major band of ~1.7 and ~2 kb, respectively. No faint band of smaller size could be detected even after long film exposure (data not shown). This demonstrated full methylation of all *HpaII* sites within the PSTVd sequence (see maps in Fig. 1a). As for the SRI-SB2(+)-I-4 DNA, an additional PSTVd-hybridizing fragment

of a larger size was observed for all the SRI-3(+) and SRI-4(-) DNAs, which clearly suggested that the PSTVd RNA-directed methylation was not strictly restricted to the PSTVd cDNA part of the transgene sequences. The ~2 and ~2.3 kb fragments detected with the SRI-3(+) and SRI-4(-) DNAs, respectively, were expected if there was only partial cleavage of the *HpaII* site located in the pAnos region (Fig. 1a).

DNA methylation status of the PSTVd-SB2(+) transgene in SRI-SB2(+)-4 and SRI-SB2(+)/I-4 plants

For bisulphite analysis, we achieved similar results using either of two methods, described by Clark *et al.* (37) and Olek *et al.* (38). DNAs of the plasmid vectors carrying the PSTVd-4(-), PSTVd-3(+) and PSTVd-SB2(+) transgene constructs were used as a control for monitoring complete conversion of unmethylated C residues into U residues (for details see Materials and Methods). Each bisulphite-treated plasmid DNA control was amplified with all the primer pairs subsequently used to analyze the corresponding genomic DNA and a total of 72 independent control clones (six clones for each amplification) were characterized. Sequencing analysis demonstrated that the only detectable m⁵C residues were located at *EcoRII* sites (data not shown). This demonstrated that the bisulphite treatment was 100% efficient, because plasmid DNA is known to be methylated at *EcoRII* sites (Cm⁵C^A/TGG) by the bacterial *dcm* methylase, and confirmed that the DNA regions we intended to analyze did not contain sequences that were intrinsically resistant to bisulphite conversion.

The DNA regions analyzed corresponded to the 5' junction between the p35S promoter and PSTVd sequences and the 3' junction between the PSTVd and pAnos sequences (Fig. 1a). The Southern blot analysis indicated that a high level of methylation could exist in the viroid cDNA and in the 5' and/or 3' adjacent sequences. Therefore, PCR primers were designed to allow amplification of methylated, unmethylated and partially methylated targets at the 5' and 3' junctions (see Materials and Methods).

Using bisulphite-treated SRI-SB2(+)-4 DNA, separate PCR amplifications were performed for the upper and lower strands at the 5' and 3' junctions. Sequence analysis of 32 clones showed that these regions were virtually free of cytosine methylation (data not shown). At the 5' p35S–PSTVd junction, a total of 984 C residues, of which 529 are located within the PSTVd cDNA, were examined. Only three C residues, each at different positions in independent clones, were identified as potentially methylated. Similarly, the analysis of the PSTVd–pAnos junction showed only one potential m⁵C out of 712 C positions analyzed (176 C positions within the PSTVd part). These results correspond to a frequency of cytosine methylation of ~0.2%, indicating that the PSTVd transgene sequences are hypomethylated in the SRI-SB2(+)-4 plant.

Sequence compilation of the 5' and 3' junctions that were amplified from the bisulphite-treated SRI-SB2(+)/I-4 DNA gave a completely different pattern of cytosine methylation (Figs 2 and 3). In both the upper and lower strand of the viroid-specific part of the 5' p35S–PSTVd junction, 91.7 and 95.1% of all the C residues were found to be methylated (Fig. 2). Of the m⁵C residues, 43.8% was found in a non-symmetrical sequence context, while 29.3% was detected at CpG and 26.9% at CpNpG sites. This distribution closely reflected the relative representation of these sites in this region (42.7, 31.2 and 26.1%). Upper strand methylation ranged from 60 to 100% and lower strand from 72.5 to

100% for individual DNA molecules. Strikingly, 22 out of 28 DNA molecules displayed >90% methylation and included 11 molecules that were completely methylated. Moreover, the few non-methylated C residues identified seemed to have a random distribution along the viroid-specific sequence.

Cytosine methylation at symmetrical and non-symmetrical sites was also detected in 24 out of 28 individual DNA strands in the non-viroid-specific section of the p35S–PSTVd junction and it appeared to be restricted to the region immediately adjacent to the viroid sequence (positions –1 to –21). In this PSTVd-flanking region, the overall level of methylation rapidly decreased from 75 (–1 to –5) to 29.4% (–6 to –21) with increasing distance from the PSTVd sequence. In the proximal –1 to –21 p35S region, this corresponded to an average degree of C methylation of 38.9%, which was significantly lower than that detected for the viroid sequence (94.7%). In the region further upstream (–22 to –128), sparse methylation was detected in eight out of 28 independent clones. As expected from the Southern blot analysis (Fig. 1b, lane 1), the bisulphite analysis confirmed that the *HpaII* site located at the 5' junction between the PSTVd and p35S sequences was not fully methylated (Fig. 2).

For the SRI-SB2(+)/I-4 plant, the methylation pattern at the 3' junction is shown in Figure 3 and was essentially similar to that found at the 5' junction. However, the level of cytosine methylation in the PSTVd region was reduced compared with that at the 5' junction and only 64% of symmetrical and 69.2% of non-symmetrical C residues were methylated. In individual clones, the methylation levels ranged from 10 to 100%, but the majority of clones (75%) showed >60% methylation. The overall level of methylation decreased to 32.8% in the pAnos region immediately adjacent to the PSTVd sequence (Fig. 3, positions +1 to +20) and, interestingly, seven out of 16 clones did not show any detectable methylation in this area. In the +21 to +117 region, m⁵C residues were detected in six out of the 16 individual strands, mainly in symmetrical positions. For the viroid-free SRI-SB2(+)-4 plant, 16 individual clones were analyzed and no methylation was found in the corresponding 3' junction. This indicated that a low but significant level of methylation was present within this 3' region of the PSTVd-infected SRI-SB2(+)/I-4 plant. In particular, two out of the 16 clones displayed cytosine modification at the diagnostic *HpaII*₊₈₈ site, which was in good agreement with the low signal strength of the ~1.4 kb band identified in the Southern blot analysis (Fig. 1b, lane 1). Nevertheless, together with the situation observed in the promoter region, these results strongly suggest that in the SRI-SB2(+)/I-4 plant the PSTVd RNA-directed DNA methylation was almost entirely restricted to the PSTVd cDNA transgene sequences.

Methylation pattern of the viroid construct in the SRI-3(+)-9 plant

DNA fragments of 191 and 208 bp representing the upper and lower strands of the 5' junction from the SRI-3(+)-9 plant were analyzed. These fragments differ from those of the SRI-SB2(+)/I-4 plants by the PSTVd sequence lying directly next to the p35S sequence (see Fig. 1a). Despite this difference, their methylation patterns were essentially the same as those detected for the 5' junction fragments in the SRI-SB2(+)/I-4 plant (data not shown). Methylation was mainly limited to the viroid sequences, with an overall methylation frequency of 92% in the viroid region and of 39% in the –1 to –21 proximal p35S region.

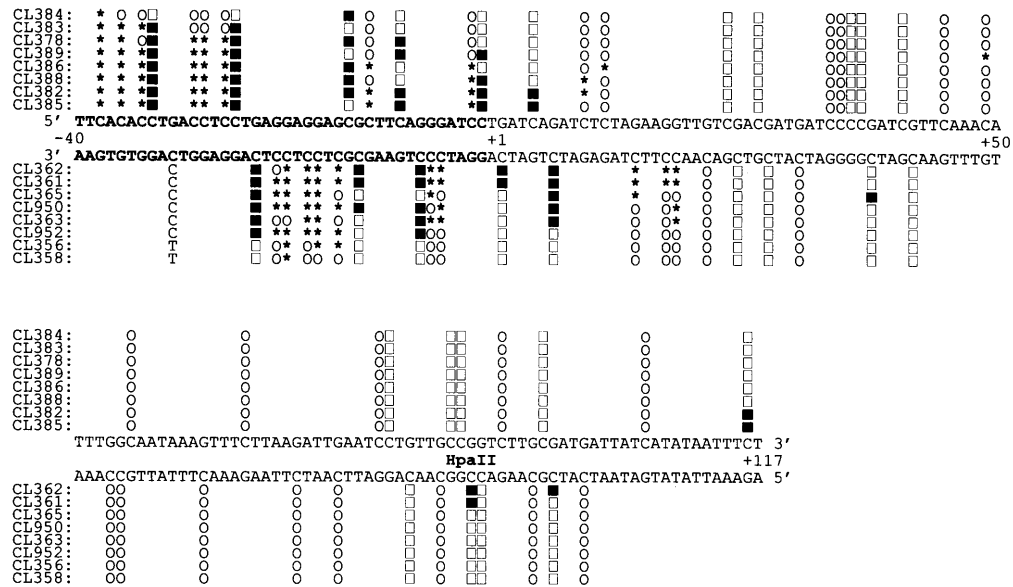


Figure 3. Methylation pattern at the PSTVd-pAnos junction in the SRI-SB2(+)/I-4 plant. The DNA region ranging from -40 to +117 relative to the border between PSTVd and pAnos sequences is shown. The PSTVd-specific sequence is presented in bold. The location of the *HpaII*₊₈₈ site is indicated. Symbols are used as in Figure 2.

region, and 54% of the m⁵C residues were found in CpG or CpNpG sites, which contain 48.6% of all the C residues. However, nine out of 24 strands did not show any methylation within this area whereas the majority of the DNA molecules displayed m⁵C residues in the +1 to +20 pAnos region. In four out of 24 DNA strands, the *HpaII*₊₈₈ site appeared to be methylated and was resistant to *HpaII* digestion, as expected from the Southern blot analysis of SRI-3(+)-9 DNA (Fig. 1c, lane 2). Within the +40 to +117 region, half of the DNA strands displayed variable m⁵C levels, mostly in the range of 7–23% (Fig. 4). DNA regions downstream of position +117 were also analyzed, but only sparse methylation could be detected (data not shown).

Methylation pattern of the PSTVd transgene in the SRI-4(-)-6 plant

The PSTVd-4(-) construct contained 4.4 PSTVd cDNA units that were arranged in an opposite orientation to the PSTVd-SB2(+) and PSTVd-3(+) constructs (Fig. 1a). The methylation level detected in the 5' junction reached 98.2% for the PSTVd sequences (data not shown), a value comparable with those observed in SRI-SB2(+)/I-4 and -3(+)-9 plants. However, cytosine methylation reached 74.3% in the proximal -1 to -21 p35S region while only ~39% of the cytosines were methylated in SRI-SB2(+)/I-4 and SRI-3(+)-9 plants (data not shown). Cytosine methylation within the upstream -22 to -95 region was also observed more frequently for the SRI-4(-)-6 transgene sequences (overall level of 16.2%) compared with the other constructs (<2.5%). In this area the methylation level reached up to 42% in individual DNA strands, but sequences upstream of the -95 position were only sparsely methylated. As observed for the SRI-SB2(+)/I-4 and -3(+)-9 plants, cytosine methylation occurred at both symmetrical as well as non-symmetrical sequences.

As for the SRI-3(+)-9 DNA, the PSTVd cDNA-specific area of the 3' junction fragment of the SRI-4(-)-6 DNA was nearly fully methylated, with an overall frequency of cytosine methylation of

98.9% (Fig. 5). The adjacent pAnos region could be separated into three sectors. (i) The 40 bp proximal sequence which was highly methylated (overall methylation 66.7%) with a range of methylation from 33 to 100% for individual strands. Full methylation was detected in one out of 22 strands, a proportion comparable with that observed for the same area in the SRI-3(+)-9 plant (2/24). (ii) The +41 to +90 pAnos region for which only 10 out of the 22 clones displayed cytosine methylation (70% for clone 502 and 10–44% for the other clones). (iii) The +90 to + 200 region which appeared virtually free of methylation (Fig. 5 and data not shown). Within the +1 to +90 pAnos region, 63.4% of the m⁵C residues are localized at CpG and CpNpG sequences while the relative representation of these sites was only 47.4%. This obviously indicates that the overall distribution of m⁵C residues was biased in favour of symmetrical sequences. The diagnostic *HpaII*₊₈₈ site appeared to be methylated in a limited number of clones, which was suggested by the Southern blot analyses (Fig. 1c, lane 3).

To verify the bisulphite results, the methylation status of the SRI-4(-)-6 PSTVd sequence was examined using *Bam*HI or *Ava*II endonuclease. *Bam*HI is sensitive to cytosine methylation at non-symmetrical positions whereas *Ava*II is sensitive to methylation on both symmetrical and non-symmetrical cytosines (see Fig. 6a). An aliquot of 15 µg of genomic DNA was spiked with 5 ng of plasmid DNA (~400 gene copy equivalents) prior to enzymatic digestion. A Southern blot of *Dra*I/*Bam*HI and *Dra*I/*Ava*II restricted DNAs was first probed with a PSTVd-specific probe and autoradiographed for 72 h (Fig. 6b). After removal of the probe, the blot was rehybridized with a plasmid-specific probe and autoradiographed for 5 min (Fig. 6c). The hybridization patterns obtained with the plasmid probe perfectly correspond to those expected for the fully digested plasmid DNA (Fig. 6c), indicating the absence of restriction endonuclease inhibitors. With the PSTVd probe (Fig. 6b), cross-hybridization to the plasmid-specific restriction fragments was detectable on the 72 h exposed autoradiograph (compare Fig. 6b and c). The cross-hybridization resulted either from a contamination

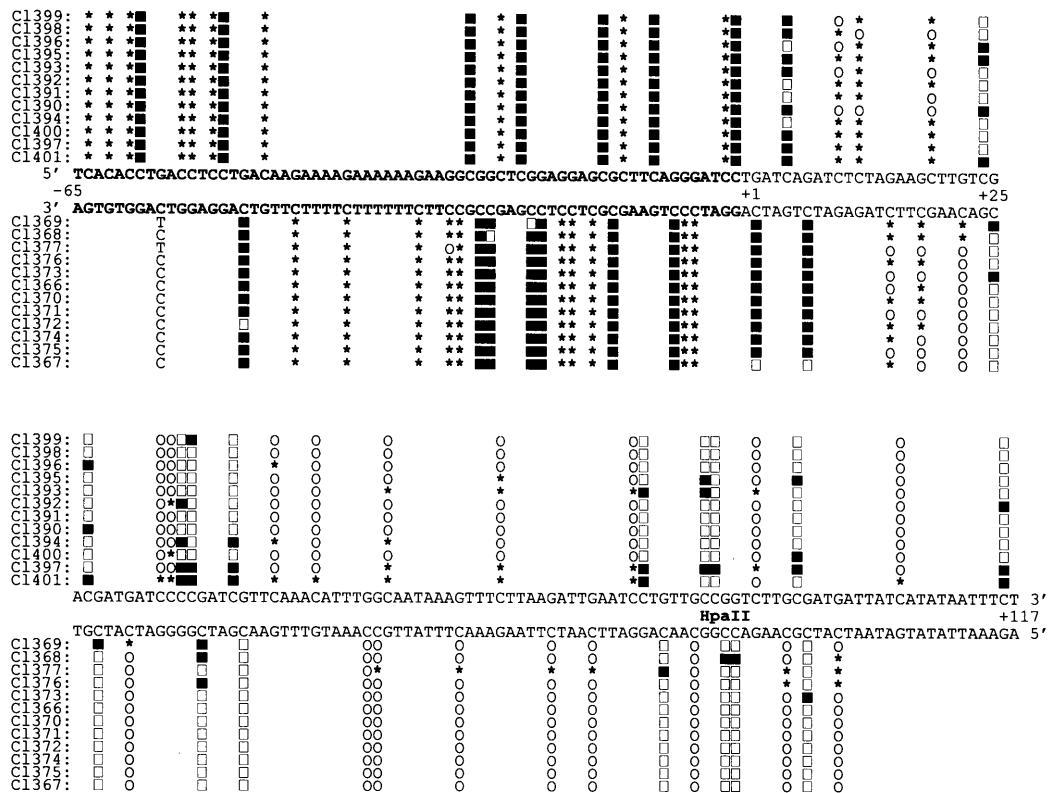


Figure 4. State of cytosine methylation at the PSTVd-pAnos junction of the SRI-3(+)-9 plant. The DNA region ranging from -65 to +117, relative to the border between PSTVd and pAnos sequences, is presented. The PSTVd-specific sequence is presented in bold. The location of the *HpaII*₈₈ site is indicated. Symbols are used as in Figure 2.

of the probe with traces of plasmid DNA or from the high CG content of the PSTVd sequence. For the *DraI/BamHI* digested DNA, detection of a major band of ~8.8 kb indicated that the *BamHI* sites within the viroid sequence were extensively methylated (Fig. 6b, lane 1). However, the presence of an ~6.8 kb faint band showed that, at least in some DNA molecules, the *BamHI* site located at the 3' boundary of the PSTVd sequence was cut (Fig. 6a). The *DraI/AvaII* digestion revealed a unique PSTVd-specific fragment of ~3.2 kb (Fig. 6b, lane 2), demonstrating that the *AvaII* sites within the viroid sequence were also extensively methylated (see Fig. 6a). The absence of any larger hybridizing band further indicated that the *AvaII* site close to the 5' junction (position -180) was not methylated. This result was in agreement with the bisulphite analysis (see above, Fig. 5). As observed for the *DraI/BamHI* restricted DNA, no hybridizing fragment corresponding to PSTVd monomers or multimers was detected. Thus, the Southern blot results strongly support the conclusion that a heavy methylation pattern was associated with the PSTVd transgene sequence.

DISCUSSION

DNA methylation is a well-characterized phenomenon which is involved in many diverse biological events, such as genomic imprinting and gene silencing. However, little is known about the molecular signals that regulate *de novo* methylation. Here, we demonstrate that *de novo* methylation of symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA

methylation. We previously reported that during PSTVd RNA replication, PSTVd cDNA copies integrated into the tobacco genome are methylated (33). Viroids are plant pathogens consisting of circular, single-stranded, non-encapsidated RNA molecules, which do not encode any protein. The demonstration that viroids use an RNA-RNA pathway to autonomously replicate in the nucleus (reviewed in 40) led to the hypothesis that an RNA-DNA interaction could serve as a signal to direct *de novo* methylation (33).

In this study, the methylation pattern associated with viroid replication was characterized in transgenic plants using the genomic sequencing method (34). In the viroid-free SRI-SB2(+)-4 plant line, no significant methylation could be detected within the PSTVd transgene sequences. In contrast, the viroid-infected SRI-SB2(+)-I-4, SRI-3(+)-9 and SRI-4(-)-6 plants displayed methylation levels ranging from 90 to 100% within the PSTVd cDNA sequences. The only notable exception was observed at the 3' junction in the SRI-SB2(+)-I-4 plant containing a lower level of m⁵C. This observation was not surprising because the PSTVd sequence comprised a 26 bp deletion (Fig. 1a), unlike the wild-type replicating viroid. Thus, formation of a RNA-DNA hybrid will be less efficient within this region than within the 5' part where the PSTVd RNA and the viroid cDNA are entirely complementary. This would then lead to a lower level of C methylation at the SRI-SB2(+)-I-4-specific 3' junction.

Although the PSTVd cDNAs of all the different constructs were not entirely sequenced, it is most likely that internal PSTVd

sequences are also heavily methylated at symmetrical and non-symmetrical positions. As shown for the SRI-4(-)-6 DNA, this is supported by evidence that the *Ava*II, *Bam*HI and *Hpa*II restriction sites, which are distributed along the PSTVd-specific transgene sequence, are resistant to digestion (Figs 1 and 6). For most of the individual DNA strand molecules, heavy methylation was almost entirely restricted to the PSTVd sequences. The flanking DNA regions were either sparsely methylated or unmodified (see below). The demonstration of heavy methylation within the PSTVd-specific sequences shows the high specificity of the methylation machinery. It also provides a strong argument for *de novo* methylation being directed by unusual structures that could arise by pairing of RNA molecules with their genomic counterparts. As originally put forward (33), this process should be termed RNA-directed and not RNA-mediated DNA methylation (RdDM). This is to emphasize that only DNA sequences complementary to the directing RNA are specifically methylated.

In plants, the characterized DNA MTases are closely related to the putative mammalian maintenance enzyme (see 5,7). They show a high specificity for CpG and/or CpA/TpG sites and for hemimethylated DNA. These enzyme characteristics, together with the self-complementarity of CpG and CpNpG sites, explain the stable propagation of methylation patterns. This is in contrast to the RdDM pattern described here. Most, if not all, of the cytosines which are located in the putative RNA-DNA triplex region are methylated, irrespective of their sequence context. That this novel and strong pattern reflects the actual situation and does not result from experimental artifacts due to a lack of bisulphite conversion (41,42) was verified by the control experiments. The usage of the vectors containing the transgene constructs that had been introduced into the SRI-SB2(+), -3(+) and -4(-) plants mixed with the genomic DNAs demonstrated that none of the analysed C positions are intrinsically resistant to bisulphite conversion. In addition, the characterization of the SRI-SB2(+)-4 plant was an ideal control for the analysis of the PSTVd-infected SRI-SB2(+)/I-4 sister plant. The fact that the PSTVd transgene sequences specific for the SRI-SB2(+)-4 plant appeared nearly free of methylation strongly reinforces the validity of our results. Harrison *et al.* (42) recently reported that for different sequence motifs, C residues adjacent to methylated sites could give partial resistance to bisulphite conversion. The homogeneity of the methylation pattern observed within the viroid sequence argues against such bias in our case. Moreover, the results obtained by Southern blot analysis of the methylation status of different C residues located at symmetrical (CpG or CpNpG) and also non-symmetrical (CpA or CpC) sites are entirely consistent with the bisulphite data (see above; Figs 1 and 6).

The recognition of specific structures in DNA that are formed during the RdDM process may strongly stimulate the potential *de novo* activity of maintenance MTase(s), resulting in the loss of sequence specificity. Alternatively, putative MTases lacking the large N-terminal regulatory domain have recently been identified in *Arabidopsis* (METIII), mice, humans (Dnmt2) (7) and in the fungus *Ascobolus immersus* (Masc1) (43). These MTases might be involved in the establishment of *de novo* methylation patterns. In *Ascobolus*, Masc1 appeared to play an essential role in *de novo* methylation during the MIP process (43). In this fungus, DNA duplications are efficiently detected during the sexual reproduction cycle and are heavily methylated at cytosines (44). MIP certainly involves a DNA-DNA pairing step for the triggering of *de novo* methylation and was shown to affect all the cytosines in

duplications >300–400 bp (15), a pattern reminiscent of that we observed for the RdDM process.

Viroids autonomously replicate in the nucleus via an RNA-RNA-dependent pathway, where they can accumulate to up to 3×10^4 copies, which are mostly localized in the nucleolus (45). Mature circular viroid RNAs are plus-stranded molecules which are transcribed into an oligomeric minus-strand RNA. The minus-strand then acts as a template for the synthesis of oligomeric plus-strands which are processed to mature molecules by enzymatic cleavage and ligation steps (40). The detection of a similar pattern of heavy methylation in all the upper and lower PSTVd-specific strands could be an indication of *de novo* methylation events on both DNA strands after each replication cycle. However, the presence of both plus and minus PSTVd RNA in the nuclei of infected cells prevented the identification of the molecule(s) that directs the methylation. Furthermore, it is not known whether an RNA-DNA duplex or a triple helix structure is recognized by MTase(s). Theoretically, mature plus, oligomeric plus or minus PSTVd RNAs are all capable of binding complementary transgene DNA sequences. However, considering the large numbers of mature PSTVd molecules in the nucleus, we surmise that the primary determinant for methylation could involve interactions with partially denatured double-stranded, viroid molecules.

35S promoter-driven PSTVd replication initiation-competent cDNA constructs were introduced into SRI-3(+) and -4(-) plants. Therefore, viroid replication is expected to occur in every leaf cell. This could then lead to a strong *de novo* methylation of the PSTVd transgene sequences, a situation which corresponds well to our observations (see above). In the SRI-SB2(+)/I-4 plant, PSTVd infection was obtained by mechanical inoculation, but a pattern of methylation similar to those observed in the SRI-3(+)-9 and -4(-)-6 plants was detected throughout the cell population of infected leaves. *In situ* hybridization studies of PSTVd distribution in mechanically infected tomato leaves led Harders *et al.* (45) to conclude that the majority of cells (>80%) were uninfected and viroid free. However, our data indicated that viroid RNA-specific molecules were present in the majority of leaf cells. Several lines of evidence may explain the differences between our observations and those of Harders *et al.*

We cannot exclude that some classes of leaf cells may be resistant to our genomic DNA extraction procedure. However, it seems more likely that only a fraction of cells from infected leaves contain high levels of actively replicating PSTVd molecules. Systemic viroid infection requires cell-to-cell movement via plasmodesmata and long-distance movement through the phloem (46,47). We speculate that PSTVd RNA does reach most leaf cells, but is at a concentration too low to be detected by *in situ* hybridization. Importantly, the presence of low concentrations of PSTVd RNA molecules could still be sufficient to target specific *de novo* methylation of the genome-integrated viroid sequences. Alternatively, initiation of methylation in highly infected cells could produce a novel sequence-specific methylation signal, which could then systemically propagate. This signal could be comparable with the putative ribonucleoprotein complex that allows propagation of infectious viroid molecules. At least the nucleic acid component, required for the signal specificity, could differ from the mature viroid RNA and consist of partially degraded RNA molecules incapable of autonomous replication. In plants, such sequence-specific diffusible signals have been recently reported to be involved in the systemic spread of post-transcriptional gene silencing (48–50). A last, but less likely, possibility would implicate

high level maintenance of C methylation in the non-symmetrical context, a hypothesis which is difficult to reconcile with the general pattern of maintenance observed so far in plants (7).

Besides heavy methylation of the viroid sequences, most of the individual DNA strands displayed a significant but lower level of methylation within the 5' and 3' PSTVd-flanking regions. In all the three different constructs the extent of methylation appeared to be mainly restricted to the -1 to -21 promoter region and to the +1 to +40 pAnos region. It is conceivable that MTase(s) which is directed to the place of RNA-DNA interactions may spread into the adjacent sequences before the enzyme is released from the template. The spreading was more efficient in the 3' region than in the promoter sequence, indicating an effect of the sequence and/or structural context of the adjacent regions. Differential spreading would occur after each cell division leading to a random methylation of various C residues. Because maintenance methylation is favoured for symmetrical m⁵C residues, the non-symmetrical patterns will be lost. This imperfect maintenance of methylation patterns would be consistent with the observed bias of symmetrical methylation within the p(A)nos region (see for example Fig. 5). Similarly, the absence of CpG/CpNpG motifs in the -9 to -60 promoter region may explain the limited spreading observed in this region.

Our results show that RNA can induce epigenetic changes substantiating the possible existence of an RNA-mediated transcriptional gene silencing system (51). This idea becomes even more attractive since it was shown that RdDM is not restricted to the viroid system. In particular, experiments in which the *nopaline synthase* promoter (*NOSpro*) was transcribed by p35S strongly suggested that specific methylation of an unlinked homologous *NOSpro* could be directed *in trans* by a specific *NOSpro* RNA (52). Another example of non viroid-specific RdDM was found in pea plants containing a single copy of a pea seed-borne mosaic virus (PSbMV) replicase transgene (53). Some of the transgenic plants that had been challenged with PSbMV exhibited an inducible resistance which was based on post-transcriptional gene silencing. Interestingly, the viral transgene was found to be specifically methylated in all the resistant plants. This was shown using endonucleases that are sensitive to methylation on both symmetrical or non-symmetrical C residues. These results imply that sequence-specific communication between cytoplasmic RNA and homologous nuclear sequences can occur and affects both methylation and gene expression (53).

Although RdDM was observed in the artificial viroid system, RNA seemed to gain increasing influence with respect to *de novo* DNA methylation in higher plants. Further studies will show whether a similar mechanism actually acts in other higher eukaryotes and whether RdDM is a widespread process that is involved in homology-dependent gene regulation.

ACKNOWLEDGEMENTS

The authors thank N. Emans for critical reading of the manuscript (www.sciwriters.org), J. M. Deragon and M. F. Mette for providing their manuscript prior to publication. T.P. was supported by a fellowship from the Human Frontier Science Program (LT-0319/1996-M). Our research was supported by the Deutsche Forschungsgemeinschaft (grants Wa1019/1-1-1-4).

REFERENCES

- Siegfried,Z. and Cedar,H. (1997) *Curr. Biol.*, **7**, R305-R307.
- Kass,S.U., Pruss,D. and Wolffe,A.P. (1997) *Trends Genet.*, **13**, 444-449.
- Flavell,R.B. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3490-3496.

- Razin,A. and Cedar,H. (1994) *Cell*, **77**, 473-476.
- Yoder,J.A. and Bestor,T.H. (1996) *Biol. Chem.*, **377**, 605-610.
- Richards,E.J. (1997) *Trends Genet.*, **13**, 319-323.
- Finnegan,E.J., Genger,R.K., Peacock,W.J. and Dennis,E.S. (1998) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 223-247.
- Gruenbaum,Y., Naveh-Manly,T., Cedar,H. and Razin,A. (1981) *Nature*, **292**, 860-862.
- Holliday,R. and Pugh,J.E. (1975) *Science*, **187**, 226-232.
- Riggs,A.D. (1975) *Cytogenet. Cell Genet.*, **14**, 9-25.
- Gruenbaum,Y., Cedar,H. and Razin,A. (1982) *Nature*, **295**, 620-622.
- Crowther,P.J., Cartwright,A.L., Hocking,A., Jefferson,S., Ford,M.D. and Woodcock,D.M. (1988) *Nucleic Acids Res.*, **17**, 7229-7239.
- Toth,M., Müller,U. and Doerfler,W. (1990) *J. Mol. Biol.*, **214**, 673-683.
- Selker,E.U., Fritz,D.Y. and Singer,M.J. (1993) *Science*, **262**, 1724-1728.
- Goyon,C., Nogueira,T.I.V. and Faugeron,G. (1994) *J. Mol. Biol.*, **240**, 42-51.
- Ingelbrecht,I., Van Houdt,H., Van Montagu,M. and Depicker,A. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 10502-10506.
- Meyer,P., Niedenhof,I. and ten Lohuis,M. (1994) *EMBO J.*, **13**, 2084-2088.
- Oakeley,E.J. and Jost,J.P. (1996) *Plant Mol. Biol.*, **31**, 927-930.
- Wang,L., Heinlein,M. and Kunze,R. (1996) *Plant Cell*, **8**, 747-758.
- Gosselink,C., Arnaud,P., Heslop-Harrison,J.S. and Deragon,J.M. (1999) *Plant Mol. Biol.*, **39**, 243-255.
- Diéguez,M.J., Bellotto,M., Afsar,K., Mittelsten Scheid,O. and Paszkowski,J. (1997) *Mol. Gen. Genet.*, **253**, 581-588.
- Assaad,FF., Tucker,K.L. and Signer,E.R. (1993) *Plant Mol. Biol.*, **22**, 1067-1085.
- Vaucheret,H. (1993) *C.R. Acad. Sci. Paris*, **316**, 1471-1483.
- Matzke,A.J.M., Neuhuber,F., Park,Y.-D., Ambros,P.F. and Matzke,M.A. (1994) *Mol. Gen. Genet.*, **244**, 219-229.
- Faugeron,G., Rhounim,L. and Rossignol,J.-L. (1990) *Genetics*, **124**, 585-591.
- Barry,C., Faugeron,G. and Rossignol,J.-L. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 4557-4561.
- Meyer,P., Heidmann,I. and Niedenhof,I. (1993) *Plant J.*, **4**, 89-100.
- Bestor,T.H. and Tycko,B. (1996) *Nature Genet.*, **12**, 363-367.
- Colot,V., Maloisel,L. and Rossignol,J.-L. (1996) *Cell*, **86**, 855-864.
- Stam,M., Mol,J.N.M. and Kooter,J.M. (1997) *Ann. Bot.*, **79**, 3-12.
- Meyer,P. and Heidmann,I. (1994) *Mol. Gen. Genet.*, **243**, 390-399.
- Jost,J.-P. and Bruhat,A. (1997) *Prog. Nucleic Acid Res. Mol. Biol.*, **57**, 217-248.
- Wassenegger,M., Heimes,S., Riedel,L. and Sängler,H.L. (1994) *Cell*, **76**, 567-576.
- 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.
- Bedbrook,J. (1981) *Plant Mol. Biol. News*, **2**, 24.
- Amasino,R.M. (1986) *Anal. Biochem.*, **152**, 304-307.
- Clark,S.J., Harrison,J., Paul,C.L. and Frommer,M. (1994) *Nucleic Acids Res.*, **22**, 2990-2997.
- Olek,A., Oswald,J. and Walter,J. (1996) *Nucleic Acids Res.*, **24**, 5064-5066.
- Tabler,M. and Sängler,H.L. (1984) *EMBO J.*, **3**, 3055-3062.
- Sängler,H.L. (1987) In Diener,T.O. (ed.), *The Viroids*. Plenum Press, New York, NY, pp. 117-166.
- Rein,T., Zorbas,H. and DePamphilis,M.L. (1997) *Mol. Cell. Biol.*, **17**, 416-426.
- Harrison,J., Stirzaker,C. and Clark,S.J. (1998) *Anal. Biochem.*, **264**, 129-132.
- Malagnac,F., Wendel,B., Goyon,C., Faugeron,G., Zickler,D., Rossignol,J.L., Noyer-Weidner,M., Vollmayr,P., Trautner,T.A. and Walter,J. (1997) *Cell*, **91**, 281-290.
- Goyon,C., Barry,C., Grégoire,A., Faugeron,G. and Rossignol,J.L. (1996) *Mol. Cell. Biol.*, **16**, 3054-3065.
- Harders,J., Lukacs,N., Robert-Nicoud,M., Jovin,T.M. and Riesner,D. (1989) *EMBO J.*, **8**, 3941-3949.
- Ding,B., Kwon,M.O., Hammond,R. and Owens,R. (1997) *Plant J.*, **12**, 931-936.
- Palukaitis,P. (1987) *Virology*, **158**, 239-241.
- Voinnet,O., Vain,P., Angell,S. and Baulcombe,D.C. (1998) *Cell*, **95**, 177-187.
- Palauqui,J.C., Elmayan,T., Pollien,J.M. and Vaucheret,H. (1997) *EMBO J.*, **16**, 4738-4745.
- Jorgensen,R.A., Atkinson,R.G., Forster,R.L.S. and Lucas,W.J. (1998) *Science*, **279**, 1486-1487.
- Wassenegger,M. and Pélassier,T. (1998) *Plant Mol. Biol.*, **37**, 349-362.
- Mette,M.F., van der Winden,J., Matzke,M.A. and Matzke,A.J.M. (1999) *EMBO J.*, **18**, 241-248.
- Jones,A.L., Thomas,C.L. and Maule,A.J. (1998) *EMBO J.*, **17**, 6385-6393.