Preferential de Novo Methylation of Cytosine Residues in Non-CpG Sequences by a Domains Rearranged DNA Methyltransferase from Tobacco Plants*

Received for publication, April 14, 2003, and in revised form, August 1, 2003
Published, JBC Papers in Press, August 12, 2003, DOI 10.1074/jbc.M303892200

From the Research and Education Center for Genetic Information, Nara Institute of Science and Technology, Nara 630-0192, Japan

Yuko Wada, Hitoshi Ohya, Yube Yamaguchi, Nozomu Koizumi, and Hiroshi Sano†

In plant DNA, cytosines in symmetric CpG and CpNpG (N is A, T, or C) are thought to be methylated by DNA methyltransferases, MET1 and CMT3, respectively. Cytosines in asymmetric CpNpN are also methylated, and genetic analysis has suggested the responsible enzyme to be domains rearranged methyltransferase (DRM). We cloned a tobacco cDNA, encoding a novel protein consisting of 608 amino acids, that resembled DRMs found in maize and Arabidopsis and designated this as Nt-DRM1. The protein could be shown to be localized exclusively in the nucleus and exhibit methylation activity toward unmethylated synthetic as well as native DNA samples upon expression in Sf9 insect cells. It also methylated hemimethylated DNA, but the activity was lower than that for unmethylated substrates. Methylation mapping of a 962-bp DNA, treated with NtDRM1 in vitro, directly demonstrated methylation of ~70% of the cytosines in methylatable CpNpN and CpNpG sequences but only 10% in CpG. Further analyses indicated that the enzyme apparently non-selectively methylates any cytosines except in CpG, regardless of the adjacent nucleotide at both 5' and 3' ends. Transcripts of NtDRM1 ubiquitously accumulated in all tissues and during the cell cycle in tobacco cultured BY2 cells. These results indicate that NtDRM1 is a de novo cytosine methyltransferase, which actively excludes CpG substrate.

The most commonly modified base in DNA among the eukaryotes through animals and plants is 5-methylcytosine (m5C). The resultant changes affect suppression of invading sequences of DNA, such as those in transgenes and retroelements (1), and management of endogenous gene expression via transcriptional repression (2). In plants, cytosine methylation has been proven to be essential for inactivation of transposable elements (3). Also, methylated cytosines are often seen in inactivated transgenes (4) and epigenetically silenced endogenes (5).

Methylation of cytosine residues in DNA is enzymatically catalyzed by DNA methyltransferases, which transfer a methyl group from S-adenosyl-L-methionine (AdoMet) to the 5-position. In mammals, two types of DNA methyltransferase, which differ in DNA substrate preference, have been reported. Those belonging to the Dnmt1 group prefer cytosines in hemimethylated CpG sites, i.e., CpG with m5C in only one strand (1), and are considered to be associated with the DNA replication complex in vivo (6) functioning in maintenance of methylation patterns. Those belonging to the Dnmt3 group are reported to methylate cytosines in unmethylated CpG (7) and have been suggested to establish the methylation pattern during embryonic development (7).

In plants, genes encoding three types of DNA methyltransferases have been reported so far. For example, in Arabidopsis, MET1, chromomethylase (CMT), and domains rearranged methyltransferase (DRM) are distinct (8). MET1 is homologous to mammalian Dnmt1 (9), and its suppression results in a drastic reduction of global methylation in transgenic Arabidopsis (10, 11). A similar reduction of global methylation and altered phenotypes was also observed in transgenic tobacco plants expressing an antisense tobacco MET1 (NAMET1) gene (12). Thus, MET1 has been suggested to function in maintenance of global genomic methylation in plants (10–12). The other two types are unique to plants; putative proteins belonging to the CMT group have a chromodomain in catalytic motifs and are reported to be responsible for maintenance of cytosine methylation at CpNpG sites in, for example, retrotransposons (where N is A, T, or C) (13, 14). Amino acid sequence analysis indicated that the DRM type has catalytic motifs, thus resembling mammalian de novo enzymes such as Dnmt3 (15), although they differ in possessing a characteristic rearrangement in catalytic motifs, between I-V and VI-X. A recent genetic analysis with mutant lines suggested that Arabidopsis DRMs might be responsible for the methylation of cytosines in CpNpG and asymmetric sequences of transgenes (16). Asymmetric cytosine methylation was also seen in epigenetically silenced loci, with DRM further suggested to function in epigenetic gene silencing (15, 16). However, no studies on biochemical properties of these proteins have been reported so far. In this article, we describe isolation of tobacco DRM. The enzyme expressed in insect cells could be shown to preferentially methylate cytosine residues in CpNpG and also CpNpG, providing concrete evidence for the predicted function of DRMs.

**EXPERIMENTAL PROCEDURES**

**Plant Materials**—Tobacco plants (Nicotiana tabacum cv. Xanthi) were grown in a greenhouse under natural day length conditions. Tobacco BY2 cells (N. tabacum cv. Bright Yellow 2) were maintained in suspension culture using modified Linsmaier and Skoog medium (LS medium) (17) at 23 °C continuously in the dark. The synchronization of
BY2 cells was performed as described (12) with modification. After culture in a modified LS medium containing 5 μg/ml 1-aminohexalin (Wako, Tokyo) for 24 h, cells were collected and washed with 3% sucrose solution, transferred to fresh medium, and cultured further. They were then harvested by centrifugation and stained with 1% orcinol to allow determination of the mitotic index.

Isolation of NtDRM1 cDNA—Initially, a DNA fragment encoding NtDRM1 was obtained by differential display of a cDNA population derived from wounded tobacco leaves (18). A full-length 2.4-kb cDNA of NtDRM1 was isolated by colony hybridization of a tobacco (N. tabacum) cDNA library using the NtDRM1 fragment as a probe and sequenced for both strands using a Big Dye terminator sequencing kit (Applied Biosystems, Foster City, CA). Motif prediction was performed using the online PROSITE (http://www.expasy.org/psicat and ScanProsite (hits.isb-sib.ch/cgi-bin/PSICAN). Phylogenetic analysis was accomplished with ClustalW (clustalw.genome.ad.jp).

Transformation of BY2 Cells and Histochemical Analysis—The NtDRM1 coding region (1852 bp) was fused in-frame via an engineered NcoI site to the N terminus of the green fluorescent protein (GFP) open reading frame. The construct was then cloned into the SinI and EcoRI sites of pBI121 (19) and transformed into Agrobacterium tumefaciens strain EHA105 (20) by the heat-shock method. Following a 2-day co-

DNA Methyltransferase Assay—DNA methyltransferase activity was measured as described above, was incubated at 37 °C for 16 h and then further with 110 ng of RNase A (Nacalai Tesque, Kyoto, Japan) at 37 °C for 2 h. DNA was extracted with phenol/chloroform, precipitated with ethanol, and digested with 2 units of nuclease P1 (Sigma) in 100-μl buffer containing 3 mM sodium acetate (pH 5.4) and 0.5 mM ZnSO4 at 37 °C for 2 h. The resultant nucleotides were dephosphorylated with 20 units of calf intestine alkaline phosphatase (Takara, Osaka, Japan) at 37 °C for 2 h. Samples were then fractionated by ultratrafuge (Ultrafree-MC PL-10 microcentrifuge tubes, Millipore, Bedford, MA), and the permeate was injected into a Superoil LC-18S-NaAm (Supelco, Bellefonte, PA). DNA methylation was performed as described with 2.5–20% methanol gradient in the presence of 50 mM K3PO4, (pH 4.3).

DNA Methylation Mapping—Substrate DNA was prepared by PCR-amplifying a 962-bp fragment of pGEX-4T1 (positions 4521–515, containing the position 1) then methylated in vitro in a reaction mixture containing 25 μg of DNA, 2 μg AdoMet (Sigma), and 93.75 μg of NtDRM1 in methylation buffer, as described above, at 37 °C for 18 h. After phenol extraction and precipitation with ethanol, DNA was subjected to bisulfite modification (24). The method was developed to identify mC in arbitrary sequences, based on the resistance of mC to bisulfite, which changes C into U. After bisulfite treatment, the modified DNA can be amplified by PCR, cloned, and directly sequenced, the C and mC in the initial sequence being replaced with T and G, respectively. Experimentally, modified DNA was subjected to first PCR using EcoTaq<sup>®</sup> enzyme (Takara) and specific forward, 5′-GGTTGTTGATTTG-GAATGTAATGG-3′, and reverse, 5′-CAACACCACAAATATTTACTATACAAAC3′ primers. Both were designed for predicted sequences after modification, in which C was assumed to be converted into T in the former and G into A in the latter. The resulting product was subjected to subsequent second PCR with the same forward primer and another reverse primer, 5′-CAGAAACCTTTAATCCTACTACCTAC-3′ (G was replaced with A). Amplified 594-bp DNA samples were ligated to the pGEM-T easy vector<sup>®</sup> (Promega, Madison, WI) and cloned in DH5α (Stratagene, La Jolla, CA). Sequences were determined with an ABI PrismDye Terminator Ready Reaction Kit and an ABI PRISM 377 DNA Analyzer automated sequencer (Applied Biosystems). Fifteen clones were sequenced, and the average methylation at each site was calculated.

DNA and RNA Isolation and Gel Blot Hybridization—Genomic DNA was extracted by the cetyltrimethylammonium bromide method (25). Total RNA was isolated from indicated tissues or from BY2 cells by the acid guanidinium/phenol/chloroform method (26), and hybridization analyses were performed as described (27) with probes synthesized with a primer of primers specific for each gene.

RESULTS

Identification of NtDRM1—During screening for genes whose transcripts accumulate in the early stage after wounding of tobacco leaves by the modified differential display, a particular fragment of 1.6 kb was identified. Although the clone was found later not necessarily to be specific to the wound response by Northern hybridization, it was further characterized because homology searches indicated resemblance to genes for DNA methyltransferases. Subsequent screening of a tobacco cDNA library yielded a full length of 2,540-bp cDNA, encoding a protein of 608 amino acids. Sequence analysis showed it to contain at least six highly conserved motifs found in DNA methyltransferases, but their order was unusual. In contrast to the majority of eukaryotic DNA methyltransferases,
having conserved motifs that are arranged in the order of I-IV-VIII-X in the C terminus, the present putative protein was found to possess motifs in the rearranged order of VI-VIII-IX-X-I-IV (Fig. 1A). A homology search indicated high homology to DRMs from Arabidopsis (DRM1 and DRM2) and maize (Zmet3) (15) (Fig. 1B). Consequently, we concluded that the isolated gene encodes a DRM, and we designated it as NtDRM1 (N. tabacam domains rearranged methyltransferase 1). In the N-terminal region, two domains (amino acid positions 61–97 and 166–204) were found that resembled the ubiquitin association domain; UBA) domains of human p62 and yeast RAD23 (Fig. 1A). UBA domains, considered to function in protein-protein interactions, have also been identified in DRMs of Arabidopsis and maize (15). A nuclear localization signal is present at amino acid 234–237 (Fig. 1A). A phylogenetic tree generated reflecting the structural features. The apparent Km values for poly(dI-dC) were 2.58 ± 0.03 and 6.36 ± 0.53 μM (in methylatable cytosine mononucleotides), respectively, and the Vmax values were calculated to be 178.9 ± 10.5 and 320 ± 58.9 fmol/min/nmol protein, respectively (Fig. 3C). The ratio of methyl incorporation to the substrate cytosines in poly(dI-dC) was estimated to be around 1:10^5 under the experimental conditions with low concentrations of enzyme and AdoMet and a reaction time of only 30 min. The results indicate de novo cytosine methylation by NtDRM1.

Sequence Specificity—In order to determine the sequence specificity for NtDRM1-mediated methylation, three synthetic

FIG. 1. Comparison of plant DNA methyltransferases. A, schematic illustration of DNA methyltransferase structures. The size of each protein is indicated in amino acid numbers (aa), and conserved motifs in the catalytic region are indicated by closed boxes with numbers. Specific regions in the regulatory region are indicated by shaded boxes with appropriate names. Glu-rich, glutamine-rich acidic region; BAH, bromo-adjacent homology domain; CD, chromodomain; NLS, nuclear localization signal; UBA, ubiquitin association domain. B, phylogenetic relationships among DNA methyltransferases. Sequence data were obtained from the data base; accession numbers are as follows: AF242320 (Zmet3), AAT5M21 (DRM1), AF240695 (DRM2), AF068625 (Dnmt3a), AF068628 (Dnmt3b), P13864 (Dnm1), P23866 (Fmu), AF348971 (DIM-2), P34881 (MET1), AF039372 (CMT1), and AL021711 (CMT3). E. coli RNA methylase Fmu is used as an external reference.

having conserved motifs that are arranged in the order of I-IV-VIII-X in the C terminus, the present putative protein was found to possess motifs in the rearranged order of VI-VIII-IX-X-I-IV (Fig. 1A). A homology search indicated high homology to DRMs from Arabidopsis (DRM1 and DRM2) and maize (Zmet3) (15) (Fig. 1B). Consequently, we concluded that the isolated gene encodes a DRM, and we designated it as NtDRM1 (N. tabacam domains rearranged methyltransferase 1). In the N-terminal region, two domains (amino acid positions 61–97 and 166–204) were found that resembled the ubiquitin association (UBA) domains of human p62 and yeast RAD23 (Fig. 1A). UBA domains, considered to function in protein-protein interactions, have also been identified in DRMs of Arabidopsis and maize (15). A nuclear localization signal is present at amino acid positions 234–237 (Fig. 1A). A phylogenetic tree generated reflecting the structural features. The apparent Km values for poly(dI-dC) were 2.58 ± 0.03 and 6.36 ± 0.53 μM (in methylatable cytosine mononucleotides), respectively, and the Vmax values were calculated to be 178.9 ± 10.5 and 320 ± 58.9 fmol/min/nmol protein, respectively (Fig. 3C). The ratio of methyl incorporation to the substrate cytosines in poly(dI-dC) was estimated to be around 1:10^5 under the experimental conditions with low concentrations of enzyme and AdoMet and a reaction time of only 30 min. The results indicate de novo cytosine methylation by NtDRM1.

Sequence Specificity—In order to determine the sequence specificity for NtDRM1-mediated methylation, three synthetic
oligonucleotides, with different sequences but the same base composition, were prepared. Each palindromic 24-mer contained 6 sites for CpG, CpNpG, or CpNpN, respectively, where N is A or T. The highest methylation was seen with the CpNpN substrate, with H incorporation increasing almost linearly with incubation period up to 1.5 h, whereas methyl transfer efficiency was lower with CpNpG than CpNpN and almost null with the CpG substrate (data not shown). The effect of substrate concentration on enzymatic activity was then examined with a fixed concentration of the enzyme and varying concentrations of DNA (Fig. 4A). The best substrate was CpNpN, with a substrate-dependent hyperbolic increase in the reaction velocity (Fig. 4A). The apparent K_m and V_max values were 45.9 ± 3.04 nM and 224.4 ± 4.12 fmol/min/mmol protein (based on the methylatable cytosine mononucleotides), respectively. The molar ratio of methyl group incorporated to the substrate was 1/10^3. The transfer rate of the methyl group into CpNpG was 2/3 of that into CpNpN and almost zero into CpG (Fig. 4A). The apparent K_m and V_max values were 78.3 ± 5.26 nM and 192.2 ± 4.62 fmol/min/mmol protein (based on the methylatable cytosine mononucleotides), respectively. These kinetic experiments suggested NtDRM1 to preferentially methylate cytosines in non-CpG sequences.

Effects of hemimethylation were then analyzed. A synthetic 28-mer oligonucleotide containing either CpG or CpNpG was prepared, in which all cytosines were substituted with m^3C. A non-methylated complementary strand was also generated and annealed to form a double-stranded substrate. Methyl transfer activity was then assayed with a fixed concentration of the enzyme and varying concentrations of the substrate (Fig. 4B). These kinetic analyses showed the enzyme to be active on the hemimethylated CpNpG at a lower velocity than on the unmethylated CpNpG substrate. By taking account of the fact that the number of available cytosines in the former is only half that in the latter, the K_m and V_max values were calculated to be 21.2 ± 2.37 nM (based on the methylatable cytosine mononucleotides) and 90.9 ± 10.8 fmol/min/mmol protein, respectively, for the hemimethylated substrate. The K_m and V_max values for the unmethylated substrate were 10.9 ± 1.50 nM (based on the methylatable cytosine mononucleotides) and 136.1 ± 13.4 fmol/min/mmol protein, respectively. These values indicate that the specificity of the enzyme is lower toward the hemimethylated CpNpG than toward the unmethylated CpNpG substrate. The methylation patterns of both hemimethylated and unmethylated CpG substrates were similar to but lower than that of CpNpG, showing apparent K_m values of 37.0 ± 8.54 and 27.3 ± 3.34 nM, respectively (Fig. 4B). The results suggested that the enzyme does not recognize the hemimethylated state of CpG and CpNpG and
untreated sequence were converted into thymine (T) (Fig. 5). Sequences and found to be nearly complete as all cytosines in the cytosine (C) into uracil (U) was directly estimated by aligning clones, and directly sequenced. The conversion efficiency of cytosines, modified by the bisulfite method, amplified with PCR, was therefore confirmed its non-maintenance, de novo properties.

Direct Methylation Mapping—The sequence specificity was directly determined by methylation mapping. Native DNA was intensively methylated in vitro with excess AdoMet and the enzyme, modified by the bisulfite method, amplified with PCR, cloned, and directly sequenced. The conversion efficiency of cytosine (C) into uracil (U) was directly estimated by aligning sequences and found to be nearly complete as all cytosines in the untreated sequence were converted into thymine (T) (Fig. 5A). The methylation frequencies were estimated for 15 clones, as-

FIG. 3. Production and properties of NtDRM1. A, in vitro expression of NtDRM1 protein. NtDRM1 was fused to the baculovector, pDEST20, transformed into Sf9 cells, and amplified. Then crude extracts (lane 1) or purified proteins after passage through a GST column (lane 2) were subjected to SDS-PAGE and stained with Coomassie Brilliant Blue (CBB). Proteins were blotted onto cellulose membrane and subjected to immunostaining with anti-GST antibodies (lane 3). The sizes of proteins are indicated at the left side in kDa. B, DNA methyltransferase assay. To estimate enzyme concentration, crude extracts were subjected to immunoblot staining with anti-GST antibodies, and signals were densitometrically quantified (upper panel). When the intensity for Dmnt3a was taken as 1, the relative intensities for NtDRM1 and β-glucuronidase (GUS) proteins were 2.37 and 2.16, respectively. Enzymatic activities were normalized based on these values. For activity assays, a reaction mixture containing 2 μl [3H]AdoMet, 2 μg of poly(dI-dC)-poly(dI-dC), and 10 μl of crude protein preparation as indicated was incubated for 6 h, spotted onto DEAE filter paper, washed, and counted for radioactivity. A mixture containing GST-fused GUS was used as the control. Values are expressed as the means of triplicate experiments, and error bars represent the standard deviations (lower panel). C, DNA methylation activity of GST-NtDRM1 fusion protein. A reaction mixture containing the indicated concentrations of poly(dG-dC)-poly(dG-dC) (open circles) or poly(dI-dC)-poly(dI-dC) (closed circles), 2 μM [3H]AdoMet, and 36 ng purified GST-NtDRM1 protein was incubated at 37°C for 30 min. The amounts of transferred methyl group were calculated based on the specific activity of [3H]AdoMet (7.7 × 10^4 cpm/mmol). Experiments were repeated three times, and mean values were estimated with standard deviations.

FIG. 4. Sequence specificity of NtDRM1. A, sequence specificity and effects of substrate concentration. The sequences of 24-mer oligonucleotides used in this assay were 5'-ACGATCGATCGATCGATCGATCGATCGT-3' for CpG (open circles), 5'-ACTCGATCTGATCGATCGATCG- CAGT-3' for CpNpG (open squares) (where N is A or T), and 5'- AGATCGATCGATCGATCG-3' for CpNpN (open triangles). A reaction mixture containing 36 ng of purified NtDRM1 protein, 2 μM [3H]AdoMet, and the indicated amount of synthetic oligonucleotides was incubated for 30 min and processed as described above. Experiments were repeated three times, and mean values were estimated with standard deviations. B, effects of hemimethylated substrate. The hemimethylated duplex sequences of 28-mer oligonucleotides used in this assay were synthesized as described in the text. The number of methylatable cytosines was 5 in one strand, and thus contained 5 and 10 sites in each duplex of hemimethylated and unmethylated substrate, respectively. Substrates were unmethylated CpG (umCpG, open circles), hemimethylated CpG (hmCpG, closed circles), unmethylated CpNpG (umCpNpG, open squares), and hemimethylated CpNpG (hmCpNpG, closed squares). Assays were performed as described above.

signed to each cytosine, and expressed as percentages (Fig. 5B). The enzyme showed the highest methylation of cytosines in CpNpG followed by CpNpP and least in CpG. The percentages of methylated sites in the total in 15 clones were ~75% for CpNpP and 60% for CpNpG, in sharp contrast with the 10% for CpG (Table I). When nucleotide triplets were classified, and the nearest neighbor was taken into account, the average methylation rates for CpT, CpA, and CpC were 87, 75, and 70%, respectively (Table II). It appears that the third nucleotide, including guanine, does not appreciably influence the methylation activity, although T located at the 3' end of CpC and CpT reduced the frequency (Table II). The nucleotide located at the 5' end of target cytosines showed no apparent effects on site specificity (Table III). These observations suggest that NtDRM1 essentially recognizes and methylates all cytosines, although some particular combinations, such as CpG and CpCpT, appear to be less favorable.
In Vivo Methylation—To examine whether or not NtDRM1 functions in vivo, native DNA from Sf9 was analyzed by HPLC (Fig. 6). Samples from Sf9 cells expressing NtDRM1 were isolated, digested with nuclease P1, dephosphorylated with alkaline phosphatase, and assayed by HPLC. As the control, DNA from wild-type Sf9 cells was used. The positive control was prepared by methylating DNA in vitro by NtDRM1, and the negative control was intact Sf9 DNA. HPLC fractionation of the resulting nucleosides showed that DNA samples prepared from transformed cells yielded m^5C at 1.8% of total cytosines, showing a similar elution profile with the control in vitro methylated DNA containing m^5C, at 2.7% of total cytosines (Fig. 6, B and C). The low amount of m^5C from transformed cells may be due to the expression system, in which Sf9 cells are dying when the protein is produced. However, since intact DNA from untransformed cells did not show any m^5C (Fig. 6D), the results suggest Nt-DRM1 to be active as a de novo cytosine methyltransferase in vivo.

**Table II**

Methylation frequency in triplets containing m^5C at 5′ end

The sums of total m^5C numbers in the total methylatable cytosines for the indicated triplets in 15 clones are presented as numbers observed and as percentages, as described in the legend for Table I. Asterisk indicates methylated cytosines.

| Triplet       | m^5C/total C | Percentage |
|---------------|--------------|------------|
| C*pApA        | 78/120       | 65.0       |
| C*pApC        | 130/165      | 78.8       |
| C*pApG        | 111/150      | 74.0       |
| C*pApT        | 110/135      | 81.5       |
| Total C*pA    | 429/570      | 75.2       |
| C*pCpA        | 90/120       | 75.0       |
| C*pCpC        | 105/120      | 87.5       |
| C*pCpG        | 88/120       | 73.3       |
| C*pCpT        | 51/120       | 42.5       |
| Total C*pC    | 334/480      | 69.6       |
| C*pGpA        | 6/165        | 3.6        |
| C*pGpC        | 6/135        | 4.4        |
| C*pGpG        | 10/180       | 5.5        |
| C*pGpT        | 43/165       | 26.1       |
| Total C*pG    | 66/145       | 10.1       |
| C*pTpA        | 70/75        | 93.3       |
| C*pTpC        | 85/90        | 94.4       |
| C*pTpG        | 224/255      | 87.8       |
| C*pTpT        | 38/60        | 63.3       |
| Total C*pT    | 417/480      | 86.9       |

**Table III**

Methylation frequency in triplets containing m^5C at 3′ end

The sums of total m^5C numbers in the total methylatable cytosines for the indicated triplets in 15 clones are presented as numbers observed and as percentages, as described in the legend for Table I. Asterisk indicates methylated cytosines.

| Triplet       | m^5C/total C | Percentage |
|---------------|--------------|------------|
| ApApC*        | 48/90        | 53.3       |
| CpApC*        | 96/165       | 58.2       |
| GpApC*        | 33/75        | 44.0       |
| TpApC*        | 84/120       | 70.0       |
| Total ApC*    | 261/450      | 58.0       |
| ApCpC*        | 44/60        | 73.3       |
| CpCpC*        | 89/120       | 74.2       |
| GpCpC*        | 51/150       | 34.0       |
| TpCpC*        | 73/135       | 57.8       |
| Total CpC*    | 282/465      | 58.3       |
| ApGpC*        | 70/120       | 58.3       |
| CpGpC*        | 87/135       | 64.4       |
| GpGpC*        | 105/240      | 42.9       |
| TpGpC*        | 105/180      | 58.3       |
| Total GpC*    | 365/675      | 54.1       |
| ApTpC*        | 109/165      | 66.1       |
| CpTpC*        | 54/90        | 60.0       |
| GpTpC*        | 54/120       | 45.0       |
| TpTpC*        | 133/150      | 88.7       |
| Total TpC*    | 350/525      | 66.7       |
Genomic Organization and Transcript Accumulation—
Genomic DNA of tobacco (N. tabacum) was digested with appropriate restriction enzymes and hybridized with a specific probe prepared from the 3'-untranslated region of NtDRM1 (positions 2064–2493). Under high stringency conditions, the probe showed four distinct signals after digestion with BamHI, EcoRI, and HindIII, respectively (Fig. 7A). Because none of these restriction sites was present in the probe sequence, and since N. tabacum is an amphidiploid, it is conceivable that NtDRM1 forms a multi-gene family existing probably as pairs in each chromosome set originating from its ancestor lines, Nicotiana sylvestris and Nicotiana tomentosiformis. RNA hybridization analyses indicated Nt-DRM1 transcripts to accumulate ubiquitously in leaves, stems, flowers, and roots (Fig. 7B). Levels were also high in all floral organs except pistils (Fig. 7B). During the cell cycle of synchronously cultured BY2 cells, NtDRM1 transcripts accumulated throughout (Fig. 7C), in marked contrast to those for NtMET1 encoding a maintenance methyltransferase, expressed predominantly in the S-phase (Fig. 7C).

DISCUSSION
Methylation of DNA is characterized by two distinct features in eukaryotes: maintenance of the preexisting methylation patterns and methylation of previously unmethylated sites, the responsible enzymes being referred to as maintenance and de novo methyltransferases, respectively. In plants, proteins belonging to the former type have been biochemically characterized in several species (8), but examples of the latter have not been isolated so far. The presently identified NtDRM1 catalyzed methylation of both synthetic oligonucleotides and native DNA in vitro. Hemimethylated substrates were also methylated at a low efficiency. When expressed in insect cells, it methylated host DNA in vivo. It was thus concluded that Nt-
One of the notable features of NtDRM1 is its de novo methyltransferase function, which allows for the reversion of DNA methylation through cell division. This process is mediated by demethylases, which convert m^5^C to cytosine (30). In contrast, the maintenance type methyltransferase, which recognizes m^5^CpG, is required for the propagation of DNA methylation through cell division and is almost exclusively methylated in symmetric CpG sequences, at a frequency of over 70% (31). The identification of a cytosine-recognizing domain in the MeCP2 protein has suggested that some amino acid sequences in MeCP2 may interact specifically with Arg (Arg-133 and Arg-111) in MeCP2 (35). It is of interest to examine whether some amino acid sequences of MeCP2 provide biochemical evidence that, at least in tobacco plants, the DRMs so far identified from maize, Neurospora crassa, and soybean Arabidopsis are capable of catalyzing cytosine methylation in asymmetrical sites. Reverse genetic analyses have suggested that methylation of CpG sites and have DRMs. Identification of a cytosine-recognizing domain in NtDRM1 would substantiate this speculation.

The biological significance of selective methylation at asymmetric cytosines is not completely clear but may be of particular importance for establishment of cell-specific methylation patterns (36). Asymmetric cytosine methylation would provide an opportunity for individual cells to establish independent methylation patterns in response to environmental conditions. An example is the de novo cytosine methylation at CpNpG sites in DNA-dependent DNA methylation, which has been inferred to function in cosuppression of plant genes upon introduction of foreign DNA (37, 38). It was reported recently that the demethylation enzyme, m^5^C-specific glycosylase, regulates gene expression (39, 40). In this case, methylation-mediated suppression is alleviated by demethylation, thereby reversibly controlling gene expression in the cell. A similar cycle of methylation-demethylation of genomic DNA can also be achieved through asymmetric methylation and subsequent passive demethylation through cell division.

It must be mentioned that the putative amino acid sequences of DRMs so far identified from maize, Arabidopsis and soybean indicated the presence of a UBA domain (15), shown to be a conserved region for protein-protein interactions (41). Because NtDRM1 also possesses a UBA domain, it is constitutively expressed throughout the cell cycle and in all tissues, it is conceivable that it forms complexes with multiple proteins that contribute to chromatin structure and thereby methylate particular regions of DNA in a de novo fashion.
Preferential de Novo Methylation of Cytosine Residues in Non-CpG Sequences by a Domains Rearranged DNA Methyltransferase from Tobacco Plants
Yuko Wada, Hitoshi Ohya, Yube Yamaguchi, Nozomu Koizumi and Hiroshi Sano

J. Biol. Chem. 2003, 278:42386-42393.
doi: 10.1074/jbc.M303892200 originally published online August 12, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M303892200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 40 references, 14 of which can be accessed free at http://www.jbc.org/content/278/43/42386.full.html#ref-list-1