Multiple Isoforms of DNA Methyltransferase Are Encoded by the Vertebrate Cytosine DNA Methyltransferase Gene*

(Received for publication, May 26, 1998, and in revised form, July 2, 1998)

Jing Deng and Moshe Szyf‡
From the Department of Pharmacology and Therapeutics, McGill University, Montreal, Quebec H3G 1Y6, Canada

This manuscript tests the hypothesis that multiple forms of cytosine-DNA methyltransferase (MeTase) are expressed in vertebrates in vivo. Vertebrate genomes are distinguished by tissue- and gene-specific DNA methylation patterns. Specific methylation patterns are believed to encode epigenetic information. In distinction from the remarkable diversity of DNA methylation patterns, only one functional DNA MeTase cDNA has been identified to date in different vertebrate organisms. Using reverse transcription-polymerase chain reaction and RNase protection analyses, we show that the methyltransferase domain of the rat DNA MeTase is alternatively spliced in vivo, generating different in-frame variants of DNA MeTase in specific tissues. This process is developmentally regulated and is induced in PC12 cells by a known inducer of neuronal differentiation, nerve growth factor. The data presented here point toward a new mechanism for generating diversity of DNA MeTases and possibly diverse DNA methylation patterns.

Tissue-specific DNA methylation patterns are a hallmark of vertebrate genomes (1, 2). Approximately 80% of the cytosines residing at the dinucleotide CG sequence are methylated; however, the identity of CG sequences that are methylated varies from tissue to tissue and at different developmental stages of the same cell type (2). It is now clear that DNA methylation patterns play an important role in encoding the epigenetic information of vertebrate genomes by regulating gene expression as well as other genome functions (3). Recent data have also associated aberrations in regulation of DNA methyltransferase with the oncogenic process (4, 5). It is clear that to understand how epigenetic information is generated and maintained, it is critical to unravel what regulates the formation of DNA methylation patterns. The principal mystery that has not been resolved yet is the fact that only one DNA MeTase has been identified in mammalian cells by molecular cloning, which can potentially methylate any CpG site (6). An additional putative DNA methyltransferase has recently been cloned, but its ability to transfer methyl groups to DNA has not yet been demonstrated (7). How can one gene or possibly a small number of genes encoding DNA MeTase activity be responsible for generating all these diverse patterns of methylation?

The most attractive and obvious explanation for the diversity of DNA methylation patterns is that a large number of DNA MeTases are encoded by the vertebrate genome, each with its distinct specificity and pattern of expression. Nevertheless, any previous attempts to identify additional DNA MeTase genes has been unsuccessful. However, there is an additional widespread biological principle that might allow the vertebrate genome to encode different mRNAs for a given gene, alternative splicing (8). It has recently been shown that sex-specific 5′ exons of DNA MeTase are expressed in mammalian germ cells (9).

The known vertebrate DNA MeTase is a large protein that is believed to be composed based on its sequence similarity to other cytosine DNA methyltransferases, from at least three structural components (10): A DNA methyltransferase domain at the 3′ end, an N terminus domain that is responsible for localization of the protein to the nucleus and the replication fork, and a poorly characterized domain in its central part (10, 11). It is generally believed that one can distinguish three functional regions in bacterial MeTases as well as in the methyltransferase domain of the mammalian enzyme: the evolutionarily conserved catalytic motif and S-adenosyl L-methionine-binding regions and the variable target recognition region (10, 12). Swapping of alternative variable domains has been shown to alter the target recognition sequence of some bacterial DNA MeTases (13). The answer to the question of whether alternative splicing of vertebrate DNA MeTase mRNA is a feasible mechanism is dependent on whether the gene structure is interrupted by a number of introns, generating alternative potential splice partners. The physical structure of the human DNA MeTase gene has recently been resolved (14) and is consistent with this hypothesis. The 3′ DNA methyltransferase catalytic domain is comprised of eleven exons. Four exons encode the previously proposed variable domain, which is believed to be responsible for sequence specificity. This domain is located in between the conserved domains of DNA MeTase and is a potential hot spot for splicing.

We have used the rat as a model system because it allows easy access to in vivo material and the rat PC12 cell line as a well defined in vitro differentiation model (15) to test the hypothesis that developmentally regulated alternative splicing is involved in the processing of DNA MeTase transcripts. The data presented in this paper suggest that one mechanism utilized by the vertebrate genome for generating diversity of DNA MeTases is alternative splicing.

MATERIALS AND METHODS

Cloning of the Rat cDNA and Genomic Fragment Encoding DNA MeTase—The rat cDNA bearing the DNA methyltransferase domain was amplified from PC12 cells (ATCC CRL 1721) and different rat tissue RNA (1 μg) by reverse transcription with Superscript reverse transcriptase (Life Technologies, Inc.) followed by PCR amplification (Promega Taq) using a sense primer corresponding to the mouse DNA MeTase cDNA sequence (3372–3391) 5′-CTGTGGGCCCCATCGAGATGTT-3′ (oligo 146), an antisense primer (4420–4400) 5′-TAGAGCC-CAGCCACAGGTGTT-3′ (oligo 147 A) (16). The amplification products

* This work was supported by the National Cancer Institute of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Pharmacology and Therapeutics, McGill University, 3655 Drummond St., Montreal, PQ H3G 1Y6, Canada.

The abbreviations used are: MeTase, methyltransferase; PCR, polymerase chain reaction; RT, reverse transcription; NGF, nerve growth factor; bp, base pair(s).
Multiple Isoforms of DNA MeTase

were subcloned into the vector PCR 2.1, sequenced, and found to show as expected high identity to human and mouse DNA MeTase (17, 18). To clone the genomic fragment, 3 ng of PC12 DNA were subjected to PCR amplification with oligos 146 and 147A described above as primers using Taq plus long from Promega according to the manufacturer’s specifications. The amplification product was subcloned into PCR 2.1, and the exons and exon-intron boundaries were sequenced using exon-specific oligonucleotides as primers (data not shown).

RESULTS

New Isoforms of the DNA Methyltransferase Domain of DNA MeTase Are Expressed upon Differentiation of PC12 Cells with NGF and in Somatic Tissues in Vivo—We tested the hypothesis that differential splicing of DNA MeTase methyltransferase domain occurs during differentiation by reverse transcribing and amplifying the DNA methyltransferase domain of DNA MeTase mRNA from NGF-treated PC12 cells at different time points after induction of differentiation. Surprisingly, as differentiation progresses, PC12 cells express multiple messages of different sizes than the one predicted by the cloned cDNA sequence (Fig. 1A). To demonstrate that the changes in size reflect alternative splicing of the DNA MeTase mRNA, the Southern blots were hybridized to the two nested labeled oligonucleotides (233 and 235). As shown in Fig. 1A, some fragments bear both oligonucleotides, whereas others bear only the 5’ 233 oligonucleotide. The fact that the amplified DNA MeTase cDNAs bear the 3′ and 5′ amplification primers (by sequencing) and different nested primers and are shorter than predicted is strongly consistent with alternative splicing of the DNA methyltransferase domain occurring during differentiation of PC12 cells.

To test the hypothesis that developmentally regulated splicing occurs in vivo, DNA MeTase mRNAs expressed in embryonal tissues were compared with adult tissues using a similar RT-PCR assay. As observed in Fig. 1B, the majority of the mRNA encoding the DNA methyltransferase domain in embryonal brain and liver is of the expected size (1047 bp). However, multiple fragments of different sizes appear in different adult tissues, suggesting that multiple splice variants of the DNA methyltransferase domain of MeTase are expressed upon full differentiation into somatic tissues, similar to the results observed with PC12 cells (Fig. 1A).

Variant Isoforms of DNA Methyltransferase Domain Are Detected by an RNase Protection Assay—To verify that the changes observed using PCR are not a consequence of a flaw in our PCR-based assay, we have performed an RNase protection assay on RNA prepared from the adult rat tissues indicated in the figure using an RNA probe bearing exon 32 of the DNA methyltransferase domain of DNA MeTase mRNA (Fig. 2). DNA MeTase mRNA spliced at alternative positions in exon 32 will not protect the entire probe but rather the parts of it that are still present on the splice variant (Fig. 3A). The DNA methyltransferase domains 4, 6, and 8 and the variable domain are encoded by six exons, which is identical to the structure of the human gene (14).

The cloned DNA MeTase catalytic domain variants (PCR products) were sequenced to determine whether the different variants are bona fide products of alternative splicing and whether they are in frame with the previously characterized coding sequence of DNA MeTase. We have identified seven different splice variants of DNA MeTase (Fig. 3A and B). The position of the splice donor and acceptor sites of the different splice products are indicated in Fig. 3A and listed in Table I and show that both skipping of exons in the DNA methyltrans-
ferase domain and use of alternative splice sites within the same exon occur during the processing of the DNA MeTase transcript. Six different splice donor sites were identified within exon 32 encoding conserved motif IV (Fig. 2 for RNase protection of exon 32), and four different splice acceptor sites were identified in exon 36 encoding part of the variable domain. All splice variants are in frame with the rest of the coding sequence of the DNA MeTase. The alternative choice of splice sites results, in one instance, in a subtle change in two amino acids (SF4: RM to VC). However, in all other cases the alternative splicing results in deletion of different segments of the variable domain without alteration in the amino acid composition of the remaining sequence. We have not identified as of yet new exons in the DNA methyltransferase domain. Most splice variants bear the previously suggested catalytic proline cysteine dipeptide (Fig. 3B). However, in three examples (SF4, SF5, and SF7) the proline cysteine motif is deleted by the alternative splicing event. It is possible that the proteins encoded by these mRNAs do not bear MeTase activity. However, one cannot rule out that alternative proline cysteine dipeptides in the protein (17) substitute for the deleted dipeptide.

Tissue-specific Splicing of the DNA Methyltransferase Domain—The splice variants exhibit tissue specificity as indicated in Table I. Whereas splice variant SF2 is present in muscle, testis, and NGF-treated PC12 cells, SF3 is only found in the lung. It stands to reason that the splice variants presented in this study are just a sample of multiple putative DNA MeTase splice variants and that the list shown here is by no means a comprehensive list. However, both the RNase protection assay shown in Fig. 2 and the PCR analysis (Fig. 1B) support the hypothesis that DNA MeTase is differentially spliced in vivo.

**DISCUSSION**

One of the long-standing and fundamental questions in DNA methylation is “How are cell- and gene-specific patterns of methylation generated and maintained?” The specificity of the known cytosine DNA MeTase is more lax, and it can recognize any sequence that bears the dinucleotide CpG (6). Despite the indiscriminate nature of the enzyme, DNA methylation patterns exhibit remarkable gene and tissue specificity. It is tempting to speculate that diversification of the variable domain of mammalian DNA MeTases can lead to generation of multiple DNA MeTases with a common core sequence specificity to CpG with different sequence context specificities. There are different biological mechanisms that can lead to such diversification, gene duplication (20), gene rearrangement (21), or alternative splicing (22). For example, thousands of neurexin isoforms are generated from three genes by usage of alternative promoters and alternative splicing (22).

The data presented in this paper are the first body of evidence supporting the hypothesis that alternative in-frame splicing of the DNA methyltransferase domain encompassing the variable region is utilized by mammals to generate different forms of DNA MeTase mRNAs (Fig. 3). Thus, multiple DNA MeTases exist in mammals, but they are encoded by the same
Multiple Isoforms of DNA MeTase

The sequences of the alternative 5’ and 3’ splice boundaries and the tissues expressing them are indicated.

| Spliced forms | Size (bp) | Splicing boundaries | Tissues and cell lines | Number (sequenced) |
|---------------|-----------|---------------------|------------------------|--------------------|
| SF1           | 804       | CTTGCCGTGCA\(^a\)   | Cerebellum             | 1                  |
|               |           | GCTGGCAGGG          |                        |                    |
| SF2           | 354       | AAAAACTCCTC        | Muscle, testis, pc12/NGF-48h | 3, 1, 1            |
| SF3           | 291       | GTGGCGGAGAG        | Lung                   | 1                  |
| SF4           | 414       | AGAGGCTGCC          | Spleen, brain primary culture | 3, 1              |
| SF5           | 420       | GCTCTCTAAG         | Brain primary culture  | 1                  |
| SF6           | 378       | GGGCATCTCTC        | PC12/Control           | 1                  |
| SF7           | 180       | CCTCGTGCC          | PC12/NGF-48h           | 1                  |
|               |           | GGCGCAAGGG         |                        |                    |
|               |           | CGGACACACTG        |                        |                    |

\(^a\) The top sequence in each splice variant is the 5’ boundary, and the bottom sequence is the 3’ boundary.

A number of observations suggest the hypothesis that the results reported here do not reflect ectopic and biologically irrelevant phenomenon. First, alternative splicing is verified by two independent methods, RT-PCR and RNase protection. Second, all the splice variants are in frame with the known DNA MeTase coding sequence. It is highly improbable that all the in-frame splice variants reported in this study are generated randomly. Third, the alternative splicing exhibits tissue and developmental specificity.

One obvious question is the relative abundance of the new forms versus the known published DNA MeTase sequence (17, 18). The fact that a similar DNA MeTase has been cloned from different organisms suggests that the known cloned cDNA sequence is the most abundant form. This might seemingly lead one to discount the other forms that are presented at lower frequencies. However, if alternative splice variants play a role in delineating discrete methylation patterns, the relative abundance of each variant should be limited. We speculate that the difference in the relative abundance of the known DNA MeTase versus the splice variants reflects a bimodal control of DNA methylation patterns. The abundant form of the DNA MeTase is responsible for the bulk of maintenance DNA methylation, whereas the discrete splice variants are involved in the fine tuning of the methylation pattern. This might explain why the relative frequency of specific splice variants is increased during differentiation.

Lei et al. (23) have previously reported that embryonic cells bearing a null mutation of the known DNA MeTase exhibit stable low level methylation and are able to methylate viral DNA as well as wild-type cells, thus providing evidence for a second de novo DNA MeTase. The null mutation described by Lei et al. (23) removes the exons corresponding to exons 32 and 33 (Fig. 3A), which are naturally spliced in some of the alternative spliced DNA MeTase cDNA described in this study (Fig. 3). Thus, the mutated allele might still encode some of the normal alternatively spliced DNA MeTase isoforms. It is possible that some of these alternatively spliced DNA MeTase mRNAs bear the de novo methylation activity described in the null mutant. The main problem with this hypothesis is that these splice variants lack the predicted catalytic PC dipeptide. A possible explanation is that an additional conserved PC dipeptide that resides upstream from the catalytic domain at amino acids 584–585 might substitute for the bona fide PC dipeptide motif in the alternatively spliced DNA MeTase isoforms. Our data, however, do not exclude the possibility suggested by Lei et al. (23) that another DNA MeTase gene encodes the de novo DNA MeTase. Whereas our study opens up a large number of questions that should be resolved in future experiments, the data presented here lay out a new framework for understanding the diversity of DNA methylation patterns.

REFERENCES

1. Razin, A., and Riggs A. D. (1980) Science 210, 604–610
2. Razin, A., and Szyf, M. (1984) Biochim. Biophys. Acta 782, 331–342
3. Razin, A., and Cedar, H. (1991) Microbiol. Rev. 55, 451–458
4. El-Deiry, W. S., Nelkin, B. D., Celano, P., Yen, R. W., Falco, J. P., Hamilton, S. R., and Baylin, S. B. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 3470–3474
5. Szyf, M. (1994) Trends Pharmacol. Sci. 15, 233–238
6. Yoder, J. A., Suman, N. S., Verdone, G. L., and Bestor, T. H. (1997) J. Mol. Biol. 276, 385–395
7. Yoder, J. A., and Bestor, T. H. (1998) Hum. Mol. Genet. 7, 279–284
8. Choi, E., Kuehl, M., and Wall, R. (1980) J. Biol. Chem. 255, 776–779
9. Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M., and Bestor, T. H. (1998) Development 125, 889–897
10. Bestor, T. H. (1992) EMBO J. 11, 2611–2617
11. Leonard, H., Page, A. W., Weier, H. U., and Bestor, T. H. (1992) Cell 71, 865–873
12. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., and Wilson, G. G. (1994) Nucleic Acids Res. 22, 1–10
13. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) J. Mol. Biol. 253, 618–632
14. Ramchandani, S., Bigey, P., and Szyf, M. (1998) Biol. Chem. 379, 535–540
15. Greene, L. A., and Tischler, A. S. (1976) Proc. Natl. Acad. Sci. U. S. A. 73, 2424–2428
16. Kawasaki, E. W., and Wang, A. M. (1989) in PCR Technology (Erlich, H. A., ed.) pp. 89–97, M Stockton Press, New York
17. Bestor, T. H., Laudauno, A., Mattaliano, R., and Ingram, V. (1968) J. Mol. Biol. 203, 971–983
18. Yen, R. W., Vertino, P. M., Nelkin, B. D., Yu, J. J., el-Deiry, W., Cumarsawamy, A., Lennon, G. G., Trask, B. J., Celano, P., and Baylin, S. B. (1992) Nucleic Acids Res. 20, 2287–2291
19. Szyf, M., Rouleau, J., Theberge, J., and Bozovic, V. (1992) J. Biol. Chem. 267, 12813–12816
20. Nebert, D. W. (1979) Mol. Cell. Biochem. 27, 27–46
21. Robertson, M. (1982) Nature 297, 184–186
22. Missler, M., and Sudhof, T. C. (1998) Trends Genet. 14, 20–26
23. Lei, H., Oh, S. P., Okano, M., Juttnerman, R., Goss, K. A., Jaenisch, R., and Li, E. (1996) Development 122, 3195–3205