

Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues

(Dnmt1 transcript/alternative splicing/Alu family repeat/chimpanzee)

DUEN-WEI HSU*†‡, MENG-JAU LIN*†‡, TAI-LIN LEE*, SHAU-CHING WEN*, XIN CHEN*, AND C.-K. JAMES SHEN*†§¶

*Institute of Molecular Biology, Academia Sinica, Nankang, Taipei, Taiwan 115, Republic of China; †Institute of Genetics, National Yang-Ming University, Taiwan 112, Republic of China; and ‡Section of Molecular and Cellular Biology, University of California, Davis, CA 95616

Communicated by Sherman M. Weissman, Yale University School of Medicine, New Haven, CT, June 11, 1999 (received for review January 20, 1999)

ABSTRACT Thus far, only one major form of vertebrate DNA (cytosine-5) methyltransferase (CpG MTase, EC 2.1.1.37) has been identified, cloned, and extensively studied. This enzyme, *dnmt1*, has been hypothesized to be responsible for most of the maintenance as well as the *de novo* methylation activities occurring in the somatic cells of vertebrates. We now report the discovery of another abundant species of CpG MTase in various types of human cell lines and somatic tissues. Interestingly, the mRNA encoding this CpG MTase results from alternative splicing of the primary transcript from the *Dnmt1* gene, which incorporates in-frame an additional 48 nt between exons 4 and 5. Furthermore, this 48-nt exon sequence is derived from the first, or the most upstream, copy of a set of seven different Alu repeats located in intron 4. The ratios of expression of this mRNA to the expression of the previously known, shorter *Dnmt1* mRNA species, as estimated by semiquantitative reverse transcription-PCR analysis, range from two-thirds to three-sevenths. This alternative splicing scheme of the *Dnmt1* transcript seems to be conserved in the higher primates. We suggest that the originally described and the recently discovered forms of CpG MTase be named *dnmt1-a* and *dnmt1-b*, respectively. The evolutionary and biological implications of this finding are discussed in relation to the cellular functions of the CpG residues and the CpG MTases.

In vertebrates, including mammals, chromosomal DNAs are modified by C-methylation at a limited number of CpG dinucleotides, resulting in methylation at the 5' position of the C residues. Many studies have indicated that this methylation process and its product, ^mCpG, play indispensable regulatory roles for a number of genetic activities in vertebrate cell differentiation and embryo development. These activities include tissue-specific gene transcription, chromatin modification, and genomic imprinting (for reviews, see refs. 1–3). DNA methylation has also been implicated in neoplastic transformation (4–6). The structural and functional analysis of the enzymes responsible for the above methylation process, the CpG methyltransferases (MTases), is thus a critical step toward understanding vertebrate development.

Thus far, only one abundant CpG MTase enzyme has been identified in the somatic tissues of different vertebrates. The cDNA and gene encoding this enzyme, *Dnmt1*, have been cloned from different vertebrates including mouse (7, 8), human (8–10), rat (11), chicken (12), *Xenopus* (13), and others. *Dnmt1* shows a marked preference for hemimethylated DNA as its *in vitro* substrate of CpG methylation and, hence, is believed to be the enzyme responsible for the propagation of the methylation patterns through the cell generations (maintenance methylation). The C-terminal domain of *Dnmt1* is built on the same architecture as the bacterial C-methylation enzymes, which contain 10 conserved motifs involved in catalytic activities (7, 14). The

N-terminal domain (Fig. 1) contains the nuclear localization signal, the replication foci-targeting sequences (15–17), and a region interacting with the replication protein PCNA (18). The critical function of *dnmt1* in development has been shown by the targeted mutation of the mouse gene. Embryonic stem cells homozygous for a disruption of *Dnmt1* proliferate normally but with their DNA highly demethylated. They also die on induced differentiation *in vitro* (19).

Recently, several minor species of mammalian CpG MTase have been identified and cloned. They have been designated *dnmt2*, *dnmt3-a*, and *dnmt3-b* (20–23). Of these, *dnmt2* contains 6 of the 10 conserved motifs mentioned above, but it lacks the large N-terminal domain of the vertebrate CpG MTases. *Dnmt2* mRNA is expressed in all cell types at very low levels, and it is essential for neither *de novo* nor maintenance methylation activities in embryonic stem cells (20, 21). *Dnmt3-a* and *Dnmt3-b* are expressed abundantly in embryonic stem cells, but, similar to *Dnmt2*, they are at very low levels in differentiated embryoid bodies and the adult tissues. The C-terminal regions of *dnmt3* enzymes contain the conserved motifs of C-methylation enzymes, but their N-terminal regions show no sequence homology to *dnmt1* (23). Further diversity of mammalian CpG MTase has been established in studies of the murine *Dnmt1* gene. Two isoforms of the *dnmt1* enzyme resulting from alternative initiation of translation were observed, the longer one being the well known mouse somatic form that is also expressed in embryonic cells. The short form is detected only in oocytes and preimplantation embryos (24–26). In addition, reverse transcription (RT)-PCR and RNase protection assays have detected low frequencies of occurrence of alternative splicing in the C-terminal region of the rat *Dnmt1* transcript in several tissues, including neuronal cells (11).

We now report the incidental discovery of a second major form of CpG MTase that is abundantly expressed in different human somatic tissues and cell lines. The mRNA coding for this CpG MTase apparently originates from an alternative splicing scheme of the primary *Dnmt1* transcript, which inserts an additional 16 amino acids, encoded by a human Alu family repeat, between exons 4 and 5 of the well characterized *dnmt1* protein. Interestingly, analysis of chimpanzee cells suggests that two abundant forms of somatic CpG MTase are conserved in the higher primates.

MATERIALS AND METHODS

RNA Samples. The acid guanidinium thiocyanate phenol-chloroform extraction method (27) was used to isolate total

Abbreviations: MTase, methyltransferase; RT, reverse transcription; NK, natural killer; kb, kilobase; ³H-SAM, S-adenosyl-L-[methyl-³H]methionine.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF169120–AF169122).

‡D.-W.H. and M.-J.L. contributed equally to this work.

¶To whom reprint requests should be addressed. E-mail: ckshen@ccvax.sinica.edu.tw or cjshen@ucdavis.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

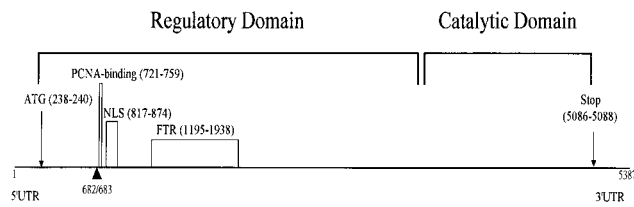


FIG. 1. Structural map of human dnmt1. The structure of the human CpG MTase dnmt1 is shown schematically at the cDNA level. The numbering of nucleotide sequences (1–5,387) follows that used in ref. 8. Besides the initiation and termination codons, the locations of the cDNA regions coding for the PCNA-binding domain, the nuclear localization signal, and the replication foci-targeting sequence are also indicated. The site of insertion of the 48-bp Alu repeat segment through alternative splicing is also marked with a black triangle. It is between nucleotides 682 and 683. UTR, untranslated region.

RNAs from different human cell lines and tissues. These include K562 (erythroid), HeLa (cervical carcinoma), Jurkat (T lymphoblast), Molt 3 (T cells), Raji (Burkitt lymphoma), Wil2 (B lymphocyte), mononuclear cells, peripheral blood T cells, and natural killer (NK) cells. In addition, RNAs from human stomach, liver, placenta, and thymus were kindly provided by T.-J. Chang and C.-K. Chou at the Veterans Hospital (Taipei, Republic of China). Chimpanzee RNA was isolated from blood samples acquired from the Yerkes Primate Center (Atlanta). Poly(A)-RNA was isolated by using the Oligotex mRNA Mini Kit column (Qiagen, Chatsworth, CA).

RT-PCR. RT was performed as described (28). Briefly, RT was performed in a 20- μ l final volume of 50 mM Tris-HCl, pH 8.3/75 mM KCl/3 mM MgCl₂/10 mM DTT/2 μ M dNTP/500 ng of oligo(dT)₁₅/200 units of Superscript II reverse transcriptase (GIBCO/BRL)/20 units RNasin (Promega) and appropriate amounts of total RNA or poly(A)-RNA. The reaction mixtures were incubated at 42°C for 60 min. They were denatured by boiling for 5 min before PCR analysis.

In general, PCR was carried out in a total reaction volume of 100 μ l containing standard PCR buffer (50 mM KCl/20 mM Tris-HCl, pH 8.4), 1.5 mM MgCl₂, 1 μ l of cDNA from the RT reaction, 0.15 μ M each of the two primers, 0.1 mM dNTP, and

2.5 units of *Taq* DNA polymerase (GIBCO/BRL). The extension temperature is usually 72°C, and the duration time for extension is approximately 1 min/kilobase (kb). For example, PCR amplifications of the human Dnmt1 cDNA between nucleotides 502 and 1,001 (Fig. 2A) and the human β -actin cDNA between nucleotides 864 and 1,166 were carried out in a HYBAID OminGene system (Hybaidd, Middlesex, U.K.) with the following temperature profiles: an initial denaturation at 95°C for 5 min; followed by 35 cycles of 95°C for 1 min, 55°C for 30 s, and 72°C for 1 min; and, finally, an elongation step at 72°C for 10 min. Each PCR analysis was done in duplicate. One-fifth of the Dnmt1 product was analyzed by electroporation on a 2% agarose/ethidium bromide gel. The gel patterns were documented with the IS1000 Digital Imaging System and saved in computer-tagged image file format; the band intensities were quantitated with the IS-2000 Documentation and Analysis System (Alpha Innotech, San Leandro, CA).

For semiquantitative analysis (see Figs. 4 and 5), amplification of the β -actin cDNA was used initially to test different PCR conditions. It was found that under the conditions described above, linear responses of the PCR signals were obtained for both the 303-bp β -actin cDNA fragment and the 500-bp Dnmt1 cDNA fragment over a range of serial dilutions of the RT samples used as the template for PCR.

To amplify the 5-kb Dnmt1 cDNA between nucleotides 238 and 5,130, PCR was carried out by the Advantage-GC cDNA PCR Kit (PT3091-1, CLONTECH) in the HYBAID OminGene system with the following temperature profiles: an initial denaturation at 95°C for 5 min; followed by 38 cycles of 95°C for 1 min and 68°C for 8 min; and, finally, a denaturation step at 95°C for 1 min and an elongation step at 68°C for 20 min.

DNA Methylation Assay. The CpG methylation activity of Dnmt1-b was compared with Dnmt1-a by using a transient DNA transfection assay. Two plasmids, pCMV-Dnmt1-a and pCMV-Dnmt1-b, were constructed by cloning the human Dnmt1 cDNA from nucleotide 234 to nucleotide 5,130 with or without the 48-bp insert into the pRc/CMV vector (Invitrogen) downstream of the *Hind*III site. pRc/CMV, pCMV-Dnmt1-a, or pCMV-Dnmt1-b (20 μ g each) were cotransfected with 1 μ g of a reporter construct, pCMV- β gal, into 10⁶ human 293T cells by using the calcium phosphate precipitation

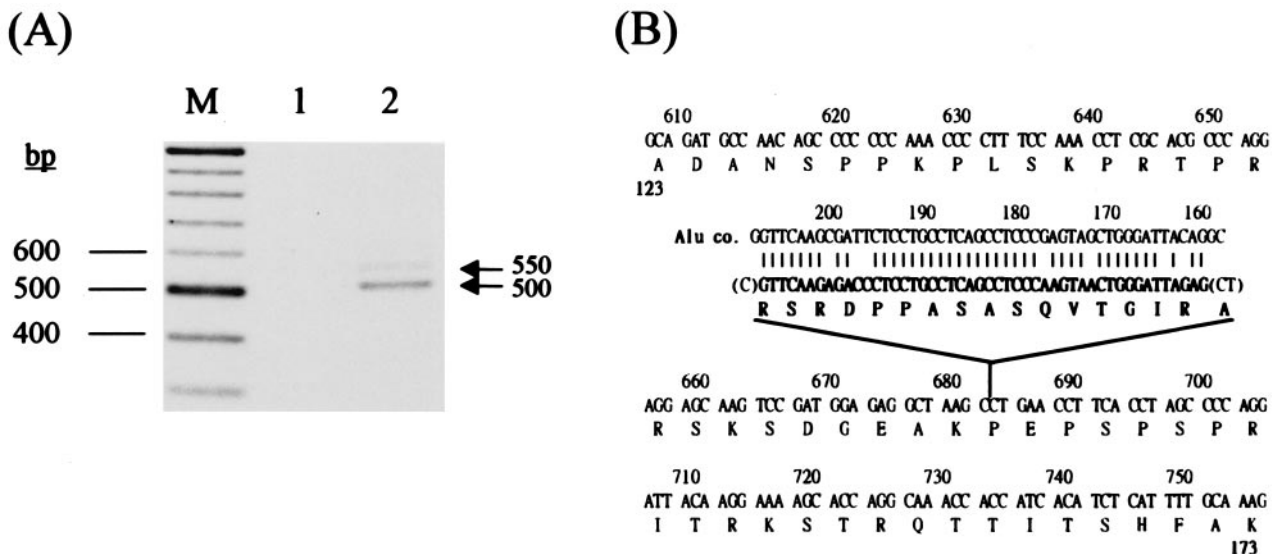


FIG. 2. Identification of alternative splicing of human Dnmt1 transcript. (A) A 2% agarose gel electrophoresis of RT-PCR products of human K562 RNA. RT-PCR was carried out as described in *Materials and Methods* by using the PCR primer pair 502–521 and 1,001–981. Lane M, DNA length marker; lane 1, PCR product without RT reaction; lane 2, RT-PCR product. (B) Nucleotide sequences of the two bands, 550 bp and 500 bp, from lane 2 in A. The sequence of the coding strand of the 500-bp fragment from position 607 to 756 is shown together with the corresponding amino acids. Sequence of the inserted 48-bp segment in the 550-bp fragment is aligned with the homologous region of the antisense strand of the Alu consensus sequence described in ref. 33. Note that the 48-bp insertion results in the substitution of proline (P) at codon 149 by arginine (R), as well as the in-frame insertion of another 16 amino acids, serine (S) through alanine (A).

method. The total cell extract was prepared 2 days after transfection in a buffer similar to the buffers described in refs. 29 and 30. The CpG MTase activities were detected by the incorporation of *S*-adenosyl-L-[methyl-³H] methionine (³H-SAM) into a 60-bp DNA oligomer containing either a hemimethylated or unmethylated CpG dinucleotide residue (29). Each reaction mixture in a total volume of 25 μ l contains 5 μ g of the oligomer substrate, 20 μ g of extract, and 1.25 μ M ³H-SAM (3 μ Ci). The mixtures were incubated at 37°C for 2 h, and stopped with the addition of 0.6% (vol/vol) SDS. After proteinase K digestion and NaOH hydrolysis of RNA, the samples were ice cooled and neutralized with HCl and Tris-HCl. A tenth of each sample was mixed with 20 μ g of salmon sperm DNA, transferred to DE-81 filter (Whatman), and washed as described (30). The filters were air dried and counted.

PCR of Genomic DNA. The isolation of genomic DNA from human K562 cells and chimpanzee blood sample was performed as described by Sambrook *et al.* (31). Genomic DNA (100 ng) was used as the template for amplification in the HYBAID OmniGene system with the following thermal cycles profiles: an initial denaturation at 95°C for 5 min; followed by 30 cycles at 95°C for 1 min, 64°C for 30 s, and 72°C for 3 min; and, finally, 72°C for 10 min. The two primers used are 502–521 and 718–698 as shown in Table 1. One-fifth of the PCR product was analyzed on a 1% agarose gel (Fig. 3A).

Table 1. Nucleotide sequences of DNA primers

Primers	Sequences
Dnmt1	
502–521	5'–GATTGTGCCTTGGAGAACGG–3'
1,001–981	5'–TTGGGTGTTGGTTCTTTGGTT–3'
238–154	5'–GAATTCGAGATGCCGGCGCGTACCGC–3'
5,130–5,101	5'–GGGATTCCTGGTACCAGAAACAGGGG–TGAC–3'
502–521	5'–GATTGTGCCTTGGAGAACGG–3'
718–698	5'–TTTTCTTGTAACTCTGGGGC–3'
238–254	5'–GAATTCGAGATGCCGGCGCGTACCGC–3'
718–698	5'–TTTTCTTGTAACTCTGGGGC–3'
601–620	5'–GAATTCATGGCAGATGCCAACAGCCC–3'
1,001–981	5'–TTGGGTGTTGGTTCTTTGGTT–3'
881–901	5'–GACCGCTTCTGCAGAGAAGAC–3'
1,351–1,331	5'–GTGGGTGCTGCCCATATTTGA–3'
1,231–1,251	5'–CCAACGGAGAAAAAATGGCT–3'
1,701–1,681	5'–CAGAATGTATTGGCAAATGA–3'
1,511–1,531	5'–GCCTCATCGAGAAGAATATCG–3'
1,994–1,974	5'–TTGATCAGTCCCAGCATGCAG–3'
1,931–1,951	5'–ACGAGGCCGGGACAGTGATG–3'
2,401–2,381	5'–TCCCCTGGTGCATTTTTTTGG–3'
2,281–2,301	5'–AGCAAGCAGGCTTGCCAAGAG–3'
2,751–2,731	5'–GGCTTTGTAGATGACTTTCAC–3'
2,631–2,651	5'–CGCTGGGACAGACACAGTCCT–3'
3,101–3,081	5'–GGACTGGACAGCTTGATGTTG–3'
2,981–3,001	5'–AGGACCTGGATAGCCGGTCC–3'
3,451–3,431	5'–CGCACTCGGGCAGGTCTCTCC–3'
3,331–3,351	5'–CCAGCGAGCTACCACGCAGAC–3'
3,801–3,781	5'–TGTGAACACTGTGGAGCCGGG–3'
3,681–3,701	5'–CGGGGGTGTGCGGAGGGATT–3'
4,151–4,131	5'–CCGTACTGACCGGCTGCAGC–3'
4,031–4,051	5'–TGGAAATGTCAGGAACTTTG–3'
4,501–4,481	5'–CCACCAATGCACACTCATGTCCT–3'
4,381–4,401	5'–TCGGCACTGGAGATCTCCTAC–3'
4,851–4,831	5'–GCGGCCCTGTTGCCCATGGG–3'
4,731–4,751	5'–GTGCCTGCCCCACACCGGGAA–3'
5,130–5,101	5'–GGGATTCCTGGTACCAGAAACAGGGG–TGAC–3'
β-Actin	
864–885	5'–CACGAAACTACCTTCAACTCCA–3'
1,166–1,146	5'–GAAGCATTGCGGTGGACGAT–3'

Plasmid, Cloning, and DNA Sequence Analysis. All recombinant DNA works were done according to the standard procedures (31). Specific PCR bands were eluted from agarose gel and cloned into pGEM-T vector (Promega). The cloned DNAs were sequenced by using primers specific for the flanking vector sequences plus different synthetic oligonucleotides. Sequencing was performed with the *Taq* Dideoxy Terminator Cycle Sequencing Kits (Applied Biosystems). The reaction products were electrophoresed and recorded on the ABI PRISM 377-96 DNA Sequencer (Perkin–Elmer). Sequences were analyzed with MACVECTOR sequence analysis software (Oxford Molecular Group, Oxford).

RESULTS

An Isoform of Human Dnmt1 mRNA with an Extra 48-nt Insertion. During the preparation of yeast two-hybrid screening baits by RT-PCR from the total RNA of human K562 cells, it was found that use of the primers 502–521 from exon 4 and primers 1,001–981 from exon 5 (Table 1) of the human *Dnmt1* coding sequence (Fig. 1) gave rise to two bands—instead of one—on the agarose gel (Fig. 2A, lane 2). The length of the lower band, approximately 500 bp, is similar to that expected from the human *Dnmt1* cDNA sequence, and the upper band is approximately 50 bp longer. Omitting the RT step eliminated both bands (Fig. 2A, lane 1), suggesting that they are derived from the RNA but not from contaminating genomic DNA.

Both RT-PCR bands were cloned into plasmid vector and subjected to dideoxynucleotide sequencing. The sequence of the lower band matches perfectly with the published *Dnmt1* cDNA from codon 89 to codon 255. The upper band, on the other hand, contains an extra 48-bp block inserted in-frame between exons 4 and 5 (Fig. 2B). This 48-bp block encodes a 16-amino acid stretch without a termination codon.

The *Dnmt1* mRNA, without counting the poly(A) tail, is about 5,400 nt long (8, 9). In agreement with this length, RT-PCR of K562 RNA with the primers 238–254 and 5,130–5,101 (Table 1) generates a DNA band of an approximate length of 5 kb (Fig. 3A,

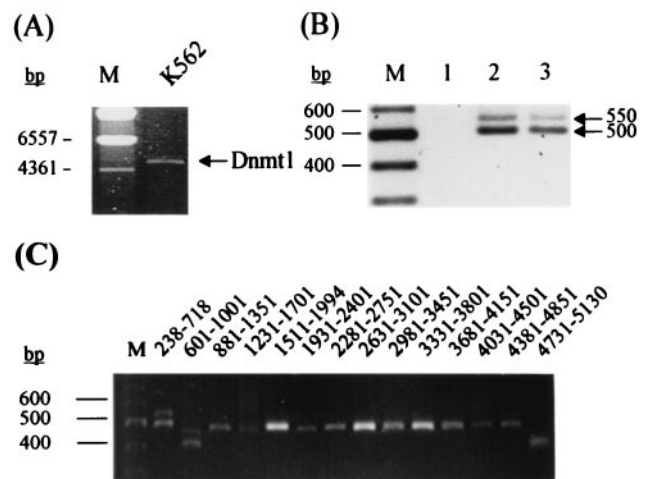


Fig. 3. (A) A 2% agarose gel electrophoresis of RT-PCR products of human K562 RNA with the primer pair 238–254 and 5,130–5,101. (B) The incorporation of the 48-nt segment into a *Dnmt1* transcript of an approximate length of 5,000 nt. Lane 1, PCR of the combined eluate from gel regions immediately above and below the *Dnmt1* band in lane K562 in A; lane 2, PCR of the *Dnmt1* band from lane K562 in A; lane 3, the same RT-PCR sample as in lane 2 of Fig. 2A. (C) RT-PCR “scanning” of *Dnmt1* transcript. PCRs were carried out with different pairs of primers. The numbers above each lane indicate the nucleotide positions of the ends of DNA fragments amplified. Note that the alternative splicing of 48 nt is confirmed further by the appearance of double bands in both the second lane (nucleotides 238–718) and the third lane (nucleotides 601–1,001).

lane K562). Interestingly, PCR of the 5-kb band eluted from the gel shown in Fig. 3A again gave rise to two bands (Fig. 3B, lane 2) of molecular masses similar to those deduced from RT-PCR of the K562 total RNA (Fig. 3B, lane 3). In a negative control experiment, PCR of the eluates from gel regions immediately above and below the 5-kb band shown in Fig. 3A revealed no bands on the gel (Fig. 3B, lane 1). The data shown in Figs. 2 and 3 together suggest that at least two different species exist in the apparently 5,000-nt-long Dnmt1 mRNA population previously detected by Northern blot analysis (9). One of them is the well known Dnmt1, which already had been characterized extensively by cloning and sequencing (8, 9). The other one contains the extra 16-amino acid block (Fig. 2B) inserted between exons 4 and 5.

To determine whether there are other isoforms of Dnmt1 mRNA with additional insertions or deletions, we have carried out RT-PCR of K562 RNA with other sets of primers (Table 1) spanning the human *Dnmt1* cDNA region. As shown in Fig. 3C, all of these reactions, except for the one that used the primer set already described in Figs. 2A and 3B, reveal only one band with lengths the same as expected from the Dnmt1 cDNA sequence. Hereafter, we refer to the originally described human CpG MTase and its isoform with the extra 16 amino acids as Dnmt1-a and Dnmt1-b, respectively.

Tissue Specificity and Relative Levels of Dnmt1-b Expression.

To examine the tissue specificity of expression of Dnmt1-b RNA, we have carried out RT-PCR of total RNA samples prepared from a variety of human cell lines and tissues. These include the cell lines K562, HeLa, Jurkat, Molt3, Raji, and Wil2, as well as different tissues including the liver, stomach, placenta, thymus, mononuclear cells, peripheral blood T cells, and NK cells. It turns out that Dnmt1-a and Dnmt1-b coexist as two abundant mRNA species in all of the cell types we have examined (data not shown). Furthermore, as exemplified by semiquantitative RT-PCR analysis of three RNA samples in Fig. 4, the expression levels of Dnmt1-b range from 70% to 40% of those of Dnmt1-a.

Dnmt1-b Is Enriched in Poly(A)-RNA and It Encodes CpG MTase Activity. Three lines of evidence suggest that Dnmt1-b also encodes a CpG MTase. First, similar to Dnmt1-a, it exists mainly as a poly(A)-RNA species (Fig. 5). Second, coupled *in vitro* transcription/translation has shown that Dnmt1-b RNA could be translated efficiently (data not shown). Finally, extract prepared from 293T cells transiently transfected with Dnmt1-b expression plasmid (pCMV-Dnmt1-b) had MTase activity *in vitro* as did that from pCMV-Dnmt1-a-transfected cells. Also, the activities from both Dnmt1-a and Dnmt1-b seem to prefer hemimethylated CpG as substrate (data not shown).

Dnmt1-b Results from Alternative Splicing of a 48-nt Alu Repeat Sequence Between Exons 4 and 5 of Dnmt1-a. The simplest explanation for the generation of Dnmt1-b is that an alternative splicing scheme inserts an 48-nt exon sequence from intron 4 of the *Dnmt1* gene between exons 4 and 5 of Dnmt1-a. To test this hypothesis, we used PCR to amplify the human genomic region containing the 3' portion of exons 4, the intron 4, and the 5' portion of exon 5 of Dnmt1-a as described in *Materials and Methods*. The resulting fragment, which is approximately 3 kb long, was cloned and subjected to automated DNA sequencing analysis.

As shown in Fig. 6A, intron 4 of the human *Dnmt1* gene contains at least seven copies of human Alu family repeats, Alu1 through Alu7, all of which are pointing in the opposite direction of Dnmt1 transcription. Because of the repetitive nature of the Alu family sequences and the distances of some of these repeats from the sequencing primers, we have been able to determine neither the sequence of a small portion of the sequence of Alu4 nor that of approximately 300 bp of DNA downstream of Alu2. Among the seven Alu repeats, Alu1 and Alu2 are arranged in tandem. Alu3 is apparently the product of a retroposition event, with the target site being the sequence 5'-TGGAA-3'. Alu4, Alu5, Alu6, and Alu7 together form a tetrameric tandem array flanked at the boundaries by the direct repeats 5'-ACTTCAT(C/

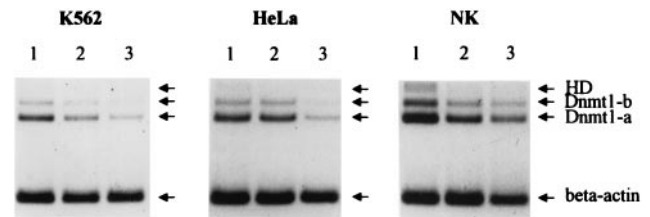


FIG. 4. Relative expression of Dnmt1-a and Dnmt1-b in total RNAs from K562, HeLa, and NK cells. The analysis was performed as described in detail in *Materials and Methods*. HD indicates the band resulting from heteroduplex formation between Dnmt1-b and Dnmt1-a during the PCR cycles (data not shown). The RT-PCR samples loaded were derived from 10 ng (lanes 1), 5 ng (lanes 2), and 2.5 ng (lanes 3) of total RNA.

T)TTT-3' (Fig. 6A). Thus, the array is most likely generated by sequential insertion of four Alu repeats into the same genomic site, similar to those identified in the primate α -globin loci (32). However, without comparison to other primate species, the time of occurrence of these four retroposition events during evolution could not be determined. It should be noted that the sequence of Alu5 corresponds only to the first half of the 290-bp monomeric Alu consensus (Fig. 6B). The origin of this unusually short Alu repeat is also unknown.

It is interesting that the only sequenced region of intron 4 identical to the 48-nt block of Dnmt1-b mentioned above is located in the antisense strand of Alu1, between position 208 and 161 relative to the Alu consensus sequence (33). Furthermore, detailed comparison of the sequences immediately flanking both sides of this segment of Alu1 indicates that they would provide functional acceptor and donor sites (34), respectively, for splicing of the 48-nt segment between exons 4 and 5 of Dnmt1 (Fig. 6B).

Dnmt1-a and Dnmt1-b Also Exist in Chimpanzee. To determine whether Dnmt1-a and Dnmt1-b coexist as two abundant RNA species in higher primates, we have used the same primers described for Fig. 2A to carry out RT-PCR of chimpanzee blood RNA. Two fragments of 500 bp and 550 bp also were generated (Fig. 7). Nucleotide sequencing of the two bands indicates that they are identical to the corresponding regions of human *dnmt1-a* and *dnmt1-b*, respectively (data not shown).

DISCUSSION

As summarized in Fig. 8, we have discovered that, *in vivo*, the primary transcript of the human *Dnmt1* gene (Fig. 8A) is alternatively spliced in two—and most likely only two—different ways. One of the spliced mRNA species (Fig. 8B Lower) would be translated to generate *dnmt1*, the well known enzyme thought to be the only major form of CpG MTase in vertebrate somatic tissues. The newly identified mRNA species, on the other hand, includes an additional 48 nt in-frame between exons 4 and 5 of the Dnmt1 mRNA (Fig. 8B Upper). The expression level of the newly identified mRNA species is approximately 40–70% of that of the Dnmt1 mRNA (Fig. 4). Previous approaches of isolation and characterization of the human CpG MTase would not detect the existence of this enzyme species (8, 9). Thus, it seems that there

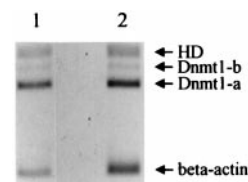


FIG. 5. Enrichment of both Dnmt1-a and Dnmt1-b in poly(A)-RNA. RT-PCR product from 0.25 ng of K562 poly(A)-RNA (lane 2) was loaded, along with, for purposes of comparison, that from 2 ng of K562 total RNA (lane 1).

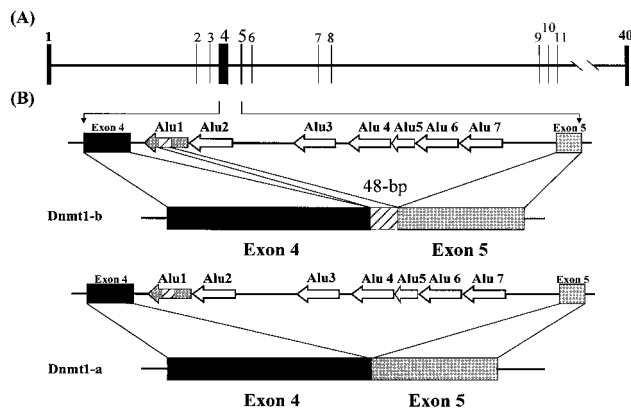


FIG. 8. Summary of the alternative splicing schemes of human Dnmt1 transcript for the generation of Dnmt1-a (*B Lower*) and Dnmt1-b (*B Upper*). The genomic map of the *Dnmt1* gene is shown in *A* with exons 1–11 and 40 indicated.

sequence similarity could be found among these spliced Alu segments. As expected, the 48-nt Alu sequence of Dnmt1-b mRNA also has no similarity to those documented cases with respect to both its location within the Alu repeat consensus and its nucleotide sequence (comparative data not shown).

Now that Dnmt1-b has been identified, the functional roles of the human Dnmt1 and possibly those of other primate genes need to be redefined. Methylation patterns are, in general, stably maintained in the somatic cells of animals but change dramatically during early development when the genomes of the mammalian embryos undergo consecutive waves of demethylation and *de novo* methylation (36). In particular, most CpG methylation present in the zygote DNA is erased during early embryo cleavage, most likely by DNA demethylase(s) (37). After implantation, the embryo undergoes a wave of global *de novo* methylation that restores the genomic methylation levels of the gastrulating embryo to those seen in the adult. It is unknown whether *dnmt1* participates in the global *de novo* methylation processes. However, at least one other CpG MTase enzyme may be involved, because mouse embryonic stem cells homozygous for a null mutation of the *Dnmt1* gene still have a detectable level of ³mCpG as well as *de novo* methylation activity toward proviral DNA (38).

The *dnmt1* CpG MTases encoded by human and murine *Dnmt1* genes have been well characterized (39–42). The enzyme is capable of methylating both unmethylated and hemimethylated DNA substrates but with a great preference toward the latter *in vitro*. It was believed that *dnmt1* is the single major enzyme establishing and propagating the methylation patterns through cell generations of the mammalian tissues and that the N-terminal, regulatory domain of *dnmt1* modulates the *de novo* and maintenance methylation activities of the enzyme (39). The current discovery of the abundantly expressed *dnmt1-b* raises interesting questions regarding many previous cellular and genetic phenomena associated with somatic CpG methylation, all of which have been thought to be largely responsible for *dnmt1-a*.

We are grateful to many people who have kindly provided us with different samples: Mr. J.-M. Lee and Dr. N.-S. Liao for human peripheral blood T cells and NK cells, Dr. L.-I. Lin for normal human blood samples, Drs. T.-J. Chang and C.-K. Chou for human tissue RNAs, and Dr. N. Gavva for the chimpanzee RT cDNA. This research was supported by the Academia Sinica and by the National Science Council, Taipei, Taiwan, Republic of China.

- Jost, J. P. & Saluz, H. P., eds. (1993) *DNA Methylation: Molecular Biology and Biological Significance* (Birkhauser, Basel).
- Bestor, T. H. & Verdine, G. L. (1994) *Curr. Opin. Cell Biol.* **6**, 386–389.

- Bestor, T. H. (1998) *Nature (London)* **393**, 311–312.
- Jones, P. A. & Buckley, J. D. (1990) *Adv. Cancer Res.* **54**, 1–23.
- Baylin, S. B. (1997) *Science* **277**, 1948–1949.
- Laird, P. W. & Jaenisch, R. (1994) *Hum. Mol. Genet.* **3**, 1487–1495.
- Bestor, T. H., Laudano, A., Mattaliano, R. & Ingram, V. (1988) *J. Mol. Biol.* **203**, 971–983.
- Yoder, J. A., Yen, R.-W., Vertino, P. M., Bestor, T. H. & Baylin, S. B. (1996) *J. Biol. Chem.* **271**, 31092–31097.
- Yen, R.-W., Vertino, P. M., Nelkin, B. D., Yu, J. J., El-Deiry, W., Cumaraswamy, A., Lennon, G. G., Trask, B. J., Celano, P. & Baylin, S. B. (1992) *Nucleic Acids Res.* **20**, 2287–2291.
- Ramchandani, S., Bigey, P. & Szyf, M. (1998) *Biol. Chem.* **379**, 535–540.
- Deng, J. & Szyf, M. (1998) *J. Biol. Chem.* **273**, 22869–22872.
- Tajima, S., Tsuda, H., Wakabayashi, N., Asano, A., Mizuno, S. & Nishimori, K. (1995) *J. Biochem.* **117**, 1050–1057.
- Kimura, H., Ishihara, G. & Tajima, S. (1996) *J. Biochem.* **120**, 1182–1189.
- Posfai, J., Bhagwat, A. S., Posgai, G. & Roberts, R. J. (1989) *Nucleic Acids Res.* **17**, 2424–2435.
- Lenonhardt, H., Page, A. W., Weier, H.-U. & Bestor, T. H. (1992) *Cell* **71**, 865–874.
- Blow, J. (1993) *Nature (London)* **361**, 684–685.
- Liu, Y., Oakeley, E. J., Sun, L. & Jost, J.-P. (1998) *Nucleic Acids Res.* **26**, 1038–1045.
- Chuang, L. S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G. & Li, B. F. L. (1997) *Science* **277**, 1996–2000.
- Li, E., Bestor, T. H. & Jaenisch, R. (1992) *Cell* **69**, 915–926.
- Van den Wyngaert, I., Sprengel, J., Kass, S. U. & Luyten, W. H. M. L. (1998) *FEBS Lett.* **426**, 283–289.
- Yoder, J. A. & Bestor, T. H. (1998) *Hum. Mol. Genet.* **7**, 279–284.
- Okano, M., Xie, S. & Li, E. (1998) *Nucleic Acids Res.* **26**, 2536–2540.
- Okano, M., Xie, S. & Li, E. (1998) *Nat. Genet.* **19**, 219–220.
- Carlson, L. L., Page, A. W. & Bestor, T. H. (1992) *Genes Dev.* **6**, 2536–2541.
- Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M. & Bestor, T. H. (1998) *Development (Cambridge, U.K.)* **125**, 889–897.
- Gaudet, F., Talbot, D., Leonhard, H. & Jaenisch, R. (1998) *J. Biol. Chem.* **273**, 32725–32729.
- Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
- Foley, K. P., Leonard, M. W. & Engel, J. D. (1993) *Trends Genet.* **9**, 380–385.
- Tollefsbol, T. O. & Hutchison, C. A. (1995) *J. Biol. Chem.* **270**, 18543–18550.
- Vertino, P. M., Yen, R.-W. C., Gao, J. & Baylin, S. B. (1996) *Mol. Cell. Biol.* **16**, 4555–4565.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Bailey, A. D. & Shen, C.-K. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7205–7209.
- Jurka, J. & Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4775–4778.
- Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
- Makalowski, W., Mitchell, G. A. & Labuda, D. (1994) *Trends Genet.* **10**, 188–193.
- Razin, A. & Shemer, R. (1995) *Hum. Mol. Genet.* **4**, 1751–1755.
- Bhattacharya, S. K., Ramchandani, S., Cervoni, N. & Szyf, M. (1999) *Nature (London)* **397**, 579–583.
- Lei, H., Oh, S. P., Okano, M., Jüttermann, R., Goss, K. A., Jaenisch, R. & Li, E. (1996) *Development (Cambridge, U.K.)* **122**, 3195–3205.
- Bestor, T. H. (1992) *EMBO J.* **11**, 2611–2517.
- Tollefsbol, T. O. & Hutchison, C. A., III (1995) *J. Biol. Chem.* **270**, 18543–18550.
- Pradhan, S., Talbot, D., Sha, M., Benner, J., Hornstra, L., Li, E., Jaenisch, R. & Roberts, R. J. (1997) *Nucleic Acids Res.* **25**, 4666–4673.
- Tucker, K. L., Talbot, D., Lee, M. A., Leonhardt, H. & Jaenisch, R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12920–12925.