One-codon alternative splicing of the CpG MTase Dnmt1 transcript in mouse somatic cells

Meng-Jau Lin[a], Tai-Lin Lee[a], Duen-Wei Hsu[a], C.-K. James Shen[b,*]

[a] *Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 115, Taiwan*
[b] *Institute of Genetics, National Yang-Ming University, Shih-Pai, Taipei 112, Taiwan*

Received 19 October 1999; received in revised form 25 January 2000

Edited by Takashi Gojobori

Abstract The genomic methylation patterns in the mammalian somatic cells are presumably maintained by a single enzyme, dnmt1. In mouse, this DNA (cytosine-5)-methyltransferase, or CpG MTase, is encoded by the *Dnmt1* gene. We now present evidence that in different tissues and cell types, the primary transcript of mouse dnmt1 is alternatively spliced to generate two poly-(A) RNAs of approximately similar abundance. This alternative splicing most likely originates from the existence of two tandemly arranged acceptor sites separated by only 3 nt. The two Dnmt1 mRNAs thus encode two CpG MTases differing by two amino acids. We discuss the implications of the discovery of two dnmt1 isozymes, instead of one enzyme as previously thought, in the somatic cells of both mouse and human.

Key words: Genomic methylation pattern; *Dnmt1*; CpG MTase

1. Introduction

C-methylation at CpG dinucleotide residues in the vertebrate genomes is closely associated with a number of biological processes. These processes include the regulation of transcription, X chromosome inactivation, genomic imprinting, cellular defense against viral agents and tumorigenesis (see [1–4] and references there in). The methylation patterns change dramatically during early development when the genomes of the mammalian embryos undergo consecutive waves of demethylation and de novo methylation [5,6]. In the adult somatic cells, approximately 60% of their CpG residues are methylated and the covalent addition of a methyl group to the 5-position of cytosine is predominantly a post-replicative reaction. More recent experiments have suggested that associations between CpG methylation and the various processes mentioned above are in part modulated through the recruitment of histone deacetylases by a class of methyl-CpG-binding proteins [7,8].

Thus far, the best characterized enzyme responsible for the methylation of the CpG residues is the somatic form of mouse CpG MTase. This enzyme, dnmt1 (Fig. 1), is 1617 amino acids long, and it is composed of a C-terminal catalytic domain closely related to the bacterial type II C-methylating enzymes. The N-terminal region of mouse dnmt1, which is absent in the bacteria1 enzymes, contains the nuclear localization signal [9], the replication foci-targeting sequences [9,10] and a cystine-rich zinc-binding region [11,12]. Both in vitro and in vivo assays have indicated that dnmt1 prefers hemimethylated CpG residues as its reaction substrate [11,13–15]. Dnmt1 is encoded by an approximately 5-kb long gene consisting of 40 exons [16]. Mouse embryos homozygous for deletion of the *Dnmt1* gene die at early stages of development [17]. In germ line cells, different forms of Dnmt1 mRNAs are generated through alternative splicing of sex-specific 5′ exons [18].

As in other vertebrates, dnmt1 has been believed to be the single major enzyme responsible for the stable maintenance of the DNA methylation patterns in mouse somatic cells [19]. We report below that in different mouse somatic tissues, the primary Dnmt1 transcript is alternatively spliced to generate two similarly abundant Dnmt1 mRNAs encoding proteins differing by only two amino acids.

2. Materials and methods

2.1. RNA isolation and RT-PCR analysis

Total RNAs from various adult mouse tissues and cell lines were isolated as described [20]. The Oligotex mRNA Mini Kit column (QIAGEN) was used to isolate poly-(A) RNA. cDNA synthesis was performed with the use of 0.5–5 μg of total RNA or 0.5 μg of poly-(A) RNA and SuperScript II Reverse Transcriptase (Gibco/BRL) in a total reaction volume of 20 μl. After incubation for 1 h at 37°C, the reactions were heated at 70°C for 15 min and then diluted to 100 μl.

Each PCR mixture contained 1 μl cDNA from the RT reaction, 0.5 μM each of the forward and reverse primers, 0.2 mM each of the four deoxyribonucleotides, 1.5 mM $MgCl_2$ and 2.5 units of *Taq* DNA polymerase (Gibco/BRL) in a total volume of 100 μl. The conditions for amplifications were: 95°C for 3 min; followed by 30 cycles of 94°C for 30 s, annealing at a primer-specific temperature for 45 s and extension at 72°C for a duration of 1 min per 1 kb of DNA to be synthesized. Finally, an additional 10 min incubation at 72°C was allowed for elongation. To amplify the 5-kb long Dnmt1 cDNA, PCR was carried out with the use of the Advantage 2 PCR Enzyme System (Clontech). The nucleotide sequences of the primers used are listed in Table 1.

2.2. Genomic amplification of mouse Dnmt1 intron 4

Genomic DNA was isolated from the mouse tails using the Nucleo-Spin-Tissue Kit from Clontech. 100 ng of the genomic DNA were then subjected to PCR with the use of the primer set 446–467/588–568. After incubation of the reaction mixture(s) at 95°C for 3 min, PCR amplification was performed with 30 cycles of 94°C for 45 s, 60°C for 30 s and 72°C for 2 min. Finally, elongation was allowed at 72°C for 10 min.

2.3. Plasmid cloning and DNA sequence analysis

The individual PCR bands were isolated from the agarose gels with an Agarose Gel DNA Extraction Kit (Boehringer Mannheim). They were subcloned into the vector pGEM-T (Promega). The cloned DNAs were sequenced with primers specific for the vector sequence flanking the inserts and analyzed as described by Hsu et al. [21].

*Corresponding author. Fax: (88)-62-27884177.
E-mail: ckshen@ccvax.sinica.edu.tw or cjshen@ucdavis.edu

PII: S0014-5793(00)01254-0

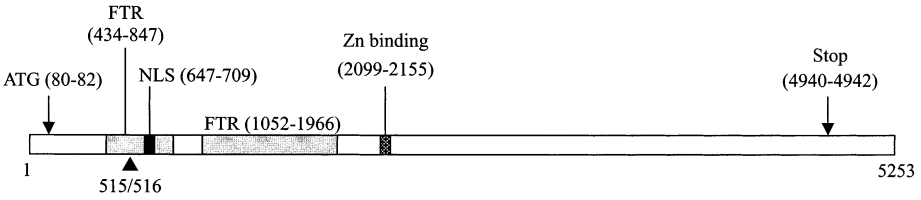

Fig. 1. Structural map of mouse dnmt1. The structure of the mouse CpG MTase, dnmt1, is shown at the cDNA level. The numbering of the nucleotides follows that of Yoder et al. [14]. Besides the initiation and termination codons, the regions coding for the nuclear localization signal (NLS), the replication foci-targeting sequence (FTR) and the cystine-rich Zn-binding region (Zn binding) are indicated. The site of the 3-bp insertion through alternative splicing is also marked with the black triangle. It is between nucleotides 515 and 516 of the original cDNA sequence reported for mouse Dnmt1 [14], now renamed as Dnmt1-a.

3. Results

3.1. A mouse Dnmt1 RNA with an out-of-frame 3-nt insertion

Several sets of DNA primers were used to amplify different regions of the mouse Dnmt1 cDNA. Two bands were observed after PCR of the mixture containing primer set 78–96/941–923 (lanes 1, 2 and 4–9, Fig. 2A). Amplification with other sets of primers all result in only one PCR band (Fig. 2B), the lengths of which are as expected from the published mouse Dnmt1 cDNA (GenBank accession No. X14805). The PCR doublet bands seen in Fig. 2A were further characterized by polyacrylamide gel electrophoresis, DNA sequence analysis after cloning into pGEM-T and by re-amplification with primer set 78–96/941–923.

Nucleotide sequence analysis of either PCR band of Fig. 2A indicated that they both contained the two types of sequences shown in Fig. 2C. One of the sequences is identical to that reported in the literature for mouse Dnmt1 [14,22]. The other category, however, contains an extra CAG triplet inserted between nucleotides 515 and 516. The 3-nt insertion appears to be out-of-frame and it changes the amino acid phenylalanine (F) at codon 146 of Dnmt1 to serine (S)–valine (V) (Fig. 2C). Finally, PCR reamplification of either PCR band of Fig. 2A generated the same doublet pattern (data not shown).

The above data together suggest that there exist two Dnmt1 RNA species, which are with or without the 3-nt insertion (Fig. 2C). During the denaturation–renaturation cycles of PCR, two different homoduplexes were generated from the two RNA species and they co-migrated as the lower band seen in Fig. 2A. The heteroduplexes between the two sequences were also formed during RT-PCR, which migrated slower than the homoduplexes.

Table 1
Primers for RT-PCR analysis of mouse Dnmt1 transcripts

Primer	Sequence
78–96	5′-CCA TGG AGA TGC CAG CGC GAA CAG-3′
79–99	5′-GGT ACC GAT GCC AGC GCG AAC AGC TCC-3′
941–923	5′-GTC GAC CTC TGG GCT GGC TCC TGG G-3′
446–467	5′-AAT AGA TCC CCA AGA TCC AGG C-3′
588–568	5′-TTC GTG AAG TGA GCC GTG ATG-3′
822–844	5′-CAT TGA GGC GGA AAT CAA AGG AG-3′
1670–1648	5′-CTT CAT ATA CAG CAT CAG GAT TG-3′
1639–1659	5′-CCT GCA AAA CAA TCC TGA TGC-3′
2391–2370	5′-TCA TCT GGA ATG ACC GAG ACG C-3′
2341–2362	5′-CGA TGA GGA GAT GCT AGA GGT G-3′
3196–3175	5′-GTG ATA GGA TCC GTT GTA GGA C-3′
3142–3163	5′-CTA CAA GTT CTA CAG GCC TGA-3′
4008–3989	5′-GCC ACG CCA TAC TGT CCA GC-3′
3954–3973	5′-GCA TGG GCT ACC AGT GCA CC-3′
4942–4925	5′-CTA GTC CTT GGT AGC AGC-3′
4942–4922	5′-GAT ATC CTA GTC CTT GGT AGC AGC CTC-3′

3.2. The mouse Dnmt1 RNA with the out-of-frame insertion is nearly 5000 nt long and poly-adenylated

The coding region of the mouse Dnmt1 mRNA is 5253 nt long. Indeed, RT-PCR of MEL cell RNA with the primer set 79–99 and 4942–4922 generated an approximate 5-kb PCR

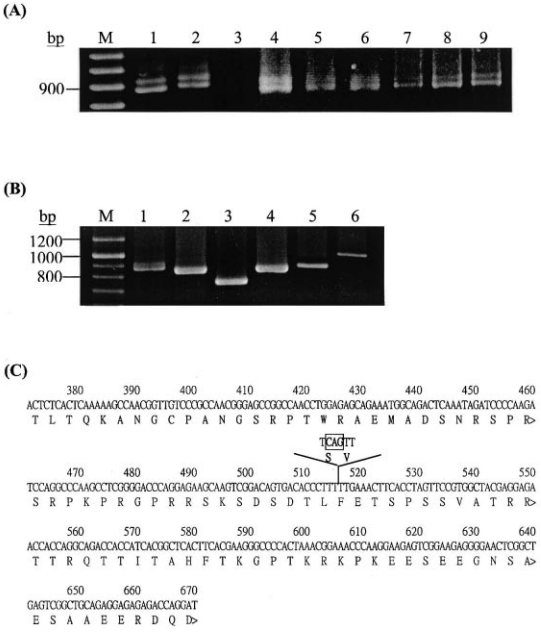

Fig. 2. Identification of alternative splicing of the mouse Dnmt1 transcript. A: 2% agarose gel electrophoresis of PCR products of mouse cDNA. PCR was carried out with the PCR primer pair 78–96 and 941–923. Lane M, 100-bp DNA ladders as the length marker; lane 1, ES cell; lane 2, MEL cell; lane 3, the negative control, without RT; lane 4, brain cDNA; lane 5, liver cDNA; lane 6, heart cDNA; lane 7, lung cDNA; lane 8, spleen cDNA; lane 9, NIH3T3 cell cDNA. B; RT-PCR scanning of mouse Dnmt1 transcript. The reactions were carried out as described in (A), using MEL cell RNA as the template. The primer sets used are: lane 1, 78–96/941–923; lane 2, 822–844/1670–1648; lane 3, 1639–1659/2391–2370; lane 4, 2341–2362/3196–3175; lane 5, 3142–3163/4008–3989; lane 6, 3954–3973/4942–4925. C: Nucleotide sequences of the two PCR bands from lane 2 of (A). The nucleotide sequence of the coding strand from position 371 to 670 of the lower PCR band is shown together with the corresponding amino acids. The boxed sequence is the 3-base insertion found in the sequence of the upper PCR band in (A). It is shown together with the disrupted 3-T codon from 515–517. The 3-base insertion thus results in the substitution of the phenylalanine (F) at codon 146 by serine (S) plus valine (V).

band (lane 1, Fig. 3A). This 5-kb band, as well as the gel eluates from immediately above and below it, were used as the templates, respectively, for PCR using the primer set 446–467/588–568. As shown in Fig. 3B, two bands were again generated, but only with use of the 5-kb band as the template (lane 1, Fig. 3B).

Whether the Dnmt1 RNA species with the 3-nt insertion is polyadenylated has been analyzed. The significantly higher intensities of the doublet by using reverse transcripts (RT) from poly(A)-RNA as the PCR templates (lane 2, Fig. 3C), in comparison to that from the total RNA (lane 1, Fig. 3C), indicated that similar to the previously characterized Dnmt1 RNA, the species with the out-of-frame 3-nt insertion is also polyadenylated. We now term the two mRNAs as Dnmt1-a and Dnmt1-b, respectively.

3.3. Generation of Dnmt1-a and Dnmt1-b is most likely due to alternative splicing

The genomic structure of the human *Dnmt1* gene has been determined previously [16]. However, only a partial sequence of mouse intron 4 has been reported, which includes 67 bp near the 5′ end of the intron [14,23]. Interestingly, the site of the 3-nt insertion is right in between the reported exons 4 and 5 of Dnmt1. To further understand the relationship between Dnmt1-a and Dnmt1-b, we have isolated intron 4 of mouse Dnmt1 by PCR amplification of mouse genomic DNA with the primer set 446–467/588–568. The nucleotide sequence of this approximately 1-kb PCR product is depicted in Fig. 4. As shown, there are a total of 17 CAG triplets in the intron. Careful sequence examination indicated that only the CAG triplet immediately upstream of exon 5 (boxed, Fig. 4) is flanked at its 5′ side with the splicing acceptor dinucleotide AG. Furthermore, none of the other 16 CAG triplets is followed at their 3′ side with the splicing donor dinucleotide sequence GT. We thus suggest that Dnmt1-b results from splicing between exon 4 and this downstream CAG triplet, instead of between exon 4 and the previously reported exon 5 as in Dnmt1-a.

4. Discussion

In summary, we have provided evidence that the primary transcript of the well-known mouse *Dnmt1* gene undergoes efficient alternative splicing to generate two mRNA species differing by one amino acid substitution plus one amino acid insertion. This in turn suggests that there exist two, not just one. major CpG MTases, dnmt1-a and dnmt1-b, in the mouse somatic cells. It is not surprising that the previous studies have missed the detection of Dnmt1-b mRNA or the dnmt1-b protein. First, the messages or proteins of the two isoforms are of relatively large molecular weights and they would not be separable on the routinely used gel systems. Second, although peptide mapping has been carried out on purified mouse dnmt1 before, only its C-terminal region was analyzed [24]. Third, antibodies directed against either of the two isozymes would have interacted with both of them, in biochemical experiments as well as in immunocytological studies. Finally, the inserts of mouse dnmt1 cDNA clones isolated previously by different laboratories might not all have been completely sequenced.

The generation of dnmt1-a and dnmt1-b apparently results from alternative splicing reactions of the Dnmt1 primary transcript, which unite two acceptor sites tandemly arranged at the 3′ end of intron 4 (Fig. 4). The upstream acceptor site may be utilized more efficiently since it is closer to the polypyrimidine tract. However, other nucleotide sequences around the two acceptor sites could also affect the relative efficiencies of splicing [28]. The data of Figs. 2 and 3 indicate that when enough PCR products were allowed to denature and renature sufficient times, the heteroduplexes and homoduplexes formed were of similar abundance. This in turn suggests that the amounts of Dnmt1-a and Dnmt1-b are of the same order. Analysis of ^{32}P-end-labeled PCR products has confirmed this point (data not shown). It should be noted here that several other studies have also detected the use of a 3-bp exon in alternative splicing reactions [25–27]. In particular, a

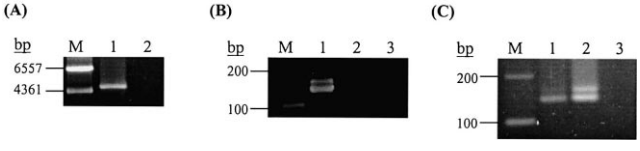

Fig. 3. A: 1% agarose gel electrophoresis of nearly full-length Dnmt1 cDNA. PCR was carried out using primer set 79–99/4942–4922 and MEL RNA, with (lane 1) or without (lane 2) prior reverse transcription. M, DNA length marker. B: Incorporation of the CAG triplet within the Dnmt1 transcript of an approximate length of 5000 nt. PCRs were carried out with the primer set 446–467/588–568. The templates used are eluates from gel regions of lane 1 in (A) containing the 5-kb cDNA (lane 1), immediately above the 5-kb band (lane 2) and below the cDNA (lane 3), respectively. C: Enrichment of both Dnmt1-a and dnmt1-b in the poly-(A) RNA fraction. RT-PCR products from 5 ng ES cell total RNA (lane 1), 5 ng ES cell poly-(A) RNA (lane 2) and from 5 ng non-poly-(A) RNA (lane 3) were separately loaded. The PCR primers used were 446–467 and 588–568.

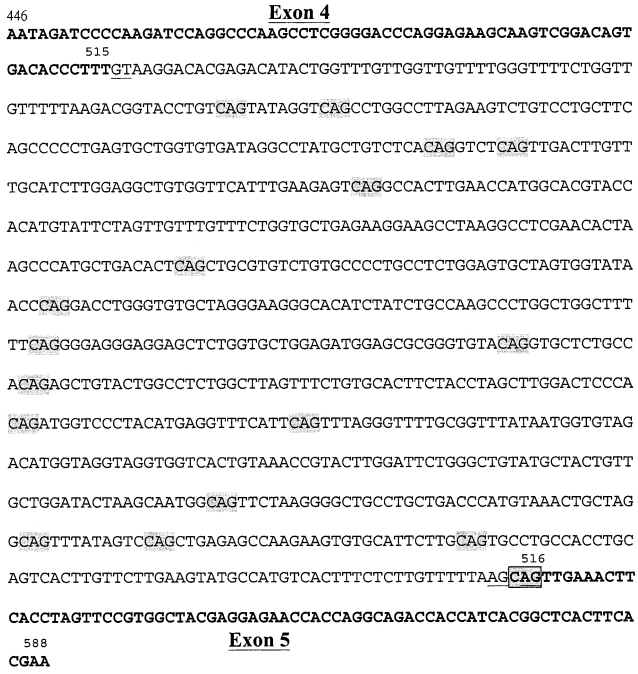

Fig. 4. Nucleotide sequence of intron 4 of the mouse Dnmt1 gene. The cDNA sequences of the flanking exons 4 and 5 are in bold letters. The splicing donor site (GT) and acceptor site (AG) are underlined. The 17 CAG sequences within intron 4 are shaded. The putative CAG triplet alternatively spliced into Dnmt1-b is boxed.

similar arrangement of tandem acceptor sites has previously been found for the *S. pombe* telomerase reverse transcriptase (GenBank accession No. AF015783; reference [26]).

Interestingly, mouse is not the only vertebrate species in which the Dnmt1 transcript is alternatively spliced to generate two similarly abundant mRNAs. Recently, we have discovered that two Dnmt1 mRNA species also exist in the somatic cells of higher primates including human. In those cases, the alternative splicing event has inserted out-of frame a 48-nt segment from an Alu family within the previously reported intron 4 in between exon 4 and exon 5 [21]. The longer variant RNA is polyadenylated and could be translated in vitro. Thus, it is expected to code for a CpG MTase differing from the dnmt1 by 17 amino acids [21].

The conservation, from mouse to the higher primates, of an alternative splicing scheme generating two major dnmt1 isoforms is intriguing. In mouse, the two differ by two amino acids; in human, they differ by 17 amino acids. Yet, the relative locations within the proteins of the extra/different amino acid blocks of dnmt1-b are identical for mouse, human and chimpanzee. This suggests that evolution of Dnmt1 and its encoded proteins are under a common functional constrain maintaining the existence of two similarly abundant dnmt1 isoforms in the mammalian somatic cells.

The differential functions of the two isoforms could only be speculated at this time. The dnmt1-a enzyme has been suggested to function as the major somatic maintenance enzyme responsible for the stabilization of the CpG methylation patterns of the mammalian genomes [11,19]. The in vitro methylation assay [19] and in vivo association with the replication foci [9] support the above assigned function of dnmt1 through post-replicative processes of methylation. However, it has also been postulated that dnmt1 may be responsible for certain de novo methylation activities in the somatic cells [11]. Furthermore, the patterns of methylation at genomic regions with sparse CpG residue and those with high densities of CpG, such as the centromeres and the inactive X chromosome, might be maintained by different enzymes. The dnmt1 protein has also been suggested to be involved in the regulation of DNA replication and repair (for references, see [29]). Different isoforms could thus be utilized separately in these different programs. A thorough understanding of the possible differential functions of dnmt1-a and dnmt1-b, in both human and mouse, awaits further molecular and genetic investigations.

Acknowledgements: We thank our laboratory colleagues, Dr. Woan-Yuh Tarn and Dr. Tom Cech for their helpful scientific communications. Dr. T. Bestor has generously provided some materials for other relevant studies of the mouse dnmt1. This research was supported by the Academia Sinica, the National Science Council, the National Health Research Institute and the Foundation of Biomedical Sciences, Taipei, Taiwan.

References

[1] Baylin, S.B. (1997) Science 277, 1948–1949.
[2] Kass, S.N., Pruss, D. and Wolffe, A.P. (1997) Trends Genet. 13, 444–449.
[3] Schmutte, C. and Jones, P.A. (1998) Biol. Chem. 379, 377–388.
[4] Walsh, C.P. and Bestor, T.H. (1999) Genes Dev. 13, 26–34.
[5] Razin, A. and Shemer, R. (1995) Hum. Mol. Genet. 4, 1751–1755.
[6] Jaenisch, R. (1997) Trends Genet. 13, 323–329.
[7] Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Nat. Genet. 19, 187–191.
[8] Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Nature 393, 386–389.
[9] Lenonhardt, H., Page, A.W., Weier, H.-U. and Bestor, T.H. (1992) Cell 71, 865–874.
[10] Liu, Y., Oakeley, E.J., Sun, L. and Jost, J.P. (1998) Nucleic Acids Res. 26, 1038–1045.
[11] Bestor, T.H. (1992) EMBO J. 11, 2611–2617.
[12] Chuang, L.S.-H., Ng, H.-H., Chia, J.-N. and Li, B.F.L. (1996) J. Mol. Biol. 257, 935–948.
[13] Bestor, T.H. and Ingram, V.M. (1983) Proc. Natl. Acad. Sci. USA 80, 5559–5563.
[14] Yoder, J.A., Yen, R.-W., Vertino, P.M., Bestor, T.H. and Baylin, S.B. (1996) J. Biol. Chem. 271, 31092–31097.
[15] Prafhan, S., Talbot, D., Benner, J., Hornstra, L., Li, E., Jaenisch, R. and Roberts, R.J. (1997) Nucleic Acids Res. 25, 4666–4673.
[16] Ramchandani, S., Bigey, P. and Szyf, M. (1998) Biol. Chem. 379, 535–540.
[17] Li, E., Bestor, T.H. and Jaenisch, R. (1992) Cell 69, 915–926.
[18] Mertineit, C., Yoder, J.A., Taketo, T., Laird, D.W., Trasler, J.M. and Bestor, T.H. (1998) Development 125, 889–897.
[19] Yoder, J.A., Soman, N.S., Verdine, G.L. and Bestor, T.H. (1997) J. Mol. Biol. 270, 385–395.
[20] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[21] Hsu, D.-W., Lin, M.-J., Lin, T.-L., Wen, S.-C., Chen, X. and Shen, C.-K.J. (1999) Proc. Natl. Acad. Sci. USA 96, 9751–9756.
[22] Tucker, K.L., Talbot, D., Lee, M.A. and Leonhardt, H. (1996) Proc. Natl. Acad. Sci. USA 88, 9578–9582.
[23] Rouleau, J., Tanigawa, G. and Szyf, M. (1992) J. Biol. Chem. 267, 7368–7377.
[24] Glickman, J.F., Pavlovich, J.G. and Reich, N.O. (1997) J. Biol. Chem. 272, 17851–17857.
[25] Santoni, M.J., Barthels, D., Vopper, G., Boned, A., Goridis, C. and Wille, W. (1989) EMBO J. 8, 385–392.
[26] Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B. and Cech, T.R. (1997) Science 277, 955–959.
[27] Sampson, M.J., Ross, L., Decker, W.K. and Craigen, W.J. (1998) J. Biol. Chem. 273, 30482–30486.
[28] Berget, S.M. (1995) J. Biol. Chem. 270, 2411–2414.
[29] Robertson, K.D. and Jones, P.A. (1997) Am. J. Hum. Genet. 61, 1220–1224.