Cloning and analysis of a novel human putative DNA methyltransferase

Ilse Van den Wyngaert[a], Jörg Sprengel[b], Stefan U. Kass[a,*], Walter H.M.L. Luyten[a]

[a] *Department of Experimental Molecular Biology, Janssen Research Foundation, Turnhoutseweg 30, 2340 Beerse, Belgium*
[b] *Department of Applied Molecular Biology, Janssen Research Foundation, Turnhoutseweg 30, 2340 Beerse, Belgium*

Received 12 February 1998; revised version received 18 March 1998

Abstract DNA methylation is intricately involved in a variety of cellular processes, such as differentiation, cell cycle progression, X-chromosome inactivation and genomic imprinting. However, little is known about how specific DNA methylation patterns are established and maintained. Previously one mammalian DNA methyltransferase has been described, but there has been considerable speculation about the presence of a second activity capable of methylation. Here we report the identification and characterization of a novel human putative DNA methyltransferase. Using a bioinformatics screen we have identified several expressed sequence tags which show high sequence similarity to the *Schizosaccharomyces pombe* gene pmt1$^+$. The cDNA for PuMet (for *put*ative DNA *met*hyltransferase) was cloned and the predicted amino acid sequence deduced. The gene is ubiquitously expressed, albeit at low levels. Like several other DNA methyltransferases, the bacterially overexpressed protein is not active in methylation assays.

Key words: DNA methylation; Transcriptional repression; Cancer

1. Introduction

Methylation at position 5 of the cytosine ring in the dinucleotide CpG is the major modification of DNA in higher eukaryotes. DNA methylation correlates with silencing of gene expression [1,2] and it was proposed as a mechanism to reduce transcriptional noise [3] and to suppress the activation of parasitic DNA sequences [4]. Methylation patterns are inherited from one cell generation to the next by the postreplicative action of the DNA methyltransferase (DNMT) [5]. DNMT recognizes hemimethylated DNA and adds a methyl group to the newly synthesized daughter strand, thus maintaining the methylation pattern of the parental DNA [6]. However, at least in early development, methylation patterns are much more dynamic. After fertilization, the DNA in the blastocyst undergoes almost complete demethylation, which is followed by a wave of global de novo methylation [7]. During the subsequent implantation of the embryo and differentiation of the somatic cells, genes receive their tissue-specific methylation pattern [8]. Thus, subsequent phases of demethylation, de novo methylation and maintenance methylation are required to establish the methylation pattern of DNA in somatic cells.

So far, the only known mammalian DNA methyltransferase is DNMT [9,10]. DNMT has a long N-terminal domain which targets the enzyme to the nucleus and to replication foci [11]. The C-terminus contains the catalytic domain, which is clearly homologous to bacterial DNA methyltransferases. The in vitro preference of the enzyme for hemimethylated DNA is consistent with its role in maintenance methylation [5]. Embryonic stem (ES) cells carrying a null mutation of the *dnmt* gene contain no maintenance methylation activity, but retain the ability to methylate integrated provirus DNA de novo, suggesting the presence of at least one additional DNA methyltransferase in mammalian cells [12,13].

This report describes the identification and analysis of a novel human putative DNA methyltransferase. Using a bioinformatics screen several overlapping expressed sequence tags (ESTs) were identified which show high similarity to all bacterial and eukaryotic DNA methyltransferases. Interestingly, the highest similarity is to the PMT1 protein, a 'pseudo DNA methyltransferase' identified in *Schizosaccharomyces pombe* [14,15]. The putative methyltransferase (PuMet) is expressed at very low levels in a variety of human tissues. The PuMet gene is located on chromosome 10p13–22. However, bacterially overexpressed protein does not show any methyltransferase activity in vitro.

2. Materials and methods

2.1. Bioinformatics strategy for the identification of DNA methyltransferase ESTs from databases

A multiple sequence alignment was generated using the protein sequences from human, mouse, and *Arabidopsis* DNA methyltransferases together with the sequence of the prokaryotic CpG methyltransferase from *Spiroplasma* sp. The SwissProt database identifiers are MTDM_HUMAN, MTDM_MOUSE, MTDM_ARATH and MTSI_SPISQ, respectively. Based on the multiple sequence alignment a similarity profile was calculated and applied to subsequent Profilesearches [16] in the SwissProt release 35 protein database [17]. Positive hits were used to optimize and refine the multiple alignment before searching in DNA databases. The similarity profile contained amino acids 1020–1220 with respect to human DNMT (see Fig. 1C). The EMBL DNA database release 52 [18] was probed with the proframesearch program. Searches were performed on a dedicated hardware accelerator (Bioccelerator, Compugen, Israel). All human ESTs obtained by the similarity search were subjected to reverse screening of the protein database with Framesearch and BlastX similarity search programs.

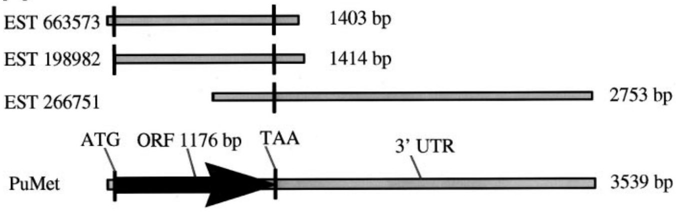
2.2. Cloning and overexpression of PuMet

Clones containing ESTs encoding the putative DNA methyltransferase were obtained from the Image consortium (clones 663573, 198982 and 266751). DNA sequencing was carried out on double-stranded plasmid DNA with dye-terminator chemistry as described by the manufacturer (Perkin Elmer/Applied Biosystems) and the products were resolved on an ABI Prism 377 Automated Sequencer. The EMBL/GenBank accession number for the PuMet sequence is AJ223333. Two primers (forward: 5′-AGG AGC TCG GCG CGG GGA TGG AGC CC-3′; reverse: 5′-GGC CAT GGC AGA ATT ACT CTC TGA AAA TG-3′), containing a *Sac*I and a *Nco*I restriction site respectively, were designed based on the EST 663573 sequence, which contains the full PuMet coding region. PCR reactions were performed using the EST 663573 plasmid and the resulting 1261 bp fragment was cloned into the expression vector pRSET (Invitro-

*Corresponding author. Fax: (32) (14) 60 6111.
E-mail: skass@janbe.jnj.com

PII S0014-5793(98)00362-7

A



B

```
                                            TGCCTGGCCGTCCGGATCGAGGCCCCGACGCCGCCGACTCCGCGCCCC
       M   E   P   L   R   V   L   E   L   Y   S   G   V   G   G   M   H   H   A   L   R   E
  49  ATG GAG CCC CTG CGG GTG CTG GAG CTA TAC AGC GGC GTG GGC GGC ATG CAC CAC GCG CTG AGA GAA
       S   C   I   P   A   Q   V   V   A   A   I   D   V   N   T   V   A   N   E   V   Y   K
 115  AGC TGT ATA CCT GCA CAA GTG GTG GCT GCC ATT GAT GTC AAC ACT GTC GCT AAT GAA GTA TAC AAG
       Y   N   F   P   H   T   Q   L   L   A   K   T   I   E   G   I   T   L   E   E   F   D
 181  TAT AAT TTT CCT CAC ACA CAG TTA CTT GCC AAG ACG ATT GAA GGC ATT ACA CTC GAA GAG TTT GAC
       R   L   S   F   D   M   I   L   M   S   P   P   C   Q   P   F   T   R   I   G   R   Q
 247  AGA TTA TCT TTT GAT ATG ATT TTA ATG AGC CCT CCC TGC CAG CCA TTC ACA AGG ATT GGC CGG CAG
       G   D   M   T   D   S   R   T   N   S   F   L   H   I   L   D   I   L   P   R   L   Q
 313  GGT GAT ATG ACT GAT TCA AGG ACG AAT AGC TTC TTA CAT ATT CTA GAT ATT CTC CCA AGA TTA CAA
       K   L   P   K   Y   I   L   L   E   N   V   K   G   F   E   V   S   S   T   R   D   L
 379  AAA TTA CCA AAG TAT ATT CTT TTG GAA AAT GTT AAA GGT TTT GAA GTA TCT TCT ACA AGA GAC CTC
       L   I   Q   T   I   E   N   C   G   F   Q   Y   Q   E   F   L   L   S   P   T   S   L
 445  TTG ATA CAA ACA ATA GAA AAT TGT GGC TTT CAG TAC CAA GAG TTT CTA TTA TCT CCA ACC TCT CTT
       G   I   P   N   S   R   L   R   Y   F   L   I   A   K   L   Q   S   E   P   L   P   F
 511  GGC ATT CCA AAT TCA AGG CTA CGA TAT TTT CTT ATT GCA AAG CTT CAG TCA GAG CCA TTA CCC TTT
       Q   A   P   G   Q   V   L   M   E   F   P   K   I   E   S   V   H   P   Q   K   Y   A
 577  CAA GCC CCT GGT CAG GTA CTG ATG GAG TTC CCC AAA ATT GAA TCT GTA CAT CCA CAA AAA TAT GCA
       M   D   V   E   N   K   I   Q   E   K   N   V   E   P   N   I   S   F   D   G   S   I
 643  ATG GAT GTA GAA AAT AAA ATT CAA GAA AAG AAC GTT GAA CCA AAT ATT AGC TTT GAT GGC AGC ATA
       Q   C   S   G   K   D   A   I   L   F   K   L   E   T   A   E   E   I   H   R   K   N
 709  CAG TGT TCT GGA AAA GAT GCC ATT CTT TTT AAG CTT GAA ACT GCA GAA GAA ATT CAC AGG AAA AAT
       Q   Q   D   S   D   L   S   V   K   M   L   K   D   F   L   E   D   D   T   D   V   N
 775  CAA CAA GAT AGT GAT CTC TCT GTG AAA ATG CTA AAA GAT TTT CTT GAA GAT GAC ACT GAC GTG AAC
       Q   Y   L   L   P   P   K   S   L   L   R   Y   A   L   L   L   D   I   V   Q   P   T
 841  CAG TAT CTT TTA CCA CCA AAG TCA TTG CTG CGA TAT GCT CTT CTG TTA GAC ATT GTT CAG CCC ACT
       C   R   R   S   V   C   F   T   K   G   Y   G   S   Y   I   E   G   T   G   S   V   L
 907  TGT AGA AGG TCC GTG TGC TTT ACC AAA GGA TAT GGA AGC TAC ATA GAA GGG ACA GGG TCT GTG TTA
       Q   T   A   E   D   V   Q   V   E   N   I   Y   K   S   L   T   N   L   S   Q   E   E
 973  CAG ACT GCA GAG GAT GTG CAG GTT GAG AAT ATC TAC AAA TCC CTT ACC AAT TTG TCA CAA GAA GAA
       Q   I   T   K   L   L   I   L   K   L   R   Y   F   T   P   K   E   I   A   N   L   L
1039  CAG ATA ACA AAG CTG TTA ATA CTT AAA CTG CGA TAT TTC ACT CCT AAA GAA ATA GCA AAT CTC CTT
       G   F   P   P   E   F   G   F   P   E   K   I   T   V   K   Q   R   Y   R   L   L   G
1105  GGA TTT CCT CCA GAG TTC GGA TTT CCT GAG AAG ATA ACA GTG AAA CAG CGT TAT CGC CTA CTT GGA
       N   S   L   N   V   H   V   V   A   K   L   I   K   I   L   Y   E   *
1171  AAT AGT CTC AAC GTG CAT GTA GTA GCT AAA CTA ATC AAA ATC TTA TAT GAA TAA TTTTGAAATAACTCT
1240  GAAAGATGGTCATATGATATTCCTTCATTTTCAGAGAGTAATTCTGAAATTCTGTTTTGAACTAATTCTGGTGAAATTTAACTAAAT
1327  TATTTTAATCTGTCCTTATTAAGAAATTTGGATTTTATTAAAAAAATCCATGTGTTTCATCAAATTTATATTACTGTATTTTATAAA
1414  ATACGAACTATTGTTATGCTTTATGGAGAGTATATTTTCATATGAAACTGATAAATATATATGTGGAATATTCTTTTTAAAAACAAT
1501  TTTATAAACAAATCAAGGAGCACTAGAGCTTTAGTGTCTACATGAGTATCTTGTGAAGTGTCATAACAGGCATAACATATTTTTATG
1688  TTAATTAAAGAAGTTTTTTATTTCTAATTTTAAAATAATGATTCTAATTCTAGCCCATTGAATGTTTTACTTAATTTCCAATCCTGT
1675  AAATATTTCTTTATTTCATCTAAACAAATGCATATGATATGTAAACCAAATTATCATCATTCAGTTTATACAGCAGACTGAAGAATA
1762  TGTCAGGAAAAATAATCCTAAATGTTTGATTACCTCTCCATGAAGCGGCCACAGAAACTGTTGGTGCACATTATCTACATTTTCCTC
1849  AGCCTCTATCTGAAGGACAAAACTCATTTTTCCTTCTTCATTTTGTATGTACTTTCCTTTAGTTGACTATTATCAAAGCTTGTTTTT
1936  GTCTTCTCAGCTTCTGAGTTGAATTTGCCTTTGCACGCTGAGGAATGTAGCATGACGTTAAGATTTCAGGGAGGAAGCTGCCCTCTG
2023  ATCCCATGTCATATGACAGTACTCCCCTCCCCCCAGTTGTTCTTTATTGTAGCAACATACACTCTCAATGAACTTTTATGACTTTTA
2110  ACTTTCTCTGTCTCTCTTCCTCAGCAGACCTCTTGGAATTAGGGTTTCTAATCAAACTAAGTCTTGCATCATAAATTCTATGGTTTG
2197  TGATAAAAGCAAGTGACAAAGAGAAGAGAGATGAAAAGAAACAAGTCCAGAAATCATATTACAAGATTTTAATTCAAATTTTTCTTC
2284  TTAAAAATCAGCTATTTAGAGGTGGATAAATTGGAAGTTTGTGCATTGCTTGTGGGAACGTAAAATGGTTCAGCTGCTGTTGGAAAA
2371  CGCTATGGAGGTTCGTCACAAAATTCAACATAGAATTACCCTATGATTCAGCGATTCCACCTCTGGCTATAGATCCAAAAGAATTGA
2458  AAGCAGGGACTTGAACAGATATTTGTACACCAGTGTTTATAACACCATTATTTACAGTAGCCAAAAGGTAGAAGCAGCCCAAGTGTC
2545  ATCCCATTCACCCATGAATGGATGCAGGAATTGTGTTCCGTGCATGCAGTGGAATAGTACCAGCCTTCCGCAGGAAGGAAATTCTGA
2632  CATATGCGGTGACATGGATGAACCTTGAGTACATTATGCGAAGTGAAATAAGCCACCCTGTGGGGATGGAGAATGGGTAGTTAGTGC
2719  ATAATGGATACAAAGTTTCTGATTGGGGTGATGAGAAAGTTCTGGAGGTGGATGGGGAGGATGGTTGCACAACAGTGTGAGTGTACT
2806  TAATGCCACTGCAACTGTATATTTAAAAATAGTTAAAATGATAAATTTTACATTATGTATATTTTACTGCAATTGAAAAAATCATAC
2893  TCTATTACCAATTTTTCAGAACTCATAGATAAATATGAATGTAGGGAAGCTGTAAGCCAGCCCATATACAGAAACTATCACTATCTA
2980  CATTTTTCATATGTGTTCAGACAGAGCCTATTACAAGTCTCATAAATTTCAGTAAACAATCACAAATTATTTTGGCACACTTTCATG
3067  TAGGTTCTGTTCACTGCTGAAATATTCTAAATAAAATGGTTTAGAGTTTTACTGTAACTATGAATAATTATGCTATAATTATAGTAT
3154  CCCACATTCGGTTGAGCAACATTATACAAAATAAGTGTCGAGTTACTGTGCACCAACTATAAATACTTGAAAGGTCTGTAGAGACAT
3241  TTGCTTACAAACTGGTAGACAGAATGTAGCTTAGTGGAATTCTTAAATTATGGGAATGTATTCACAGGAAGACGGTGATTCTTGCCC
3328  CTAATTCATAATTGAAAATATGCAGTAACAAGCATAAAACTCACATCTTAATGAAGATGTATCTACCAGTGTCATTCAATTAGCAGA
3415  TGTATATGTATGCATAATTAATTTGGAAAATGTTATTTTATGGAATGTCTGCTATATGCTTTTGGTAATAATGTTGTATATAAAATC
3502  TGCACTGGAAATTCTGTTTCCTGATAAAAAAAAAAAAA
```

Fig. 1. A: Alignment of EST clones identified in a bioinformatics screen for novel human DNA methyltransferases. Clone numbers and clone lengths in bp are indicated. The position of the open reading frame (ORF) of PuMet is indicated by the dark arrow. B: cDNA sequence and deduced amino acid sequence of PuMet. GenBank accession number is AJ223333. C: Alignment of the amino acid sequence of PuMet, pmt1^{+} (accession number X82444) and the human DNA methyltransferase (DNMT, accession number X63692). Amino acid sequences were aligned using the Clustal W multiple sequence algorithm and conserved amino acid residues were annotated using the GeneDoc program. The 10 conserved motifs found in 5-methylcytosine methyltransferases are indicated according to Wilkinson et al. [14].

gen). Overexpression was in *Escherichia coli* BL21(DE3) (Novagen). 100 ml of 2×YT medium containing 50 mg/ml ampicillin was inoculated with 1 ml overnight culture and incubated at 37°C with vigorous shaking. Optimal induction for maximum expression of soluble protein was found at OD_{600}=0.5. Overexpression was induced by addition of IPTG to 1 mM and cells were further incubated at 30°C for 2 h. Cells were collected by centrifugation (4000×*g*, 15 min, 4°C) and the pellet resuspended in lysis buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 10 mM imidazole) at 2–5 ml per gram wet weight. Lysozyme was added to 1 mg/ml and after a 30 min incubation on ice the suspension was sonicated using ten 10 s bursts with a 10 s cooling period between bursts. The lysate was cleared by centrifugation (10 000×*g*, 30 min, 4°C) and the protein purified from the supernatant. 1 ml of 50% (w/v) Ni-NTA silica (Qiagen) was added to 4 ml of the supernatant. The slurry was mixed gently at 4°C for 60 min and loaded onto a column. The column was washed two times with 4 ml wash buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 20 mM imidazole) and protein eluted with 4×0.5 ml elution buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 250 mM imidazole). All fractions were collected and analyzed by SDS-PAGE. Pooled fractions of PuMet protein were dialyzed overnight against 2 l PBS. Control lysates from bacteria grown with an insert-less pRSET vector were prepared in the same way.

2.3. Methylation assay

The activity of the purified PuMet protein was measured using *S*-adenosyl-L-[methyl-^{3}H]methionine (^{3}H-SAM) as methyl group donor. DNA substrates were either plasmid DNA or poly d[d(I-C) d(I-C)] (Sigma, 13 units/mg). 5 μg DNA was methylated in a 50 μl reaction in NEBuffer 2 (New England Biolabs) containing 64 μM ^{3}H-SAM (Amersham, 79 Ci/mmol) and 13.75 μg of the purified PuMet protein for 2 h at 37°C. The reaction was stopped by adding an equal volume of stop mix containing 2% SDS, 10 mM EDTA and 1 mg/ml proteinase K followed by a 30 min incubation at 55°C. To each sample 40 μl 1.05 M NaOH was added and the samples were incubated at 60°C for 10 min. Samples were transferred to a 96 well GF/C Millipore filtration plate. 20 μl salmon sperm DNA (Promega, 1 mg/ml) was added and the DNA precipitated with 80 μl 30% trichloroacetic acid (TCA), 15 mM sodium pyrophosphate. Tritiated SAM was removed by washing the wells 10 times with 200 μl 5% TCA, 5 mM sodium pyrophosphate and twice with ethanol. After drying, 50 μl MicroscintO (Packard) was added and samples were counted in a Packard microplate scintillation counter. The negative control contained equivalent amounts of bovine serum albumin (BSA) instead of purified protein, the positive control contained DNA methyltransferase *Sss*I (New England Biolabs). In addition, extracts from bacteria grown with a vector containing no PuMet insert were assayed in the same way as described above.

2.4. Expression analysis of PuMet

Two human multiple tissue Northern (MTN) blots, a cancer cell line and a human endocrine system multiple tissue Northern blot (all from Clontech and containing 2 μg poly(A)$^+$ RNA per lane) were probed with a 1093 bp *Nsp*I fragment corresponding to the PuMet ORF, labeled with ^{32}P-dCTP by random priming using the *redi*prime DNA labeling system (Amersham). The blots were washed at high stringency (2×20 min with 0.1×SSC, 0.1% SDS, 65°C) and exposed at −70°C to Kodak Bio Max autoradiography film with intensifying screens. The blots were then rehybridized without stripping as described above with a random primed 3890 bp *Apa*I fragment from the human DNMT cDNA. Blots were stripped with boiling SDS solution (0.5%) for 10 min and rehybridized with a β-actin probe (Clontech) as described.

Variations in the subsequent hybridization of the Northern blots were kept at a minimum in order to allow an approximate comparison of expression signals. The same amount of probe (30 ng) was used in the labeling reactions which resulted in closely matched specific activities for the three different probes (5×10^8 dpm/μg). Hybridization and washing conditions were kept identical for all three probes. The Northern blots were exposed for 16 h to a phosphoimager screen and hybridization signals were quantified using a Storm phosphoimaging system (Molecular Dynamics).

2.5. Fluorescence in situ hybridization studies (FISH)

Chromosomal localization was carried out by SeeDNA Inc. (Ontario, Canada). Lymphocytes isolated from human blood were cul-

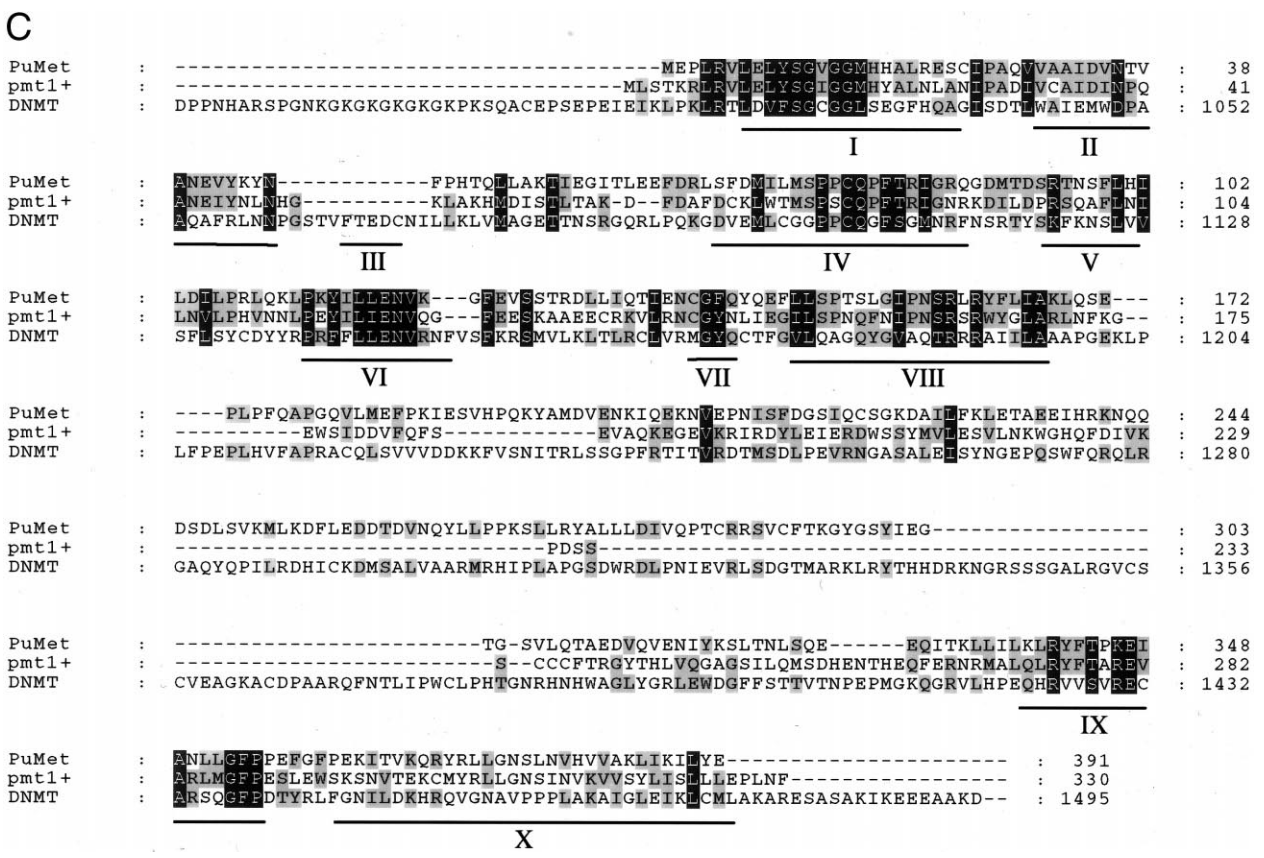


Fig. 1. (*continued*)

tured in α-minimal essential medium (α-MEM) supplemented with 10% fetal calf serum and phytohemagglutinin at 37°C for 68–72 h. The lymphocyte cultures were treated with bromodeoxyuridine (0.18 mg/ml) to synchronize the cell population. The synchronized cells were washed three times with serum-free medium to release the bromodeoxyuridine block and recultured at 37°C for 6 h in α-MEM with thymidine (2.5 mg/ml). Cells were harvested and slides were prepared by using standard procedures including hypotonic treatment, fixation and air-drying.

A PuMet cDNA fragment was gel-purified and biotinylated with dATP using the BRL BioNick labeling kit (15°C, 1 h) [19]. The procedure for FISH detection was performed as previously described [19,20]. Briefly, slides were baked at 55°C for 1 h. After RNase treatment, the slides were denatured in 70% formamide in 2×SSC for 2 min at 70°C followed by dehydration with ethanol. Probes were denatured at 75°C for 5 min in a hybridization mix consisting of 50% formamide and 10% dextran sulfate. Probes were loaded on the denatured chromosomal slides. After overnight hybridization, slides were washed and the signal detected. FISH signals and the 4′,6-diamidino-2-phenylindole (DAPI) banding pattern were recorded separately by taking photographs, and the assignment of the FISH mapping data to chromosomal bands was achieved by superimposing FISH signals with DAPI-banded chromosomes [21].

3. Results and discussion

3.1. Cloning and sequence analysis of PuMet

Using the mouse, human, *Arabidopsis* and *Sss*I DNA methyltransferase protein sequences, a 200 amino acid profile was created which contained the active site and motifs I–VIII (Fig. 1C) of the known DNA methyltransferases. This profile was then used as a probe to screen human EST databases. Several hits were identified, most of which were ESTs from the known DNA methyltransferase gene, DNMT [9,10]. However, some of the identified sequences were distinctly different from the DNMT gene, but instead were more similar to the *pmt1*$^+$ gene from *S. pombe* [14]. The *pmt1*$^+$ gene from *S. pombe* encodes a protein clearly homologous to prokaryotic and eukaryotic DNA methyltransferases. However, this protein does not show any significant methyltransferase activity in vitro, which is consistent with the lack of any detectable 5-methylcytosine in *S. pombe* DNA [14].

The identified ESTs formed one contig encoding a protein which we named PuMet (for putative methyltransferase). Fig. 1A shows a map of the position of the EST clones and Fig. 1B shows the cDNA sequence with the deduced amino acid sequence. The cDNA contains an open reading frame of 391 amino acids and has a long 3′ untranslated region of 2.3 kb. Several 5′ RACE reactions did not yield clones which extend the sequence beyond that of EST 663573, suggesting that clone 663573 contains the full 5′ untranslated region of the PuMet cDNA. Comparison of the predicted amino acid sequence of PuMet with both prokaryotic and eukaryotic 5-methylcytosine methyltransferases (5mC-MTases) reveals a high degree of similarity (Fig. 1C). The predicted molecular mass of 45 kDa is within the range found for known bacterial 5mC-MTases. This is in contrast to the known mammalian 5mC-MTases which have an additional N-terminal domain of approximately 1000 amino acids long, which is involved in nuclear localization and targeting of the enzyme to replication foci [11] and is also a target for the cell cycle machinery [22]. The C-terminal domain of the mammalian enzymes contains the catalytic domain, and is similar to the bacterial enzymes. Ten conserved motifs have previously been identified in this domain, which are also found in the amino acid sequence of PuMet, with strongest conservation in motifs I, IV, VI and IX (see Fig. 1C). The active site, a PPCQ tetrapeptide in motif IV, is perfectly conserved between the human DNMT and PuMet. However, in the *S. pombe* pmt1$^+$ sequence a PSCQ motif is present and it was concluded that this may cause catalytic inactivity of the pmt1$^+$ protein [14]. Subsequently, Pinarbasi et al. [15] deleted the serine residue in the pmt1$^+$ protein, which restored the Pro-Cys motif, and as a consequence also restored methyltransferase activity of the protein.

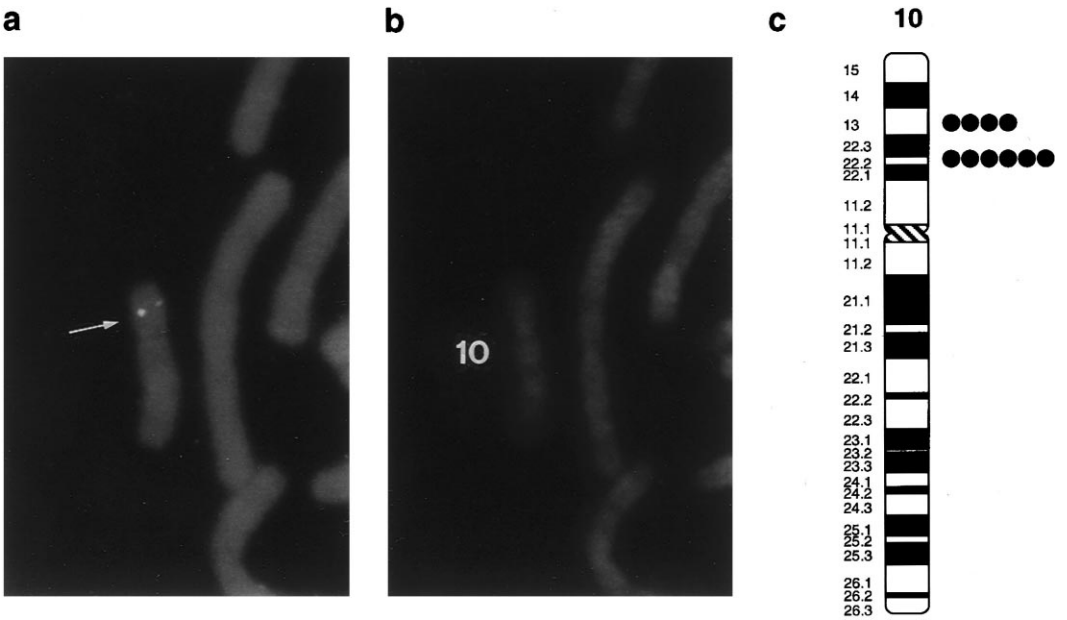


Fig. 2. Chromosomal localization of the PuMet gene determined by FISH analysis. A: Photograph of metaphase chromosome spread labeled with fluorescent PuMet cDNA probe. B: Photograph of the same mitotic figure stained with DAPI to identify chromosomes and reveal chromosome band patterns. Comparison with A shows the FISH signal is located on chromosome 10. C: Idiogram of chromosome 10. The signal was further localized to position p13–22 by determination of the distribution of signals from 10 independent photographs.

3.2. Chromosomal localization of the PuMet gene

The chromosomal position of the PuMet gene was determined to establish whether loss of heterozygosity (LOH) associated with PuMet might be linked with any known disease. The insert of EST clone 663573 was used as a probe for FISH mapping. Under the conditions used, the hybridization efficiency was approximately 71% for the probe (among 100 checked mitotic figures, 71 showed signals on one pair of chromosomes). The DAPI banding pattern was used to establish that the PuMet gene localizes to the short arm of chromosome 10. The detailed position was further determined based upon the analysis of 10 photographs leading to the conclusion that the PuMet gene is located on human chromosome 10, region p13–22 (Fig. 2). However, queries of the Genome Data Base (Johns Hopkins University) for this region did not result in any significant markers which may be relevant to a certain disease.

3.3. Overexpression of PuMet

The coding region of the PuMet gene was cloned into the bacterial expression vector pRSET-A which resulted in the addition of an N-terminal 6×His tag to the recombinant protein. Previously, this has not led to an impairment of DNA methylation activity for prokaryotic and eukaryotic DNA methyltransferases [23]. Subsequently the overexpressed protein was purified by Ni-NTA chromatography (Fig. 3A). Purified fractions were used in methylase assays according to Adams et al. [24]. As shown in Fig. 3B, the purified protein did not show any methyltransferase activity compared to the positive control methyltransferase *Sss*I, using poly(dIdC) and plasmid DNA (Fig. 3B) or lambda and calf thymus DNA (data not shown) as DNA substrates. In addition, the PuMet cDNA was cloned into the mammalian expression vector pcDNA3 (Pharmacia). However, COS cells did not show elevated levels of DNA methyltransferase activity upon transient

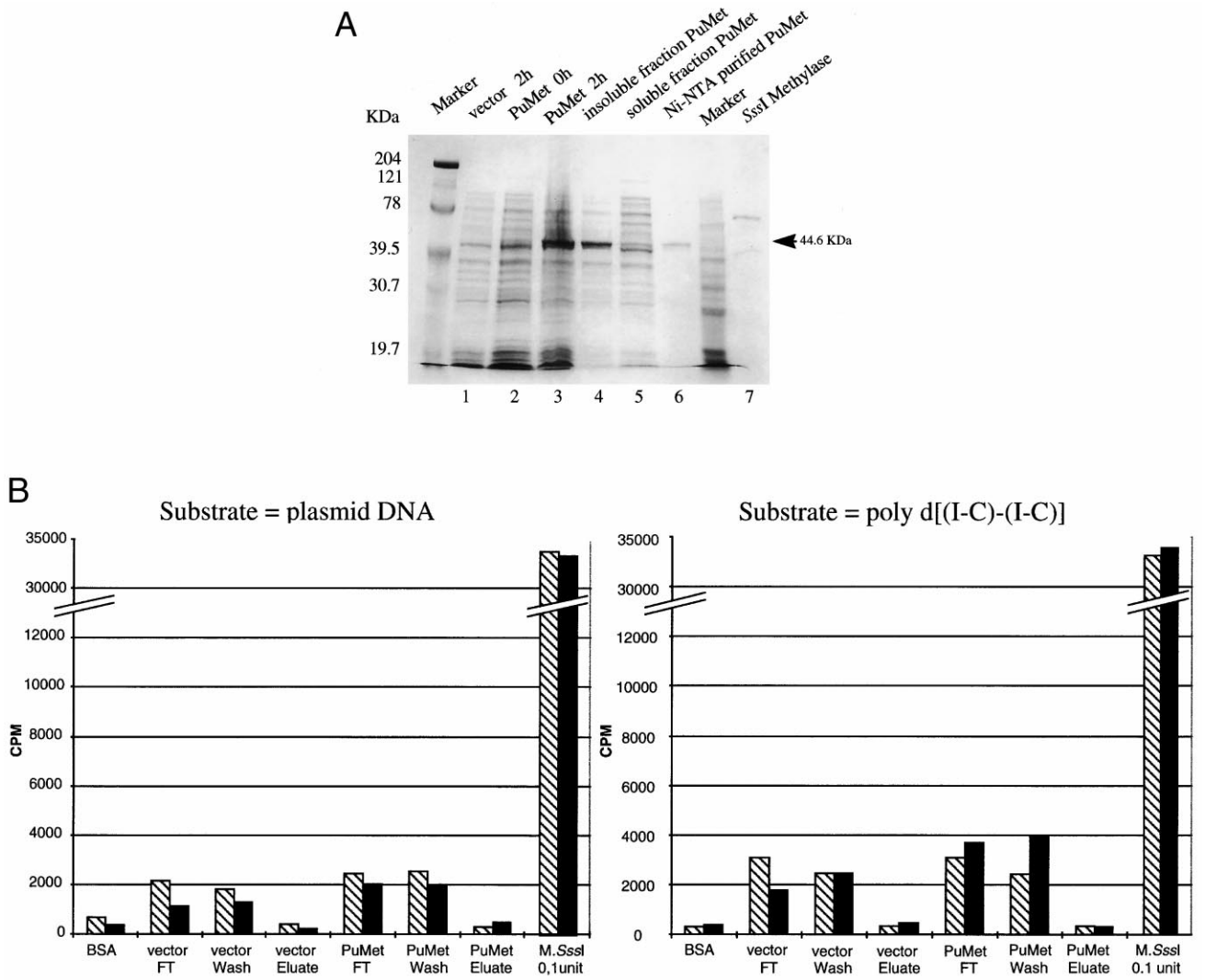


Fig. 3. A: Purification of bacterially overexpressed PuMet protein. SDS-PAGE of fractions from a typical protein purification. Lane 1: protein extract from bacteria grown with an insert-less vector control; lane 2: protein extract from uninduced PuMet-expressing bacteria; lane 3: protein extract from induced (2 h) PuMet-expressing bacteria; lanes 4 and 5: insoluble and soluble fraction of protein extract from induced PuMet-expressing bacteria; lane 6: Ni-NTA-purified PuMet; lane 7: equivalent of one unit of methylase *Sss*I (New England Biolabs). Size markers are indicated. B: Methylation assay of fractions obtained during purification of bacterially overexpressed PuMet; methylase *Sss*I (New England Biolabs) was included as a positive control and is a CpG-specific 5-methylcytosine methyltransferase. Hatched and filled bars represent duplicate samples of the same experiment. FT: flow-through. BSA indicates the use of an equivalent amount of BSA in the methylation reaction as a negative control.

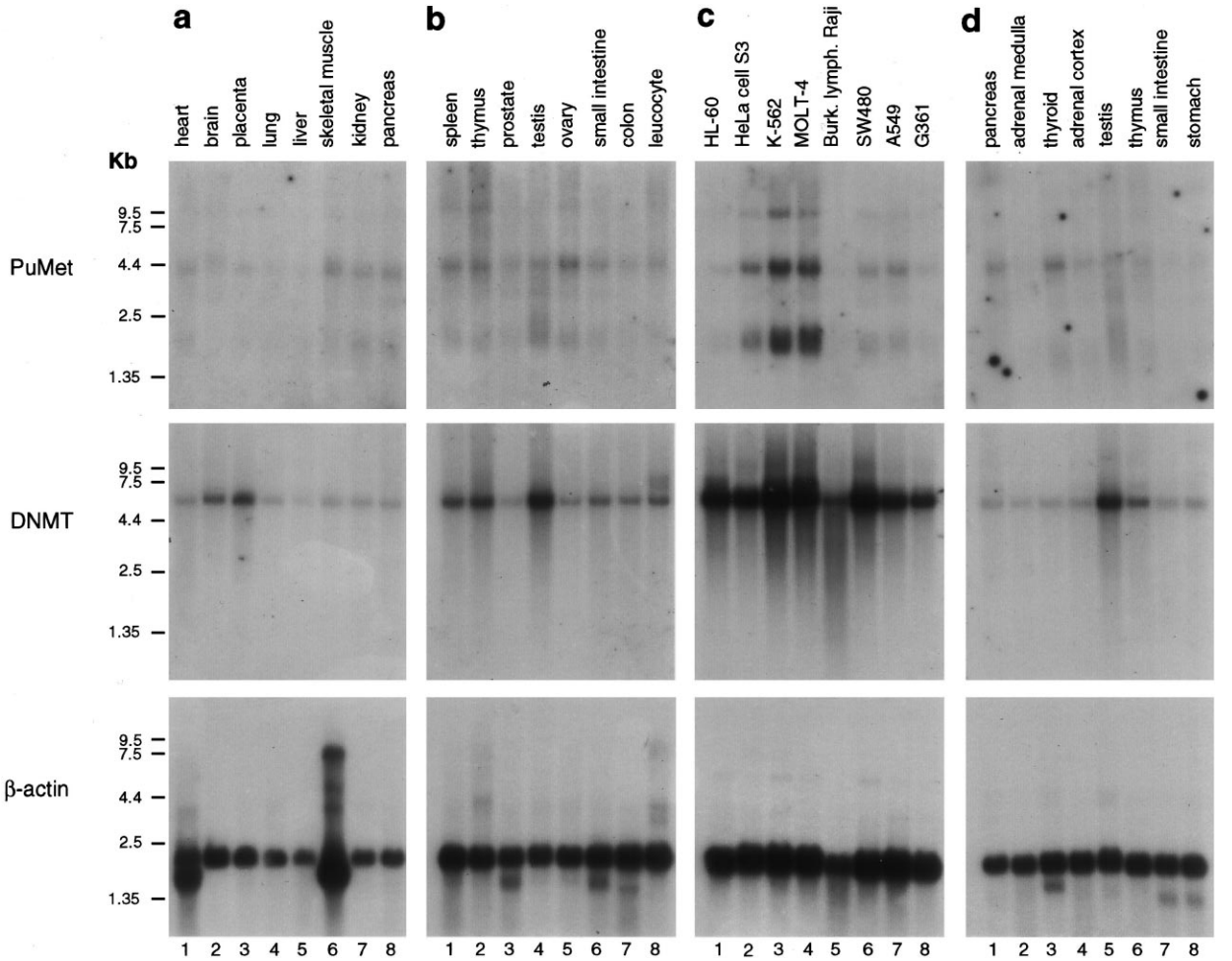


Fig. 4. Northern blot analysis of PuMet and DNMT. Multiple tissue Northern blots were hybridized as described in Section 2. The RNA source is indicated above the lanes. Exposure time of X-ray film was 80 h for PuMet, 5 h for DNMT and 4 h for β-actin, respectively. See Section 2 for details.

transfection with these constructs (data not shown). We conclude that, under our assay conditions, PuMet has no DNA methyltransferase activity. A recently identified DNA methyltransferase from *Ascobolus* was similarly overexpressed and was also inactive using these heterologous expression systems [25]. Conceivably, an additional protein or cofactor may be required for the activity of these types of eukaryotic DNA methyltransferases, because they lack the large N-terminal domain present in DNMT. However, the protein may also lack any DNA methyltransferase activity. In a recent study, Yoder et al. [26] used a mechanism-based probe to detect de novo methyltransferases in lysates from ES cells and from cells derived from different developmental stages of mouse embryos. They concluded that the predominant de novo methyltransferase activity in those tissues which are known to be active in de novo methylation is DNMT. Thus, DNMT but not PuMet may actually be responsible for de novo methyltransferase activity during early development. However, it cannot be excluded that an as yet unidentified protein is responsible for de novo methylation in mammalian systems.

3.4. Tissue-specific expression of the PuMet gene

In order to establish the expression profile of PuMet a series of Northern hybridization experiments was carried out. Three mRNA species of approximately 8, 4 and 1.8 kb are detected on Northern blots containing poly(A)$^+$ mRNA (Fig. 4). Whereas the band of approximately 4 kb probably corresponds to the identified PuMet cDNA, the nature of the larger and smaller mRNA is unknown. Overall, the gene for PuMet is expressed in most tissues analyzed; however, expression levels are very low. Expression appears to be slightly elevated in thymus, thyroid and testis (Fig. 4B, lanes 2 and 4 and Fig. 4D, lanes 3 and 5). Highest expression is seen in three cancer cell lines (Fig. 4C, lanes 2–4). The same blots were also hybridized to a probe derived from the human DNMT gene. Hybridization signals were compared using phosphoimaging analysis (see Section 2). Interestingly, expression of the DNMT gene is about 20–30-fold higher compared to PuMet, with highest expression in placenta (Fig. 4A), thymus and testis (Fig. 4B,D). Importantly, DNMT appears to be overexpressed in all cancer lines analyzed (Fig. 4C), which is consistent with the role of DNA methylation in the onset and progression of cancer [27]. Based on the low levels of expression of PuMet in the various tissues it is difficult to predict a distinct role for PuMet. Considering the approximately 20–30-fold higher expression levels of DNMT and the recent evidence for its involvement in de novo methylation of DNA

[26], it appears that PuMet may only play a role in a certain developmental stage or under conditions where it is highly upregulated.

The targeted deletion of the mouse DNMT gene resulted in embryonic lethality [12]; however, undifferentiated ES cells are able to grow. These ES cells retain the ability to methylate newly integrated retroviruses [13,28]. Thus PuMet may mediate this de novo methylation of retroviral DNA. This could be tested by generating the targeted deletion of PuMet in $DNMT^{-/-}$ ES cells.

4. Note added in proof

While this work was being written, the sequence of PuMet was published as DNA methyltransferase 2 (DNMT2) by J.A. Yoder and T.H. Bestor [Hum. Mol. Genet. 7 (1998) 279–284] (accession number AF012128).

Acknowledgements: The authors wish to thank Peter Verhasselt for excellent sequencing services. Tim Bestor is thanked for helpful discussions and communicating data prior to publication.

References

[1] Kass, S.U., Landsberger, N. and Wolffe, A.P. (1997) Curr. Biol. 7, 157–165.
[2] Kass, S.U., Pruss, D. and Wolffe, A.P. (1997) Trends Genet. 13, 444–449.
[3] Bird, A.P. (1995) Trends Genet. 11, 94–100.
[4] Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Trends Genet. 13, 335–340.
[5] Adams, R.L.P. (1995) BioEssays 17, 139–145.
[6] Bestor, T.H. and Verdine, G.L. (1994) Curr. Opin. Cell Biol. 6, 380–389.
[7] Razin, A. and Shemer, R. (1995) Hum. Mol. Genet. 4, 1751–1755.
[8] Bestor, T.H. and Tycko, B. (1996) Nature Genet. 12, 363–367.
[9] Yen, R.-W.C., Vertino, P.M., Nelkin, B.D., Yu, J.J., El-Deiry, W., Cumaraswarmy, A., Lennon, G.G., Trask, B.J., Celano, P. and Baylin, S.B. (1992) Nucleic Acids Res. 20, 2287–2291.
[10] Yoder, J.A., Chiu Yen, R.-W., Vertino, P.M., Bestor, T.H. and Baylin, S.B. (1996) J. Biol. Chem. 271, 31092–31097.
[11] Leonhardt, H., Page, A.W., Weier, H.U. and Bestor, T.H. (1992) Cell 71, 865–873.
[12] Li, E., Bestor, T.H. and Jaenisch, R. (1992) Cell 69, 915–926.
[13] Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R. and Li, E. (1996) Development 122, 3195–3205.
[14] Wilkinson, C.R.M., Bartlett, R., Nurse, P. and Bird, A.P. (1995) Nucleic Acids Res. 23, 203–210.
[15] Pinarbasi, E., Elliot, J. and Hornby, D.P. (1996) J. Mol. Biol. 257, 804–813.
[16] Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) Proc. Natl. Acad. Sci. USA 84, 4355–4358.
[17] Bairoch, A. and Apweiler, R. (1997) Nucleic Acids Res. 25, 31–36.
[18] Rodriguez-Tome, P., Stoehr, P.J., Cameron, G.N. and Flores, T.P. (1996) Nucleic Acids Res. 24, 6–13.
[19] Heng, H.H.Q., Squire, J. and Tsui, L.-C. (1992) Proc. Natl. Acad. Sci. USA 89, 9509–9513.
[20] Heng, H.H.Q. and Tsui, L.-C. (1993) Chromosoma 102, 325–332.
[21] Heng, H.H.Q. and Tsui, L.-C. (1994) Methods Mol. Biol. 33, 35–49.
[22] Chuang, L.S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G. and Li, B.F.L. (1997) Science 277, 1996–2000.
[23] Pradhan, S., Talbot, D., Sha, M., Benner, J., Hornsta, L., Li, E., Jaenisch, R. and Roberts, R.J. (1997) Nucleic Acids Res. 25, 4666–4673.
[24] Adams, R.L.P., Rinaldi, A. and Seivwright, C. (1991) J. Biochem. Biophys. Methods 22, 19–22.
[25] Malagnac, M., Wendel, B., Goyon, C., Faugeron, G., Zickler, D., Rossignol, J.-L., Noyer-Weidner, M., Vollmayr, P., Trautner, T.A. and Walter, J. (1997) Cell 91, 281–290.
[26] Yoder, J.A., Soman, N.S., Verdine, G.L. and Bestor, T.H. (1997) J. Mol. Biol. 270, 385–395.
[27] Szyf, M. (1997) Pharmacol. Ther. 70, 1–37.
[28] Tucker, K.L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E. and Jaenisch, R. (1996) Genes Dev. 10, 1008–1020.