*4666–4673* *Nucleic Acids Research, 1997, Vol. 25, No. 22* 

# Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase

**Sriharsa Pradhan, Dale Talbot[1], Mi Sha[2], Jack Benner, Linda Hornstra, En Li[3], Rudolf Jaenisch[1] and Richard J. Roberts***

New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA, [1]Whitehead Institute for Biomedical Research, Massachusetts Institutes of Technology and Department of Biology, Cambridge, MA 02142, USA, [2]Genetics Institute, 87 Cambridge Park Drive, Cambridge, MA 02140, USA and [3]Cardiovascular Research Center, Massachusetts General Hospital-East, 149 13th Street, Charlestown, MA 02129, USA

Received July 1, 1997; Revised and Accepted October 1, 1997

## ABSTRACT

**The original cDNA sequence reported for the murine DNA methyltransferase (MTase) was not full length. Recently, additional cDNA sequences have been reported that lie upstream of the original and contain an extended open reading frame with three additional ATGs in frame with the coding region [Tucker *et al.* (1996) *Proc. Natl. Acad. Sci. USA*, 93, 12920–12925; Yoder *et al.* (1996) *J. Biol. Chem.* 271, 31092–31097]. Genomic DNA upstream of this ATG contains two more ATGs in frame and no obvious splice site. We have constructed, and expressed in baculovirus, MTase clones that begin at each of these four ATGs and examined their properties. Constructs beginning with any of the first three ATGs as their initiator methionines give a predominant DNA MTase band of ~185 kDa on SDS–PAGE corresponding to translational initiation at the third ATG. The fourth ATG construct gives a much smaller protein band of 173 kDa. The 185 kDa protein was purified by HPLC, characterized by mass spectrometry and has a measured molecular mass of 184 ± 0.5 kDa. All of these MTases were functional *in vitro* and steady state kinetic analysis showed that the recombinant proteins exhibit similar kinetic properties irrespective of their length. The homogeneous recombinant enzyme from the fourth ATG construct shows a 2.5-fold preference for a hemi-methylated DNA substrate as compared to an unmethylated substrate, whereas the 185 kDa protein is equally active on both substrates. The kinetic properties of the recombinant enzyme are similar to those reported for the native MTase derived from murine erythroleukemia cells. The new clones are capable of yielding large quantities of intact MTases for further structural and functional studies.**

## INTRODUCTION

The diploid mammalian genome contains ~1 × $10^8$ CpG dinucleotides (1), of which ~60% are methylated at the 5-position of the cytosine residue (2). Cytosine methylation is the most common form of DNA methylation known so far in vertebrates and is implicated in the control of developmental processes (3), gene expression (4), X chromosome inactivation (5,6), transposition (7), recombination (8), replication (9) and genomic imprinting (10). The importance of DNA methylation during embryonic development in mice has been well-documented (11). Apart from normal biological processes it has been postulated that DNA methylation plays a role in carcinogenesis (12), genetic diseases (13) and evolution (14).

The mechanism of (cytosine-5) DNA methylation has been elucidated in detail for the prokaryotic MTases and involves the formation of a covalent intermediate between the methyltransferase (MTase) and the 6-position of the target cytosine (15,16). Crystallographic evidence has shown that an unusual mode of DNA binding is employed in which the target cytosine residue swings completely out of the DNA helix and is positioned in the active site, where the chemistry can proceed without steric hindrance (17,18). It is believed that the murine MTase follows the same mechanism, because there is great sequence similarity between the C-terminal 500 amino acid residues of the murine MTase and the bacterial MTases (19). However, at its N-terminus the murine MTase has more than 1000 additional amino acids whose function is mostly unknown. One small segment has been implicated in co-localizing the MTase with the replication machinery (20).

DNA methylation is sequence specific and in the case of both mouse and human MTases, CpG dinucleotides are methylated in a symmetrical manner. After each round of DNA replication the methylated parental strands give rise to a pair of hemi-methylated daughter strands. To preserve the methylation pattern, the mammalian enzyme methylates the newly synthesized strand. The MTase localizes with replication foci in the nucleus as shown by confocal microscopy in Swiss 3T3 cell lines (20). These properties have led to this MTase being considered the prime activity responsible for maintaining the methylation pattern of the DNA. Apart from its maintenance activity, the previously-characterized MTase also shows weak activity on unmethylated target sites. There is evidence that this maintenance MTase can act on unmethylated, single-stranded DNA at a comparable rate with that of hemi-methylated DNA (21). Furthermore, structural abnormalities in DNA such as loops and hairpins also lead to *de novo* methylation by the enzyme *in vitro* (21,22). It is not known whether these are normal activities *in vivo*. Recently, it has been shown that an additional *de novo* MTase activity is present in mutant murine embryonic stem (ES) cell lines, which have been

*To whom correspondence should be addressed. Tel: +1 508 927 3382; Fax: +1 508 921 1527; Email: roberts@neb.com

disrupted for the maintenance MTase (23). Thus mammals appear to have more than one MTase.

The previously-reported, truncated DNA MTase cDNA failed to complement mutant ES cell lines, in which the gene for the maintenance MTase had been disrupted (24). However, complementation could be achieved by constructs that contained sequences upstream of the previously reported start site (25,26). These upstream sequences contained an open reading frame with a coding capacity of at least 171 amino acids. The resulting constructs, when transfected into mutant ES cells, expressed a protein that co-migrated with the endogenous maintenance MTase. Functional rescue was also observed in these transfected cells, since the genomic DNA became re-methylated and the cells regained the ability to form teratomas displaying a variety of differentiated cell types (24). This suggests that these new N-terminal sequences may have an essential function or that the correct translation of the gene must begin at the newly-discovered upstream ATG. To test the properties of these additional sequences we have prepared four different MTase constructs in baculovirus transfer vectors such that each has either the first, second, third or fourth ATG available as its start site for translation. These recombinant viruses have been purified, amplified and used for the production of recombinant protein. The proteins have been purified and sequenced, and their steady state kinetics have been studied *in vitro* using poly (dI.dC:dI.dC) and either unmethylated or hemi-methylated synthetic oligonucleotides as substrates.

## MATERIALS AND METHODS

### Transfer vectors

DNA MTase expression constructs were derived from pMG (T.Bestor), p7B (K.Tucker) and pSX121 (E.Li) and are presented schematically in Figure 1. The pMMT4 construct was made by inserting pMG, which is a 4930 bp DNA MTase cDNA into a baculovirus transfer vector pVL1393 in which recombinant gene expression is driven by the polyhedrin promoter (27). This construct expresses a truncated version of the MTase beginning at the fourth ATG and is essentially identical to constructs described previously (26,28,29). pMMT3 contains a version of the MTase gene, which begins at the third ATG. It was prepared by ligating the newly-isolated upstream cDNA fragment from p7B (*Bss*HII–*Eco*RI) into pMMT4 giving a plasmid that is ~200 bp larger than pMMT4. This third ATG construct was also cloned as an in-frame fusion with a hexahistidine leader peptide and the enterokinase site of the vector pBlueBac2C (Invitrogen). This construct, pHisMMT3, expresses an MTase that is easily purified on an IMAC column containing chelated nickel. The first ATG construct, pMMT1, was prepared by ligating the *Bss*HII–*Bgl*I fragment of pSX121 to pMMT3. This construct is 150 bp longer than pMMT3 and has two additional ATGs in-frame with the ATG of pMMT3 (third ATG). The additional DNA sequences were derived from the corresponding genomic DNA of the murine MTase (25). The first ATG of pMMT1 overlapped an *Sph*I (GC**ATG**C) site so that digestion of pMMT1 with *Sph*I followed by blunting of the ends by T4 polymerase and subsequent ligation destroyed this first ATG. The resulting construct, pMMT2, was thus missing the first ATG, but did contain the second ATG as a potential initiation site for translation.

The transfer vectors, pMMT1, pMMT2, pMMT3 and pMMT4 were cotransfected with baculogold DNA, a modified, linearized *Autographa californica* nuclear polyhedrosis virus (AcNPV) DNA (Pharmingen) onto a monolayer of SF9 insect cells. pHisMMT3 was cotransfected using the Bac-N-Blue transfection kit (Invitrogen). Transfection supernatants were screened for recombinant baculovirus using the agarose overlay technique (27). For the pVL1393-derived transfer vectors neutral red was used, whereas bluo-gal (Gibco-BRL) was used for pHisMMT3 to isolate pure plaques using standard techniques (27). Each purified plaque was amplified to reach a viral titer between $2 \times 10^8$ and $1 \times 10^9$ pfu/ml.

### Expression of recombinant murine DNA MTase in SF9 cells

Insect cells (SF9) were routinely used for the expression of recombinant protein. SF9 cells were grown and maintained in TNM-FH complete medium supplemented with 10% fetal calf serum (27). For initial expression studies 25 $cm^2$ flasks were used. However, for routine purification of protein, spinner culture flasks were used. SF9 cells at a density of $2 \times 10^6$/ml were infected at a multiplicity of infection (MOI) of 10. The cells were incubated with gentle stirring. Cells were harvested 72 h post-infection (PI) unless otherwise stated. Infected cells were centrifuged at 4°C at 1200 *g*. The pelleted cells were washed once with phosphate buffered saline.

### Purification of recombinant murine DNA MTase

Typically 3 g of infected cells were resuspended in 15 ml lysis buffer M (20 mM Tris–HCl, pH 7.4, 1 mM $Na_2EDTA$, 10% sucrose, 10 µg/ml TPCK, 10 µg/ml TLCK, 5 µg/ml E64, 5 µg/ml leupeptin, 7 µg/ml PMSF) supplemented with 250 mM NaCl. The cell suspension was sonicated on ice for 30 s using a model W-225R (Heatsystem-Ultrasonics, USA) sonicator in pulsed mode with 50% duty cycle. The extract was incubated on ice for 30 min and then centrifuged at 11 000 *g* in a JA18 rotor at 4°C for 45 min. The supernatant was collected and an equal volume of buffer M, without NaCl, was added to bring the final concentration of NaCl to 125 mM. This mixture was centrifuged for a further 30 min at 11 000 *g* to remove the precipitate. The supernatant was loaded on a 70 ml Q Sepharose HR (Pharmacia) column pre-equilibrated with buffer A (M with 100 mM NaCl) at 1 ml/min using a superloop. The column was washed with 200 ml of buffer A. The bound proteins were eluted with a linear gradient of 0–1.0 M NaCl in 300 ml of buffer B (buffer M supplemented with 1.0 M NaCl). Fractions containing MTase activity were pooled and the final salt concentration of the protein solution was brought to 100 mM NaCl by the addition of buffer M.

This protein solution was loaded on a 5 ml Heparin–Sepharose (Pharmacia) column. The column was washed with buffer A containing 100 mM NaCl and the bound protein was eluted with 90 ml of a linear gradient from 0.1 to 1.5 M NaCl in buffer M. Fractions containing MTase were pooled and dialysed against buffer M containing 100 mM NaCl. The dialysed MTase was loaded on a 1 ml Source 15Q (Pharmacia) column. The unbound proteins were washed with buffer M containing 100 mM NaCl and the bound proteins were eluted with 60 ml of a linear gradient from 0.1 to 1.0 M NaCl in buffer M.

The MTase fraction was further purified by gel filtration on a Hiload Superdex 200 prep grade 16/60 (Pharmacia) column equilibrated with buffer M containing 100 mM NaCl at 0.2 ml/min. The peak fractions were collected and dialysed against buffer M containing 50% glycerol.

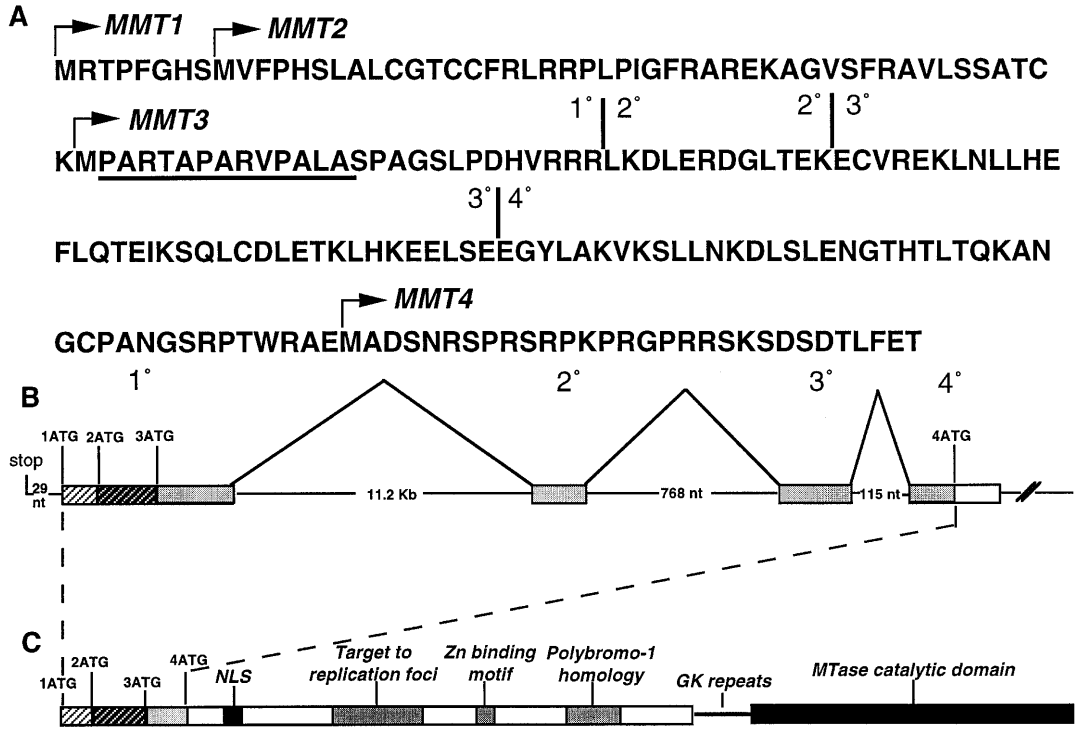


**Figure 1.** Murine DNA MTase with additional N-terminal amino acids. (**A**) Conceptual translation of the newly discovered 5′-sequences of the murine DNA MTase. MMT1, 2, 3 or 4 represent the murine MTase protein translated from the first, second, third or fourth initiator methionine respectively. The underlined amino acid sequence was confirmed by N-terminal sequence analysis of MMT3 by Edman degradation. Vertical bars between amino acids denotes the location of introns in the genomic DNA sequence. The exons are numbered from 1° to 4°. Methionine 1 and 2 are derived from the genomic sequence preceding the newly discoved 5′-end of the cDNA, which only contains the third methionine. (**B**) Schematic of the 5′-end of the murine MTase gene showing the relative locations of the four potential initiator ATGs within the first four exons. The first stop codon is TAG and is located 29 bases before the first ATG. The first three methionines all lie in the first exon. Previously the fourth ATG, located in exon 4° was believed to be the translation initiator (25). More recently the third ATG is believed to be the first methionine for translation (24,35 and this work). (**C**) Schematic showing the complete murine MTase gene. NLS shows the putative nuclear localisation signal. Grey boxes mark the sequences responsible for targeting the enzyme to replication foci, a putative Zn finger and a region that shares homology with the chicken polybromo-1. The regulatory and catalytic domains are linked by six GK dipeptide repeats.

## DNA substrates for MTase assay

Poly (dI.dC:dI.dC) with an average length of 7000 bp was obtained from Sigma. Oligonucleotides were made by the Organic Synthesis division of NEB and their purity was checked by $^{32}$P-labelling and PAGE. Duplex oligonucleotides were prepared by adding equimolar amounts of the complementary strands. The annealing was checked by PAGE. The following oligonucleotides were used, M represents 5-methylcytosine.

1. HM 30mer: 5′-AGACMGGTGCCAGMGCAGCTGAGCMGGATC-3′
   3′-TCTGGCCACGGTCGCGTCGACTCGGCCTAG-5′
2. UM 30mer: 5′-AGACCGGTGCCAGCGCAGCTGAGCCGGATC-3′
   3′-TCTGGCCACGGTCGCGTCGACTCGGCCTAG-5′
3. DM 30mer: 5′-AGACMGGTGCCAGMGCAGCTGAGCMGGATC-3′
   3′-TCTGGMCACGGTCGMGTCGACTCGGMCTAG-5′
4. Poly (dI.dC: dI.dC): poly deoxyinosinic-deoxycytidylic acid.

## MTase assay

When screening clones for expression, the MTase was assayed in a crude extract prepared from infected cells. The cell pellet was briefly sonicated in buffer M and the cell debris was pelleted by centrifugation. The supernatant was used to assay the incorporation of $^{3}$H methyl groups into poly (dI.dC:dI.dC). The reaction mix (70 µl) contained 3.3 µM $^{3}$H AdoMet (1.08 µCi), 1 µg DNA and protein in M buffer. The reaction was incubated at 37°C, for 1 h and then cooled in ethanol/dry ice and processed as described earlier (30).

To study the steady state kinetic parameters of the purified enzymes, various DNA substrates were used (as indicated above) and the reaction mix contained 20 mM Tris–HCl, pH 7.4, 1 mM $Na_2EDTA$, 10 µg/ml TPCK, 10 µg/ml TLCK, 5 µg/ml E64, 5 µg/ml leupeptin and 7 µg/ml PMSF. The concentration of protein was estimated using the Bradford reagent with BSA as the protein standard.

## Determination of molecular mass and sequencing of recombinant murine MTase

Murine MTase preparations were further purified for determination of molecular mass. A sample (2 µg) of MTase in 20 µl buffer M plus 10% sucrose, was injected on a microbore HPLC system containing a 250 × 0.8 mm, POROS/R1/H particle column (LC Packings) equilibrated with 95% buffer A (water, 0.1% TFA) and 5% solvent B (acetonitrile, 0.1% TFA). The column was developed for 2 min with a flow rate of 200 µl/min followed by a 20 min linear gradient from 5 to 100% solvent B. Protein was detected by UV absorption at 214 nm. The major peaks were collected and subjected to Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectroscopy on a Perseptive

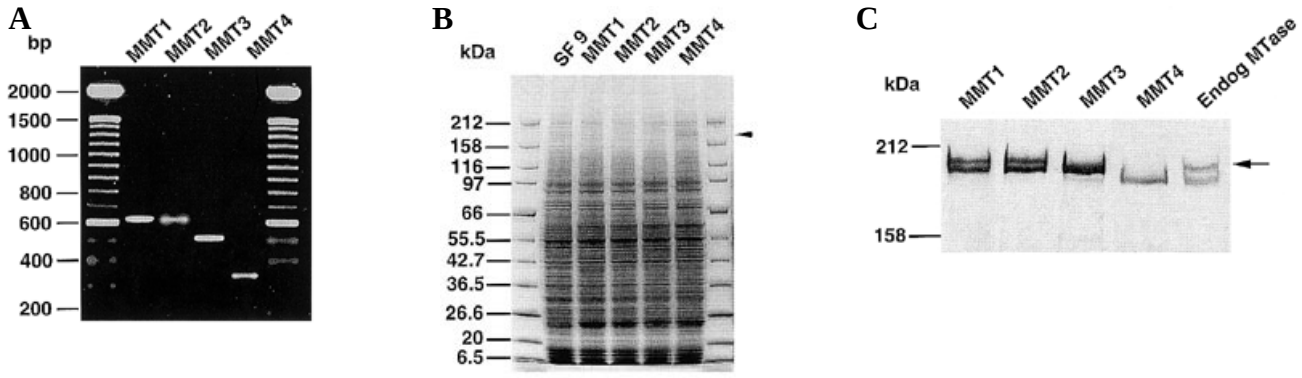


**Figure 2.** Generation of recombinant baculovirus constructs with various length of murine MTase gene. (**A**) PCR analysis of the recombinant baculovirus DNAs. The forward primer binds from –44 (nt 4049) to –21 (nt 4072) in front of the start of the polyhedrin gene, using the nomenclature of reference 27. The reverse primer, CTGCCATTTCTGCTCTCCAGGTTG, is located 337 bp downstream of the third ATG spans the region containing the fourth ATG. The outer lanes contain a 100 bp DNA ladder (BRL) as a MW marker. Lane 2, pMMT1 template; lane 3, pMMT2 template; lane 4, pMMT3 template; lane 5, pMMT4 template. (**B**) Extracts from SF9 cell infected with various MTase constructs were resolved on a 4–20% SDS–PAGE and stained with Coomassie Blue. (**C**) Western blot analysis of the same protein mixtures shown in (B). The top of the gel is enlarged. An antibody (Ab 334) specific for the residues just downstream of the fourth ATG of the MTase was used to detect the MTase protein. The MW markers are in kDa. The arrow indicates the MTase specific bands.

Biosystems Voyager DE instrument using recrystallized sinapinic acid (Sigma) as the matrix (31). A maltose binding protein–β galactosidase fusion protein of 158 194 Da was used as the standard. The protein fractions were also analysed by western detection with Ab 334 (as described below).

For N-terminal sequence analysis the murine MTases prepared as described above, were subjected to electrophoresis on an 8% Tris–glycine polyacrylamide gel (Novex) and electroblotted according to the procedure of Matsudaira (32), with modifications as previously described (33,34). The membrane was stained with Coomassie Blue R-250 and the appropriate protein bands of ~173 or 185 kDa were excised and subjected to sequential degradation (33).

### Immunoblot analysis of different MTases

Cell extracts or purified protein samples were mixed with SDS–PAGE sample loading buffer and incubated at 90°C for 3 min. The protein mixtures were separated in a 4–20% polyacrylamide ISS miniplus SepraGel. The protein bands were blotted on to a PVDF membrane and probed using a polyclonal rabbit antibody (Ab 334) raised against a peptide, RSPRSRPKPRGPRRSK, located four amino acids downstream of the fourth ATG in the murine MTase protein (24).

## RESULTS

### Expression of recombinant murine MTase in SF9 insect cells

Co-transfection of the four pMMT plasmids (Fig. 1A) with linear AcNPV DNA resulted in homologous recombination of the MTase cDNAs carried by each construct. The transfection supernatants were harvested and viral DNAs isolated. To check that the recombination had proceeded correctly, a small sample of each viral DNA was used for PCR amplification using one primer from the polyhedrin gene and a reverse primer from the MTase cDNA. All the recombinant viral DNAs produced the expected length of PCR product (Fig. 2) confirming that the cDNA has recombined downstream of the polyhedrin promoter. These supernatants were used for plaque purification to isolate single recombinant viruses. From each transfection, six putative recombinant virus plaques were isolated. Each viral plaque was amplified at least twice to reach a titre of $1 \times 10^{11}$ pfu/ml.

An SF9 insect cell line was used for the expression studies. This cell line offered the following advantages. It had a short doubling time of between 18 and 24 h and it grew well in monolayer and suspension culture. Three days after infection the cell extracts were tested for recombinant MTase expression either by assaying for the transfer of $^{3}$H methyl groups into poly (dI.dC:dIdC) or by western analysis using an N-terminal specific antibody (Ab 334). This allowed an assessment of the level of expression of functional protein, it also indicated that the recombinant MTase was intact. All four pMMT constructs produced catalytically active recombinant MTase as predicted by the cDNA sequence. However, western analysis of cell extracts revealed that the constructs pMMT1 or pMMT2 produce proteins that co-migrate with the MTase that is produced by pMMT3. The molecular weight of the MTase produced by pMMT3 was estimated from SDS gel analysis to be ~185 kDa (Fig. 3). pMMT4, which has a much smaller cDNA, expressed a protein with an estimated molecular weight of ~173 kDa. This is slightly larger than the theoretical mass of unmodified MMT4, which is 169 955 Da. This small discrepancy may be the result of post-translational modification of the MTase, such as phosphorylation, which is known to occur in insect cells (27). The mouse MTase is thought to be phosphorylated in its native state (36).

### Different translational initiation/processing of proteins?

A closer look at the western blot reveals that pMMT1 and pMMT2 give at least one additional protein band cross-reacting with Ab 334. This band is larger than the common MTase band at 185 kDa, which is the sole product from pMMT3. The larger protein band could be the initial, intact form of the protein, which is subsequently processed to give the 185 kDa band. Alternatively the double bands could be the result of two translational start sites being used. While we suspect the latter possibility is more likely, at present we cannot distinguish between these alternatives.

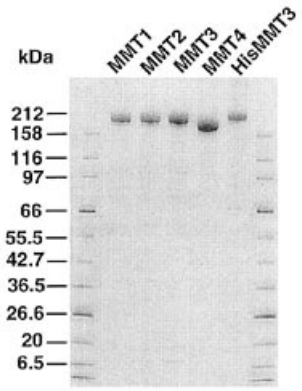


**Figure 3.** Purified MTases. Recombinant murine MTases were purified from various constructs to apparent homogeneity and resolved by 4–20% SDS–PAGE. Proteins were visualised by Coomassie Blue stain. Lanes 2–6, contain the indicated purified MTases. Protein markers are in the outer lanes 1 and 7.

## Purification of recombinant MTases

The mammalian DNA MTase is reported to be a highly unstable protein that is extremely susceptible to proteolytic breakdown (37). We attempted to minimize proteolysis by carrying out all steps of the purification at 4°C and by adding a cocktail of protease inhibitors (E64, TPCK, TLCK, leupeptin, PMSF). These protease inhibitors have been used previously to facilitate the purification of intact murine MTase from both mouse erythroleukemia cells and baculovirus expression systems (26,28). Harvested cells were processed within 12 h to minimize proteolysis. During sonication, 250 mM NaCl was present to ensure that most of the MTase, which binds tightly to cellular DNA, was rendered soluble. Following the removal of debris by centrifugation, >80% of the MTase was in the supernatant in soluble form.

A Q-sepharose HR (Pharmacia) column was used as the first step for purification. Typically 700–1000 mg of the protein was applied to it. MTase binds to this column at low salt concentration and was eluted as a single peak between 375 and 475 mM. This purification step removed the majority of the cellular protein. This pooled fraction was diluted with buffer M to lower the NaCl concentration to 100 mM and then applied to a Heparin–Sepharose column. The MTase binds strongly to Heparin–Sepharose and elutes at high salt between 700 and 800 mM. This step resulted in a high degree of purification and 80% of the eluted protein was MTase. The pooled protein solution was dialysed against buffer M containing 100 mM NaCl for 6 h to remove excess NaCl. This sample was loaded on to a Source Q 15 (Pharmacia) column. The bound protein eluted between 150 and 200 mM NaCl. At this step the protein was apparently pure as observed by Coomassie staining. If additional protein bands were present, then a gel filtration step was carried out. For most purposes the protein eluting from the Source Q purification step was used.

## Molecular mass of the murine MTases

To characterize the recombinant protein preparations, we attempted to determine the N-terminal sequences of MMT3 and MMT4. For MMT3 the first 12 residues of the protein matched the amino acid sequence predicted from the cDNA (Fig. 1A). However, for MMT4 no sequence was obtained, suggesting that the N-terminus is blocked. It should be noted that the N-terminal sequence of the endogenous MTase from mouse embryonic stem cells matches that of MMT3 (D.Talbot, unpublished). The molecular mass of MMT3 and MMT4 were determined by MALDI-TOF to be 184 ± 0.5 and 170 ± 0.5 kDa, respectively (Fig. 4). The theoretical molecular masses of the unmodified proteins are 183 482 Da (1619 amino acids) for MMT3 and 170 002 Da (1501 amino acids) for MMT4.

## Enzymological properties of recombinant MTases

Since only two predominant MTases with apparent molecular weights of 185 and 173 kDa were observed, the kinetic properties of these two forms of the enzyme were examined in detail. The first substrate used for the study of steady state kinetics was a poly (dI.dC:dI.dC) duplex. This substrate is classically used for monitoring the activity of mammalian MTases during purification

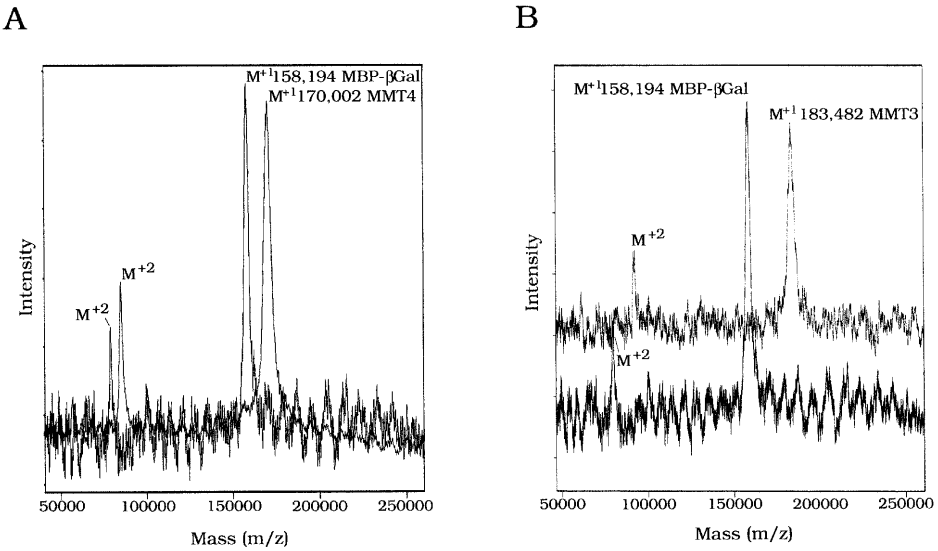


**Figure 4.** Mass spectroscopy of processed murine MTases. Spectra of the standard protein sample, maltose binding protein–β galactosidase fusion protein (MBP-βGal) and murine MTases MMT3 (**A**) or MMT4 (**B**). The observed molecular masses are indicated on the top of each peak. The theoretical mass of MBP-βGal is 158 194 Da and the unmodified forms of MMT3 and MMT4 are 183 482 and 170 002 Da respectively.

and also behaves as a substrate whose methylation rate is comparable to hemi-methylated DNA substrates (39). The reaction proceeded in a linear fashion for ~80 min in the presence of 20–100 nM of enzyme using poly (dI.dC:dI.dC) as substrate (data not shown). Each molecule of poly (dI.dC:dI.dC) undergoes a series of catalytic turnovers during this time and has a large number of potential target sites. Table 1 summarizes the results obtained using poly (dI.dC:dI.dC) as substrate. Both the MMT3 and MMT4 enzymes have similar $K_m$ values and convert 4.5 cytosines to 5-methylcytosine on the substrate per hour as determined by $K_{cat}$. This result is comparable to the finding of a $K_{cat}$ in the range of 4.9–6.4 for the MTase isolated from murine erythroleukemia cells (40). The specificity constant ($K_{cat}/K_m$) was also comparable. The kinetic parameters for MMT3 containing a His tag are slightly lower. The $K_m$ for AdoMet by both the recombinant MTases, MMT3 and MMT4 were 2.4 and 2.5 μM, respectively. This result is consistent with the finding of Bolden *et al*., who report a $K_m$ of 8.0 for the crude native enzyme using hemi-methylated ΦX174 DNA as substrate (41). Overall, the two recombinant MTases behave in a very similar fashion when poly (dI.dC:dI.dC) is used as substrate.

To test whether the recombinant enzymes show a preference for hemi-methylated substrates, two 30mer oligonucleotide duplexes, containing three CpG sites, were used as substrates. These three sites were either uniformly hemi-methylated, unmethylated or doubly methylated and thus contained either three, six or zero CpG sites available for methylation. Both the 185 and 173 kDa enzymes have comparable values for $K_m$ and $K_{cat}$ for the unmethylated substrate (Table 2B). The hemi-methylated duplexes are preferentially methylated (2.4-fold) by MMT4 over MMT3 (Table 2B and C). Surprisingly the longer MTase (MMT3) has a very similar catalytic turnover number on both unmethylated and hemi-methylated duplexes.

## Methylation of single- and double-stranded oligonucleotide substrates by recombinant MTase

The mammalian DNA MTases can use both single- and double-stranded DNAs as substrates, although the biological significance of single-stranded DNA methylation is not well understood. Smith *et al*., have shown that the mammalian MTase can use as substrate CpG sites that are present in a long stem region of a stem–loop structure (42,43). Christman *et al*., using an extract from Friend erythroleukemia nuclei, have demonstrated that the presence of 5-methylcytidine (5mC) in single-stranded DNA can serve as a *cis*-acting signal for *de novo* methylation (21). These findings are based on both crude nuclear extracts and partially purified MTase preparation. It was of interest, therefore, to test the two purified MTases, MMT3 and MMT4, to see if they differed in their action on single- and double-stranded substrates. Oligonucleotide substrates were prepared that were identical to those used previously by other groups (21,29). The results are presented in Table 3 and show that both MMT3 or MMT4 can act in a similar manner on all substrates tested. These results support previous findings (21) and show that 5mC can stimulate methylation even if it is present at a non-CpG site.

**Table 1.** Comparison of the steady state kinetic parameters of recombinant MTases using ds poly (dI.dC:dI.dC)

| Enzyme | $K_m$ (μM) | $K_{cat}$ ($h^{-1}$) | $K_{cat}/K_m$ ($h^{-1}$ $M^{-1} \times 10^6$) | $K_m$ (AdoMet) |
|---|---|---|---|---|
| MMT3 | 1.86 ± 0.07 | 4.48 ± 0.55 | 2.62 | 2.50 ± 0.53 |
| HisMMT3 | 1.06 ± 0.28 | 1.58 ± 0.17 | 1.49 | – |
| MMT4 | 2.24 ± 0.4 | 4.40 ± 0.5 | 2.0 | 2.40 ± 0.25 |

Reaction mixtures contained 75 nM MTase and poly (dI.dC:dI.dC). The concentration of CpI dinucleotides were 0.098, 0.14, 0.194, 0.29, 0.39 and 0.486 μM. Reaction conditions were as indicated in the Materials and Methods. Each reaction was performed in duplicate and the data were plotted based on a Michaelis–Menten steady state equation. For the determination of $K_m$ values, the reaction mixtures contained 75 nM enzyme, 19.48 μM CpI and AdoMet at concentrations of 0.5, 0.75, 1.0, 1.25, 1.50, 1.75, 2.0, 3.0 and 4.0 μM. ± represents standard error.

**Table 2.** Steady state kinetics of recombinant murine MTase with oligonucleotides

| Enzyme | $K_m$ (μM) | $K_{cat}$ ($h^{-1}$) | $K_{cat}/K_m$ ($h^{-1}$ $M^{-1} \times 10^6$) |
|---|---|---|---|
| **A.** Hemimethylated 30mer, ds oligonucleotide substrate | | | |
| MMT3 | 0.22 ± 0.02 | 0.29 ± 0.01 | 1.32 |
| MMT4 | 0.60 ± 0.08 | 0.67 ± 0.07 | 1.16 |
| **B.** Unmethylated 30mer, ds oligonucleotide substrate | | | |
| MMT3 | 0.42 ± 0.03 | 0.29 ± 0.02 | 0.7 |
| MMT4 | 0.41 ± 0.06 | 0.28 ± 0.04 | 0.68 |
| Enzyme | $K_{cat}$ ($h^{-1}$) HM/UM | Relative $K_{cat}/K_m$ HM/UM | |
| **C.** Preference for hemi-methylated DNA by murine MTases | | | |
| MMT3 | 1.06 ± 0.23 | 1.88 | |
| MMT4 | 2.32 ± 0.32 | 1.70 | |

The olignucleotide sequences are in Materials and Methods. The reaction mix contained 75 nM enzyme and 3.3 μM AdoMet and ds oligonucleotides. The concentrations of CpG dinucleotides were 0.011, 0.022, 0.054, 0.108, 0.216 and 1.08 μM for hemi-methylated and 0.022, 0.043, 0.108, 0.216, 0.432 and 2.16 μM for unmethylated oligonucleotide duplex reactions with the MMT3 enzyme. The background incorporation of tritium was measured under identical conditions with the doubly methylated oligonucleotide duplexes. ± represents standard error.

**Table 3.** (**A**) Stimulation of *de novo* methylation on single-stranded oligonucleotides and methylationon on double stranded oligonucleotides by recombinant MTases. (**B**) Methylation on double stranded oligonucleotides by recombinant murine MTase

| Oligonucleotide | $^3$H incorporated/30 min | |
|---|---|---|
| | 3rd ATG MMT | 4th ATG MMT |
| **A** | | |
| ATTGCGCATTCCGGATCCGCGATC | 290 | 630 |
| ATTG**M**GCATTC**M**GGATC**M**GMGATC | 230 | 530 |
| ATTGCG**M**ATT**M**CGGAT**M**CGCGATC | 1550 | 2560 |
| ATTG**M**GCATTCCGGATCCGCGATC | 1095 | 1840 |
| ATTGCG**M**ATTCCGGATCCGCGATC | 3360 | 5090 |
| ATTGCGCATTCCGGAT**M**CGCGATC | 420 | 760 |
| ATTGCGCATTCCGGATC**M**GCGATC | 800 | 1430 |
| ATTGCGCATTCCGGATCCG**M**GATC | 510 | 1185 |
| GATCGCGGATCCGGAATGCGCAAT | 260 | 581 |
| poly dI.dC:dI.dC 7000bp | 22990 | 70160 |
| **B** | | |
| a) GTGAATTCACATAGTACCGGATGTCGACTAATCGATATTGCGCATCTCGAGTGAATTCTGCACTTA<br>AGTGTATCATGGCCTACAGCTGATTAGCTATAACGCGTAGAGCTCACTTAAGAC | 1500 | 3000 |
| b) GTGAATTCACATAGTAC**M**GGATGTCGACTAAT**M**GATATTGCGCATCT**M**GAGTGAATTCTGCACTTA<br>AGTGTATCATGG**M**CTACAGCTGATTAG**M**TATAACGCGTAGAG**M**TCACTTAAGAC | 2000 | 4800 |
| c) GTGAATTCACATAGTAC**M**GGATGTCGACTAAT**M**GATATTGCGCATCT**M**GAGTGAATTCTGCACTTA<br>AGTGTATCATGGCCTACAGCTGATTAGCTATAACGCGTAGAGCTCACTTAAGAC | 14000 | 33000 |

(**A**) Reaction mixture contained 250 pmol of the oligonucleotides. Reaction conditions were as indicated in Materials and Methods. Each reaction was performed in duplicate and the average of the two independently obtained values was taken. M represents 5mC. Poly dI.dC:dI.dC is double stranded. The sequences of oligonucleotides are based on Christman *et al.* (1995).

(**B**) Reaction mixture contained 50 pmol of oligo duplex and was performed as per Materials and Methods. Each reaction was performed in duplicate and the average of the two independently obtained values was taken. M represents 5mC. Oligo a) is unmethylated, b) is methylated at three CpGs on both strands and c) is the hemi-methylated version of b. The oligonucleotide sequences are based on Tollefsbol and Hutchison (1995).

## DISCUSSION

We have used a baculovirus expression system to produce high levels of murine MTases suitable for biochemical studies. From the four constructs we prepared, two predominant forms of the MTase were produced. The MMT4 construct was based on the previously published sequence (25) and thus gave a truncated version of the enzyme. The incorporation of the newly discovered N-terminal sequence allowed us to express the MMT3 construct, which produced a protein comparable to the endogenous enzyme (24,26). The recombinant enzyme was phosphorylated at serine/threonine residues (data not shown). Further addition of genomic sequences lying upstream of these cDNA sequences, potentially could give yet longer forms of the enzyme. However, both upstream constructs gave rise to proteins that were mixtures of the 185 kDa protein and at least one longer product. Enterokinase mapping and antibody identification of N-terminal peptides has shown that the J1 ES cell MTase appears identical to the MMT3 construct (D.Talbot and S.Pradhan, unpublished results).

Using the expression and purification methods described in this paper, yields of either form of the mouse MTase were ~10 mg/l of infected SF9 cells ($2 \times 10^9$ cells). From the larger constructs, the isolated MTase protein is comparable in size to the endogenous MTase protein (37). The previously reported protein expressed in *Eschericia coli* (29) was a truncated protein and was made in only low yield. Furthermore, the protein produced in *E.coli* lacked any post-translational modification such as phosphorylation. There has also been one report of the expression in baculovirus of this same truncated protein with a leader peptide sequence and a hexahistidine N-terminal tag (28). We now describe a comparable construct, but which expresses the much longer protein containing the additional N-terminal sequences found recently.

The recombinant MTase has the same remarkably slow turnover number as the native enzyme (44). The turnover number of 4.5/h for the mammalian MTase is much less than prokaryotic M.*Hha*I, which is about 78/h (15). Apart from the above findings both recombinant proteins showed $K_{cat}$ and $K_{cat}/K_m$ values comparable with those of the MTase, with a reported mass of ~190 kDa, derived from murine erythroleukemia cells suggesting that the recombinant MTase is functionally similar.

The mammalian MTase consists of two functionally independent domains, the N-terminal or regulatory domain and the C-terminal domain, which contains the catalytic center. The N-terminal region of the mouse enzyme has been shown to contain a nuclear localisation signal (NLS), disruption of which leads to failure of the enzyme to relocate to the nucleus (20). Once it moves into the nucleus it associates with several other proteins and enzymes at the replication foci and DNA methylation occurs soon after DNA synthesis begins (20). This enzyme behaves as a classical maintenance DNA MTase. Separation of the N-terminal domain by cleavage with V8 protease greatly increases the *de novo* activity of the murine enzyme (39). This raised the possibility that the *de novo* MTase activity of the enzyme is under the regulatory control of intra- or inter-molecular interactions mediated by the N-terminal domain. The previously published sequence (25) failed to provide functional complementation of the hypomethylation phenotype of ES cells homozygous for various mutations in the DNA MTase. However the methylation pattern was restored when

the previously reported cDNA was fused to some additional upstream sequences discovered recently (24,26). The ability of this construct to rescue the MTase function of these mutant ES cells indicates that the additional N-terminal segment plays a role *in vivo*. However, it is not clear if this indicates true *de novo* methylation or whether it is the concerted action of this MTase with a separate *de novo* MTase that leads to the global increase in methylation.

The recombinant proteins described here were tested and compared to see if the new, longer product showed an increased *de novo* activity relative to the truncated product. Based on the results of steady state kinetics on oligonucleotide substrates, both the 185 kDa and the truncated MTases have similar catalytic properties *in vitro*. Overexpression of DNA MTase cDNA in fibroblasts leads to an increased level of DNA methylation, particularly methylation of previously unmethylated CpG islands (45). These observations suggest the role of other *cis* or *trans*-acting factors mediating the *de novo* activity of the MTase *in vivo*. It was previously reported that a single 5mC on DNA could act as a signal for *de novo* methylation (21). It has been well-documented that in COS cells a patch of methylated cytosine can nucleate methylation spreading (46). A closer look at the catalytic turnover number for both the enzymes for either hemi-methylated or unmethylated duplexes shows that the turnover number of the longer MTase, with additional amino acids, is unaffected by the methylation status of the substrate. This suggests that the newly-identified N-terminal sequence (24,35) may have a regulatory function, resulting in the MMT3 enzyme being less proficient in maintenance methylation than the truncated MMT4 enzyme.

The poorer performance of the MMT3 MTase as a maintenance MTase might be interpreted as suggesting that *in vivo* it serves as both a maintenance and a *de novo* MTase. This *in vitro* observation might be sufficient to explain the observation that *in vivo*, MMT4 cannot complement cells in which the normal gene has been inactivated (24). Nevertheless, the finding that residual MTase activity can be detected in cells in which this MTase gene has been inactivated (23) means that it cannot be the sole *de novo* MTase.

## ACKNOWLEDGEMENTS

The authors thank W.Jack for discussions and help with the steady-state kinetic analysis, T.Bestor for the murine MTase cDNA clone (pMG) and K.Tucker for the 5′-MTase cDNA clone, p7B. This research was supported by grants from the National Institutes of Health (GM46127 to R.J.R.; CA44339 to R.J.; GM52106 to E.L.). D.T. was supported by the Human Frontier Science Program.

## REFERENCES

1 Schwartz, M.N., Trautner, T.A. and Kornberg, A. (1962) *J. Biol. Chem.*, **237**, 1961–1967.
2 Bestor, T., Hellewell, S.B. and Ingram, V.M. (1984) *Mol. Cell. Biol.*, **4**, 1800–1806.
3 Monk, M. (1990) *Phil. Trans. R. Soc. Lond.*, **B 326**, 299–312.
4 Graessmann, M. and Graessmann, A. (1993) in Jost, J.P. and Saluz, H.P. (eds), *DNA Methylation: Molecular Biology and Biological Significance*. Birkhäuser Verlag, Switzerland, pp. 404–425.
5 Gartler, S.M. and Riggs, A.D. (1983) *Annu. Rev. Genet.*, **17**, 155–190.
6 Migeon, B.R. (1994) *Trends Genet.*, **10**, 230–235.
7 Fedoroff, N.V. (1989) *Cell*, **77**, 473–476.
8 Engler, P., Weng, A. and Storb, U. (1993) *Mol. Cell. Biol.*, **13**, 571–577.
9 Rein, T., Zorbas, H. and DePamphilis, M. (1997) *Mol. Cell. Biol.*, **17**, 416–426.
10 Barlow, D.P. (1995) *Science*, **270**, 1610–1613.
11 Li, E., Bestor, T. and Jaenisch, R. (1992) *Cell*, **69**, 915–926
12 Jones, P.A. and Buckley, J.D. (1990) *Adv. Cancer Res.*, **54**, 1–23.
13 Cooper, D.N. and Youssoufian, H. (1988) *Human Genet.*, **78**, 151–155.
14 Sved, J. and Bird, A.P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4692–4696.
15 Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.*, **262**, 4778–4786.
16 Chen, L., MacMillan, A.M. and Verdine, G.L. (1993) *J. Am. Chem. Soc.*, **115**, 5318–5319.
17 Klimasauskas, S., Kumar, S., Roberts, R. J. and Cheng, X. (1994) *Cell*, **76**, 357–369.
18 Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) *Cell*, **82,** 143–153.
19 Kumar, S.K., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. and Wilson, G.G. (1994) *Nucleic Acids Res.*, **22**, 1–10.
20 Leonhardt, H., Page, A.W., Weier, H.U. and Bestor, T.H. (1992) *Cell*, **71**, 865–873.
21 Christman, J., Shikhnejad, G., Marasco, C.J. and Sufrin, J.R. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7347–7351.
22 Smith, S.S., Kaplan, B.E., Sowers, L.C. and Newman, E.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4744–4748.
23 Lei, H., Oh, S., Okano, M., Jutterman, R., Goss, K.A., Jaenisch, R. and Li, E. (1996) *Development*, **122**, 3195–3205.
24 Tucker, K.L., Talbot, D., Lee, M.A., Leonhardt, H. and Jaenisch, R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12920–12925.
25 Bestor, T.H. Laudano, A., Mattaliano, R. and Ingram, V. (1988) *J. Mol. Biol.*, **203**, 971–983.
26 Czank, A., Hauselmann, R., Page, A.W., Leonhardt, H., Bestor, T.H., Schaffner, W. and Hergersberg, M. (1991) *Gene*, **109**, 259–263.
27 O'Reilly, D., Miller, L.K. and Lucknow, V.A. (1992) in *Baculovirus Expression Vectors: A Laborarory Manual*. W. H. Freeman and Co., New York. pp. 216–234.
28 Glickman, J.F. and Reich, N.O. (1994) *Biochem. Biophy. Res. Comm.*, **204**, 1003–1008.
29 Tollefsbol, T.O. and Hutchison, C.A. (1995) *J. Biol. Chem.*, **270**, 18543–18550.
30 Kumar, S., Cheng, X., Pflugrath, J.W. and Roberts, R.J. (1992) *Biochemistry*, **31**, 8648–8653.
31 Brown, R.S. and Lennon, J.J. (1995) *Anal. Chem.*, **167**, 1998–2003.
32 Matsudaira, P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.
33 Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J. and Benner, J.S. (1991) *J. Bacteriol.*, **173**, 5207–5219.
34 Looney, M.C., Moran, L. S., Jack, W.E., Feehery, G.R., Benner, J.S., Slatko, B. E. and Wilson, G.G. (1989) *Gene*, **80**, 193–208.
35 Yoder, J.A. Yen, R.W.C., Vertino, P.M., Bestor, T.H. and Baylin, S.B. (1996) *J. Biol. Chem.*, **271**, 31092–31097.
36 DePaoli-Roach, A., Roach, P.J., Zucker, K.E. and Smith, S.S. (1986) *FEBS Lett.*, **197**, 149–153.
37 Bestor, T.H. and Ingram, V.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2674–2678.
38 Xu, G., Flynn, J., Glickman, F. and Reich, N.O. (1995) *Biochem. Biophy. Res. Comm.*, **207**, 544–551.
39 Bestor, T.H. (1992) *EMBO J.*, **11**, 2611–2617.
40 Flynn, J., Glickman, J.F., and Reich, N. O. (1996) *Biochemistry*, **35**, 7308–7315.
41 Bolden, A., Ward, C., Siedlecki, J.A. and Weissbach, A. (1984) *J. Biol. Chem.*, **259**, 12437–12443.
42 Smith, S.S., Kan, J.L.C., Baker, D.J., Kaplan, B.E. and Dembek, P. (1991) *J. Mol. Biol.*, **17**, 39–51.
43 Smith, S.S., Lingeman, R.G. and Kaplan, B.E. (1992) *Biochemistry*, **31**, 850–854.
44 Glickman, J.F., Flynn, J. and Reich, N.O. (1997) *Biochem. Biophys. Res. Comm.*, **230**, 280–284
45 Vertino, P.M., Yen, R.W., Gao, J. and Baylin, S.B. (1996) *Mol. Cell. Biol.*, **16**, 4555–4565.
46 Kass, S., Goddard, J.P. and Adams, R.L.P. (1993) *Mol. Cell. Biol.*, **13**, 7372–7379.