

# Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation

(DNA methylation/nucleophilic attack/transition mutations/hydrolytic deamination)

STEVEN S. SMITH\*, BRUCE E. KAPLAN, LAWRENCE C. SOWERS, AND EDWARD M. NEWMAN

City of Hope National Medical Center and Beckman Research Institute, 1500 East Duarte Road, Duarte, CA 91010

Communicated by Robert L. Sinsheimer, February 18, 1992 (received for review December 27, 1991)

**ABSTRACT** The properties of the methyl-directed DNA (cytosine-5-)-methyltransferase (EC 2.1.1.37) suggest that it is the enzyme that maintains patterns of methylation in the human genome. Proposals for the enzyme's mechanism of action suggest that 5-methyldeoxycytidine is produced from deoxycytidine via a dihydrocytosine intermediate. We have used an oligodeoxynucleotide containing 5-fluorodeoxycytidine as a suicide substrate to capture the enzyme and the dihydrocytosine intermediate. Gel retardation experiments demonstrate the formation of the expected covalent complex between duplex DNA containing 5-fluorodeoxycytidine and the human enzyme. Formation of the complex was dependent upon the presence of the methyl donor *S*-adenosylmethionine, suggesting that it comprises an enzyme-linked 5-substituted dihydrocytosine moiety in DNA. Dihydrocytosine derivatives are extremely labile toward hydrolytic deamination in aqueous solution. Because C-to-T transition mutations are especially prevalent at CG sites in human DNA, we have used high-performance liquid chromatography to search for thymidine that might be generated by hydrolysis during the methyl transfer reaction. Despite the potential for deamination inherent in the formation of the intermediate, the methyltransferase did not produce detectable amounts of thymidine. The data suggest that the ability of the human methyltransferase to preserve genetic information when copying a methylation pattern (i.e., its fidelity) is comparable to the ability of a mammalian DNA polymerase to preserve genetic information when copying a DNA sequence. Thus the high frequency of C-to-T transitions at CG sites in human DNA does not appear to be due to the normal enzymatic maintenance of methylation patterns.

Patterns of DNA methylation are somatically inherited and tissue specific. Proposals for the maintenance of these patterns (1, 2) have postulated the existence of a methyl-directed enzyme that would recognize asymmetrically methylated CG dinucleotide pairs generated by DNA replication and rapidly convert them back to symmetrically methylated pairs (Fig. 1), thus copying the parental pattern of methylation into the DNA of each daughter cell. Work with the murine enzyme (3) has shown that it responds to DNA in which all cytosines on one strand have been replaced with 5-methylcytosine by actively methylating the cytosines in CG dinucleotides on the opposite strand. Subsequent work with the human enzyme (4, 5) demonstrated that the enzyme had a clear specificity for the CG dinucleotide pair and that it responded to either of the two forms of an asymmetrically methylated CG dinucleotide pair by sensing the resident methyl group and rapidly converting the symmetrically placed cytosine to 5-methylcytosine. These properties clearly suggest that this enzyme activity is responsible for the maintenance of patterns of methylation in human cells according to the cycle depicted in Fig. 1.

Proposals for the mechanism of action of DNA methyltransferases (6) suggest production of a dihydrocytosine intermediate (Fig. 2). These derivatives are extremely susceptible to hydrolytic deamination, since saturation of the 5-6 bond increases the  $pK_a$  of the ring significantly (7). This increase will facilitate protonation of dihydrocytosine and thereby increase the rate of subsequent hydrolytic deamination in a form of the addition-elimination mechanism (8, 9) depicted in Fig. 2A. Although the expected enzyme-DNA intermediate has been detected with a prokaryotic methyltransferase (10), prior to the experiments described here, direct evidence for its formation by a eukaryotic methyltransferase had not, to our knowledge, been reported. Even so, we have pointed out that the production of this nonplanar intermediate in DNA would be consistent with the observed preference of the human enzyme for unusual DNA structures (11).

Since the putative intermediate (Fig. 2) might also be susceptible to deamination during catalysis (12), the enzymatic maintenance of methylation patterns could contribute to the high frequency of C-to-T transitions at the CG dinucleotide pair that are thought to be the result of deamination of methylcytosine to thymidine (13–15). Since the C-to-T mutation frequency observed *in vivo* at 5-methylcytosine relative to that at C (14, 15) is significantly greater than expected on the basis of the deamination rates observed in solution for 5-methylcytosine and C in DNA (16), we have asked two related questions: first, is the dihydrocytosine intermediate in fact formed by the human enzyme? and second, does significant deamination occur as a result of the enzymatic formation of the intermediate?

## MATERIALS AND METHODS

**DNA Preparation.** Procedures used for the chemical synthesis of foldback oligodeoxynucleotides from  $\beta$ -cyanoethylphosphoramidite precursors were as previously described (17). Procedures for  $^{32}P$ -end-labeling were also as previously described (18). As noted in ref. 17, uneven foldback molecules are self-priming substrates for DNA polymerase I and could be extended to full length in a reaction mixture containing approximately 5.0  $\mu$ g of DNA (0.64  $\mu$ M 47mer), 250  $\mu$ M dATP, 250  $\mu$ M dCTP, 250  $\mu$ M dGTP, 250  $\mu$ M dTTP, 100 mM Tris-HCl at pH 7.5, 10 mM  $MgCl_2$ , 80 mM NaCl, 10 mM 2-mercaptoethanol, and 10 units of the Klenow fragment of DNA polymerase I (BRL) in a final volume of 500  $\mu$ l. Incubation was for 30 min at 37°C. 5-Fluorodeoxycytidine (5FdC) was introduced into the foldback 65mer by substituting 5-fluorodeoxycytidine 5'-triphosphate (5FdCTP) for dCTP in the DNA polymerization reaction.

Abbreviations: AdoMet, *S*-adenosylmethionine; AdoHcy, *S*-adenosylhomocysteine; 5FC, 5-fluorocytosine; 5FdC, 5-fluorodeoxycytidine; 5FdCTP, 5-fluorodeoxycytidine 5'-triphosphate; nt, nucleotides.

\*To whom reprint requests should be addressed.

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

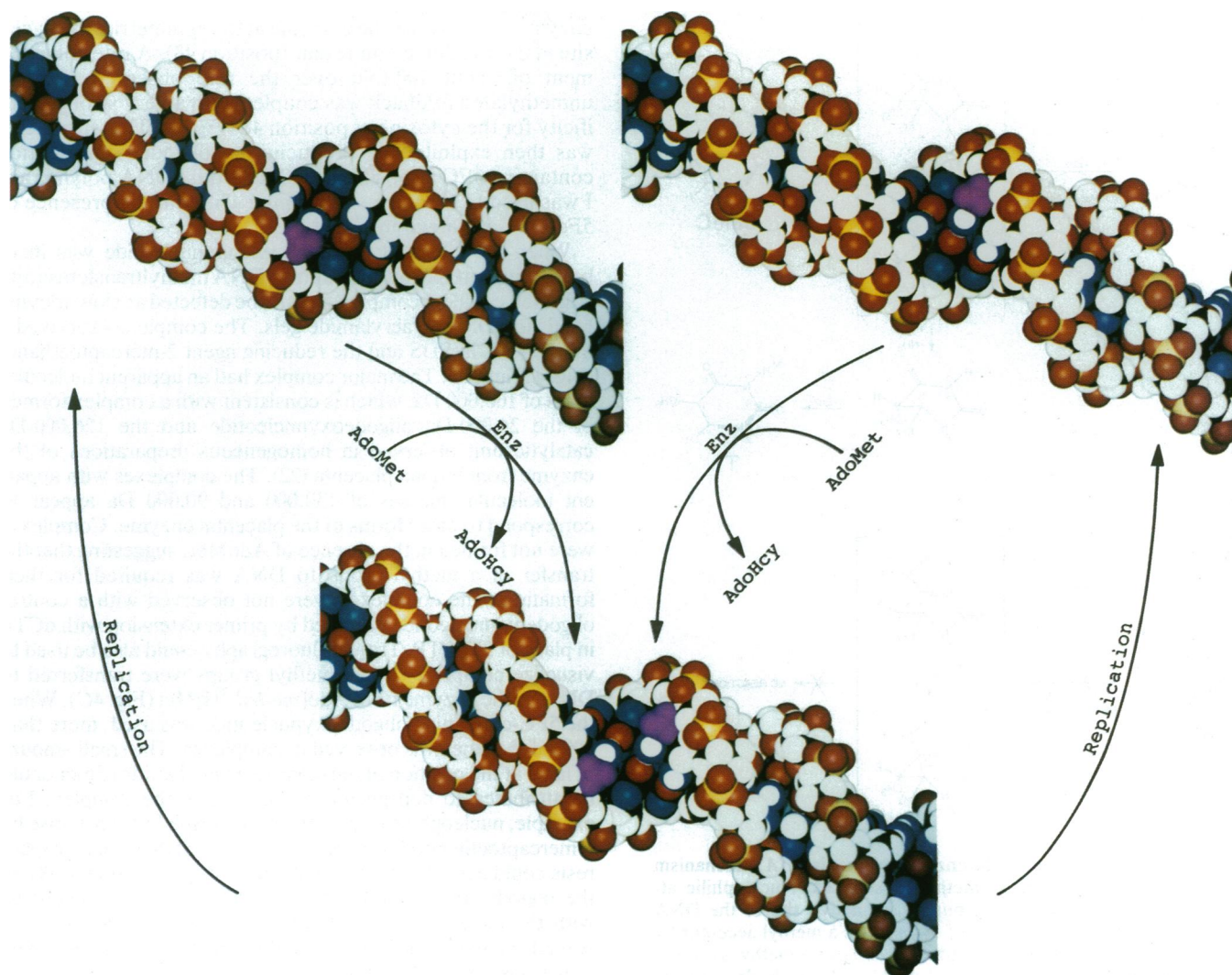


FIG. 1. Methyl-directed DNA (cytosine-5-)-methyltransferase reaction (EC 2.1.1.37). Space-filling models of the two possible asymmetrically methylated B-DNA molecules and the symmetrically methylated B-DNA molecule depict the reaction. Methyl groups on 5-methylcytosine residues at the CG dinucleotide pair are shown in purple. AdoMet, *S*-adenosylmethionine; AdoHcy, *S*-adenosylhomocysteine; Enz, enzyme.

5FdCTP was synthesized from 5FdC by using the method of Ruth and Cheng (19). The identity of the purified 5FdCTP was verified with mass spectroscopy as previously described (20).

Extension of the oligodeoxynucleotide to full length [65 nucleotides (nt)] was demonstrated by gel electrophoresis on 20% polyacrylamide sequencing gels, using a chemically synthesized foldback 65mer as a standard. The product 65mer was purified by HPLC on a Gen-Pak anion-exchange column (Waters). The presence of 5FdC, 5-methyldeoxycytidine, and other deoxynucleosides in appropriate stoichiometry was demonstrated by HPLC analysis of the nucleosides produced after digestion of the 65mer with nuclease P1 and bacterial alkaline phosphatase by the method of Palmgren *et al.* (21).

**Electrophoretic Separation of Enzyme-DNA Complexes.** Preparation of partially purified DNA (cytosine-5-)-methyltransferase from human placentas was as previously described (18, 22). The methyltransferase reaction was carried out in a mixture containing 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Hepes) at pH 7.0, 50 mM NaCl, 2 mM dithiothreitol, 75  $\mu$ M spermine, 10% (vol/vol) glycerol, and 6.0  $\mu$ M Ado[methyl- $^3$ H]Met. Reaction times were as indicated in the text.

The reaction product was concentrated by precipitation with trichloroacetic acid and resuspended in buffer containing 0.74 M 2-mercaptoethanol and 4% (wt/vol) SDS. After being heated to 95°C for 5 min, the sample was separated by

electrophoresis through a 6–20% polyacrylamide gradient in a 1.5-mm-thick minigel (Tall-Mighty-Small; Hoefer) as described by the manufacturer.

$^{32}$ P was visualized by direct autoradiography of the wet gel.  $^3$ H was visualized by fluorescence-enhanced autoradiography after impregnating the gel with EN $^3$ HANCE as previously described for DNA sequencing gels (18).

**Molecular Modeling.** Molecular models of duplex 30mers having the sequence of the complementary duplexes studied here were constructed in BIOGRAF (BioDesign, Pasadena, CA). Energy minimizations were performed by using the Dreiding parameter set. Methyl groups were introduced at the 5 position of cytosine with the C—C bond length constrained to that previously determined in crystalline Z-DNA (23). As with thymine methyls, cytosine methyls are not expected to perturb B-DNA conformation, since both pyrimidine methyls occupy roughly the same position in the major groove of the B-DNA helix (24, 25). This expectation was consistent with the molecular modeling carried out here, in which the energy-minimized structure was found to be roughly isomorphous to the B-DNA conformation adopted by the structure prior to minimization. Models were rendered in CPK (Corey–Pauling–Koltun) format for photography. Photographs, taken directly from the cathode ray tube of an Evans and Southerland (Salt Lake City) PS-340 graphics system, depict the region of the oligodeoxynucleotide in the vicinity of the site of methylation in oligodeoxynucleotides.

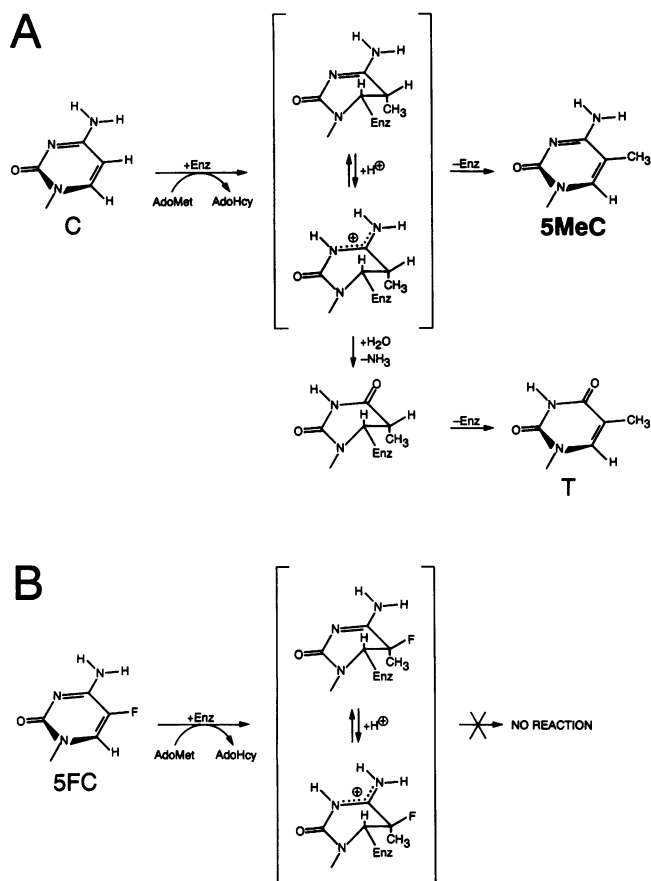


FIG. 2. Consequences of the enzyme mechanism. (A) Mechanism of human DNA (cytosine-5-)methyltransferases: Nucleophilic attack at C-6 of cytosine by a group at the active site of the DNA methyltransferase activates C-5 of cytosine as a methyl acceptor by saturating the 5-6 double bond. AdoMet donates a methyl group to C-5 of cytosine to form AdoHcy and the bracketed dihydrocytosine intermediate, which is unable to stack in duplex DNA because it is both nonplanar and nonaromatic. Elimination of the enzyme from the complex results in the regeneration of the 5-6 double bond to produce 5-methylcytosine. In the alternative pathway shown below the brackets, the protonated form of the dihydrocytosine intermediate undergoes hydrolytic deamination to generate dihydrothymine (bottom center). Elimination of the enzyme from this intermediate results in the regeneration of the 5-6 double bond to form thymine. (B) Stable enzyme-DNA complex with 5-fluorocytosine (5FC): C-5 of 5FC, like that of cytosine, can be activated as a methyl acceptor through nucleophilic attack by the DNA methyltransferase at C-6. However, transfer of the methyl group from AdoMet results in a stable covalent complex comprising DNA and the enzyme, because neither the C—F nor the C—CH<sub>3</sub> bond at C-5 can be broken to regenerate the 5-6 double bond in the cytosine ring.

## RESULTS AND DISCUSSION

As noted in Fig. 2A, methylated pyrimidines are generated from the dihydropyrimidine intermediates by elimination of the hydrogen atom at C-5 and the enzyme moiety at C-6 to regenerate the 5-6 double bond. This should not be possible when the pyrimidine being attacked is 5FC, because the 5-fluorine cannot be lost by elimination. For this reason nucleophilic attack and methyl-group transfer at a 5FC moiety will produce a stable complex (6) comprising DNA and the enzyme (Fig. 2B).

To detect the putative intermediate, we began by studying methylation of the asymmetrically methylated foldback depicted in Fig. 3. Restriction analysis of the enzymatically tritiated product by previously established methods (5) showed that the 5-methylcytosine at position 17 directed the

enzyme to methylate the cytosine at the symmetrically placed site in the CG dinucleotide pair (position 48). A rate enhancement of about 100-fold over the rate observed with an unmethylated foldback was coupled with a very strong specificity for the cytosine at position 48 (Fig. 3). This specificity was then exploited by producing an oligodeoxynucleotide containing 5FC at position 48. To do this, DNA polymerase I was used to extend a self-priming 47mer in the presence of 5FdCTP, as shown in Fig. 4A.

When the <sup>32</sup>P-end-labeled oligodeoxynucleotide was incubated with partially purified human DNA methyltransferase and AdoMet, covalent complexes could be detected as slow-moving bands in SDS/polyacrylamide gels. The complexes survived 5 min at 95°C in SDS and the reducing agent 2-mercaptoethanol (Fig. 4B and C). The major complex had an apparent molecular mass of 160,000 Da, which is consistent with a complex formed of the 24,000-Da oligodeoxynucleotide and the 126,000-Da catalytic unit observed in homogeneous preparations of the enzyme from human placenta (22). The complexes with apparent molecular masses of 130,000 and 90,000 Da appear to correspond to other forms of the placental enzyme. Complexes were not formed in the absence of AdoMet, suggesting that the transfer of a methyl group to DNA was required for their formation. The complexes were not observed with a control oligodeoxynucleotide prepared by primer extension with dCTP in place of 5FdCTP. Tritium fluorography could also be used to visualize complexes when methyl groups were transferred to DNA by the enzyme from Ado[methyl-<sup>3</sup>H]Met (Fig. 4C). When the 5FC-containing oligodeoxynucleotide was used, more than 80% of the label was observed in complexes. The small amount of label at the position of the oligodeoxynucleotide 65mer could be attributed to nonspecific breakdown of the complex. For example, nucleophilic displacement of the methyltransferase by 2-mercaptoethanol during preparation for SDS/gel electrophoresis could account for the small amount of breakdown. When the oligodeoxynucleotide 65mer lacking 5FC was incubated with the enzyme and Ado[methyl-<sup>3</sup>H]Met, no label was observed in complexes. The bulk of the tritium label was observed at the position of the 65mer.

Incorporation into trichloroacetic acid-insoluble material when 5FC was present in the oligodeoxynucleotide was about 1/40th of that when it was absent, suggesting that the reaction with the 5FC-containing oligodeoxynucleotide was stoichiometric, while that with the oligodeoxynucleotide lacking 5FC was catalytic. The tritium results also permit an estimation of the efficiency of complex formation as follows. We observed about 9800 cpm/pmol of tritiated methyl groups when AdoMet was spotted directly onto a glass fiber filter and its radioactivity was measured in scintillation fluid. The specific activity of the homogeneous enzyme is about 40,000 units/mg; its molecular mass is about 126,000 Da. When 5.7 units of enzyme (143 ng) was incubated with an excess of 5FC-containing DNA, 2200 cpm of <sup>3</sup>H was incorporated into trichloroacetic acid-insoluble material after 20 min of incubation. If we assume one mole of substrate bound per mole of enzyme, then [2200 cpm/(9800 cpm/pmol)] (126,000 pg/pmol) = 28.3 ng of enzyme is contained in the complex. This suggests that about 20% of the available DNA methyltransferase (i.e., 28.3 ng/143 ng) was converted into stable complexes in 20 min when the 5FC-containing oligodeoxynucleotide was used.

On the basis of the data described above, we conclude that the enzyme operates by the mechanism shown in Fig. 2. Several of the consequences of the formation of this intermediate by eukaryotic methyltransferases have been discussed previously (5, 6, 11, 12), but the potential for hydrolytic deamination during catalysis has not yet been explored. In light of this potential, we investigated the possibility that a low level of deamination might occur during enzymatic methylation. After enzymatic methylation, a tritium-labeled oligodeoxynucleotide was digested to deoxynucleosides and the digest was analyzed by HPLC. A

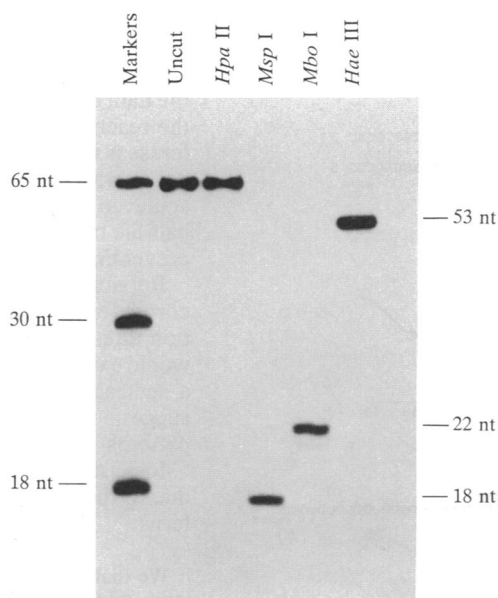
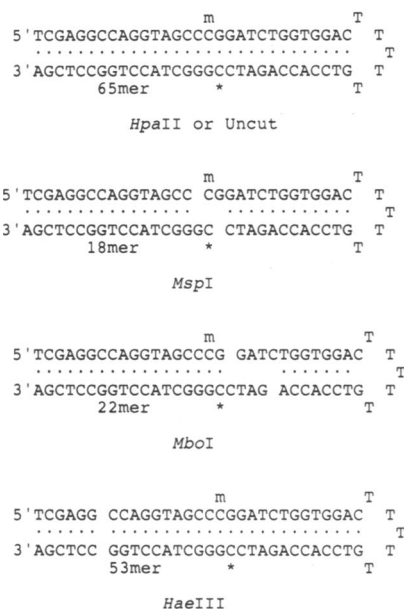
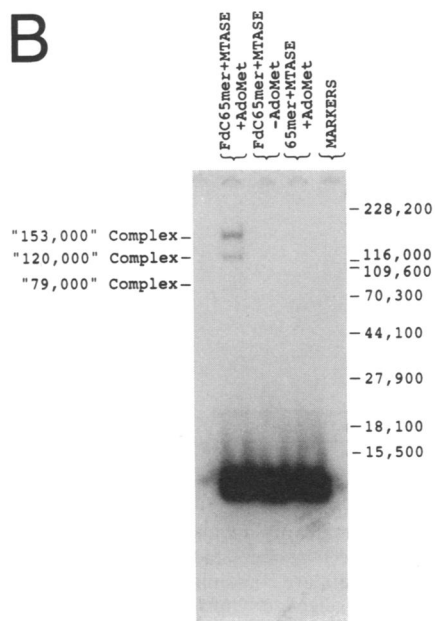


FIG. 3. Restriction analysis of the enzymatically labeled foldback 65mer. (Left) Labeling pattern predicted for the product of the methyl-directed reaction. When the continuous chain is numbered from the 5' end, the 5-methylcytosine introduced synthetically, indicated with an m, is at position 17; the 5-methylcytosine generated enzymatically, indicated with an asterisk, is expected to be at position 48. Cleavage sites for the indicated restriction enzymes are depicted as gaps. (Right) Autoradiograph showing sizes of the labeled restriction fragments observed after enzymatic labeling for 4 hr with Ado[methyl-<sup>3</sup>H]Met. This labeling pattern clearly indicates that the cytosine at position 48 is in fact the target of the enzymatic reaction.

A



B



C

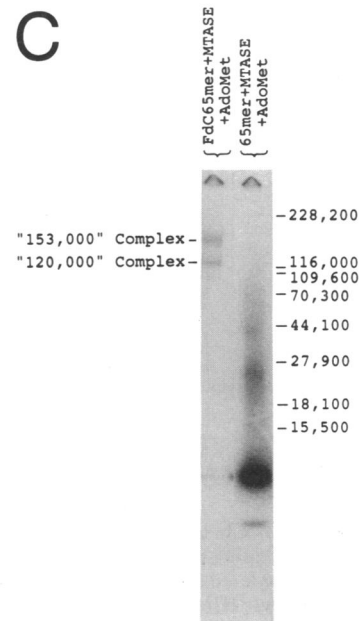


FIG. 4. Mechanism-based labeling of the human DNA (cytosine-5)-methyltransferase. (A) As starting material, a 47mer was constructed so as to link a long block of DNA to a shorter complementary block of DNA through a tether consisting of five thymidine residues. A 5-methylcytosine residue (indicated with m) was introduced at position 17 during synthesis. Molecules of this type form unimolecular foldbacks and are self-priming substrates for DNA polymerase I (17). To generate an oligodeoxynucleotide in which position 48 was occupied by 5FdC, 5FdCTP was used in place of dCTP during extension with the Klenow fragment of DNA polymerase I. This method places 5FdC at residues (indicated by F) between positions 48 and 65 that are normally occupied by dC. The 5FdC residue at position 48 is indicated with an asterisk. (B) Primer extension products were end-labeled with <sup>32</sup>P and incubated for 4 hr with a partially purified preparation of DNA methyltransferase (MTASE) from human placenta (5) under the conditions given in the text. The products were separated by electrophoresis through SDS-containing polyacrylamide gels. <sup>32</sup>P was visualized by autoradiography. From the left: lane 1, 5FdC-containing 65mer + complete reaction mixture; lane 2, 5FdC-containing 65mer + reaction mixture lacking AdoMet; lane 3, dC-containing 65mer + complete reaction mixture; lane 4, markers (in Da). (C) Unlabeled primer extension products were incubated with the methyltransferase preparation and separated by electrophoresis through SDS/polyacrylamide gels as in B except that the tritiated methyl groups applied to the foldback were visualized by fluorography after the gels were impregnated with EN<sup>3</sup>HANCE (5). Lane 1, complete reaction mixture + 5FdC-containing 65mer; lane 2, complete reaction mixture + dC-containing 65mer.

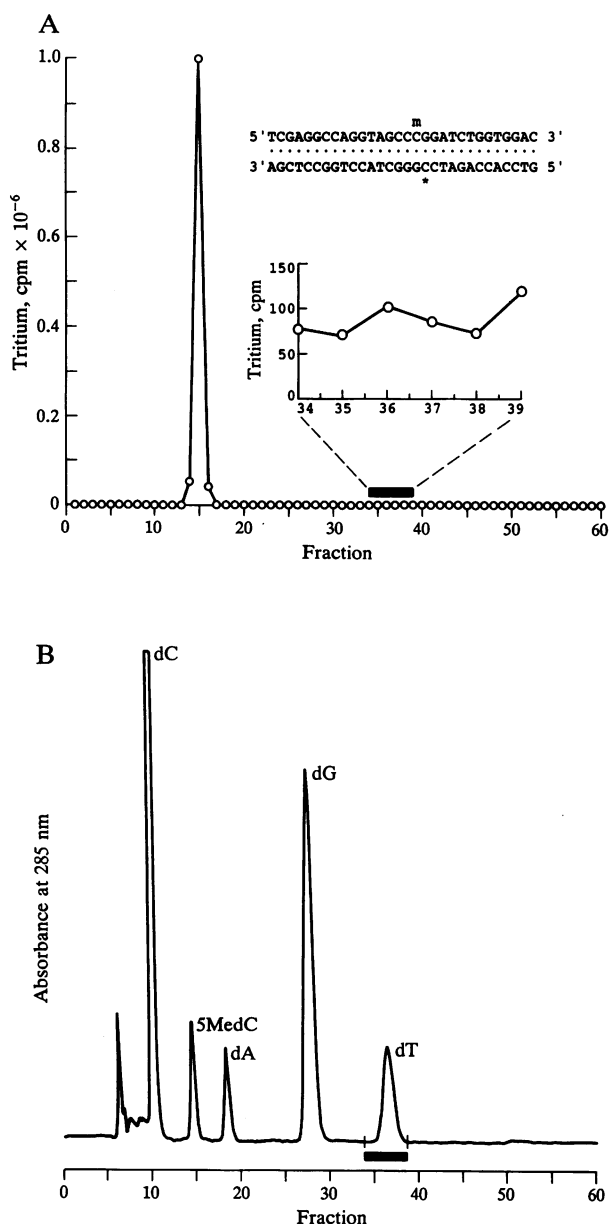


FIG. 5. Chromatographic analysis of enzymatically labeled DNA. A duplex oligodeoxynucleotide 30mer (Inset) was enzymatically methylated for 1 hr in the complete reaction mixture containing Ado[methyl-<sup>3</sup>H]Met. After digestion of the labeled oligodeoxynucleotide with nuclease P1 and bacterial alkaline phosphatase, the liberated nucleosides were separated by HPLC on a  $\mu$ Bondapak C<sub>18</sub> column (Waters). Fractions were collected and tritium was quantified by scintillation counting. (A) Tritium profile. (B) Absorbance profile.

comparison of the tritium profile with the absorbance profile in Fig. 5 shows that tritium-labeled methyl groups were detected in 5-methyldeoxycytidine but not in thymidine. The average cpm for the 5-methylcytidine peak was about 980,000, while the average background in the region of the thymidine peak was about 93 cpm. If we assume a detection limit of 50% above background, then less than 1 thymidine would be produced for every 21,000 cytosine residues methylated ( $980,000/47 \approx 21,000$ ). This means that the *in vitro* fidelity (i.e., the ability of the enzyme to accurately preserve genetic information) observed with the human DNA methyltransferase in copying a

methylation pattern would be comparable to the fidelity (1/2930 to 1/47,500) observed for mammalian polymerases in copying a DNA sequence (26). Since we have no reason to suspect from the data that thymidine residues were actually produced during the reaction, it is possible that the fidelity of the methyltransferase is in fact considerably better than that of the mammalian polymerases. This high degree of fidelity could have been achieved by the evolution of rapid turnover to minimize the half-life of the intermediate and by the evolution of a means of preventing exposure of the intermediate to water.

Replication fidelity is improved by proofreading systems that correct errors introduced by DNA polymerases, so that mutation rates can be several orders of magnitude lower than one would expect from *in vitro* measurements (26). The existence of a specialized G/T repair system in mammalian cells (27–29) suggests that this is also the case for the somatic replication of patterns of DNA methylation. Thus the high frequency of C-to-T transitions at CG sites in human DNA is not likely to be due to the normal enzymatic maintenance of methylation patterns.

We thank David Baker for his help with these experiments. This work was supported primarily by Grants GM-38350 from the National Institutes of Health and 1571B from The Council for Tobacco Research U.S.A., Inc., with additional support from Grants GM-41336 and CA-33572, also from the National Institutes of Health.

- Holliday, R. & Pugh, J. E. (1975) *Science* **187**, 226–232.
- Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14**, 9–25.
- Gruenenbaum, Y., Cedar, H. & Razin, A. (1982) *Nature (London)* **295**, 620–622.
- Bolden, A. H., Nalin, C. M., Ward, C. A., Poonian, M. S. & Weissbach, A. (1986) *Mol. Cell. Biol.* **6**, 1135–1140.
- Smith, S. S., Kan, J. L. C., Baker, D. J., Kaplan, B. E. & Dembek, P. (1991) *J. Mol. Biol.* **217**, 39–51.
- Santi, D. V., Garrett, C. E. & Barr, P. J. (1983) *Cell* **33**, 9–10.
- Kochetkov, N. K. & Budovskii, E. I. (1972) *Organic Chemistry of Nucleic Acids* (Plenum, New York), pp. 159–306.
- Shapiro, R. & Klein, R. S. (1966) *Biochemistry* **5**, 2358–2362.
- Sowers, L. C., Sedwick, W. D. & Shaw, B. R. (1989) *Mutat. Res.* **215**, 131–138.
- Osterman, D. G., DePillis, G. D., Wu, J. C., Matsuda, A. & Santi, D. V. (1988) *Biochemistry* **27**, 5204–5210.
- Baker, D. J., Kan, J. L. C. & Smith, S. S. (1988) *Gene* **74**, 207–210.
- Selker, E. U. (1990) *Annu. Rev. Genet.* **24**, 579–613.
- Salser, W. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 985–1002.
- Cooper, D. N. & Krawczak, M. (1990) *Hum. Genet.* **85**, 55–74.
- Rideout, W. M., III, Coetzee, G. A., Olumi, A. F. & Jones, P. A. (1990) *Science* **249**, 1288–1290.
- Ehrlich, M., Zhang, X.-Y. & Inamdar, N. M. (1990) *Mutat. Res.* **238**, 277–286.
- Smith, S. S., Lingeman, R. G. & Kaplan, B. E. (1992) *Biochemistry* **31**, 850–854.
- Smith, S. S., Hardy, T. A. & Baker, D. J. (1987) *Nucleic Acids Res.* **15**, 6899–6916.
- Ruth, J. L. & Cheng, Y.-C. (1981) *Mol. Pharmacol.* **20**, 415–422.
- Hardy, T. A., Baker, D. J., Newman, E. M., Sowers, L. C., Goodman, M. F. & Smith, S. S. (1987) *Biochem. Biophys. Res. Commun.* **145**, 146–152.
- Palmgren, G., Mattsson, O. & Okkels, F. T. (1990) *Biochim. Biophys. Acta* **1049**, 293–297.
- Zucker, K. E., Riggs, A. D. & Smith, S. S. (1985) *J. Cell. Biochem.* **29**, 337–349.
- Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H. & Rich, A. (1982) *Nucleic Acids Res.* **10**, 7879–7892.
- Ivarie, R. (1987) *Nucleic Acids Res.* **15**, 9975–9983.
- Farrance, I. K. & Ivarie, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1045–1049.
- Loeb, L. A. & Kunkel, T. A. (1982) *Annu. Rev. Biochem.* **52**, 429–457.
- Jiricny, J., Hughes, M., Corman, N. & Rudkin, B. B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8860–8864.
- Wiebauer, K. & Jiricny, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5842–5845.
- Wiebauer, K. & Jiricny, J. (1989) *Nature (London)* **339**, 234–236.