Interaction of Oligonucleotides Containing 6-*O*-Methylguanine with Human DNA (Cytosine-5-)-methyltransferase†

Ngee-Wah Tan* and Benjamin F. L. Li

Institute of Molecular and Cell Biology, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore

Received March 6, 1990; Revised Manuscript Received June 21, 1990

ABSTRACT: Thirty-base-pair synthetic oligonucleotide duplexes containing a single meG·C (meG = 6-*O*-methylguanine) or A·C base pair at the 16th position (i.e., 5′-CCCGTTTAAATATACXTATACCCGGGTACC-3′, where *X* = A or meG) were used to study de novo methylation by the purified human DNA (cytosine-5)-methyltransferase isolated from CEM cells. Both duplexes containing meG·C and A·C base pairs show enhanced methyl group acceptor properties. Subsequent introduction of hemimethylated sites at the 15th position of the top strand (the C residue next to the abnormal base pair) and the 7th, 15th (which represents the C residue in the 6meG·C and A·C base pairs), and 27th positions of the bottom strand were used to study the maintenance methylation of the hemimethylated duplexes by the methylase. This revealed striking differences in the rate, amount, and sites of methylation, which are dependent on the position of the hemimethylated site in the duplex. The possible mechanism of action of the methylase is discussed. The data show that 6-*O*-methylguanine residues in DNA can have other genetic effects apart from their miscoding behavior and that meG·C and A·C base pairs exert different effects in terms of methylation.

Although 6-*O*-alkylguanine (i.e., formed by alkylating agents with DNA) is known to produce point mutations in vitro (Abbott & Saffhill, 1979) and in vivo (Mitra et al., 1989), it remains to be established whether this lesion in DNA can exert other genetic effects, such as perturbation of the endogenous DNA methylation process (i.e., the formation of 5-methylcytosine). This may exist for the following reasons: (1) 6-*O*-Alkylguanine has a reasonably long biological half-life (in cells that have been treated with alkylating agents), so that the lesion may persist through cell cycles (Pegg, 1984). (2) Apart from its miscoding behavior (Abbott & Saffhill, 1979), it perturbs the regular structure of DNA (Li & Swann, 1989; Kalnik et al., 1989a,b). (3) It may be formed in CpG doublets, where 5-methylcytosine (meC) residues are often found in mammalian DNA. (4) 6-Methylguanine (meG) can also be formed from nonenzymatic reaction of *S*-adenosylmethionine (SAM) with DNA in vitro (Rydberg & Lindahl, 1982) and in vivo when animals are treated with the hepatotoxin hydrazine (Barrows et al., 1983). SAM is a cofactor for the methylase.

For the above reasons, we were particularly interested to determine whether DNA containing this lesion can have an effect on the methylation of cytosine residues by the human DNA (cytosine-5)-methyltransferase. The methylation of DNA is known to have important biological consequences (Razin et al., 1984; Adam & Burdon, 1985; Keshet et al., 1985). Thirty years ago, Chargaff (1955) pointed out the possible importance of 5-methylcytosine because of its nonrandom distribution in the mammalian genome. Since then, numerous experiments have shown that this modified base plays a role in gene suppression by a mechanism that is not yet established. Recently, Holliday (1989) has suggested that this base may play a part in the "epigenetic" mechanism of cell differentiation, aging, and carcinogenesis. This hypothesis still remains to be established. In the past, the study of DNA methylation has been hampered by the practical problem of the precise location of the 5-methylcytosine residues in DNA, but this problem has now been overcome (Saluz & Jost, 1989). Hopefully, more definitive experimental data will emerge to prove this hypothesis.

The following are important experimental observations that prompted our study: (1) Wilson and Jones (1983) reported that chemical carcinogens (i.e., nitrogen mustard, ethylnitrosourea, 1,3-bis(2-chloroethyl)-1-nitrosourea, etc.) interact with DNA and the resulting damaged DNA is a poor substrate for the methylase as compared to the original DNA. (2) Smith et al. (1987a–c, 1988, 1989) reported the unusual methylation of DNA containing mismatches and G_4 structure (a complex composed of four parallel strands of guanine-rich DNA) by the methylase.

All these experiments indicate that the structure of DNA is crucial for the function of the methylase, which is reponsible for the maintenance of methylation patterns in genomic DNA. We are intrigued by these observations and infer that DNA containing 6-alkylguanine residues may have an effect on the methylase. However, we also include an A·C base pair in our study for the following reasons: (1) 6-*O*-Methylguanine is a substrate for adenosine deaminase (Pegg & Swann, 1979). (2) NMR (nuclear magnetic resonance) study of dodecanucleotides containing meG·C and A·C base pairs shows that there is similarity between these two base pairs (Kalnik et al., 1989a,b). (3) 6-*O*-Alkylguanine is highly lipophilic (Li & Swann, 1989). Point 1 indicates that 6-*O*-alkylguanine may resemble adenine in vivo. In this respect, it is important to know whether they are the same in terms of methylation by the methylase. However, the high lipophilicity of 6-*O*-alkylguanine distinguishes it from adenine. It is not clear whether this highly lipophilic residue in DNA may have some unspecific interaction with DNA-binding proteins, such as inhibition of the methylase processing along the DNA. We hope that this comparison will therefore give us some insight into the problem. We also use the G·T base pair in one of the experiments to show that thymine which can be formed from

† This research is funded by the National University of Singapore.
* Address correspondence to this author.

0006-2960/90/0429-9234$02.50/0

deamination of 5-methylcytosine during deprotection of synthetic oligonucleotides at high pH and temperature will not cause unusual methylation.

Materials and Methods

(*1*) *Chemicals and Enzymes.* Poly(dI·dC) (Boehringer Mannheim), [^{3}H]SAM (Amersham; 15 Ci/mmol), SAM (New England Biolabs; 32 mM), T4 DNA ligase (Amersham), T4 kinase (New England Biolabs), snake venom alkaline phosphatase (Boehringer Mannheim), snake venom phosphodiesterase (Pharmacia), DE-52 (Whatman), and hydroxyapatite (Bio-Rad) were used according to the manufacturer-supplied protocols. N^4-Benzyl-5-methyl-2′-deoxycytosine phosphoramide was obtained from Pharmacia.

(*2*) *Oligonucleotides.* The following oligonucleotides were used in our experiments: (1) 5′-TATAC*meG*TATA, (2) 5′-TATA*meCmeG*TATA, (3) 5′-CCCGTTTAAA, (4) 5′-CCCGGGTACC, (5) 5′-CCCGTTTAAATATAC*G*TATACCCGGGTACC (CT), (6) 5′-CCCGTTTAAATATA*CA*TATACCCGGGTACC (ACT), (7) 5′-CCCGTTTAAATATA*meCA*TATACCCGGGTACC (15meCT), (8) 5′-CCCGTTTAAATATAC*meG*TATACCCGGGTACC (16meGT), (9) 5′-CCCGTTTAAATATA*meCmeG*TATACCCGGGTACC (15meC16meGT), (10) 5′-CCCGTTTAAAATAT*meCA*TATACCCGGGTACC (15meCAT), (11) 5′-GGTACCCGGGTATAC*G*TATATTTAAACGGG (CB), (12) 5′-GGTAC*meC*GGGTATA*CG*TATATTTAAACGGG (7meCB), (13) 5′-GGTACCCGGGTATA*meCG*TATATTTAAACGGG (15meCB), (14) 5′-GGTACCCGGGTATAC*G*TATATTTAAA*meCG*GG (27meCB), and (15) 5′-GGTACCCGGGTATA*TG*TATATTTAAACGGG (GTCB). The abbreviations used are as follows: meC, 5-methylcytosine; meG, 6-methylguanine; CT, control top strand; ACT, A·C mismatch top strand; 15meCT, meC at the 15th position of the top strand; 16meGT, meG at the 16th position of the top strand; CB, control bottom strand; 7meCB, meC at the 7th position of the bottom strand.

All the normal and 5-methylcytosine-containing oligonucleotides were synthesized on a Milligene 7500 DNA synthesizer (Millipore) with cyanoethyl phosphoramidite blocks. Oligomers containing meG (i.e., oligomers 1 and 2) were synthesized by the phosphotriester approach in solution (Li & Swann, 1989). Oligomers 8 and 9 were synthesized by ligation of oligomers 3, 5′-phosphorylated 1 (or 2), and 4, with oligomer 11 used as template. The ligated products were purified by chromatography on a strong anion exchanger (MonoQ, Pharmacia) at pH 12.

(*3*) *Isolation of DNA (Cytosine-5-)-methyltransferase.* The methylase was purified from a nuclear extract of CCRF-CEM cells (human peripheral blood acute lymphoblastic leukemia cells). The cells were cultured in roller bottles at a high population density in RPMI (47.5%), MEM (47.5%), and fetal bovin serum (5%). Cells were harvested at a cell density of (2–3) × 10^6 cells/mL.

(*A*) *Nuclear Extract Preparation.* All purification steps were carried out at 0–4 °C. The nuclear extract was prepared following the procedure of Wildemann et al. (1984). Six batches of cell culture (3.5 L) were harvested and processed independently. The final purified protein fractions were then combined as the main stock.

(*B*) *DEAE-cellulose Chromatography.* The dialyzed nuclear extract (~20 mL) was batch-adsorbed onto DE-52 cellulose (20 mL wet volume, preequilibrated with the starting buffer: 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, and 20% glycerol). The tube containing the matrix and the extract (in a 50-mL screw-cap tube) was then put in a roller for 30 min. The tube was then spun at 1000*g* and the supernatant was collected.

(*C*) *Hydroxyapatite Chromatography.* The supernatant from the DEAE-cellulose column (~30 mL) was stirred for 30 min with 0.5 g of preswollen hydroxyapatite (in starting buffer plus 20 mM K_2HPO_4). The gel was then poured into a column (1 × 5 cm) and the column was washed with two column volumes of starting buffer. The column was then eluted with a gradient of phosphate (20–500 mM K_2HPO_4 in starting buffer, flow rate = 0.5 mL/min, total elution volume = 54 mL). The methylase was eluted as a single activity peak at about 0.15 M phosphate. The combined active fractions were then used directly in our experiment. The specific activity of the fraction is 1496 units/mg [defined as the transfer of 1 pmol of [^{3}H]CH_3 from SAM to 1 μg of poly(dI·dC) under standard conditions]. This preparation shows a major band of ~190 kD on an SDS gel and methylates with a slight preference for properly annealed rather than heat-denatured poly(dI·dC). No detectable 6-methylguanine alkyltransferase activity was observed by using the assay procedure described by Graves et al. (1987).

(*4*) *DNA (Cytosine-5-)-methyltransferase Assay.* The standard reaction mixture [200 μL, in a buffer of 20 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.25 mM DDT, 1.5 μM (≃2 μCi) [^{3}H]SAM, and 10 μg of heat-denatured poly(dI·dC) (Pedrali-Noy & Weissback, 1986)] was incubated at 37 °C. Aliquots (45 μL) were withdrawn from the reaction mixture at various periods of time. The reaction was terminated by addition of proteinase K and SDS to final concentrations of 0.1 mg/mL and 0.1% (w/v), respectively. After further incubation for 30 min at 60 °C, the samples were chilled on ice, and sodium pyrophosphate (0.2 mL, 0.2 M, pH 6.5) and carrier DNA (0.6 mg, 100 μL) were added. The DNA was then precipitated with ice-cold perchloric acid (0.2 mL, 5 N). After centrifugation (Eppendorf, 1000 rpm) for 10 min, the pellet was redissolved in NaOH (0.2 mL, 0.5 M) and incubated for 10 min at 60 °C to remove RNA. The DNA was finally recovered by precipitation at 0 °C with cold TCA (trichloroacetic acid; 15%, 2 mL) and filtered onto GFC filters (Whatman). The filter was then washed five times with TCA (5%, 5 mL), twice with EtOH (95%, 5 mL), and once with absolute EtOH (to dehydrate the filters). The filters were then put into scintillation vials containing Filtron X (5 mL, National Diagnostic) for tritium counting.

(*5*) *Analysis of Oligonucleotides.* (*A*) *Base Analysis (Quantification of DNA).* Oligonucleotide (~0.1 A_{260}) in water (50 μL) was added to the digestion buffer (100 μL; 50 mM Tris-HCl, pH 8.3, and 5 mM $MgCl_2$). The solution was then heated at 90 °C for 3 min and cooled in ice immediately. The solution was then left for equilibration at 37 °C for 10 min. Phosphodiesterase (20 μL) was added for 90 min, followed by alkaline phosphatase (10 μL) for 25 min. The samples were then heated at 80 °C for 3 min. The solution was then analyzed by reverse-phase chromatography, with conditions as reported (Li & Swann, 1989). The A_{260nm} peak areas of the formed nucleosides were quantified with reference to the standard solution. The concentration of DNA was then quantified by averaging the concentrations of A, C, G, and T.

(*B*) *Strand Separation.* The top and bottom strands of the 30mers studied (irrespective of whether the 16th position is substituted by A or meG or whether meC is present) can be separated by strong anion-exchange chromatography (MonoQ, Pharmacia) with a gradient of NaCl: 20% buffer B for 8 min,

20–40% buffer B over 3 min, and then 40–55% buffer B over 15 min at a flow rate of 1 mL/min (buffer A, 0.01 M NaOH; buffer B, 0.01 M NaOH and 1.5 M NaCl). Retention times for the top and bottom strands are 17 and 20 min, respectively. For investigation of the ^3H labeling, the eluent from the MonoQ column was passed directly to a radiometric detector (Flo-One/β Radiometric Instruments, 1000-μL cell with Flo-scint IV, 3 mL/min). The radioactive peaks observed for the separated strands were quantified by integration of the radioactive peak area.

(*6*) *Typical Kinetic Experiments.* The experiments that involved different 30mers were carried out essentially as described in the assay procedure but with an additional step of annealing. Annealing buffer (10×, 2 μL: 1 M NaCl, 100 mM Tris-HCl, pH 7.8, and 10 mM EDTA) was added to a known concentration of DNA (i.e., 40 pmol each of bottom and top strands in 20 μL of water). The mixture was then warmed to 70 °C for 10 min in a water bath. The water bath was then switched off and left to cool to room temperature (~4 h). Nondenaturing gel electrophoresis experiments (data not shown) showed that under this condition all these sequences, i.e., those containing A·C or meG·C base pairs, form proper duplexes and no unusual large-sized fragments were observed. [^3H]SAM (2 μL), protein extract (20 μL), assay buffer (20 μL, 10×), glycerol (40 μL, 50%) and water (added to give a total volume of 200 μL) were then added to the reaction mixture. Aliquots (45 μL) were withdrawn at various times and heat-denatured at 80 °C for 3 min. The samples were then analyzed by TCA precipitation as described above. For chromatography, the heat-denatured aliquoted samples were diluted to 400 μL with NaOH (0.01 M) before injection onto the MonoQ column. Comparison experiments were done under identical conditions.

(*7*) *Chemical Sequencing.* Oligonucleotide 11 (i.e., 40 pmol of the control bottom strand) was 5′-phosphorylated with [$^{32}\gamma$]ATP according to the blunt-end labeling protocol, with an additional step of chase by cold ATP (1 μL, 1 mM solution) for 5 min, as described by Maniatis et al. (1989). After two cycles of phenol–chloroform extraction, the oligomer was desalted on a NAP5 column (Pharmacia). The desalted solution was then dried down and annealed to the top strand (i.e., oligomer 6) as described in the assay procedure above. After incubation for 1 h, the reaction mixture was extracted twice with phenol–chloroform and desalted on a NAP5 column. The control experiment was done exactly as the experimental except no methylase was added. The desalted reaction mixture was then aliquoted to five Eppendorf tubes for chemical sequencing. The sequencing was carried out essentially as described by Maxam and Gilbert (1980) with a few modifications: ethanol precipitation was replaced by adsorption of the cleavage solution onto C18 silica, the G reaction used dimethyl sulfate in 1 M KH_2PO_4 at pH 3.7, the C reaction used hydroxylamine at pH 6.0, and the T reaction used $KMnO_4$ in 1 M KH_2PO_4 at pH 3.7. Details of this sequencing protocol and the high rate of cleavage of the cytosine in this abnormal base pair will be published shortly.

Results and Discussion

TCA precipitation has been widely used as a method to assay DNA (cytosine-5-)-methyltransferase. The activity of the enzyme is estimated by counting the amount of radioactive methyl groups incorporated into the TCA-precipitated DNA. This method gives the overall radioactive methyl group incorporation but not the sites or the strands of methylation. Although chemical sequencing is the best way to identify the sites of methylation, it is difficult to visualize in these in vitro

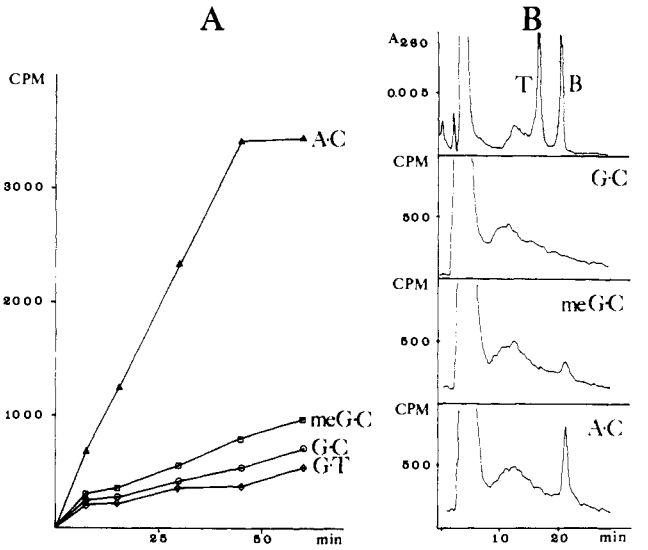

FIGURE 1: Methylation of nonmethylated oligonucleotides (de novo methylation). (A) Kinetics of methylation, measured after TCA

5′-CCCGTTTAAATATACXTATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCATGG-5′

precipitation. (B) Product analysis by chromatography (MonoQ): the top panel shows the optical trace, and the remaining panels show the on-line tritium traces. X represents A, G, or meG. The designations meG·C, A·C, G·T, and G·C represent duplexes containing meG·C, A·C, G·T, and G·C base pairs at the 16th position; the first residue of the base pair is on the top strand and the second residue is on the bottom strand.

experiments, where the incorporation of the methyl group is low. Many workers in this field have tried to overcome the problem by introducing restriction enzyme sites into the substrates. As the substrates can be digested to smaller fragments, it is possible to locate the incorporated radioactivity. However, this tends to limit the choice of substrates available for investigation. In the experiments reported here, we analyze our reaction mixtures by both TCA precipitation and liquid chromatography on line with a radioactivity detector. Using this approach, we are able to estimate both the overall radioactive methyl group incorporation into the oligonucleotides (TCA precipitation) and into which strand it is incorporated (liquid chromatography). This information is essential for our interpretation.

(*1*) *Kinetics of Methylation of Nonmethylated DNA (de Novo Methylation).* Figure 1A summarizes the kinetics of methylation of the 30mers in the duplex state. Separate experiments on the single-stranded oligonucleotides (with or without 5-methylcytosine or 6-*O*-methylguanine) show no apparent difference in the methylation as compared to the double strands (data not shown). It is clear that these oligonucleotides are not good substrates for the methylase (see Figure 2A). However, both duplexes containing A·C and meG·C base pairs show enhanced methyl group incorporation. Bolden et al. (1986) reported that high CG content is required for de novo methylation; the 30mers studied here have only 47% CG content. However, the rate of methylation of this sequence is similar to that of the sequence (66% CG content) studied by Smith et al. (1987c) (data not shown). The following sequences have been studied with the human enzyme: 5′-CCGGCCATT*ACGGATCCGT*CCTGGGC (70% CG; Bolden et al., 1986), 5′-GTCCACCAGATCCGGGCTACCTGGCCTCGA (66% CG; Smith et al., 1987c), and 5′-CCCGTTTAAA*TATACGTATA*CCCGGGTACC (47% CG; this paper). The rates of methylation of all these se-

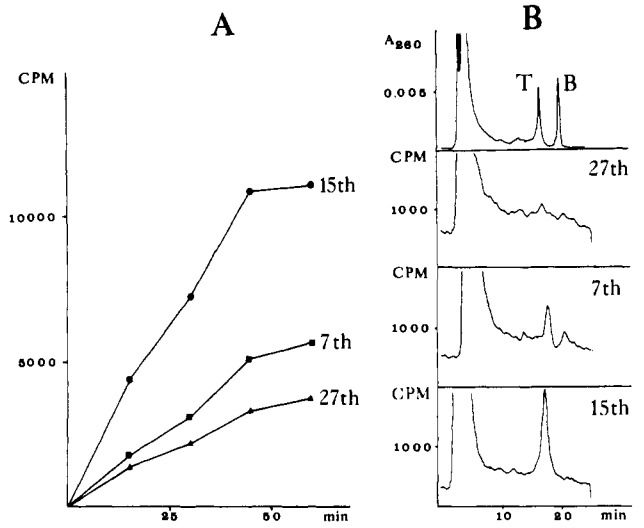

FIGURE 2: Comparison of the efficiency of methylation of hemimethylated sites at different positions in the 30mer duplex (maintenance methylation). (A) Kinetics of methylation, measured after

5'-CCCGTTTAAATATACGTATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCATGG-5'
* * *

TCA precipitation. (B) Product analysis by chromatography (MonoQ): the top panel shows the optical trace, and the remaining panels show the on-line tritium traces. The designations 7th, 15th, and 27th represent the position of the 5-methylcytosine residue, *C, from the 5' end of the bottom strand. The slightly lower radioactivity observed in this experiment is due to the 10-fold higher concentration of DNA used.

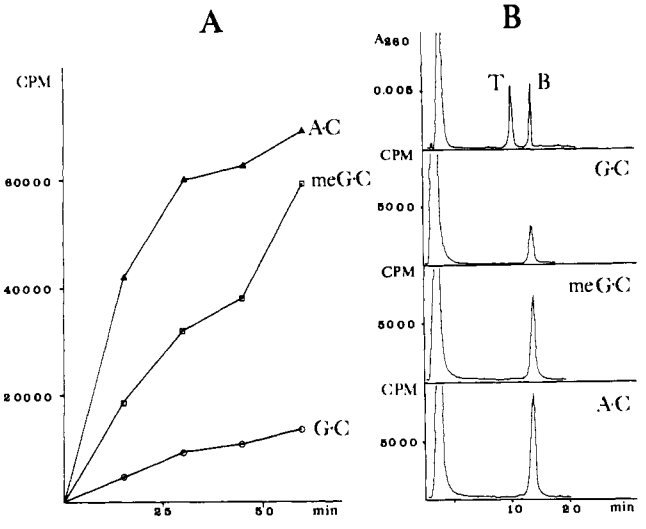

FIGURE 3: Methylation of hemimethylated site 15 (i.e., the 5-methylcytosine residue, *C, is at the 15th position of the top strand next to the G·C, A·C, or meG·C base pair). (A) Kinetics of me-

5'-CCCGTTTAAATATACXTATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCATGG-5'
*

thylation, measured after TCA precipitation. (B) Product analysis by chromatography (MonoQ): the top panel shows the optical trace, and the remaining panels show the on-line tritium traces.

quences are comparable (on the basis of the reported conditions). From these data, it seems that palindromic sequences are also good substrates for de novo methylation. However, there appears to be a lack of autocatalysis when studying de novo methylation. One would expect that when one methyl group is incorporated into the DNA, hemimethylated sites will be formed. These hemimethylated sites should subsequently enhance the incorporation of the second methyl group into the DNA. As a result, both strands should be methylated. Pedrali-Noy and Weissbach (1986) suggested that the lack (or slower rate) or methylation of nonmethylated DNA could possibly be due to substrate inhibition, but this was only observed for poly(dI·dC) (which is known to be the best substrate for de novo methylation). Alternatively, de novo methylation might be only on the non-CpG sites (Hubrich-Kühner et al., 1989), which could explain the lack of autocatalysis. Surprisingly, both the A·C and meG·C base pair containing duplexes show enhanced methylation, and all the radioactivity is exclusively on the bottom strand (see Figure 1B). The data clearly show that the hemimethylated site, formed from the de novo methylation of the bottom strand, could not be further fully methylated by the methylase. As the incorporation of [^{3}H]methyl groups was low, it is not possible to investigate further the sites of methylation by chemical sequencing. However, we believe that one could be the C residue of the A·C and meG·C base pairs (see later discussion). The presence of a G·T base pair in the duplex has no effect on de novo methylation, as previously reported by Smith et al. (1987b).

(2) *Kinetics of Methylation of Hemimethylated Oligonucleotides (Maintenance).* (*A*) *Site Efficiency.* We have investigated the efficiency of methylation of different hemimethylated sites along the DNA duplex. Figure 2A summarizes the kinetics of methylation of duplexes containing the different hemimethylated sites; in all cases the meC residues are on the bottom strand but at different positions. It is clear that site 15 (i.e., in the middle of the helix) is a much better substrate than the 7th and 27th sites. From the product analysis (Figure 2B), it is likely that both sites 7 and 27 also cause incorporation of counts to the bottom strands (above noise level) as compared to the 15th site. We cannot explain this observation. Since the 7th and 27th sites are close to the ends of the duplex, this may limit the contact of the methylase with the DNA substrate, therefore lowering the efficiency of methylation. Nevertheless, the 7th and 27th hemimethylated sites are much better methyl acceptors than the nonmethylated control duplex (compare with Figure 1A).

(*B*) *Effect of A·C and meG·C on Methylation of Hemimethylated Sites.* In these experiments, we are particularly interested in looking at the effect of the A·C and meG·C base pairs either close to (15th site) or some distance away from (7th and 27th sites) the methylated sites.

(*i*) *Neighboring Effect (15th Site).* Figures 3 and 5 summarize two separate experiments; first, when the hemimethylated site (i.e., the 5-methylcytosine residue) is on the top strand next to the A·C or 6meG·C base pair (this may resemble the in vivo situation, where the 6-*O*-methylguanine residue is formed in a fully methylated site and the DNA is replicated) and second, when the hemimethylated site is on the bottom strand (the C residue of the A·C and meG·C base pairs). Dramatic differences were observed in the two cases. In the first case, when the hemimethylated site was on the top strand next to the A·C or meG·C base pair, a dramatic increase in the methyl acceptor property of the duplex was observed. Both the rate and the amount of ^{3}H incorporated were substantially higher than in the control hemimethylated duplex (Figure 3A). Analysis of ^{3}H incorporation by chromatography of the reaction mixture after 1 h shows that the ^{3}H counts are exclusively on the bottom strand (see Figure 3B). Because of the high incorporation of the [^{3}H]methyl group into the bottom strand, it is possible to use chemical sequencing (Kossel et al., 1987) to locate the 5-methylcytosine residue formed in the bottom strand. The 5-methylcytosine residue was identified exclusively on the cytosine residue in the meG·C and A·C base pairs (see Figure 4). In the second case, when the bottom

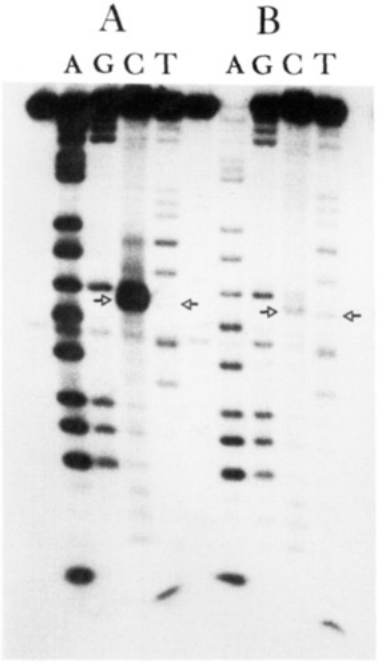

FIGURE 4: Chemical sequencing gel of the duplex containing a 5-methylcytosine residue, *C, at the 15th position of the top strand next to the A·C base pair. This gel shows the cleavage pattern of the

5′-CCCGTTTAAATATACATATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCTAGG-^{32}P-5′

labeled bottom strand (see Figure 3B). (A) Before (control experiment). (B) After incubation with the methylases and SAM for 1 h (arrows indicate the disappearance of the C band on the C ladder and the formation of the T band on the T ladder after the enzymatic methylation).

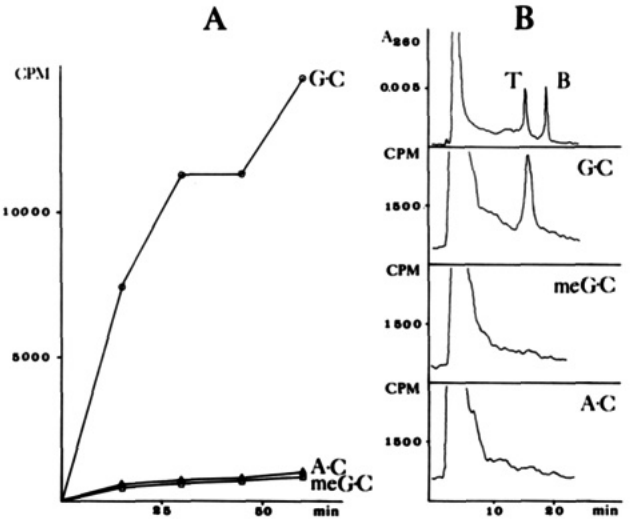

FIGURE 5: Methylation of hemimethylated site 15 (i.e., the 5-methylcytosine residue, *C, is at the 15th position of the bottom strand, replacing the C residue of the G·C, A·C and meG·C base pairs). (A)

5′-CCCGTTTAAATATACXTATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCATGG-5′

Kinetics of methylation, measured after TCA precipitation. (B) Product analysis by chromatography (MonoQ): the top panel shows the optical trace, and the remaining panels show the on-line tritium traces.

strand serves as the hemimethylated site, poor ^{3}H incorporation into the duplexes containing A·C and meG·C base pairs was observed (Figure 5A and 5B). This result is unexpected. Unfortunately, unlike the mouse methylase (Bestor et al., 1988; Spiess et al., 1988) and other prokaryotic methylases (Wu & Santi, 1987), the primary structure of the human DNA (cytosine-5)-methyltransferase is still unknown. However, Posfai et al. (1989) made an extensive computer analysis of the protein sequences of all the DNA (cytosine-5-)-methyltransferases known so far. They showed the existence of five highly conserved regions among all these prokaryotic and eukaryotic methylases. The mouse methylase may serve as a model for the human enzyme because the comparative molecular weight and high specificity toward hemimethylated sites suggest they are homologous enzymes. On the basis of the structure obtained from electron microscopy, Spiess et al. (1988) proposed that the mouse methylase contains a "handle" region that is responsible for recognition and a "cup" region for the methylase activity. Because of the high specificity of the mammalian protein toward hemimethylated sites, it is possible that the handle region binds strongly to the 5-methylcytosine residue in the hemimethylated site (this may be due to the asymmetry of the site). This binding will inevitably expose the CpG doublet of the complementary strand toward the cup region, where it is methylated. This may partially explain our first observation. The second observation can also be explained in that when the 5-methylcytosine is on the bottom strand, the CpA and CpmeG doublets of the top strand become exposed to the cup region. The lack of methylation is simply because the CpA and CpmeG doublets are not the substrate of the cup region (i.e., it is specific for CpG doublets) where the methylase activity is located. This implies that the meG residue is not recognized as a guanine residue by the methylase. However, it is still difficult to explain the excessive methylation observed in the first experiment. In principle, the methylation in vivo is an all-or-none process. It is not clear whether these data are purely artifacts of in vitro experiments, in that the methylase is not acting on a long strand of DNA (i.e., in vivo) along which it processes (i.e., possibly it always attaches to the DNA). Bolden et al. (1986) reported the observation of product inhibition in the in vitro experiments; for example, poly(meC·G) is a good inhibitor for the human methylase. It could be possible that the fully methylated site that is formed, i.e., 5′-meCmeG-3′/3′-GmeC-5′ or 5′-meCA-3′/3′-GmeC-5′, is simply not an inhibitor for the methylase. This interpretation suggests that the presence of an A·meC or meG·meC base pair [resembling an A·C or meG·C base pair (Kalnik et al., 1988)] destabilizes the fully methylated site. This may effectively cause the enzyme to detach from the DNA. Perhaps the observed excessive methylation is therefore due to the loss of product inhibition. The support for this argument may come from the theoretical study by Housheer et al. (1989), who show, by computational analysis, that a localized, highly stable, and lipophilic structure can be formed in the site of cytosine methylation. This is entirely opposite to the fully methylated unstable mismatch site. However, we do not know the relevance of this observed excessive methylation to the in vivo system. It is likely that loss of methylation may occur in vivo if the 6-methylguanine residues are formed in the CpG doublets of the native newly synthesized daughter strand DNA opposite the parent hemimethylated sites during DNA replication.

(*ii*) *Distance Effect* (*7th and 27th Sites from the 5′ End of the Bottom Strand*). (*a*) *Site 7*. Site 7 is essentially a *Sma*I (5′-CCCGGG-3′) site and is characterized by its pure CG base pairs. The result is summarized in Figure 6. From the kinetic data in Figure 6A, it is clear that the presence of an A·C or meG·C base pair at the 15th position (i.e., eight base pairs away from the hemimethylated site) enhances [^{3}H]methyl group incorporation into the duplex. Product analysis (Figure

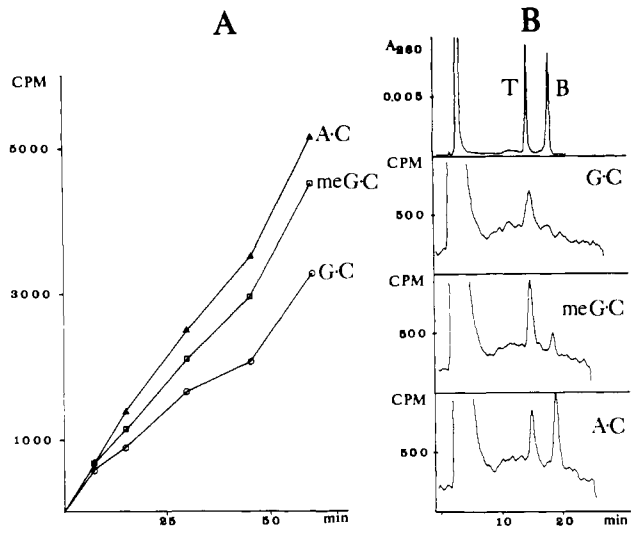

FIGURE 6: Methylation of hemimethylated site 7 (i.e., the 5-methylcytosine residue, *C, is at the 7th position from the 5′ end of the bottom strand; X = A, meG, or G). (A) Kinetics of methylation,

5′-CCCGTTTAAATATACXTATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCATCC-5′
*

measured after TCA precipitation. (B) Product analysis by chromatography (MonoQ): the top panel shows the optical trace, and the remaining panels show the on-line tritium traces.

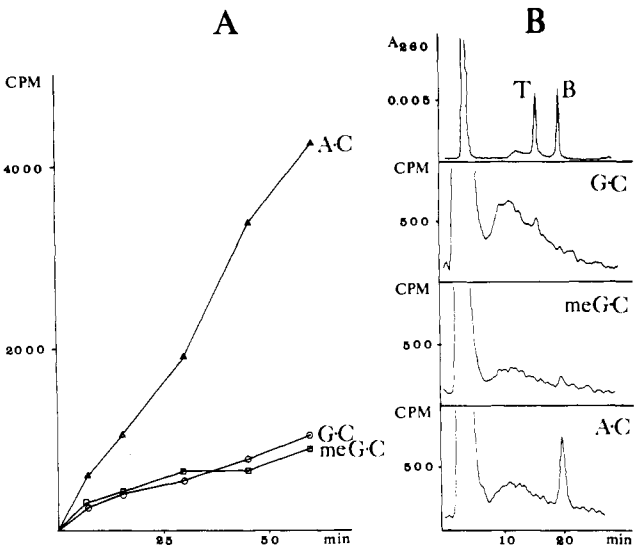

FIGURE 7: Methylation of hemimethylated site 27 (i.e., the 5-methylcytosine residue, *C, is at the 27th position from the 5′ end of the bottom strand; X = A, meG, or G). (A) Kinetics of meth-

5′-CCCGTTTAAATATACXTATACCCGGGTACC
GGGCAAATTTATATGCATATGGGCCCATGG-5′
*

ylation, measured after TCA precipitation. (B) Product analysis by chromatography (MonoQ): the top panel shows the optical trace, and the remaining panels show the on-line tritium traces.

6B) of the reaction mixture after a 1-h incubation clearly shows that the incorporation of the methyl group is "aberrant". Both A·C and meG·C cause extra methylation of the bottom strand: 60% of the total counts for the A·C mismatch and 20% of the total counts for the meG·C sequences. Only the meG·C sequence causes excessive methylation of the top strand. It is likely that the data observed in the A·C duplex is additive from two effects: it is the sum of de novo methylation (i.e., the bottom strand incorporation; see Figure 1) and maintenance methylation (i.e., the top strand incorporation due to the hemimethylated site). We are sure that the majority of incorporation on the top strand must be at the hemimethylated site because none of the nonmethylated sites are susceptible to methylation (see Figure 1). Particularly, the 15th C residue of the top strand cannot be methylated because it is not a CpG doublet (i.e., a CpA doublet is not a substrate for the cup region of the methylase; see Figure 5). The data suggest that the two strands are methylated independently. In the meG·C duplex, the incorporated counts on the bottom strand can be due to de novo methylation (see Figure 1). However, there is apparent enhancement of incorporation on the top strand. These data clearly show that the A·C and meG·C base pairs are different. We cannot explain this observation. Surprisingly, this aberrant methylation can exist when the hemimethylated site is eight base pairs away from the meG·C base pair.

(*b*) *Site 27.* This hemimethylated site is three base pairs away from the 5′-end. As shown in Figure 2A, it is the least effective methyl group acceptor. The kinetics of methylation and product analysis (after 1 h) are summarized in Figure 7. It is clear that the A·C mismatch enhances the methylation of the bottom strand as compared to the control. As this enhancement is comparable to that from de novo methylation (see Figure 1), we conclude that the methylation on the bottom strand is from de novo methylation. Although the incorporated count is low, the product analysis (Figure 7B) shows apparently that there are no observable counts on the top strand (i.e., the methylation of the hemimethylated site is inhibited) in the duplexes containing meG·C and A·C base pairs. Unfortunately, because of the poor incorporation of radioactivity at the hemimethylated site as compared to sites 15 and 7 (i.e., close to the 5′ end), the data are insufficient to show that the effect is significant. These data are important because this site is on the 5′ side (in contrast to the 7th site, which is on the 3′ side) of the abnormal base pairs. This indicates that the methylase may have a directional property. Obviously, study of a longer sequence is required. Nevertheless, it is interesting that A·C and meG·C base pairs can have such a distance effect (i.e., 11 base pairs away) on the hemimethylated site.

Conclusion

In these experiments, we observed the following: (1) 30mer duplexes containing A·C and meG·C base pairs at the 15th position enhance de novo methylation as compared to the control duplex. (2) When an A·C or meG·C base pair was next to a hemimethylated site, we observed (a) excessive methylation of the cytosine residue of A·C and meG·C base pairs of the bottom strand when the hemimethylated site is on the top strand (i.e., site 15 experiments) and (b) complete loss of the methyl acceptor property of the top strand in duplexes containing A·C or meG·C base pairs when the hemimethylated site is on the bottom strand (i.e., when the C residue of the A·C or meG·C base pair is replaced by 5-methylcytosine). (3) An meG·C base pair that is eight base pairs away from the 3′ hemimethylated site causes excessive methylation of the top strand (site 7 experiment). (4) Both A·C and meG·C base pairs that are 11 base pairs away from the 5′ hemimethylated site inhibit the methylation of the hemimethylated site (site 27 experiments showed that there was no methylation on the top strand).

The important conclusion from these experiments is that methylation sites can be destroyed or created as a consequence of abnormal base pairing in the helix. It is clear from these experiments that 6-*O*-methylguanine residues in DNA can exert additional effects other than miscoding behavior. This

may relate to the mechanism of chemical carcinogenesis and, perhaps, Holliday's hypothesis (Holliday, 1989) of the epigenetic mechanism in carcinogenesis. However, this requires in vivo experimental data. Obviously, further in vitro experiments need to address the following questions: (1) Can meG residues be formed in DNA containing 5-methylcytosine? (2) Can meG residues be repaired effectively in DNA containing 5-methylcytosine? (3) What DNA sequences are important, for example, in the 5′ regulatory region, where the presence of 5-methylcytosine residues is often inversely proportional to gene activity (Toth et al., 1989)?

ACKNOWLEDGMENTS

We thank Dr. Thomas Leung and Dr. Edward Manser for helpful discussions.

REFERENCES

Abbott, P. J., & Saffhill, R. (1979) *Biochim. Biophys. Acta 562*, 51–61.

Adams, R. L. P., & Burdon, R. H. (1985) *Molecular Biology of DNA Methylation*, Springer-Verlag, New York.

Barrows, L. R., Shank, R. C., & Magee, P. N. (1983) *Carcinogenesis 4*, 953–957.

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988) *J. Mol. Biol. 203*, 971–983.

Bolden, A., Ward, C. A., Nalin, C. M., & Weissbach, A. (1986) *Prog. Nucleic Acid Res. Mol. Biol. 33*, 231–250.

Chargaff, E. (1955) *The Nucleic Acids* (Chargaff, E., & Davidson, J. N., Eds.) Vol. pp 1, 307–371, Academic Press, New York.

Graves, R. J., Li, B. F. L., & Swann, P. F. (1987) *IARC Sci. Publ. 84*, 41–43.

Hausheer, F. H., Rao, S. N., Gamcsik, M. P., Kollman, P. A., Colvin, O. N., Saxe, J. D., Nelkin, B. D., McLennan, I. J., Barnett, G., & Baylin, S. B. (1989) *Carcinogenesis 10*, 1131–1137.

Holliday, R. (1989) A different kind of inheritance. *Sci. Am. 260* (June), 40–48.

Hubrich-Kühner, K., Buhk, H.-J., Wagner, H., Kröger, H., & Simon, D. (1989) *Biochem. Biophys. Res. Commun. 160*, 1175–1182.

Kalnik, M. W., Kouchakdjian, M., Patel, D. J., Li, B. F. L., & Swann, P. F. (1988) *Biochemistry 27*, 100–108.

Kalnik, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1989a) *Biochemistry 28*, 6170–6181.

Kalnik, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1989b) *Biochemistry 28*, 6182–6192.

Keshet, I., Yisraeli, J., & Cedar, H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2560–2564.

Kossel, H., Fritzche, E., Hayatsu, H., & Iida, G. L. (1987) *Nucleic Acids Res. 15*, 5517–5528.

Li, B. F. L., & Swann, F. P. (1989) *Biochemistry 28*, 5779–5786.

Maniatis, T., Sambrook, J., & Fritsch, E. F. (1989) Molecular *Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499.

Mitra, G., Pauly, G. T., Kumar, R., Pei, G. K., Hughes, S. H., Moschel, R. C., & Barbacid, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8650–8654.

Pedrali-Noy, G., & Weissbach, A. (1986) *J. Biol. Chem. 261*, 7600–7602.

Pegg, A. E. (1984) *Cancer Invest. 2*, 223–231.

Pegg, A. E., & Swann, P. F. (1979) *Biochim. Biophys. Acta 565*, 241–252.

Posfai, J., Bhagwat, A. S., Posfai, G., & Roberts, R. J. (1989) *Nucleic Acids Res. 17*, 2421–2435.

Razin, A., Cedar, H., & Riggs, A. (1984) *DNA Methylation: Biochemistry and Biological Significance*, Springer-Verlag, New York.

Rydberg, B., & Lindahl, T. (1982) *EMBO J. 1*, 211–216.

Saluz, H., & Jost, J.-P. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2602–2606.

Smith, S. S., Hardy, T. A., Baker, D. J., Newman, E. M., Sowers, L. C., & Goodman, M. F. (1987a) *Biochem. Biophys. Res. Commun. 145*, 146–152.

Smith, S. S., Baker, D. J. & Hardy, T. A. (1987b) *Biochem. Biophys. Res. Commun. 146*, 596–602.

Smith, S. S., Hardy, T. A., & Baker, D. J. (1987c) *Nucleic Acids Res. 15*, 6899–6916.

Smith, S. S., Baker, D. J., & Kan, J. L. C. (1988) *Gene 74*, 207–210.

Smith, S. S., Baker, D. J., & Jardines, L. A. (1989) *Biochem. Biophys. Res. Commun. 160*, 1397–1402.

Spiess, E., Tomassetti, A., Hernaiz-Driever, P., & Pfeifer, G. P. (1988) *Eur. J. Biochem. 177*, 29–34.

Toth, M., Lichtenberg, U., & Doerfler, W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3728–3732.

Wildeman, A., Sassone-Corsi, P., Grundström, T., Zenke, M., & Chambon, P. (1984) *EMBO J. 3*, 3129–3133.

Wilson, V. L., & Jones, P. A. (1983) *Cell 32*, 239–246.

Wu, J. C., & Santi, D. V. (1987) *J. Biol. Chem. 262*, 4778–4786.