Examination of the DNA substrate selectivity of DNA cytosine methyltransferases using mass tagging

Valaiporn Rusmintratip¹, Arthur D. Riggs² and Lawrence C. Sowers^{1,3,*}

¹Division of Molecular Medicine, ²Division of Biology and ³Division of Pediatrics, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, CA 91010, USA

Received May 16, 2000; Revised and Accepted July 25, 2000

ABSTRACT

The biological significance of cytosine methylation is as yet incompletely understood, but substantial and growing evidence strongly suggests that perturbation of methylation patterns, resulting from the infidelity of DNA cytosine methyltransferase, is an important component of the development of human cancer. We have developed a novel *in vitro* assay that allows us to quantitatively determine the DNA substrate preferences of cytosine methylases. This approach, which we call mass tagging, involves the labeling of target cytosine residues in synthetic DNA duplexes with stable isotopes, such as ¹⁵N. Methylation is then measured by the formation of 5-methylcytosine (5mC) by gas chromatography/mass spectrometry. The DNA substrate selectivity is determined from the mass spectrum of the product 5mC. With the non-symmetrical duplex DNA substrate examined in this study we find that the bacterial methyltransferase *HpaII* (duplex DNA recognition sequence CCGG) methylates the one methylatable cytosine of each strand similarly. Introduction of an A-C mispair at the methylation site shifts methylation exclusively to the mispaired cytosine residue. In direct competition assays with *HpaII* methylase we observe that the mispaired substrate is methylated more extensively than the fully complementary, normal substrate, although both have one *HpaII* methylation site. Through the use of this approach we will be able to learn more about the mechanisms by which methylation patterns can become altered.

INTRODUCTION

Cytosine methylation is a covalent modification of DNA generated enzymatically at the fifth carbon position of cytosine residues within CpG dinucleotides after DNA replication (1,2). Although 5-methylcytosine (5mC) makes up <1% of all nucleotides in the human genome (3), 5mC may play a critical role in tumorigenesis through its ability to influence gene expression (3,4).

The mechanisms by which methylation patterns are established are still unknown. Once established, however, the distribution

of 5mC residues may be heritably transmitted from parent to progeny cells. The physical basis for the transmission of methylation patterns is based upon the preferential methylation of CpG dinucleotides in which the cytosine residue on the parental strand is methylated. *In vitro* studies have demonstrated that the predominant methylase found in most eukaryotic cells, which is encoded by the DNMT1 gene and is considered to be the maintenance methylase, will methylate hemimethylated sites at 10–100 times the rate for unmethylated sites (5,6). If the methylase methylates all hemimethylated sites and no unmethylated sites, the methylation pattern will be transferred from parent to progeny cells with high fidelity *in vivo*. If the cytosine DNA methylase does not methylate all hemimethylated sites, methylates previously unmethylated sites or induces the deamination of cytosine or 5mC, then methylation patterns will be perturbed.

Factors that may promote the perturbation of methylation patterns in human cancer are not well understood. A correlation has been established between reduced methylation in promoter regions and the expression of transforming oncogenes (7). Other studies report a correlation between promoter region hypermethylation and silencing of tumor suppressor genes (8). Although the precise relationship between perturbation of cytosine methylation patterns and the development of human cancer is currently unresolved, there is a substantial body of studies that clearly indicates the significance of understanding this relationship (9–11).

Several techniques have been developed to measure total 5mC levels and to map the location of 5mC residues in genomic DNA (12,13). The most sensitive chemical methods currently available for analyzing the modification of DNA bases utilize mass spectrometry, including gas chromatography/mass spectrometry (GC/MS) (14,15). However, in general, chromatographic methods cannot be used to examine the methylation of a specific cytosine residue because the DNA must be converted to either the free base or deoxynucleoside prior to analysis, resulting in the loss of sequence-specific information.

We have developed a novel approach for studying the methylation of target cytosine residues which exploits the sensitivity of GC/MS but retains the sequence information following oligonucleotide hydrolysis. This approach requires the synthesis of oligonucleotides in which target cytosine residues are labeled with stable isotopes. In the experiments conducted below, specific cytosine residues are selectively enriched with ¹⁵N in the exocyclic amino group, rendering these residues one mass unit heavier. In the *in vitro* methylase

*To whom correspondence should be addressed. Tel: +1 626 359 8111; Fax: +1 626 301 8862; Email: lsowers@coh.org

5' gtccaccagatcCgggctacctggcctga Duplex 1
3' caggtggtctaggCccgatggaccggact

5' gtccaccagatccgggctacctggcctga Duplex 2
3' caggtggtctaggCccgatggaccggact

5' gtccaccagatccAggctacctggcctga Duplex 3
3' caggtggtctaggCccgatggaccggact

5' gtccaccagatccAggctacctggcctga Duplex 4
3' caggtggtctaggcccgatggaccggact

Figure 1. Sequences of the DNA duplexes used in these assays.

assay, if the mass-tagged cytosine residues in the duplexes are methylated, the corresponding 5mC will also be one mass unit heavier. Indeed, the entire ion abundance profile will be shifted one mass unit higher. The significance of this mass tagging approach is that the mass spectrum of the product 5mC can be used to determine if the target, isotopically enriched cytosine, or another cytosine was methylated, irrespective of the length or sequence of the DNA substrate.

MATERIALS AND METHODS

Oligonucleotide synthesis

Oligonucleotides (29mer, Fig. 1) were prepared by automated phosphoramidite synthesis and purified by reverse phase HPLC. The mass-tagged cytosine precursor, (4-¹⁵NH₂)-2'-deoxycytidine, was prepared using a method developed by Divakar and Reese (16) that was modified by this laboratory (17). Molar equivalents were estimated from absorbance at 260 nm. Complementary oligonucleotides were combined in equal amounts in 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂ and 1 mM DTT, pH 7.9. This mixture was heated to 95°C and slowly cooled to room temperature to generate duplexes (40 nM). To confirm complete duplex formation, ³²P-end-labeled duplexes were digested to completion with *Bst*NI, *Hpa*II and *Sau*3AI. The products were sized on non-denaturing 12% polyacrylamide gels. The solutions of duplexes were dried under reduced pressure.

Assays with cloned enzymes

Three cloned prokaryotic C5 methyltransferases (*Sss*I methylase, *Msp*I methylase and *Hpa*II methylase) and their respective reaction buffers were obtained from New England Biolabs. *Sss*I methylates all cytosine residues at the C5 position within

the double-stranded dinucleotide recognition sequence CG. *Msp*I recognizes CCGG and methylates the external cytosine residue at the C5 position. *Hpa*II also recognizes CCGG but methylates at the C5 position of the second cytosine residue. Enzymatic reactions with these commercially available methylases were carried out under the conditions described by the manufacturer. Duplexes (3.6 nmol) were incubated with methylase (50 U) with the corresponding buffers and 80 μM (160 μM for *Sss*I) *S*-adenosylmethionine at 37°C for 4 h. The reaction mixtures were then transferred to derivatization vials supplied by Pierce and dried. No extraction procedures were conducted.

GC/MS studies

The GC/MS studies were conducted with a Hewlett Packard 5970 GC interfaced with a 5890 mass selective detector. The 88% formic acid solution was obtained from Aldrich Chemical Co. The dried duplexes were hydrolyzed by treatment with 200 μl of 88% formic acid in Teflon-sealed 1 ml Hypo-vials supplied by Pierce. Formic acid was removed under reduced pressure and the liberated free bases were converted to their trimethylsilyl (TMS) ether derivatives with 100 μl of dry acetonitrile and 100 μl of bis(trimethylsilyl)trifluoroacetamide containing 1% trimethylchlorosilane. Formic acid hydrolysis and derivatization were conducted at 140°C for 30 min each (18,19). The derivatized bases were then analyzed by GC/MS. GC/MS analysis was performed at least three times on each reaction. The product of the methylation reaction, 5mC, was chromatographically separated from cytosine by gas chromatography before mass spectral analysis. GC/MS has proven to be an extremely sensitive technique, in which modified DNA bases can be measured at a level of one modified DNA base in 100 000 normal DNA bases (14,15,18,19).

Theoretical ion abundance profile calculations

The mass spectrum of a given compound is comprised of the parent ion and various fragment ions that are generated by electron impact ionization. For the TMS derivatives of the DNA bases the predominant ions observed are the parent ion (P), or intact molecule, and the fragment resulting from loss of a 15 mass unit methyl group from TMS (P-15).

In the mass spectrum of the P-15 ion of unenriched cytosine there is a line at 240 a.m.u. as well as smaller lines at 241, 242, etc. The mass lines greater than 240 a.m.u. result from the natural abundance of heavier isotopes. The presence of a heavier isotope of nitrogen increases the mass of the molecule by ~1 mass unit. A cytosine molecule carrying one ¹⁵N would then be detected as 241 a.m.u. Based upon the known abundance of the heavier isotopes of each of the elements present in the molecule examined and the molecular formula, the relative sizes of the P (240 a.m.u.), P+1 (241 a.m.u.), P+2 (242 a.m.u.), etc. peaks can be calculated (20). For the P-15 fragment of silylated cytosine (C₉H₁₈N₃OSi₂) the calculation of its isotope abundance ratios is given below, where x is the polynomial variable:

$$P_{(\text{cytosine}-15)}(x) = 1 = (0.98892x^{12} + 0.01108x^{13})^9 \times (0.99984x^1 + 0.00016x^2)^{18} \times (0.99635x^{14} + 0.00365x^{15})^3 \times (0.99759x^{16} + 0.00037x^{17} + 0.000204x^{18}) \times (0.92210x^{28} + 0.04700x^{29} + 0.03090x^{30})^2 = 0.7567x^{240} + 0.1642x^{241} + 0.0677x^{242} + 0.0096x^{243}$$

Thus, with the TMS derivative of the P-15 fragment of cytosine, the predicted contribution of molecules with masses of 240, 241, 242 and 243 a.m.u. to the mass spectrum would be 75.67, 16.42, 6.77 and 0.96%, respectively. All of the theoretical mass spectra in this study were calculated using this simple mathematical method described by Hugentobler and Löliger (20).

RESULTS AND DISCUSSION

The sequences of the DNA duplexes used in this study are shown in Figure 1. As an initial test of our synthesis conditions and analytical methods, we prepared and analyzed one oligonucleotide of sequence 5'-d(ATGC)₃-3' using all standard phosphoramidite reagents (Fig. 2A). A second oligonucleotide of the same sequence in which all the cytosine residues were replaced with ¹⁵N-enriched cytosine residues was also prepared and analyzed by GC/MS (Fig. 2B). In the mass spectrum of cytosine the parent ion is observed at 255 a.m.u. and the most prominent fragment ion is P-15 at 240 a.m.u. For this reason, the P-15 ion was chosen for study over the parent ion.

The theoretical ion abundance profiles for the cytosine 240 a.m.u. fragment ion and ¹⁵N-enriched cytosine 241 a.m.u. fragment ion in the above samples are shown in Figure 2A and B. As explained in Materials and Methods, the predicted contribution of molecules with masses of 240, 241, 242, 243 and 244 a.m.u. to the mass spectrum of cytosine, based upon natural abundance, would be 75.67, 16.42, 6.77, 0.96 and 0.16%, respectively. The experimental percentages are 75.60, 16.53, 6.78, 0.85 and 0.24%, respectively (Fig. 2A). The correlation coefficient (*r*²) between theory and experiment is 0.9996. For the mass-tagged cytosine the entire ion abundance profile is shifted one mass unit higher and thus there would be 0% molecules with mass 240 a.m.u., 75.67% with mass 241 a.m.u., etc. The experimental percentages are <1 (240 a.m.u.), 75.68 (241 a.m.u.), 16.32 (242 a.m.u.), 7.03 (243 a.m.u.) and 0.97% (244 a.m.u.) (*r*² = 0.9998) (Fig. 2B).

For all of the ion abundance profiles in this paper calculated and experimental abundances at each mass are presented on the same figure to allow direct visual comparison. The experimentally derived ion abundance at each mass is the solid line to the left of the pair of lines, whereas the dashed line to the right of each pair is the theoretical calculated line. The agreement between the experimentally determined profile and the theoretical profile is excellent.

We then synthesized Duplex 1 (Fig. 1) in which the cytosine residues indicated in upper case bold were mass tagged with ¹⁵N and all the other cytosines were not tagged. The sequence of Duplex 1 was described previously by Smith *et al.* (21) as a good substrate for methylases of both prokaryotic and eukaryotic origin. Following deprotection and purification, the DNA duplex was hydrolyzed, derivatized and analyzed by GC/MS as described above. The abundance profile of the cytosine P-15 fragment ion peak is shown in Figure 2C. The percentages of the mass lines expected, based upon 99% ¹⁵N-amino enrichment of two of the 19 cytosine residues in Duplex 1, is 68.79 (240 a.m.u.), 21.81 (241 a.m.u.), 7.64 (242 a.m.u.) and 1.49% (243 a.m.u.). The experimental percentages are 68.18, 21.49, 8.57 and 1.77%, respectively (*r*² = 0.9989).

Duplex 1 was then treated with *Hpa*II methylase, which has been shown to methylate the inner cytosine residues in the sequence CCGG, dried, hydrolyzed, derivatized and analyzed

Figure 2. (A) The ion abundance profile for the P-15 peak of silylated unenriched cytosine. (B) The ion abundance profile for the P-15 peak of silylated (4-¹⁵NH₂)-cytosine. (C) The ion abundance profile for the P-15 peak of the silylated cytosine peak derived from Duplex 1, a 29 bp duplex in which two of 19 cytosine residues are selectively enriched with ¹⁵N. The experimentally derived ion abundance at each mass is the solid line to the left of the pair of lines, whereas the dashed line to the right of each pair is the theoretical line calculated.

by GC/MS (Fig. 3). In contrast to the above experiments in which we were observing the mass spectrum of cytosine, here we were interested in the product of the methylation reaction, 5mC, which has a retention time of 6.0 min. The mass spectrum of 5mC is dominated by the parent ion (269 a.m.u.) and the P-15 fragment (254 a.m.u., Fig. 3B). In Duplex 1 there is only one potentially methylatable cytosine in each strand. As only the internal cytosines are mass tagged, we anticipated that the product 5mC would also be mass tagged. The ratio of the 254 to 255 a.m.u. lines in the 5mC spectrum would then indicate the relative degree of methylation at the inner tagged (255 a.m.u.) and other cytosine (254 a.m.u.) sites.

Methylation of the duplex by *Hpa*II methylase was indicated by the appearance of a 5mC peak at 6.0 min. In Figure 3A we show the ion chromatogram for the 270 a.m.u. ion. The 270 a.m.u. ion was chosen because thymine at natural abundance and mass-tagged 5mC, arising from methylation of the inner, target site are both of mass 270, but have different retention times. Comparison of the integrated peak areas for the thymine and 5mC peaks indicates that ~38% of the total methylatable sites have been methylated. In order to determine

Figure 3. (A) The ion chromatogram at 270 a.m.u. showing the silylated thymine and 5mC peaks derived from treatment of Duplex 1 with *HpaII* methylase. (B) The ion abundance profile of the silylated 5mC P-15 peak, derived from treatment of Duplex 1 with *HpaII* methylase. The experimentally derived ion abundance at each mass is the solid line to the left of the pair of lines, whereas the dashed line to the right of each pair is the theoretical line calculated.

which cytosines were methylated, we examined the ion abundance profile of the P-15 fragment ion of 5mC (Fig. 3B). By inspection it is apparent that only the mass-tagged inner cytosines of the central CCGG site were methylated, as there is only a 255 a.m.u. line, and the 254 a.m.u. line is not observed.

Parallel studies were also conducted with cloned *SssI* methylase and *MspI* methylase, which have been shown to methylate only CpG dinucleotides and the outer cytosine residues in the sequence CCGG, respectively. Duplex 1 was separately treated with both enzymes, then dried, hydrolyzed, derivatized and analyzed by GC/MS (Fig. 4). For *SssI* methylase, in Duplex 1 there is only one potentially methylatable site, the central CG site. As only the internal cytosines are mass tagged, we anticipated that the ion abundance profile would be identical to that of the *HpaII* methylase experiment, and Figure 4A confirms this expectation. For *MspI* methylase, in Duplex 1 there is also only one potentially methylatable site, the outer cytosines of the central CCGG site on both strands, which are not mass tagged. Indeed, Figure 4B reveals that the 5mC produced from this reaction was not mass tagged.

To determine whether there is any strand preference to the methylation reaction, we prepared Duplex 2 (Fig. 1), a duplex of the same sequence as Duplex 1 in which only the upper strand is mass tagged. Duplex 2 was incubated with *HpaII* methylase under the previously described conditions, dried, hydrolyzed, derivatized and analyzed by GC/MS. The ion

Figure 4. (A) The ion abundance profile of the silylated 5mC P-15 peak, derived from treatment of Duplex 1 with *SssI* methylase. (B) The ion abundance profile of the silylated 5mC P-15 peak, derived from treatment of Duplex 1 with *MspI* methylase. The experimentally derived ion abundance at each mass is the solid line to the left of the pair of lines, whereas the dashed line to the right of each pair is the theoretical line calculated.

abundance profile of the silylated 5mC P-15 peak in Figure 5A illustrates that there may be a slight preference for methylation of the lower, tagged strand. However, this difference corresponds to a preference for one strand of only a few percent, slightly above experimental error.

Previously it had been demonstrated that the presence of an A-C base mismatch at the target methylation site substantially increases the degree of methylation, relative to an unmethylated target site, by *HpaII* methylase, perhaps by facilitating base flipping (22). We wished to repeat this experiment with the mass tagging approach to verify these findings. We constructed Duplex 3, a duplex of the same sequence as Duplex 1 except that an A-C base pair, in which the mismatched cytosine was mass tagged, was inserted into the middle of the central CCGG site. Treatment with *HpaII* methylase resulted in predominant methylation at the cytosine in the mismatch, as lack of the 254 a.m.u. line in the ion abundance profile in Figure 5B indicates. This observation is of particular significance, within the context of the relationship between perturbation of methylation patterns and tumorigenesis, in that the occasional formation of mismatches could possibly stimulate methylation of previously unmethylated sites (21), perhaps accounting in part for the observed methylation of promoter regions of tumor suppressor genes. The presence of mismatches might also direct methylation exclusively to one strand, destroying the symmetry of the methylation pattern.

In previous studies the degree of methylation of an unmethylated, normal target site versus that of an unmethylated target site containing a mismatch has been measured independently for

Figure 5. (A) The ion abundance profile of the silylated 5mC P-15 peak, derived from treatment of Duplex 2 with *HpaII* methylase. (B) The ion abundance profile of the silylated 5mC P-15 peak, derived from treatment of Duplex 3 with *HpaII* methylase. (C) The ion abundance profile of the silylated 5mC P-15 peak, derived from treatment of Duplexes 1 and 4 with *HpaII* methylase. The experimentally derived ion abundance at each mass is the solid line to the left of the pair of lines, whereas the dashed line to the right of each pair is the theoretical line calculated.

each substrate (23). In order to investigate methylase preference using our approach, we conducted a direct competition assay in which the two different substrates were simultaneously present. Due to mass tagging of selected cytosine residues, we can unequivocally determine the preference of the methylase. We are unaware of any assays, apart from those described here, which would allow such measurements to be made directly.

We synthesized Duplex 4 (Fig. 1), a duplex of the same sequence as Duplex 1 except that an A-C base pair, in which the mispaired cytosine was not mass tagged, was inserted into the middle of the central CCGG site. Equimolar amounts of Duplex 1 and Duplex 4 and *HpaII* methylase were all reacted together, and the outcome of the GC/MS analysis is presented in Figure 5C. In this competition experiment there are three potentially methylatable cytosine residues: two tagged cytosines on Duplex 1 and one untagged cytosine on Duplex 4. If all three cytosines are methylated equally, the amount of tagged 5mC (255 a.m.u.) observed would be twice that of untagged 5mC (254 a.m.u.). The theoretical abundance profile for equal methylation at the three potential sites is indicated by

the dashed lines in Figure 5C. Experimentally we observed that methylation occurred preferentially at the untagged, mispaired cytosine residue (254 a.m.u.), as indicated by the solid line in Figure 5C. In this experiment the mispaired cytosine was methylated approximately seven times more frequently than the cytosine residues in the corresponding normal duplex.

CONCLUSIONS

In this paper we have demonstrated how a single mass-tagged site can be used to examine base selectivity, strand selectivity, consequences of a mispair at the methylation site and competition between a normal and mispaired DNA duplex. We found that the bacterial methyltransferase *HpaII* methylates the two strands of a normal, complementary duplex similarly. In our assay we observed that the presence of an A-C base pair at the target methylation site enhanced the capacity of the mispaired cytosine residue to serve as a substrate for the methyltransferase while preventing methylation of the opposing strand. Furthermore, in direct competition assays we discovered that the methyltransferase preferentially methylates duplex DNA structures containing a mismatched A-C base pair at the target site 7-fold over normal, unmethylated, complementary DNA duplexes.

The mass tagging approach presented here is extremely powerful for examining the substrate selectivity of cytosine DNA methyltransferases. With this approach it is possible to quantitatively probe the consequences of various forms of DNA structural perturbations, including base mispairs and damaged bases, on the methylation of specific cytosine residues. With the addition of more isotope labels (17,24), methylation could be measured simultaneously at multiple sites on the same or competing duplexes and the relative rates of productive versus non-productive encounter of the methyltransferase with its target could be determined. Mass tagging studies could contribute substantially to the understanding of the mechanisms by which methylation patterns become altered.

ACKNOWLEDGEMENT

This work was supported by National Institutes of Health Grant CA84487.

REFERENCES

1. Riggs, A.D. (1975) *Cytogenet. Cell Genet.*, **14**, 9-25.
2. Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.*, **262**, 4778-4786.
3. Laird, P.W. and Jaenisch, R. (1994) *Hum. Mol. Genet.*, **3**, 1487-1495.
4. Schmutte, C. and Jones, P.A. (1998) *Biol. Chem.*, **379**, 377-388.
5. Gruenbaum, Y., Cedar, H. and Razin, A. (1982) *Nature*, **295**, 620-622.
6. Hardy, T.A., Baker, D.J., Newman, E.M., Sowers, L.C., Goodman, M.F. and Smith, S.S. (1987) *Biochem. Biophys. Res. Commun.*, **145**, 146-152.
7. Feinberg, A.P. and Vogelstein, B. (1983) *Nature*, **301**, 89-92.
8. Nyce, J. (1989) *Cancer Res.*, **49**, 5829-5836.
9. Wainfan, E. and Poirier, L.A. (1992) *Cancer Res.*, **52**, 2071s-2077s.
10. Smith, S.S., Kaplan, B.E., Sowers, L.C. and Newman, E.M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 4744-4748.
11. Shen, J.-C., Rideout, W.M. and Jones, P.A. (1992) *Cell*, **71**, 1073-1080.
12. Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) *Nucleic Acids Res.*, **22**, 2990-2997.
13. Rein, T., DePamphilis, M.L. and Zorbas, H. (1998) *Nucleic Acids Res.*, **26**, 2255-2264.
14. Crain, P.F. (1990) *Methods Enzymol.*, **193**, 857-865.
15. Crain, P.F. and McCloskey, J.A. (1983) *Anal. Biochem.*, **132**, 124-131.

16. Divakar,K.J. and Reese,C.B. (1982) *J. Chem. Soc. Perkin Trans.,I.*, 1171–1176.
17. LaFrancois,C.J., Fujimoto,J. and Sowers,L.C. (1998) *Chem. Res. Toxicol.*, **11**, 75–83.
18. Dizdaroglu,M. (1985) *Anal. Biochem.*, **144**, 593–603.
19. LaFrancois,C.J., Yu,K. and Sowers,L.C. (1998) *Chem. Res. Toxicol.*, **11**, 786–793.
20. Hugentobler,E. and Lölliger,J. (1972) *J. Chem. Ed.*, **49**, 610–612.
21. Smith,S.S., Hardy,T.A. and Baker,D.J. (1987) *Nucleic Acids Res.*, **15**, 6899–6916.
22. Klimasauskas,S. and Roberts,R.J. (1995) *Nucleic Acids Res.*, **23**, 1388–1395.
23. Laayoun,A., Baker,D.J., Riley,J. and Smith,S.S. (1994) *Gene*, **150**, 195–196.
24. Ahmadian,M. and Bergstrom,D.E. (1998) *Nucl. Nucl.*, **17**, 1183–1190.