Recognition of Foldback DNA by the Human DNA (Cytosine-5-)-methyltransferase†

Steven S. Smith,* Robert G. Lingeman, and Bruce E. Kaplan

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

Received January 30, 1991; Revised Manuscript Received August 9, 1991

ABSTRACT: In order to specify the recognition requirements of the human DNA (cytosine-5-)-methyltransferase, two isomeric 48mers were synthesized so as to link a long block of DNA with a shorter complementary block of DNA through a tether consisting of five thymidine residues. These isomeric foldback molecules, differing only in the location of the 5-methyldeoxycytosine, were shown to be unimolecular, to contain a region of duplex DNA, and to contain a region of single-stranded DNA. When used as substrates for the DNA methyltransferase, only one of the isomers was methylated. A comparison of the structures of the two isomers allows us to begin to define the potential sites of interaction between the enzyme and the three nucleotides forming a structural motif consisting of 5-methyldeoxycytosine, its base-paired deoxyguanosine, and a deoxycytosine 5′ to the paired deoxyguanosine.

One of the novel properties of the human DNA (cytosine-5-)-methyltransferase is its preference for unusual DNA structures. Since this preference suggests that unusual DNA structures could provide biological signals for the methylation of DNA in vivo, it is important to more clearly define this specificity. Previous work has shown that the enzyme actively methylates DNA containing mispaired bases, gaps, O^6-methyldeoxyguanosine (O^6MedG)[1] residues, and abasic sites (Smith et al., 1987, 1991; Baker et al., 1987, 1988; Hardy et al., 1987; Kan & Smith, 1989; Tan & Li, 1990; Hepburn et al., 1991). The action of the enzyme on each of the above substrates suggests that only a three-nucleotide motif in DNA is required for recognition. The putative recognition motif in duplex DNA containing 5-methyldeoxycytosine (5MedC) is composed of a 5MedC residue, the deoxyguanosine (dG) residue to which it is base paired, and a deoxycytosine (dC) residue 5′ to the paired dG (Smith et al., 1991).

In what follows, we have used structurally isomeric uneven foldback molecules containing a short region of duplex DNA and a short region of single-stranded DNA to test this proposed recognition specificity. The response of the enzyme to the placement of 5MedC in these structural isomers was consistent with the recognition (Smith et al., 1991) of the three-nucleotide motif for methyl-directed methylation by the enzyme. These studies allow us to begin to define required points of enzyme–DNA interaction within the motif itself.

Experimental Procedures

Partially purified DNA (cytosine-5-)-methyltransferase was prepared from human placentas by using previously published methods (Zucker et al., 1985; Smith et al., 1987). Procedures used for oligodeoxynucleotide synthesis, as well as those used for tritium labeling and determination of methylation rate, were also as previously described (Smith et al., 1991).

Fill-In Reaction. The fill-in reaction was a modification of the method of Wu (1970). Fill-in reaction conditions for a 50-μL reaction mixture were as follows: approximately 500 ng of ^{32}P-end-labeled foldback DNA (0.63 μM), 250 μM dATP, 250 μM dTTP, 250 μM dCTP, 250 μM dGTP, 100 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 80 mM NaCl, and 1 unit of the Klenow fragment of DNA polymerase I (BRL Inc., Gaithersburg, MD). Reaction mixtures were incubated at 37 °C, and samples were taken at 0.5, 2, and 4 h. Fill-in reaction products were separated by electrophoresis on 20% polyacrylamide/8 M urea gels at 60 V/cm.

Fluorescence-Enhanced Autoradiography of Enzymatically Tritiated Oligodeoxynucleotides. Tritium-labeled DNA fragments were also separated by electrophoresis on DNA sequencing gels (20% polyacrylamide/8 M urea) and visualized by fluorescence autoradiography as described previously (Smith et al., 1987). Digestion of oligodeoxynucleotides with *Hpa*II, *Msp*I, and *Mbo*I was carried out as described by Smith et al. (1991).

Molecular Modeling. Molecular models were constructed in BIOGRAF (Biodesign, Pasadena, CA). Energy minimizations were performed by using the Dreiding parameter set for molecular mechanics and dynamics calculations. Photographs were taken directly from the screen of a Silicon Graphics workstation.

Determination of T_m Values. Oligodeoxynucleotides were annealed as described before (Smith et al., 1991). DNA at various concentrations was suspended in the buffer used for enzyme assay but with dithiothreitol omitted. Thermal melting profiles were obtained with the Response II spectrophotometer system (Gilford Instruments, Oberlin, OH). T_m values were determined from the maximum of the first derivative of the melting profile at 260 nm obtained with the Responstar software package.

Results and Discussion

The two structural isomers shown in Figure 1 can be used to study methyltransferase recognition specificity. If the enzyme requires only three nucleotides for substrate recognition

† This work was supported by grants from the National Institutes of Health (GM38350) and the Council For Tobacco Research (1571B), with additional support from the City of Hope Cancer Center (CA33572-09).

* Address correspondence to this author.

[1] Abbreviations: O^6MedG, O^6-methyldeoxyguanosine; dA, deoxyadenosine; dC, deoxycytosine; dG, deoxyguanosine; dI, deoxyinosine; 5MedC, 5-methyldeoxycytosine; dT, thymidine; AdoMet, *S*-adenosylmethionine; DEAE-cellulose, (diethylamino)ethylcellulose.

0006-2960/92/0431-850$03.00/0

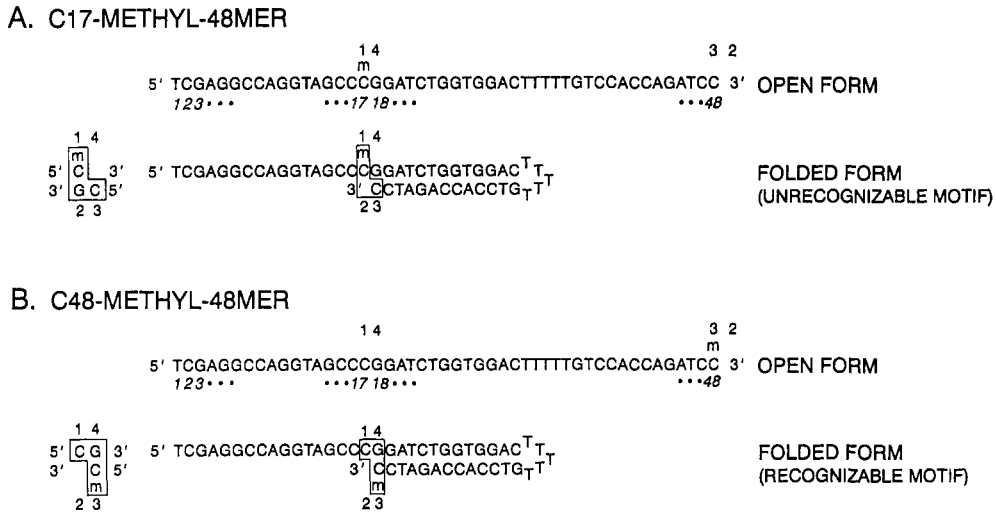

FIGURE 1: Response of DNA methyltransferase to 5MedC in gapped d(pCG) dinucleotide pairs. (A) The foldback molecule depicted has a 5MedC at the d(pCG) site near the center of the long block (position 17, numbered beneath open form). This molecule is referred to as C17-methyl-48mer in the text. In the folded form, the 5MedC occupies site 1 of the gapped d(pCG) dinucleotide pair, but it does not present the required recognition motif shown at the left. (B) The second foldback molecule (C48-methyl-48mer) is a structural isomer of the first; the 5MedC in this oligomer is at position 48. When folded, the 5MedC is at site 3 of the gapped d(pCG) dinucleotide pair. The complete recognition motif is present in this molecule and is shown at the left of the folded form.

Table I: Methyl-Directed Methylation in Foldbacks and Duplexes

substrate	substrate sequence[a]	methylation rate[b] (fmol/min)	relative rate[c]
REFERENCE SINGLE STRAND	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC3'	N.D.	---
UNMETHYLATED DUPLEX	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC3' 3'AGCTCCGGTCCATCGGGCCTAGACCACCTG5'	4.2 ± 1.0	1.0
UNMETHYLATED 48MER	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC^{T}T$_{T}$T 3'CCTAGACCACCTG$_{T}$T^{T}	10.5 ± 1.6	2.5
ASYMMETRICALLY METHYLATED DUPLEX I	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC3' 3'AGCTCCGGTCCATCGGGCCTAGACCACCTG5'	115.2 ± 18.2	54.8
C17-METHYL-48MER	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC^{T}T$_{T}$T 3'CCTAGACCACCTG$_{T}$T^{T}	N.D.	---
ASYMMETRICALLY METHYLATED DUPLEX II	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC3' 3'AGCTCCGGTCCATCGGGCCTAGACCACCTG5'	180.8 ± 35.4	86.1
C48-METHYL-48MER	5'TCGAGGCCAGGTAGCCCGGATCTGGTGGAC^{T}T$_{T}$T 3'CCTAGACCACCTG$_{T}$T^{T}	249.7 ± 25.3	118.9

[a] The positions of 5MedC moieties introduced during chemical DNA synthesis are indicated by $\overset{m}{C}$. The positions of enzymatically methylated dC moieties are depicted by asterisks. The site of the d(pCG) dinucleotide pair in each substrate oligodeoxynucleotide is numbered. The sites of enzymatic methylation were determined by using *Hpa*II, *Msp*I, and *Mbo*I. [b] Means and standard deviations were from at least six independent determinations. ND, not detectable. [c] Calculated on the basis of the rate of methylation at the relevant cytosine in the unmethylated duplex.

(see Figure 1), then it should be able to differentiate between the two isomeric 48mers. The isomer with the 5MedC residue at position 17 (C17-methyl-48mer, Figure 1A) was not expected to be actively methylated because it lacks a recognizable three-nucleotide motif in its folded form. On the other hand, the structural isomer with the 5MedC residue at position 48 (C48-methyl-48mer, Figure 1B) was expected to be actively methylated because the folded form does possess a recognizable motif.

Characterization of Foldbacks. To determine whether intramolecular foldbacks in fact formed, the melting behavior of the oligodeoxynucleotide depicted in Figure 1B was studied. Melting profiles showed clear thermal transitions, suggesting the formation of a partial duplex. The T_m of the oligodeoxynucleotide was 75 °C.

Marky et al. (1987) have shown that the T_m value for an intramolecular association is independent of the input concentration ($[C_T]$) of single strands. Intermolecular associations involving multiple strands (e.g., duplex, triplex, quadruplex, etc.) do exhibit concentration dependence for the plot of $1/T_m$ versus ln $[C_T]$. These plots are linear, with slopes equal to $(n-1)R/\Delta H_{vH}$ and intercepts that are a function of van't Hoff entropy (ΔS_{vH}) and van't Hoff enthalpy (ΔH_{vH}) as well as the molecularity (n) of the reaction (Marky et al., 1987). A plot of $1/T_m$ versus ln $[C_T]$ for the oligodeoxynucleotide depicted in Figure 1B was independent of the input DNA concentration

over the range tested ($10^{-6.5}$–10^{-5} M; data not shown). This suggested that the thermal transition is associated with the melting of a unimolecular foldback.

Each of the 48mers in Table I was an effective substrate for DNA polymerase I in the absence of exogenous primer. When the ^{32}P-end-labeled 48mers shown in Table I were extended in the presence of unlabeled nucleotide triphosphates, the product oligodeoxynucleotide was identified as a discrete 65mer in autoradiograms. No unextended molecules were detected (data not shown). DNA polymerase I is a template-directed polymerase that requires a free 3′ hydroxyl group at the end of a base-paired primer for enzymatic activity. Since this enzyme will not copy single-stranded molecules effectively, these results were also consistent with the formation of a unimolecular foldback.

Methyltransferase Recognition of Foldback Molecules. To determine whether the enzyme could in fact distinguish between the two structural isomers, enzymatic methylation rates were determined. The C48-methyl-48mer was methylated very rapidly, while only background levels of incorporation were observed with the C17-methyl-48mer (see Table I). The control rate obtained with the duplex analogue of the C48-methyl-48mer (asymmetrically methylated duplex II) was comparable to the rate obtained with the foldback. On the other hand, the control rate obtained with the duplex analogue of the C17-methyl-48mer (asymmetrically methylated duplex I) was much faster than the rate obtained with the foldback.

Additional control experiments showed that the single-stranded analogue of the long block in the foldback (reference single strand; Table I) was not a substrate for the enzyme. On the other hand, the unmethylated 48mer *was* methylated slowly. Since this 48mer forms a foldback, one could attribute the moderate enhancement of the de novo reaction rate to the recognition of an unmethylated three-nucleotide motif or to the recognition of some other structural feature in the foldback (e.g., the oligothymidine loop). A comparison of the reaction rate observed with the unmethylated 48mer with those observed for the two methylated structural isomers showed that the reaction was inhibited when a methyl group was present on cytosine-17 (C17-methyl-48mer, Table I) and stimulated when a methyl group was on cytosine-48 (C48-methyl-48mer, Table I). This strongly suggests that the availability of 5MedC in a recognizable three-nucleotide motif dominates substrate recognition.

This interpretation of the kinetic data suggests that the enzymatically methylated form of the C48-methyl-48mer is the foldback and that methylation is confined to cytosine-17. The results of experiments designed to determine whether or not the product of enzymatic methylation was a foldback are depicted in Figure 2. In the first experiment, enzymatically methylated (i.e., tritiated) C48-methyl-48mer was prepared, mixed with nondenaturing loading buffer, and separated by electrophoresis under nondenaturing conditions. In the second experiment, enzymatically tritiated C48-methyl-48mer was mixed with denaturing loading buffer and separated under denaturing conditions. A set of duplex oligodeoxynucleotide markers that had been tritium labeled on one strand were used as markers in each experiment in order to compare relative mobilities. The enzymatically labeled 48mer migrated with the relative mobility of a duplex 26mer under nondenaturing conditions (Figure 2), but it migrated with the relative mobility of a single-stranded 48mer under denaturing conditions (Figure 2). In contrast, a control oligothymidine 45mer (i.e., a sequence lacking internal homology) migrated well behind the 30mer markers under both denaturing and nondenaturing

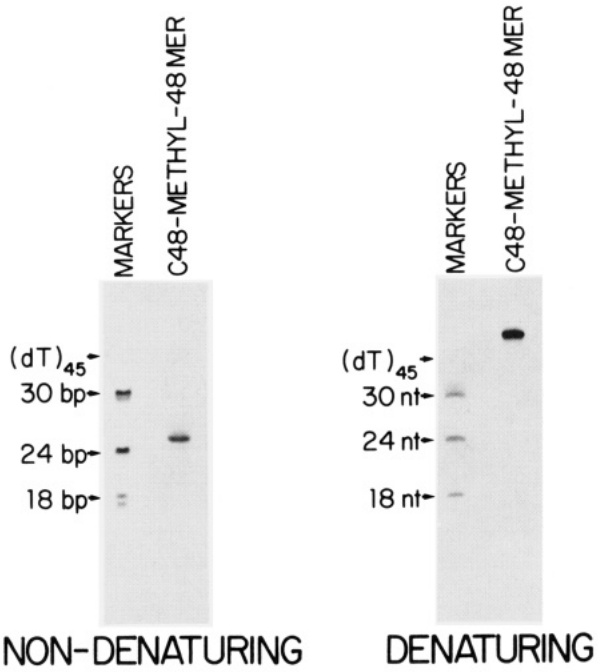

FIGURE 2: Mobilities of preferentially methylated foldbacks. Foldback molecules were enzymatically methylated using AdoMet. The product was then electrophoresed on nondenaturing or denaturing gels. Fluorescence-enhanced autoradiographs were used to detect tritium-labeled fragments. Arrows indicate the positions of a $(dT)_{45}$ marker.

conditions. Taken together, these findings suggest that the enzymatically labeled product was the foldback form of the 48mer.

Restriction analysis was used to determine whether or not the site of enzymatic labeling was cytosine-17 as expected. *Hpa*II and *Msp*I are isoschizomers that recognize the d(pCCGG) tetranucleotide in DNA. *Mbo*I recognizes the d(pGATC) tetramer. Each of these enzymes requires a short duplex region containing an intact tetrameric recognition sequence.

In order to permit a complete restriction analysis, it was necessary to denature the enzymatically methylated product and reanneal it in the presence of a 2-fold excess of a 30mer complementary to the 30-nucleotide arm of the 48mer. The 30mer readily displaced the 13-nucleotide arm of the foldback to form 30 bp of duplex with an overhanging 3′ tail (Figure 3A). On nondenaturing gels, this partial duplex had a lower mobility than that of a duplex 30mer lacking a tail (not shown).

The fragmentation patterns predicted for cleavage of the partial duplex with various restriction enzymes are given in Figure 3A. *Hpa*II is able to nick a duplex molecule when it is asymmetrically methylated at the centrally located cytosine in the d(pCCGG) site. In this case, nicking occurs only on the unmethylated strand (McClelland & Nelson, 1988). Thus, enzymatic labeling at the centrally located cytosine (cytosine-17) in the d(pCCGG) site in the 48mer should block cleavage of labeled molecules visualized by autoradiography (Figure 3A). *Hpa*II did not cleave the enzymatically labeled C48-methyl-48mer (Figure 3B, lane 3).

In contrast to *Hpa*II, *Msp*I is able to cleave duplex DNA when the centrally located cytosine residue in the d(pCCGG) site is methylated. Cleavage of the enzymatically labeled 48mer should produce an unlabeled 16mer and a tritium-labeled 32mer (Figure 3A). As can be seen from Figure 3B, lane 4, *Msp*I cleaved the enzymatically labeled 48mer. The only labeled fragment produced by this enzyme had the mobility of a single-stranded 26mer. The labeled fragment

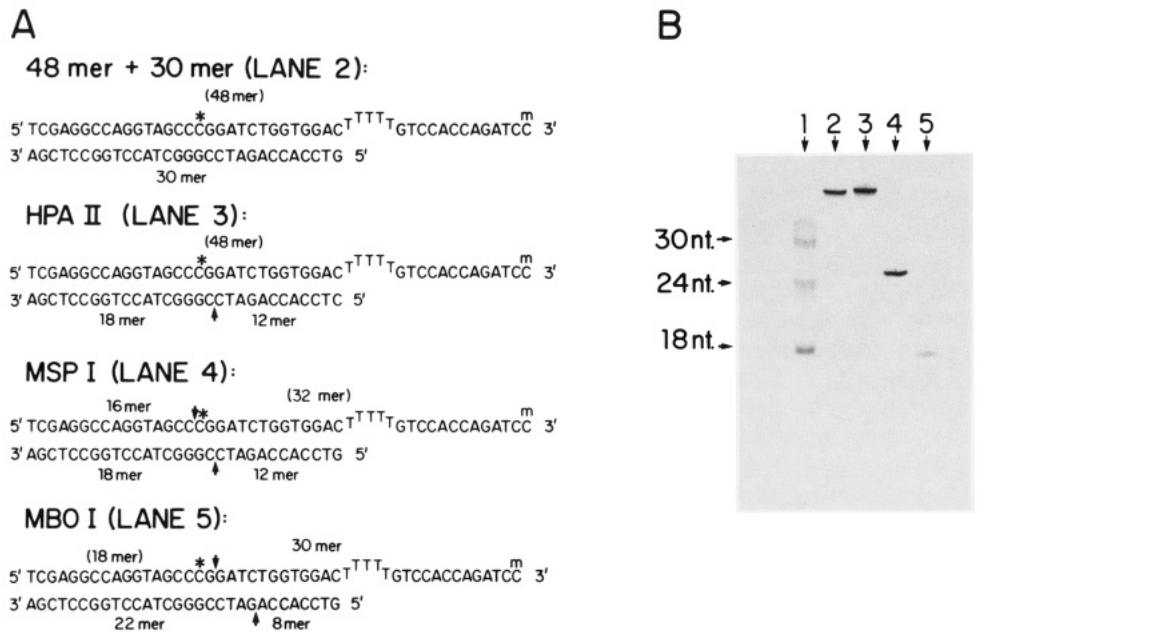

FIGURE 3: Analysis of enzymatically methylated foldbacks. (A) Fragmentation patterns predicted for restriction endonuclease cleavage are shown. The sizes of labeled fragments are shown in parentheses. (B) Fluorescence-enhanced autoradiograph of the digested DNAs separated on 20% denaturing acrylamide gels. Lane 1: tritium-labeled oligodeoxynucleotide markers 30, 24, and 18 bases long. Lane 2: tritium-labeled foldback and complementary 30mer separated without restriction enzyme treatment. Lane 3: tritium-labeled foldback plus complementary 30mer digested with *Hpa*II. Lane 4: *Msp*I-cleaved tritium-labeled foldback and complementary 30mer. Lane 5: *Mbo*I-cleaved tritium-labeled foldback and 30mer.

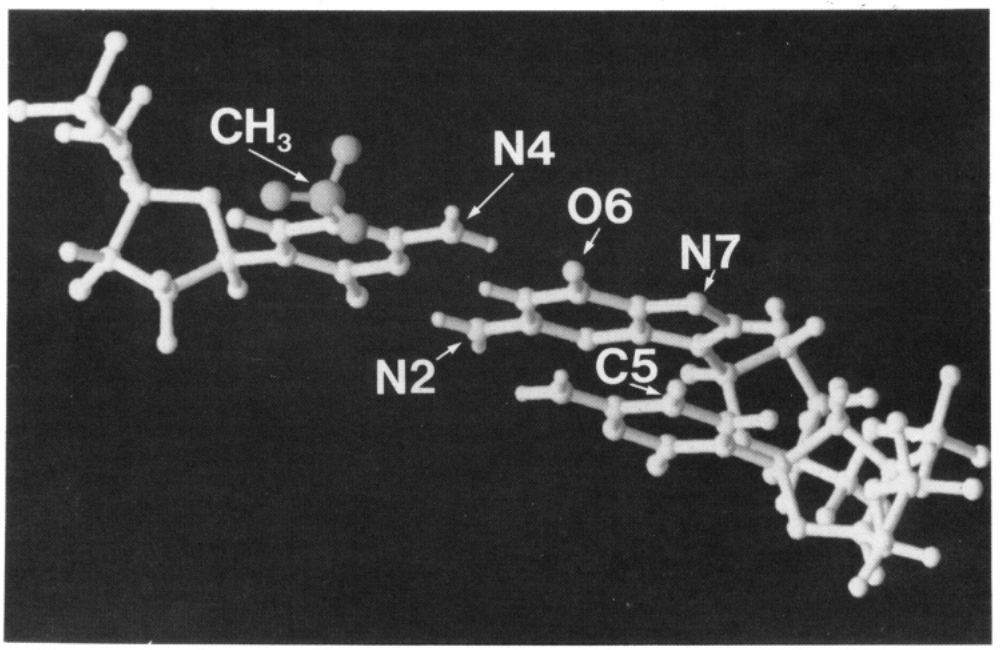

FIGURE 4: Structural features of the three-nucleotide motif. A molecular model of an asymmetrically methylated duplex 30mer (see Table I) in the B-form was constructed in BIOGRAF. The isolated three-nucleotide motif is depicted as it would appear from a perspective oblique to the major groove. The C-5 methyl and N-4 amino moieties of the directing 5MedC residue are indicated at the upper left. N-2, O-6, and N-7 of the necessary guanosine residue are also indicated, as is C-5 of the methyl-accepting cytosine moiety.

represents the 32mer containing the dT loop, since it is clearly too slow to be compatible with the 18mer, 16mer, or 12mer that can be produced by digestion with *Msp*I (Figure 3A). Given the specificity of *Hpa*II and *Msp*I, these results are sufficient to assign the site of enzymatic methylation in the C48-methyl-48mer to cytosine-17. However, additional proof of this assignment is provided by the *Mbo*I cleavage pattern given in Figure 3B, lane 5. The exclusive labeling of the 18mer originating from the 48mer (Figure 3B) confirms the assignment because cytosine-17 is the only cytosine in the sequence overlap between the 18mer produced by *Mbo*I and the 32mer produced by *Msp*I. The sites of methylation for duplex 30mers deduced from previous analyses (Smith et al., 1987, 1991) are indicated in Table I.

Recognition Requirements for the DNA Methyltransferase. The differences in the response of the methyltransferase to the structural isomers in Table I clearly showed that a base-paired guanosine residue adjacent to the 5MedC residue in an asymmetrically methylated d(pCG) dinucleotide pair was not required for full enzyme activity. Moreover, the data allow us to begin to assign potential sites of interaction between the enzyme and the three nucleotides that are required for recognition. The nucleotides required for methyl-directed methylation are the 5MedC residue, its base-paired dG, and the

dC residue 5′ to the base-paired guanosine. They are depicted in Figure 4 in the conformation that they adopt in B-DNA.

The importance of the 5MedC residue in directing methylation to the acceptor dC is demonstrated by its ability to stimulate the reaction. The nature of the interaction between the cytosine residue and the enzyme is not yet clear. However, the enzyme is known to measure the size of the substituent at C-5 of cytosine (Hardy et al., 1987).

Other sites of interaction between the enzyme and the 5MedC residue are also likely to exist. For example, dT is not an effective director of methylation when mispaired with guanosine at the d(pCG) dinucleotide pair (Baker et al., 1987). Thus, the methyl group at C-5 of dT is not sufficient to stimulate enzyme activity when it is adjacent to an oxygen in the major groove at position 4. Since 5MedC provides a strong stimulation of activity from this same position in the three-nucleotide motif, it seems reasonable to suspect that the N-4 amino group (Figure 4) provides a proton donor site for interaction with the enzyme. While this possibility exists, the requirement for *S*-adenosylmethionine (AdoMet) binding in the methylation reaction introduces the possibility that hydrogen bonding between DNA and the enzyme may involve bridging through the AdoMet moiety. In this case, the N-4 amino group of the directing 5MedC would provide a proton donor site but the proton acceptor would lie on AdoMet and not on the surface of the methyltransferase. This type of bridging cannot be ruled out by the experiments described above.

The inability of the enzyme to effectively methylate the foldback molecule with 5MedC at position 17 (Table I) shows that interaction with guanosine is required. The capacity of deoxyinosine (dI) to serve at this site in duplex molecules (unpublished observations) shows that N-2 of guanosine is not required for enzymatic activity. Although dI can effectively substitute for dG at this site, neither deoxyadenosine (dA) nor O^6MedG is an effective substitute for dG (Kan & Smith, 1989; Smith et al., 1991). Since dA carries an amino nitrogen instead of an oxygen at position 6 and since O^6MedG carries a modification at this site, O-6 may provide an important hydrogen-bond acceptor site during interaction of the enzyme (or AdoMet) with dG or dI.

The sites of interaction between the enzyme and the cytosine accepting the methyl group are not clearly defined by these experiments. Although C-5 is methylated, the mechanism proposed by Santi et al. (1983) suggests that a group on the enzyme would facilitate catalysis through nucleophilic attack at C-6. Both of these sites are easily accessible in the major groove of B-DNA. In light of the experiments described above showing that the accepting cytosine need not be base-paired, interaction with N-4, N-3, or O-2 on the accepting cytosine cannot be ruled out.

The recognition specificity established above may have significance in structural studies of DNA, since the active methylation of an unusual structure containing 5MedC may be taken as evidence that the three-nucleotide motif is present. On the other hand, the biological significance of this specificity is more difficult to assess. For the foldback structures studied here, a 5MedC residue at one site in DNA was clearly able to direct methylation to a second site *on the same strand*. This could provide a mechanism for the spreading of DNA methylation in a region capable of cruciform or foldback formation.

References

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

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

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.

Hepburn, P. A., Margison, G. P., & Tisdale, M. J. (1991) *J. Biol. Chem. 266*, 7985–7987.

Kan, J. L. C., & Smith, S. S. (1989) *J. Cell. Biochem. Suppl. 13D*, 218.

Marky, L. A., Kallenbach, N. R., McDonough, K. A., Seeman, N. D., & Breslauer, K. J. (1987) *Biopolymers 26*, 1621–1634.

McClelland, M., & Nelson, M. (1988) *Gene 74*, 291–304.

Santi, D. V., Garrett, C. E., & Barr, P. J. (1983) *Cell 33*, 9–10.

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

Smith, S. S., Kan, J. L. C., Baker, D. J., Kaplan, B. E., & Dembek, P. (1991) *J. Mol. Biol. 217*, 39–51.

Tan, N.-W., & Li, B. F. L. (1990) *Biochemistry 29*, 9234–9240.

Wu, R. (1970) *J. Mol. Biol. 51*, 501–521.

Zucker, K. E., Riggs, A. D., & Smith, S. S. (1985) *J. Cell. Biochem. 29*, 337–349.