1584–1589 Nucleic Acids Research, 1995, Vol. 23, No. 9

Methylation of slipped duplexes, snapbacks and cruciforms by human DNA(cytosine-5)methyltransferase

Ali Laayoun* and Steven S. Smith

Department of Cell and Tumor Biology, City of Hope National Medical Center, 1500 E. Duarte Road, Duarte, CA 91010, USA

Received December 9, 1994; Revised and Accepted March 16, 1995

ABSTRACT

When human DNA(cytosine-5)methyltransferase was used to methylate a series of snapback oligodeoxynucleotides of differing stem lengths, each containing a centrally located CG dinucleotide recognition site, the enzyme required a minimum of 22 base pairs in the stem for maximum activity. Extrahelical cytosines in slipped duplexes that were 30 base pairs in length acted as effective methyl acceptors and were more rapidly methylated than cytosines that were Watson–Crick paired. Duplexes containing hairpins of CCG repeats in cruciform structures in which the enzyme recognition sequence was disrupted by a C•C mispair were also more rapidly methylated than control Watson–Crick-paired duplexes. Since enzymes have higher affinities for their transition states than for their substrates, the results with extrahelical and mispaired cytosines suggest that these structures can be viewed as analogs of the transition state intermediates produced during catalysis by methyltransferases.

INTRODUCTION

DNA(cytosine-5)methyltransferases catalyze the covalent addition of methyl groups to cytosine residues, creating 5-methylcytosine in DNA (1–3). The preferred site for DNA methylation in mammals and other vertebrates is the dinucleotide sequence CG. In mammals, this biological modification of the cytosine residue may play a role in the control of gene expression (4) and may contribute to chromosome stability (5). Abnormalities in DNA methylation may contribute to the pathogenesis of certain human diseases, including cancer (6,7) and fragile X syndrome (8–10). Methyltransferases from both bacterial (11–13) and mammalian sources (14) operate by activating cytosine for methyl-transfer through a mechanism in which a cysteine moiety at the enzyme active site makes a nucleophilic attack at C-6 of the target cytosine. The dihydrocytosine intermediate thus formed participates in methyltransfer by accepting a methyl group from S-adenosyl-L-methionine (15). Work with human methyltransferase has demonstrated an exceptional acceleration of *de novo* methylation by DNA containing mispaired cytosines (16–22) and suggests that protonation at N-3 might activate the cytosine ring as a methyl acceptor by facilitating nucleophilic attack at C-6. Molecular modeling (20) of both the carbanion and the dihydrocytosine intermediate, with β-mercaptoethanol or β-mercaptan as analogs of the active-site nucleophile, showed that these intermediates cannot stack in their normal position in B-DNA regardless of stereochemistry. This finding supports the proposals that catalysis cannot proceed without a significant disruption of the structure of the B-DNA helix (19,22) and that methyltransferases have evolved to accommodate this structural perturbation during catalysis. Crystallographic data on the structure of the covalent complex formed between the *Hha*I methyltransferase and 5-fluorodeoxycytidine-containing DNA (23) show that the required conformational change is accommodated by rotating the target cytosine into a completely extrahelical position within the active site. Thus the cytosine ring is completely extrahelical during catalysis by this bacterial methyltransferase. In light of the strong sequence conservation at the active site of these enzymes (2), this appears to be a common feature of all methyltransferases.

In the present study, we sought answers to the following questions. (i) What is the minimum size of the DNA required by human methyltransferase for maximal DNA substrate activity? (ii) Can extrahelical cytosines in slipped structures (24) be actively methylated in duplexes satisfying the size requirement? (iii) Can human enzyme methylate cytosines in the hairpins or cruciform structures that satisfy the size requirement?

MATERIALS AND METHODS

Oligodeoxynucleotide synthesis

The oligodeoxynucleotides were synthesized by β-cyanoethyl phosphoramidite chemistry on a 1 μmol scale with a Cyclone Plus MilliGen/Biosearch instrument and were purified by using OligoPak columns from Millipore (South San Francisco, CA). To form duplexes, complementary strands were annealed by heating at 95°C for 5 min in 10 mM Tris, 1 mM EDTA, 100 mM NaCl. Incubation at 50°C for 1 h was followed by gradual cooling to room temperature and then to 0°C prior to use. Oligodeoxynucleotide concentrations were estimated by measuring the optical density at 260 nm for each sample.

* To whom correspondence should be addressed at present address: Laboratoire des Sondes Nucléiques, UMR bio-Mérieux, CNRS n° 103, Ecole Normale Supérieure, 46 allée d'Italie, 69364 Lyon cedex 07, France

Table 1. Methylation rates for different substrates[a]

Duplex Name	Reaction Rate[b]	Sequence[c]
I	20.3 +/- 1.8	5'CCAGAGGTACCTGGACGTGAGATCCAGTGAG 3' 3'GGTCTCCATGGACCTGCACTCTAGGTCACTC 5'
II	274.7 +/- 9.7	5'CCAGAGGTACCTGGAGTGAGATCCAGTGAG 3' 3'GGTCTCCATGGACCTCACTCTAGGTCACTC 5'
III	422.7 +/- 45.2	5'CCAGAGGTACCTGGAGTGAGATCCAGTGAG 3' 3'GGTCTCCATGGACCTCACTCTAGGTCACTC 5'
IV	206.0 +/- 20.6	5'CCAGAGGTACCTGGAGTGAGATCCAGTGAG 3' 3'GGTCTCCATGGACCTCACTCTAGGTCACTC 5'
V	666.7 +/- 55.6	5'CCAGAGGTACCTGGAGTGAGATCCAGTGAG 3' 3'GGTCTCCATGGACCTCACTCTAGGTCACTC 5'

[a]The assays were carried out as described in Materials and Methods at the indicated concentrations of duplexes and S-adenosyl-L-[methyl-^{3}H]methionine.
[b]Reaction rates are the means of five initial velocity determinations ± 1 standard deviation of the mean. They are expressed in femtomoles of [^{3}H]methyl groups incorporated per minute. Under these conditions the rate observed for comparable hemimethylated duplexes was 304 fmole/min to 631 fmole/min (20,21). Thus the *de novo* rates reported here are comparable to the methyl-directed rates observed previously (20,21).
[c]The square shaped box on duplex I highlights a representative CG dinucleotide pair. The L shaped box on duplex V highlights the 3 nt motif recognized by the human methyltransferase.

Enzymatic methylation

Human placental methyltransferase (44 U) (16) was incubated with duplex DNA (4 μM) and S-adenosyl-L-[methyl-^{3}H]methionine (6 μM; 15 Ci/mmol; Amersham) in a mixture containing 50 mM Hepes [4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid] (pH 7), 50 mM NaCl, 2 mM dithiothreitol, 75 μM spermine, and 10% (v/v) glycerol in a total vol of 100 μl. Velocities were determined with oligodeoxynucleotides substrates as described in reference 16.

Restriction analysis

After duplexes II and III were methylated (see Table 1), the two solutions were desalted and concentrated using Centricon-10 (Amicon, Beverly, MA) as described by the manufacturer. The concentrated and desalted fractions were dried, and restriction enzyme digestion was carried out, in a total volume of 50 μl, by using *Mbo*I and *Bst*NI endonucleases (New England Biolabs, Beverly, MA), also under the conditions described by the manufacturer.

Gel electrophoretic analysis of enzymatically tritiated DNA

After digestion by *Mbo*I and *Bst*NI endonucleases, 150 000 c.p.m. of labeled products was mixed with denaturing buffer, heat denatured, and separated by electrophoresis through a 20% polyacrylamide (8 M urea) sequencing gel. The gel slab was soaked in En^{3}Hance (Dupont), transferred to filter paper, and dried. For autoradiography, Kodak XAR film was exposed (for up to 1 week) at –70°C.

Gel electrophoretic analysis of ^{32}P 5′ end-labeled duplexes

Oligodeoxynucleotide (400 nmol) was incubated in kinase buffer with 5 μCi of [γ-^{32}P]ATP (3000 Ci/mmol) and 50 U of polynucleotide kinase (New England Biolabs, Beverly, MA) at 37°C for 1 h in a total volume of 10 μl. Oligodeoxynucleotide duplexes were constructed by annealing as described above and electrophoresed through a 12% polyacrylamide non-denaturing gel.

Molecular modeling and molecular orbital calculations

Energy minimized molecular models of the extrahelical structures of duplexes I, II and III in Table 1 were generated in Biograf 3.21 (Molecular Simulations Inc. Burlington, MA) using a network of Silicon Graphics Workstations. Molecular orbital calculations also carried out on a network of Silicon Graphics workstations, using Spartan 3.1 (Wavefunction Inc., Irvine, CA). Relative energies used in ordering models of reaction intermediates were taken from eigenvalues for the relevant frontier molecular orbital calculated at the Hartree–Fork level of theory neglecting configuration interaction and using the 6-31G* basis set.

RESULTS AND DISCUSSION

A kinetic footprint for the human methyltransferase

To determine the number of base pairs required by the human methyltransferase for maximum interaction with its DNA substrate, we used oligodeoxynucleotides expected to form double-stranded DNA (snapbacks) by unimolecular association (sequence described in Fig. 1). Enzymatically methylatable snapbacks form under these conditions as previously demonstrated (25). These snapback molecules, which have a hemimethylated CG site in the center of the stem and four T residues in the loop, differ in their stem lengths: 2 base pairs (bp) in a linear 8mer, 6 bp in a linear 16mer, 10 bp in a linear 24mer, 14 bp in a linear 32mer, 22 bp in a linear 48mer, and 46 bp in a linear 96mer. They were annealed and methylated as described in Materials and Methods. The results are reported in Figure 1. Under these conditions, methylation activity increased roughly with substrate length until a stem length of ~22 bp was reached; no further increase was afforded by substrates with longer stems.

Electrophoretic analysis of slipped duplexes and cruciforms

With duplex substrates it is important to be certain that each substrate is present as an intact duplex and is not dissociated into single strands or into several duplex structures. To verify the formation of the appropriate duplexes, the two component single strands were separately ^{32}P 5′ end-labeled and annealed under conditions expected to give stoichiometric formation of the appropriate duplex. A non-denaturing 12% polyacrylamide gel showing this to be the case is shown in Figure 2. In the case of

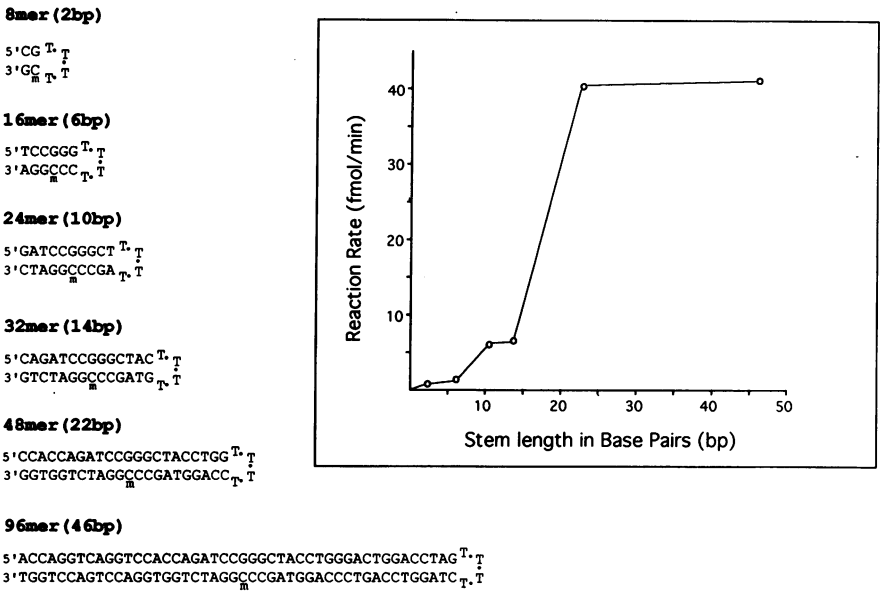

Figure 1. Effect of oligodeoxynucleotide stem length on methylation activity. Six oligodeoxynucleotides expected to form double-stranded DNA by unimolecular association were synthesized and annealed as described in Materials and Methods. These snapback molecules, which have a 5-methylcytosine (C_m) at the CG site in the center of the folded block and four T residues in the loop, differ in their stem lengths: 8mer, 2 bp; 16mer, 6 bp; 24mer, 10 bp; 32mer, 14 bp; 48mer, 22 bp; 96mer, 46 bp. The hemimethylated CG site in each duplex was methylated as described in Materials and Methods, and the reaction rates are reported here.

standard duplex I and slipped duplexes II and III, all the label ran at the position expected for the length of each duplex. The markers are single-stranded poly(T) oligodeoxynucleotides. The autoradiogram (Fig. 2) also shows that constrained hairpins (cruciforms) IV and V form a unique duplex structure in each case. These structures migrate more slowly than predicted by the number of their base pairs. Duplex IV (36 bp) migrates as a 55mer, and duplex V (45 bp) migrates as a 75mer. As discussed previously (26,27), the stem structures are telomere-like fold-backs (28). This electrophoretic mobility is characteristic of the T (or Y) form of duplexes with three arms (29,30). Given the evidence in Figure 1, we can conclude that each synthetic oligodeoxynucleotide duplex used in this study is present in the form given in Table 1.

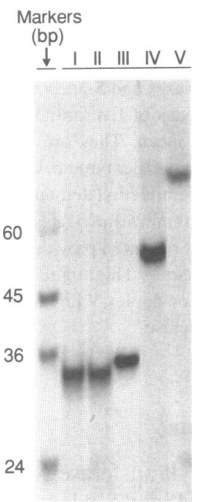

Figure 2. Non-denaturing (12%) polyacrylamide gel for standard duplex (I), slipped duplexes (II and III), and cruciforms (IV and V) used in methylation assays. Substrates I–V were prepared as described in Materials and Methods by annealing single-stranded oligodeoxynucleotides that were synthesized and ^{32}P 5′ end-labeled also as described in Materials and Methods. Labeled oligodeoxynucleotides poly(dT) of the indicated lengths served as single-stranded electrophoretic markers.

Methylation of slipped and cruciform structures

Reaction rates with human methyltransferase, determined for each substrate as described in Materials and Methods, are given in Table 1. Each reaction rate is the mean of five initial velocity determinations ±1 standard deviation of the mean. The data show that duplex II [containing an extrahelical cytosine (C)] and duplex III (containing an extrahelical CG dinucleotide) are methylated by human enzyme at *de novo* rates that are respectively 13 and 20 times faster than that of the comparable Watson–Crick-paired duplex (duplex I). Duplexes IV and V, incorporating hairpins of CCG triplet repeats, are also excellent substrates for the human enzyme, with relative rates 10 and 32 times faster than that of duplex I. This is comparable to the 15- to 31-fold stimulations observed with hemimethylated 30mers under these conditions (20,21). To localize the cytosine residue target of enzymatic methylation in duplexes II and III, we analyzed restriction patterns produced from DNA that had been enzymatically labeled with [^{3}H]methyl groups provided by S-adenosyl-L-[methyl-^{3}H]-methionine. The tritiated DNA product was cleaved with *Mbo*I and *Bst*NI restriction endonucleases, and the labeled fragments were visualized by low-temperature autoradiography (Fig. 3). In the case of duplex II which contains an extrahelical C, the production of a labeled 20mer with *Mbo*I (recognition site: 5′↓GATC-3′) or *Bst*NI (recognition site: 5′-CC↓-(A/T)GG-3′) indicated that the label was on the upper strand, uniquely specifying the extrahelical cytosine as the target of the methyl-

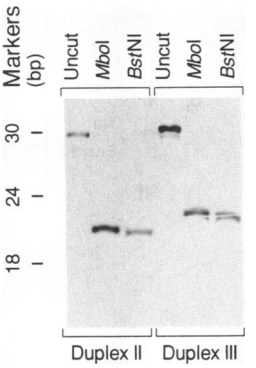

Figure 3. Autoradiograph from the restriction analysis of duplexes II and III. These duplexes, methylated by human DNA(cytosine-5)methyltransferase in the presence of S-adenosyl-L-[methyl-^{3}H]methionine, were cleaved with the indicated restriction endonucleases and separated by electrophoresis through a 20% polyacrylamide sequencing gel. A schematic of the restriction analysis of labeled oligodeoxynucleotides is shown here.

transferase, since this is the only cytosine common to the two labeled fragments. This was also true in the case of the extrahelical CG pair of duplex III, since its cleavage by *Mbo*I and by *Bst*NI produced labeled 21mers (Fig. 3).

The restriction enzyme *Bso*FI has been used to make a preliminary assignment of the site of methylation in the CCG hairpin (27). This enzyme recognizes the sequence 5′-GCNG-C-3′and is inhibited by the presence of 5-methylcytosine at the cytosines specified in its recognition site. It is not sensitive to methylation at the central nucleotide in its recognition sequence which can be occupied by any base. After methylation with the human enzyme *Bso*FI, cleaves the sequence showing that the methyl groups are applied to the cytosines in the CG dinucleotide (27) corresponding to those sites in the hairpin in slipped duplex characterized by the C•C mispairs. We have not yet attempted to assign methylation to particular locations in the stem of the slipped cruciform.

Previous work with foldback oligodeoxynucleotides of several types (25) showed that these structures are unimolecular and that they can be excellent substrates for human methyltransferase. The curve reported in Figure 1 gives the relationship between stem length and reaction rate for a series of foldback oligodeoxynucleotides with centrally located recognition sites for the methyl-directed activity of the human methyltransferase. Maximum activity is obtained with a 48mer (22 bp in the stem) and a 96mer (46 bp in the stem). We interpret the maximum for stem lengths >22 bp as the kinetic footprint associated with the enzyme; i.e., the minimum size required by human enzyme for maximal enzyme-DNA contact during catalysis is ~22 bp as determined kinetically.

Recognition, structure and mechanism

Duplex 30mers II and III which contain extrahelical C and CG residues satisfy this size requirement. The data in Table 1 show that these two duplexes are excellent substrates for the methyltransferase, since they are methylated respectively 13 and 20 times faster than the corresponding standard substrate (duplex I) which contains a Watson–Crick-paired CG site. In DNA duplex II, the cytosine ring that acts as the methyl acceptor is expected to be completely extrahelical and therefore is not base paired or stacked (24). Since this slipped duplex is readily cleaved at restriction sites by *Mbo*I and *Bst*NI, it is clear that the regions neighboring the unpaired cytosine adopt a B-DNA structure. Indeed, molecular modeling of the extrahelical cytosine residue in duplex II (data not shown) suggests that the slipped duplex may be a very close structural analog of the duplex carrying the extrahelical cytosine in the enzyme–DNA complex formed between the *Hha*I methyltransferase and the duplex DNA containing 5-fluorodeoxycytidine at the enzyme target site (23). Since enzymes have a higher affinity for their transition states (31), we conclude that the increase in reaction rate afforded by the structural perturbation of the substrate shows that it is a productive transition-state analog of the methyltransferase.

In the case of duplex III which contains an extrahelical CG pair, the stimulation observed again suggests that the substrate mimics the transition state for the enzyme. Consistent with this hypothesis, molecular modeling studies suggest that the conformation of the cytosine ring in duplex III is comparable to that in duplex II (data not shown). However, the difference in the level of stimulation noted between duplex II (13-fold stimulation) and duplex III (20-fold stimulation) suggests that the cytosine ring in the extrahelical CG pair has more conformational degrees of freedom than the corresponding cytosine in duplex I.

Recognition of these anomalous structures is also supported by previous work suggesting that the human methyltransferase(s) exhibits a relaxed sequence specificity for substrate recognition (22). The sequence specificity can be viewed as relaxed because only three of the four nucleotides (nt) in the CG/CG dinucleotide pair are recognized by the enzyme: a cytosine on one strand base paired with a guanine on the other strand that is adjacent to a 5′ target cytosine. Thus the enzyme recognize only a 3 nt motif in DNA (see Table 1). Since the second guanine site in the dinucleotide pair is not recognized by the enzyme, the target cytosine can be opposite any base, an abasic site or the 3′ end of a partial duplex (3). Rapid rates of methylation observed for the extrahelical C and CG are consistent with the presence of this 3 nt recognition motif in each of the disrupted CG sites in duplexes II and III. In the 3 nt motif, the cytosine 5′ to the paired guanine is the methyl acceptor (26). The results reported in Figure 3 show that this is also the case for extrahelical cytosines of duplexes II and III.

The excellent stimulation observed with the cruciform duplex V which incorporates 15 bp in its hairpin suggests that the enzyme shifts its orientation on the cruciform so that it now recognizes the CCG repeat in the foldback region (26). Compared to the rate observed with the duplex the rate enhancement is ~33-fold. In previous work we noted that a single C•C mispair gave a rate enhancement of ~7-fold (16). Since there are five C•C sites in the stem of this cruciform, the data are consistent with the processive attack of each site by the enzyme to yield the cumulative rate enhancement. It is also possible that the enzyme(s) have an enhanced affinity for the cruciform structure itself. This later interpretation is suggested by the kinetic footprint of the enzyme on duplex DNA which would tend to confine methylation available sites near the center of the footprint.

Since the 3 nt recognition motif mentioned above (3) targets a cytosine in each C•C mispair (26,27), stimulation of catalysis may be provided by the conformational freedom imparted by the mispair. The stimulation might also result from those mispaired cytosines that are protonated at N-3. Preliminary *ab initio* molecular orbital calculations for 1-methylcytosine and

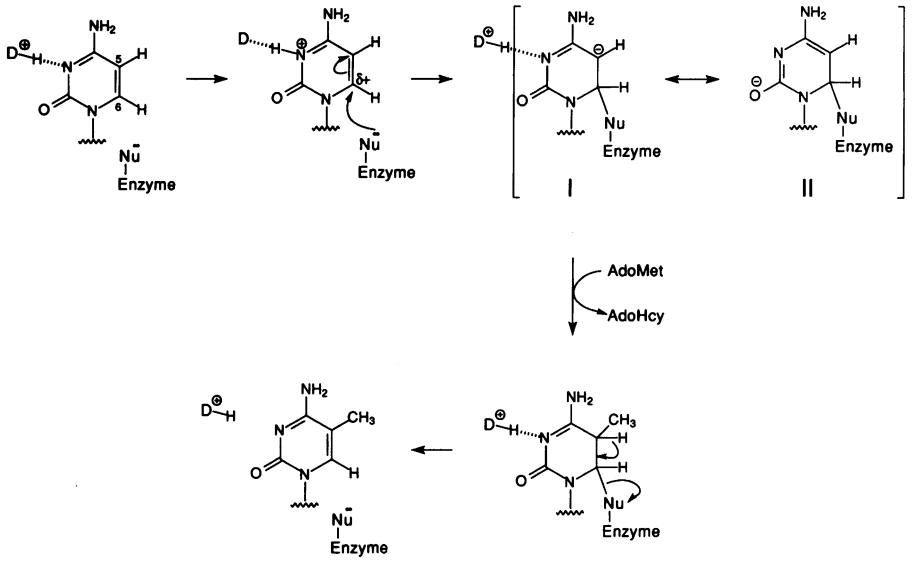

Figure 4. Possible mechanism of DNA methylation by DNA(cytosine-5)methyltransferase. Binding of the cytosine methyl acceptor at the active site establishes several hydrogen bonds between cytosine, conserved hydrogen acceptors, and hydrogen donors [$D^{(+)}$-H] provided by amino acid side chains at the active site of the enzyme (23). Since protonation at N-3 can evidently occur, transient protonation at this site may activate the C-6 for nucleophilic attack to produce carbanion I. This carbanion is stabilized by resonance with anion II. Activated C-5 is thus able to attack the methyl group on S-adenosyl-L-methionine (AdoMet). The 5,6-dihydrocytosine intermediate is transiently formed, and β-elimination driven by resonance stabilization of the product generates 5-methylcytosine and active enzyme. Protonated dynamic cytosines like those in the C•C mispair would be transition-state analogs of the second intermediate shown here.

1-methyl-5-fluorocytosine as models of cytosine in DNA suggest that the ease of nucleophilic attack should be: cytosine < 5-fluorocytosine < N-3-protonated cytosine < N-3-protonated 5-fluorocytosine based on the energies for the lowest unoccupied molecular orbitals (LUMO energies).

As one might expect this calculation is consistent with the possibility that an exogenously supplied proton at N-3 of cytosine would accelerate the reaction (3,16,19,22). Extensive studies with the bisulfite-catalyzed deamination of cytosine also suggest that the electronic reorganization associated with protonation at N-3 facilitates nucleophilic attack at C-6 (32).

While this protonation at N-3 could account for the observed stimulation of the reaction, NMR structural studies of the foldbacks in duplexes IV and V are most consistent with unprotonated C•C base-pairs in which both bases are *anti* and the N-4 of one base is hydrogen bonded to N-3 the other (27). This would suggest that the rate enhancement that we observe at C•C mispairs is largely due to conformational similarity to the transition state for the catalysis (16,19,22). Nevertheless, the NMR data cannot rule out the presence of protonated $C•C^+$ pairs in low concentration, since the rate enhancement due protonated $C•C^+$ pairs may be quite large, based on the calculated LUMO energies we cannot rule out a contribution from such base-pairs to the rate enhancement.

N-3 protonation might also enhance nucleophilic attack during the course of the normal methylation reaction if the groups at the active site can affect protonation (3,19,22). In fact the crystal structure of the *Hha*I methyltransferase (23) suggests that a glutamic acid that is conserved in all sequenced methyltransferases including the human methyltransferase, appears to be hydrogen bonded to N-3 of cytosine, indicating that N-3 is protonated in the crystal structure. If this proton is transiently donated during initial binding but prior to nucleophilic attack during catalysis, as depicted in Figure 4, then a protonated cytosine in a $C•C^+$ pair would present a doubly activated substrate to the methyltransferase, i.e. one that presents a protonated and unstacked cytosine, approximating the putative protonated-open-form-transition-state-intermediate of the methyltransferase reaction.

However, it is important to recognize that the protonated dihydrocytosine intermediate is at extremely high risk for hydrolytic deamination (14). Thus, unless the active site of the enzyme can exclude water with exceptional efficiency, it is unlikely that enzyme will have evolved to permit the proton to remain at N-3 during the subsequent catalysis. This leads us to favor a mechanism in which the proton is not normally present at N-3 or is eliminated from N-3 during nucleophilic attack and prior to methyltransfer yielding the resonance-stabilized carbanion as originally proposed by Santi and co-workers (11,15).

Verdine has proposed that the proton is donated by a group on the enzyme and that it survives at N-3 to produce an enamine as an activated intermediate (34). Glutamic acid 119 might be expected to donate a proton to N-3 of cytosine in the M•*Hha*I reaction, but it forms a salt bridge with arginine 165 in the absence of bound DNA (35). Since this requires that it be ionized prior to DNA binding it is not likely to be the proton donor. In the crystal structure of the *Hha*I enzyme–DNA complex (23) arginine 165 forms a hydrogen bond to O-4 of the 5-fluorodihydrocytosine. This arginine, which is conserved in all sequenced methyltransferases including the human enzyme (2), is in a position to ligate charge on O-2 that would be present on a resonance stabilized carbanion. Moreover it might be expected to promote tautomerization of the enamine to the O-2 enol. Preliminary calculations for 1-methyl-6-suphydryl cytosines as models of these intermediates at the Hartree–Fock level of theory suggest that the ease of methyltransfer for these intermediates will be: cytosine carbanion > 5-fluorocytosine carbanion >> cytosine-O-2-enol > 5-fluorocytosine-O-2-enol > cytosine N-4-enamine > 5-fluorocytosine-

N-4-enamine, based on calculated energies of the highest occupied molecular orbitals (HOMO energies).

The calculation predicts that 5-fluorocytosine will slow the rate of methyltransfer regardless of which of the intermediates functions in methyltransfer. Recently Som and Friedman (35) have reported that 5-fluorocytosine slows the rate of the reaction for the bacterial *Eco*RII methyltransferase and we have also noted this effect with the human enzyme (unpublished observations). While this result appears to identify methyltransfer as the rate limiting step in the catalysis our preliminary calculation does not allow us to distinguish between the anion, the enol or the enamine as the catalytic intermediate on energetic grounds. Additional experiments and more detailed calculations which include the effects of hydrogen bonding on the energetics and stereochemistry of the proposed intermediates will be required to address this question.

Methylation site of slippage

As noted above, our results clearly suggest that CCG hairpins having stem lengths comparable to the kinetic footprint of the methyltransferase are better substrates than those with smaller stems. This finding may be of considerable biological importance because both the potential for repeat expansion and the development of the Fragile X syndrome at the *FMR1* locus increase with increasing repeat size. Normal individuals inherit triplet repeat elements between ~18 and 150 nt in length while transmitting carriers inherit triplet repeat elements that are ~150–690 nt in length (8–10). Thus normal individuals have a very low probability of forming a hairpin in the region that can accommodate the methyltransferase footprint.

The increased probability for the formation of hairpins long enough to accommodate the methyltransferase may explain the occurrence of *de novo* methylation that has been observed in the region containing the repeat in affected individuals (10). Under normal conditions, cytosine methylation might serve to mark slipped structures for repair (5), in the absence of repair, asymmetric methylation of heteroduplexes would establish a clonally inheritable pattern of *de novo* methylation on the expanded strand resulting from replication of the slipped heteroduplex (26,27).

Although *de novo* cytosine methylation is predicted by the enzymology, the function of this methylation in the evolution of the disease is more difficult to assess. One possibility is that it plays an indirect role in silencing the *FMR1* gene (33). Alternatively, normal methylation patterns might serve to mark the parental strand during loop repair at slipped intermediates formed during replication. In this context *de novo* methylation of loops formed on the nascent strand would present the hypothetical system with an excision dilemma associated with attendant changes in repeat length.

ACKNOWLEDGEMENTS

This work was supported by grant 0388 from the Smokeless Tobacco Research Foundation, and by grant N00014-94-1-1116 from the Office of Naval Research.

REFERENCES

1 Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
2 Kumar,S., Cheng,X., Klimasauskas,S., Mi,S., Posfai,J., Roberts,R.J. and Wilson,G.G. (1994) *Nucleic Acids Res.*, **22**, 1–10.
3 Smith,S.S. (1994) *Prog. Nucleic Acids Res. Mol. Biol.*, **49**, 65–111.
4 Razin,A. and Cedar,H. (1991) *Microbiol. Rev.*, **55**, 451–458.
5 Smith,S.S. (1991) *Mol. Carcinog.*, **4**, 91–92.
6 Jones,P.A. and Buckley,J.D. (1990) *Adv. Cancer Res.*, **54**, 1–23.
7 Doerfler,W., Toth,M., Kochanek,S., Achten,S., Freisem-Rabien,U., Behn-Krappa,A. and Orend,G. (1990) *FEBS Lett.*, **268**, 329–333.
8 Craig,I. (1991) *Nature*, **349**, 742–743.
9 Hwu,W-L., Lee,Y-M., Lee,S-C. and Wang,T-R. (1993) *Biochem. Biophys. Res. Commun.*, **193**, 324–329.
10 Oberlé,I., Rousseau,F., Heitz,D., Fretz,C., Devys,D., Hanauer,A., Boué,J., Bertheas,M.F. and Mandel,J.L. (1991) *Science*, **252**, 1097–1102.
11 Osterman,D.G., DePillis,G.D., Wu,J.C., Matsuda,A. and Santi,D.V. (1988) *Biochemistry*, **27**, 5204–5210.
12 Friedman,S. and Ansari,N. (1992) *Nucleic Acids Res.*, **20**, 3241–3248.
13 Chen,L., MacMillan,A.M., Chang,W., Ezaz-Nikpay,K., Lane,W.S. and Verdine,G.L. (1991) *Biochemistry*, **30**, 11018–11025.
14 Smith,S.S., Kaplan,B.E., Sowers,L.C., and Newman,E.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4744–4748.
15 Santi,D. V., Garrett,C.E. and Barr,P.J. (1983) *Cell*, **33**, 9–10.
16 Smith,S.S., Hardy,T.A. and Baker,D.J. (1987) *Nucleic Acids Res*, **15**, 6899–6916.
17 Tan,N-W. and Li,B.F.L. (1990) *Biochemistry*, **29**, 9234–9240.
18 Hepburn,P.A., Margison,G.P. and Tisdale,M.J. (1991) *J. Biol. Chem.*, **266**, 7985–7987.
19 Baker,D.J., Kan,J.L.C. and Smith,S.S. (1988) *Gene*, **74**, 207–210.
20 Baker,D.J., Laayoun,A. and Smith,S.S. (1993) *Biochem. Biophys. Res. Commun.*, **196**, 864–871.
21 Laayoun,A., Baker,D.J., Riley,J. and Smith,S.S. (1994) *Gene*, **150**, 195–196.
22 Smith,S.S., Kan,J.L.C., Baker,D.J., Kaplan,B.E. and Dembek,P. (1991) *J. Mol. Biol.*, **217**, 39–51.
23 Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) *Cell*, **74**, 299–307.
24 Wurdeman,R.L., Douskey,M.C. and Gold,B. (1993) *Nucleic Acids Res.*, **21**, 4975–4980.
25 Smith,S.S., Lingeman,R.G. and Kaplan,B.E. (1992) *Biochemistry*, **31**, 850–854.
26 Smith,S.S., Laayoun,A., Lingeman,R.G., Baker,D.J. and Riley,J. (1994) *J. Mol. Biol.*, **243**, 143–151.
27 Chen,X., Mariappan,S.V.S., Catasti,P., Ratliff,R., Moyzis,R.K., Laayoun,A. Smith,S.S., Bradbury,E.M. and Gupta,G. (1995) *Proc. Natl. Acad. Sci. USA*, in press.
28 Ahmed,S. and Henderson,E. (1992) *Nucleic Acids Res.*, **20**, 507–511.
29 Kallenbach,N.R., Ma,R.I. and Seeman,N. C. (1983) *Nature*, **305**, 829–831.
30 Duckett,D.R., Murchie,A.I.H., Diekmann,S., Von Kitzing,E., Kemper,B. and Lilley,D.M.J. (1988) *Cell*, **55**, 79–89.
31 Fersht,A.R. (1974) *Proc. R. Soc. Lond*, **187**, 397–407.
32 Hayatsu,H. (1976) *Prog. Nucleic Acids Res. Mol. Biol.*, **16**, 75–124.
33 Hansen,R.S., Gartler,S.M., Scott,C.R., Chen,S.-C. and Laird,C.D. (1992) *Hum. Mol. Genet.*, **1**, 571–578.
34 Verdine,G.L. (1994) *Cell*, **76**, 197–200.
35 Som,S. and Friedman,S. (1994) *J. Biol. Chem.*, **269**, 25986–25991.