Critical Reviews in Biochemistry and Molecular Biology, 37(3):167–197 (2002)

Structure, Function, and Mechanism of *Hha*I DNA Methyltransferases

*Umesh T. Sankpal and Desirazu N. Rao**

Department of Biochemistry, Indian Institute of Science, Bangalore 560012, INDIA

* To whom correspondence should be addressed. Tel: 91 80 3092538. Fax: 91 80 3600683 / 3600814. E-mail: dnrao@biochem.iisc.ernet.in

TABLE OF CONTENTS

ABSTRACT: A vast amount of literature has accumulated on the characterization of DNA methyltransferases. The *Hha*I DNA methyltransferase, a C5-cytosine methyltransferase, has been the subject of investigation for the last 2 decades. Biochemical and kinetic characterization have led to an understanding of the catalytic and kinetic mechanism of the methyltransfer reaction. The *Hha*I methyltransferase has also been subjected to extensive structural analysis, with the availability of 12 structures with or without a cofactor and a variety of DNA substrates. The mechanism of base flipping, first described for the *Hha*I methyltransferase, is conserved among all DNA methyltransferases and is also found to occur in numerous DNA repair enzymes. Studies with other methyltransferase reveal a significant structural and functional similarity among different types of methyltransferases. This review aims to summarize the available information on the *Hha*I DNA methyltransferase.

1040-9238/02/$.50

KEY WORDS: DNA methylation, *Hha*I methyltransferase, base flipping, C5-cytosine DNA methyltransferases.

I. INTRODUCTION

Methyl transfer reactions play an important role in many aspects of biology. These reactions are catalysed by enzymes called methyltransferases, predominantly using the cofactor *S*-adenosyl-L-methionine (AdoMet) as the methyl group donor. AdoMet serves as a cofactor for enzymes that act on a variety of substrates, including proteins, DNA, RNA, lipids, polysaccharides, and a number of small molecules. This range of biological methylation reactions involving AdoMet-dependent enzymes makes AdoMet the second most frequently used cofactor after ATP. A number of AdoMet-dependent methyltransferases have been structurally characterized (DNA, RNA, protein, and small molecules). Structural comparisons revealed that they all harbor a common structural fold referred to as the 'AdoMet dependent methyltransferase fold' (Cheng and Roberts, 2001) (Plate 1). This comprises of seven β-sheets of which six are arranged in a parallel orientation with the seventh strand in an antiparallel orientation. However, the similarity among these diverse methyltransferases is limited to this structural fold and is not reflected at the level of sequence conservation.

DNA modification by methylation occurs in diverse organisms ranging from archebacteria, viruses, plant, and mammals. DNA methylation has been shown to play an important role in numerous cellular processes. In prokaryotes, the methylation of DNA is primarily responsible for distinguishing it from foreign viral DNA, thus protecting the host genome from the cognate, defensive restriction endonucleases. The methylation of DNA predominantly occurs at cytosine and adenine residues. Cytosine methylation can take place at the C5 position or at the N4 amino group, whereas adenine methylation occurs at the N6 amino group. DNA methyltransferases can be broadly classified into two groups, the exocyclic amino methyltransferases and the endocyclic carbon methyltransferases. The exocyclic amino methyltransferases comprise of the N4-cytosine and N6-adenine methyltransferase families, whereas the endocyclic methyltransferases comprise the C5-cytosine methyltransferase family. All three forms of methylation are found in prokaryotes, whereas in eukaryotes only C5-cytosine methylation prevails.

II. *Hha*I DNA METHYLTRANSFERASE — CLONING AND PURIFICATION

*Hha*I DNA methyltransferase is a part of the restriction modification system of *Haemophilus haemolyticus* and is involved in the protection of the host genome from restriction by the *Hha*I endonuclease (Roberts *et al.*, 1976; Smith, 1979). The gene encoding *Hha*I methyltransferase was cloned, sequenced, and found to encode a 327 amino acid protein. The transcription initiation sites, the putative ribosome binding site, and promoter sequences were also identified (Caserta *et al.*, 1987). The protein was expressed in *E. coli*, under the control of its natural promoter (Caserta *et al.*, 1987), and found to have a molecular weight of 37 kDa, as determined by gel filtration and SDS-PAGE analysis (Wu and Santi, 1987). In order to obtain

* Plates are placed following page 180.

large amounts of the protein, M.*Hha*I was cloned under the T7 promoter (Wu and Santi, 1988). Although a significant amount of protein was in the pellet fraction, the increased percentage of protein in the supernatant resulted in higher yields. The protein was purified using DE52 and Heparin-Sepharose column chromatography. N-terminal sequence of the purified recombinant protein (Met1-Phe18) and its molecular weight were in agreement with the deduced amino acid sequence. Plasmids containing the *Hha*I gene were not able to propagate on all bacterial strains due to the presence of methylation sensitive restriction systems. A more stringent regulation on the expression of M.*Hha*I was achieved by cloning the gene under a *lac*I repressible expression system (pHSH0-1) (Klimasauskas *et al.*, 1991). However, in the process of cloning, Ile at position 2 was changed to Leu. The protein was purified from such an IPTG-inducible expression system. The purification protocol was slightly modified and the insoluble protein present in the pellet was solubilized using high salt conditions (400 m*M*) (Kumar *et al.*, 1992). High-salt-extracted protein was purified further to homogeneity using cation and anion exchange (Mono S and Mono Q, respectively) column chromatography. The mutant protein (I2L) had no appreciable difference in terms of activity compared with the wild-type protein (Kumar *et al.*, 1992).

III. SEQUENCE ANALYSIS AND CONSERVED MOTIFS

With the continuous discovery of restriction-modification systems, and the cloning, sequencing, and characterization of a number of methyltransferases, interest in studying these enzymes has increased. This is partly due to the possibility of using these enzymes as model systems to study DNA-protein interactions. The sequencing of a number of DNA methyltransferases has facilitated the study of their sequence homology, which led to the identification of a number of conserved motifs. These motifs are fairly conserved among all DNA methyltransferase families (C5-cytosine, N6-adenine, and the N4-cytosine) (Lauster *et al.*, 1989) (Posfai *et al.*, 1989; Kumar *et al.*, 1994). The genes for more than 100 methylases have been cloned and their sequence determined. This provides a basis for comparative studies aimed at elucidating the functional organization of these motifs. Preliminary analysis of the sequences of methyltransferases (protein, DNA, and RNA) revealed three short patches of sequence similarity (Ingrosso *et al.*, 1989). Conservation of sequence among such diverse methyltransferases suggested a common functional and evolutionary origin. Sequence analysis of 11 C5-cytosine methyltransferases, both prokaryotic and eukaryotic, led to the identification of four homologous regions (Lauster *et al.*, 1989). A similar analysis with 13 C5-cytosine methyltransferases demonstrated the presence of five highly conserved motifs (Posfai *et al.*, 1989). A more rigorous analysis performed using 35 to 45 methyltransferase sequences indicated the presence of six more motifs (Cheng *et al.*, 1993a; Kumar *et al.*, 1994). Therefore, among the C5 methylase family, 10 conserved motifs (I to X) could be identified. When such an analysis was done with the N4-cytosine and N6-adenine methyltransferases, nine out of the ten (I to VIII and X) conserved motifs of C5-cytosine methyltransferase family could be identified in these enzymes (Klimasauskas *et al.*, 1989; Malone *et al.*, 1995).

All 10 conserved motifs identified in the C5-methyltransferase family could be detected in the *Hha*I methyltransferase (M.*Hha*I). The consensus sequence for these motifs are listed in Table 1, and their location in M.*Hha*I is shown in Plate 2. Among these motifs, two have been characterized

TABLE 1
Conserved motifs

Motif	Function	Consensus sequence *
I	AdoMet binding	$\mathbf{F}_{18}$xGxG
IV	Catalysis	P$\mathbf{C}_{81}$Q
VI	DNA binding	$\mathbf{E}_{119}$NV
VII		DY
VIII	DNA binding	QxRxR
IX		R-----E
X	AdoMet binding	G$\mathbf{N}_{304}$
II, III, V	AdoMet binding	$\mathbf{E}_{40}$,$\mathbf{W}_{41}$ / $\mathbf{D}_{60}$ /$\mathbf{L}_{100}$
TRD	DNA binding	—

* sequence shown is for M.*Hha*I; amino acids in bold face are directly involved in the interaction with cofactor AdoMet or DNA, as seen from the crystal structures.

Note: The ten conserved motifs in the C5-cytosine methyltransferase family are listed with their consensus sequence and possible role in methylation reaction.

reasonably well. Crystallographic and biochemical evidence is available, from both M.*Hha*I and other methyltransferases, to assign a functional role for these motifs.

Motif I, the FxGxG motif, is involved in cofactor AdoMet binding. This motif was identified in studies carried out with various types of methyltransferases (Ingrosso *et al.*, 1989; Klimasauskas *et al.*, 1989; Lauster *et al.*, 1989; Posfai *et al.*, 1989). This is the only motif conserved among all AdoMet-dependent enzymes. The evidence for the involvement of this motif in AdoMet binding also came from structural studies with M.*Hha*I. The crystal structure of the binary M.*Hha*I-AdoMet complex clearly demonstrated the involvement of this motif in cofactor binding (Cheng *et al.*, 1993a). Apart from motif I, the AdoMet binding pocket is comprised of conserved residues from motifs II, III, and V (Table 1) (Kumar *et al.*, 1994).

Motif IV, the PC motif, is involved in catalysis. The highly conserved cysteine residue is involved in the formation of a covalent enzyme–DNA complex (discussed later). This covalent complex was clearly visualized in the ternary M.*Hha*I-AdoMet-DNA complex (Klimasauskas *et al.*, 1994). Mutational analysis of this residue in M.*Hha*I (Mi and Roberts, 1993) and other systems such as M.*Eco*RII (Wyszynski *et al.*, 1993), M.*Hae*III (Chen *et al.*, 1993), and Dcm (Hanck *et al.*, 1993) led to the abolishment of methyltransferase activity, thus clearly demonstrating the importance of this residue in catalysis. Earlier experi-

ments done with DNA containing 5-azacytosine and 5-fluorocytosine (Friedman, 1985; Osterman *et al.,* 1988) in the presence of AdoMet resulted in trapping of the covalent complex. Using such a strategy the nucleophilic cysteine, of the PC motif, was identified in M.*Hae*III (Chen *et al.,* 1991).

The target recognition domain (TRD) is the variable region between motifs VIII and IX. This domain was first identified in multispecific C5-cytosine methyltransferases (Balganesh *et al.,* 1987; Trautner *et al.,* 1988). Mutations within the target recognition domain resulted in a complete loss of activity. Based on mutational analysis (Wilke *et al.,* 1988) and construction of chimeric enzymes (Trautner *et al.,* 1988), the variable region was assigned a role in target recognition. Hybrid enzymes were also constructed between M.*Hha*I and M.*Hpa*II, and it was seen that specificity of the hybrids was related to the parent methylase that contributed the TRD (Klimasauskas *et al.,* 1991). Similar experiments with M.*Hpa*II and M.*Msp*I, which recognize identical sequences but methylate at different positions (CCGG), demonstrated that the TRD determines the base to be methylated (Mi and Roberts, 1992). The work with both monospecific and multispecific enzymes demonstrated that the TRD, apart from having a role in sequence recognition, is also involved in recognition of the specific base to be methylated. Sequence analysis of C5-cytosine methyltransferases (including M.*Hha*I) revealed the presence of a highly conserved core sequence and a variable region comprising the target recognition domain (Lauster *et al.,* 1989). The predicted target recognition domain was actually found to overlap with the recognition domain observed in the ternary M.*Hha*I-DNA-AdoMet crystal structure (Klimasauskas *et al.,* 1994; Kumar *et al.,* 1994).

IV. STRUCTURAL ANALYSIS OF MOTIFS

The final validation of the role of these motifs came from the structural analysis of methyltransferases. The crystal structure of the binary M.*Hha*I-AdoMet complex (Cheng *et al.,* 1993a) was the first report of a structure for any AdoMet-dependent enzyme and also the first DNA methyltransferase to be crystallized. Structural analysis of the conserved motifs, using the M.*Hha*I structure, can be applied to all systems because the *Hha*I methylase contains all 10 conserved motifs. Subsequent to this structure, M.*Hha*I was crystallized with cofactor AdoMet/AdoHcy and a variety of DNA substrates (Table 2) (Klimasauskas *et al.,* 1994; O'Gara *et al.,* 1996a; O'Gara *et al.,* 1998). A number of other methyltransferases belonging to the C5-cytosine, N6, and N4-adenine methyltransferase families have been crystallized: *Hae*III (Reinisch *et al.,* 1995), *Pvu*II (Gong *et al.,* 1997), *Dpn*I (Tran *et al.,* 1998), *Taq*I (Labahn *et al.,* 1994; Schluckebier *et al.,* 1997), *Rsr*I (Scavetta *et al.,* 2000), and Dnmt2 (Dong *et al.,* 2001). Structural comparison of some of these revealed that they share significant homology, especially at the orientation of active site amino acids (Cheng and Blumenthal, 1996).

Structural studies revealed that M.*Hha*I is organized into two domains. The larger domain comprises 9 out of the 10 conserved motifs and has the cofactor binding motif and the catalytic motif. The variable region, between motifs IX and X, serves as the target recognition domain and forms bulk of the smaller domain (Cheng *et al.,* 1993b). The cleft formed between the two domains is responsible for binding DNA, with the major groove facing the smaller recognition domain and the minor groove facing the larger catalytic domain.

Table 2
Structures of *Hha*I Methyltransferases

Structures	Components		
1hmy	M.*Hha*I	+ AdoMet	
1mht	M.*Hha*I	+ AdoMet	+ DNA (5-fluoro deoxycytidine)
2mhy	M.*Hha*I	+ AdoMet	(in the presence of non-specific DNA)
3mht	M.*Hha*I	+ AdoHcy	+ DNA (unmethylated)
4mht	M.*Hha*I	+ AdoHcy	+ DNA (methylated)
5mht	M.*Hha*I	+ AdoHcy	+ DNA (hemimethylated)
6mht	M.*Hha*I	+ AdoMet	+ DNA (4'-thio-2'-deoxycytidine)
7mht	M.*Hha*I	+ AdoHcy	+ DNA (mismatch G:A)
8mht	M.*Hha*I	+ AdoHcy	+ DNA (mismatch G:U)
9mht	M.HhaI	+ AdoHcy	+ DNA (abasic oligonucleotide G:AP)
10mht	M.*Hha*I	+ AdoHcy	+ DNA (hemimethylated, 5'- azacytidine)
1fjx	M.*Hha*I (T250G)	+ AdoHcy	

The catalytic domain is largely an α/β structure, comprising of about two-thirds of the protein from N-terminal. It consists of a core seven β-sheet structure around which the six helices are folded. One of these helices comes from the C-terminus. The core β-sheet structure is now recognized as the 'AdoMet-dependent methyltransferase fold' based on the presence of similar structural fold in different types of AdoMet-dependent methyltransferases (Cheng and Roberts, 2001). The methyltransferase fold basically comprises of seven β-sheets, six parallel, and one antiparallel. In the case of M.*Hha*I, the sheets are arranged in the following manner: (6↑ 7↓ 5↑ 4↑ 1↑ 2↑ 3↑), the seventh sheet being in an antiparallel orientation (Plate 1). The smaller domain consists of five sheets arranged in an antiparallel fashion (↑↓↑↓↑), forming a structure that resembles the blades of a propeller. It consists of two glycine-rich sequences (233–240, and 250–257) that form surface loops between the β-strands facing the cleft from the small domain and contact the major groove of the DNA (Cheng *et al.*, 1993a; Klimasauskas *et al.*, 1994; Cheng, 1995).

A. AdoMet Binding Motif

The AdoMet binding pocket primarily consists of amino acids from motif I (FxGxG) and also residues from motif II, III, and V. All these are present in the large domain of M.*Hha*I and form a pocket adjacent to the catalytic site facing the cleft between the two domains. Motif I, which was implicated in AdoMet binding long before structural studies were initiated (Ingrosso *et al.*, 1989; Klimasauskas *et al.*, 1989), forms a tight loop in the first turn of a β_1-α_A-β_2 structural unit and appears to be important in the correct positioning of AdoMet. The glycine residues play an important role in AdoMet binding because substitutions at this position with bulky residues as in the *Bacillus* phage SPR methyltransferase, resulted in a loss of methylation activity. Although some of the motifs, II, III, and V are not absolutely conserved among the C5-cytosine MTase family they contain amino acids that are highly conserved and are quite tolerant to altered flanking sequences (Kumar *et al.*, 1994). In the tertiary complex of M.*Hha*I-AdoMet-DNA, the adenosyl moiety of AdoMet is surrounded by the hydrophobic amino acids Phe18 (motif I), Trp41 (motif II), Pro80 (motif IV), and Leu100 (motif V). The adenosyl ring is sandwiched between Phe18 and Leu100 residues on one side and Trp41 on the other side. Phe18 interacts with AdoMet, with its ring perpendicular to that of the adenosyl ring (Plate 3). Apart from these hydrophobic residues, acidic amino acids Glu40 and Asp60 also interact with AdoMet. The side chain of Asp 60 makes contacts with the N6-amino group of the adenosyl moiety, whereas Glu40 is hydrogen bonded to the ribose hydroxyl groups. Besides these amino acids described above, for which an obvious role can be proposed, the AdoMet binding pocket is comprised of several other residues that could be play an important role in the binding event. These include the side chains of Asn304 (motif X) and Pro80 and the main chain atoms of Leu21, Gly23, and Ser305. Site-directed mutagenesis of residues involved in AdoMet binding in M.*Hha*I demonstrated that the enzyme is tolerant to changes at these residues (Sankpal *et al.*, 2002). Mutagenesis at Phe18, Trp41, Asp60, and Leu100, which are fairly conserved among the methyltransferase families, did not have drastic effects on activity, thus suggesting that methyltransferases may be using alternate modes of interaction with cofactor (Sankpal *et al.*, 2002).

The AdoMet binding pocket, comprising of residues from motif I, II, III, and V has a strong resemblance to the Rossmann fold of the dinucleotide-binding motif that is characterized by a β-α-β structure. This β-α-β structure contains a glycine-rich turn and acidic amino acids at similar positions. This is the basis for a hypothesis that the NAD-dependent enzymes and the C5-MTase could have a common evolutionary origin (Cheng *et al.*, 1993b; Schluckebier *et al.*, 1995; Fauman *et al.*, 1999).

Comparison of the initial structures, the binary M.*Hha*I-AdoMet complex (Cheng *et al.*, 1993a), and the tertiary M.*Hha*I-AdoMet-DNA complex (Klimasauskas *et al.*, 1994), showed that the orientation of AdoMet in both these structures is quite different. In fact, comparison of the two binary complexes of M.*Hha*I and AdoMet (1 mhy) and M.*Hha*I-AdoMet in the presence of nonspecific DNA (2 mhy) also showed a difference in AdoMet orientation (O'Gara *et al.*, 1999) (Plate 4). Although nonspecific DNA was used during crystallization, no DNA was found in the 2 mhy crystal structure. The orientation of AdoMet in the 2 mhy structure is actually equivalent to that in the tertiary structure. Comparison of the two structures revealed that the

AdoMet molecule was oriented 180 degrees with respect to each other in the two structures (O'Gara *et al.,* 1999). The two orientations are termed primed and unprimed, on the basis of the presence or absence of DNA in the crystal structure (or during crystallization). In the primed orientation the methyl group is closer to the catalytic site and is in an active conformation (Kumar *et al.,* 1997; O'Gara *et al.,* 1999). In the unprimed orientation observed in the initial binary complex the AdoMet is in an inactive conformation. The positive charged sulfonium ion S^{+}-CH_3 is in a favorable orientation with respect to the π ring of Trp41 for cation-π interactions (Dougherty, 1996). Although the significance of these two orientations is debatable, it has been proposed that the unprimed orientation is not a random event and might have a physiological role in sequestering AdoMet and preventing its hydrolysis. The implications of these findings on the catalytic mechanism of M.*Hha*I will be dealt with in later sections.

A recent study (Swaminathan *et al.,* 2002) has shown that the existence of a dual mode of interaction of M.*Hha*I with cofactor AdoMet. The two modes were observed in an isothermal titration calorimetry analysis of cofactor binding, as an exothermic and an endothermic mode. The titration of AdoMet with M.*Hha*I initially shows the presence of a high-affinity exothermic mode, which is rapidly saturated. This is followed by a low-affinity endothermic mode that does not get saturated even at the highest concentration of AdoMet used. The dual mode is conserved with the reaction end product AdoHcy. Also, the conservation of the dual mode in two Trp41 mutants (Trp41Ile and Trp41Tyr) undermines the role of the proposed cation-π interaction between the aromatic ring of Trp41 and the charged sulfonium group of AdoMet (Cheng *et al.,* 1993a; O'Gara *et al.,* 1999). A coupled osmotic-calorimetric analysis demonstrated the involvement of water molecules in the interaction of AdoMet/AdoHcy with M.*Hha*I (8 and 10, respectively).

B. Catalytic Motif

The catalytic motif (motif IV) containing the invariant PC dipeptide is situated in a loop region in the larger catalytic domain. Comparison of the binary and tertiary M.*Hha*I structures revealed that this loop, containing the nucleophilic Cys81, undergoes a massive conformational change toward the DNA binding cleft and lies in the minor groove (Plate 5). Also seen is a small shift in the smaller recognition domain, toward the DNA binding cleft. The movement of the catalytic loop thus ensures that the catalytic nucleophile is in proximity to the target cytosine, allowing a direct attack at the C6 position. For the target cytosine to be accessible for methylation, *Hha*I methyltransferase uses a novel mechanism of base flipping (discussed in later sections).

The cytosine residue flips out of the helix and occupies a position near the catalytic motif. Cys81 and the cofactor AdoMet lie on opposite sides of the flipped out base, suggesting that the thiolate addition and methyl group transfer occur from opposite sides of the ring. Cys81 reacts with the C6 of cytosine resulting in the formation of a covalent complex and activation of carbon at the fifth position. The active site is also composed of residues Phe79 (motif IV), Arg165 (motif VIII), and Glu119 (motif VI) that interact with the cytosine residue. The side chains of these amino acids form hydrogen bonds with the polar groups of cytosine. Both Arg165 and Glu119 are absolutely conserved among the C5-methyltransferase family, while Phe79 is not. The gap created in the DNA helix by the flipped cytosine is occupied by the side chain of

Gln237, which provides hydrogen bonds to the orphan guanine, thus maintaining helical stacking. Ser87 (motif IV) interacts with the Gln237 stabilizing its conformation. However, mutational studies at the Gln237 did not have any significant effect on the methyltransferase activity, undermining the importance of this residue (Mi *et al.*, 1995). Gln237 was changed to all 19 amino acids, and it was seen that the mutations affected the stability of M.*Hha*I-DNA complex but not the specificity for target sequence, and the mutants retained detectable levels of activity. Glu119 interacts with both N3 and N4 atoms of the cytosine ring. The hydrogen bonding between N3 of cytosine and the carboxylate oxygen of Glu119 occurs through an ordered water molecule (Klimasauskas *et al.*, 1994; Cheng 1995; O'Gara *et al.*, 1996a). Thus, these observations favor the proposed protonation at N3 position (see section on catalysis).

C. DNA Binding Motif

The DNA binding motif of M.*Hha*I is localized to the variable region between motif VIII and IX, essentially comprising the entire small domain of M.*Hha*I. Sequence comparison among methyltrans-ferases predicted a target recognition domain with a consensus $T(V/I/L)X_{(5-6)}(V/L)$ (where X is any amino acid) (Lauster *et al.*, 1989). The TRD of M.*Hha*I comprises mainly of two glycine-rich loops (233–240 and 250–258). The two loops lie within the predicted TRD and one of them (the second loop) has a sequence similar to the consensus sequence. These loops formed between β-strands interact with DNA through the major groove. The minor groove is oriented toward the catalytic domain. The extensive contacts of the second loop (especially the Thr250-Leu251 dipeptide), therefore appear to be important for the correct positioning of the first loop with respect to the target cytosine residue. This then results in the infiltration of the helix by the enzyme and flipping of the target base (Cheng, 1995). The first loop makes six specific contacts with DNA. Gln237 is hydrogen bonded with the orphan guanine, Arg240 makes two contacts with the first guanine base, and Lys234 makes a salt bridge with the phosphate group of the first cytosine on the opposite strand.

Apart from these contacts, there are numerous contacts that the protein makes with the sugar-phosphate backbone (Figure 1). Recent crystal structure of M.*Hha*I with an abasic oligonucleotide demonstrated that the rotation about the sugar phosphate backbone is responsible for base flipping and that the base per se is not required. In light of this finding, the phosphate contacts that the protein makes with the DNA are proposed to play an important role in base flipping (O'Gara *et al.*, 1999).

Thr250, present as a part of the highly conserved dipeptide T(LVI), was subjected to mutagenesis (Lauster *et al.*, 1989) (Kumar *et al.*, 1994; Vilkaitis *et al.*, 2000). No change in the kinetic parameters was seen with mutants containing smaller side chain, whereas substitution with bulky side chain resulted in increased K_m values for both substrates. These results suggested that Thr250 was involved in constraining the conformation of the DNA backbone and the target base during its flipping into the catalytic site of M.*Hha*I (Vilkaitis *et al.*, 2000).

Protein-DNA interactions of M.*Hha*I were studied using both dimethylsulfate protection and hydroxyl-radical footprinting (Renbaum and Razin, 1995a). It was seen that M.*Hha*I covers a stretch of 9 to 10 bases on the sugar-phosphate backbone. The sugar phosphate backbone was proposed to participate in nonspecific interactions. These nonspecific interactions could allow the primary recognition of double-stranded DNA and suggested a processive mode of action for M.*Hha*I (Renbaum and Razin, 1992; Renbaum and Razin, 1995a).

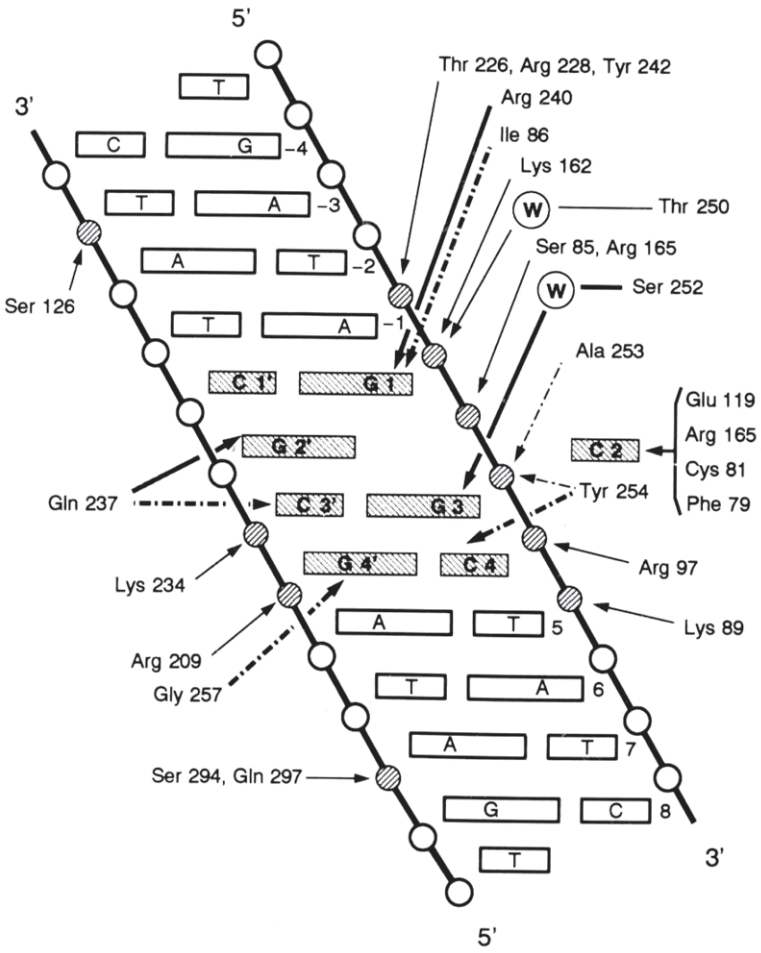

FIGURE 1. Schematic drawing showing the contacts that M.*Hha*I makes with DNA. (Adapted from Klimasauskas *et al.*, 1994.)

V. CHEMISTRY OF DNA METHYLATION

Understanding the chemistry of C5-methylation by M.*Hha*I has been greatly enhanced by structural studies, as well as numerous biochemical studies that have identified the residues involved in the reaction. The methylation reaction can be broadly separated into three steps (1) binding of the substrate DNA and cofactor AdoMet, (2) catalysis (the transfer of methyl group), and (3) product release.

A. Catalysis

Methylation of the cytosine residue at the C5 position seems difficult due to the inert nature of the C5 atom. Earlier biochemical studies by Wu and Santi (Wu and Santi, 1987) led to the proposal that the catalytic mechanism involves the formation of a transient covalent adduct between a cysteine residue of the protein and C6 of target cytosine. The formation of this complex results in the activation of C5 allowing the transfer of methyl group from AdoMet. These results were supported by the observation that DNA containing 5-azacytosine forms covalent complexes with methyltrans-ferases. The mechanism for C5-methyltrans-ferase was proposed to be similar to that of enzymes that catalyze one-carbon transfers to the C5 of pyrimidines, like thymidylate synthase, tRNA-(uracil-5) methylase, and dCMP hydroxymethylase (Santi and Hardy, 1987; Carreras and Santi, 1995). Sequence comparisons among the C5-MTase family had identified an invariant Pro-Cys

dipeptide (Wu and Santi, 1987; Lauster *et al.*, 1989). The thiol group of this dipeptide in M.*Hha*I, Cys81, was suggested to serve as a catalytic nucleophile that attacks the C6 of cytosine resulting in covalent complex formation (Wu and Santi, 1987). Subsequent to the formation of the covalent complex, there is generation of a high-energy carbanion at the C5 position that is either resonance stabilized or is completely avoided by the protonation at the N3 position (Figure 2) (Chen *et al.*, 1993).

It was proposed that methylation occurs preferentially on cytosine residues protonated at N3, as this leads to increased charge at C6 position (Baker *et al.*, 1988). The increased charge at C6 predisposes it to nucleophilic attack by the cysteine thiol group. This leads to the formation of a carbanion, with the concurrent loss of proton from the N3 position. The carbanion is then resonance stabilized, leading to increased charge at C5 and O2 positions. Structural studies show that Arg165 is in close proximity to O2 of cytosine and is likely to accept the charge (Klimasauskas *et al.*, 1994) (Figure 1).

Although it is difficult to distinguish between the two proposed mechanisms, strong evidence for the protonation event exists in the form of crystal structures, in which distance between carboxylate oxygen of Glu119 and N3 of cytosine suggest the presence of a hydrogen bond (O'Gara *et al.*, 1996a). It has been proposed that the protonation of the carboxylate group of Glu119 might be through water-mediated interaction with the amino group of the methionine moiety of cofactors AdoMet or AdoHcy. Therefore, this implies that AdoMet has two roles in the methylation reaction: that of a methyl group donor to the

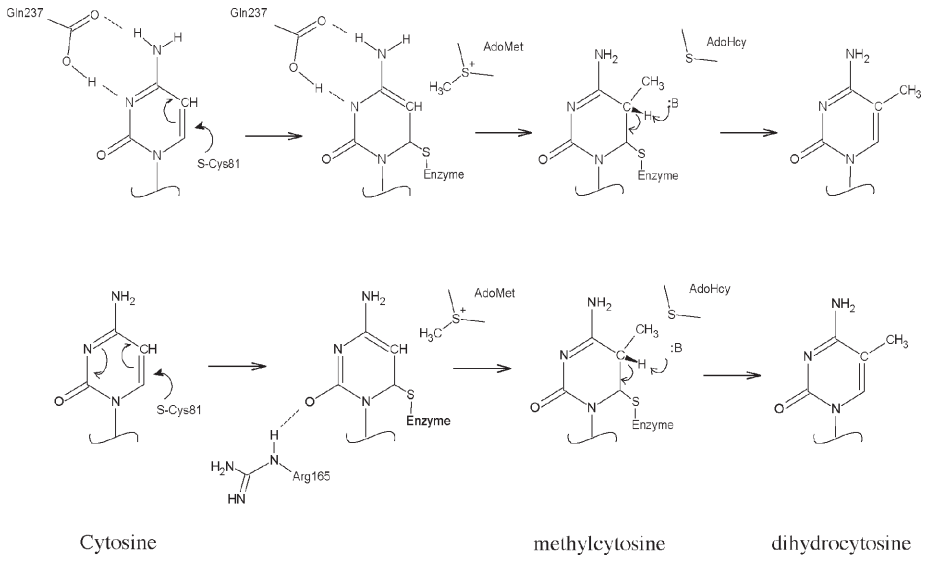

Figure 2. Proposed reaction mechanisms for C5-methylation. The mechanism involves the attack at the C6 position by the catalytic thiol group. This leads to the formation of a covalent enzyme-DNA complex, with the generation of a high-energy carbanion at C5 position. The carbanion is entirely avoided by the protonation at N3 by glutamic acid (**A**) or is resonance stabilized as shown in (**B**). Both pathways finally lead to the formation of a dihydrocytosine intermediate, from which the C5 proton is abstracted resulting in the formation of 5-methylcytosine.

C5 of cytosine and a proton donor to Glu119 (O'Gara *et al.,* 1996a).

A molecular dynamic study was performed on M.*Hha*I to gain insights into the mechanism of cytosine methylation. The simulation provided two possible sources for the protonation of N3 residue (Lau and Bruice, 1999). One was the amine group of AdoMet, which provides a proton via a water bridge, as was proposed from earlier structural data (O'Gara *et al.,* 1996a). An alternate source of proton was the Arg163 residue that could activate a water molecule, which in turn would cause protonation of N3.

The attack at C6 of cytosine by the protein thiol group results in the formation of a covalent intermediate and a reactive enamine. The activated C5 then attacks the methyl group of AdoMet, resulting in the formation of a dihydrocytosine intermediate and the release of AdoHcy. The proton at the C5 position is then abstracted by a basic residue of the enzyme with a simultaneous β-elimination of the enzyme from the C6 position. There are several lines of evidence that support the above-mentioned mechanism. All one carbon transfers at the C5 position of pyrimidines follow a similar mechanism with covalent complex formation at C6. Oligonucleotides containing 5-fluoro-2'-deoxycytidine act as inhibitors of methylation reaction by trapping the enzyme in the covalent intermediate state. This is due to the inability of the final β-elimination step to occur, because fluorine being a highly electronegative atom cannot be abstracted easily (Osterman *et al.,* 1988). Sequence comparison among the C5 methylase family had identified the presence of an absolutely conserved Pro-Cys dipeptide (Lauster *et al.,* 1989). A similar Pro-Cys dipeptide is present in thymidylate synthase, and extensive work had demonstrated its role in catalysis. This was later confirmed by X-ray crystallography. Mutagenesis in the *Bacillus* phage SPR multispecific methylase and *Eco*RII methyltransferase (Wilke *et al.,* 1988; Wyszynski *et al.,* 1992) and photoaffinity crosslinking studies (Som and Friedman, 1990) showed the importance of Cys in catalysis by C5 methyltransferases. Using the mechanism-based inhibitor 5-fluoro-2'-deoxycytidine Chen, (Chen *et al.,* 1991) showed the direct involvement of Cys in catalysis.

In the case of M.*Hha*I, Wu and Santi proposed the involvement of Cys81 in the nucleophillic attack at C6 of cytosine based on precedence in other systems and its susceptibility to modification by *N*-ethylmaliamide (Wu and Santi, 1987; Osterman *et al.,* 1988; Smith *et al.,* 1992). Site-directed mutagenesis of Cys81 to Gly, Arg, His, or Ser resulted in loss of enzyme activity (Mi and Roberts, 1993). The Cys81 to Gly mutant was also found to be cytotoxic to *E. coli* cells, as was the case with analogous mutations in *Eco*RII (Wyszynski *et al.,* 1992). This cytotoxicity was due to its increased affinity for DNA (threefold), which could interfere with cellular processes such as replication or transcription. The final conformation for the role of Cys81in catalysis came for the crystal structure of M.*Hha*I with 5-fluoro-2'-deoxycytidine. It was seen that Cys81 is in close proximity to the C6 atom, clearly indicating a covalent link through thioester bond formation (Klimasauskas *et al.,* 1994).

In all the ternary structures of M.*Hha*I, the hydrophobic edge of the flipped cytosine residue is accessible to the solvent, and water molecules are observed in the vicinity of C5 and C6. Two water molecules were found to interact with the conserved Glu82 and Asn304 and the negatively charged phosphate group. These water molecules are likely candidates for the abstraction of the proton at C5 position of cytosine in a manner similar to that observed in thymidylate synthase (Kalman and Matthews, 1995). Comparison of the structure containing oligonucleotides with FdC

and normal unmethylated oligonucleotide (N13 structure) revealed that the C5 proton must occupy a planar position. Therefore, the planar position makes it inaccessible for abstraction by the thioether group of Cys81, as was proposed earlier (Verdine, 1994). It was shown in the case of *Eco*RII that the Cys81 is unlikely to have any role in proton abstraction (Gabbara *et al.*, 1995). Based on the simulation studies of the methylation reaction catalyzed by M.*Hha*I (Lau and Bruice, 1999), it was proposed that a water network at the active site could provide a means for the C5 proton to diffuse into the solvent. This water network is stabilized by residues Asn304 and Gln82 and at least four water molecules were found to be necessary for the activation and abstraction of the C5 proton.

B. Structural Distortion — Base Flipping

In a normal B-DNA, carbon 5 of the target cytosine is buried too deep in the helix to allow the reaction to proceed. For the reaction mechanism described above to proceed, a distortion in the DNA would have to be evoked. The attack of the thiolate on C6 is blocked by the DNA backbone and the neighboring bases block the trajectory for methyl delivery at C5 (Wu and Santi, 1987; Chen *et al.*, 1991). In order to access the target cytosine, this distortion would have to be in the form of a strand separation event. The strand separation event was proposed to be a localized event, as DNA crosslinked by disulfide bonds was still a substrate for *Hae*III methyltransferase (Erlanson *et al.*, 1993). This assumption was proven by the crystallization of M.*Hha*I with DNA containing the mechanism based inhibitor 5'-fluoro-2-deoxycytidine in the presence of cofactor AdoMet (1mht) (Klimasauskas *et al.*, 1994). The use of 5'-fluoro-2-deoxycytidine resulted in trapping of the enzyme-DNA complex, with the target cytosine in an extrahelical conformation (Plate 6).

It was seen from the ternary structure of M.*Hha*I-DNA-AdoHcy (1 mht), that the target cytosine was flipped out of the DNA helix into a concave active site, with the enzyme linked to DNA through thioether bond to the fluoro deoxycytidine residue. During the process of DNA binding and base flipping, a major conformational change occurs in the protein. The catalytic loop containing the conserved Cys81 moves nearly 2.5 nm toward the cleft and into the minor groove of the DNA, making contacts with the target cytosine (Plate 5). The two domains of M.*Hha*I move slightly closer to each other toward the cleft with the two glycine-rich loops of the variable region making contacts with the helix from the major groove side. Concomitant with base flipping, Gln237 from the first recognition loop infiltrates the DNA duplex and provides hydrogen bonds to the orphan guanine residue. The second recognition loop interacts predominantly with the strand containing the target cytosine. Ser87 from the catalytic loop interacts with Gln237 through hydrogen bonding and stabilizes its conformation. The glycine residues flanking Gln237 and Ser87 appear to be crucial in positioning the side chains of these residues to ensure hydrophobic stacking with the bases of the methylated strand and hydrogen bonding with the complementary strand (Klimasauskas *et al.*, 1994; Kumar *et al.*, 1994).

Solution studies on base flipping in M.*Hha*I were also carried out using NMR as a tool (Klimasauskas *et al.*, 1998). Two ^{19}F nuclei (5-fluorocytosine) were introduced into the DNA at the target position and two bases away from the recognition sequence (as an internal control). Formation of the ternary enzyme-DNA-cofactor complex resulted in the trapping of the target cytosine

in an extrahelical conformation. This was noted by the appearance of a distinct ^{19}F signal, corresponding to the ^{19}F atom at the target position.

Apart from the conformational change occurring in the protein after binding to DNA, there is some amount of distortion in the DNA structure also. Besides the obvious distortion in the form of flipped cytosine, an insignificant distortion is seen in the phosphodiester backbone of the target strand. Both phosphates flanking the target cytosine are shifted away from their corresponding positions in the unbound state. This distortion of phosphates continues along the backbone in the 5' direction. It was proposed that these changes in the phosphates led to an increase in interstrand phosphate group distances, thus facilitating the flipping of cytosine residue through the minor groove of the helix (Klimasauskas *et al.*, 1994).

The C5-methyltransferase family shares a significant amount of similarity at the level of sequence, structure, and function. Therefore, it was not surprising when the phenomenon of base flipping was found in *Hae*III DNA methyltransferase. Crystal structure of M.*Hae*III, a member of the C5-methylase family, was solved in the presence of DNA and cofactor AdoHcy (Reinisch *et al.*, 1995). The enzyme recognizes the sequence 5'-GGCC-3' and methylates the internal cytosine residue. Crystal structure revealed that the protein is organized into a two-domain structure similar to M.*Hha*I. The target cytosine was found to be in an extrahelical position, and in an enzyme pocket. In contrast to the M.*Hha*I structure, significant distortion in the DNA at the recognition site was seen in the M.*Hae*III structure. The distortion was in the form of rearrangement of bases. In the absence of a residue analogous to Gln237 in M.*Hha*I, the orphan guanine of M.*Hae*III, base pairs with the second cytosine residue on the methylated strand. The altered base pairing leaves the first guanine on the opposite strand without a complementary base. This unpaired guanine is stabilized by the interaction with the Arg243. The unusual base pairing of the orphan guanine results in the opening up of a cleft in the helix, into which the side chain of Ile 221 projects (Reinisch *et al.*, 1995).

C. Substrate Specificity

Several crystal structures of M.*Hha*I are available in complex with various forms of DNA substrates and cofactor AdoMet or AdoHcy. All structures contain a 13-mer oligonucleotide with different modifications at the target cytosine within the recognition sequence. The structure with 5-fluoro-2'-deoxycytidine (FdC) was the first structure that clearly demonstrated the phenomenon of base flipping (described above). Subsequently, M.*Hha*I was crystallized in the presence of an unmethylated (N13), fully methylated (M13) and hemimethylated oligonucleotide (HM13) (O'Gara *et al.*, 1996a and b). In all these structures, M.*Hha*I was essentially organized in a similar manner. The phosphate backbone of these structures could be superimposed. The target nucleotide in each of these structures was flipped out of the helix and found to occupy the enzyme active site. It was surprising to see that the methylated cytosine (reaction end product) of the fully methylated oligonucleotide was also flipped out analogous to the cytosine residue in the unmethylated oligonucleotide. In both the N13 and M13 complexes, there is a continuous electron density between Cys81 and C6 position of cytosine, suggesting a partial covalent bond formation (O'Gara *et al.*, 1996a and b). These structures essentially validated the earlier proposals about the involvement of Cys81 in catalysis.

Analysis of the M13 structure revealed that the methyl group of the flipped methylcytosine residue was oriented below the plane of the ring toward the sulfur atom of AdoHcy. The catalytic Cys81 was located on the opposite face of the ring. The space between the sulfur atom of AdoHcy and C6 of cytosine was sufficient to hold a single methyl group. This could be one of the reasons for the absence of a corresponding structure in the presence of AdoMet as cofactor and could also contribute to a mechanism by which AdoMet distinguishes between methylated and unmethylated DNA.

An important aspect of DNA methyltransferases is their ability to distinguish between hemimethylated and unmethylated DNA. Most bacterial enzymes (type II methyltransferases) function as maintenance and *de novo* methyltransferases, that is, they act on both hemimethylated and unmethylated DNA. However, it has been observed that methyltransferases show stronger binding to hemimethylated DNA over unmethylated DNA in the presence of cofactors, AdoHcy, or sinefungin, [M.*Hha*I (Klimasauskas and Roberts, 1995a); M.*Eco*RV (Szczelkun and Connolly, 1995); M.*Hae*III (Chen *et al.*, 1993); M.*Msp*I (Dubey and Roberts, 1992); M.*Eco*RI (Reich and Mashhoon, 1990)]. The mechanistic clues for the preference of hemimethylated DNA were obtained by the analysis of the structure of M.*Hha*I with a 13 mer, nonpalindromic hemimethylated substrate (HM13), and AdoHcy (O'Gara *et al.*, 1996b). The two strands are nonequivalent because only the unmethylated cytosine flips out, whereas the 5-methyl cytosine on the complementary strand remains within the helix. The methyl group of this 5-methylcytosine (5MC) interacts with the side chain of Glu239 residue through van der Waals' contacts. The methyl group of 5MC also interacts through van der Waals' contact, with the main-chain carbonyl oxygen of Gln237, which is hydrogen bonded to the N4 amino group of 5MC. Therefore, it was proposed that these interactions with (N4) amino group and the methyl group of 5MC might enable the enzyme to distinguish methylated and unmethylated cytosine. Hence, the preference for hemimethylated DNA over unmethylated DNA largely results from van der Waals' contacts between the planar Glu239 carboxylate and the methyl group of the intrahelical 5MC. Using the HM13 structure, the electrostatic potential for the protein-DNA complex was generated by partial atomic charges of AdoHcy and the formal charge of +1e for AdoMet. It was proposed that the increase in DNA binding affinity results from long-range electrostatic interaction between cofactor and the negatively charged phosphate (O'Gara *et al.*, 1996b). A similar electrostatic effect on DNA binding by AdoMet has been reported for the *E. coli* methionine repressor (Phillips and Phillips, 1994). Studies with M.*Rsr*I demonstrate that 5-methylthioadenosine is also capable of increasing the enzymes affinity for the target (Scavetta *et al.*, 2000).

D. Structural Studies with Base Analogs

An analysis of crystal structures F13, N13, M13, and HM13 suggests that M.*Hha*I enzyme is indifferent to the nature of base present at the target position. Therefore, this allows for the incorporation of nucleoside analogs into synthetic oligonucleotides for probing structural and chemical interactions involved in sequence-specific recognition and catalysis. *Hha*I has been crystallized in the presence of two such analogs: 4'-thio-2'-deoxycytidine (Kumar *et al.* 1997) and 5,6-dihydro-5-azacytosine (Sheikhnejad *et al.*, 1999).

The methyltransferase reaction was strongly inhibited by oligonucleotides containing 4'-thio-2'-deoxycytidine at the target

position. However, the same oligonucleotide was susceptible to cleavage by the cognate endonuclease R.*Hha*I. The inhibitory effect could be partially overcome by the long time in the crystal environment. DNA binding studies using oligonucleotides containing 4'-thio-2'-deoxycytidine showed that binding of M.*Hha*I to the oligonucleotide was not affected. On the other hand, the stability of the protein DNA complex was greatly reduced, as seen by the increased off-rate of the bound oligonucleotide. Although the rate of base flipping was not measured, the base flipping process per se was not affected. The crystal structure of M.*Hha*I in complex with 4'-thio-2'-deoxycytidine containing oligonucleotide in the presence of AdoMet clearly showed the flipped out nucleotide, with no apparent distortion introduced into the DNA. However, the presence of thio group at the 4'-deoxyribose atom seemed to sterically hinder its interaction with Arg165. The interaction of this residue with the O2 carbonyl and charged phosphate group is thought to play an important role in the methylation reaction pathway. The distances between the methyl group and sulfur atom of AdoMet or distances between the methyl group and C5 carbon are indistinguishable, therefore suggesting the presence of a partial covalent bond. A similar situation also arises with the distance between Cys81 thiol group and the C6 atom, suggesting partial covalent character. These results and the fact that the increased off-rate occurred only in the presence of AdoHcy and not AdoMet suggests that the effect of the thio substitution is on steps prior to methyl transfer.

The enzymatic mechanism of M.*Hha*I involves the formation of a transient dihydrocytosine intermediate, linked to the active site cysteine (Wu and Santi, 1987; Klimasauskas *et al.*, 1994). The cytosine analog 5,6-dihydro-5-azacytosine, having an sp^3-hybridized carbon at C6 and a NH group at position 5, would resemble this intermediate. The presence of this analog in oligonucleotides resulted in the inhibition of methylation reaction. This oligonucleotide was found to be more potent than the 5-fluorocytosine substituted oligonucleotide, forming tight reversible complexes with M.*Hha*I. M.*Hha*I was crystallized in the presence of this oligonucleotide and cofactor AdoHcy, and it was seen that the dihydro azacytosine residue was flipped out of the helix as observed with the earlier structures containing cytosine, methylcytosine, or fluorocytosine. An analysis of this structure revealed that there was no covalent bond formation between the active site Cys81 residue and the C6 carbon of the flipped base. The strong affinity of M.*Hha*I for this analog, which mimics the transition state, is the cause for its strong inhibitory effect on methylation.

E. Mechanism of Base Flipping

While the work with the crystallization of M.*Hha*I was in progress, there were reports on the interaction of M.*Hha*I with oligonucleotides containing a mismatch at the target position (Klimasauskas and Roberts, 1995a; Klimasauskas and Roberts, 1995b; Renbaum and Razin, 1995b; Yang *et al.*, 1995). It was demonstrated using gel mobility shift assays that M.*Hha*I was capable of binding to mismatched substrates with greater affinity. The target cytosine was replaced with adenine, guanine, thymine, uracil, and 5-methylcytosine. It was observed that the binding affinities (m5C < C< G, A < T, U < gap) were inversely correlated to the strength of the base pairs, as judged from the T_m values (m5C > C >> G, A, T). In other words, the weaker the base pairing at the target position, the stronger the binding. Similar observations were made with

substrates having substitutions at the guanine residue (Y:C) of the target G:C base pair. AdoMet and the reaction end product AdoHcy are known to favor the formation of specific ternary complexes (Dubey and Roberts, 1992; Chen *et al.*, 1993; Wyszynski *et al.*, 1993). The presence of AdoHcy selectively enhanced complex formation with cytosine and to a lesser extent with uracil or methylcytosine at the target position. Although M.*Hha*I was able to bind mismatch substrates with greater affinity, it specifically methylated only the cytosine residue. The methylation of uracil (conversion to thymine) was observed, albeit with lower efficiency. The Y:C oligonucleotides were all good substrates for the methytransferase reaction, with a slight variation in their methylation efficiencies. No apparent correlation could be drawn with either base pair strength or with the formation of binary and ternary complexes. However, it has been reported that the human DNA methyltransferase shows increased methylation in the presence of mismatches (Smith *et al.*, 1987; Tan and Li, 1990; Smith *et al.*, 1991). The ability of M.*Hha*I to interact strongly with mismatch substrates can be rationalized thermodynamically. The energy required to disrupt the G:C base pair is offset by the total energy gained after the formation of protein DNA complex. The weaker the base pair, the lesser the energy required to disrupt it and hence tighter the binding, and therefore the inverse correlation between base pair strength and binding affinity. Also, the fact that the enzyme binds to gap (abasic site) substrates suggests that the interactions of the catalytic site with the flipped base have little effect on binding specificity.

In a parallel study with M.*Hha*I and M.*Hpa*II, it was shown that these C5-methyltransferases are capable of sequence specific, hydrolytic deamination of target cytosine, and this was proposed to be responsible for the high mutability of CpG dinucleotides in DNA (Shen *et al.*, 1992; Yang *et al.*, 1995). The deamination process results in the production of premutagenic DNA mismatches, G:U and G:T. It was observed that both M.*Hha*I and M.*Hpa*II were capable of binding these mismatches with increased affinity and were proficient in methylating uracil, leading to the creation of thymine. *In vitro* repair of the G:U mismatch was inhibited by the binding of methyltransferase to the mismatch sequence.

The studies with mismatch oligonucleotides do not provide direct evidence for base flipping, although it seems likely to occur. The crystal structures of M.*Hha*I complexed with the mismatches, G:U, G:A and G:AP (AP — abasic or apuric/apyrimidinic) are now available (O'Gara *et al.*, 1999). All the structures clearly revealed that the mismatched adenine, uracil, and the abasic sites were completely flipped out from the DNA helix and found in the enzyme active site. The structures were essentially identical to previous structures with normal substrate, cytosine.

The adenine of the G:A mismatch interacts with the amino acids of M.*Hha*I through hydrogen bonds and hydrophobic interactions. The backbone carbonyl oxygen of Phe79 and carboxylic oxygen of Glu119 interacts with the N6 amino group and N1 atom through hydrogen bonds, respectively. The side chains of Gln82 and Asn304 interact with the N7atom through two water molecules. The sulfur atoms from Cys81 and AdoHcy approach the C8 and N7 atoms from opposite directions. The fact that adenine can flip out of the helix and interact with the active site amino acids is not surprising. This is because a comparative analysis of C5 cytosine, N4 cytosine, and N6 adenine methyltransferases revealed comparable folding, almost identical cofactor binding sites as well as a remarkable

similarity in the spatial orientation of the active site amino acids (Malone *et al.*, 1995; Schluckebier *et al.*, 1995; Gong *et al.*, 1997). The flipped uracil of the G:U mismatch interacts with M.*Hha*I in a manner similar to the flipped cytosine. The sulfur atom of Cys81 and the C6 of uracil are close enough to possibly catalyze the methylation reaction. The O4 atom of the uracil is located out of the plane of the uracil ring, which could be a reason for the low efficiency of methylation reaction (Klimasauskas and Roberts, 1995a; Yang *et al.*, 1995).

The most surprising finding, however, was from the structure of M.*Hha*I with G:AP mismatch, which showed that the deoxyribose is flipped out and interacts with M.*Hha*I through two sets of water molecules. This water network connects the 5' phosphate, the O4 of deoxyribose and the side chain of Arg165. The Glu119 and Arg165 ion pair observed in the binary M.*Hha*I-AdoMet complex is preserved in the complex with G:AP mismatch, suggesting that formation of active site requires a conformational change, and does not indicate a preformed active site complex. The fact that the abasic oligonucleotide also flips out indicates that the base per se in not essential for the structural change in the DNA. Therefore, it can be concluded that the enzyme targets the backbone for rotation and the base is merely carried along with it. Similar results were observed in several glycosylase-DNA complexes, *E. coli* mismatch-specific uracil glycosylase (MUG), uracil DNA glycosylase (UDG), alkyladenine glycosylase (AAG), and the alkylation glycosylase (AlkA) wherein the backbone of AP site was in a flipped conformation (Barrett *et al.*, 1998; Lau *et al.*, 1998; Parikh *et al.*, 1998; Barrett *et al.*, 1999; Hollis *et al.*, 2000).

A NMR approach was also undertaken to study the behavior of the target cytosine during its interaction with M.*Hha*I in solution state (Klimasauskas *et al.*, 1998). This study demonstrated that the interaction of M.*Hha*I with DNA leads to a dynamic equilibrium between an ensemble of flipped out forms of cytosine (complex II and III) and the stacked state (complex I). The data also suggest that there is no specific recognition or trapping of the target cytosine by M.*Hha*I in the absence of cofactor. The results correlate with the earlier observation that complex formation is insensitive to the target base, but is inversely correlated to the base pair strength (Klimasauskas and Roberts, 1995a; Yang *et al.*, 1995). The result was also supported by gel mobility shift assays where two complexes could be resolved in the absence of cofactor (II and III). The addition of cofactor AdoHcy resulted in the trapping of the cytosine in a unique extrahelical conformation, as seen by the appearance of a unique ^{19}F signal, with the protein also undergoing compaction (Klimasauskas *et al.*, 1998).

Base flipping observed with M.*Hha*I is now a universally acceptable phenomenon for enzymes that perform chemistry on DNA bases (Roberts and Cheng, 1998; Cheng and Roberts, 2001). It has been shown directly or indirectly to occur in a number of DNA methyltransferase and DNA repair enzymes. The surprising results of M.*Hha*I binding to mismatches (Klimasauskas and Roberts, 1995a; Yang *et al.*, 1995), and the subsequent structure demonstrating flipping of the mismatched base (O'Gara *et al.*, 1998) led to the proposal that the methyltransferases could have evolved from DNA mismatch binding proteins (Roberts, 1995; Roberts and Cheng, 1998). The ancient repair systems would have had individual polypeptides carrying out different steps during repair. It was proposed that the methyltransferases could have evolved by the combination of domains involved in the recognition of mismatch binding, sequence specific recognition, and AdoMet binding (Roberts and Cheng, 1998).

F. Base Flipping — Biochemical Studies

Although the ultimate evidence for base flipping comes from structural studies, a number of other approaches are available that would indirectly indicate the possible occurrence of base flipping. It was shown that both M.*Hha*I and M.*Hpa*II bind to mismatched oligonucleotides with higher affinity (discussed earlier) (Klimasauskas and Roberts, 1995a; Yang *et al.*, 1995). Binding to substrates with reduced hydrogen bonding at the target site is a strong indication for base flipping. Similar studies were reported for M.*Eco*P15I and M.*Eco*RV (Reddy and Rao, 2000; Gowher and Jeltsch, 2000). These studies indicate that although the methylation reaction is sensitive to the base at the target position, the base flipping process is not. Therefore, this allows the use of fluorescent base analog 2-aminopurine at the target position, to study conformational changes in DNA (base flipping) (Millar, 1996). The fluorescence of 2-aminopurine is quenched in a double-stranded context due to base stacking interactions and would increase if the base were flipped out of the helix or in a single-stranded conformation. The excitation and emission maxima of 2-aminopurine are separated from that of the aromatic amino acids tyrosine and tryptophan. Using oligonucleotides containing 2-aminopurine at various positions (5'-2CGC-3', 5'-G2GC-3' and 5'-GC2C-3'), titrations were carried out with M.*Hha*I (Holz *et al.*, 1998). It was seen that the addition of M.*Hha*I leads to a dramatic increase (54-fold) in fluorescence only when 2-aminopurine is present at the target position (5'-G2GC-3'). Because structural studies for M.*Hha*I had shown that base flipping results in complete removal of stacking interactions, the increase in 2-aminopurine fluorescence therefore is entirely due to the base-flipping phenomenon. Analogous titration studies performed with M.*Taq*I also showed an increase in fluorescence with oligonucleotides containing 2-aminopurine at the target position. Recently, the crystal structure of M.*Taq*I was solved in the presence of DNA and cofactor, and as expected from the 2-aminopurine studies the target base was found in an extrahelical environment (Goedecke *et al.*, 2001). Duplex oligonucleotides containing 2-aminopurine have been used to study M.*Eco*RI (Allan and Reich, 1996; Allan *et al.*, 1998), M.*Eco*P15I (Reddy and Rao, 2000), M.*Eco*RV (Gowher and Jeltsch, 2000), M.*Rsr*I (Szegedi *et al.*, 2000), and T4 dam (Malygin *et al.*, 2001). An increase in fluorescence is not always an indication of base flipping, because an increase was also observed when 2-aminopurine was present away from the target site in M.*Eco*P15I (Reddy and Rao, 2000) and M.*Eco*RV (Gowher and Jeltsch, 2000), suggesting a local DNA distortion other than base flipping. Also, only a small increase in fluorescence was seen with M.*Eco*RV and M.*Rsr*I when the analog was present at the target position (Gowher and Jeltsch, 2000; Szegedi *et al.*, 2000).

The structural peculiarities of pyrimidine bases in DNA can be probed by various chemical methods. Potassium permanganate was used as an efficient probe for the display of the flipped out base (mismatched thymine) in M.*Hha*I-DNA complexes (Serva *et al.*, 1998). Oxidation of thymine with $KMnO_4$ leads to the formation of *cis*-thymine glycol, which then undergoes cleavage after piperidine treatment. Solvent accessibility of the thymine residue is important, because the attack is at the C6-C5 double bond, and therefore double-stranded DNA is relatively resistant to $KMnO_4$ oxidation. Experiments with mismatched oligonucleotide (G**T**GC) revealed an enhanced susceptibility of thymine residues after modifi-

cation with $KMnO_4$. The addition of M.*Hha*I selectively enhanced the reactivity of the target base. These results are consistent with structural data and 2-aminopurine studies and demonstrate the use of potassium permanganate as a probe to study base flipping. Similar experiments were done with the M.*Taq*I and M.*Eco*P15I (both belonging to the *N*6-methyltransferase family), which showed increased reactivity of thymine at the target position, suggesting that the target base does flip out of the helix. This was indeed the case with M.*Taq*I, where direct structural evidence for base flipping was obtained recently (Goedecke *et al.*, 2001).

The base flipping property of M.*Hha*I was used to study the long-range oxidative damage to DNA (Arkin *et al.*, 1997; Holmlin *et al.*, 1997). Oxidative damage to the guanine of GC sequence in DNA occurs due to electron migration through π-stacked DNA base pairs. M.*Hha*I, due to its ability to flip bases, was used to investigate the effect of gaps within the π-stack on long-range oxidative damage (Rajski *et al.*, 1999). Indeed, strong inhibition of damage was observed in the presence of M.*Hha*I due to the insertion of Gln237 in the helix (occupying the position of the flipped cytosine). Mutants of Gln237 were also used in this study, and it was seen that the tryptophan mutant was able to restore the electron migration. These results suggested that proteins can modulate the electron-transfer chemistry in DNA, which could play an important role in cellular events.

Base flipping is a well-characterized phenomenon that is found in several DNA methyltransferases as well as repair enzymes. In general, it has been proposed for enzymes that do chemistry on DNA bases. Although a vast amount of information has accumulated showing directly (structural studies) or indirectly (mismatched /fluorescence base analogs) the occurrence of this phenomenon, it is still not clear whether it is a passive or an active process. The active process implicates a role for the protein in 'pushing' the target base out of the helix and later 'pulling' it back into the helix. Such a push-and-pull mechanism has been proposed for uracil DNA glycosylase (UDG) (Slupphaug *et al.*, 1996). A similar mechanism has been envisaged for M.*Hha*I, where Gln237 pushes the base out of the helix. A three-step pathway has been proposed for DNA methyltransferases. The first step is the recognition of the target base pair. The second step is the increase in interstrand phosphate-phosphate distance, and the third step is the protein-mediated base flipping (Cheng and Blumenthal, 1996). On the other hand, a passive theory suggests that the enzyme captures the base during the normal breathing of the DNA when the base is transiently in an extrahelical conformation. The support for this theory comes form NMR spectroscopy and molecular dynamics of a synthetic duplex DNA (Goljer *et al.*, 1995; Dornberger *et al.*, 1999). Rapid base pair opening dynamics of GC tracts were observed in contrast to the AT tracts. This was proposed to account for the modest deformation of DNA helix during base flipping by M.*Hha*I and M.*Hae*III compared with M.*Eco*RI (A tract type recognition sequence), where a 52° bend in the DNA was observed (Allan *et al.*, 1999).

Structural data available for M.*Hha*I indicate that the enzyme contacts the DNA through the major groove and flips the target cytosine through the minor groove (Klimasauskas *et al.*, 1994). This is in contrast to modeling studies (Ramstein *et al.*, 1988; Chen *et al.*, 1998a,b) that suggest a major groove pathway for base flipping, thus favoring an enzyme-induced base flipping mechanism. However, studies done to calculate the energetic cost of base flipping indicate that although base flipping occurs through the major groove, the flipped base appears to oriented toward the minor groove

similar to the flipped base by M.*Hha*I, with energy costs comparable to premelting thermal fluctuations (Chen *et al.*, 2000). Also, assuming that the flipping process is reversible, the enzyme-induced base flipping through the minor groove imposes restrictions on the movement of the base, therefore suggesting that this pathway is unlikely to occur (Chen *et al.*, 2000). Further studies would be needed to resolve this issue.

VI. KINETIC MECHANISM

Understanding the kinetics mechanism of the methyltransfer reaction by DNA methyltransferases remains an important and challenging problem to investigate. A number of methyltransferases belonging to the C5 cytosine and *N*6-adenine methyltransferase families have been studied. Among the C5-cytosine methyltransferases, the kinetic mechanism has been well studied for M.*Msp*I (Bhattacharya and Dubey, 1999), human Dnmt1 (Bacolla *et al.*, 1999), and the murine Dnmt1 (Flynn and Reich, 1998). The M.*Msp*I and murine Dnmt1 follow an ordered bi-bi mechanism with DNA binding first followed by AdoMet. The human Dnmt1 is proposed to follow a random bi- bi mechanism with either AdoMet or DNA binding first. Although DNA methyltransferases share a significant amount of sequence and structural homology, they display a wide variety of mechanism to interact with their substrates and carry out the methylation reaction (Cheng, 1995; Malone *et al.*, 1995; Cheng and Roberts, 2001). Therefore, no generalizations can be made concerning the kinetics of the methyltransferase catalyzed reactions.

The kinetics of methylation by M.*Hha*I has also been well studied and serves as a paradigm for the C5-methyltransferase family. Wu and Santi first documented the details of the kinetic and catalytic properties of M.*Hha*I (Wu and Santi, 1987). Using poly(dG-dC) as substrate, the M.*Hha*I reaction was shown to follow an ordered bi-bi mechanism, in which DNA binds to the enzyme first followed by AdoMet. The product AdoHcy is released prior to methylated DNA. This conclusion was based on the following results: (1) the methylation reaction was found to decrease progressively as reaction proceeded. This was attributed to competitive inhibition by the end product AdoHcy, (2) initial velocities were determined using various concentrations of [^{3}H]AdoMet and ^{32}P-poly(dG-dC) and double reciprocal plots when analyzed indicated equilibrium ordered mechanism, (3) M.*Hha*I catalyzed the exchange of ^{3}H from C5 position of cytosine with a K_m(DNA) similar to that for the methylation reaction. This suggested the formation of a productive enzyme-DNA complex, (4) the exchange of ^{3}H is inhibited by AdoHcy and shows uncompetitive inhibition with DNA, both in the exchange reaction and methylation reaction. This demonstrates the formation of enzyme-DNA-AdoHcy complex but not the binary enzyme-AdoHcy complex. In other words, both AdoMet and AdoHcy significantly interact with the enzyme-DNA complex bind weakly to the enzyme-methylated DNA complex and do not bind to the free enzyme. In light of these findings, the binary structure of M.*Hha*I with AdoMet was surprising (Cheng *et al.*, 1993a). It was also seen that during the purification of M.*Hha*I on a cation exchange column, the protein eluted in two forms, native M.*Hha*I alone and AdoMet bound M.*Hha*I (Kumar *et al.*, 1992). Therefore, these observations questioned the validity of the mechanism proposed by Wu and Santi.

The mechanism of M.*Hha*I was reinvestigated, and it was proposed that a binary enzyme-AdoMet complex could be formed, but is catalytically incompetent. Steady state initial velocity analysis was

performed under different substrate conditions, using both methylated and unmethylated DNA. With unmethylated DNA, the steady state data fitted into the rapid equilibrium-ordered kinetic mechanism, similar to that proposed by Wu and Santi. This mechanism predicts that either the binary enzyme-AdoMet complex is not formed or if formed is catalytically inactive (Lindstrom *et al.*, 2000). The assumption was tested further by isotope-partitioning studies. M.*Hha*I was preincubated with radiolabeled AdoMet, followed by the addition of DNA with either radiolabeled AdoMet or unlabeled AdoMet. The addition of the labeled AdoMet led to an initial burst in product formation followed by a slower product formation rate. The addition of unlabeled AdoMet and DNA resulted in a reduction in the rate of product formation that was proportional to the dilution of radiolabeled AdoMet. This would mean that either the enzyme-AdoMet complex was not formed, or if formed was catalytically inactive. However, an analysis of the native protein fluorescence in the presence of AdoMet suggested the formation of enzyme-AdoMet complex. The K_d (AdoMet) determined using fluorescence was greater than the K_m (AdoMet), indicating that AdoMet binding is weaker in the binary complex, than in the presence of DNA, that is, ternary complex. Based on these results, it was proposed that M.*Hha*I could form binary complex with AdoMet, but in a catalytically incompetent manner. Therefore, this would, imply that at higher concentrations of AdoMet, a decrease in burst magnitude would be observed. The failure to see any inhibitory effect of AdoMet on the reaction rates was attributed to the rapid association and dissociation rates of AdoMet with M.*Hha*I compared with the dead time of the experiment (Lindstrom *et al.*, 2000). The converse experiment was done to find the competency of M.*Hha*I-DNA complex (molecular partitioning analysis). To preformed M.*Hha*I-DNA complex, excess of DNA (of longer length) and AdoMet were added. The formation of methylated DNA (smaller length) indicated the competency of the M.*Hha*I-DNA complex.

Two distinct orientations of the cofactor are seen in the binary M.*Hha*I-AdoMet complex and the M.*Hha*I-AdoMet complex, crystalized in the presence of DNA (Plate 4) (Cheng *et al.*, 1993b; O'Gara *et al.*, 1999). The latter orientation (primed) is analogous to the orientation of AdoHcy in ternary complexes. This suggested that DNA is required for the correct orientation of AdoMet to form a catalytically active complex. Taken together the study by Lindstrom *et al*. (2000) essentially refined the Wu and Santi mechanism, stating that M.*Hha*I can bind AdoMet to form a binary complex, but DNA was necessary for the proper orientation of the cofactor and for it to be catalytically competent.

One of the major factors cited in favor of an ordered mechanism was the catalytic incompetency of the M.*Hha*I-AdoMet complex. However, in a recent study (Vilkaitis *et al.*, 2001) it was argued that the catalytic incompetency of the binary M.*Hha*I-AdoMet complex was due to the presence of only a small fraction of *M.HhaI* and AdoMet in the binary complex. Thus, the decreased burst was merely a dilution effect, that is, the decrease in rate was proportional to the dilution factor (Swaminathan *et al.*, 2002). Therefore, when the experiment was repeated at concentrations that favored more binary M.*Hha*I-AdoMet complex formation, a significant fraction (up to 7%) was converted into product without going through the association, dissociation, and rebinding steps, suggesting catalytic competency of the complex (Vilkaitis *et al.*, 2001). The results were in contrast to the earlier study that suggested the formation of a catalytically incompetent dead end binary

M.*Hha*I-AdoMet complex (Lindstrom *et al.*, 2000).

Structural studies have revealed that AdoMet binds to M.*Hha*I in two different orientations (primed and unprimed). Fluorescence quenching studies indicated that AdoMet binds *M.HhaI* in one orientation, which probably is the primed orientation. Both AdoMet and AdoHcy gave similar fluorescence result suggesting that they both formed similar reversible complexes. A rapid off rate (20 s^{-1}) was observed for M.*Hha*I-AdoMet complex indicating that the complex was highly unstable, contrary to the highly stable complex, observed during purification (Kumar *et al.*, 1992). This could be due to high protein concentration (above the K_d) or an effect of the cation exchange column. The latter case is likely because no *Hha*I-AdoMet complex was observed even when different purification protocols were used (Wu and Santi, 1988; Lindstrom *et al.*, 2000).

More recently, an alternate mechanism was proposed based on studies with the T4-Dam methyltransferase (Evdokimov *et al.*, 2002). This involves an isomerization of the inactive ternary *Hha*I-DNA-AdoMet complex to an active form, which then follows the Wu and Santi mechanism (Evdokimov *et al.*, 2002). The enzyme is in an inactive form that randomly interacts with substrates AdoMet and DNA to form a ternary E-AdoMet-DNA complex. This undergoes an isomerization to an active form, F-AdoMet-DNA, which carries out subsequent catalytic reactions strictly according to an ordered mechanism as proposed by Wu and Santi. This mechanism is consistent with both the mechanisms of Lindstrom *et al.* (Lindstrom *et al.*, 2000) and Vilkaitis *et al.* (Vilkaitis *et al.*, 2001).

An interesting aspect of methyltransferases is their ability to distinguish between unmethylated, hemimethylated, and methylated DNA. Equilibrium binding studies had shown that the order of binding in the case of M.*Hha*I is hemimethylated > unmethylated >methylated. An analysis of M.*Hha*I structures with hemimethylated or methylated DNA revealed a possible mechanism by which the enzyme discriminates between these substrates (described earlier). Single turnover kinetics was performed with excess substrates using limiting amounts of hemimethylated duplex. The complete methylation of the target strand (unmethylated) was observed, suggesting that the enzyme binds hemimethylated sites in a productive manner with high selectivity. However, contradictory results were observed when similar single turnover experiments were done using unmethylated and hemimethylated DNA (Lindstrom *et al.*, 2000). The ratio of product to substrate for unmethylated DNA was twice that observed for hemimethylated DNA, suggesting that only half of the binding events with hemimethylated DNA were in a productive orientation (Lindstrom *et al.*, 2000). Taking all the results into account, it was proposed that M.*Hha*I follows a partial rapid equilibrium random bi-bi mechanism with DNA at steady state. This mechanism does not support the formation of an incompetent M.*Hha*l-AdoMet complex (Vilkaitis *et al.*, 2001).

A quantitative understanding of the relative contributions of various steps in the catalytic cycle toward k_{cat} can be used to correlate functional aspects with the structural data. Steady-state time course assays revealed an initial burst and subsequent linear increase of product formation, suggesting that product release could be a rate-limiting step. Both single turnover and presteady state burst gave identical k_{chem} values (rate for the methyl transfer step). Also, the k_{off} and k_{cat} values, estimated under steady state conditions, were similar to those obtained by pre-steady state burst analysis. The k_{chem}/k_{cat} ratio of 6–8 (as opposed to 2–3 from an earlier study (Lindstrom *et al.* 2000), therefore suggested

that the methylation occurred rapidly and catalysis was limited by steps after the chemical step. Using substrates containing 5-fluorocytosine (FdC) at the target position, the k_{chem} increased by 400 fold. The effect was enhanced further in the case of the Thr250 mutants (which had increased K_m for AdoMet) (Lindstrom *et al.*, 2000). This suggested that with the FdC substrates the methyltransfer step would be rate limiting and be responsible for the 400-fold reduction in the methylation rate $k_{chem.}$ However, this is not the case with the normal substrate (cytosine), where formation of covalent intermediate or other steps prior to it would be rate limiting (Lindstrom *et al.*, 2000).

VII. CONCLUDING REMARKS AND PERSPECTIVES

Understanding the functioning of methyltransferases has been a topic of intense research in light of the numerous proposed roles for methylation. Among the DNA methyltransferases, the C5 cytosine methylases, and in particular the M.*Hha*I, has been the most extensively studied enzymes. Although a significant amount of our understanding of DNA methylases comes from studies with M.*Hha*I, a complete clear picture is yet to emerge. Structural alignment studies have revealed a remarkable similarity in the orientation of amino acids and motifs among different classes of DNA methyltransferases, suggesting that they share a common domain structure and an evolutionary origin. The conservation of the base flipping mechanism, first identified in DNA methyltransferases, among a variety of repair enzymes also has evolutionary implications. A number of issues pertaining to the *Hha*I methyltransferase need to be looked into (1) detailed mechanistic information for base flipping is lacking, and it is still unclear whether the process is an active or passive one; (2) the functional amino acids in M.*Hha*I are fairly conserved among the C5-methyltransferase family, and therefore a need to biochemically assign a role for them; (3) the kinetic mechanism for the methyltransferase reaction needs to be looked into as contradictory results exists in literature. (Novel approaches will have to be used to elucidate the mechanistic details); (4) thermodynamic studies have revealed a dual mode of interaction of cofactor AdoMet with M.*Hha*I and the involvement of water molecules in cofactor-M.*Hha*I interactions. The role of water molecules in the methyltransferase reaction and in cofactor interaction needs to be investigated further.

The availability of a number of structures for M.*Hha*I brings research on C5 methyltransferases into a new era. Aided by single molecule manipulations, the study of structure-function relationship of these enzymes will pose an exciting challenge for chemists and biologists.

ACKNOWLEDGMENTS

Work on *Hha*I DNA methyltransferase in our laboratory was supported by a research grant from DBT, Government of India. We thank Rich Roberts and New England Biolabs, USA, for all the help and support. D.N.R. thanks the Rockefeller Foundation, USA, for the award of the Biotechnology Career Award Fellowship.

REFERENCES

Allan, B. W. and Reich, N. O. 1996. Targeted base stacking disruption by the *Eco*RI DNA methyltransferase. *Biochemistry* **35**: 14757–62.

Allan, B. W., Beechem, J. M., Lindstrom, W. M., and Reich, N. O. 1998. Direct real time observation of base flipping by the *Eco*RI DNA methyltransferase. *J. Biol. Chem.* **273**: 2368–73.

Allan, B. W., Garcia, R., Maegley, K., Mort, J., Wong, D., Lindstrom, W., Beechem, J. M., and Reich, N. O. 1999. DNA bending by *Eco*RI DNA methyltransferase accelerates base flipping but compromises specificity. *J. Biol. Chem.* **274**: 19269–75.

Arkin, M. R., Stemp, E. D., Pulver, S. C. and Barton, J. K. 1997. Long-range oxidation of guanine by Ru(III) in duplex DNA. *Chem. Biol.* **4**: 389–400.

Bacolla, A., Pradhan, S., Roberts, R. J., and Wells, R. D. 1999. Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated DNA. *J. Biol. Chem.* **274**: 33011–9.

Baker, D. J., Kan, J. L., and Smith, S. S. 1988. Recognition of structural perturbations in DNA by human DNA(cytosine-5)methyltransferase. *Gene* **74**: 207–10.

Balganesh, T. S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K., and Trautner, T. A. 1987. Construction and use of chimeric SPR/phi 3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *EMBO J.* **6**: 3543–9.

Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. 1998. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92**: 117–29.

Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. 1999. Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J.* **18**: 6599–609.

Bhattacharya, S. K. and Dubey, A. K. 1999. Kinetic mechanism of cytosine DNA methyltransferase *Msp*I. *J. Biol. Chem.* **274**: 14743–9.

Carreras, C. W. and Santi, D. V. 1995. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **64**: 721–62.

Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G. G., and Wells, R. D. 1987. Cloning, sequencing, *in vivo* promoter mapping, and expression in *Escherichia coli* of the gene for the *Hha*I methyltransferase. *J. Biol. Chem.* **262**: 4770–7.

Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S., and Verdine, G. L. 1991. Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* **30**: 11018–25.

Chen, L., MacMillan, A. M., and Verdine, G. L. 1993. Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase. *J. Am. Chem. Soc.* **115:** 5318–19.

Chen, Y. Z., Mohan, V., and Griffey, R. H. 1998a. The opening of a single base without perturbations of neighboring nucleotides: a study on crystal B-DNA duplex d(CGCGAATTCGCG)2. *J. Biomol. Struct. Dyn.* **15**, 765–77.

Chen, Y. Z., Mohan, V., and Griffery, R. H. 1998b. Effect of backbone ζ torsion angle on low energy single base opening in B-DNA crystal structures. *Chem. Phys. Lett.* **287**, 570.

Chen, Y. Z., Mohan, V., and Griffey, R. H. 2000. Spontaneous base flipping in DNA and its possible role in methyltransferase binding. *Phys. Rev. E. Stat. Phys. Plasmas. Fluids Relat. Interdiscip. Topics* **62**, 1133–7.

Cheng, X. 1995. Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**: 293–318.

Cheng, X. and Blumenthal, R. M. 1996. Finding a basis for flipping bases. *Structure* **4**: 639–45.

Cheng, X. and Roberts, R. J. 2001. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* **29**: 3784–95.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. 1993a. Crystal structure of the *Hha*I DNA methyltransferase complexed with *S*-adenosyl-L-methionine. *Cell* **74**: 299–307.

Cheng, X., Kumar, S., Klimasauskas, S., and Roberts, R. J. 1993b. Crystal structure of the *Hha*I DNA methyltransferase. *Cold Spring Harb. Symp. Quant. Biol.* **58**: 331–8.

Dong, A., Yoder, J. A., Zhang, X., Zhou, L., Bestor, T. H., and Cheng, X. 2001. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res.* **29**: 439–48.

Dornberger, U., Leijon, M., and Fritzsche, H. 1999. High base pair opening rates in tracts of GC base pairs. *J. Biol. Chem.* **274**: 6957–62.

Dougherty, D. A. 1996. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**: 163–8.

Dubey, A. K. and Roberts, R. J. 1992. Sequence-specific DNA binding by the *Msp*I DNA methyltransferase. *Nucleic Acids Res.* **20**: 3167–73.

Erlanson, D. A., Chen, L., and Verdine, G. L. 1993. DNA methylation through a locally unpaired intermediate. *J. Am. Chem. Soc.* **115**: 12583–12584.

Evdokimov, A. A., Zinoviev, V. V., Malygin, E. G., Schlagman, S. L., and Hattman, S. 2002. Bacteriophage T4 Dam DNA-[*N*6–adenine]methyltransferase. Kinetic evidence for a catalytically essential conformational change in the ternary complex. *J. Biol. Chem.* **277**: 279–86.

Fauman, E. B., Blumenthal, R. M., and Cheng, X. 1999. Structure and evolution of AdoMet-dependent methyltransferases. **In** *S-Adenosylmethionine-Dependent Methyltransferases: Structure and Function*. pp. 1–38. X. Cheng and R. M. Blumenthal, Eds. World Scientific Publishing Co. Pvt. Ltd.

Flynn, J. and Reich, N. 1998. Murine DNA (cytosine-5–)-methyltransferase: steady-state and substrate trapping analyses of the kinetic mechanism. *Biochemistry* **37**: 15162–9.

Friedman, S. 1985. The irreversible binding of azacytosine-containing DNA fragments to bacterial DNA(cytosine-5)methyltransferases. *J. Biol. Chem.* **260**: 5698–705.

Gabbara, S., Sheluho, D., and Bhagwat, A. S. 1995. Cytosine methyltransferase from Escherichia coli in which active site cysteine is replaced with serine is partially active. *Biochemistry* **34**: 8914–23.

Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J., and Weinhold, E. 2001. Structure of the N6–adenine DNA methyltransferase M.*Taq*I in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**: 121–5.

Goljer, I., Kumar, S., and Bolton, P. H. 1995. Refined solution structure of a DNA heteroduplex containing an aldehydic abasic site. *J. Biol. Chem.* **270**: 22980–7.

Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. 1997. Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.* **25**: 2702–15.

Gowher, H. and Jeltsch, A. 2000. Molecular enzymology of the *Eco*RV DNA-(Adenine-*N* (6))-methyltransferase: kinetics of DNA binding and bending, kinetic mechanism and linear diffusion of the enzyme on DNA. *J. Mol. Biol.* **303**: 93–110.

Hanck, T., Schmidt, S., and Fritz, H. J. 1993. Sequence-specific and mechanism-based crosslinking of Dcm DNA cytosine-C5 methyltransferase of *E. coli* K-12 to

synthetic oligonucleotides containing 5–fluoro-2'-deoxycytidine. *Nucleic Acids Res.* **21**: 303–9.

Hollis, T., Ichikawa, Y., and Ellenberger, T. 2000. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J.* **19**: 758–66.

Holmlin, R. E., Dandliker, P. J., and Barton, J. K. 1997. Charge transfer through the DNA base stack. *Angew. Chem. Int. Ed. Engl.* **36**: 2714–2730.

Holz, B., Klimasauskas, S., Serva, S., and Weinhold, E. 1998. 2–Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. *Nucleic Acids Res.* **26**: 1076–83.

Ingrosso, D., Fowler, A. V., Bleibaum, J., and Clarke, S. 1989. Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule *S*-adenosylmethionine-dependent methyltransferases. *J. Biol. Chem.* **264**: 20131–9.

Kalman, T. I. and Matthews, D. A. 1995. An analogue of the first covalent intermediate of the thymidylate synthase catalyzed reaction. *FASEB J.* **9**: A1325.

Klimasauskas, S. and Roberts, R. J. 1995a. M.*Hha*I binds tightly to substrates containing mismatches at the target base. *Nucleic Acids Res.* **23**: 1388–95.

Klimasauskas, S. and Roberts, R. J. 1995b. Disruption of the target G-C base-pair by the *Hha*I methyltransferase. *Gene* **157**: 163–4.

Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. 1989. Sequence motifs characteristic of DNA[cytosine-*N*4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res.* **17**: 9823–32.

Klimasauskas, S., Nelson, J. L., and Roberts, R. J. 1991. The sequence specificity domain of cytosine-C5 methylases. *Nucleic Acids Res.* **19**: 6183–90.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. 1994. *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **76**: 357–69.

Klimasauskas, S., Szyperski, T., Serva, S., and Wuthrich, K. 1998. Dynamic modes of the flipped-out cytosine during *Hha*I methyltransferase-DNA interactions in solution. *EMBO J.* **17**: 317–24.

Kumar, S., Cheng, X., Pflugrath, J. W., and Roberts, R. J. 1992. Purification, crystallization, and preliminary X-ray diffraction analysis of an M.*Hha*I-AdoMet complex. *Biochemistry* **31**: 8648–53.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., and Wilson, G. G. 1994. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **22**: 1–10.

Kumar, S., Horton, J. R., Jones, G. D., Walker, R. T., Roberts, R. J., and Cheng, X. 1997. DNA containing 4'-thio-2'-deoxycytidine inhibits methylation by *Hha*I methyltransferase. *Nucleic Acids Res.* **25**: 2773–83.

Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I., and Saenger, W. 1994. Three-dimensional structure of the adenine-specific DNA methyltransferase M.*Taq* I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 10957–61.

Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. 1998. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**: 249–58.

Lau, E. Y. and Bruice, T. C. 1999. Active site dynamics of the *Hha*I methyltransferase: insights from computer simulation. *J. Mol. Biol.* **293**: 9–18.

Lauster, R., Trautner, T. A., and Noyer-Weidner, M. 1989. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* **206**: 305–12.

Lindstrom, W. M., Jr., Flynn, J., and Reich, N. O. 2000. Reconciling structure and function in *Hha*I DNA cytosine-C-5 methyltransferase. *J. Biol. Chem.* **275**: 4912–9.

Malone, T., Blumenthal, R. M., and Cheng, X. 1995. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**: 618–32.

Malygin, E. G., Evdokimov, A. A., Zinoviev, V. V., Ovechkina, L. G., Lindstrom, W. M., Reich, N. O., Schlagman, S. L., and Hattman, S. 2001. A dual role for substrate *S*-adenosyl-L-methionine in the methylation reaction with bacteriophage T4 Dam DNA-[*N*6–adenine]-methyltransferase. *Nucleic Acids Res.* **29**: 2361–9.

Mi, S. and Roberts, R. J. 1992. How M.*Msp*I and M.*Hpa*II decide which base to methylate. *Nucleic Acids Res.* **20**: 4811–6.

Mi, S. and Roberts, R. J. 1993. The DNA binding affinity of *Hha*I methylase is increased by a single amino acid substitution in the catalytic center. *Nucleic Acids Res.* **21**: 2459–64.

Mi, S., Alonso, D., and Roberts, R. J. 1995. Functional analysis of Gln-237 mutants of *Hha*I methyltransferase. *Nucleic Acids Res.* **23**: 620–7.

Millar, D. P. 1996. Fluorescence studies of DNA and RNA structure and dynamics. *Curr. Opin. Struct. Biol.* **6**: 322–6.

O'Gara, M., Klimasauskas, S., Roberts, R. J., and Cheng, X. 1996a. Enzymatic C5–cytosine methylation of DNA: mechanistic implications of new crystal structures for *Hha*I methyltransferase-DNA-AdoHcy complexes. *J. Mol. Biol.* **261**: 634–45.

O'Gara, M., Roberts, R. J., and Cheng, X. 1996b. A structural basis for the preferential binding of hemimethylated DNA by *Hha*I DNA methyltransferase. *J. Mol. Biol.* **263**: 597–606.

O'Gara, M., Horton, J. R., Roberts, R. J., and Cheng, X. 1998. Structures of *Hha*I methyltransferase complexed with substrates containing mismatches at the target base. *Nat. Struct. Biol.* **5**: 872–7.

O'Gara, M., Zhang, X., Roberts, R. J., and Cheng, X. 1999. Structure of a binary complex of *Hha*I methyltransferase with *S*-adenosyl-L-methionine formed in the presence of a short nonspecific DNA oligonucleotide. *J. Mol. Biol.* **287**: 201–9.

Osterman, D. G., DePillis, G. D., Wu, J. C., Matsuda, A., and Santi, D. V. 1988. 5–Fluorocytosine in DNA is a mechanism-based inhibitor of *Hha*I methylase. *Biochemistry* **27**: 5204–10.

Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. 1998. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**: 5214–26.

Phillips, K. and Phillips, S. E. 1994. Electrostatic activation of *Escherichia coli* methionine repressor. *Structure* **2**: 309–16.

Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. 1989. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**: 2421–35.

Rajski, S. R., Kumar, S., Roberts, R. J., and Barton, J. K. 1999. Protein-modulated DNA electron transfer. *J. Am. Chem. Soc.* **121**: 5615–5616.

Ramstein, J. and Lavery, R. 1988. Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7231–5.

Reddy, Y. V. and Rao, D. N. 2000. Binding of *Eco*P15I DNA methyltransferase to DNA

reveals a large structural distortion within the recognition sequence. *J. Mol. Biol.* **298**: 597–610.

Reich, N. O. and Mashhoon, N. 1990. Inhibition of *Eco*RI DNA methylase with cofactor analogs. *J. Biol. Chem.* **265**: 8966–70.

Reich, N. O. and Mashhoon, N. 1991. Kinetic mechanism of the *Eco*RI DNA methyltransferase. *Biochemistry* **30**: 2933–9.

Reich, N. O. and Mashhoon, N. 1993. Presteady state kinetics of an *S*-adenosylmethionine-dependent enzyme. Evidence for a unique binding orientation requirement for *Eco*RI DNA methyltransferase. *J. Biol. Chem.* **268**: 9191–3.

Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. 1995. The crystal structure of *Hae*III methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**: 143–53.

Renbaum, P. and Razin, A. 1992. Mode of action of the *Spiroplasma* CpG methylase M.*Sss*I. *FEBS Lett.* **313**: 243–7.

Renbaum, P. and Razin, A. 1995a. Footprint analysis of M.*Sss*l and M.*Hha*l methyltransferases reveals extensive interactions with the substrate DNA backbone. *J. Mol. Biol.* **248**: 19–26.

Renbaum, P. and Razin, A. 1995b. Interaction of M.*Sss*I and M.*Hha*I with single-base mismatched oligodeoxynucleotide duplexes. *Gene* **157**: 177–9.

Roberts, R. J., Myers, P. A., Morrison, A., and Murray, K. 1976. A specific endonuclease from *Haemophilus haemolyticus*. *J. Mol. Biol.* **103**: 199–208.

Roberts, R. J. 1995. On base flipping. *Cell* **82**: 9–12.

Roberts, R. J. and Cheng, X. 1998. Base flipping. *Annu. Rev. Biochem.* **67**: 181–98.

Sankpal, U. T. and Rao, D. N. 2002. Mutational analysis of conserved residues in *Hha*I DNA methyltransferase *Nucleic Acids Res.* **30:** 2628–2638.

Santi, D. V. and Hardy, L. W. 1987. Catalytic mechanism and inhibition of tRNA (uracil-5–) methyltransferase: evidence for covalent catalysis. *Biochemistry* **26**: 8599–606.

Scavetta, R. D., Thomas, C. B., Walsh, M. A., Szegedi, S., Joachimiak, A., Gumport, R. I., and Churchill, M. E. 2000. Structure of *Rsr*I methyltransferase, a member of the N6–adenine beta class of DNA methyltransferases. *Nucleic Acids Res.* **28**: 3950–61.

Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**: 16–20.

Schluckebier, G., Kozak, M., Bleimling, N., Weinhold, E., and Saenger, W. 1997. Differential binding of *S*-adenosylmethionine *S*-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M.*Taq*I. *J. Mol. Biol.* **265**: 56–67.

Serva, S., Weinhold, E., Roberts, R. J., and Klimasauskas, S. 1998. Chemical display of thymine residues flipped out by DNA methyltransferases. *Nucleic Acids Res.* **26**: 3473–9.

Sheikhnejad, G., Brank, A., Christman, J. K., Goddard, A., Alvarez, E., Ford, H., Jr., Marquez, V. E., Marasco, C. J., Sufrin, J. R., O'Gara, M., and Cheng, X. 1999. Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6–dihydro-5–azacytosine. *J. Mol. Biol.* **285**: 2021–34.

Shen, J. C., Rideout, W. M., 3rd, and Jones, P. A. 1992. High frequency mutagenesis by a DNA methyltransferase. *Cell* **71**: 1073–80.

Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. 1996. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**: 87–92.

Smith, H. O. 1979. Nucleotide sequence specificity of restriction endonucleases. *Science* **205**: 455–62.

Smith, S. S., Hardy, T. A., and Baker, D. J. 1987. Human DNA (cytosine-5)methyltransferase selectively methylates duplex DNA containing mispairs. *Nucleic Acids Res.* **15**: 6899–916.

Smith, S. S., Kan, J. L., Baker, D. J., Kaplan, B. E., and Dembek, P. 1991. Recognition of unusual DNA structures by human DNA (cytosine-5)methyltransferase. *J. Mol. Biol.* **217**: 39–51.

Smith, S. S., Kaplan, B. E., Sowers, L. C., and Newman, E. M. 1992. Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 4744–8.

Som, S. and Friedman, S. 1990. Direct photolabeling of the *Eco*RII methyltransferase with *S*-adenosyl-L-methionine. *J. Biol. Chem.* **265**: 4278–83.

Swaminathan, C. P., Sankpal, U. T., Rao, D. N., and Surolia, A. 2002. Water-assisted dual mode cofactor recognition by *Hha*I DNA methyltransferase. *J. Biol. Chem.* **277**: 4042–4049.

Szczelkun, M. D. and Connolly, B. A. 1995. Sequence-specific binding of DNA by the *Eco*RV restriction and modification enzymes with nucleic acid and cofactor analogues. *Biochemistry* **34**: 10724–33.

Szegedi, S. S., Reich, N. O., and Gumport, R. I. 2000. Substrate binding in vitro and kinetics of *Rsr*I [*N*6–adenine] DNA methyltransferase. *Nucleic Acids Res.* **28**: 3962–71.

Tan, N. W. and Li, B. F. 1990. Interaction of oligonucleotides containing 6–*O*-methylguanine with human DNA (cytosine-5–)-methyltransferase [published erratumm appears in *Biochemistry* 1992 Aug 4;31(30):7008]. *Biochemistry* **29**: 9234–40.

Tran, P. H., Korszun, Z. R., Cerritelli, S., Springhorn, S. S., and Lacks, S. A. 1998. Crystal structure of the *Dpn*M DNA adenine methyltransferase from the *Dpn*II restriction system of streptococcus pneumoniae bound to *S*-adenosylmethionine. *Structure* **6**: 1563–75.

Trautner, T. A., Balganesh, T. S., and Pawlek, B. 1988. Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Res.* **16**: 6649–58.

Verdine, G. L. 1994. The flip side of DNA methylation. *Cell* **76**: 197–200.

Vilkaitis, G., Dong, A., Weinhold, E., Cheng, X., and Klimasauskas, S. 2000. Functional roles of the conserved threonine 250 in the target recognition domain of *Hha*I DNA methyltransferase. *J. Biol. Chem.* **275**: 38722–30.

Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E., and Klimasauskas, S. 2001. The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of *Hha*I methyltransferase. *J. Biol. Chem.* **276**: 20924–34.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., and Trautner, T. A. 1988. Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J.* **7**: 2601–9.

Wu, J. C. and Santi, D. V. 1987. Kinetic and catalytic mechanism of *Hha*I methyltransferase. *J. Biol. Chem.* **262**: 4778–86.

Wu, J. C. and Santi, D. V. 1988. High-level expression and purification of *Hha*I methyltransferase. *Nucleic Acids Res.* **16**: 703–17.

Wyszynski, M. W., Gabbara, S., and Bhagwat, A. S. 1992. Substitutions of a cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. *Nucleic Acids Res.* **20**: 319–26.

Wyszynski, M. W., Gabbara, S., Kubareva, E. A., Romanova, E. A., Oretskaya, T. S.,

Gromova, E. S., Shabarova, Z. A., and Bhagwat, A. S. 1993. The cysteine conserved among DNA cytosine methylases is required for methyl transfer, but not for specific DNA binding. *Nucleic Acids Res.* **21**: 295–301.

Yang, A. S., Shen, J. C., Zingg, J. M., Mi, S., and Jones, P. A. 1995. *Hha*I and *Hpa*II DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucleic Acids Res*. **23**: 1380–7.