J Mol Model (2010) 16:1213–1222
DOI 10.1007/s00894-009-0632-9

ORIGINAL PAPER

Homology modeling and molecular dynamics simulations of HgiDII methyltransferase in complex with DNA and S-adenosyl-methionine: Catalytic mechanism and interactions with DNA

Juan A. Castelán-Vega · Alicia Jiménez-Alberto · Rosa M. Ribas-Aparicio

Received: 25 September 2009 / Accepted: 23 November 2009 / Published online: 22 December 2009

Abstract M.HgiDII is a methyltransferase (MTase) from *Herpetosiphon giganteus* that recognizes the sequence GTCGAC. This enzyme belongs to a group of MTases that share a high degree of amino acid similarity, albeit none of them has been thoroughly characterized. To study the catalytic mechanism of M.HgiDII and its interactions with DNA, we performed molecular dynamics simulations with a homology model of M.HgiDII complexed with DNA and S-adenosyl-methionine. Our results indicate that M. HgiDII may not rely only on Glu119 to activate the cytosine ring, which is an early step in the catalysis of cytosine methylation; apparently, Arg160 and Arg162 may also participate in the activation by interacting with cytosine O2. Another residue from the catalytic site, Val118, also played a relevant role in the catalysis of M. HgiDII. Val118 interacted with the target cytosine and kept water molecules from accessing the region of the catalytic pocket where Cys79 interacts with cytosine, thus preventing water-mediated disruption of interactions in the catalytic site. Specific recognition of DNA was mediated mainly by amino acids of the target recognition domain, although some amino acids (loop 80–88) of the catalytic domain may also contribute to DNA recognition. These interactions involved direct contacts between M.HgiDII and DNA, as well as indirect contacts through water bridges. Additionally, analysis of sequence alignments with closely related MTases helped us to identify a motif in the TRD of M.HgiDII that may be relevant to specific DNA recognition.

Keywords DNA-methyltransferase · DNA recognition · Homology modeling · M.HgiDII · Molecular dynamics · S-adenosyl-methionine

J. A. Castelán-Vega
Center for Biologics Evaluation and Research, US FDA,
CBER/DBPAP [HFM-443],
1401 Rockville Pike,
Rockville, MD 20852, USA

A. Jiménez-Alberto · R. M. Ribas-Aparicio (✉)
Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional,
Prolongación de Carpio y Plan de Ayala s/n, Distrito Federal,
CP 11340 Mexico, Mexico
e-mail: rmrj@encb.ipn.mx

Present Address:
J. A. Castelán-Vega
Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional,
Prolongación de Carpio y Plan de Ayala s/n, Distrito Federal,
CP 11340 Mexico, Mexico

Introduction

DNA-methyltransferases (MTases) catalyze the transfer of methyl groups from cofactor S-adenosyl-methionine (SAM) to adenines or cytosines in DNA [1]. Some MTases methylate exocyclic amino groups to form either N6-methyl adenine or N4-methyl cytosine, whereas others methylate the C5 carbon to form C5-methyl cytosine (C5mC). These modifications have important functions in the cell, ranging from DNA repair in bacteria, to genetic regulation and embryonic development in vertebrates [2]. Most of the MTases described thus far are from bacterial origin and associate with restriction enzymes to form restriction-modification (RM) systems [3]. These systems are classified in four types (I–IV) [4], of which type II are

the most abundant and most studied systems. Type II restriction enzymes are well known for their relevant applications in biotechnology, whereas type II MTases have been increasingly used to study the effect of methylation on gene expression [5] and for sequence-specific labeling of DNA [6].

In MTases, catalysis and substrate recognition functions are located in two separate domains: the catalytic domain, where the transfer of a methyl group from SAM to the target base occurs, and the target recognition domain (TRD), which determines the DNA sequence specificity. Several conserved amino acid motifs have been identified in the catalytic domain (9 for adenine MTases and 10 for cytosine MTases); these motifs have been implicated in SAM binding (motifs I-III, V, X), in the catalysis of methylation (motifs IV and VI), and in correct positioning of the target nucleotide within the active site of the enzyme (motifs IV, VI, VIII) [7, 8]. Aditionally, a structural motif involving a positively charged or polar group and a hydrophobic residue has been implicated in extrahelical base stabilization and catalysis [9].

A key step in the modification of DNA is translocation of the target base from the double stranded DNA configuration into the catalytic pocket (base flipping). Structural comparisons of M.HhaI (a C5mC MTase), with or without target DNA, showed a large displacement of one protein loop located in the catalytic domain in response to DNA binding [10]. The active role of the protein in base flipping was demonstrated by analysis of free-energy profiles obtained through molecular dynamics simulations in M.HhaI. These profiles showed that the protein lowers the free-energy barrier for base flipping and stabilizes the fully flipped state [11]. It has been proposed that M.HhaI binds to DNA and moves linearly along the DNA; once it reaches its cognate recognition sequence, a loop from the catalytic domain shifts toward the DNA [12, 13]. This rearrangement contributes to the correct assembly of the active site and is coupled with base flipping [14].

The mechanism of methyl group transfer has been widely studied in M.HhaI. Active site residues Glu119 and Arg165 are required to orient and stabilize the cytosine for nucleophilic attack by Cys81 [15, 16]. The role of Glu119 in catalysis is still debated; while computational studies found that Glu119 plays a secondary role in catalysis, only contributing to create the ground state conformation [17], mutational analyses showed that Glu119 is essential in the first steps of the catalysis [15]. According to the latter mechanism, protonation at N3 by Glu119 activates the cytosine ring, allowing the attack at C6 by Cys81 (Fig. 1a), which ultimately leads to attack of C5 to the methyl group of SAM (Fig. 1b). The final reaction step involves elimination of a hydrogen atom from C5. Analysis of molecular dynamics simulations in the catalytic site of M.HhaI showed that a water channel in the active site provides hydroxide anions required to deprotonate the Cys81-cytosine complex (Fig. 1c) and free the methylated cytosine (Fig. 1d) [17, 18].

While the mechanism of methylation is well understood and has been found to be conserved among MTases, specific recognition of DNA is still not well understood due to the high divergence of the TRD, which makes comparative studies difficult to perform [19]. Despite this, MTases that recognize similar sequences show some degree of conservation at the TRD [19, 20]. Moreover, it has been shown that several cytosine MTases contain a conserved motif (TL motif), which is involved in the constraining and positioning of the target cytosine inside the catalytic pocket [21]. Conservation of the amino acids that surround the TL motif of MTases that recognize similar sequences has raised the possibility of using this region as predictor of sequence specificity [22].

M.HgiDII is a C5mC-MTase that recognizes the sequence GTCGAC, and belongs to the HgiDII RM system from *Herpetosiphon giganteus* D [23]. No structural studies have been performed in this enzyme, and although the recognition sequence has been determined, it is unknown

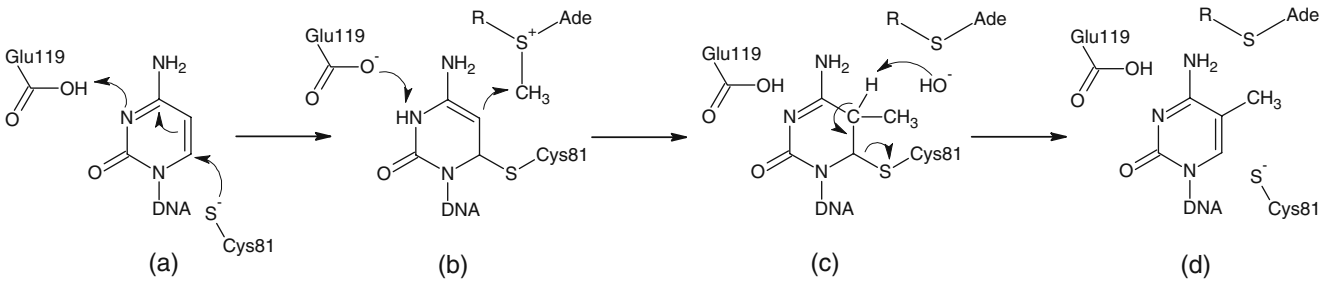


Fig. 1 Methyl-group transfer mechanism of M.HhaI. (**a**) Activation of the cytosine ring by Glu119-mediated protonation at N3 facilitates nucleophilic attack at C6 by Cys81. (**b**) Electron flow to the pyrimidine ring activates cytosine C5, which attacks SAM methyl group. (**c**) A hydroxide anion from water deprotonates the C5 position, causing breakage of the covalent complex between Cys81 and cytosine. (**d**) Final products of the reaction: C5mC, and S-adenosylhomocysteine (SAH). Based on references [15] and [17]

which cytosine from the recognition sequence is methylated. Additionally, several MTases identified in genome sequence projects are highly similar to M.HgiDII at the amino acid level, thus discoveries made in this enzyme could be extrapolated to these MTases. In this work, we performed molecular dynamics simulations with a model of M.HgiDII complexed with SAM and target DNA. Our results indicate that the catalytic mechanism of M.HgiDII differs from that of M.HhaI because interactions at cytosine N3 may not be necessary for the catalytic mechanism. We also found that Val118 may have an important role in the catalysis, by impeding the access of water molecules to the catalytic site. Finally, we identified the amino acids that may be responsible for the specific recognition of DNA.

Methods

Homology modeling of M.HgiDII complexed with DNA and SAM

The templates used to model M.HgiDII (Entrez accession code: P25265) were M.HaeIII complexed with DNA, and M.HhaI complexed with DNA and S-adenosyl-l-homocysteine (PDB accession codes: 1DCT and 3MHT, respectively). The amino acid sequences of M.HgiDII and the templates were aligned with ClustalX [24]. Secondary structure information, predicted by PSI-PRED [25], was used to add structural penalties, thus avoiding the introduction of gaps in α-helices or β-strands. The alignment was modified manually to match all motifs in the three MTases (catalytic motifs and the TL motif) [8, 22].

Homology models were generated with Modeller 9.3 [26], adding secondary structure restraints for the TRD region (amino acids 207–352); loops were optimized with Modeller 9.3. Local energy profile of the initial model was evaluated with the ANOLEA server [27], and manual modifications of the input alignment were done until none of the amino acids in the model had a score higher than 10 E/kT units. DNA coordinates were obtained from 1DCT structure, mutating the sequence GGCCAC to GTCGAC (C is the flipped cytosine) with Chimera [28]. The DNA chain was extended with coordinates from a non-related DNA structure (1BNA). The ternary complex was created by structural alignment of M.HgiDII to M.HhaI and M.HaeIII; DNA coordinates were positioned using their correspondent structures in 1DCT and 3MHT as guide. Finally, SAM was constructed with the AutoPSF plug-in of VMD [29] from an adenine residue positioned in a location equivalent to the adenine moiety from 3MHT and a patch included in the topology and parameter stream files from the c35b2 and c36a2 releases of CHARMM [30, 31]. The model was refined by molecular dynamics simulations (described below). The final structure was obtained by minimizing (10000 steps) the average structure in one nanosecond during the last third of the simulation.

Model quality assessment

Quality of the models was assessed with respect to their energy and stereochemical geometry. Procheck [32] was used to evaluate the geometry, ProSA-Web [33] to evaluate the energy potential, and Verify3D to evaluate the local compatibility of the model relative to a set of good protein structures [34]. ANOLEA was used to calculate local energies according to the "non-local environment" of each heavy atom in the protein [27].

Molecular dynamics simulations

All simulations were performed with NAMD v. 2 [35] installed in the high-performance Biowulf Linux cluster at NIH (http://www.biowulf.nih.gov). CHARMM22 and CHARMM27 [30, 31] force fields were used for the simulations. The all-atom model for the ternary complex was solvated with water (TIP3P model) in a cubic box, forcing a minimum distance between the solute and the box walls of 12 Å; the system net charge was neutralized addition of Na^+ and Cl^- ions randomly placed in the simulation box. The system was simulated under periodic boundary conditions at 310 K, with a cutoff radius of 10 Å for non-bonded interactions and a time step of 2 fs; SHAKE algorithm was applied to constrain the covalent bonds of all hydrogen atoms. Coordinates were saved every 5 ps during equilibration, and every 10 ps in the production run.

The system was equilibrated by fixing the solute and relaxing the solvent with 20,000 minimization steps and 20 ps of simulation time. The solute was gradually relaxed in 10ps runs with decreasing harmonic constraints (100, 10, and 1 kcal mol^{-1} $Å^{-2}$), minimizing for 1000 steps between each simulation period. The relaxed system was simulated for 5 ns at 310 K and a pressure of 1 atm.

Results and discussion

Homology modeling of M.HgiDII complexed with SAM and DNA

We selected M.HaeIII and M.HhaI (1DCT and 3MHT, respectively) as templates for homology modeling of M. HgiDII. These MTases shared a low amino acid sequence identity (approximately 19%) with M.HgiDII, which made the sequence alignment the critical step in the modeling procedure. Based on the knowledge that all MTases described thus far contain several conserved catalytic amino

acid motifs [8], and that all MTases have common structural catalytic domains [7], we included secondary structure penalties in the alignments, and made modifications to match all catalytic motifs in the three MTases. This approach rendered an alignment of the catalytic domain with an identity percentage of approximately 27%, which is better suited for homology modeling [36]. The alignment in the TRD region was modified manually while maintaining the TL motif fixed. We created a structural model for each modification made to the alignment, and continued the modifications until the ANOLEA profile of the resulting model indicated favorable energies in all amino acids. Similar approaches involving reiterate modifications of the alignments based on the score of the resulting models have been successfully applied to model other MTases [37, 38] due to the intrinsic variability in the TRD of these enzymes. The final alignment (Fig. 2) showed a high degree of similarity in the catalytic domain of the MTases at the primary and at the secondary structure level. The amino acid sequence of M.HgiDII is longer than that of the templates, and most of the extra residues lied within the TRD (Fig. 2, cross hatched box).

Although the DNA recognition sequence (GTCGAC) of M.HgiDII has been determined experimentally [23], it is still unknown which cytosine is the target for methylation. Structural comparison of M.HgiDII and the templates, which methylate the inner cytosine of their recognition sequence, showed that the access to the catalytic site lies nearly in the middle of the TRD, suggesting that the inner cytosine was better suited as target for methylation. Thus, we created a DNA structure containing the M.HgiDII recognition sequence with the inner cytosine flipped out of the DNA helix. The complex was subjected to molecular dynamics simulations, and the system was monitored through the evolution of Cα RMSD and radius of gyration (Rg) (Fig. 3). The final model was obtained by averaging and minimizing a 1-ns trajectory in the late term of the simulation (Fig. 3, dashed-line rectangle).

Model quality assessment

Global analysis of the model with ProSA showed a *z* score = -7.18, indicating no significant deviation from typical native structures of similar size as M.HgiDII [33]. The

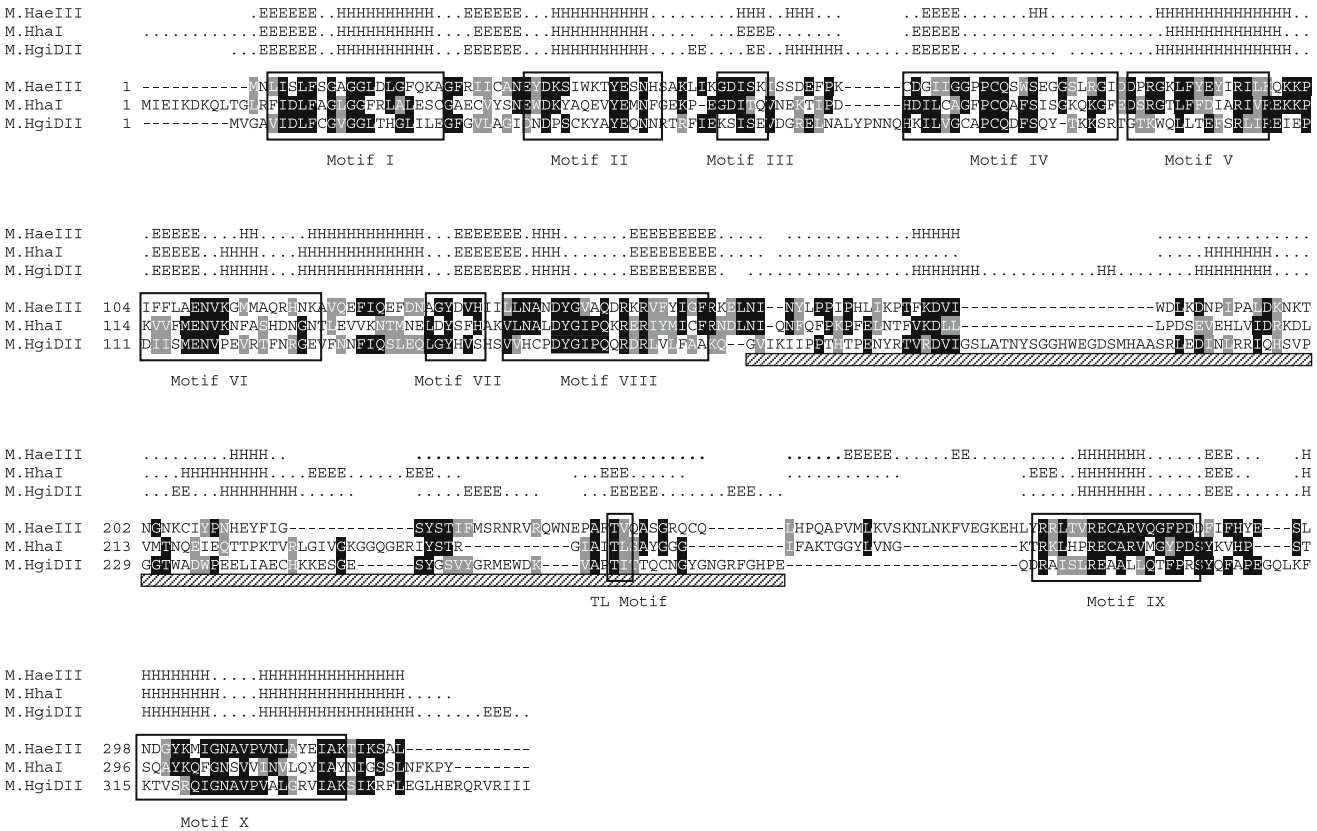


Fig. 2 Multiple sequence alignment of M.HgiDII with templates M. HhaI and M.HaeIII. Secondary structure information is shown above the amino acid sequences (H for α-helix, and E for β-strand). Conserved amino acid motifs and M.HgiDII TRD are shown under the sequences. Alignments were formatted with BoxShade (www.ch.embnet.org/software/BOX_form.html)

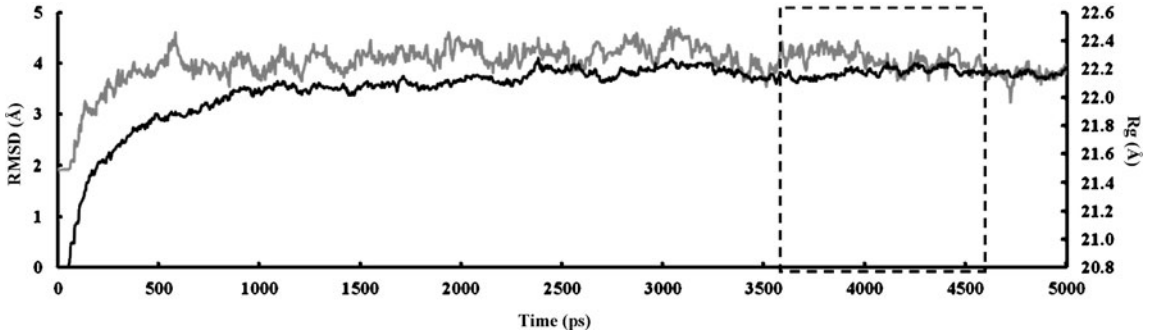


Fig. 3 Cα RMSD and radius of gyration (Rg) of the ternary complex. RMSD, black line; Rg, gray line. The square represents the simulation period used to calculate the average structure of the ternary complex

Verify3D analysis indicated a reasonably good sequence-to-structure agreement because none of the amino acids had a negative score (average score=0.36). The Procheck stereochemical analysis showed no bad contacts and no bad scores for main-chain or side-chain parameters, and the Ramachandran plot showed that 97.3% of the residues lied in allowed regions (87.8% in the most favored regions). Thus, the approach that we followed to model M.HgiDII resulted in a reliable structure.

Overall three-dimensional structure

The final model of M.HgiDII contains two separate domains: the catalytic domain and the TRD; the union of these domains forms a cleft where interactions with DNA take place (Fig. 4). The catalytic domain contains a β-sheet core surrounded by alpha helices (Fig. 4a). The β sheet contains seven strands (6 ↑ 7 ↓ 5 ↑ 4 ↑ 1 ↑ 2 ↑ 3 ↑) arranged in two sub-domains. The first sub-domain forms the SAM binding site and is built of a string of parallel β strands (1 ↑ 2 ↑ 3 ↑), which are the three leftmost β strands shown in Fig. 4b. The second domain forms the cytosine binding pocket and is composed of four β strands (6 ↑ 7 ↓ 5 ↑ 4 ↑); strand 7 is inserted in an antiparallel fashion. The β sheet contains one layer of α helices on either side of the sheet, forming a Rossmann-like fold characteristic of SAM-dependent methyltransferases [1].

The TRD of M.HgiDII resembles that of M.HaeIII [39] because both MTases show low secondary structure organization. Most of the interactions with the target DNA sequence were mediated by several loops from the TRD. One loop (amino acids 80–88) from the catalytic domain contacted the DNA helix through the minor groove (Fig. 4b), constraining the DNA backbone and trapping the flipped cytosine inside the catalytic pocket. In M.HhaI, the equivalent loop undergoes an extensive conformational change in response to DNA binding [10]; this displacement is coupled to base flipping and takes place only when the enzyme is bound to its target DNA [14]. Thus, M. HgiDII may possess the same induced fit mechanism as in M.HhaI, given the high structural similarity between both enzymes.

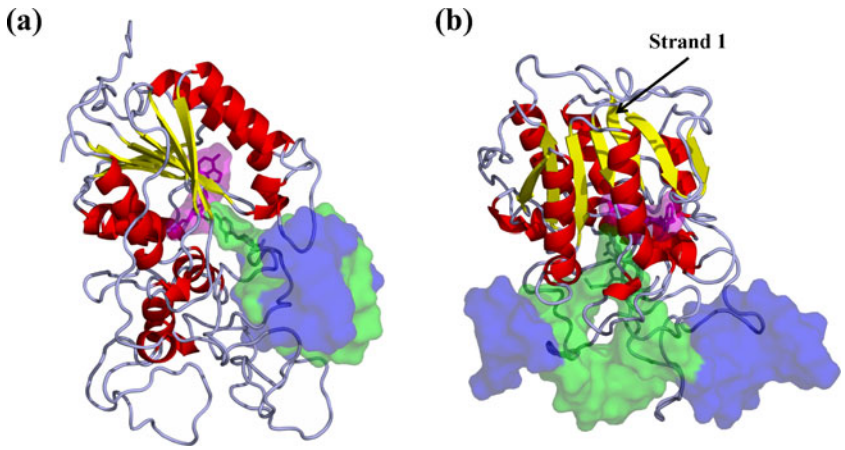


Fig. 4 Model of M.HgiDII complexed with DNA and SAM. Protein is shown as cartoon colored by secondary structure (red=α-helix; yellow=β-strand). SAM is displayed as magenta sticks with semi-transparent surface; DNA is displayed as semi-transparent surface, the recognition sequence is green, and the rest of the DNA is blue. (**a**) and (**b**) are 90-degree rotated images of the same structure. Images were rendered with PyMOL (http://www.pymol.org)

Interactions within the catalytic pocket

Trajectory analysis showed that M.HgiDII established extensive contacts with the flipped base (Fig. 5a). Arg160 (motif VIII), Arg162 (motif VIII), and Thr269 (TRD) interacted with cytosine's exocyclic oxygen O2 (average distance of 2.8, 3.1, and 3.1 Å, respectively); Arg162 also interacted with the cytosine backbone (avg. dist. 3.3 Å) and established a stable hydrogen bond with cytosine N3 (avg. dist. 3.1 Å). Ala77 (main chain; motif IV) and Glu116 (motif VI) made hydrogen bonds with the exocyclic nitrogen N4 (avg. dist. 2.9 and 2.8Å, respectively). These interactions held the cytosine properly oriented for optimal contacts with Cys79 and SAM. The importance of interactions with cytosine's backbone and O2 has been demonstrated in M.HhaI, in which substitution of Arg165 (equivalent to Arg162 of M.HgiDII) by alanine had a negative impact on methylation activity [16]. The role of Arg160 in the catalysis has been demonstrated in M.SssI, another cytosine MTase, in which a mutation of Arg230 (equivalent to Arg160 of M.HgiDII) with alanine impairs methyltransferase activity [40].

Although the conserved valine residue from motif VI is not directly involved in catalysis, its substitution causes alterations in the methylation activity [9, 40]. Mutation analysis and DNA binding experiments in M.HhaI have suggested that this residue could be important for the correct assembly of the catalytic site, and also could participate in the base flipping mechanism. In the case of M.HgiDII, Val118 (motif VI) was positioned at an average distance of 5.9Å from the target cytosine (C4), interacting with the base through hydrophobic interactions. Trajectory analysis showed that in addition to interacting with cytosine, Val118 also blocked the passage of water molecules from the DNA entry site into the catalytic site. This blockage creates a water-free environment in the region where the interaction of C6 with Cys79 takes place (Fig. 5b). Thus, we propose that besides its contribution to the positioning of cytosine, Val118 may also participate as a solvent barrier that prevents the passage of molecules into the catalytic site, which otherwise could interfere with catalysis.

According to the catalytic mechanism shown in Fig. 1, protonation of N3 by a glutamic acid residue (Glu119 in M. HhaI) allows the attack of cysteine to C6; electron flow back into the cytosine ring activates C5 to attack SAM methyl group. However, the trajectory of M.HgiDII showed that although Cys79 contacted C6 (avg. dist. 3.5Å), and that the methyl group of SAM interacted with C5 (avg. dist. 3.5Å), Glu116 did not make direct contact with N3; instead, it interacted with cytosine N4 and the amino group of SAM (Fig. 5a). Apparently, direct interactions at N3 may not be necessary for methylation in M.HgiDII; however, the cytosine ring could still be activated through water-mediated interactions at N3 (Fig. 5b). Participation of water molecules in the activation of N3 has been proposed in a molecular dynamics study on M.HhaI [18]; furthermore, a water molecule has been shown to be responsible of N3 activation in a partially active mutant of HhaI (Glu119Ala) [15]. In addition, extensive contacts at cytosine O2 (Fig. 5a) may alleviate the lack of direct interactions at N3. It has been proposed that interactions at O2 can imbalance electron density in the cytosine ring and hence allow attack at C6 by the cysteine residue [41, 42]. This mechanism seems plausible for M.HgiDII due to the extensive hydrogen bonds established between cytosine O2, Arg160, and Arg162. Glu116 was still indispensable for the catalysis because it stabilized the target cytosine and SAM.

The final step in the methylation process is the release of cytosine catalyzed by deprotonation at C5. Molecular dynamics simulations of M.HhaI have shown that a water channel in the catalytic site provides hydroxide molecules

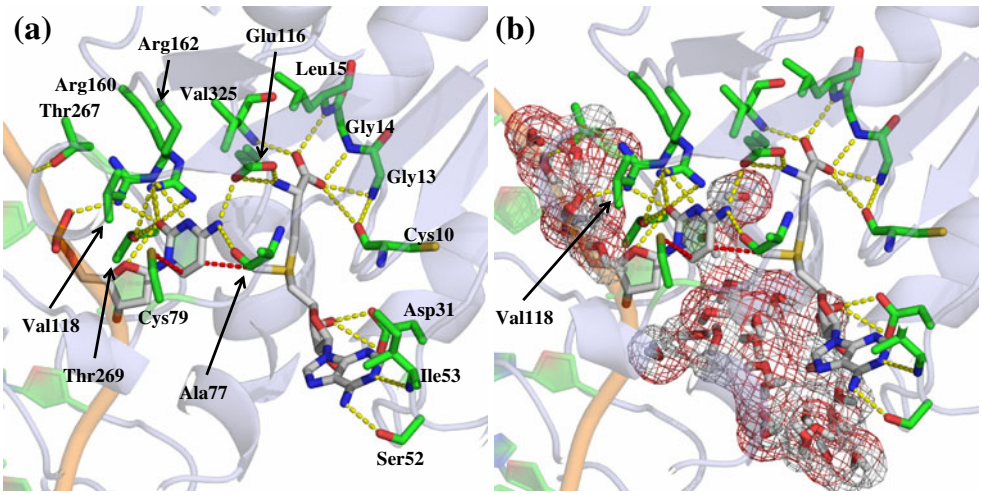


Fig. 5 Interactions in the catalytic site of M.HgiDII. Cytosine and SAM are shown as white-carbon sticks and amino acids as green-carbon sticks. Hydrogen bonds are shown in yellow; lines connecting Cys79 and C6, and C5 and SAM methyl group are shown in red. (**a**) Interactions in the catalytic site of M.HgiDII. (**b**) Water channels formed during simulations

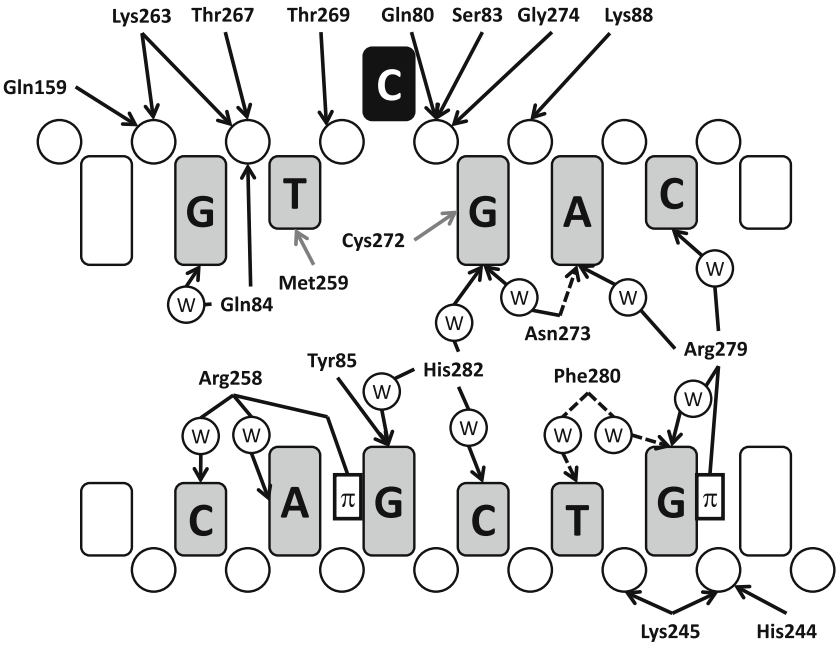


Fig. 6 Interactions with the target DNA sequence. Only interactions with bases other than the flipped cytosine are shown. M.HgiDII recognition sequence (GTCGAC) is colored gray with exception of the flipped cytosine, which is colored black. Polar contacts are shown as follows: direct side-chain interactions, continuous black arrows; direct main-chains interactions, dashed black arrows; van der Waals contacts are shown as gray continuous arrows; circles with W inside represent water-mediated contacts; π represents cation-π interactions

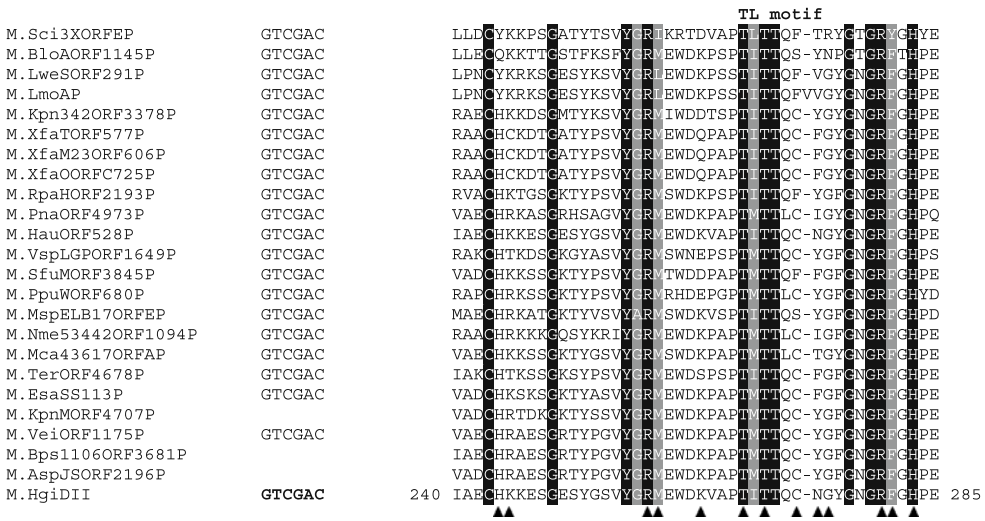
from solvent that can deprotonate C5 [17, 19]. We suggest that the same mechanism takes place in M.HgiDII, which also showed a water channel that reached cytosine C5 (Fig. 5b).

M.HgiDII interacted with SAM, orienting it to allow proper interactions with the flipped cytosine. The adenine moiety was lining part of the water channel in the catalytic site; SAM N6 interacted with Ser52 (motif III), N1 with Ile53 (main chain; motif III), and the ribose hydroxyls with Asp31 (motif II) (Fig. 5a). Additionally, the purine ring made hydrophobic interactions with Phe9 (motif I). The methionine moiety was buried in a solvent inaccessible area, and was firmly attached to the catalytic site by interactions at its amino and carboxyl groups; SAM amino group interacted with Glu116, whereas SAM carboxyl group interacted with the main chains of amino acids from motif I (Cys10, Gly13, Gly14, Leu15), and motif X (Val325).

Interactions with the target DNA sequence

M.HgiDII interacted with all the bases of its recognition sequence, establishing contacts with the DNA backbone as well as with the nucleobases (Fig. 6). Almost all the contacts with DNA were established through the major groove, which was located in the protein-DNA interface (Fig. 3); one loop (amino acids 225–246) interacted with the minor groove several bases downstream of the target sequence (Fig. 3); these interactions outside the recognition sequence are not likely to contribute to specificity, although

Fig. 7 Multiple sequence alignment of the TRD of M.HgiDII and the TRD of MTases closely related to M.HgiDII. Putative recognition sequences, as predicted by REBASE, are shown after the MTase name. Shading represents 100% identity (black background), or highly conservative changes (gray background). The amino acids that interacted with DNA are flagged with black triangles. The TL motif is shown above the alignment



they may be helpful to stabilize the union with DNA. Amino acids from loop 80–88 interacted through the minor groove, mainly with the DNA backbone; Tyr85, however, established direct contacts with the orphan guanine. This is the first report indicating an involvement of this loop in the recognition of cognate DNA, and further studies are needed to elucidate the role of Tyr85 in this process. In M.HhaI, the equivalent loop moves into the DNA only when the MTase is bound to its specific DNA sequence [12, 13].

Thr267 and Ile268 form the "TL" motif of M.HgiDII. Thr267 interacted with the thymine backbone adjacent to the flipped cytosine (Fig. 5a); similarly as in M.HhaI [22], this amino acid may be important for the correct positioning of the target cytosine inside the catalytic site of M.HgiDII. Amino acids located in the carboxyl-terminal side of Thr267 interacted with the three bases in the right side of the recognition sequence (GTC, or their corresponding GAC in the complementary strand), whereas amino acids located in the amino-terminal side of Thr267 interacted with the rest of the bases. This distribution of interactions involving amino acids at both sides of the TL motif have been identified in several cytosine MTases [22, 43].

Arg258 and Arg279 participated in cation-π interactions with two guanines of the recognition sequence (Fig. 6). Cation-π interactions are established between an aromatic ring (nucleobase) and a positive charge located above it (*e.g.*, arginine guanidinium group); although arginine residues can also establish cation-π interactions with the other DNA nucleobases, they have preference for guanines [44]. An arginine residue (Arg240) in M.HhaI is essential for sequence discrimination, loop closure and catalysis [12]. Further studies are required to establish if Arg258 or Arg279 in M.HgiDII have the same function as in M.HhaI.

Water molecules played a key role in the interactions with DNA. A network of water molecules in the interface DNA-protein contributed to stabilize the interaction DNA-protein by eliminating electrostatic repulsions between the two macromolecules [45]. Additionally, many of the specific contacts with the nucleobases were water-mediated hydrogen bonds (Fig. 6). Water molecules can act as an extension of the amino acid side-chains to reach nucleobases otherwise unavailable due to geometric or packing restrictions, or allow interactions between two acceptor or donor atoms by eliminating unfavorable electrostatic interactions [45]. Water-mediated interactions have been described in several DNA binding proteins, and are a common mechanism in the specific recognition of DNA [45, 46].

Conserved amino acids in the TRD of MTases closely related to M.HgiDII

To better characterize the interactions of M.HgiDII with DNA, we made sequence alignments with MTases closely related to M.HgiDII retrieved from the REBASE. All MTases were putative enzymes, and 20 of them had an assigned recognition sequence identical to that of M. HgiDII. The sequence alignment (Fig. 7) showed a highly-conserved motif [Cx_5Gx_6Y(GA)R(MLI)x_7T(IML) TTx_{5-6}GxGR(FY)xH], which includes the TL motif and all the amino acids of the TRD that interacted with DNA. These results indicate that these MTases could interact with DNA in a similar fashion as M.HgiDII, and possibly may recognize the same sequence. A previous work [21] revealed that conservancy of amino acids neighboring the TL motif does not necessarily predict specificity in 6-pb recognizing MTases because an MTase, M.TspMI (recognition sequence CCCGGG), has a similar motif as the one they described for M.HgiDII [YGRx_8T(LIM)x_9GRxGH]. Our analysis of interactions between M.HgiDII and the sequence alignments with related MTases suggest that a longer motif could be more useful to predict DNA specificity in MTases; moreover, the finding that loop 80–88 also made specific contacts with DNA indicates that this region should also be taken into account to predict sequence specificity in new MTases.

Conclusions

Structural studies in DNA MTases provide a challenge due to their highly variable TRDs. The approach that we followed in the homology modeling procedure allowed us to obtain a reliable model of M.HgiDII in complex with DNA and SAM. Trajectory analysis suggested that interactions with cytosine O2 and water-mediated interactions with N3 may be relevant to catalysis in M.HgiDII. Additionally, the finding that loop 80–88 interacted specifically with the target sequence and the identification of regions of the TRD that are highly conserved in MTases closely related to M.HgiDII, are expected to contribute to the design and refinement of algorithms aimed to predict sequence specificities in new MTases.

Acknowledgments RMR-A is grateful for support to this work from the Instituto Politécnico Nacional (IPN), SIP grant number 20091190, and is recipient of COFAA and EDD fellowships, granted by the IPN. JAC gives thanks to the Consejo Nacional de Ciencia y Tecnología (CONACyT, Mexico) and the IPN for graduate study fellowships. JAC thanks support from the Comité Técnico de Prestaciones a Becarios (COTEPABE, IPN). During the realization of this work, JAC received financial support from the Research Participation Program at the Center for Biologics Evaluation and Research, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration. The authors are greatly indebted to Juan L. Arciniega (Food and Drug Administration) and the NIH Helix Systems group for providing access to high-performance computational resources.

References

1. Kozbial PZ, Mushegian AR (2005) Natural history of S-adenosylmethionine-binding proteins. BMC Struct Biol 5:19. doi:10.1186/1472-6807-5-19
2. Bheemanaik S, Reddy YVR, Rao DN (2006) Structure, function and mechanism of exocyclic DNA methyltransferases. Biochem J 399:177–190
3. Bujnicki JM (2001) Understanding the evolution of restriction-modification systems: clues from sequence and structure comparisons. Acta Biochim Pol 48:935–967
4. Tock MR, Dryden DTF (2005) The biology of restriction and anti-restriction. Curr Opin Microbiol 8:466–472
5. Buryanov Y, Shevchuk T (2005) The use of prokaryotic DNA methyltransferases as experimental and analytical tools in modern biology. Anal Biochem 338:1–11
6. Klimasauskas S, Weinhold E (2007) A new tool for biotechnology: AdoMet-dependent methyltransferases. Trends Biotechnol 25:99–104
7. Jeltsch A (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. ChemBioChem 3:274–293
8. Malone T, Blumenthal RM, 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 253:618–632
9. Estabrook RA, Lipson R, Hopkins B, Reich N (2004) The coupling of tight DNA binding and base flipping: identification of a conserved structural motif in base flipping enzymes. J Biol Chem 279:31419–31428
10. Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. Cell 76:357–369
11. Huang N, Banavali NK, MacKerell AD (2003) Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. Proc Natl Acad Sci USA 100:68–73
12. Estabrook RA, Nguyen TT, Fera N, Reich NO (2009) Coupling Sequence-specific Recognition to DNA Modification. J Biol Chem 284:22690–22696
13. Zhou H, Purdy MM, Dahlquist FW, Reich NO (2009) The recognition pathway for the DNA cytosine methyltransferase M. HhaI. Biochemistry 48:7807–7816
14. Estabrook RA, Reich N (2006) Observing an induced-fit mechanism during sequence-specific DNA methylation. J Biol Chem 281:37205–37214
15. Shieh F, Reich NO (2007) AdoMet-dependent methyl-transfer: Glu119 is essential for DNA C5-cytosine methyltransferase M. HhaI. J Mol Biol 373:1157–1168
16. Shieh F, Youngblood B, Reich NO (2006) The role of Arg165 towards base flipping, base stabilization and catalysis in M.HhaI. J Mol Biol 362:516–527
17. Zhang X, Bruice TC (2006) The mechanism of M.HhaI DNA C5 cytosine methyltransferase enzyme: a quantum mechanics/molecular mechanics approach. Proc Natl Acad Sci USA 103:6148–6153
18. Lau EY, Bruice TC (1999) Active site dynamics of the HhaI methyltransferase: insights from computer simulation. J Mol Biol 293:9–18
19. Roberts RJ (1994) An amazing distortion in DNA induced by a methyltransferase. Biosci Rep 14:103–117
20. Szegedi SS, Gumport RI (2000) DNA binding properties in vivo and target recognition domain sequence alignment analyses of wild-type and mutant RsrI [N6-adenine] DNA methyltransferases. Nucleic Acids Res 28:3972–3981
21. Vilkaitis G, Dong A, Weinhold E, Cheng X, Klimasauskas S (2000) Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. J Biol Chem 275:38722–38730
22. Neely RK, Roberts RJ (2008) The BsaHI restriction-modification system: cloning, sequencing and analysis of conserved motifs. BMC Mol Biol 9:48. doi:10.1186/1471-2199-9-48
23. Kröger M, Hobom G, Schütte H, Mayer H (1984) Eight new restriction endonucleases from Herpetosiphon giganteus–divergent evolution in a family of enzymes. Nucleic Acids Res 12:3127–3141
24. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948
25. Bryson K, McGuffin LJ, Marsden RL, Ward LL, Sodhi JS, Jones DT (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33:W36–38
26. Eswar N, Eramian D, Webb B, Shen M, Sali A (2008) Protein structure modeling with MODELLER. Methods Mol Biol 426:145–159
27. Melo F, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: a www server to assess protein structures. Proc Int Conf Intell Syst Mol Biol 5:187–190
28. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
29. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
30. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
31. Mackerell A, Banavali N (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. J Comput Chem 21:105–120
32. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
33. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–410
34. Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404
35. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
36. Bourne PE, Weissig H (2003) Structural bioinformatics. Wiley-Liss, Wilmington
37. Koudan EV, Bujnicki JM, Gromova ES (2004) Homology modeling of the CG-specific DNA methyltransferase SssI and its complexes with DNA and AdoHcy. J Biomol Struct Dyn 22:339–345
38. Bujnicki JM (2000) Homology modelling of the DNA 5mC methyltransferase M.BssHII. Is permutation of functional subdomains common to all subfamilies of DNA methyltransferases? Int J Biol Macromol 27:195–204
39. Reinisch K (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: An extrahelical cytosine and rearranged base pairing. Cell 82:143–153
40. Darii MV, Cherepanova NA, Subach OM, Kirsanova OV, Raskó T, Slaska-Kiss K, Kiss A, Deville-Bonne D, Reboud-Ravaux M,

Gromova ES (2009) Mutational analysis of the CG recognizing DNA methyltransferase SssI: insight into enzyme-DNA interactions. Biochim Biophys Acta. doi:10.1016/j.bbapap. 2009.07.016
41. Gabbara S, Sheluho D, Bhagwat AS (1995) Cytosine methyltransferase from Escherichia coli in which active site cysteine is replaced with serine is partially active. Biochemistry 34:8914–8923
42. Kumar S, Horton JR, Jones GD, Walker RT, Roberts RJ, Cheng X (1997) DNA containing 4′-thio-2′-deoxycytidine inhibits methylation by HhaI methyltransferase. Nucleic Acids Res 25:2773–2783
43. Lauster R, Trautner TA, Noyer-Weidner M (1989) Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. J Mol Biol 206:305–312
44. Wintjens R, Liévin J, Rooman M, Buisine E (2000) Contribution of cation-pi interactions to the stability of protein-DNA complexes. J Mol Biol 302:395–410
45. Reddy CK, Das A, Jayaram B (2001) Do water molecules mediate protein-DNA recognition? J Mol Biol 314:619–632
46. Rhodes D, Schwabe JW, Chapman L, Fairall L (1996) Towards an understanding of protein-DNA recognition. Philos Trans R Soc Lond B Biol Sci 351:501–509