Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation

Jikui Song,* Marianna Teplova, Satoko Ishibe-Murakami, Dinshaw J. Patel†

DNMT1, the major maintenance DNA methyltransferase in animals, helps to regulate gene expression, genome imprinting, and X-chromosome inactivation. We report on the crystal structure of a productive covalent mouse DNMT1(731-1602)–DNA complex containing a central hemimethylated CpG site. The methyl group of methylcytosine is positioned within a shallow hydrophobic concave surface, whereas the cytosine on the target strand is looped out and covalently anchored within the catalytic pocket. The DNA is distorted at the hemimethylated CpG step, with side chains from catalytic and recognition loops inserting through both grooves to fill an intercalation-type cavity associated with a dual base flip-out on partner strands. Structural and biochemical data establish how a combination of active and autoinhibitory mechanisms ensures the high fidelity of DNMT1-mediated maintenance DNA methylation.

In mammals, DNA methylation predominantly occurs on the carbon-5 position of cytosines within CpG dinucleotides (*1*–*4*) and is critically maintained by DNMT1 during DNA replication (*5*). DNMT1 is composed of a C-terminal methyltransferase domain and an N-terminal regulatory domain, linked by a conserved $(Gly\text{-}Lys)_n$ repeat (fig. S1A). The N-terminal domain contains sequences that mediate interactions of DNMT1 with other proteins (*6*, *7*), a nuclear localization sequence, a target recognition sequence that localizes DNMT1 to the DNA replication fork (*8*), a zinc finger CXXC (Cys-X-X-Cys) domain that specifically recognizes unmethylated CpG DNA (*9*, *10*), and a pair of bromo-adjacent homology (BAH) domains (*11*). The methyltransferase domain of DNMT1 is further folded into two subdomains, designated as the catalytic core and the target recognition domain (TRD).

We have determined the crystal structure of a productive complex of truncated mouse DNMT1 (mDNMT1) with DNA (Fig. 1A), which contrasts with our previously determined autoinhibitory structure of the mDNMT1-DNA complex (*12*). The complex contains mDNMT1(731-1602), which encompasses both BAH domains and the methyltransferase domain, bound to DNA within a (mCpG)-(fCpG) dinucleotide context. The complementary 12-mer DNA duplex is composed of a central 5-methylcytosine (mC)–containing hemi-mCpG step on the parental strand, positioned opposite to a 5-fluorocytosine (5fC)–containing fCpG step (*13*) on the target strand. The complex was produced by enzymatically cross-linking the 5fC on the DNA target strand to the reactive cysteine positioned in the catalytic pocket of mDNMT1 (fig. S1B).

Superposition of the structures of the productive (this study) and autoinhibited (*12*) mDNMT1-DNA complexes demonstrates that the largest conformational change is centered within the catalytic loop of mDNMT1 (Fig. 1B and fig. S2). In the productive covalent complex, the catalytic loop (in green) is followed by a straight α helix (green helical segment in inset of Fig. 1B) and penetrates into the DNA minor groove (Fig. 1C), forming extensive protein-DNA contacts (see below), whereas in the autoinhibitory structure of the complex, the α helix following the catalytic loop has a kink at the N-terminal end (purple helical segment in inset of Fig. 1B), and the catalytic loop (in purple) is excluded from the DNA minor groove by the autoinhibitory linker (in blue) (Fig. 1D).

Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.

*Present address: Department of Biochemistry, University of California, Riverside, CA 92521, USA.
†To whom correspondence should be addressed. E-mail: pateld@mskcc.org

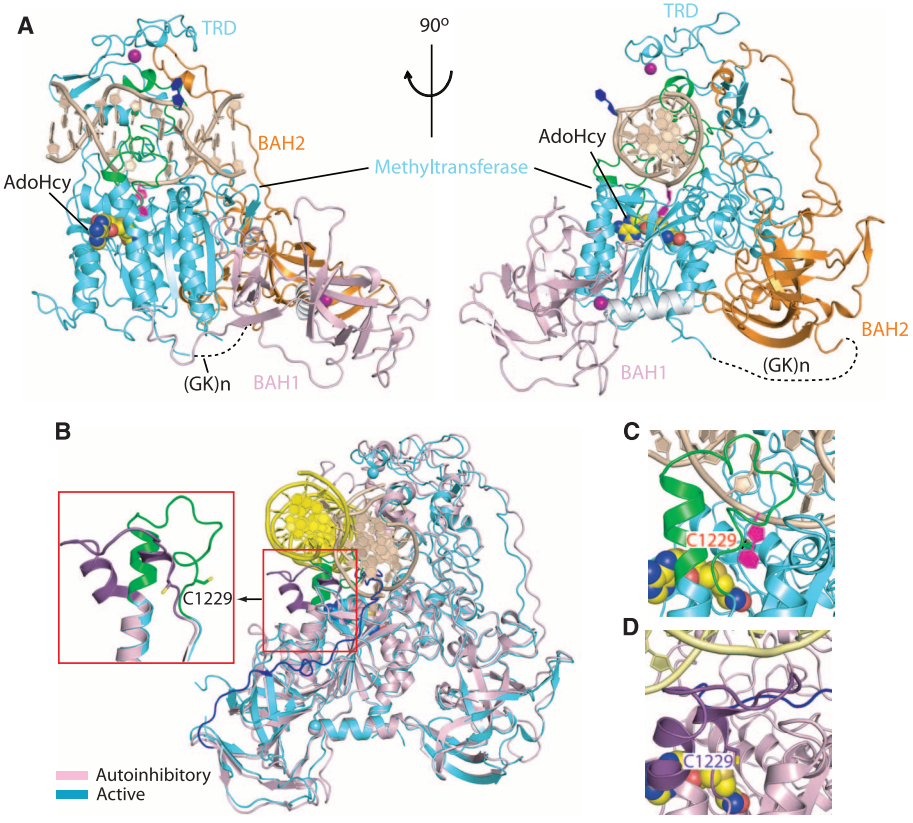

Fig. 1. Structural overview of mDNMT1(731-1602) bound covalently to a hemi-mCpG site. (**A**) Ribbon representation of the covalent complex in two orthogonal views. The BAH1, BAH2, and methyltransferase domains are colored light pink, orange, and cyan, respectively; DNA and zinc ions are barley and purple, respectively. The disordered $(Gly\text{-}Lys)_n$ linker [(GK)n] is shown as black dashed lines, the 5-methyl group from mC6 is green, the flipped-out target fC7′ is purple, the flipped-out C from parental strand is blue, the catalytic loop and TRD loops 1 and 2 are green, and the bound AdoHcy is in a space-filling representation. (**B**) Structural superposition of the covalent mDNMT1(731-1602)–DNA complex with the autoinhibited mDNMT1(650-1602)–DNA complex. The bound DNA in the covalent complex is in beige with the protein in cyan; they are yellow and cyan in the autoinhibited complex. The catalytic loop (expanded view in inset) is colored purple in the autoinhibited complex and green in the covalent complex. The CXXC-BAH1 domain linker is in dark blue in the autoinhibited complex. The inset highlights the transition from a straight helix (green) in the productive covalent complex to a kinked helix (purple) in the autoinhibited complex. (**C**) Close-up view of the catalytic loop (in green) in the covalent mDNMT1 (731-1602)–DNA complex (this study). (**D**) Close-up view of the catalytic loop (in dark brown) in the autoinhibited mDNMT1(650-1602)–DNA complex (PDB: 3PT6). The side chain of residue Cys^{1229} is shown in stick representation.

mDNMT1(731-1602)–mediated cytosine methylation kinetics on DNA 14-mer substrates containing single central hemi-mCpG versus unmodified CpG sites (fig. S3A) yielded kinetic (k_{cat} and k_{cat}/K_M) parameters similar to those previously measured for mDNMT1(717-1602) (fig. S3B and table S1) (*12*); the relative velocity of methylation of hemi-mCpG DNA substrate was greater than that of its CpG DNA counterpart by a factor of ~10 (fig. S3C).

The DNA duplex (Fig. 2A) is embedded in the catalytic cleft of mDNMT1, with the fC7′ of the target strand looped out of the helix and inserted into the catalytic pocket of the methyltransferase domain, where it is methylated and anchored through covalent bond formation with the reactive cysteine (Cys^{1229}) (Fig. 2B and fig. S4A). The fC7′ is surrounded by strictly conserved residues from the catalytic core (fig. S4), as initially observed for bacterial M.*Hha*I-DNA (*14*) and M.*Hae*III-DNA (*15*) complexes, and is in proximity to the *S*-adenosyl homocysteine (AdoHcy) (fig. S5).

Although the CXXC domain and the CXXC-BAH1 linker together play an autoinhibitory role in preventing methylation of unmethylated CpG sites by DNMT1 (*12*), DNMT1 retains considerable substrate preference toward hemi-mCpG steps in the absence of these two elements (fig. S3), which suggests that the methyltransferase domain by itself encodes capabilities for evaluating the methylation status of the substrate.

In the current structure of the productive mDNMT1-DNA complex, the methyl group attached to the C-5 position of mC6 on the parental strand is directed toward a hydrophobic segment within the TRD domain (Fig. 2C). In particular, the indole ring of a tryptophan (Trp^{1512}) slides into the DNA major groove upon complex formation and is partially stacked with the base of mC6, and, together with other hydrophobic residues (Cys^{1501}, Leu^{1502}, Leu^{1515}, and Met^{1535}), forms a shallow concave surface harboring the 5-methyl group of mC6 (Fig. 2C). This observation establishes the molecular basis for intrinsic preference of DNMT1 toward hemi-mCpG DNA substrates.

The formation of the productive mDNMT1-DNA complex involves infiltration of the DNA from both major (two TRD loops) and minor (catalytic loop) grooves (Fig. 1A and fig. S6), with the TRD domain undergoing concerted movement toward the DNA major groove by about 2 to 3 Å upon complex formation (Fig. 1B). We observe a large helical distortion around the central hemi-mCpG site resulting in the opening of an intercalation-type cavity 7.5 Å in length along the DNA helical axis (fig. S7, A and B). The side chain of Met^{1235} from the catalytic loop inserts into the DNA from the minor groove and occupies the space vacated by the extruded fC7′ on the target strand (Fig. 2, D and E). The side chain of Lys^{1537} from TRD loop 2 inserts into the DNA from the major groove and occupies the adjoining empty space on the parental strand (Fig. 2, D and E), which results from translation of the stacked orphan guanine (G7) by one step along the DNA helix toward the 3′ end, thereby flipping out the immediate downstream residue C8 (Fig. 2D). The antiparallel alignment of the side chains of Met^{1235} and Lys^{1537} is stabilized by a single hydrogen bond and further buttressed by the indole ring of Trp^{1512} (Fig. 2, D and E), with these distortions (fig. S8A) distinct from helical distortions reported for bacterial DNA methyltransferases (fig. S8, B and C). The mC6-G6′ Watson-Crick pair (Fig. 2F) and G7 of the noncanonical G7•G8′ pair (Fig. 2G) that bracket the intercalation site are recognized

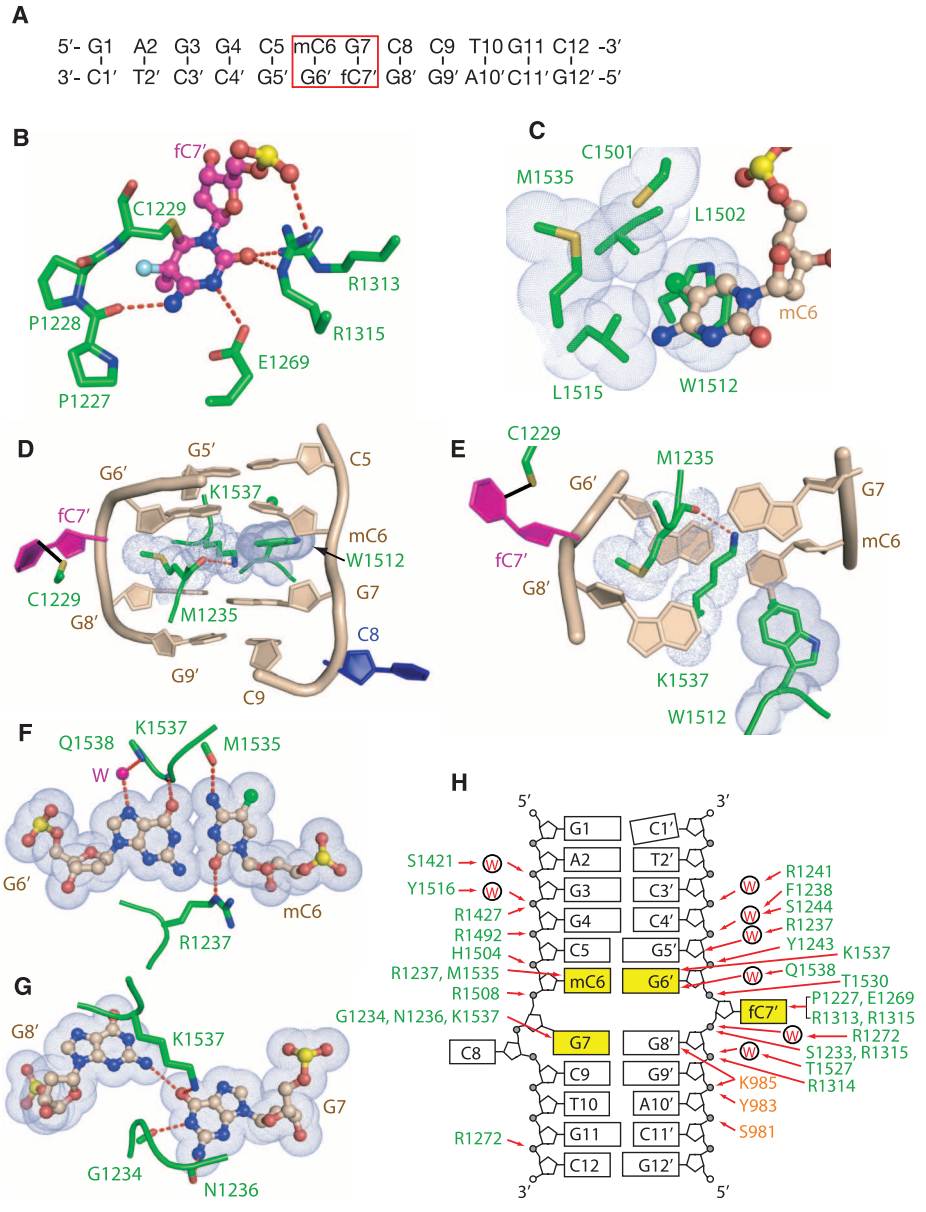

Fig. 2. Insertion of amino acid side chains into the intercalation-like distortion of the bound DNA and base-specific recognition at the hemi-mCpG site in the covalent mDNMT1-DNA complex. (**A**) Sequencing and numbering system of the hemi-mCpG–containing 12-mer duplex. (**B**) Stabilization of the flipped-out target cytosine fC7′ (light blue ball) in the active site of the enzyme. (**C**) The 5-methyl group (green ball) of mC6 is anchored within a hydrophobic concave surface of mDNMT1. (**D**) Insertion of side chains of Met^{1235} and Lys^{1537} into the intercalation-type space, and buttressing by indole ring of Trp^{1512}. The covalent bond between the sulfur atom of Cys^{1229} and the C6 atom of the target cytosine fC7′ is shown as a dark line. (**E**) End-on view of Met^{1235} and Lys^{1537} side chain insertion in the covalent complex. (**F**) Direct and water-mediated (labeled with purple W) hydrogen bond interactions target the Watson-Crick mC6-G6′ base pair from both grooves. (**G**) Stabilization of the repositioned guanine G7 through hydrogen-bonding interactions. (**H**) Schematic tabulation of direct and water-mediated hydrogen bond and electrostatic interactions between mDNMT1 and DNA in the covalent complex. Abbreviations for amino acids: C, Cys; E, Glu; F, Phe; G, Gly; H, His; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; W, Trp; Y, Tyr.

by elements from both the TRD and the catalytic core (Fig. 2H; stereo views in fig. S9, A and B).

The catalytic loop (residues 1227 to 1243) inserts into and spans both strands of the minor groove (Fig. 3A), making a network of direct and water-mediated contacts with mC6 and G7 on the parental strand, fC7′ on the target strand, and the phosphate backbone on either side of fC7′ (stereo view in fig. S10). In addition, both TRD loop 1 (residues 1501 to 1516) and loop 2 (residues 1530 to 1537) insert into and span both strands of the major groove centered about the hemi-mCpG step (fig. S6). TRD loop 1 contacts the phosphate backbone of the parental strand (Fig. 3B; stereo view in fig. S11), whereas TRD loop 2 is involved in base-specific recognition of the hemi-mCpG step and the DNA cavity (Fig. 3C; stereo view in fig. S12). Both TRD loops participate in formation of the hydrophobic concave surface that recognizes the 5-methyl group of mC6. Other DNA-interacting residues in the methyltransferase domain include Arg^{1272}, Arg^{1314}, Ser^{1421}, Arg^{1427}, Arg^{1492}, and Thr^{1527} (Fig. 2H and fig. S13).

Formation of the covalent mDNMT1-DNA complex is further accompanied by the conformational change within the BAH2 domain, whose helical tip moves toward the DNA along with the TRD (fig. S14). As a result, the side chains of three residues (Ser^{981}, Tyr^{983}, and Lys^{985}) immediately C-terminal to the tip, which are disordered in both free mDNMT1 and the autoinhibited mDNMT1-DNA complex, become ordered and engage in interactions with the phosphate backbone of the DNA target strand (Fig. 3D).

Guided by the structural analysis, we mutated a number of key and evolutionarily conserved residues (fig. S15) to monitor their impact on DNA methylation. First, we mutated residues lining the shallow concave surface that harbors the 5-methyl group of mC6. We observe that the Trp^{1512} → Ala mutant is essentially catalytically dead (Fig. 3E and fig. S16, A and B), consistent with the observation by a previous study (*16*) and the role of Trp^{1512} in base-stacking with the 5-methylcytosine. All other single mutants had their methylation rates reduced by factors of 2 to 4 for hemi-mCpG DNA substrate (Fig. 3E and fig. S16A) and by factors of 3 to 4 on unmethylated CpG substrate (Fig. 3E and fig. S16B); these findings suggest that the overall hydrophobic environment generated by a cluster of residues, rather than individual residues, is likely to play a dominant role in dictating the substrate specificity of mDNMT1.

Among residues that insert into the DNA intercalation site, the Met^{1235} → Ala mutant lost major catalytic activity on both hemi-mCpG and CpG DNA substrates (Fig. 3E and fig. S16, C and D). By contrast, the Lys^{1537} → Ala mutant exhibited a factor of 2 reduction in methylation rate for hemi-mCpG DNA substrate but an unanticipated factor of 2 increase in methylation rate on unmethylated CpG substrate (Fig. 3E and fig. S16, C and D). We have no definitive explanation for the latter result, but it is conceivable that such altered enzymatic behavior could arise from the impact of the Lys^{1537} → Ala mutation on the side chain orientations of Trp^{1512} and mC6 (Fig. 2C), which in turn could influence the substrate recognition.

Additional methylation kinetics were also performed for the Arg^{1237} → Ala mutant and for the Ser^{981} → Ala, Tyr^{983} → Ala, Lys^{985} → Ala triple mutant. In accordance with the structural analysis, both mutations show reduced catalytic activities upon DNA methylation (Fig. 3E and fig. S16, E and F). Taken together, our enzymatic

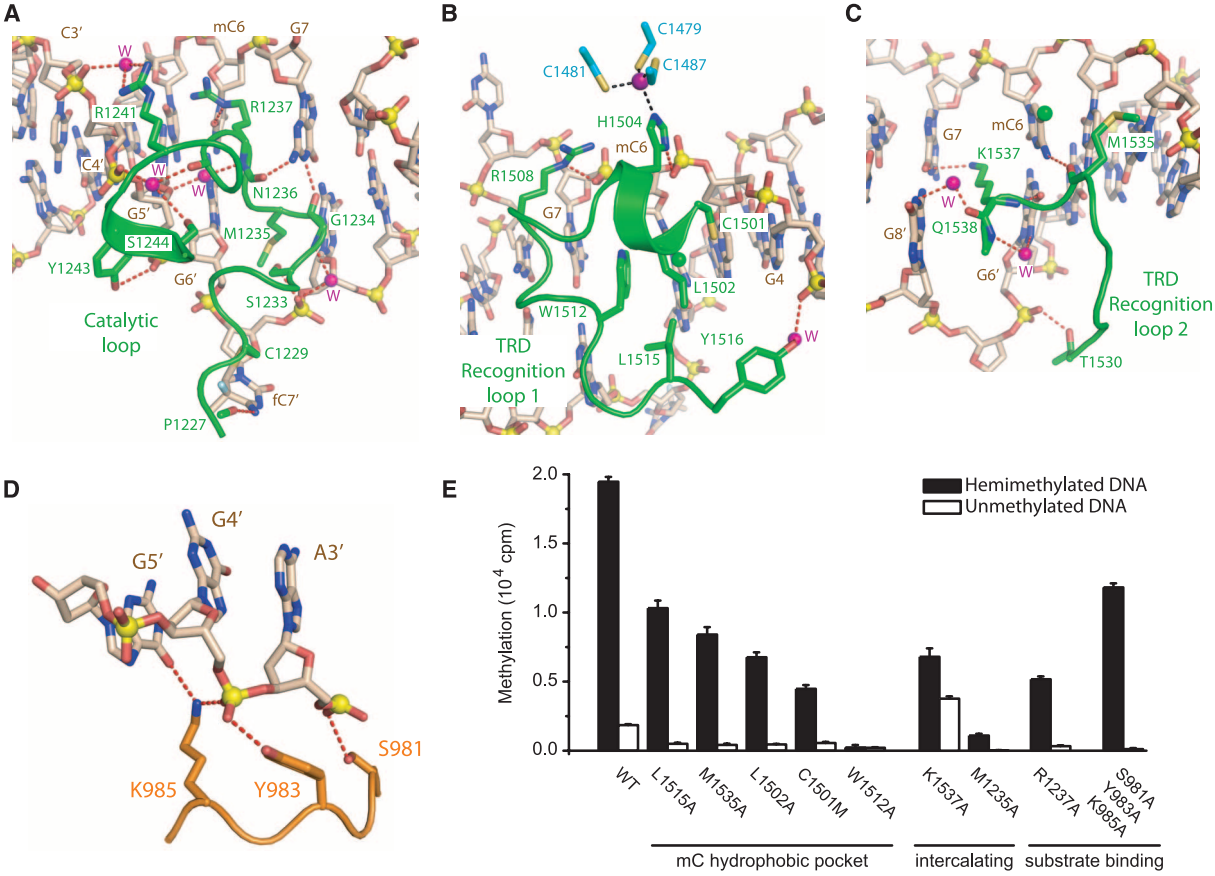

Fig. 3. Insertion of the catalytic loop and a pair of TRD loops into the grooves centered about the hemi-mCpG site in the covalent mDNMT1-DNA complex. (**A**) Catalytic loop–DNA interactions. (**B**) TRD loop 1–DNA interactions. The imidazole ring of His^{1504} bridges the zinc finger with the backbone phosphate of mC6. (**C**) TRD loop 2–DNA interactions. (**D**) Intermolecular hydrogen-bond interactions between side chains of a segment of the BAH2 loop and DNA in the complex. (**E**) Methylation activities of mDNMT1(731-1602) and its mutants involving amino acids that line the mC-recognizing hydrophobic surface, or insert into the intercalation-type cavity, and catalytic and BAH2 loop residues involved in substrate recognition. Methylation activities were monitored after reaction on hemi-mCpG DNA substrate (black bars) or unmodified CpG DNA substrate (white bars) for 10 min. Error bars represent SD calculated from two measurements.

methylation assays on selective mDNMT1 mutants appear to support structure-based molecular insights into substrate recognition by DNMT1.

Maintenance of DNA methylation patterns during cell division is crucial to animal development. DNMT1 is tethered to the DNA replication fork and is the major player in maintaining DNA methylation patterns during replication. The current structure of the productive mDNMT1-DNA complex, when combined with our earlier structure of the corresponding autoinhibited complex (*12*), establishes that two distinct but mutually supportive mechanisms are used by DNMT1 to ensure faithful maintenance of DNA methylation patterns after replication. As unmethylated CpG dinucleotides emerge from the replication complex, such sites are protected from de novo methylation by the autoinhibitory mechanism. On the other hand, as hemimethylated CpG sites emerge from the replication complex, specific recognition by the TRD domain of DNMT1 makes such sites optimal targets for maintenance DNA methylation. These two mechanisms, working in tandem and mutually supportive of each other, together with anticipated regulation by other proteins in the replication complex (such as UHRF1), ensure the fidelity of DNMT1-mediated maintenance DNA methylation.

References and Notes

1. X. Cheng, R. M. Blumenthal, *Structure* **16**, 341 (2008).
2. M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **74**, 481 (2005).
3. J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **11**, 204 (2010).
4. M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **9**, 465 (2008).
5. E. Li, T. H. Bestor, R. Jaenisch, *Cell* **69**, 915 (1992).
6. L. S. Chuang *et al.*, *Science* **277**, 1996 (1997).
7. M. R. Rountree, K. E. Bachman, S. B. Baylin, *Nat. Genet.* **25**, 269 (2000).
8. H. Leonhardt, A. W. Page, H. U. Weier, T. H. Bestor, *Cell* **71**, 865 (1992).
9. J. H. Lee, K. S. Voo, D. G. Skalnik, *J. Biol. Chem.* **276**, 44669 (2001).
10. M. Pradhan *et al.*, *Biochemistry* **47**, 10000 (2008).
11. I. Callebaut, J. C. Courvalin, J. P. Mornon, *FEBS Lett.* **446**, 189 (1999).
12. J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **331**, 1036 (2011).
13. D. G. Osterman, G. D. DePillis, J. C. Wu, A. Matsuda, D. V. Santi, *Biochemistry* **27**, 5204 (1988).
14. S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **76**, 357 (1994).
15. K. M. Reinisch, L. Chen, G. L. Verdine, W. N. Lipscomb, *Cell* **82**, 143 (1995).
16. K. Takeshita *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9055 (2011).

Acknowledgments: Supported by funds from the Abby Rockefeller Mauze Trust and STARR and Maloris Foundations (D.J.P.). We thank T. Bestor and M. Goll for helpful discussions, and the staff of the NE-CAT beamlines at the Advanced Photon Source, Argonne National laboratory, for access and assistance with data collection. The structure of the complex has been deposited under accession no. 4DA4 in the Protein Data Bank.

Supporting Online Material

www.sciencemag.org/cgi/content/full/335/6069/709/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 and S2
References (*17*–*21*)

26 September 2011; accepted 12 January 2012
10.1126/science.1214453

Structure and Allostery of the PKA RIIβ Tetrameric Holoenzyme

Ping Zhang,[1] Eric V. Smith-Nguyen,[2] Malik M. Keshwani,[1,3] Michael S. Deal,[2] Alexandr P. Kornev,[2,3] Susan S. Taylor[1,2,3]*

In its physiological state, cyclic adenosine monophosphate (cAMP)–dependent protein kinase (PKA) is a tetramer that contains a regulatory (R) subunit dimer and two catalytic (C) subunits. We describe here the 2.3 angstrom structure of full-length tetrameric RIIβ$_2$:C$_2$ holoenzyme. This structure showing a dimer of dimers provides a mechanistic understanding of allosteric activation by cAMP. The heterodimers are anchored together by an interface created by the β4-β5 loop in the RIIβ subunit, which docks onto the carboxyl-terminal tail of the adjacent C subunit, thereby forcing the C subunit into a fully closed conformation in the absence of nucleotide. Diffusion of magnesium adenosine triphosphate (ATP) into these crystals trapped not ATP, but the reaction products, adenosine diphosphate and the phosphorylated RIIβ subunit. This complex has implications for the dissociation-reassociation cycling of PKA. The quaternary structure of the RIIβ tetramer differs appreciably from our model of the RIα tetramer, confirming the small-angle x-ray scattering prediction that the structures of each PKA tetramer are different.

Cyclic adenosine monophosphate (cAMP)–dependent protein kinase (PKA), a ubiquitous serine/threonine protein kinase, exists in mammalian cells as an inactive tetrameric holoenzyme composed of a regulatory (R) subunit dimer and two catalytic (C) subunits. cAMP binding to the R subunits releases the active C subunits, allowing them to phosphorylate specific substrate proteins. The two classes of R subunit, RI and RII, each have α and β isoforms, and these functionally nonredundant isoforms are a primary mechanism for achieving specificity in PKA signaling (*1*). Deletion of RIα, for example, is embryonically lethal (*2*), whereas RIIβ knockout mice have a lean phenotype and are not susceptible to diet-induced insulin resistance (*3*, *4*). Depletion of RIIβ reverses the obesity syndrome of agouti mice (*5*). RIIβ is the predominant isoform in brain and adipose tissue (*2*, *6*). RII subunits are typically anchored to membrane proteins through high-affinity binding to adenosine kinase–anchoring proteins (AKAPs).

The crystal structure of the C subunit revealed the conserved kinase core shared by all members of the protein kinase superfamily (*7*, *8*). It is bilobal with a large, helical C lobe (Fig. 1A) that facilitates substrate recognition and provides the catalytic machinery for phosphoryl transfer and a smaller, more dynamic beta-rich N lobe that is associated mostly with adenosine triphosphate (ATP) binding (*7*, *9*, *10*). The C-terminal tail (C tail), a conserved feature of kinases belonging to the AGC subfamily, serves as a cis-regulatory element that wraps around both lobes and primes the C subunit for catalysis (*11*).

Each R subunit (Fig. 1A) contains an N-terminal dimerization/docking (D/D) domain, followed by a flexible linker containing an inhibitor site (IS) that docks to the active-site cleft of the C subunit in the holoenzyme. At the C terminus are two tandem highly conserved cyclic nucleotide-binding domains (CNB-A and CNB-B) (*12*). PKA is anchored to specific sites in the cell by binding of an AKAP amphipathic helix to the D/D domain (*13*, *14*). RII ISs have a Ser at their phosphorylation site (P site) and are both substrates and inhibitors, whereas RI subunits with Gly or Ala at their P site are inhibitors and pseudosubstrates. Phosphorylation of the P-site Ser in RII slows the rate of association with C subunit (*15*, *16*), and formation of holoenzyme in cells is influenced substantially depending on whether the P-site residue is a substrate or a pseudosubstrate (*12*, *17*). In contrast to RII subunits, forming a high-affinity type I holoenzyme with RI subunits requires Mg_2ATP (*18*).

Complexes of R and C subunits, first described with truncated monomeric R subunits, showed for the first time how the C subunit was inhibited by R and how the complex was activated by cAMP (*19*–*22*). These RIα, RIIα, and RIIβ heterodimeric complexes also showed that the R subunits undergo a dramatic conformational change as they release cAMP and bind to C subunit. However, in the absence of cAMP, PKA in cells exists as a tetramer, and only a full-length tetrameric holoenzyme structure will explain how PKA is assembled in its physiological state and how it acquires its allosteric properties. A model of the (RIα)$_2$:C$_2$ holoenzyme was proposed recently (*23*),

[1]Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093–0654, USA. [2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093–0654, USA. [3]Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093–0654, USA.

*To whom correspondence should be addressed. E-mail: staylor@ucsd.edu

www.sciencemag.org/cgi/content/full/335/6069/709/DC1

Supporting Online Material for

Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation

Jikui Song, Marianna Teplova, Satoko Ishibe-Murakami, Dinshaw J. Patel

*To whom correspondence should be addressed. E-mail: pateld@mskcc.org

Published 10 February 2012, *Science* **335**, 709 (2012)
DOI: 10.1126/science.1214453

This PDF file includes:

SUPPLEMENTARY ONLINE MATERIALS (SOM)

Materials and Methods

Protein expression and purification. The synthesized gene encoding residues 731-1602 of mouse DNMT1 were purchased from Geneart Inc, and was inserted into a modified pRSFDuet-1 vector (Novagen), in which mDNMT1 was separated from the preceding His_6-SUMO tag by a ubiquitin-like protease (ULP1) cleavage site. Expression and purification of mDNMT1(731-1602) used protocols described previously (*12*). To generate the covalent mDNMT1-DNA complex, a 12-bp DNA containing a single central hemi-mCG site (supplied by Keck Oligonucleotide Synthesis Resource at Yale University; upper parental strand: 5'- GAGGCmCGCCTGC-3', mC = 5-methyl-deoxycytosine; lower target strand 5'-GCAGGfCGGCCTC-3', fC = 5-fluoro-deoxycytosine) (Figure 2A), was enzymatically methylated in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 20% Glycerol, 4 mM DTT at room temperature with mDNMT1 (731-1602) in the presence of AdoMet. The reaction product was purified through a Heparin column, followed by gel filtration on a 16/60 G200 Superdex column. The final sample for crystallization of the productive mDNMT1-DNA complex contains about 0.1-0.2 mM covalent mDNMT1(731-1602)-DNA complex, 20 mM Tris-HCl, 250 mM NaCl, 5 mM DTT, 5% glycerol, pH 7.5.

Crystallization conditions. The initial crystallization condition (0.1 M sodium citrate, pH 4.8) for the productive covalent mDNMT1(731-1602)-DNA complex was identified using sparse-matrix screens (Qiagen inc). The crystals were subsequently reproduced by hanging-drop vapor-diffusion method at 4 °C, from drops mixed from 1 μl of mDNMT1-DNA solution and 1 μl of precipitant solution [0.1 M sodium citrate (pH 4.8), 50 mM sodium malonate (pH 4.0), 3 mM TCEP]. The quality of crystals was further improved by seeding into drops mixed from 1 μl of mDNMT1-DNA solution and 1 μl of precipitant solution containing 3 mM TCEP, 0.1 M sodium citrate, pH 4.5. The crystals were soaked in cryoprotectant made of mother liquor supplemented with 30% glycerol, before flash frozen in liquid nitrogen.

Structure determination. X-ray diffraction data sets for the covalent mDNMT1(731-1602)-DNA complex were collected at selenium peak wavelength on the 24-IDE NE-CAT beamline at the Advanced Photo Source (APS), Argonne National

Laboratory. The diffraction data were indexed, integrated and scaled using the HKL 2000 program. The structure of the productive covalent mDNMT1(731-1602)-DNA complex was solved using the molecular replacement method in PHASER (*17*) using the free structure of mDNMT1(731-1600) (PDB 3PT9) as a search model. There are two molecules of the mDNMT1-DNA complex in the asymmetric unit (ASU). The Fo-Fc map that resulted from molecular replacement clearly revealed a DNA molecule bound to each mDNMT1. Further modeling of the covalent mDNMT1(731-1602)-DNA was carried out using COOT (*18*), and was then subject to refinement using the PHENIX (*19*). The same R-free test set was used throughout the refinement. The refinement was restrained using secondary structure and non-crystallographic symmetry (NCS). The final model was refined to 2.6 Å resolution. During the final cycles of refinement, the NCS term was left out to account for different interaction environments of the two complexes. The B-factors were refined with individual B values.

The statistics for data collection and structural refinement for the productive covalent mDNMT1-DNA complexes is summarized in table S2.

Enzymatic methylation kinetics assays. The DNMT1 cytosine-5 methylation assays were carried out at 37 °C, using protocols described previously (*12*). A typical 25 µL reaction mixture contained 5 µM S-adenosyl-l-[methyl-^{3}H]methionine (AdoMet) (specific activity 18 Ci/mmol, PerkinElmer), 40 nM of DNMT1 methyltransferase, 0.6 µM of 14-mer DNA duplex (upper strand: 5'-GGAGGCXGCCTGCT-3, X = unmethylated C or m5C) that is unmethylated or hemimethylated on the upper strand in 50 mM Tris-HCl, pH 7.8, 1 mM DTT, 5% glycerol and 100 µg/mL BSA. The methyltransferase reactions were stopped at certain time points by transferring the reaction tubes to an ethanol-dry ice bath and were processed by spotting the reaction mix on DE81 paper circles (Whatman). The DE81 circles were washed sequentially with 2 x 5 ml of cold 0.2 M ammonium bicarbonate, 2 x 5 ml of Milli Q water, and 5 ml of ethanol using a manifold vacuum. The filters were dried, transferred to scintillation vials; 5 ml of ScintiVerse (Fisher) were added to each vial and tritium scintillation was detected with a Beckman LS6500 counter. Points were plotted as the mean of the two samples at 5, 10, 20, 30 40, 50 and 60 min and fit to a linear progression for unmodified DNA substrate, and fit to

a first-order exponential for hemimethylated DNA substrate as described (*20*) using Microcal Origin software.

To estimate the steady-state Michaelis–Menten parameters, the DNMT1 cytosine-5 methylation assays were carried out at 37 °C, for 50 min in triplicate. The useful range of the CpG site concentration was between 0.02 μM and 0.75 μM for unmethylated DNA and between 0.02 μM and 1.5 μM for hemimethylated DNA. The reaction rates (v, nM/h) were plotted as a function of the DNA concentration (Microcal 6 Software Inc.), and the data were fit by nonlinear regression of the Michaelis-Menten equation,

$v = V_{max} \times [DNA]/(K_M + [DNA])$

to calculate K_M and V_{max}; k_{cat} was calculated as V_{max} divided by the concentration of DNMT1 protein. DNMT1 (731-1602) wild type and mutants were purified at concentrations ranging from 12 to 85 μM. In typical experiment 2 μM protein samples were prepared in the reaction buffer and the concentration of each mutant was quantitated by comparison of its band intensity in SDS polyacrylamide gel with that of wild type using ImageGauge software (**fig. S17**). The efficiency of [^{3}H]–DNA detection was estimated at approx. 20% as described previously (*12*) and tritium counts in cpm were converted to nM accordingly.

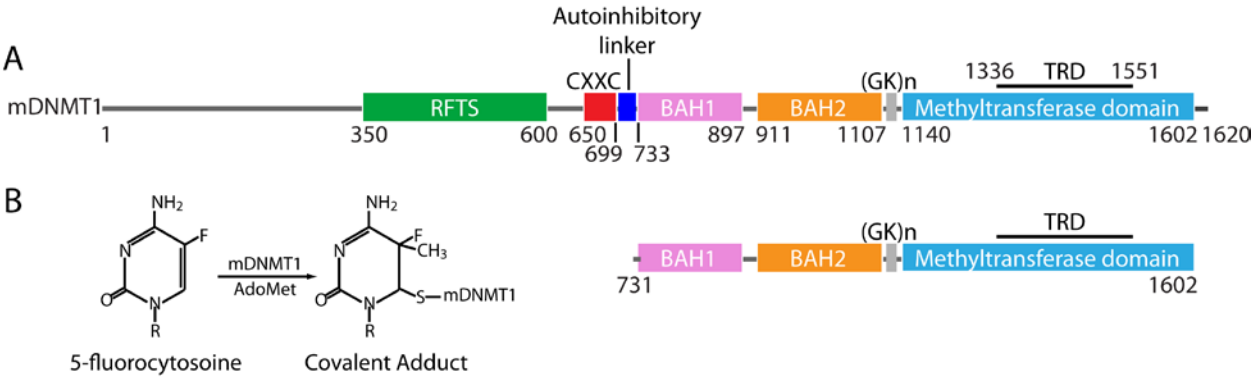

Figure S1. **Domain architecture of mDNMT1.**

A, Color-coded domain architecture and numbering of full-length mDNMT1. **B**, Chemical formula of the covalent cytosine adduct (left panel) and domain architecture and numbering system of the 731-1602 segment of mDNMT1 (right panel). mDNMT1(731-1602) shares about 85% sequence identity with its human counterpart.

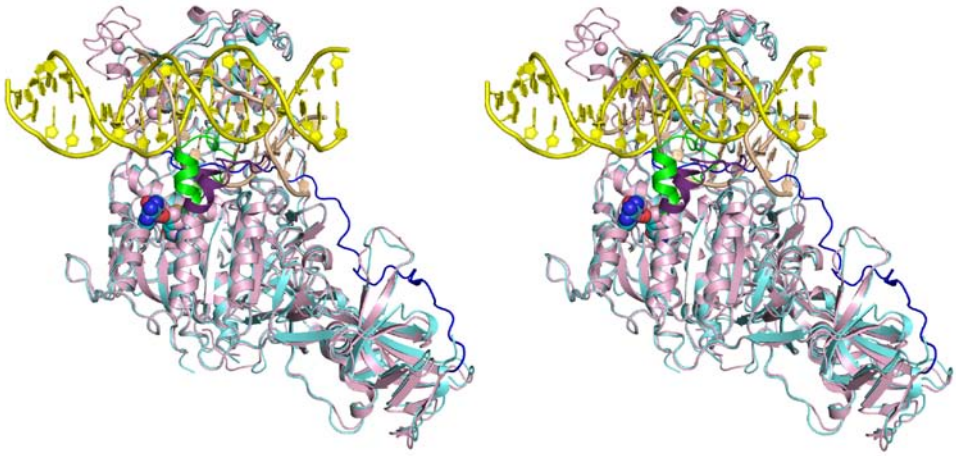

Figure S2. **Stereo view of the superimposed productive and autoinhibition mDNMT1-DNA complexes.**

The color scheme is the same as in Fig. 1B. The structure of mDNMT1 in complex with a 19-mer DNA duplex containing unmethylated CpG sites (3PT6) reveals that the CXXC domain specifically recognizes the unmethylated CpG dinucleotide, and in synergy with this recognition, the CXXC-BAH1 linker inhibits the DNMT1-mediated *de novo* DNA methylation through an autoinhibitory mechanism. Comparison of the currently determined productive mDNMT1-DNA complex with its autoinhibited counterpart identifies key conformational differences between the two complexes. First, the DNA duplex is positioned within the catalytic pocket (in beige) in the productive complex, but occluded (in yellow) in the autoinhibition complex. Second, there are extensive helical distortions in the DNA duplex spanning the hemi-mCG site in the productive complex associated with the dual flip-out of bases on partner strands and the resulting intercalation-type cavity, whereas the DNA duplex is unperturbed in the autoinhibition complex. Accordingly, base-specific intermolecular hydrogen bonds are only observed for the productive complex (Fig. 2, F and G), reflecting the much more extensive and penetrating protein-DNA contacts in the productive complex compared to the autoinhibition complex. Third, a helix within the catalytic domain, that is kinked (in purple, insert to Fig. 1B) in the autoinhibition complex, straightens out (in green, insert to Fig. 1B) in the productive complex, and in so doing, participates in redirecting the adjacent catalytic loop towards the catalytic pocket in the productive complex (see insert in Fig. 1B).

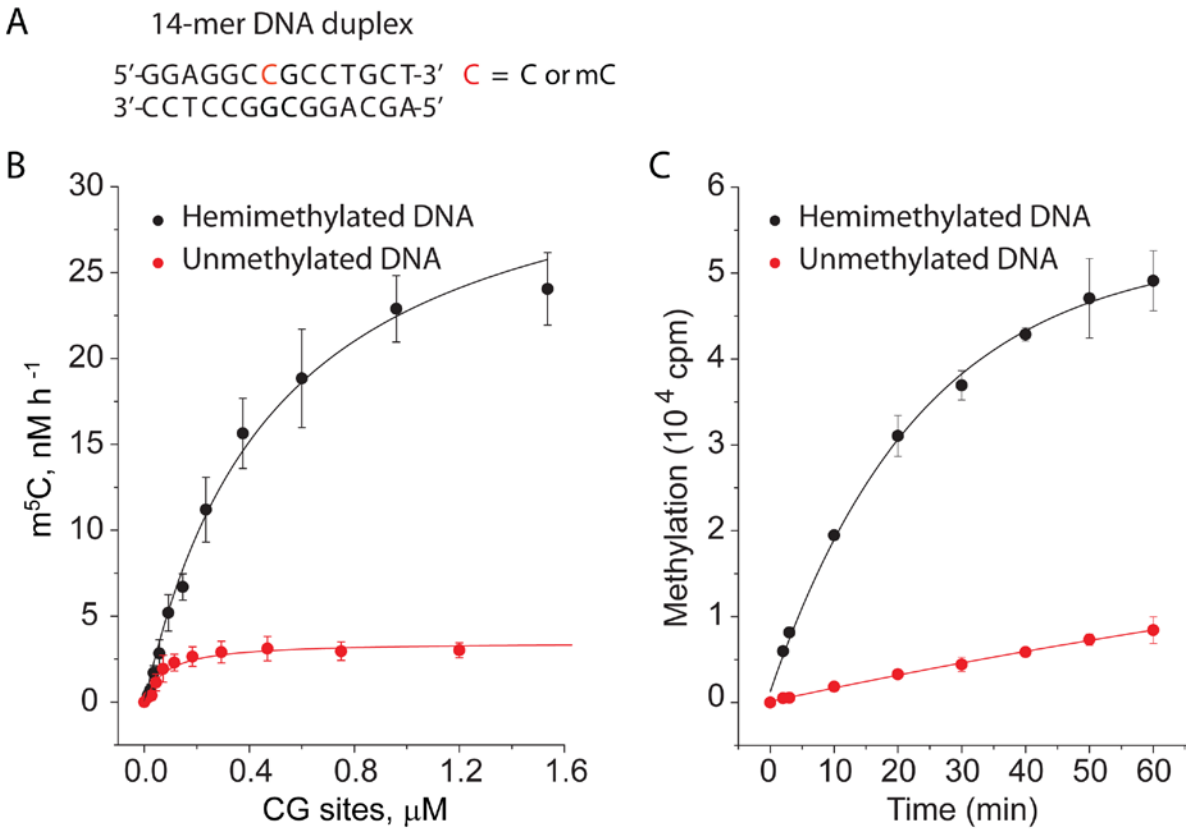

Figure S3. Efficiency of cytosine methylation of unmethylated and hemimethylated DNA by mDNMT1 (731-1602) protein.

A, Sequence of 14-mer DNA duplex. **B**, Amounts of $^{3}H\text{-}CH_3$ in nM per hr transferred to unmethylated (red circles) and hemimethylated (black circles) 14-mer DNA duplexes by 1 nM of DNMT1 plotted as a function of CpG site concentration; the steady-state Michaelis–Menten parameters estimated from these plots are listed in table S1. Each reaction point was repeated in triplicate; mean and SD values are shown. **C,** mDNMT1 (731-1602)-mediated cytosine methylation kinetics on DNA 14-mer substrates containing a central hemi-mCG (black circles) versus unmodified CG sites (red circles).

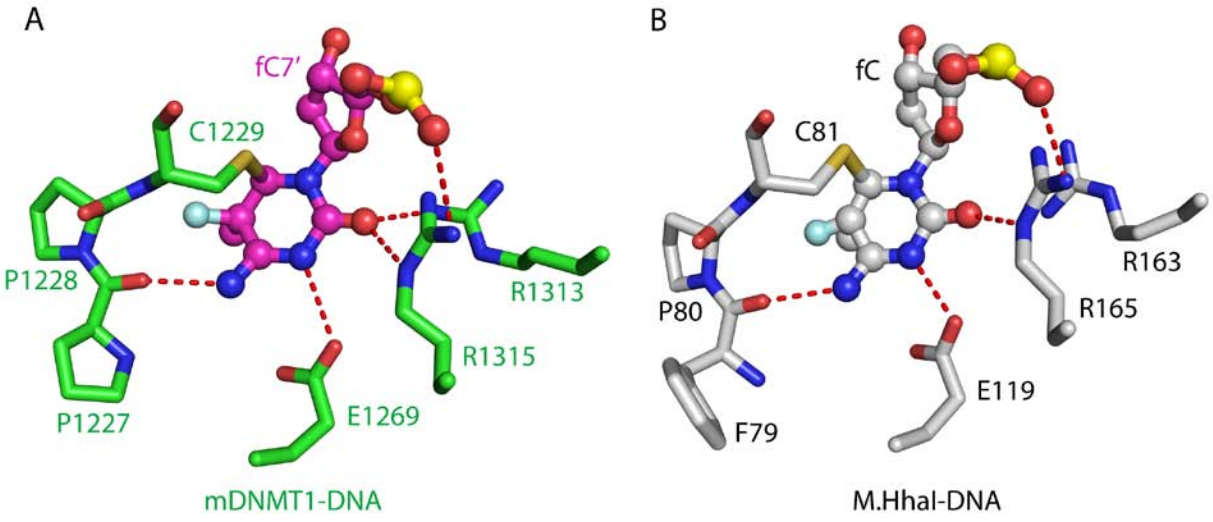

Figure S4. Recognition of flipped-out target cytosine (fC7' with attached 5-methyl group) from the hemi-mCG step by mDNMT1 in the productive covalent mDNMT1-DNA complex and comparison with bacterial M.*Hha*I-DNA complex.

A, B, The flipped-out target cytosine (fC7') is stabilized by residues in the active site of both mDNMT1 (**A**) and M.*Hha*I (**B**, PDB 1MHT) through intermolecular hydrogen bonds (red dashed lines). Inferring from their counterparts in M.*Hha*I (*14*), the functions of these residues are presumably tightly coupled to the enzymatic function of DNMT1. For instance, Arg1315 (equivalent to Arg165 in M.*Hha*I) is responsible for proper orientation of the flipped-out fC (*21*), while Glu1269 (equivalent to Glu119 in M.*Hha*I), is essential for activating the flipped out fC ring for nucleophilic attack through protonation of its N3 atom, and contributing to tight DNA binding.

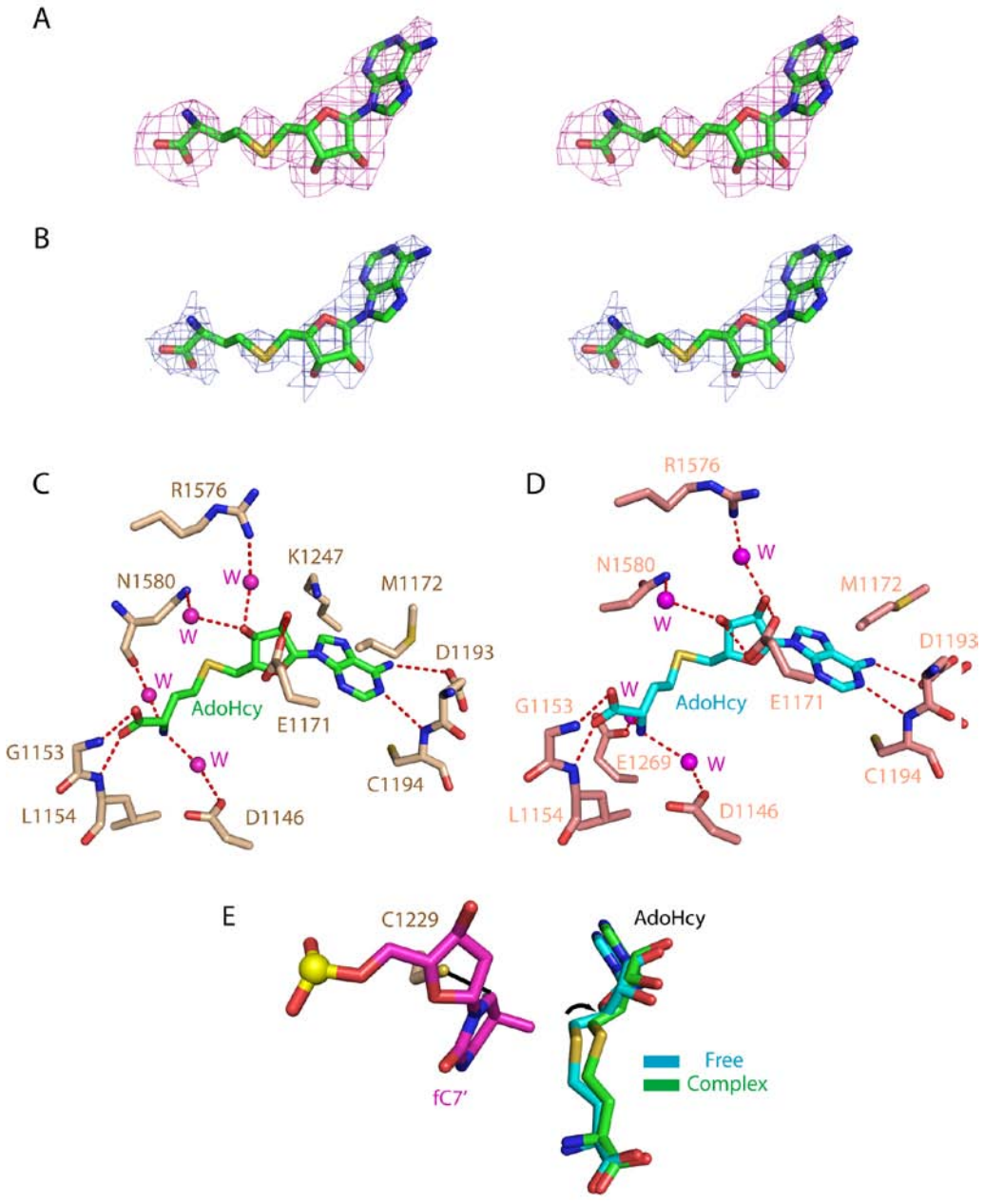
A
B
C
R1576
N1580
W
K1247
M1172
W
D1193
W
AdoHcy
E1171
G1153
W
C1194
L1154
D1146
D
R1576
W
N1580
M1172
W
D1193
G1153
W
AdoHcy
E1171
E1269
W
C1194
L1154
D1146
E
C1229
fC7'
AdoHcy
Free
Complex

Figure S5. **Comparison of bound AdoHcy in the productive covalent mDNMT1-DNA complex with that of mDNMT1 in the free state.**

A, Stereo view of the omit Fo-Fc map of the mDNMT1-bound AdoHcy (contoured at 2.0 σ) in the complex. **B**, Stereo view of the refined 2Fo-Fc map of the mDNMT1-bound AdoHcy (contoured at 1.0 σ) in the complex. **C**, Bound AdoHcy in the productive covalent mDNMT1-DNA complex is stabilized through a network of direct or water-mediated hydrogen bonds with the enzyme. Water molecules, shown in magenta spheres, are labeled "W". **D**, Interactions between AdoHcy and binding pocket residues in free mDNMT1 (PDB: 3PT9). **E**, Superposition of bound AdoHcy in free mDNMT1 and the productive covalent mDNMT1-DNA complex indicates a slight positional shift in response to the covalent link between Cys1229 and the flipped-out 5-fluorocytosine. We observe small conformational transitions on comparison of the cofactor-binding site on proceeding from mDNMT1 in the free state to its productive covalent complex with DNA (this study). Notably, the water-mediated hydrogen bond between the side chain carboxylate of Glu1269 and the NH_2 group in the methionine moiety of AdoHcy in free mDNMT1, is substituted by the water-mediated hydrogen bond from the backbone carbonyl oxygen of Asn1580 in the covalent mDNMT1-DNA complex. Another distinct difference involves a slight backward shift in the AdoHcy bound to the covalent complex so as to accommodate the newly added 5-methyl group in the flipped-out fC7'.

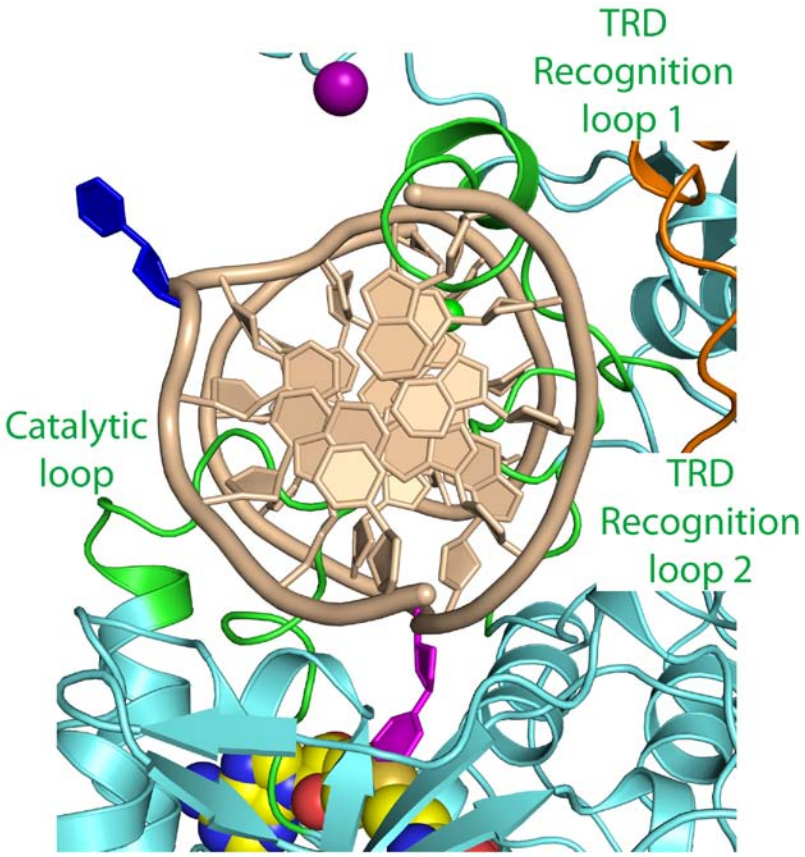

Figure S6. Penetration of the DNA grooves by catalytic and TRD loops in the productive covalent mDNMT1-DNA complex.

Overview of penetration of the catalytic loop into the DNA minor groove and the TRD loops 1 and 2 into the DNA major groove. All inserting loops are colored in green.

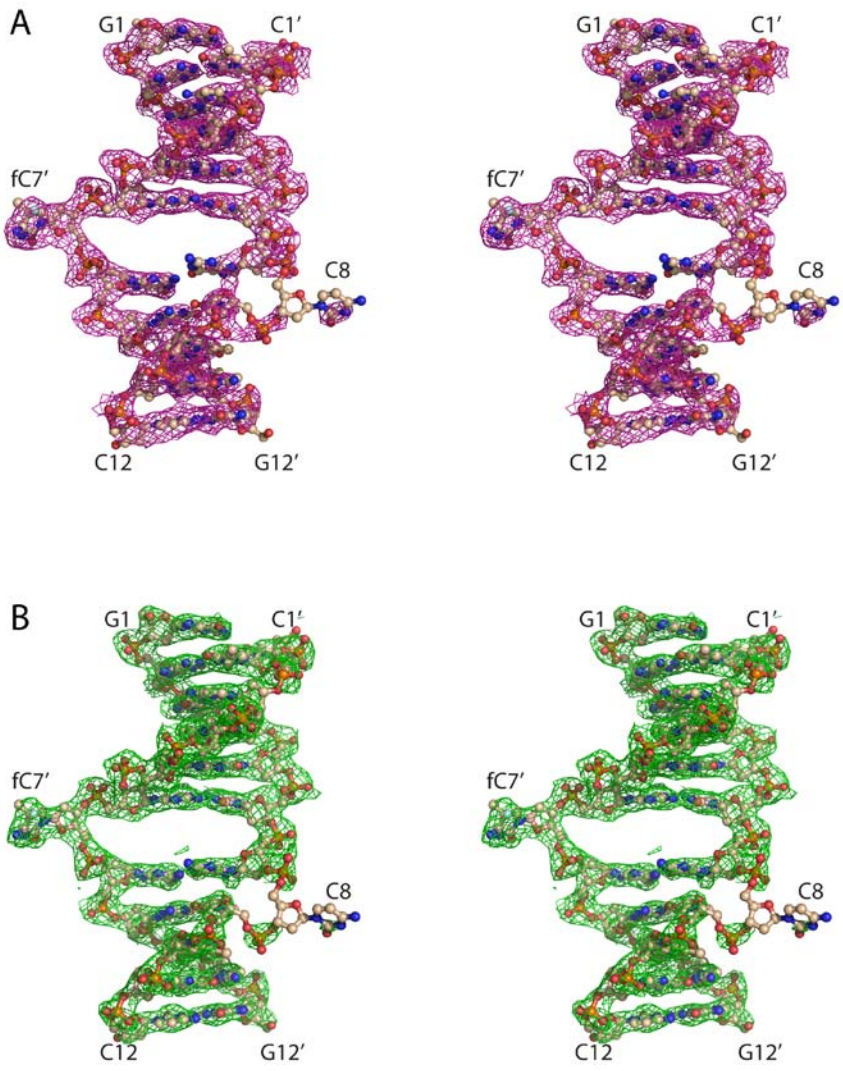

Figure S7. Electron density maps of bound DNA in the productive covalent mDNMT1-DNA complex.

A, Stereo view of the omit Fo-Fc map of the DNA in the productive covalent mDNMT1-DNA complex (contoured at 2.0 σ). **B**, Stereo view of the refined 2Fo-Fc map of the DNA in the productive covalent mDNMT1-bound DNA (contoured at 1.0 σ). Beyond the central CG site, the DNA molecule largely retains a B-form conformation towards both ends of the duplex, except that the terminal base pair at one end is disrupted due to crystal packing.

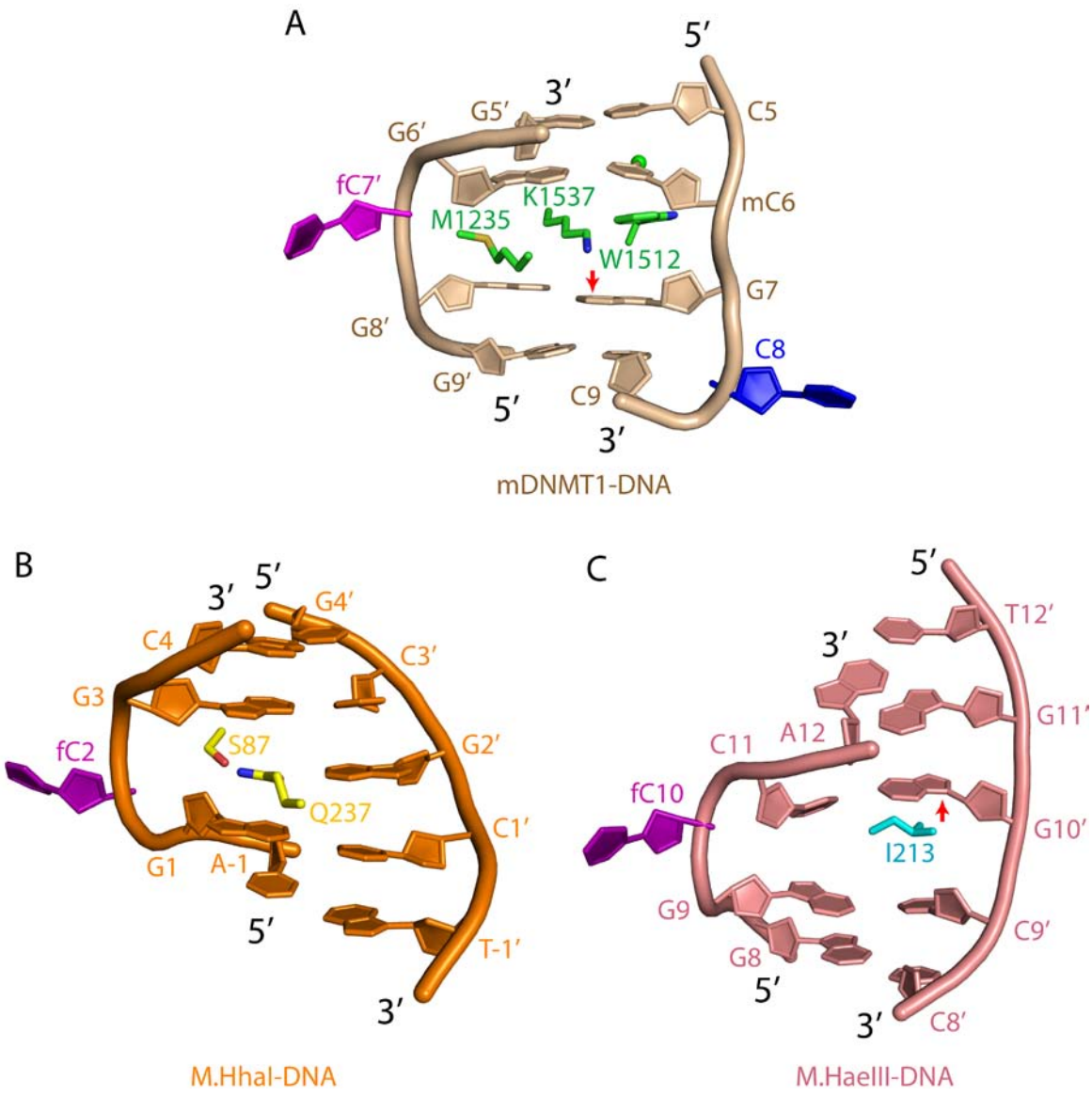

Figure S8. Conformational rearrangements of the recognition bases in the covalent A, mDNMT1-DNA, B, M.*Hha*I-DNA (pdb: 1MHT) and C, M. *Hae*III-DNA (pdb: 1DCT) complexes.

The flipped-out target cytosine is shown in purple. The protein residues that are inserted into DNA cavity are shown in stick representation. Notably, residue I221 of M. HaeIII slightly juts into the DNA cavity, without deep insertion. The numbering systems for (B) and (C) are based on original papers (*14, 15*). The directions toward which the orphan

guanine that normally pairs with the target cytosine is translated is marked by red arrows in panels A and C.

We compare below the covalent mammalian mDNMT1-DNA complex (panel A) with its bacterial counterparts (panels B and C). Similar to bacterial DNA methyltransferases, DNMT1 infiltrates DNA through insertion of both catalytic and TRD loops. However, the molecular basis underlying the distinct substrate recognition by DNMT1 has diverged significantly from that of its bacterial counterparts. In particular, the TRD loop 1 of DNMT1, which is responsible for forming the hydrophobic concave surface harboring the methyl group of 5-methyl cytosine on the parental strand, is a unique feature of DNMT1. The dramatic base-pairing realignments within the central hemi-mCG site observed for the mDNMT1-DNA complex (Fig. 2, D and E) is in striking contrast to the structures of bacterial methyltransferase-DNA complexes reported to date. To our knowledge, the structure of mDNMT1-DNA complex represents the first example where a dual base flip-out on partner strands (Fig. 2E) is involved in the enzymatic cytosine methylation reaction. In the bacterial M.*Hha*I complex, the residues equivalent to inserting residues Met1235 and Lys1537 of the DNMT1 complex, are Ser87 and Gln237, respectively (*14*). With smaller side chains, these residues did not perturb the position of the stacked orphan guanine significantly, and instead formed stabilizing hydrogen-bonding interactions, following flip-out of the target cytosine. It should be noted that a larger local rearrangement in base-pairing alignments centered on the DNA recognition sequence was observed for the bacterial M. *Hae*III-DNA complex (*15*), in which the orphan guanine partner of the flipped-out cytosine formed a base pair with the cytosine immediately downstream of the extruded cytosine on the target strand. Distinct from the dual flip-out observed in the mDNMT1-DNA complex, the opening of gaps in the DNA in the single base flip-out M. *Hae*III-DNA complex remained largely unfilled, resulting in a wide solvent channel.

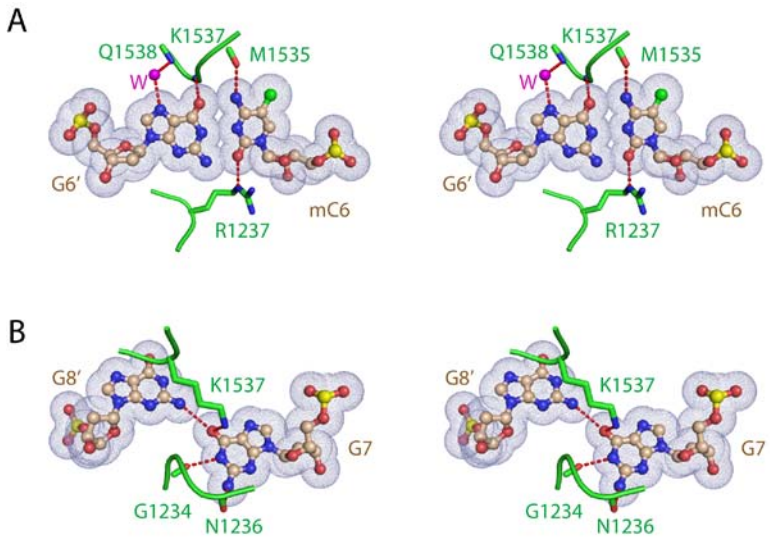

Figure S9. Base-pairing alignment recognition at the hemi-mCpG step in the productive covalent mDNMT1-DNA complex.

A, Stereo view of Watson-Crick mC6-G6' pair recognition. On the DNA minor groove side, Arg1237 forms hydrogen bonds to the O2 atom of mC6 through its guanidinium NεH proton. On the DNA major groove side, TRD loop 2 comes in close proximity to the mC6-G6' pair, initiating a network of backbone hydrogen bonds to mC6 and G6'. These include the backbone carbonyl oxygen of Met1535 that hydrogen bonds to the N4 atom of mC6, the backbone amide of Lys1537 that hydrogen bonds to the O6 atom of G6', and the backbone amide of Gln1538 that forms a water-mediated hydrogen bond with the N7 atom of G6'. **B**, Stereo view of the recognition of G7 of the non-canonical G7•G8' pair. The backbone carbonyls of Gly1234 and Asn1236 form hydrogen bonds with the N1 and N2 atoms of orphan G7, respectively. In addition, the side chain of Lys1537, which is inserted into the DNA cavity, forms a hydrogen bond with the O6 atom of G7 through its side chain ammonium group. Note that such interactions partly compensate for the enthalpy cost associated with flip-out of C8.

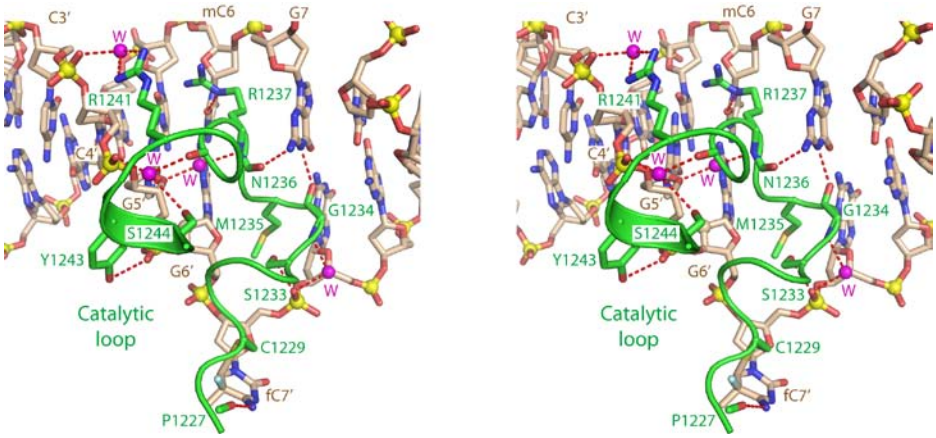

Figure S10. Stereo view of intermolecular contacts associated with insertion of catalytic loop into the DNA minor groove in the productive covalent mDNMT1-DNA complex.

The N-terminal end of the catalytic loop interacts with the extruded target fC7', while the rest of the loop involving Gly1234, Gln1236 and Arg1237 form base-specific contacts with G7 and mC6, with these interactions supplemented by a network of phosphate backbone contacts along the target strand. The nucleophilic Cys1229 forms a covalent bond with the C6 atom of the target cytosine C7', whose position is further fixed by the hydrogen bond between the backbone carbonyl oxygen of Pro1227 and the N4 atom of C7'. Phosphate backbone contacts on the target strand include the side chain hydroxyl group of Ser1233 and backbone amide of Gly1234 that form direct or water-mediated hydrogen bonds with the backbone phosphate of C7', while the side chain hydroxyl group of Tyr1243 forms a hydrogen bond to the backbone phosphate of G5'. In addition, the backbone carbonyl oxygen of Phe1238 and the side chain hydroxyl group of Ser1244, mediated by the same water molecule, hydrogen bond to the backbone phosphate of C4', while the guanidinium group of Arg1241 forms a water-mediated hydrogen bond with the backbone phosphate of C3'. An additional interaction includes a water-mediated hydrogen bond between the backbone amide of Arg1237 and the O4 sugar ring atom of G5'.

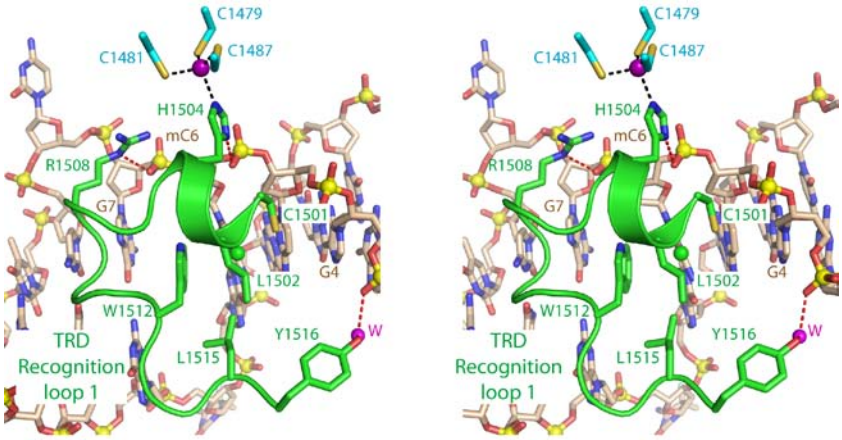

Figure S11. Stereo view of intermolecular contacts associated with insertion of TRD loop 1 into DNA major groove in the productive covalent mDNMT1-DNA complex.

Residues from TRD loop 1 are mainly engaged in recognition of the 5-methyl group of mC6 that is positioned within a hydrophobic concave surface lined by Cys1501, Leu1502, Trp1512 and Leu1515. This alignment is anchored by three hydrogen bonds with the DNA phosphate backbone along the parental strand. These include a hydrogen bond between the imidazole ring of His1504 and the backbone phosphate of mC6, a hydrogen bond between the guanidinium NεH proton of Arg1508 and the backbone phosphate of G7, and a water-mediated hydrogen bond between the side chain hydroxyl group of Tyr1516 and the backbone phosphate of G3

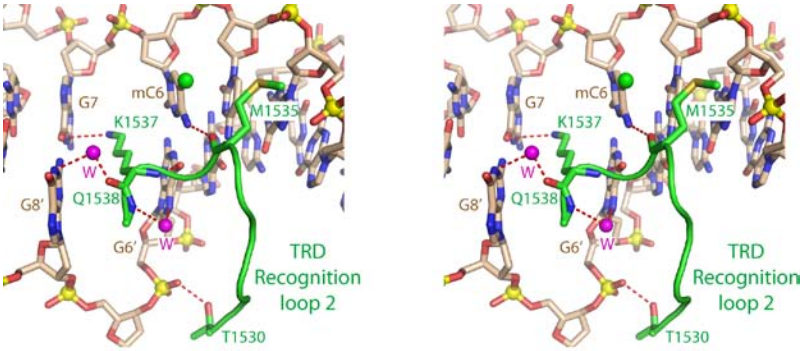

Figure S12. Stereo view of intermolecular contacts associated with insertion of TRD loop 2 into the DNA major groove in the productive covalent mDNMT1-DNA complex.

Residues from TRD loop 2 are responsible for base-specific recognition of the Watson-Crick mC6-G6' base pair, as well as the orphaned G7 base involved in non-canonical G7•G8' pairing. The positioning of this loop is further facilitated by a hydrogen bond formed between the side chain hydroxyl group of Thr1530 and the backbone phosphate of G6'. In addition, the side chain of Met1535 is brought in close proximity to the 5-methyl group of mC6 by a hydrogen bond formed between its backbone carbonyl oxygen and the N4 atom of mC6, thereby contributing to the formation of a hydrophobic environment around the 5-methyl group of mC6.

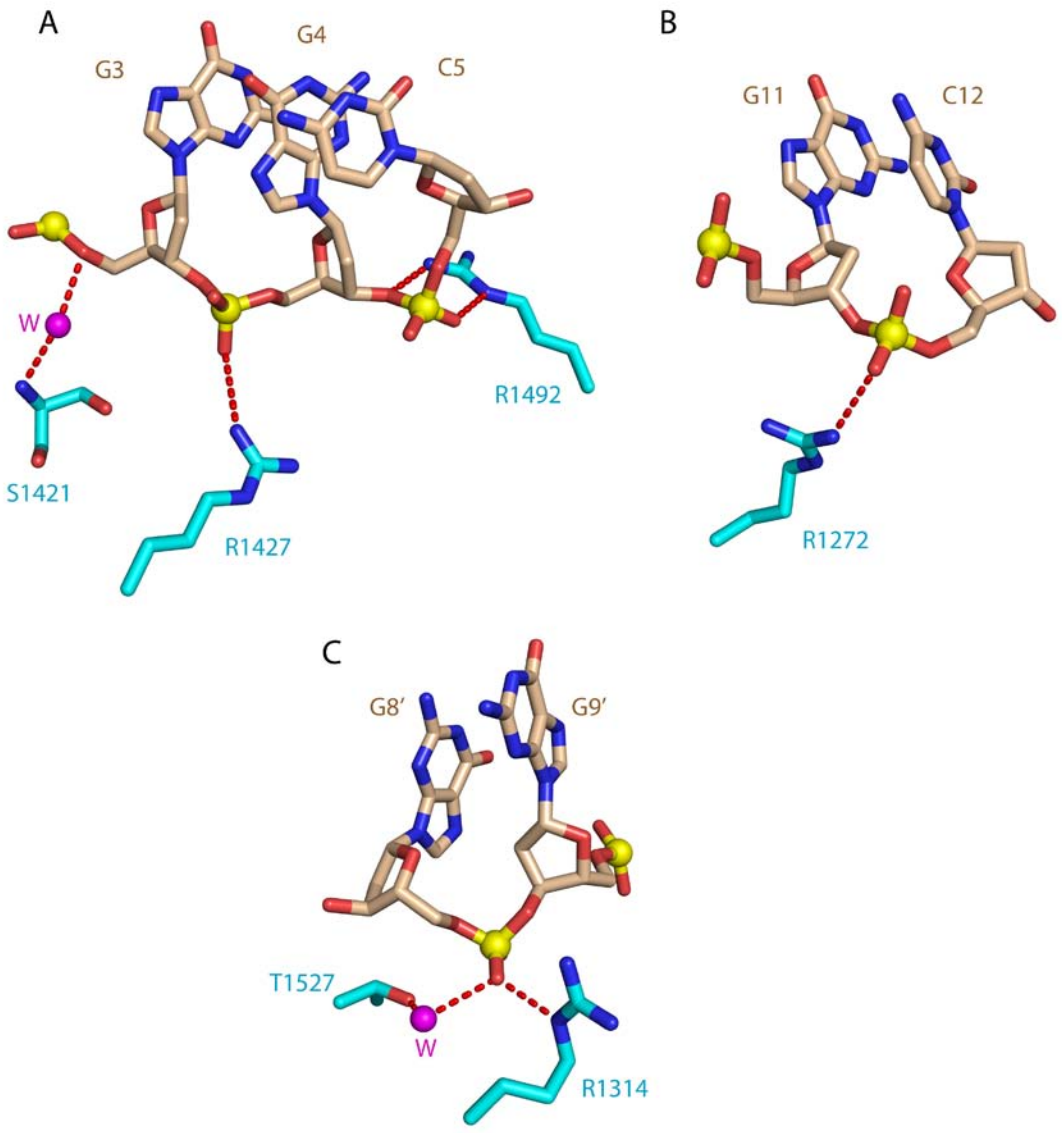

Figure S13. **Hydrogen-bonding interactions between mDNMT1 and the DNA phosphate backbone that extend beyond the hemi-mCG site in the productive covalent mDNMT1-DNA complex.**

A, Residues Ser1421, Arg1427 and Arg1492 make direct or water-mediated contacts with the DNA phosphate backbone of the parental strand upstream to mC6. **B**, At one end of the DNA, the side chain of Arg1272 forms a hydrogen bond with the backbone phosphate of C12 of the parental strand. **C**, Side chains of Arg1314 and Thr1527 form direct or water-mediated hydrogen bonds with the backbone phosphate of G8’ on the target strand.

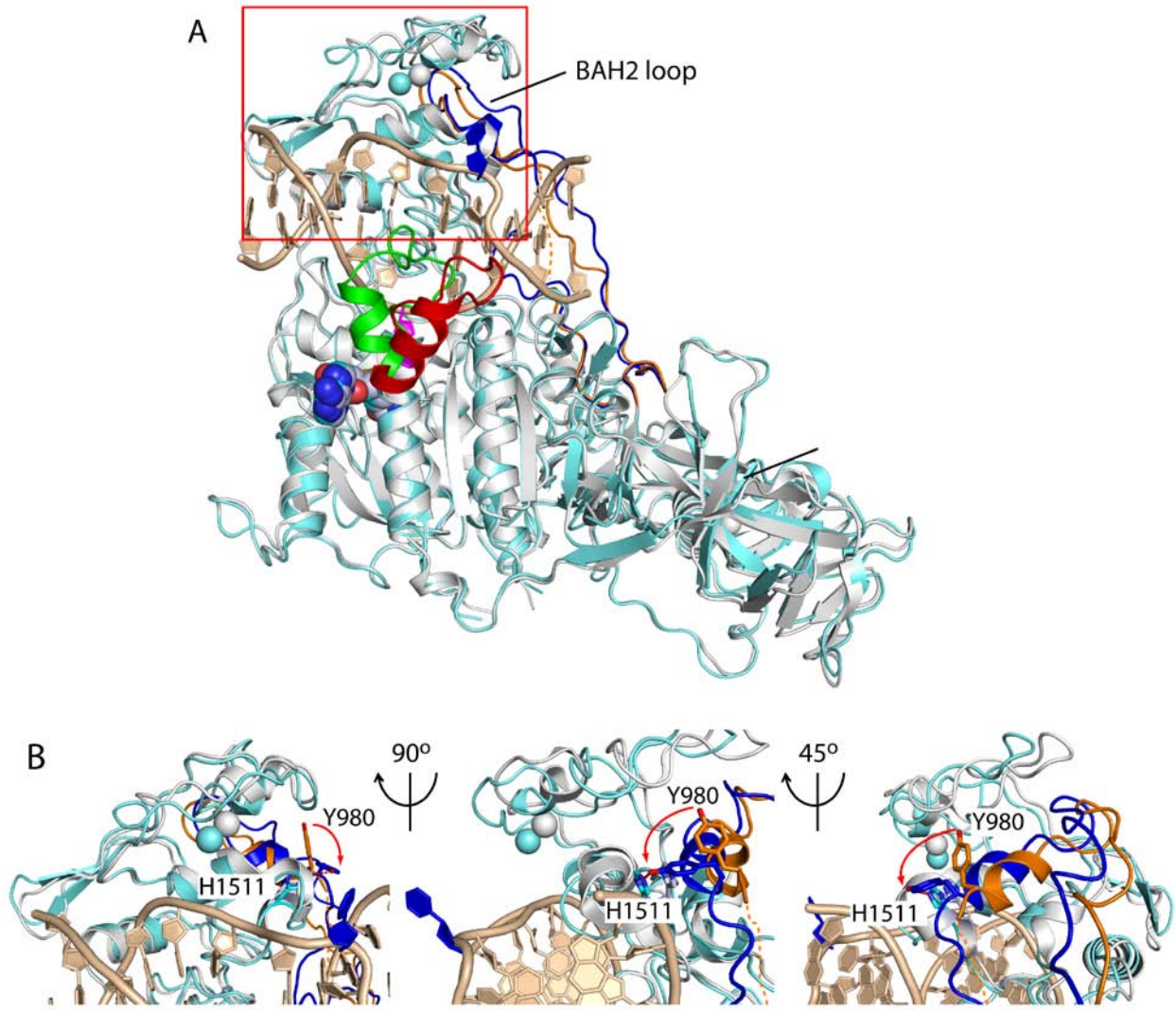
A
BAH2 loop
B
90°
Y980
H1511
Y980
H1511
45°
Y980
H1511

Figure S14. **Conformational change of TRD and BAH2 loops of mDNMT1 upon complex formation with DNA.**

A, Structural superposition of the productive covalent mDNMT1-DNA complex and free mDNMT1. The bound DNA in the covalent complex is in beige, and the protein in cyan, while free mDNMT1 in silver. The TRD is boxed in a red square. The catalytic loop and the N-terminal end of the following α-helix show the largest conformational change between the two states of mDNMT1. They are colored red in free mDNMT1 and green in the covalent complex. The BAH loop is in orange in free mDNMT1 and blue in the covalent complex. **B**, Close-up views showing the conformational rearrangement of the region around the tip of the BAH2 loop. Tyr980 realigns its aromatic ring and forms a hydrogen bond with the side chain of His1511 upon complex formation.

In the context of substrate specific recognition, it appears that the BAH2 loop does not discriminate the methylation state of the substrates. Instead, it binds to the hemimethylated substrates in synergy with the TRD.

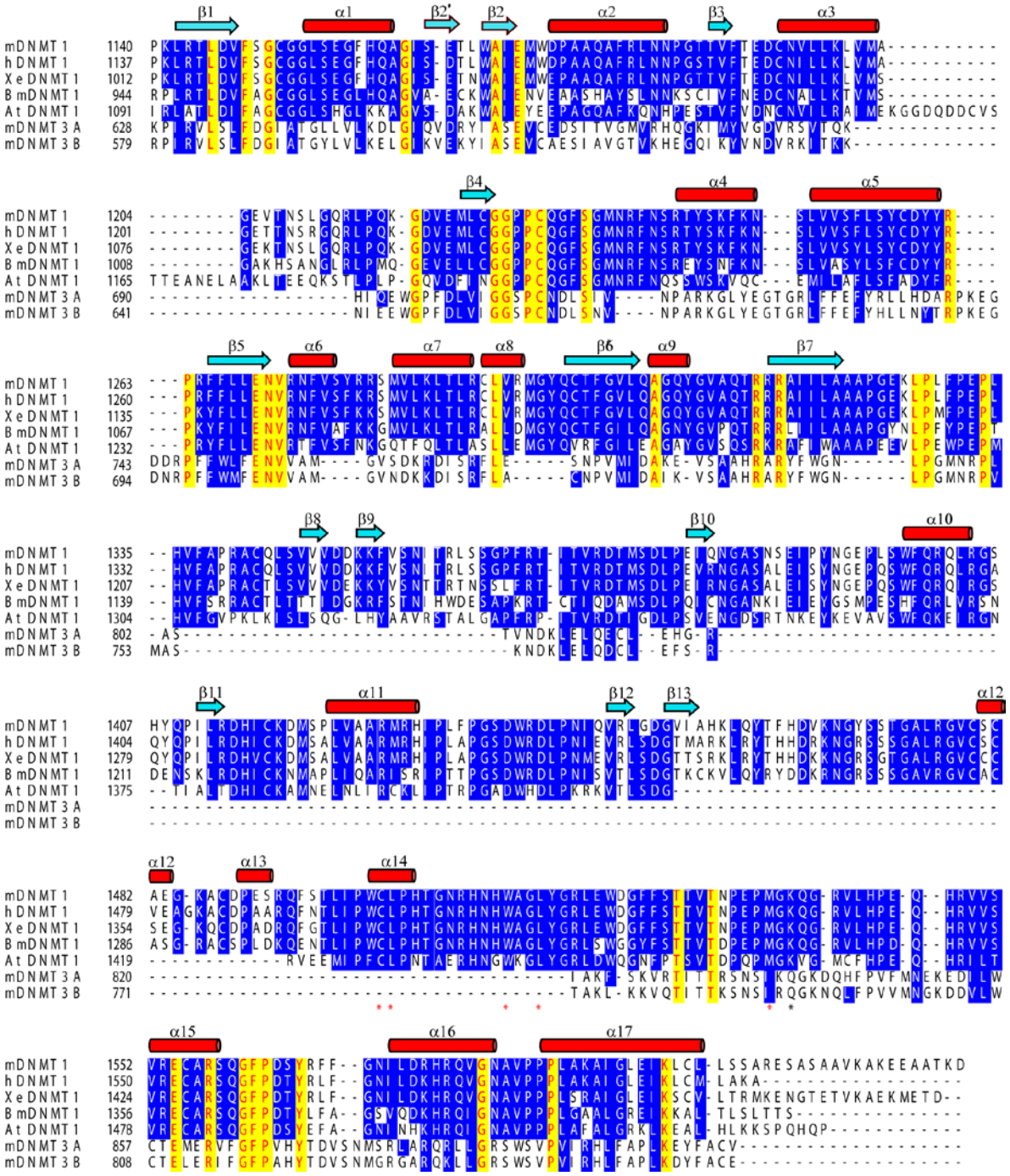
β1 α1 β2' β2 α2 β3 α3
mDNMT1 1140 PKLRTLDVFSGCGGLSEGFHQAGISETLWAIEMWDPAAQAFRLNNPGTTVFTEDCNVLLKLVMA----------
hDNMT1 1137 PKLRTLDVFSGCGGLSEGFHQAGISDTLWAIEMWDPAAQAFRLNNPGSTVFTEDCNILLKLVMA----------
XeDNMT1 1012 PKLRTLDVFSGCGGLSEGFHQAGISETNWAIEMWEPAAQAFRLNNPGTTVFTEDCNILLKLVMS----------
BmDNMT1 944 RPLRTLDVFAGCGGLSEGLHQAGVAECKWAIENVEAASHAYSLNNKSCIVFNEDCNALLKTVMS----------
AtDNMT1 1091 IRLATLDIFAGCGGLSHGLKKAGVSDAKWAIEYEEPAGQAFKQNHPESTVFVDNCNVILRALMEKGGDQDDCVS
mDNMT3A 628 KPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQK----------
mDNMT3B 579 RPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYVNDVRKITKK----------
β4 α4 α5
mDNMT1 1204 --------GEVTNSLGQRLPQK-GDVEMLCGGPPCQGFSGMNRFNSRTYSKFKN---SLVVSFLSYCDYYR----
hDNMT1 1201 --------GETTNSRGQRLPQK-GDVEMLCGGPPCQGFSGMNRFNSRTYSKFKN---SLVVSFLSYCDYYR----
XeDNMT1 1076 --------GEKTNSLGQRLPQK-GDVEMLCGGPPCQGFSGMNRFNSRTYSKFKN---SLVVSYLSYCDYYR----
BmDNMT1 1008 --------GAKHSANGLRLPMQ-GEVELLCGGPPCQGFSGMNRFNSREYSNFKN---SLVASYLSFCDYYR----
AtDNMT1 1165 TTEANELAAKLTEEQKSTLPLP-GQVDFINGGPPCQGFSGMNRFNQSSWSKVQC---EMILAFLSFADYFR----
mDNMT3A 690 ------------------HIQEWGPFDLVIGGSPCNDLSIV----NPARKGLYEGTGRLFFEFYRLLHDARPKEG
mDNMT3B 641 ------------------NIEEWGPFDLVIGGSPCNDLSNV----NPARKGLYEGTGRLFFEFYHLLNYTRPKEG
β5 α6 α7 α8 β6 α9 β7
mDNMT1 1263 ---PRFFLLENVRNFVSYRRSMVLKLTLRCLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPL
hDNMT1 1260 ---PRFFLLENVRNFVSFKRSMVLKLTLRCLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPL
XeDNMT1 1135 ---PKYFLLENVRNFVSFKKSMVLKLTLRCLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPMFPEPL
BmDNMT1 1067 ---PKYFILENVRNFVAFKKGMVLKLTLRALLDMGYQCTFGILQAGNYGVPQTRRRLIILAAAPGYNLPFYPEPT
AtDNMT1 1232 ---PRYFLLENVRTFVSFNKGQTFQLTLASLLEMGYQVRFGILEAGAYGVSQSRKRAFIWAAAPESVLPEWPEPM
mDNMT3A 743 DDRPFFWLFENVVAM----GVSDKRDISRFLE-----SNPVMIDAKE-VSAAHRARYFWGN------LPGMNRPL
mDNMT3B 694 DNRPFFWMFENVVAM----GVNDKKDISRFLA-----CNPVMIDAIK-VSAAHRARYFWGN------LPGMNRPV
β8 β9 β10 α10
mDNMT1 1335 --HVFAPRACQLSVVVDDKKFVSNITRLSSGPFRT-ITVRDTMSDLPEVRNGASNSEIPYNGEPLSWFQRQLRGS
hDNMT1 1332 --HVFAPRACQLSVVVDDKKFVSNITRLSSGPFRT-ITVRDTMSDLPEVRNGASALEISYNGEPQSWFQRQLRGA
XeDNMT1 1207 --HVFAPRACTLSVVVDEKKYVSNTTRTNSSLFRT-ITVRDTMSDLPEIRNGASALEISYNGEPQSWFQRQIRGS
BmDNMT1 1139 --HVFSRRACTLTTTIDGKRFSTNIHWDESAPKRT-CTIQDAMSDLPQICNGANKIELEYGSMPESHFQRLVRSN
AtDNMT1 1304 --HVFGVPKLKISLSQG-LHVAAVRSTALGAPFRP-ITVRDIIGDLPSVENGDSRTNKEYKEVAVSWFQKELRGN
mDNMT3A 802 -AS----------TVNDKLELQECL--EHG-R----------
mDNMT3B 753 MAS----------KNDKLELQDCL--EFS-R----------
β11 α11 β12 β13 α12
mDNMT1 1407 HYQPILRDHICKDMSPLVAARMRHIPLFPGSDWRDLPNIQVRLGDGVIAHKLQYTFHDVKNGYSSIGALRGVCSC
hDNMT1 1404 QYQPILRDHICKDMSALVAARMRHIPLAPGSDWRDLPNIEVRLSDGTMARKLRYTHHDRKNGRSSSGALRGVCSC
XeDNMT1 1279 QYQPILRDHVCKDMSALVAARMRHIPLAPGSDWRDLPNMEVRLSDGTTSRKLRYTHHDKKNGRSGTGALRGVCCC
BmDNMT1 1211 DENSKLRDHICKNMAPLIQARISRIPTTPGSDWRDLPNISVTLSDGTKCKVLQYRYDDKRNGRSSSGAVRGVCAC
AtDNMT1 1375 --TIALTDHICKAMNELNLIRCKLIPTRPGADWHDLPKRKVTLSDG----------
mDNMT3A ----------
mDNMT3B ----------
α12 α13 α14
mDNMT1 1482 AEG-KACDPESRQFSTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPMGKQG-RVLHPE-Q--HRVVS
hDNMT1 1479 VEAGKACDPAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPMGKQG-RVLHPE-Q--HRVVS
XeDNMT1 1354 SEG-KQCDPADRQFGTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPMGKQG-RVLHPE-Q--HRVVS
BmDNMT1 1286 ASG-RACSPLDKQENTLIPWCLPHTGNRHNHWAGLYGRLSWGYFSTTVTDPEPMGKQG-RVLHPD-Q--HRVVS
AtDNMT1 1419 ----------RVEEMIPFCLPNTAERHNGWKGLYGRLDWQGNFPTSVTDPQPMGKVG-MCFHPE-Q--HRILT
mDNMT3A 820 ----------TAKL-SKVRTITTRSNSIKQGKDQHFPVFMNEKEDILW
mDNMT3B 771 ----------TAKL-KKVQTITTKSNSLRQGKNQLFPVVMNGKDDVLW
α15 α16 α17
mDNMT1 1552 VRECARSQGFPDSYRFF--GNILDRHRQVGNAVPPPLAKAIGLEIKLCL-LSSARESASAAVKAKEEAATKD
hDNMT1 1550 VRECARSQGFPDTYRLF--GNILDKHRQVGNAVPPPLAKAIGLEIKLCM-LAKA----------
XeDNMT1 1424 VRECARSQGFPDTYRLF--GNILDKHRQVGNAVPPPLSRAIGLEIKSCV-LTRMKENGTETVKAEKMETD--
BmDNMT1 1356 VRECARSQGFPDTYLFA--GSVQDKHRQIGNAVPPPLGAALGREIKKAL-TLSLTTS----------
AtDNMT1 1478 VRECARSQGFPDSYEFA--GNINHKHRQIGNAVPPPLAFALGRKLKEAL-HLKKSPQHQP----------
mDNMT3A 857 CTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV----------
mDNMT3B 808 CTELERIFGFPAHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE----------

Figure S15. **Structure-based sequence conservation analysis of DNMT1 and DNMT3 methyltransferase domains.**

Identical or similar residues are colored in white and highlighted in blue. Completely conserved residues are colored in red and highlighted in yellow. The DNMT1 sequences are selected from *mus musculus* (mDNMT1), *homo sapiens* (hDNMT1), *Xenopus laevis* (XeDNMT1), *Bombyx mori* (BmDNMT1) and *Arabidopsis thaliana* (AtDNMT1). The DNMT3A and DNMT3B sequences are selected from *mus musculus* (mDNMT3A and mDNMT3B, respectively). Under the alignment, the residues lining the 5-methyl group of mC6 are marked with red asterisks, while those residues that are inserted into the intercalative-like gap in the DNA are marked with black asterisks. Note that these residues, while highly conserved in DNMT1 family, are not conserved in DNMT3A and DNMT3B.

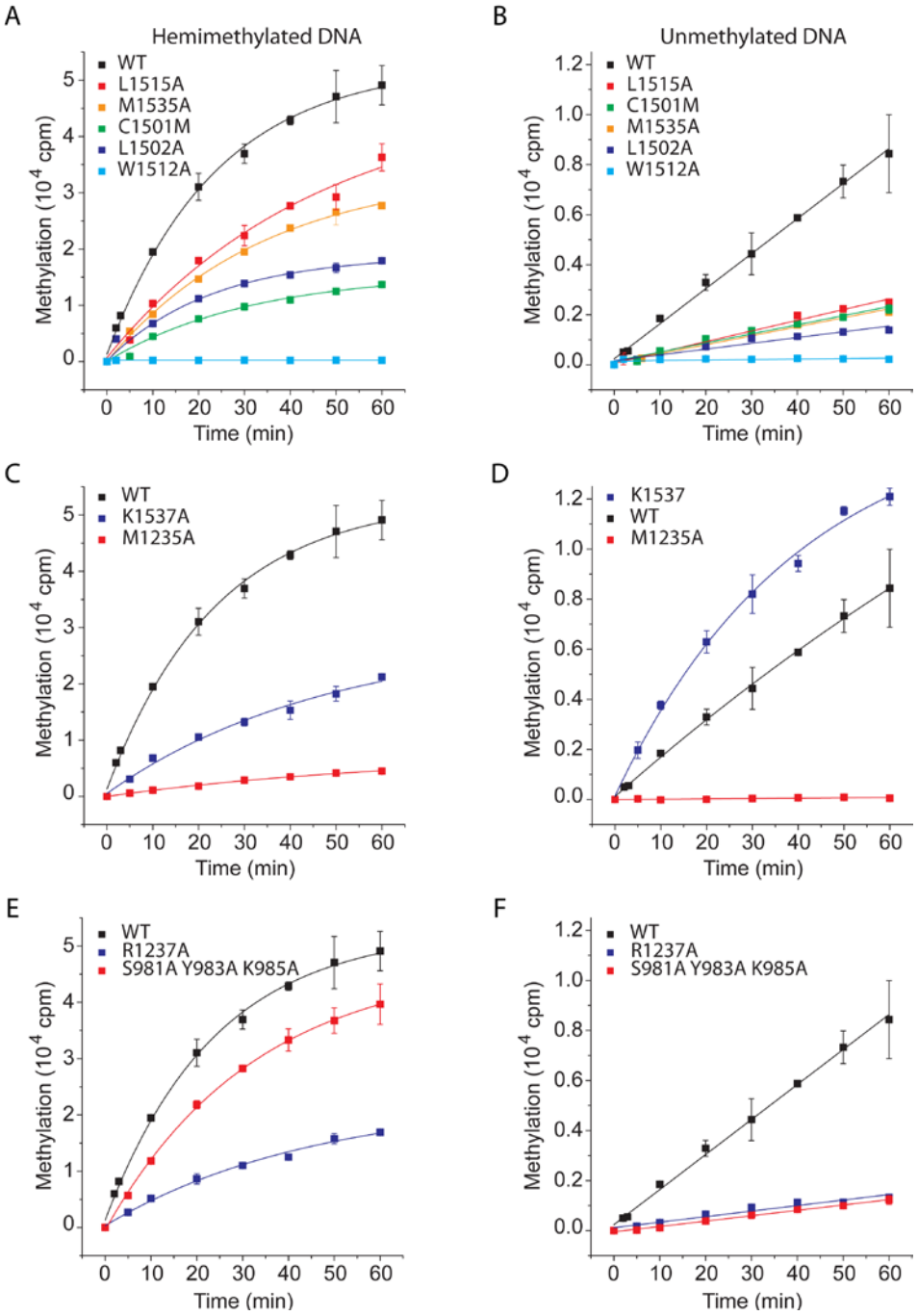
A
Hemimethylated DNA
WT
L1515A
M1535A
C1501M
L1502A
W1512A
Methylation (10^4 cpm)
Time (min)
B
Unmethylated DNA
WT
L1515A
C1501M
M1535A
L1502A
W1512A
Methylation (10^4 cpm)
Time (min)
C
WT
K1537A
M1235A
Methylation (10^4 cpm)
Time (min)
D
K1537
WT
M1235A
Methylation (10^4 cpm)
Time (min)
E
WT
R1237A
S981A Y983A K985A
Methylation (10^4 cpm)
Time (min)
F
WT
R1237A
S981A Y983A K985A
Methylation (10^4 cpm)
Time (min)

Figure S16. **mDNMT1(731-1602)-mediated cytosine methylation kinetics of selected amino acid mutants on hemimethylated and unmethylated CpG-containing DNA substrates.**

A, Methylation kinetics of amino acid mutants lining the mC-recognizing hydrophobic surface on hemi-mCG DNA substrate. **B**, Corresponding kinetic data to panel A on unmodified CG DNA substrate. **C**, Methylation kinetics of amino acid mutants inserting into the intercalation-type cavity on hemi-mCG DNA substrate. **D**, Corresponding kinetic data to panel C on unmodified CG DNA substrate. **E**, Methylation kinetics of amino acid mutants from catalytic, TRD and BAH2 loops on hemi-mCG DNA substrate. **F**, Corresponding kinetic data to panel E on unmodified CG DNA substrate.

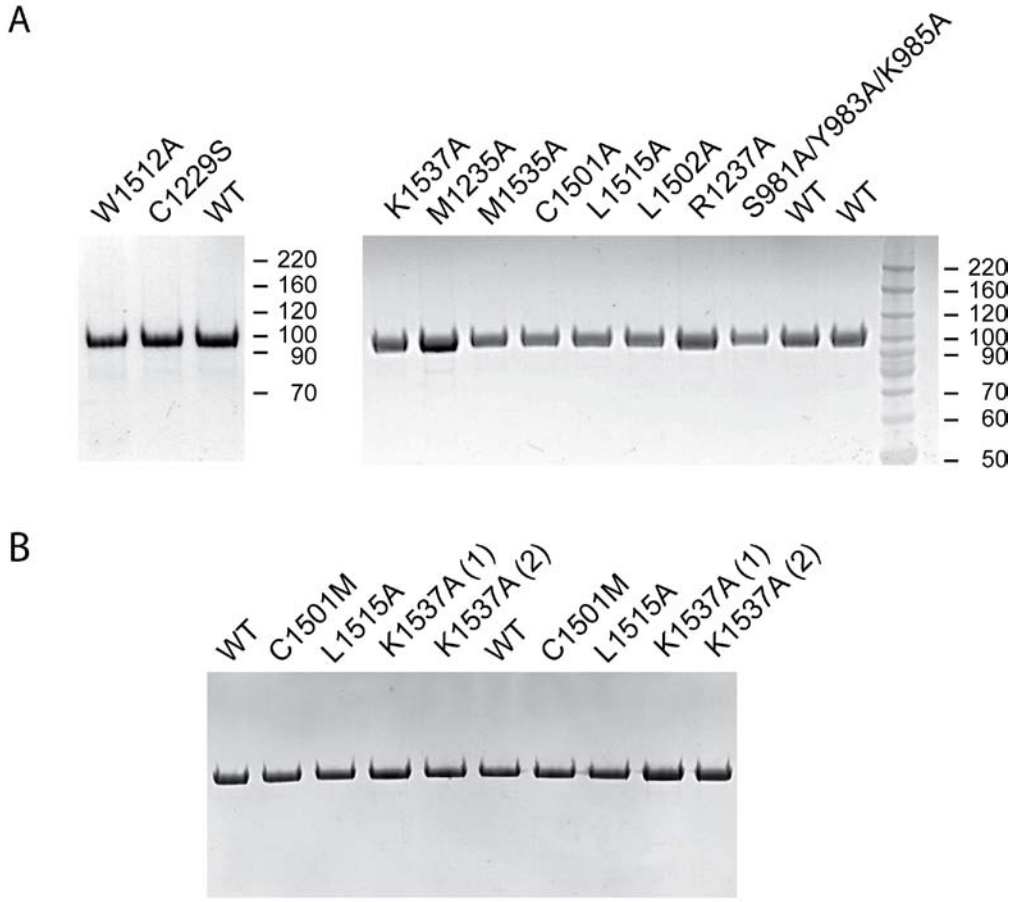

Figure S17. Quantitative SDS polyacrylamide gel (NuPage 4-12% Bis-Tris PAG, Invitrogen) electrophoresis of mDNMT1 (731-1602) and mutants used in enzymatic assays.

A, Representative gels showing that mDNMT1 (731-1602) wild type (WT) and mutants indicated above each lane were at comparable concentration. Standard molecular mass BenchMark protein ladder (Invitrogen) is in the right lane. **B**, The SDS polyacrylamide gel comparing concentrations of the WT, C1501M, L1515A and two K1537A mutant preparations K1537A (1) and K1537A (2) used for kinetics parameters calculations (Fig. 3E and fig. S15). Equal amounts of 2 μM protein samples were loaded on the gel in duplicate and concentrations of the mutants were determined by measuring average band intensity relative to that of WT.

table S1. **DNMT1 (731-1602) catalyzed enzymatic methylation assays on unmethylated (UM) and hemimethylated (HM) 14-mer DNA**

DNA substrate	k_{cat} h^{-1}	K_M μM	k_{cat}/K_M h^{-1} μM^{-1}
UM single CpG 14-mer	3.6 ± 0.3	0.09 ± 0.02	40 ±10
HM single CpG 14-mer	34 ± 3	0.5 ± 0.1	68 ± 15

Relative preference for hemimethylated over unmethylated DNA for 14-mer DNA

Relative k_{cat} (HM/UM) 14-mer	Relative k_{cat}/K_M (HM/UM) 14-mer
9 ± 1	1.7 ± 0.6

table S2. Crystallographic statistics for the covalent mDNMT1-DNA complex.

Data collection	
Wavelength (Å)	0.9792
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	89.7, 152.0, 96.1
α, β, γ (°)	90, 94.5, 90
Resolution (Å)*	50-2.6 (2.69-2.6)
R_{sym} or R_{merge}*	9.6 (47.4)
$I/\sigma I$	12.4 (1.9)
Completeness (%)*	97.9 (98.0)
Redundancy*	2.7 (2.5)
Unique reflections	77,126 (7,658)
Refinement	
Resolution (Å)	39.5-2.6
No. reflections	77,070
R_{work}/ R_{free}	19.6/24.6
No. atoms	
Protein	13,154
DNA	978
AdoHcy	52
Zinc	4
Water	860
B-factors	
Protein	42.2
DNA	61.4
AdoHcy	34.1
Zinc	39.8
Water	35.1
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (º)	1.190

*Highest resolution shell is shown in parenthesis.

References

1. X. Cheng, R. M. Blumenthal, Mammalian DNA methyltransferases: A structural perspective. *Structure* **16**, 341 (2008). doi:10.1016/j.str.2008.01.004 Medline

2. M. G. Goll, T. H. Bestor, Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481 (2005). doi:10.1146/annurev.biochem.74.010904.153721 Medline

3. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204 (2010). doi:10.1038/nrg2719 Medline

4. M. M. Suzuki, A. Bird, DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465 (2008). doi:10.1038/nrg2341 Medline

5. E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915 (1992). doi:10.1016/0092-8674(92)90611-F Medline

6. L. S. Chuang *et al*., Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**, 1996 (1997). doi:10.1126/science.277.5334.1996 Medline

7. M. R. Rountree, K. E. Bachman, S. B. Baylin, DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* **25**, 269 (2000). doi:10.1038/77023 Medline

8. H. Leonhardt, A. W. Page, H. U. Weier, T. H. Bestor, A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865 (1992). doi:10.1016/0092-8674(92)90561-P Medline

9. J. H. Lee, K. S. Voo, D. G. Skalnik, Identification and characterization of the DNA binding domain of CpG-binding protein. *J. Biol. Chem.* **276**, 44669 (2001). doi:10.1074/jbc.M107179200 Medline

10. M. Pradhan *et al*., CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry* **47**, 10000 (2008). doi:10.1021/bi8011725 Medline

11. I. Callebaut, J. C. Courvalin, J. P. Mornon, The BAH (bromo-adjacent homology) domain: A link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* **446**, 189 (1999). doi:10.1016/S0014-5793(99)00132-5 Medline

12. J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036 (2011). doi:10.1126/science.1195380 Medline

13. D. G. Osterman, G. D. DePillis, J. C. Wu, A. Matsuda, D. V. Santi, 5-Fluorocytosine in DNA is a mechanism-based inhibitor of HhaI methylase. *Biochemistry* **27**, 5204 (1988). doi:10.1021/bi00414a039 Medline

14. S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357 (1994). doi:10.1016/0092-8674(94)90342-5 Medline

15. K. M. Reinisch, L. Chen, G. L. Verdine, W. N. Lipscomb, The crystal structure of HaeIII methyltransferase covalently complexed to DNA: An extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143 (1995). doi:10.1016/0092-8674(95)90060-8 Medline

16. K. Takeshita *et al*., Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9055 (2011). doi:10.1073/pnas.1019629108 Medline

17. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658 (2007). doi:10.1107/S0021889807021206 Medline

18. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126 (2004). doi:10.1107/S0907444904019158 Medline

19. P. D. Adams *et al*., PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948 (2002). doi:10.1107/S0907444902016657 Medline

20. J. Flynn, J. F. Glickman, N. O. Reich, Murine DNA cytosine-C5 methyltransferase: Pre-steady- and steady-state kinetic analysis with regulatory DNA sequences. *Biochemistry* **35**, 7308 (1996). doi:10.1021/bi9600512 Medline

21. F. K. Shieh, B. Youngblood, N. O. Reich, The role of Arg165 towards base flipping, base stabilization and catalysis in M.HhaI. *J. Mol. Biol.* **362**, 516 (2006). doi:10.1016/j.jmb.2006.07.030 Medline