Structural Basis of the Versatile DNA Recognition Ability of the Methyl-CpG Binding Domain of Methyl-CpG Binding Domain Protein 4*

The methyl-CpG binding domain (MBD) protein MBD4 participates in DNA repair as a glycosylase that excises mismatched thymine bases in CpG sites and also functions in transcriptional repression. Unlike other MBD proteins, MBD4 recognizes not only methylated CpG dinucleotides (5mCG/5mCG) but also T/G mismatched sites generated by spontaneous deamination of 5-methylcytosine (5mC/TG). The glycosylase activity of MBD4 is also implicated in active DNA demethylation initiated by the deaminase-catalyzed conversion of 5-methylcytosine to thymine. Here, we report the crystal structures of the MBD of MBD4 (MBD_{MBD4}) complexed with 5mCG/5mCG and 5mCG/TG. The crystal structures show that the DNA interface of MBD4 has flexible structural features and harbors an extensive water network that supports its dual base specificities. Combined with the results of biochemical analyses, the crystal structure of MBD4 bound to 5-hydroxymethylcytosine further demonstrates that MBD_{MBD4} is able to recognize a wide range of 5-methylcytosine modifications through the unique water network. The versatile base recognition ability of MBD_{MBD4} implies multifunctional roles for MBD4 in the regulation of dynamic DNA methylation patterns coupled with deamination and/or oxidation of 5-methylcytosine.

DNA methylation is the most prominent epigenetic modification in higher eukaryotic genomes (1, 2). In mammals, DNA methylation mainly occurs at the C5 position of symmetrically arranged cytosines in CpG dinucleotides, and plays essential roles in various cellular events such as gene repression, imprinting, X-chromosome inactivation, suppression of repetitive genomic elements, and carcinogenesis (3). Recent studies have shown that DNA methylation can be actively reversed and that its pattern is dynamically altered in mammalian cells (4–7). Although the underlying molecular mechanism is not fully understood, active DNA demethylation has been proposed to involve further oxidation or deamination of 5-methylcytosine (5mC) followed by base excision repair (6–13). Successive oxidation of 5mC to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC) is catalyzed by TET proteins and, has attracted much attention as a crucial process in DNA demethylation. Furthermore, demethylation pathways are thought to involve the spontaneous or enzymatic deamination of 5mC or hmC and subsequent base excision repair of the mismatched thymine or 5-hydroxymethyluracil (hmU) base (6, 9, 11–13). Therefore, precise interpretation and regulation of the modification status of 5mC are required for various epigenetic events in cells.

**MBD** (methyl-CpG binding domain) proteins are archetypal mediators of DNA methylation marks. They recognize methyl-CpG sites (5mCG/5mCG) through a conserved MBD and recruit transcriptional repressors or chromatin modifiers to these sites (14). One of the MBD family proteins, MBD4 contains a C-terminal DNA glycosylase domain in addition to an N-terminal MBD domain. MBD4 is involved in DNA mismatch repair as a T/G or U/G mismatch glycosylase and also in transcriptional repression via its recruitment of Sin3A and HDAC1 (15, 16).

---

*This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology and the Japan Science and Technology Agency (MEXT) (to M. S.) and by JST, PRESTO and **CREST, Global COE Program “International Center for Integrated Research and Technology Agency (MEXT) (to M. S.) and by JST, PRESTO and **CREST.

The methyl-CpG binding domain (MBD) protein MBD4 participates in DNA repair as a glycosylase that excises mismatched thymine bases in CpG sites and also functions in transcriptional repression. Unlike other MBD proteins, MBD4 recognizes not only methylated CpG dinucleotides (5mCG/5mCG) but also T/G mismatched sites generated by spontaneous deamination of 5-methylcytosine (5mC/TG). The glycosylase activity of MBD4 is also implicated in active DNA demethylation initiated by the deaminase-catalyzed conversion of 5-methylcytosine to thymine. Here, we report the crystal structures of the MBD of MBD4 (MBD_{MBD4}) complexed with 5mCG/5mCG and 5mCG/TG. The crystal structures show that the DNA interface of MBD4 has flexible structural features and harbors an extensive water network that supports its dual base specificities. Combined with the results of biochemical analyses, the crystal structure of MBD4 bound to 5-hydroxymethylcytosine further demonstrates that MBD_{MBD4} is able to recognize a wide range of 5-methylcytosine modifications through the unique water network. The versatile base recognition ability of MBD_{MBD4} implies multifunctional roles for MBD4 in the regulation of dynamic DNA methylation patterns coupled with deamination and/or oxidation of 5-methylcytosine.

DNA methylation is the most prominent epigenetic modification in higher eukaryotic genomes (1, 2). In mammals, DNA methylation mainly occurs at the C5 position of symmetrically arranged cytosines in CpG dinucleotides, and plays essential roles in various cellular events such as gene repression, imprinting, X-chromosome inactivation, suppression of repetitive genomic elements, and carcinogenesis (3). Recent studies have shown that DNA methylation can be actively reversed and that its pattern is dynamically altered in mammalian cells (4–7). Although the underlying molecular mechanism is not fully understood, active DNA demethylation has been proposed to involve further oxidation or deamination of 5-methylcytosine (5mC) followed by base excision repair (6–13). Successive oxidation of 5mC to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxylcytosine (caC) is catalyzed by TET proteins and, has attracted much attention as a crucial process in DNA demethylation. Furthermore, demethylation pathways are thought to involve the spontaneous or enzymatic deamination of 5mC or hmC and subsequent base excision repair of the mismatched thymine or 5-hydroxymethyluracil (hmU) base (6, 9, 11–13). Therefore, precise interpretation and regulation of the modification status of 5mC are required for various epigenetic events in cells.

**MBD** (methyl-CpG binding domain) proteins are archetypal mediators of DNA methylation marks. They recognize methyl-CpG sites (5mCG/5mCG) through a conserved MBD and recruit transcriptional repressors or chromatin modifiers to these sites (14). One of the MBD family proteins, MBD4 contains a C-terminal DNA glycosylase domain in addition to an N-terminal MBD domain. MBD4 is involved in DNA mismatch repair as a T/G or U/G mismatch glycosylase and also in transcriptional repression via its recruitment of Sin3A and HDAC1 (15, 16).
The glycosylase activity of MBD4 specifically excises a mismatched thymine or $h$mU base generated by the deamination of $^{5m}C$ or $^{5m}U$ in a CpG site; thus, MBD4 is thought to participate in both DNA repair in the context of CpG and DNA demethylation (14). The functional importance of MBD4 in maintaining genomic integrity has been demonstrated by an increased frequency of C to T transitions at CpG sites in MBD4$^{-/-}$ mice (17) and the finding that frequent MBD4 mutations in various human carcinomas are characterized by microsatellite instability (18). Moreover, MBD4 contributes to the stimuli-dependent active DNA demethylation of specific genomic loci together with thymine DNA glycosylase (TDG) (19).

Previous structural studies of MBD1, MBD2, and MeCP2 demonstrated how MBDs recognize only $^{5m}C$/$^{5m}G$ sites (20–22). However, in addition to the fully methylated CpG, the MBD domain of MBD4 binds to T/G mismatched base pairs that result from asymmetrical $^{5m}C$ deamination of $^{5m}C$/$^{5m}G$ dinucleotides (16). Recent structural and biochemical studies of the glycosylase domain of MBD4 suggest that the specificity of full-length MBD4 for $^{5m}C$/$^{5m}G$ is provided by MBD$_{BD4}$ (23–25). The glycosylase domain recognizes the mismatched thymine or $^{5m}U$ base but not the adjacent $^{5m}C$/$G$ base pair. Thus, the recognition of methylated DNA by MBD$_{BD4}$ appears to be indispensable for the multifunctional roles of MBD4 in the regulation and maintenance of DNA methylation patterns.

Here, we present the crystal structures of the MBDMBD4 (MBD$_{BD4}$) in complex with a DNA fragment containing the $^{5m}C$/$^{5m}G$ site or its deamination product, $^{5m}C$/$G$. The structures reveal the unique flexible DNA interface of MBDMBD4 accompanied by an extensive water network. Our structural and biochemical data demonstrate that, in addition to $^{5m}C$/$^{5m}G$ and $^{5m}C$/$G$, the DNA interface of MBDMBD4 is able to accommodate $h$mC and its further oxidation or deamination products. We also determined the crystal structure of MBDMBD4 bound to a methylated CpG site containing $^{5m}C$ ($^{5m}C$/$^{h}mC$) and found that the water network at the DNA interface of MBDMBD4 can be finely tuned to accommodate various modified pyrimidine rings. Our structural and biochemical studies indicate the molecular basis of the broad base recognition ability of MBDMBD4, which underlies DNA methylation and gene regulation involving MBD4.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—A DNA fragment encoding MBDMBD4 (residues 69–136) was amplified by PCR and cloned into the bacterial expression vector pGEX4T-3 (GE Healthcare Biosciences), which was engineered for the expression of recombinant proteins with an N-terminal tandem fusion tag of glutathione S-transferase (GST) and small ubiquitin-like modifier-1 (SUMO-1). The GST-SUMO-1-MBD$_{BD4}$ fusion was overexpressed in *Escherichia coli* strain BL21(DE3). Cells were grown at 37°C in Luria-Bertani (LB) medium containing 50 µg/ml of ampicillin, to an optical density of 0.5–0.6 at 660 nm, and then induced with 0.2 mM isopropyl-$b$-thiogalactoside for 15 h at 18°C. Cells were harvested by centrifugation, and lysed by sonication in 50 mM Tris-HCl, pH 8.0, buffer containing 300 mM NaCl, 1 mM dithiothreitol (DTT), 5% glycerol, 0.1% Triton X-100, and 1 mM phenylmethylsulfonyl fluoride. The clarified lysate was loaded onto glutathione-Septharose 4 Fast Flow beads (GE Healthcare). GST-SUMO-1-fused MBDMBD4 was eluted from the beads with elution buffer containing 10 mM glutathione. The tag-free MBDMBD4 was prepared by SENP2 protease treatment, and was further purified by sequential column chromatography steps using HiTrap Heparin HP and HiLoad 16/60 Superdex 75 columns (GE Healthcare). Purified protein in the final elution buffer containing 10 mM Hepes-NaOH, pH 7.4, 150 mM NaCl, and 2 mM DTT was concentrated using an Amicon Ultra 3,000 cut-off membrane concentrator (Millipore). To introduce selenomethionine, Leu-116 of MBD4 was substituted with methionine. The selenomethionine containing MBDMBD4 was expressed in modified M9 medium (26). Purification of the selenomethionine-labeled L116M mutant was performed following the same procedure as that for the native protein.

**Crystallization, Data Collection, and Structure Determination**—MBDMBD4 at a concentration of 200–800 µM was mixed with each DNA fragment at a 1:1 molar ratio. Crystals of MBDMBD4 were obtained by a vapor diffusion method at 20°C using PEG 10,000 or PEG 1500 as the precipitant. Details of crystallization conditions are listed in Table 1. MBDMBD4 bound to 14- and 11-bp oligomers containing $^{5m}C$/$G$ were crystallized in orthorhombic $C22_2$ and triclinic $P1$ forms, respectively. In the orthorhombic form, a complex of one protein and DNA is contained in an asymmetric unit, whereas the triclinic form comprises two protein molecules and one DNA oligomer. The complex of MBDMBD4 with 14-bp oligomer containing $^{5m}C$/$G$ or $^{5m}U$/$G$ was crystallized in a $C22_2$ form. All crystals were flash frozen at 100 K in cryoprotectant containing 20% ethylene glycol. X-ray diffraction data sets were collected at a wavelength of 1.0000 Å on beamlines BL-5A, BL-17A, NE3A, and NW12 at Photon Factory (Tsukuba, Japan) and beamline BL-38 at SPring8 (Harima, Japan), and were processed with the program HKL2000 (27). The phases of the selenomethionine derivative MBDMBD4 L116M complexed with the 14-bp $^{5m}C$/$G$ fragment were determined by the single wavelength anomalous dispersion method using the programs SOLVE and RESOLVE (28, 29). The initial model was built using the COOT program (30) and was refined in the native data using the PHENIX suite (31), thus yielding a crystallographic $R$ factor of 18.8% and a free $R$ factor of 22.4% to 2.0 Å. The triclinic form structure of the MBDMBD4$^{5m}C$/$G$ complex and the structures of the MBDMBD4$^{5m}C$/$G$ and MBDMBD4$^{5m}U$/$G$ complexes were solved by a molecular replacement method using the orthorhombic form structure of MBDMBD4$^{5m}C$/$G$ as the search model. The stereochemical quality of the final models was assessed using MolProbity (32). The sequence information of DNA fragments used for crystallization is summarized in Table 2. The crystallographic data, data collection statistics, and refinement statistics are summarized in Table 1. All structural figures were produced using PyMOL (43).

**DNA Binding Assays**—Isothermal titration calorimetry (ITC) measurements were performed on an iTC200 microcalorimeter (MicroCal, USA) at 25°C. The protein solution was diazoxed to the ITC measurement buffer of 25 mM Hepes-NaOH, pH 7.4, containing 100 mM NaCl and 0.1 mM Tris(2-carboxy-
ethyl)phosphine. Each annealed DNA duplex was dried and dissolved in ITC buffer. The DNA solution (10–20 μM) in a calorimetric cell was titrated with a 100–400 μM protein solution. Binding constants were calculated by fitting the data using the ITC data analysis module of Origin 7.0 (OriginLab). Competitive binding assays were also performed in the ITC buffer. The upper strand of the 14-bp 5mCG/5mCG DNA fragment was radioisotope labeled at the 5′ end with T4 polynucleotide kinase (TOYOBO, Japan) and [γ-32P]ATP, where the free reflections (5% of the total used) were held aside for Rfree throughout refinement.

**Rcryst** and **Rfree** are calculated by fitting the data using

\[ R = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \]

where the free reflections (5% of the total used) were held aside for Rfree throughout refinement.

**RESULTS**

**Dual Binding Specificity of MBDM4 for 5mCG/5mCG and 5mCG/TG Sites**—The DNA binding properties of mouse MBDM4 were examined quantitatively by ITC measurements using 14-bp double-stranded DNA oligomers containing a single CpG site in various modification or mismatch states (Tables 2 and 3). In agreement with previous reports (16), MBDM4 exhibited a 5-fold greater affinity for the 5mCG/5mCG site (K_{D} = 72.5 nM) over the 5mCG/TG site (K_{D} = 458 nM). The binding

**TABLE 1**

Crystallographic data and refinement statistics

| Crystal | 5mCG/TG (SeMet) | 5mCG/TG | 5mCG/5mCG | 5mCG/5mCG | 5mCG/TG | 5mCG/CT/TG |
|---------|----------------|---------|-----------|-----------|---------|------------|
| k | 12% PEG10,000 | 12% PEG10,000 | 7% PEG10,000 | 8% PEG1,500 | 2 mM Zn acetate | 0.1 mM Na cacodylate |
| NaCl (mM) | 0.1 | 0.1 Na acetate, pH 4.4 | 0.1 Na acetate, pH 4.4 | 0.1 Na acetate, pH 4.4 | 0.1 Na acetate, pH 3.9 |
| NaCl (mM) | 0.2 | 0.2 NaCl | 0.2 NaCl | 0.2 NaCl | 0.1 Na cacodylate, pH 5.4 |
| Wavelength (Å) | 0.97923 (peak) | 1.0000 | 1.0000 | 1.0000 | SPring8-BL38B1 |
| Space group | P1 | P1 | P1 | P1 |
| Unit cell parameters | a = 88.752; b = 97.588; c = 54.959; α = β = γ = 90° | a = 89.074; b = 94.989; c = 54.738; α = β = γ = 90° | a = 89.182; b = 93.829; c = 55.357; α = β = γ = 90° | a = 88.693; b = 97.758; c = 55.725; α = β = γ = 90° |
| Resolution range (Å) | 50–2.7 (2.8–2.7) | 50–2.0 (2.07–2.0) | 50–2.2 (2.28–2.2) | 50–2.19 (2.27–2.19) |
| Total observations | 6,051 (363) | 16,062 (1,591) | 12,048 (1,192) | 11,517 (701) | 7,225 (531) |
| Multiplicity | 6.6 (5.6) | 7.2 (7.3) | 7.2 (7.3) | 6.5 (4.1) | 2.9 (1.7) |
| Rmerge (%) | 0.097 | 0.031 | 0.078 | 0.052 | 0.058 |
| Completeness (%) | 98.6 (54.0) | 99.7 (99.9) | 89.6 (54.7) | 92.8 (67.4) | 92.4 (67.4) |
| I/σ(I) | 12.7 (5.2) | 18.2 (4.3) | 12.4 (4.5) | 14.7 (4.1) | 16.3 (5.4) |
| Refinement | Resolution range (Å) | 32.5–2.00 | 28.3–2.20 | 34.7–2.40 | 33.0–2.53 |
| R_grav (3σ) | 18.8 | 19.6 | 19.06 | 19.34 |
| Rfree (%) | 22.4 | 21.4 | 21.99 | 23.65 |
| Bond root mean square deviations | Bond length (Å) | 0.018 | 0.006 | 0.004 | 0.007 |
| Ramachandran plot (%) | Bond angle (°) | 2.071 | 1.193 | 1.008 | 1.161 |
| Favored (%) | 100 | 100 | 98.36 | 99.18 |
| Allowed (%) | 100 | 100 |

Notes:

a Numbers in parentheses are the values for the highest resolution shell of each data set.

b Rmerge = Σ |I(h)| − |Σ |I(h)|, where I(h) is the intensity of reflection h, Σ is the sum of all measured reflections and Σ is the sum of i measurements of reflection.

c Rwork (%) = Σ |I(h)| − |Σ |I(h)|/Σ |I(h)|, where the free reflections (5% of the total used) were held aside for Rfree throughout refinement.

Accession Codes—Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 3VXV, 3VXX, 3VYB, and 3VYQ.
MBD<sub>MBD4</sub> Exhibits Versatility in DNA Recognition

### TABLE 2
DNA sequences used in crystallization and binding assays

| Lower strand | Upper strand |
|--------------|--------------|
| 11 bp<sup>5mCG/TG</sup> | TCAC TG GATGTC | ACAT<sup>5mCG</sup> GATG |
| 14 bp<sup>5mCG/TG</sup> | GTGC TG GATGTCAC | GTGCTAC<sup>5mCG</sup> GAC |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTAC<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GACA |
| 5mCG<sup>5mCG</sup>/TG | GTA<sup>5mCG</sup> GATGTCAC | GTGCTACT<sup>5mCG</sup> GAC
interacting guanine (Fig. 3, C and D). Additionally, the main chain carbonyl group of Arg finger-2 forms a CHO hydrogen bond (3.8 Å) with the 5-methyl group of the 5mC base in the upper strand (Fig. 3D).

In MBDMeCP2, the positions of both Arg fingers are stabilized through interactions with the conserved acidic residues (Fig. 3, E and F) (21). Similarly, the orientation of Arg finger-1 of MBDMBD4 is defined by its intramolecular interaction with a conserved acidic residue, Asp-94. The side chain carboxyl group of Asp-94 forms salt bridges with the guanidino group of Arg finger-1, resulting in an arginine side chain conformation suitable for recognition of the 5mCG sequence (Fig. 3A). Asp-94 also forms a CHO hydrogen bond (3.9 Å) with the 5-methyl group of the 5mC base in the lower strand (Fig. 3C). In contrast, Arg finger-2 lacks such an intramolecular lock because Glu-137 in MeCP2 is replaced by Ser-110 in MBD4 (Fig. 2A). Nevertheless, the position of Arg finger-2 bound to 5mCG/5mCG shows good superimposition with that in the MBDMeCP2-5mCG/5mCG complex (Fig. 3, B and F).

The Water Network in the MBDMBD4-DNA Interface—The most significant structural difference between MBDMBD4 and other MBD proteins is the orientation of the conserved tyrosine residue, Tyr-96, located on the DNA binding surface (Fig. 3B). The corresponding tyrosine residues of MBDMeCP2, MBDMBD1, and MBDMBD2 are oriented toward the 5mC base in the lower strand through hydrophobic interactions with the aliphatic side chains of their surrounding residues (20–22). In the crystal structure of the MBDMBD4-DNA complex, the side chain of the corresponding residue, Tyr-123, recognizes the 5mC base via two water-mediated interactions (Fig. 3F) (21). Previous mutational analysis of MBD1 and MeCP2 suggested that the conserved Tyr residue is critical for DNA binding (20, 21). The side chain of Tyr-96 in MBDMBD4 is flipped out of the DNA interface and makes water-mediated interactions with the phosphate backbone of the lower DNA strand (Fig. 3B). The aromatic side chain is stabilized by a stacking interaction with the compact hydrophobic side chain of Val-80 (Fig. 4). Notably, despite the absence of the tyrosine hydroxyl group at the common position, the MBDMBD4-DNA interface retains the hydration water molecules involved in the recognition of the lower strand 5mC (Fig. 3, B and F). The water molecules, W1, W2, and W3, form van der Waals interactions with the 5-methyl group of the lower strand 5mC in a similar manner to that observed in the MBDMeCP2-DNA complex. Coordination of the three other water molecules (W4, W5, and W6 in the MBDMBD4-DNA complex) surrounding the upper strand 5mC base is also conserved in the MBDMBD4-DNA and MBDMeCP2-DNA complexes (Fig. 3, A and E). MBDMBD4 and MBDMeCP2 share the recognition scheme for the upper strand 5mC base involving a water molecule (W4 in the MBDMBD4-5mCG/5mCG complex or W4/11032 in the MeCP2-DNA complex) that bridges the N4 atom of the base with the carboxyl group of the conserved Asp residue. In the vacant space generated by the flipping of Tyr-96, the molecular water network is further extended at the MBDMBD4-DNA interface. For example, W1 forms a hydrogen-bonding network with other surrounding water molecules (Figs. 3B and 4), whereas its counterpart in the MBDMeCP2-DNA complex, W1’ in Fig. 3F, is fixed by hydrogen bonds with the Tyr-123 and...
**FIGURE 2.** Overall structure of MBD_{MBD4} complexed with DNA containing 5mCG/TG. A, sequence alignment of structurally known MBD domains; mouse MBD4 (mMBD4, 69–136 amino acids), human MeCP2 (hMeCP2, 88–167 amino acids), human MBD1 (hMBD1, 1–75 amino acids), and chicken MBD2 (gMBD2, 3–72 amino acids). Asterisks indicate the conserved residues among the MBD proteins. The hydrophobic core residues are highlighted in yellow. The residues highlighted in red are involved in recognition of methylated CpG base pairs. B and C, overall structures of MBD_{MBD4}–5mCG/TG complexes. B, C222 orthorhombic form; C, P1 triclinic form. The 2mFo/H2DFc electron density map for the DNA molecule contoured at 1.5σ is shown in blue. The lower DNA strand containing a mismatched thymine is presented in orange, and the complementary upper strand is yellow. A disulfide bond linking two symmetry-related MBDs is indicated by an orange circle in B. In the P1 form, specific (green) and nonspecific (blue) MBD_{MBD4}-DNA complexes are evident (C). D, specific binding of MBD_{MBD4} to 5mCG/TG. The triclinic P1 form structure of MBD_{MBD4} is shown as a green ribbon representation. Side chains of Arg-84, Arg-106, and Asp-94 are presented as stick models. The lower DNA strand containing a mismatched thymine is presented in orange, and the complementary upper strand is yellow. The 5mCG/TG base pairs are shown as stick models. The DNA sequences are indicated below in B and D.

**FIGURE 3.** Structural comparison of DNA binding surfaces of MBD_{MBD4} and MBD_{MeCP2}. A and B, recognition of the 5mC site by Arg finger-1 (A) or Arg finger-2 (B) of MBD_{MBD4}. Hydration water molecules are represented as small red spheres. DNA backbone structures of the upper and lower strands are depicted as yellow and orange tubes, respectively. Black dotted lines indicate hydrogen bonds (<3.2 Å). C, recognition of the methyl group of 5mC in the lower strand by MBD_{MBD4}. Orange dotted lines indicate nonbonded contacts with the 5-methyl group of 5mC (<4.2 Å). D, recognition of the 5-methyl group of the lower strand 5mC by MBD_{MeCP2}. The red spheres labeled as W1–6 in A–D represent the water molecules in the coordinate file of the MBD_{MBD4}–5mCG/5mCG complex structure (PDB code 3VXX); W1, Wat-202 in chain C; W2, Wat-312 in chain A; W3, Wat-319 in chain A; W4, Wat301 in chain A; W5, Wat-214 in chain B; W6, Wat-202 in chain B. E and F, recognition of the 5mC site by Arg finger-1 (E) or Arg finger-2 (F) of MBD_{MeCP2} (21). Black dotted lines indicate hydrogen bonds. W1′ and W4′ indicated in E and F correspond to the water molecules in the MeCP2-DNA complex structure (PDB code 3C2I), Wat-32 in chain B and Wat-183 in chain A, respectively.
Arg-133 residues of the protein. The hydrogen-bonding network within the DNA interface of MBDMBD4 is also maintained through water-mediated interactions between the phosphate groups of the DNA backbone and the side chains of Asp-94 and Lys-104 (Fig. 4). Thus, the DNA interface of MBDMBD4 contains more open space filled with ordered water molecules in comparison with other MBDs.

Recognition of 5mCG/TG by the Flexible DNA Binding Surface of MBDMBD4—The hydrogen bonding pattern of the T/G mismatched base pair in the MBDMBD4-5mCG/TG complex is identical to that observed in the crystal structure of a DNA oligomer with a T/G mismatch (PDB entry 113D) (34). The T/G mismatch still allows two hydrogen bonds to form between the bases, thus creating an overall shape similar to that in Watson-Crick base pairing. However, the mismatched thymine base is shifted 1–2 Å toward the major groove side of the DNA duplex (Fig. 5A). The base stacking interactions with neighboring pairs are unaffected by the mismatched pair (35), and the entire DNA binding mode common to MBDs is retained in the MBDMBD4-5mCG/TG complex.

The guanine base in the T/G mismatch is recognized by Arg finger-2 through a hydrogen-bonding pattern analogous to that observed in the crystal structure of a DNA oligomer with a T/G mismatch (PDB entry 113D) (34). The T/G mismatch still allows two hydrogen bonds to form between the bases, thus creating an overall shape similar to that in Watson-Crick base pairing. However, the mismatched thymine base is shifted 1–2 Å toward the major groove side of the DNA duplex (Fig. 5A). The base stacking interactions with neighboring pairs are unaffected by the mismatched pair (35), and the entire DNA binding mode common to MBDs is retained in the MBDMBD4-5mCG/TG complex.

In contrast to Arg finger-2 of MBDMBD4, Arg finger-2 of MBDMBD1 or MBDMCP2 is presumably incapable of recognizing the protruding mismatched base because its side chain is fixed by the interaction with conserved acidic residues (Fig. 3F) (20–22). Indeed, MBDMBD1 exhibited significantly weaker binding to 5mCG/TG compared with 5mC/5mC (Fig. 1; Table 3). Thus, the flexibility of Arg finger-2 provided by the lack of an intra-molecular lock appears to be indispensable for T/G mismatch recognition.

The Nonspecific DNA Binding Mode of MBDMBD4—The nonspecific DNA binding mode of MBDMBD4 is observed in the crystal structure of the triclinic form of the MBDMBD4-5mCG/TG complex (Fig. 2C). In the nonspecific complex, MBDMBD4 also binds to DNA via the major groove side. The phosphate backbone recognition scheme by the α1 helix and L1 loop is essentially identical to that in the specific complex (Fig. 6A).

In the nonspecific complex, the dynamic movement of Arg finger-2 is of great interest; this movement takes place in the vacant space generated by the flipping of Tyr-96. Arg finger-2, which is directed toward the target base in the specific complexes, adopts a completely different conformation to form a hydrogen bond with an atom of the phosphate backbone (Fig. 6B) and, thereby reinforcing DNA duplex binding. The unique flexibility of Arg finger-2 in MBDMBD4 presumably facilitates nonspecific DNA interaction, which implies a sliding mode prior to target recognition (Fig. 6B). In agreement with the structural observations, MBDMBD4 exhibited more highly significant binding to nonmodified CpG than MBDMBD1 in our electrophoretic mobility shift assay (data not shown).

The DNA Binding Surface of MBDMBD4 Tolerates Binding to Oxidation and Deamination Products of 5mC—The structural features of the protein-DNA interface suggest that MBDMBD4 has the ability to bind to modifications that are more bulky than the methyl group at the 5th position of cytosine. We therefore examined the binding of MBDMBD4 to a methylated CpG fragment containing hmC, hmU, fmC, or fmC (Fig. 7A). In a competitive EMSA, the nonlabeled 5mCG/hmCG, 5mCG/hmUG, and 5mCG/fmCG fragments competed with a 32P-labeled 5mCG/foCG duplex for binding to MBDMBD4. The affinity of MBDMBD4 for 5mCG/hmCG, 5mCG/hmUG, and 5mCG/fmCG was estimated to be 2- or 3-fold weaker than its affinity for 5mCG/5mCG based on the data from the competitive EMSA and ITC binding assays (Fig. 7, B and D, and Table 3). However, the 5mCG/fmCG and hmCG/hmCG fragments exhibited weaker binding to MBDMBD4 than the other modified nucleotides (Fig. 7B). In contrast, MBDMBD1 exhibited a tight specificity for 5mCG/5mCG (Fig. 7, C and E). The affinity of MBDMBD1 for 5mCG/hmCG (K_{d} = 1.04 μM) was more than 10-fold weaker than that for 5mCG/5mCG (K_{d} = 72.5 nm) (Table 3). Combined with the structural data, these findings suggest that MBDMBD4 is capable of binding to methylated CpG sequences that have undergone further asymmetric oxidative modification.

To achieve a better understanding of the structural basis of the versatile DNA binding ability of MBDMBD4, we determined its crystal structure at 2.4Å resolution when bound to a 5mCG/hmCG fragment (Table 1). Hydroxylation of the 5-methyl group of 5mC does not perturb either the canonical hydrogen bonding pattern in the C/G base pair or the overall DNA binding mode...
An unambiguous electron density for the hydroxyl group of hmC suggests a confined rotational movement of the 5-hydroxymethyl moiety against the pyrimidine ring (Fig. 8A); intriguingly, the hydroxyl group makes an intra-base hydrogen bond with the amino group at the 4th position in addition to a hydrogen bond with a water molecule at the DNA interface. The 5-hydroxymethyl moiety also donates CHO hydrogen bonds to the carbonyl of Asp-94 and the phosphate group of the DNA backbone, which show tetrahedral coordination around the methyl carbon at the 5th position (Fig. 8B). Thus, the positional preference of the hydroxyl group is ensured by the intra-base hydrogen bond and the tetrahedral configuration around the methyl carbon despite the close contacts with the neighboring base on the 5’ side. The flexible DNA interface of MBDMBD4 is likely to have enough space to accommodate the hmU or hmC base as well as hmC. In contrast, the
FIGURE 6. Structure of the nonspecific DNA complex of MBDMBD4. A, structure of the nonspecific complex observed in the P1 form. DNA molecule is presented as a surface model. The protein residues are shown as blue stick models. B, schematic model of recognition and scanning modes of MBDMBD4. Magnified views of the structure of Arg fingers in the nonspecific and the 5mCG/TG complexes are shown in right panels.

FIGURE 7. Broad binding specificity of MBDMBD4. A, schematic representation of cytosine oxidation and deamination. B and C, DNA binding specificities of MBDMBD4 (B) and MBDMBD1 (C) analyzed by competitive electrophoretic mobility shift assay. Representative autoradiographic images of competitive assays with MBDMBD4 and MBDMBD1 are presented. The left panel of each section shows the control experiment in the absence of competitor. D and E, the relative values for each complex are plotted against the amount of competitor DNA. Each data point represents an average of three independent experiments using MBDMBD4 or MBDMBD1. Neither the 5mCG/5mCG nor the 5mCG/hmUG fragment exhibited competitive effects on the MBDMBD4-5mCG/5mCG complex.
MBD<sub>MBD4</sub> Exhibits Versatility in DNA Recognition

relatively low affinity of MBD<sub>MBD4</sub> for 5mCG/caCG is presumably caused by electrostatic repulsion between the 5-carboxyl group of the base and the side chain carboxyl of Asp-94.

DISCUSSION

The crystal structures of MBD<sub>MBD4</sub> complexed with 5mCG/TG, 5mCG/5mCG, and 5mCG/hmCG provide new insight into the structural mechanism of the versatility of base recognition by MBD4. The broad base specificity of MBD<sub>MBD4</sub> is implicated in heterochromatin localization and enzymatic activity of MBD4 associated with methylated DNA regions. In contrast to MBD<sub>MBD1</sub>-MBD<sub>MBD4</sub> binds not only 5mCG/βmCG but also various modified pyrimidine rings including deamination and/or oxidation products of the 5mC base, such as 5mCG/TG, 5mCG/hmCG, 5mCG/hmUG, and 5mCG/hmCG, in a methylated CpG site. MBD<sub>MBD4</sub> shares an overall DNA recognition mode with other MBDs. The important role of the water molecules in target base recognition is highlighted by their conserved positions in the MBD<sub>MBD4</sub>-5mCG/hmCG and MBD<sub>MeCP2</sub>-5mCG/5mCG complexes (Fig. 3, A, B, E, and F) (21). However, local structural differences between MBD<sub>MBD4</sub> and other MBDs have a large impact on the DNA binding properties of MBD<sub>MBD4</sub>. In particular, the structural features unique to MBD<sub>MBD4</sub> around the conserved Tyr-96 and the Arg finger-2 provide plasticity in the DNA binding surface and allow versatile base recognition (Fig. 4). A comparison of the water structures around the lower strand target bases (5mC, mismatched T and hmC) highlighted the plasticity in the arrangement of the ordered water molecules in the DNA interface, in which the water-mediated hydrogen-bonding network of MBD<sub>MBD4</sub> is finely tuned to accommodate each of the modified bases.

Compared with the lower strand target base recognition, the interface with the upper strand 5mC more strictly maintains the structural features conserved in other MBDs including the conformation of Arg finger-1, which is fixed by the aspartic acid and hydration water structure (Fig. 3, A and E). This structural feature obviously indicates that the symmetric oxidative modification of both 5mC bases in the CpG sequence perturbs MBD<sub>MBD4</sub> binding. In fact, our DNA binding data combined with previously reported data demonstrate that neither MBD4 nor the other MBD proteins are capable of binding to the symmetrical hmCG/hmCG site (Fig. 7, B and C) (36, 37). MBD<sub>MBD4</sub> does not make contact with bases other than the CpG sequence and are able to bind to the symmetric 5mCG/5mCG site equally in both directions as observed in the flipping motion of MBD<sub>MBD1</sub> on its target DNA (38). In contrast, the tight recognition of 5mCG by the Arg finger-1 presumably prevents the flipping motion of MBD<sub>MBD4</sub> on asymmetric target sequences, such as 5mCG/TG, 5mCG/hmCG, and 5mCG/hmUG (Fig. 5D).

Despite the broad spectrum of MBD<sub>MBD4</sub> binding targets, full-length MBD4 exhibits glycosylase activity only toward mismatched thymine and hmU bases (Fig. 9) (24, 39). The oxidative products of 5mC, such as hmC, foC, and caC, are not susceptible to digestion by MBD4, whereas TDG excises foC and caC (7, 10). These findings indicate a partial functional redundancy and a possible functional difference between MBD4 and TDG (7, 10). The glycosylase domain itself exhibits the substrate specificity for T/G or hmU/G mismatched bases regardless of the methylation status of the adjacent C/G base pair (23, 25). Therefore, the DNA binding of MBD<sub>MBD4</sub> is presumably a prerequisite for the intrinsic glycosylase activity of MBD4 toward the mismatched bases generated in methylated CpG sites. Addition-
ally, isolated MBDMBD4 has been shown to inhibit the catalytic activity of the glycosylase domain toward a single 5mCG/TG site in vitro (40), suggesting that the DNA substrate is transferred from MBDMBD4 to the glycosylase domain only in full-length MBD4. The unidirectional binding of MBDMBD4 to the 5mCG/TG or 5mCG/hmC site could facilitate its synergetic action with the C-terminal glycosylase domain in DNA mismatch repair processes. It remains unclear whether the binding of MBDMBD4 to 5mC, hmC, or 6C targets the glycosylase domain to neighboring 5mCG/TG or 5mCG/hmC sites.

Intriguingly, the active DNA demethylation of the p15ink4b tumor suppressor gene triggered by the TGF-β/Smad signaling pathway is accompanied by the accumulation of hM bases, MBDMBD4, TDG, and downstream base excision repair proteins (19). The versatile base recognition ability of MBDMBD4 demonstrated in our study may contribute to the stimuli-dependent accumulation of MBDMBD4 at hydroxymethylated regions, which leads to erasure of DNA methylation marks. Further investigation of MBDMBD4 protein complexes colocalized to hM-rich regions will be crucial for fully understanding the functional roles of MBDMBD4 in DNA demethylation pathways. Furthermore, recent studies have indicated that the hM bases, fC, and 6C bases have long lifetimes during preimplantation development (41, 42); thus they may function as bona fide epigenetic marks antagonistic to 5mC bases in vivo. MBDMBD4 may recognize these bases independently of its glycosylase activity and act as a mediator via its multifunctional capabilities, although further investigation is necessary to fully understand the role of MBDMBD4 in the biology of oxidized cytosine bases.

Acknowledgments—We thank Drs. N. Matsugaki, N. Igarashi, and Y. Yamada and Prof. S. Wakatsuki for data collection at Photo Factory, Tsukuba. The synchrotron radiation experiments at the BL38B1 were performed with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2010B1059).

REFERENCES

1. Bird, A. (2002) DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6–21
2. Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nat. Rev. Genet. 3, 662–673
3. Robertson, K. D., and Wolffe, A. P. (2000) DNA methylation in health and disease. Nat. Rev. Genet. 1, 11–19
4. Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., and Salber, G. (2008) Cyclical DNA methylation of a transcriptional active promoter. Nature 452, 45–50
5. Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R. P., Benes, V., Gannon, F., and Reid, G. (2008) Transient cyclical methylation of promoter DNA. Nature 452, 112–115
6. Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Ciglano, A., Le Coz, M., Devarajan, K., Wessels, A., Sopranz, D., Abramowitz, L. K., Bar-tolomei, M. S., Rambow, F., Bassi, M. R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A. J., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., and Bellacosa, A. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 146, 67–79
7. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxycytosine and its excision by TDG in mammalian DNA. Science 333, 1303–1307
8. Wu, S. C., and Zhang, Y. (2010) Active DNA demethylation. Many roads lead to Rome. Nat. Rev. Mol. Cell Biol. 11, 607–620
9. Gao, J. U., Su, Y., Zhong, C., Ming, G. L., and Song, H. (2011) Hydroxyla-
tion of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell 145, 423–434
10. Maiti, A., and Drohat, A. C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxycytosine. Potential implications for active demethylation of CpG sites. J. Biol. Chem. 286, 35334–35338
11. Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., and Blau, H. M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 463, 1042–1047
12. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463, 1101–1105
13. Rai, K., Huggins, J. J., James, S. R., Karpf, A. R., Jones, D. A., and Cairns, B. R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell 135, 1201–1212
14. Bogdanović, O., and Veenstra, G. J. (2009) DNA methylation and methyl-CpG binding proteins. Developmental requirements and function. Chromosoma 118, 549–565
15. Kondo, E., Gu, Z., Horii, A., and Fukushima, S. (2005) The thymine DNA glycosylase MBDMBD4 represses transcription and is associated with methylated p16INK4a and hMLH1 genes. Mol. Cell. Biol. 25, 4388–4396
16. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999) The thymine glycosylase MBDMBD4 can bind to the product of deamination at methylated CpG sites. Nature 401, 301–304
17. Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., and Bird, A. (2002) Enhanced CpG mutability and tumorigenesis in MBDMBD4-deficient mice. Science 297, 403–405
18. Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., Ruggieri, B. A., Klein-Szanto, A. J., Testa, J. R., Neri, G., and Bellacosa, A. (1999) The DNA repair gene MBDMBD4 (MED1) is mutated in human cancers with microsatellite instability. Nat. Genet. 23, 266–268
19. Thillainadesan, G., Chittian, J. J., Isovic, M., Abblack, J. N., Myrmek, J. S., Tini, M., and Torchia, J. (2012) TGF-β-dependent active demethylation and expression of the p15ink4b tumor suppressor are impaired by the ZNF217/CoREST complex. Mol. Cell. 46, 636–649
20. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) Solution structure of the methyl-CpG binding domain of human MBDMBD4 complex with methylated DNA. Cell 105, 487–497
21. Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., and Walkinshaw, M. D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol. Cell 29, 525–531
22. Scardale, J. N., Webb, H. D., Ginder, G. D., and Williams, D. C. (2011) Solution structure and dynamic analysis of chicken MBDMBD4 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res. 39, 6741–6752
23. Manvilla, B. A., Maiti, A., Begley, M. C., Toth, E. A., and Drohat, A. C. (2012) Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. J. Mol. Biol. 420, 164–175
24. Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res., 40, 4841–4849
25. Moréra, S., Grin, L., Vigouroux, A., Couvé, S., Henriot, V., Sarapuraev, M., and Ishchenko, A. A. (2012) Biochemical and structural characterization of the glycosylase domain of MBDMBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. Nucleic Acids Res., 40, 9917–9926
26. Doublié, S. (1997) Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. 276, 523–530
27. Orwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. A 276, 307–326
28. Tvergell, T., C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. 55, 849–861
MBD<sub>MBD4</sub> Exhibits Versatility in DNA Recognition

29. Terwilliger, T. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972

30. Emsley, P., and Cowtan, K. (2004) COOT. Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132

31. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221

32. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Cα geometry. π, ψ and Cβ deviation. *Proteins* **50**, 437–450

33. Morrison, J. F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim. Biophys. Acta* **185**, 269–286

34. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., and Kennard, O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* **262**, 9962–9970

35. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229–237

36. Jin, S. G., Kadam, S., and Pfeifer, G. P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38**, e125

37. Valinluck, V., Tsai, H. H., Rogstad, D. K., Burdzy, A., Bird, A., and Sowers, L. C. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **32**, 4100–4108

38. Inomata, K., Ohki, I., Tochio, H., Fujiwara, K., Hiroaki, H., and Shirakawa, M. (2008) Kinetic and thermodynamic evidence for flipping of a methyl-CpG binding domain on methylated DNA. *Biochemistry* **47**, 3266–3271

39. Liu, P., Burdzy, A., and Sowers, L. C. (2003) Repair of the mutagenic DNA oxidation product, 5-formyluracil. *DNA Repair* **2**, 199–210

40. Aziz, M. A., Schupp, J. E., and Kinsella, T. J. (2009) Modulation of the activity of methyl binding domain protein 4 (MBD4/MED1) whereas processing iododeoxyuridine generated DNA mispairs. *Cancer Biol. Ther.* **8**, 1156–1163

41. Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011) Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21**, 1670–1676

42. Inoue, A., and Zhang, Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194

43. DeLano, W. L. (2010) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York