Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4

Peiying Wu, Emory University
Chen Qiu, Emory University
Anjum Sohail, Wayne State University
Xing Zhang, Emory University
Ashok S. Bhagwat, Wayne State University
Xiaodong Cheng, Emory University

Journal Title: Journal of Biological Chemistry
Volume: Volume 278, Number 7
Publisher: American Society for Biochemistry and Molecular Biology | 2003-02-14, Pages 5285-5291
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1074/jbc.M210884200
Permanent URL: http://pid.emory.edu/ark:/25593/fjmsk

Final published version: http://www.jbc.org/content/278/7/5285.long

Copyright information:
© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.
Accessed October 17, 2024 5:13 PM EDT
MISMATCH REPAIR IN METHYLATED DNA

STRUCTURE AND ACTIVITY OF THE MISMATCH-SPECIFIC THYMINE GLYCOSYLASE DOMAIN OF METHYL-CpG-BINDING PROTEIN MBD4*

Received for publication, October 23, 2002
Published, JBC Papers in Press, November 26, DOI 10.1074/jbc.M210884200

Peiying Wu‡, Chen Qiü‡§, Anjum Sohail¶, Xing Zhang‡, Ashok S. Bhagwat‡,
and Xiaodong Cheng‡‡

From the Departments of ‡Biochemistry and §Chemistry, Emory University, Atlanta, Georgia 30322 and
the iDepartment of Chemistry, Wayne State University, Detroit, Michigan 48202-3489

MBD4 is a member of the methyl-CpG-binding protein family. It contains two DNA binding domains, an aminoproximal methyl-CpG binding domain (MBD) and a C-terminal mismatch-specific glycosylase domain. Limited in vitro proteolysis of mouse MBD4 yields two stable fragments: a 139-residue fragment including the MBD, and the other 155-residue fragment including the glycosylase domain. Here we show that the latter fragment is active as a glycosylase on a DNA duplex containing a G:T mismatch within a CpG sequence context. The crystal structure confirmed the C-terminal domain is a member of the helix-hairpin-helix DNA glycosylase superfamily. The MBD4 active site is situated in a cleft that likely orients and binds DNA. Modeling studies suggest the mismatched target nucleotide will be flipped out into the active site where candidate residues for catalysis and substrate specificity are present.

MBD4 is a mammalian DNA glycosylase that excises thymines from G:T mispairs and contains both a methyl-CpG binding domain (MBD)† and a domain found in the Escherichia coli endonuclease III class of DNA glycosylases (1). It has preference for G:T mismatches within a CpG sequence context (1), and hence this enzyme can act upon G:T mismatches that result from the deamination of 5-methylcytosines (5mC) at CpG sites. The importance of this enzyme for mutation avoidance in mammals is confirmed by an increase in 5mC to T mutations in Mbd4−/− Big Blue mouse and by increased occurrence of colon carcinoma in Mbd4−/− ApcMin+− mice (2). Additionally, studies of MBD4 (also called MED1) using the yeast two-hybrid system have shown that it interacts with MLH1 (a protein implicated in mismatch repair) and suggest a role for this enzyme in maintaining genome stability (3). Consistent with this observation, it is found that MBD4 is mutated in 26–43% of human colorectal tumors that show microsatellite instability (4).

MBD4 is not the only DNA glycosylase reported to excise thymines from G:T mismatches. Another enzyme, named thymine-DNA glycosylase (TDG), was found earlier to have this ability (5). However, TDG is unrelated to MBD4 and belongs to the same structural superfamily as the uracil-excising enzymes UDG (6) and SMUG1 (7). MBD4 also differs from TDG in its substrate preference. Whereas the preferred substrates for TDG are N4-ethenocytosine or uracil paired with a G (8), MBD4 prefers thymine over N4-ethenocytosine (9). Recombinant MBD4 can also remove uracil, 5-fluorouracil, and 5mC at a low rate, particularly when these bases are opposite a guanine within CpG dinucleotides (1, 9, 10).

The MBD domain of MBD4 is similar to domains within other mammalian proteins, MeCP2, MBD1, MBD2, and MBD3 (reviewed in Refs. 11 and 12). The latter proteins are involved in suppressing transcription in regions of heavy CpG methylation, but no such role has been ascribed to MBD4. Whereas the NMR structures of the MBD domains from MBD1 (13, 14) and MeCP2 (15) have been elucidated, no structural information regarding the glycosylase domain of MBD4 is available. Here we present the crystal structure of the C-terminal glycosylase domain of MBD4, and we show that it belongs to the helix-hairpin-helix DNA glycosylase superfamily. The glycosylase domain alone is active on DNA duplex containing a G:T mismatch within a CpG sequence context.

EXPERIMENTAL PROCEDURES

Overexpression and Purification—The full-length mouse MBD4 was expressed as a His-tagged fusion protein in vector pET6H (16). The four fragments of MBD4 (Fig. 1A), amino acids 49–187 (MBD domain), 400–554 (glycosylase domain or Δ399), 49–554 (Δ48), and 429–554 (Δ348), were cloned into a modified pET28b (Novagen) vector, which contains an N-terminal tag of MGHHHHHHH and accepts an NdeI-EcoRI insert. The Δ428 fragment was also expressed as a GST fusion in pGEX2T vector (Amersham Biosciences). * E. coli strain BL21(DE3) carrying respective plasmid was grown in LB media supplemented with appropriate antibiotics at 37 °C to A600 = 0.6, shifted to 22 °C, and induced with 0.4 mM isopropyl-1-thio-β-D-galactopyranoside overnight at 22 °C, except that the full-length MBD4 was induced at 37 °C for 1 h. The full-length and Δ48 proteins were purified from cleared lysates using three successive chromatography steps as follows: a nickel chelate column, a HiTrap heparin column, and Superdex 200 (Amersham Biosciences). The proteins were stored in 20 mM potassium phosphate, pH 7.5, 1 mM EDTA, 0.1% 2-mercaptoethanol, and 0.2 M NaCl.

The MBD domain and the glycosylase domain were purified using nickel chelate, HiTrap Q, and Superdex 75 columns. The proteins were stored in a high salt buffer for crystallization (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, 5% glycerol, and 0.5 M NaCl) or in a low salt buffer for activity assay (20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, 50% glycerol, and 50 mM NaCl).

The GST-Δ428 was purified using glutathione-Sepharose 4B (Amer-
and unit cell dimensions were a/Å = 100.5, b/Å = 110.0, and c/Å = 110.0, with β = 90°. Pressed in a methionine auxotrophic strain B834(DE3) grown in M9 medium supplemented with 50 μg/ml ampicillin (full-length MBD4) or 50 μg/ml kanamycin (all MBD4 fragments) at 37 °C until the A₅₉₀ reached 0.6. The cultures were induced by adding isopropyl-1-thio-β-D-galactopyranoside to a final concentration of 0.5 mM. After incubation at 37 °C for 3 h, the cells were recovered by centrifugation, and the cell pellets were washed with a buffer containing 20 mM Tris-Cl, pH 7.6, and 0.1 mM EDTA. Cells were resuspended in 0.5 ml of extraction buffer (20 mM Tris-Cl, pH 7.8, 0.1 mM EDTA, 5 mM 2-mercaptoethanol) containing 1 mg/ml lysozyme and incubated on ice for 15 min. Finally, the cells were broken by sonication on ice, and cell-free lysate was recovered following centrifugation at 12,000 × g for 15 min at 4 °C. Protein concentration in the extracts was determined using the Bradford Reagent (Bio-Rad).

Preparation of the Labeled DNA Substrate—The oligonucleotides were gel-purified prior to their use. The T oligo was labeled at the 5′ end using T4 polynucleotide kinase (New England Biolabs) in the presence of [γ-³²P]ATP (specific activity 600 Ci/mmol, PerkinElmer Life Sciences). The reaction was terminated by heating it to 65 °C for 20 min. The labeled T oligo was mixed with 3-fold molar excess of the unlabeled G-oligo in the STE buffer (150 mM NaCl, 10 mM Tris-Cl, pH 8.0, and 1 mM EDTA). The mixture was heated to 95 °C for 3 min and then slowly cooled to room temperature over a period of 2–3 h to promote duplex formation. The unincorporated [γ-³²P]ATP was removed from the labeled duplex by passage through a G-50 micro column (Amersham Biosciences).

DNA Glycosylase Assay—Twenty ng labeled duplex was equilibrated with nicking buffer (10 mM Tris-Cl, pH 8.0, 5 mM EDTA, 1 mM DTT, and 0.1 mg/ml bovine serum albumin), and the reactions were initiated by adding 100 ng of purified MBD4 variants (Fig. 5A) or 2 μg of cell-free extract (Fig. 5B). Following incubation at 37 °C for 1 h, the reaction was stopped by heating to 95 °C for 7 min in the presence of 0.1 mM NaOH. Subsequently, 8 μl of gel loading dye (80% formamide, 10 mM EDTA, 1 mg/ml each of xylene cyanol and bromophenol blue) was added to the samples which were then heated to 95 °C and electrophoresed in 20% sequencing gel. The gel was exposed to a PhosphorImager screen (Amersham Biosciences), and the reaction products were quantified using ImageQuant software.

RESULTS

MBD4 Glycosylase Domain Structure—The overall structure of MBD4 glycosylase domain consists of 11 helices (αA to αK) (Fig. 1B) forming a single domain with a cleft in the middle (Fig. 2). Structural comparison with other DNA glycosylases (Fig. 3, A and B) reveals that the MBD4 glycosylase domain belongs to the helix-hairpin-helix (HhH) DNA glycosylase superfamily (25), named after a conserved structural motif αH-hairpin loop-α1 (shown in red in Fig. 2A). The six helices before

| Crystal | Native | SeMet |
|---------|--------|-------|
| Wavelength (Å) | 1.54 | 0.97890 | 0.97865 |
| Resolution (Å) | 20-2.1/2.1-2.1 | 20-2.1/2.1-2.1 | 20-2.1/2.1-2.1 |
| Completeness (%) | 99.2/99.8 | 80.42/89.1 | 83.42/38.6 |
| Rmerge | 0.048/0.098 | 0.028/0.081 | 0.028/0.081 |
| I/σ(I) | 42.7 | 51.2 | 50.7 |
| Unique reflections | 12,345/585 | 9,665/230 | 9,944/228 |
| Total reflections | 107,778 | 58,654 | 55,029 |

Ramachandran plot
- Most favored regions (%) | 93.4 | 6.6 |
- Additional allowed regions (%) | 93.4 |
- Bond mean square deviation from ideality | 0.01 |
- Bond angles (°) | 1.5 |
- Dihedrals (°) | 23.4 |
- Improper (°) | 1.4 |
- G-factor | 0.21 |
- Covalent | 0.45 |
- Overall | 0.31 |
- Average thermal factor (Å²) | 19.7 |
- Side chain | 20.9 |
the HhH motif (αC to αG in green) are highly conserved structural elements among family members, forming the bottom of the cleft in the orientation shown (Fig. 2A). Among the known HhH enzymes, MBD4 has the shortest sequence following the HhH motif (Fig. 1B). The C-terminal helices αD and αK, the short N-terminal helix αA, and its 12-residue preceding loop, the HhH motif, come together to form a hydrophobic core (Fig. 2C), forming the top of the cleft.

Model of the MBD4-DNA Complex—The high degree of structural similarity among HhH glycosylases allowed us to create a model of the MBD4 glycosylase domain bound to DNA. By using the coordinates of the AlkA-DNA (26) or hOGG1-DNA (27) complexes, we superimposed the protein components, and then the DNA was positioned over the surface of MBD4 with the cleft. Previous modeling studies of other HhH glycosylases MutY and EndoIII suggested that they bind to DNA in a manner similar to that of AlkA (26). Our modeling suggests that the MBD4 glycosylase domain also binds DNA similarly to AlkA and hOGG1, which bind DNA via the minor groove and AlkA and hOGG1, which bind DNA via the minor groove and major groove of the DNA helix (26, 27). The residues that contact the DNA backbone in the hOGG1 and AlkA structures occupy similar positions in the free MBD4 structure (Fig. 3C), and the MBD4 glycosylase domain could contact bent DNA without major physical distortion of the protein component (Fig. 3D). Two important DNA-binding loops are superimposed, the loop between helices αB and αC and the Gly-rich hairpin loop of HhH motif (Fig. 3C). Arg442 of MBD4, as well as Arg47 of MIG (28), is in the same position as Leu125 of AlkA (or Asn449 of hOGG1) that fills the space in the DNA duplex vacated by the flipped nucleotide. Thr443 of MBD4 is in the same position as Asn550 of hOGG1 that forms main chain contacts to the phosphate groups 3' to the flipped nucleotide. Ser544 of MBD4 is in the position of Asn551 of hOGG1 that forms hydrogen bonds with the base 5' immediate to the flipped nucleotide. It seems that the loop between helices αB and αC contains residues (Arg442-Thr443-Ser444) important for DNA binding and base flipping.

Mechanisms for Recognition of Flipped Bases and Catalysis—First, where is the active site? In analogy to the AlkA-DNA (26) and hOGG1-DNA (27) complexes, the MBD4 cleft defines the location of the active site (Fig. 4A). The target nucleotide is likely to be flipped out from the DNA helix into the active-site cleft of the enzyme, in a similar manner to AlkA or hOGG1. The structural superimposition of the HhH glycosylase-DNA complexes and the unbound MBD4 reveals several informative features. Interestingly, the flipped base can only be docked into the active site by stacking the base between the side chains of Leu440 of αB and Lys536 of αK (Fig. 4B). Although these residues are not conserved in HhH glycosylases, similar stacking appears to be conserved: in hOGG1 8-oxoguanine is between Cys253 and Phe319 (27) and in MutY adenine soaked into the crystal lies between Leu440 of αB and Met185 (29). Leu440 of MBD4 corresponds to Leu149 of MutY (Figs. 1B and 4E), whereas MutY Met185 corresponds to Phe319 of hOGG1.

A second question is where the key catalytic residues are located. Asp314, the last residue prior to helix αK (Fig. 1B), is in...
a position structurally equivalent to the catalytically important Asp$^{328}$ of AlkA (26), Asp$^{308}$ of hOGG1 (27), Asp$^{318}$ of MutY (29), and Asp$^{138}$ of EndoIII (30). Two mechanisms have been suggested for the function of this structurally conserved aspartic acid in HhH glycosylases: (i) it activates a catalytic nucleophile, which is either a water (29) or the $\epsilon$-amino group of a lysine (27), for the attack on the deoxyribose C1’ carbon atom of the target nucleotide; or (ii) it directly assists base removal by protonating the leaving group of the substrate sugar (26). In the docking model of MBD4-thymine (Fig. 4C), the C1’ position of a modeled substrate is in direct contact (~3.0 Å) with the carboxylate of Asp$^{534}$, which would favor the second (protonation) mechanism.

A third question regarding the MBD4 action is how it distinguishes an A:T pair from a G:T. Although it is possible that the protein distinguishes G:T from an A:T because of their differing geometries, it is also possible that it may make specific contacts with the guanine in a manner similar to E. coli MUG (31) or hOGG1 (27); Arg$^{486}$ of MBD4 is in the same position as Arg$^{234}$ of hOGG1 that forms hydrogen bonds in the minor groove side with the G on the opposite strand of the flipped nucleotide. A detailed answer to this question must await the availability of a MBD4-DNA co-crystal structure.

Thymine and Uracil—How does the flipped base specifically bind in the active site? In MutY the adenine soaked into the crystal are recognized by Glu$^{422}$ and Gln$^{393}$ (29) (Fig. 4D). Structural superimposition between MutY and MBD4 (Fig. 3A) indicates the side chains of Gln$^{423}$ and Tyr$^{514}$ of MBD4 are in the vicinity of the adenine-specific interacting side chains of MutY (Fig. 4E).

In MBD4, the two polar residues (Gln$^{223}$ of $\alpha$A and Tyr$^{514}$ of $\alpha$I) and three hydrophobic residues (Val$^{422}$ prior to $\alpha$A, Gly$^{445}$, and Ile$^{449}$ of $\alpha$C) line in the cleft next to the catalytic Asp$^{334}$ (Fig. 4A). We suggest that these amino acids are the major determinants of specificity after docking the flipped thymine into the binding pocket. In the absence of the target nucleotide, the active site is occupied by ordered water molecules (Fig. 2D), which lie almost in a plane and directly interact with Tyr$^{514}$, Gln$^{423}$, and Val$^{422}$ (Fig. 4C). We docked a thymine with its Watson-Crick pairing edge (O-2, N-3, and O-4) occupying three water sites (Fig. 4C). The OH group of Tyr$^{514}$ can make one hydrogen bond with the O-2 atom, the side chain carbonyl oxygen atom of Gly$^{445}$ and Ile$^{449}$ form a surface hydrophobic patch near the end of the cleft, in a perfect position to accommodate the methyl group of thymine. Of all contacts made to the thymine base (Fig. 4E), the hydrophobic-methyl interaction will be absent for a uracil base.

Interestingly, Glu, Gln, or Tyr are often found in the active site of the HhH glycosylases. A glutamate is found in MIG (Glu$^{42}$, Ref. 28) and TAG (Glu$^{32}$, Ref. 32) in the equivalent position as Glu$^{37}$ of MutY; the corresponding main chain position in MBD4 is Thr$^{437}$ (i.e. I-equivalent helix, 90° on the vertical axis). Structural superimposition between MutY and MBD4 (Fig. 3A) indicates the side chains of Gln$^{423}$ and Tyr$^{514}$ of MBD4 are in the vicinity of the adenine-specific interacting side chains of MutY (Fig. 4E).
helix of HhH motif), and in TAG (32) (Tyr16 of an N-terminal helix). A glutamine is common to MutY (Gln182) and hOGG1 (Gln315) in recognizing their substrate base, adenine and 8-oxoguanine, respectively; both Gln182 of MutY and Gln315 of hOGG1 are located in a C-terminal helix outside of the structurally homologous regions among the HhH glycosylases shown in Fig. 1B. Although MBD4 does not have an equivalent C-terminal helix, the N-terminal and C-terminal regions of all structurally characterized HhH glycosylases are folded together, above the cleft as shown in Fig. 3; and in the case of MBD4, Gln423 is from the N-terminal helix and its side chain occupies a similar position as that of Gln from the C-terminal helix.

**DNA Glycosylase Activity of MBD4 N-terminal Truncations**—Among the known HhH enzymes, MBD4 has the longest N-terminal sequence before the glycosylase domain (for examples, see Fig. 1B). Zhu et al. (10) analyzed a series of N-terminal deletion mutants of human MBD4, and the results are consistent with our glycosylase domain structure presented here. In that study, N-terminal deletions of up to 65% of the total length of MBD4 retain the DNA glycosylase activity. The smallest fragment that retained activity, N433 (10), is very similar in size to our glycosylase domain determined by proteolysis (see Fig. 1B). We used a DNA duplex containing a G:T within a CpG sequence context as the substrate to test the glycosylase activities of purified full-length MBD4 and several of its deletion derivatives. The T-containing strand was radiolabeled, and the excision of this base was monitored by gel electrophoresis. Typical results are presented in Fig. 5A and show that in addition to the full-length MBD4, the 439 mutant used for crystallography is an active thymine DNA glycosylase. A construct missing the first 48 amino acids of the full-length protein (48) has less activity, but the construct containing only the MBD segment of the protein (amino acids 48–187) has no detectable activity (Fig. 5A). We also measured the activity of all MBD4 constructs in crude cell extract by expressing the proteins in a strain lacking the endogenous uracil glycosylase (ung) to minimize background. Full-length MBD4, 48, and 399 all have detectable activity in this assay (Fig. 5B).

Petronzelli et al. (33) have reported that a deletion of the first 454 amino acids of the human MBD4 still retained its enzymatic activity. The murine MBD4 equivalent to this deletion would be missing 428 N-terminal residues (Fig. 1B), which include helix αA and its preceding loop that provides part of the hydrophobic core above the cleft (Trp412, Pro414, Pro415, Pro418, and Phe419; Fig. 2C) and Val422 and Gln423 that are proposed to contact the target thymine (Fig. 4C). Thus the results reported by Petronzelli et al. (33) are not compatible with the crystal structure and are surprising.

To resolve these discrepancies, we attempted to duplicate the result of Petronzelli et al. (33) by making the equivalent murine MBD4 truncation (428; Fig. 1) and fusing it to a six histidine tag or GST tag. We were unable to detect any expression of the

---

**Fig. 3. Structural comparison of HhH glycosylases.** A, superimposition of MBD4 glycosylase domain (colored according to Fig. 1B) and MutY (gray; Protein Data Bank code 1MUD). MutY (29), as well as EndoIII (30) and MIG (28), contains a C-terminal [4Fe-4S] cluster (shown in space-filling model). B, superimposition of MBD4 glycosylase domain and AlkA (gray; Protein Data Bank code 1DHZ). AlkA (26) and hOGG1 (27) contain an additional N-terminal β-sheet domain, and a 13- or 15-residue insertion between αC and αD (see Fig. 1B). C, superimposition of two DNA binding loops between MBD4 (red and green) and AlkA (gray). The Gly-rich hairpin loop of HhH motif is indicated by conserved Gly510 and Gly512 of MBD4. The minor groove wedge (Leu125 in AlkA), which assists in base flipping, superimposed on Arg162 of MBD4 in the loop between helices αB and αC (see Fig. 1B). D, based on the superimposition shown in C, the MBD4 glycosylase domain is docked to DNA from the minor groove side. E, the MBD domain of MBD4 has not yet been structurally characterized; however, the NMR solution structure of the MBD domain of MBD1 was shown to bind DNA from the major groove side (13).
His-tagged Δ428 protein, either by Coomassie staining or anti-His tag antibody (data not shown), whereas all other MBD4 fragments were expressed and soluble under the same conditions. Not surprisingly, no glycosylase activity was detected in the extract of Δ428 construct using the ung− strain (Fig. 5B). The GST-tagged Δ428 was expressed to high level, but most of the protein was insoluble (data not shown). However, we did manage to partially purify some GST-Δ428 fusion protein using a glutathione affinity column and a HiTrap Q column. The protein was heavily associated with Hsp60 (data not shown), an indication that the protein may not be folded properly. When the GST-Δ428 protein was tested for glycosylase activity, none was detected (Fig. 5A). The observation that Δ428 mutant does not fold properly is consistent with the important structural roles of the missing residues. In addition, although sequence similarity of MBD4 to other glycosylases starts at helix αB, MutY, MIG, EndoIII, and TAG all have N-terminal extensions similar in size to Δ399 of MBD4 (Fig. 1B). We do not know the origin of the discrepancy between our data and that of Petronzelli et al. (33), as the sequences of human Δ454 and mouse Δ428 deletions are almost 100% identical except 4 residues (see Fig. 1B). One possibility is that the pET28b vector (Novagen) used for the human Δ454 construct would add at least 10 additional residues besides the 6 histidines at the N terminus. These residues may fortuitously substitute the natural MBD4 residues and allow folding and enzymatic activity.

The activity of the Δ399 deletion was easiest to detect in the extracts, whereas the full-length MBD4 and the Δ48 construct displayed relatively poor activity (Fig. 5B). The lower activity of the full-length MBD4 in cell-free extracts was surprising but reproducible. It is noted that the MBD domain of MBD4 binds DNA with G:T mismatches (1), and it is possible that both the MBD and the glycosylase domains compete for the DNA substrate. Regardless, it is clear from these data that the Δ399 construct of the murine MBD4, which has almost the same N-terminal extension as the MutY, MIG, EndoIII, and TAG, is a stable protein fragment with substantial glycosylase activity.

**DISCUSSION**

We have described the crystallographic structure of the glycosylase domain of the methyl-CpG-binding protein MBD4. The structure reveals that the MBD4 glycosylase domain belongs to the HhH DNA glycosylase superfamily. Modeling studies suggest that MBD4 glycosylase domain, similar to that of AlkA and hOGG1 HhH glycosylases, binds DNA from the minor groove side (Fig. 3D).

Unlike other HhH glycosylases, MBD4 contains an additional DNA binding domain, the MBD, near its N terminus. An NMR solution structure of the MBD domain from human MBD1, in complex with methylated DNA, revealed that the MBD domain contacts both methyl groups of methyl-CpG site via the major groove of B-form DNA (13) (Fig. 3E). This is consistent with the observation that of the DNA sequence tested, only the fully methylated CpG or the methylated mismatch 5mCpG/TpG (both contain two methyl groups in the major groove) is bound by the MBD of MBD4 (1). Because all structurally characterized HhH glycosylases in complex with DNA appear to bind DNA exclusively via the minor groove, it is attractive to think that the MBD and the glycosylase domains of MBD4 would come together at 5mCpG/TpG mismatches to engulf DNA from opposite directions (28). However, because the MBD domain does not bind DNA (13), whereas all HhH glycosylases appear to significantly bend DNA and flip the target, it is not clear how DNA would be bent when both domains bind together. Alternatively, perhaps the two domains separated by ~200 residues bind DNA at adjacent but non-
overlapping sites. The function of the MBD domain in MBTD4 may be to target the glycosylase activity to regions of heavily methylated DNA as methyl-CpG dinucleotides tend to occur in clusters (reviewed in Ref. 34), so the tethered glycosylase domain could sample nearby sites for G:T mismatches. This would raise the local concentration of glycosylase activity in regions where methylated mismatch 5mCpG/TpG is most likely to occur.

The active-site cleft of the glycosylase domain suggests a base flipping mechanism for accessing the damaged or mismatched base (reviewed in Ref. 35), the mismatched base should be swung completely out of the DNA helix by torsional rotation of its flanking sugar-phosphate backbones so as to occupy the active-site cleft of MBTD4. The structure also reveals candidate residues for catalysis (Asp334), for thymine (or uracil)-specific recognition hydrogen bonding (Tyr114, Gln223, and Val122), for the methyl group of thymine (Ile449 and Gly456), and for the stacking stabilization of the flipped base (Leu440 and Lys536). With this information, our structure provides useful starting points for more detailed studies of this interesting enzyme.

Acknowledgments—We thank Dr. Adrian Bird (University of Edinburgh) for providing the constructs to overexpress mouse MBTD4; Drs. Anand Saxena and Dieter Schneider (Brookhaven National Laboratory) and John R. Horton (Emory University) for help with x-ray data collection at beamlines X12-C and X26-C in the National Synchrotron Light Source; and Drs. Robert Blumenthal (Medical College of Ohio) and Paul Wade (Emory University) for critical comments on the manuscript.

REFERENCES

1. Hendrich, B., Hardeland, U., Ng, H.-H., Jiricny, J., and Bird, A. (1999) Nature 398, 301–304
2. Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. B., and Bird, A. (2002) Science 297, 483–485
3. Bellacosa, A., Cicchitelli, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M., and Neri, G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 3965–3974
4. Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Pernescia, A., Salovaara, R., Mascarello, V., Genuardi, M., Paravattu-Petsotsa, M., Bassi, D. E., Ruggieri, B. A., Klein-Staute, A. J., Testa, J. R., Neri, G., and Bellacosa, A. (1999) Nat. Genet. 23, 266–268
5. Neddermann, P., Gallinarini, P., Lattieri, T., Schimid, D., Trung, O., Haasen, J. J., Weibauer, K., and Jiricny, J. (1996) J. Biol. Chem. 271, 12767–12774
6. Krokan, H. E., Nisen, H., Skorpen, F., Otterlei, M., and Slupphaug, G. (2000) FEBS Lett. 476, 73–77
7. Nisen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdiene, G. L., and Lindahl, T. (2001) EMBO J. 20, 4278–4286
8. Saparbayev, M., and Laval, J. (1996) Proc. Natl. Acad. Sci. U. S. A. 95, 8508–8513
9. Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., Yeung, A. T., Matsumoto, Y., and Bellacosa, A. (2000) J. Biol. Chem. 275, 23242–23249
10. Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Pernescia, A., Salovaara, R., Mascarello, V., Genuardi, M., Paravattu-Petsotsa, M., Bassi, D. E., Ruggieri, B. A., Klein-Staute, A. J., Testa, J. R., Neri, G., and Bellacosa, A. (1999) Nat. Genet. 23, 266–268
11. Wakefield, R. I. D., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P., and Barlow, P. N. (1999) J. Mol. Biol. 291, 1055–1065
12. Wade, P. A. (2001) Oncogene 20, 3166–3173
13. Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M. (2002) J. Mol. Biol. 315, 487–497
14. Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999) EMBO J. 18, 6653–6661
15. Wakefield, R. I. D., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P., and Barlow, P. N. (1999) J. Mol. Biol. 291, 1055–1065
16. Hendrich, B., and Bird, A. (1998) Mol. Cell. Biol. 18, 6538–6547
17. Deng, A. Yoder, J. A., Zhang, X., Zhou, L., Bestor, T. H., and Cheng, X. (2001) Nucleic Acids Res. 29, 439–446
18. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1999) EMBO J. 18, 1665–1672
19. Oshino, M., and Minor, W. (1997) Methods Enzymol. 276, 307–328
20. Terwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 849–861
21. Terwilliger, T. C. (2000) Acta Crystallogr. Sect. D Biol. Crystallogr. 56, 863–872
22. Jones, T. A., Kjeldgard, M. (1997) Methods Enzymol. 277, 173–208
23. Brungger, A. T. (1992) X-PLOR: A System for X-ray Crystallography and NMR
24. version 3.1, Yale University, New Haven, CT
25. Beletskii, A., Grigoriev, A., Joyce, S., and Bhagwat, A. S. (2000) J. Mol. Biol. 300, 1057–1065
26. Nash, H. M., Bruner, S. D., Scherer, O. D., Kawata, T., Addona, T. A., Spooner, E., Lane, W. S., and Verdiene, G. L. (1996) Curr. Biol. 6, 986–989
27. Hollis, T., Ichikawa, Y., and Ellenberger, T. (2000) EMBO J. 13, 758–766
28. Bruner, S. D., Norman, D. P. G., and Verdiene, G. L. (2000) Nature 403, 899–903
29. Mol, C. D., Arvai, A. S., Belegrery, T. J., Cunningham, R. P., and Tainer, J. A. (2002) J. Mol. Biol. 315, 375–384
30. Guan, Y., Banda, M. A., Ayala, O. A., Parkh, S. S., Mol, C. D., Miller, J. H., Lloyd, S., and Tainer, J. A. (1998) Nat. Struct. Biol. 5, 1058–1064
31. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1999) EMBO J. 18, 4188–4193
32. Barrett, T. E., Scherer, O. D., Sawa, R., Brown, T., Jiricny, J., Verdiene, G. L., and Pearl, L. H. (1999) EMBO J. 18, 6599–6609
33. Dreshat, A. C., Kwon, K., Krosky, D. J., and Stivers, J. T. (2002) Nat. Struct. Biol. 9, 659–664
34. Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Genuardi, M., Karbowski, M., Yeung, A. T., Matsumoto, Y., and Bellacosa, A. (2000) J. Cell. Physiol. 185, 473–480
35. Jones, P. A., and Takai, D. (2001) Science 293, 1068–1070
36. Roberts, J. R., and Cheng, X. (1998) Annu. Rev. Biochem. 67, 181–198
37. Carson, M. (1997) Methods Enzymol. 277, 493–505
38. Nicholls, A., Sharp, K. A., and Honig, B. (1994) Protein Struct. Funct. Genet. 11, 281–296
Mismatch Repair in Methylated DNA: STRUCTURE AND ACTIVITY OF THE MISMATCH-SPECIFIC THYMINE GLYCOSYLASE DOMAIN OF METHYL-CpG-BINDING PROTEIN MBD4
Peiying Wu, Chen Qiu, Anjum Sohail, Xing Zhang, Ashok S. Bhagwat and Xiaodong Cheng

J. Biol. Chem. 2003, 278:5285-5291.
doi: 10.1074/jbc.M210884200 originally published online November 26, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M210884200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 36 references, 13 of which can be accessed free at http://www.jbc.org/content/278/7/5285.full.html#ref-list-1