Published in final edited form as:

J Mol Biol. 2012 July 13; 420(3): 164–175. doi:10.1016/j.jmb.2012.04.028.

Crystal Structure of Human Methyl Binding Domain IV Glycosylase Bound to Abasic DNA

Brittney A. Manvilla, Atanu Maiti, Matthew C. Begley, Eric A. Toth, and Alexander C. Drohat
Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine,
108 N Greene St., Baltimore, MD 21201

Abstract

The mammalian repair protein MBD4 (methyl-CpG-binding domain IV) excises thymine from mutagenic G-T mispairs generated by deamination of 5-methylcytosine (mC), and downstream base excision repair (BER) proteins restore a G-C pair. MBD4 is also implicated in active DNA demethylation, by initiating BER of G-T mispairs generated by a deaminase enzyme. The question of how mismatch glycosylases attain specificity for excising thymine from G-T but not A-T pairs remains largely unresolved. Here, we report a crystal structure of the glycosylase domain of human MBD4 (residues 427-580) bound to DNA containing an abasic nucleotide paired with guanine, providing a glimpse of the enzyme-product complex. The mismatched guanine remains intrahelical, nestled into a recognition pocket. MBD4 provides selective interactions with the mismatched guanine (N1H, N2H₂) that are not compatible with adenine, which likely confer mismatch specificity. The structure reveals no interactions that would be expected to provide the MBD4 glycosylase domain with specificity for acting at CpG sites. Accordingly, we find modest 1.5- to 2.7-fold reductions in G-T activity upon altering the CpG context. In contrast, 37- to 580-fold effects were observed previously for TDG (thymine DNA glycosylase). These findings suggest that specificity of MBD4 for acting at CpG sites depends largely on its methyl-CpG-binding domain, which binds preferably to G-T mispairs in a methylated CpG site. MBD4 glycosylase cannot excise 5-formylcytosine (fC) or 5-carboxylcytosine (caC), intermediates in a Tet-initiated DNA demethylation pathway. Our structure suggests that MBD4 does not provide the electrostatic interactions needed to excise these oxidized forms of mC.

Keywords

DNA glycosylase; G-T mismatch; Deamination; 5-methylcytosine; methyl-CpG-binding domain; DNA methylation

© 2012 Elsevier Ltd. All rights reserved.

To whom correspondence should be addressed: Alex Drohat Department of Biochemistry and Molecular Biology University of Maryland School of Medicine 108 N Greene St. Baltimore, MD 21201 Telephone: 410-706-8118 Fax: 410-706-8297
adrohat@som.umaryland.edu.

Accession number Coordinates and structure factors are deposited in the Protein Data Bank (<http://www.rcsb.org/>) with accession number 4DK9.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

The conversion of cytosine to 5-methylcytosine (mC) at 5'-CpG-3' dinucleotides creates an epigenetic mark for gene silencing, and DNA methylation is important for numerous cellular processes and essential for embryonic development.^{1, 2} However, deamination of m⁵C to T yields mutagenic G·T mispairs and is a major source of C→T transitions at CpG sites, which are among the most frequent point mutations in cancer and genetic disease.³⁻⁵ Mutations caused by mC deamination also effect epigenetic integrity, due to the loss of CpG sites.

While the repair of G·T mismatches arising during replication is directed at the new strand via mismatch repair (MMR), the repair of mC deamination must involve excision of T (not G). Two mammalian base excision repair (BER) enzymes can selectively remove T from G·T mispairs, methyl binding domain IV (MBD4)^{6, 7} and thymine DNA glycosylase (TDG).^{8, 9} Inactivation of MBD4 (also called MED1) in mice (*Mbd4*^{-/-}) causes increased C→T transitions and promotes tumor formation, indicating that MBD4 suppresses CpG mutability and tumorigenesis *in vivo*.^{10, 11} In addition to this important repair function, MBD4 has also been implicated in active DNA demethylation, by initiating BER of G·T mispairs generated by a deaminase enzyme (AID or APOBEC).^{12, 13} Thus, the ability of MBD4 to find and excise deaminated mC is important for maintaining genetic and epigenetic integrity.

How repair enzymes distinguish modified bases from undamaged DNA remains an important question, particularly for G·T mismatch glycosylases, which excise one of the canonical bases from DNA. These enzymes must balance the requirement for efficient excision of thymine from mutagenic G·T lesions and the need to minimize aberrant excision of T from the huge excess of normal A·T pairs, which can be mutagenic and cytotoxic.¹⁴ Such a compromise could result in suboptimal G·T repair activity, which could contribute to the high mutational frequency at CpG sites in cancer and genetic disease.^{10, 15} Our understanding of how MBD4 attains the specificity needed to excise thymine from G·T but not A·T pairs remains poor, due partly to a lack of structural information.

Full-length MBD4 (580 residues, human) contains a C-terminal DNA glycosylase domain (residues 426-580), and a methyl-CpG binding domain (MBD, residues 82-147), separated by a region of unknown function that is predicted to be largely disordered.^{6, 7, 16} Previous studies show that the glycosylase domain retains the same or slightly greater activity compared to full-length MBD4 for G·T substrates.^{6, 7, 16} The structure of DNA-free MBD4 glycosylase domain has been solved,^{16, 17} confirming that it belongs to the helix-hairpin-helix (HhH) superfamily of enzymes that also includes MutY, MIG, EndoIII, OGG1, AlkA, and TAG.^{18, 19} However, no structure of an enzyme-DNA complex has been reported for MBD4, or for the related thermophilic MIG enzymes, which also excise T from G·T mispairs.²⁰

To address this problem, we solved a crystal structure of the human MBD4 glycosylase domain (residues 426-580) bound to DNA containing a tetrahydrofuran (THF) abasic site analog paired with guanine (G·THF). The structure reveals details regarding how MBD4 glycosylase distinguishes G·T mispairs from normal A·T pairs, recognition of lesions in a CpG context, and the active site environment that mediates base excision. MBD4 cannot excise 5-formylcytosine (fC) or 5-carboxylcytosine (caC), which are intermediates in a Tet-initiated pathway for active DNA demethylation.^{21, 22} Our structure suggests that MBD4 does not provide the electrostatic interactions needed to bind and excise fC or caC, bases that can be excised by TDG.²²⁻²⁴ Alignment of our MBD4 structure with MIG reveals key

differences between these related enzymes, particularly with regard to recognition of the mismatched guanine in G-T mispairs.

Results and Discussion

Overall Structure of the MBD4 Product Complex

We obtained crystals of the human MBD4 glycosylase domain (Δ N425) bound to DNA containing a tetrahydrofuran (THF) abasic nucleotide, a chemically stable mimic of the natural abasic (or AP) reaction product. The structure was solved by molecular replacement and refined to a crystallographic *R*-factor of 23.4 % and an *R*_{free} of 25.9 % at a resolution of 2.76 Å (Table 1).

As shown in Fig. 1A, the abasic nucleotide (THF) is flipped completely out of the DNA duplex and into the MBD4 active site, providing a snapshot of the enzyme-product (E-P) complex. The complementary DNA strand contains a guanine across from the THF nucleotide, in keeping with the specificity of MBD4 for excising T from G-T mispairs, and excising thymine analogs (e.g., 5-iodouracil) that are paired with guanine.^{7, 25} While the mismatched guanine (hereafter, the “opposing G”) remains intrahelical, it is nestled into a recognition pocket, as described in more detail below. MBD4 binds the minor groove and dramatically bends the DNA at the target site by about 57°, nearly identical to the distortion imposed by MutY (55°)²⁶ and comparable to that observed for many other DNA glycosylases.²⁷

MBD4 glycosylase contains ten alpha helices, including the signature HhH motif (α 7-loop- α 8) that is conserved in all members of the HhH superfamily (Fig. 1). Notably, at 140 residues (438-577, structured region) MBD4 glycosylase is the smallest HhH enzyme that has been structurally characterized to date. As such, the structure reported here reveals the minimal protein scaffold and enzyme-DNA interactions required for glycosylase activity in the HhH superfamily.

It is of interest to compare the overall structure of MBD4 glycosylase to two other HhH enzymes that initiate repair of mismatches in DNA, the 8-oxoG-A mismatch glycosylase MutY (Fig. 1B),^{28, 29} and the thermophilic G-T mismatch glycosylase MIG (Fig. 1C).³⁰ MBD4 glycosylase is much smaller than MutY (352 residues) and MIG (221 residues), and it lacks the iron-sulfur cluster [4Fe-4S] domain found in these and other HhH enzymes. As shown in Fig. 1B, MBD4 glycosylase exhibits remarkably high structural homology to MutY, involving nearly all regions of MBD4 (except α 10). Notably, a motif of MBD4 that contributes to recognition of the opposing guanine, α 5-loop- α 6 (discussed below) is highly conserved with MutY, a mismatch glycosylase that recognizes 8-oxoG (or G) as the mismatched base. However, the structures diverge substantially beginning with the final helix (α 10) of MBD4, and some catalytic residues of MutY (E188, L191, I192)²⁹ are absent in MBD4.

Although MIG and MBD4 both process G-T mispairs, the structural homology of MBD4 is lower for MIG compared to MutY. Indeed, for MIG, the homology is limited to helices α 1- α 3 and α 7- α 8 (HhH) of MBD4 (Fig. 1C). Remarkably, the structural alignment reveals that MIG completely lacks the motif in MBD4 (α 5-loop- α 6) and MutY that forms numerous contacts with the complementary strand and the opposing guanine (8 oxoG for MutY), as discussed below.

DNA Binding, DNA Distortion, and Nucleotide Flipping

As summarized in Fig. 2A, MBD4 forms numerous electrostatic and van der Waals interactions with the lesion-containing and complementary DNA strands. Together, these

interactions mediate DNA binding, nucleotide flipping, recognition of the mismatched guanine, and base excision. As shown in Fig. 2B, MBD4 contacts five backbone phosphates of the lesion-containing strand, two located 5' of the flipped nucleotide (+1, +2 positions) and three on its 3' side (-1, -2, -3). Notably, the HhH motif provides most of these interactions, including nearly all contacts at the -2 and -3 positions. As observed for other HhH glycosylases, one phosphate (-3) is anchored in part by a cation (in this case, K⁺), which is itself coordinated by the backbone oxygen of three HhH residues (Fig 2). Another phosphate (-2 position) is anchored by interactions with backbone amides of residues in the second helix of the HhH motif (G538, K539, Y540), together with the side chain amine of Asn467.

As shown in Fig. 2B, the 5' phosphate of the flipped nucleotide (+1) is displaced from the DNA helix, and the two flanking phosphates (+2, -1) are compressed together by >5 Å, such that the P₊₂ to P₋₁ distance is reduced to 7.4 Å (compared to an average of ~12.8 Å for other regions of the DNA). This "phosphate pinch" has been observed for many DNA glycosylases and is thought to facilitate or stabilize nucleotide flipping.^{29, 31, 32} It may also facilitate the chemical step, by helping to position the anionic phosphates to stabilize the transient glycosyl cation intermediate; such interactions account for up to 50% of the rate enhancement provided by UDG.³³⁻³⁵ Notably, for MBD4, the distances between C1' of the flipped nucleotide and the putative catalytic phosphates (+1, 5.6 Å; -1, 4.9 Å; -2, 8.5 Å) are similar to those observed for UDG^{31, 34} (+1, 5.5 Å; -1, 5.0 Å; -2, 9.9 Å), suggesting that such electrostatic interactions could also be important for the chemical step of the MBD4 reaction.

As shown in Fig. 2B, the guanidinium side chain of Arg468 penetrates the DNA minor groove, where it plugs the void in the DNA helix that is created by nucleotide flipping and forms hydrogen bonds with the two "pinched" phosphates (+2, -1). These phosphate contacts made by Arg468 could stabilize the phosphate pinch and help maintain the nucleotide in a flipped state. A bulky "plug" residue is found in most, but not all DNA glycosylases, which is thought to promote and/or stabilize nucleotide flipping.^{31, 36, 37} The corresponding "plug" residue in other HhH enzymes varies widely, from Gln in MutY and Endo III, to Leu in AlkA, Asn in hOGG1, and no side-chain (Gly) in TAG³⁸. Remarkably, in addition to MBD4, an Arg plug is found in both of the other known G·T mismatch glycosylases, including the HhH enzyme MIG,³⁰ and the unrelated TDG,³² which belongs to the UDG superfamily. Notably, mutation of the Arg plug to Leu causes a 30-fold loss in G·T activity for TDG,³⁷ and an Arg to Ala mutation gives a 17-fold loss in G·T activity for MIG.³⁰ Together, these structural and biochemical findings indicate an important role for an Arg side chain "plug" in mediating excision of T from G·T mispairs.

As shown in Figs. 1A, 2A, and 3, residues from the α5-α6 loop and helix α6 of MBD4 glycosylase form numerous interactions with the complementary DNA strand. In particular, MBD4 contacts both phosphates flanking the opposing guanine (G6), and wedges a Leu side chain (L508) between the opposing G and its 5' Cyt neighbor (C5). Together, these steric and electrostatic interactions contribute to the severe DNA bending imparted by MBD4. Indeed, stacking between the opposing G and the 5' Cyt is dramatically perturbed, where the 5' Cyt is tilted and displaced away from the opposing G. In addition, Leu511 also distorts the DNA, predominantly at the backbone of the dC nucleotide located 5' of the opposing G (i.e., C5). These interactions with the complementary strand may stabilize the flipped state of the target nucleotide, and contribute to recognition of the opposing guanine, as described below.

Specificity for Guanine as the Pairing Partner of the Target Base

Our structure reveals interactions that likely confer specificity for excising thymine from G·T rather than A·T base pairs. This specificity is important, because aberrant excision of T from A·T pairs can be mutagenic and cytotoxic.¹⁴ In addition, MBD4 exhibits high specificity for excising uracil and analogs such as 5-halogenated uracils when they are paired with guanine.^{7, 25} As shown in Fig. 3, the opposing G remains intrahelical and occupies a recognition pocket, forming numerous interactions with MBD4. The opposing G remains partially stacked with its 3' neighbor (T7), though it is laterally displaced such that it stacks with both bases of this flanking base pair (T7-A17). As noted above, stacking with the 5' Cyt (C5) is completely disrupted by the Leu508 "wedge" residue. The opposing G is bound in a pocket that provides favorable steric and electrostatic interactions with its 5' face (Leu508), its minor groove edge (Leu506, Gly507), and its Watson-Crick base-pairing edge (Arg468, Thr469) (Fig. 3). The backbone oxygens of Leu506 and Arg468 contact the exocyclic amine (N2H₂) of the opposing G, and the Arg468 backbone oxygen also contacts its imino nitrogen (N1H). Importantly, these contacts are not compatible with adenine and are likely to confer specificity for excising thymine from G·T versus A·T pairs, and specificity for excising thymine analogs that are paired with guanine. Notably, the Arg468 backbone contacts with the opposing G could stabilize the insertion of Arg468 into the DNA helix and thereby help stabilize nucleotide flipping, perhaps supplementing the effect of phosphate contacts involving the Arg468 side chain.

In addition to these selective interactions with the opposing guanine, mismatch specificity may derive in part from the inherently weaker stability of G·T mispairs relative to normal A·T base pairs in DNA; T is more readily flipped from a G·T mispair, which adopts a wobble structure.³⁹⁻⁴¹ In addition, the differing geometry of a G·T mispair relative to an A·T pair could potentially play a role in the early stages of mismatch recognition by MBD4. Additional structural and biochemical studies are needed to assess the role of base-pairing stability and geometry in mismatch recognition by MBD4.

The new MBD4 structure here can account for the previous finding that MBD4 exhibits nearly the same activity for excising thymine when it is paired with either guanine or O6-methylguanine.²⁵ As shown in Fig. 3, it appears that a methyl substituent could be accommodated at the O6 oxygen of the opposing G.

It is instructive to consider how the other HhH enzymes MutY, hOGG1, and MIG recognize the "opposing base". Recall that MutY removes A from A·oxoG but not A·T base pairs. The interactions provided by MBD4 to the opposing G are largely conserved in MutY, which provides the same backbone contacts with N1H and N2H₂ of oxoG. In addition, MutY provides another N2H₂ contact and interactions with N7H and O8 of oxoG involving a C-terminal domain that is absent in MBD4.²⁶ In contrast, OGG1, which excises oxoG from oxoG·C but not oxoG·A pairs, recognizes the opposing cytosine using only side chain interactions,⁴² none of which are conserved in either MBD4 or MutY.

Remarkably, a structural alignment of MIG with MBD4 reveals that MIG lacks the motif of MBD4 (α 5-loop- α 6) that provides most of the recognition pocket for the opposing G, disrupts base stacking of the opposing G, and mediates bending of the complementary strand. While Arg468 and Thr469 of MBD4 are conserved in MIG and could potentially contact the opposing G, the absence of the α 5-loop- α 6 motif found in MBD4 and MutY raises the question of how MIG discriminates between G·T and A·T pairs. Consistent with the potential for less stringent recognition of an opposing guanine, biochemical studies indicate that MIG can act upon many types of mismatched base pairs, including T·C and U·C.^{20, 43}

Notably, TDG, the other mammalian G·T glycosylase, also uses backbone oxygens to form selective interactions with the N1H and N2H₂ of the opposing G,³² Together, these findings reveal that two unrelated glycosylases use a common strategy to recognize guanine as the base-pairing partner of the target base.

MBD4 Glycosylase Domain Has Little Preference for Acting at a CpG Site

TDG exhibits very high specificity for acting on G·T mismatches (and other substrates) in a CpG context,⁴⁴⁻⁴⁷ due likely to interactions that it forms with guanine positioned 3' of the flipped nucleotide (i.e., CpG/TpG).³² As shown in Fig. 4A, MBD4 glycosylase forms no interactions with the G·C pair (G19·C5) positioned 3' of the flipped nucleotide that would be expected to confer substantial specificity for acting at a CpG site. Consistent with this structural observation, kinetics experiments indicate that MBD4 glycosylase activity is largely independent of the base pair located 3' of G·T mismatch (Fig. 4B). Thus, compared to the activity for a G·T mismatch in a CpG context, the activity is merely 1.5-fold weaker for a TpG or GpG context, and 2.7-fold weaker for an ApG context. Similarly, full-length MBD4 exhibits modestly weaker activity for G·T mismatches in a GpG versus CpG context⁶ (activity for other contexts were not reported). In stark contrast, the G·T activity of TDG is reduced by 37-, 96-, and 582-fold for TpG, GpG, or ApG relative to a CpG context.³²

Our observation that MBD4 glycosylase does not interact selectively with the 3' G·C pair (G19·C5) is also consistent with previous findings that the glycosylase domain exhibits similar activity for G·T mismatches (and other substrates) in a methylated versus a non-methylated CpG context (i.e., mCpG·T versus CpG·T).^{6, 7, 25} Notably, previous studies show that full-length MBD4 and the isolated MBD bind preferentially to G·T mismatches in a methylated CpG context, though the glycosylase activity of full-length MBD4 does not depend substantially on whether a CpG site is methylated.^{6, 7} These findings suggest that the MBD serves to localize full-length MBD4 to G·T mismatches arising at methylated CpG sites, though it does not appear to directly modulate the specificity or activity of the bound glycosylase domain. Thus, there appears to be little biochemical evidence for a model involving *simultaneous* binding of the methyl-binding and glycosylase domains to a single G·T mismatch. While the MBD likely binds the major groove of DNA⁴⁸ and the glycosylase binds the minor groove (Fig. 1A), the severe DNA distortion imparted by the glycosylase domain could preclude simultaneous binding of both domains to a single mCpG·T site. However, the methyl-binding domain could target the full-length protein to a single lesion, even if it is subsequently displaced by the glycosylase domain. Alternatively, the MBD and glycosylase domains, which are separated by about 280 residues, could bind neighboring CpG sites.¹⁶ Additional studies are needed to address this question.

Residues Implicated in Substrate Recognition and Catalysis

As shown in Fig. 5A, the structure reveals the active site environment presented to the flipped abasic (THF) nucleotide, and suggests residues that could be important for substrate recognition and catalysis. Also shown in Fig. 5A is a model for a dT nucleotide flipped into the active site. Together, the crystal structure (THF) and model (dT) suggest that the MBD4 active site is rather accommodating, particularly near the C5 position of the flipped base, as discussed below.

The results here suggest a key role for Tyr540 (Figs. 2A, 5A), a residue that is conserved in MIG and MutY enzymes.^{29, 30} Together, our findings and previous studies suggest that the Tyr540 hydroxyl may contact the O2 oxygen of a flipped thymine base (or T analog).^{16, 29, 30} Such an interaction could help retain the flipped nucleotide in the active site and/or stabilize the negative charge that is expected to develop on the base (including O2) in the chemical step of the reaction.^{49, 50} Consistent with an important catalytic role, mutation

of the corresponding Tyr in MIG leads to greatly reduced (Y126F, 62-fold loss; Y126S, 261-fold loss) or completely depleted (Y126K) activity for a G-T substrate.³⁰

With the exception of TAG enzymes,⁵¹ all HhH glycosylases contain an Asp downstream of the HhH motif, and this conserved residue is Asp560 in MBD4. As shown in Figs. 2B and 5A, the carboxylate side chain of Asp560 is positioned close to C1' (3.4 Å) of the flipped abasic sugar (THF), as observed for the conserved Asp in other HhH enzymes.^{29, 42, 52} It has been proposed that the Asp activates the nucleophile (water) or electrostatically stabilizes the positively charged sugar (glycosyl cation) that develops upon glycosylic bond cleavage. However, support for such an essential role is questioned by findings that the Asp to Ala mutation reduces activity by only 16-fold in MIG,³⁰ and by findings of similarly modest effects caused by some mutations of the conserved Asp in some other HhH enzymes.⁵³ As noted previously,⁵³ the conserved Asp invariably resides at the N-terminus of an α -helix and forms helix-capping interactions, as it does in MBD4 (Fig. 5A), suggesting a potential structural role. Additional studies are needed to determine the importance and role of Asn560 for MBD4 activity.

As shown in Figs 2A and 5A, the Lys562 side chain NH_3^+ contacts O4' of the flipped abasic sugar. While such a contact may be favorable for binding the reaction product (abasic site), it seems likely that this would slow rather than accelerate the chemical step, as the cationic NH_3^+ side chain would likely destabilize the glycosyl cation that is generated upon glycosylic bond cleavage.^{53, 54} A reasonable role for Lys562 in the chemical step could be electrostatic stabilization of the departing anionic base, perhaps via contact with the O4 oxygen. Such a contact could also favor nucleotide flipping to give the reactive enzyme-substrate (Michaelis) complex. Notably, Lys562 is not conserved in MutY or MIG enzymes, as it resides in a region of structural divergence between MBD4 and the other enzymes.

Our results suggest that Leu466, which is conserved in MIG and MutY,^{29, 30} helps to define the active site pocket and may stack with the flipped base¹⁶ and potentially serve to limit the extent of flipping. It was previously proposed that the Gln449 side chain oxygen contacts the imino N3H of flipped thymine, and that thymine O4 is contacted by the Val448 backbone NH.¹⁶ Our model with flipped dT is consistent with the proposed role for Gln449, and perhaps Val488 (V488-NH to Thy-O4 distance of 5 Å in the model). However, determination of the role played by these residues will require a structure of MBD4 bound to a non-cleavable substrate analog and mutational studies.

Together, our results and previous biochemical studies do not support the proposal that the methyl of thymine is recognized by a hydrophobic patch comprised of Gly471 and Ile475.¹⁶ These residues are quite removed from the flipped THF in the crystal structure (distance to THF-C1' is 7.7 Å for G471-C α and 9.2 Å for I475-C γ 1), and removed from the methyl group of the modeled dT (distance to Thy-CH₃ is 7.6 Å for G471-C α and 9.2 Å for I475-C γ 1). Moreover, our results suggest the MBD4 active site could accommodate pyrimidines with C5 substituent that are substantially larger than a methyl group (Fig. 5A). Consistent with this possibility, previous biochemical studies show that MBD4 excises 5-iodouracil (5IU) at the same rate as 5-fluorouracil (5FU),²⁵ despite the much greater size of 5IU relative to 5FU.⁵⁵ Notably, 5IU is excised over 100-fold faster than thymine,²⁵ despite the larger size of the iodo versus methyl substituent.⁵⁵ Observation that the non-catalyzed hydrolyses for 5-I-dU and 5-F-dU proceed at about the same rate⁵⁵ indicates that MBD4-mediated excision of 5IU is not substantially slowed by steric hindrance, despite the bulk of the iodo substituent. Together, these observations indicate that the MBD4 active site can accommodate very bulky substituents at the C5 position of pyrimidines. In contrast, TDG activity is 3300-fold slower for 5IU versus 5FU, due largely to steric hindrance with the iodo substituent.^{46, 55}

MBD4 Cannot Excise 5-formylcytosine or Other Oxidized Forms of 5-methylcytosine

Tet enzymes can catalyze iterative oxidation of mC to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5carboxylcytosine (caC), which is important for epigenetic regulation and active DNA demethylation.^{21, 22, 56} Recent studies show that TDG can excise fC and caC from G·fC and G·caC pairs, suggesting a novel BER-mediated pathway for active DNA demethylation.^{22, 23} Previous studies indicate MBD4 may also participate in active DNA demethylation, by initiating BER of G·T mispairs generated by a deaminase.^{12, 13} Given the robust fC activity of TDG,²³ we examined the ability of MBD4 glycosylase to excise fC, caC, and hmC from DNA. As shown in Fig. 5B, we find no detectable activity for any of these oxidized mC bases, consistent previous reports for hmC and caC.²² It was proposed that the inability of MBD4 to excise caC is due to steric hindrance involving Gly471 and Ile475 with the carboxyl group of caC.²⁴ However, this is not supported by our structural observations (Fig. 5A) or by previous findings that MBD4 can rapidly excise the bulky 5IU base.²⁵ Rather, we suggest that the inability of MBD4 to excise caC is due to the lack of electrostatic interactions needed to bind caC and stabilize the additional negative charge that likely develops on the departing caC base upon cleavage of the base-sugar bond. The TDG active site provides electrostatic interactions that favor the binding of caC,²⁴ though it remains unclear how TDG activates the caC base, which harbors a negative charge (on the carboxyl group) that would be expected to suppress base-sugar bond cleavage.²³ The absence of MBD4 activity for fC is also likely due to the lack of specific electrostatic interactions needed to bind fC or stabilize negative charge that likely develops on the formyl oxygen upon cleavage of the base-sugar bond.²³ The presence of such stabilizing interactions likely confers TDG with the ability to rapidly excise fC.²³

Materials and Methods

Materials

A plasmid for expressing the glycosylase domain of human MBD4 (residues 426-580 or Δ 425) was obtained from DNA 2.0 and transformed into *Escherichia coli* BL21(DE3) cells (Stratagene). Cells were grown in LB at 37 °C to an OD₆₀₀ of 0.6, then expression was induced at 22 °C with 0.4 mM IPTG (isopropyl β -D-thiogalactoside) and continued for ~16 hrs. The enzyme was purified using Ni-affinity chromatography (Qiagen), and the N-terminal poly-His peptide was removed by thrombin cleavage (4 °C, 18 hrs), leaving two non-native N-terminal residues (Gly-Ser). Additional purification steps included ion exchange, using SP and Q sepharose chromatography columns (GE Healthcare). The purified enzyme was quantified by absorbance (using $\epsilon^{280} = 54.4 \text{ mM}^{-1}\text{cm}^{-1}$) as previously described,⁵⁷ flash frozen, and stored at -80 °C.

The duplex DNA used for the crystal structure consisted of 5'-TGTCCTA FGTCT-3', where F is a tetrahydrofuran (THF) abasic site analog, and a complement, 5'-AAGACGTGGAC-3'. The duplex is designed such that F is positioned opposite G and in a CpG context (i.e., the target F is 5' to a G-C base pair). The 19 bp G-T DNA substrates used to determine the dependence of MBD4 activity on a CpG context were previously described.⁴⁶ The 28 bp DNA substrates containing a G-T, G·fC, G·caC, or G·hmC were previously described.^{23, 58}

Crystallization and Structure Determination

The sample used for crystallization contained 0.25 mM (4.6 mg/ml) MBD4 glycosylase and 0.3 mM DNA in a buffer of 0.02 M sodium phosphate pH 6.5, 0.1 M NaCl, 0.5 mM DTT, and 0.2 mM EDTA. Crystals were grown by vapor diffusion at 22 °C in sitting drop format, using 0.5 μ l of the MBD4-DNA sample and 0.5 μ l of mother liquor (0.2 M potassium acetate, 20% PEG 3350). Crystals were grown for two days, transferred to mother liquor

supplemented with 20% glycerol, and flash-frozen in liquid nitrogen. X-ray diffraction data extending to 2.76 Å were collected (in-house) using a Rigaku Micromax 7 generator equipped with a Rigaku R-Axis IV++ image plate detector. The images were processed and scaled using the Denzo and Scalepack.⁵⁹ Initial phases were obtained by molecular replacement using the program Phaser⁶⁰ and DNA-free MBD4 (PDB ID: 3IHO)¹⁷ and DNA from a MutY-DNA complex (PDB ID: 3FSP) as search models. Model building was performed with Coot,⁶¹ and refinement of the model phases was performed using Refmac.⁶² The refinement protocol included TLS refinement, with the protein and DNA as separate TLS groups. Structural figures were made with PyMOL (<http://www.pymol.org>).

Glycosylase Activity for MBD4

The enzymatic turnover of MBD4 is suppressed by slow product release and/or product inhibition.⁷ Thus, the glycosylase reactions were performed under saturating enzyme conditions, with 0.5 μM DNA substrate and 5 μM enzyme, in HEMN.1 buffer (0.02 M HEPES pH 7.5, 0.1 M NaCl, 0.2 mM EDTA, 2.5 mM MgCl₂) at 22 °C. Reactions were initiated by mixing concentrated enzyme with buffered substrate, quenched with 50% (v:v) 0.3M NaOH, 0.03M EDTA, and then incubated at 85 °C for 15 m to cleave the DNA at nascent abasic sites. The reactions involving G-fC, G-caC, or G-hmC were quenched with 50% (v:v) 0.1M piperidine, 0.03M EDTA. The reaction progress was analyzed by electrophoresis²³ or anion-exchange HPLC.⁵⁵

For the reactions analyzed by HPLC, the data were fitted by non-linear regression to eq. 1:

$$\text{fraction product} = A (1 - \exp(-k_{\text{obs}}t)) \quad (1)$$

where A is the amplitude, k_{obs} is the rate constant, and t is reaction time (min). The saturating enzyme conditions ($[E] \gg [S] > K_d$) used here provide a rate constant (k_{obs}) that is not impacted by enzyme-substrate association or by product release or inhibition, such that k_{obs} reflects the maximal rate of base excision ($k_{\text{obs}} \approx k_{\text{max}}$). To confirm saturating enzyme conditions, experiments were in some cases repeated with two enzyme concentrations (5 μM and 10 μM), providing equivalent rate constants within experimental uncertainty.

Acknowledgments

This work was supported in part by National Institutes of Health grant R01-GM72711 (to ACD).

Abbreviations

AID	activation induced deaminase
AlkA	3-methyladenine DNA glycosylase II
APOBEC	apolipoprotein β mRNA editing complex
BER	base excision repair
caC	5-carboxylcytosine
EndoIII	endonuclease III
fC	5-formylcytosine
HhH	Helix-hairpin-Helix
hmC	5-hydroxymethylcytosine

MBD	methyl-CpG binding domain
MBD4	methyl binding domain IV
MIG	mismatch specific glycosylase
mC	5-methylcytosine
MutY	A/G-specific adenine glycosylases
OGG1	8-oxoguanine DNA glycosylase
TAG	3-methyladenine DNA glycosylase I
TDG	thymine DNA glycosylase
Tet	ten eleven translocation
THF	tetrahydrofuran
UDG	uracil DNA glycosylase

References

1. Feng S, Jacobsen SE, Reik W. *Science*. 2010; 330:622–627. [PubMed: 21030646]
2. Deaton AM, Bird A. *Genes Dev*. 2011; 25:1010–1022. [PubMed: 21576262]
3. Rideout WM 3rd, Coetzee GA, Olumi AF, Jones PA. *Science*. 1990; 249:1288–1290. [PubMed: 1697983]
4. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. *Science*. 2006; 314:268–274. [PubMed: 16959974]
5. Pfeifer GP, Besaratinia A. *Hum Genet*. 2009; 125:493–506. [PubMed: 19308457]
6. Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. *Nature*. 1999; 401:301–304. [PubMed: 10499592]
7. Petronzelli F, Riccio A, Markham GD, Seeholzer SH, Stoerker J, Genuardi M, Yeung AT, Matsumoto Y, Bellacosa A. *J. Biol. Chem*. 2000; 275:32422–32429. [PubMed: 10930409]
8. Wiebauer K, Jiricny J. *Nature*. 1989; 339:234–236. [PubMed: 2716851]
9. Neddermann P, Jiricny J. *J Biol Chem*. 1993; 268:21218–21224. [PubMed: 8407958]
10. Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A. *Science*. 2002; 297:403–405. [PubMed: 12130785]
11. Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K, Fazzari M, Jin B, Brown AM, Lipkin M, Edelmann W. *Proc Natl Acad Sci U S A*. 2002; 99:14937–14942. [PubMed: 12417741]
12. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. *Cell*. 2008; 135:1201–1212. [PubMed: 19109892]
13. Kim MS, Kondo T, Takada I, Youn MY, Yamamoto Y, Takahashi S, Matsumoto T, Fujiyama S, Shiode Y, Yamaoka I, Kitagawa H, Takeyama K, Shibuya H, Ohtake F, Kato S. *Nature*. 2009; 461:1007–1012. [PubMed: 19829383]
14. Boiteux S, Guillet M. *DNA Repair*. 2004; 3:1–12. [PubMed: 14697754]
15. Waters TR, Swann PF. *Mutat. Res*. 2000; 462:137–147. [PubMed: 10767625]
16. Wu P, Qiu C, Sohail A, Zhang X, Bhagwat AS, Cheng X. *J Biol Chem*. 2003; 278:5285–5291. [PubMed: 12456671]
17. Zhang W, Liu Z, Crombet L, Amaya MF, Liu Y, Zhang X, Kuang W, Ma P, Niu L, Qi C. *Biochem Biophys Res Commun*. 2011; 412:425–428. [PubMed: 21820404]
18. Fromme JC, Banerjee A, Verdine GL. *Curr Opin Struct Biol*. 2004; 14:43–49. [PubMed: 15102448]
19. Hitomi K, Iwai S, Tainer JA. *DNA Repair (Amst)*. 2007; 6:410–428. [PubMed: 17208522]

20. Horst JP, Fritz HJ. *EMBO J.* 1996; 15:5459–5469. [PubMed: 8895589]
21. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. *Science.* 2011; 333:1300–1303. [PubMed: 21778364]
22. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. *Science.* 2011; 333:1303–1307. [PubMed: 21817016]
23. Maiti A, Drohat AC. *J Biol Chem.* 2011; 286:35334–35338. [PubMed: 21862836]
24. Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C. *Nat Chem Biol.* Feb 12.2012 2012 [Epub ahead of print].
25. Turner DP, Cortellino S, Schupp JE, Caretti E, Loh T, Kinsella TJ, Bellacosa A. *Cancer Res.* 2006; 66:7686–7693. [PubMed: 16885370]
26. Fromme JC, Banerjee A, Huang SJ, Verdine GL. *Nature.* 2004; 427:652–656. [PubMed: 14961129]
27. Stivers JT, Jiang YL. *Chem Rev.* 2003; 103:2729–2759. [PubMed: 12848584]
28. Guan Y, Manuel RC, Arvai AS, Parikh SS, Mol CD, Miller JH, Lloyd S, Tainer JA. *Nat Struct Biol.* 1998; 5:1058–1064. [PubMed: 9846876]
29. Lee S, Verdine GL. *Proc Natl Acad Sci U S A.* 2009; 106:18497–18502. [PubMed: 19841264]
30. Mol CD, Arvai AS, Begley TJ, Cunningham RP, Tainer JA. *J Mol Biol.* 2002; 315:373–384. [PubMed: 11786018]
31. Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA. *Nature.* 1996; 384:87–92. [PubMed: 8900285]
32. Maiti A, Morgan MT, Pozharski E, Drohat AC. *Proc Natl Acad Sci USA.* 2008; 105:8890–8895. [PubMed: 18587051]
33. Dinner AR, Blackburn GM, Karplus M. *Nature.* 2001; 413:752–755. [PubMed: 11607036]
34. Jiang YL, Ichikawa Y, Song F, Stivers JT. *Biochemistry.* 2003; 42:1922–1929. [PubMed: 12590578]
35. Parker JB, Stivers JT. *Biochemistry.* 2008; 47:8614–8622. [PubMed: 18652484]
36. Jiang YL, Kwon K, Stivers JT. *J Biol Chem.* 2001; 276:42347–42354. [PubMed: 11551943]
37. Maiti A, Morgan MT, Drohat AC. *J Biol Chem.* 2009; 284:36680–36688. [PubMed: 19880517]
38. Metz AH, Hollis T, Eichman BF. *Embo J.* 2007; 26:2411–2420. [PubMed: 17410210]
39. Moe JG, Russu IM. *Biochemistry.* 1992; 31:8421–8428. [PubMed: 1327102]
40. Krosky DJ, Schwarz FP, Stivers JT. *Biochemistry.* 2004; 43:4188–4195. [PubMed: 15065862]
41. Darwanto A, Theruvathu JA, Sowers JL, Rogstad DK, Pascal T, Goddard W 3rd, Sowers LC. *J Biol Chem.* 2009; 284:15835–15846. [PubMed: 19324873]
42. Bruner SD, Norman DP, Verdine GL. *Nature.* 2000; 403:859–866. [PubMed: 10706276]
43. Yang H, Fitz-Gibbon S, Marcotte EM, Tai JH, Hyman EC, Miller JH. *J Bacteriol.* 2000; 182:1272–1279. [PubMed: 10671447]
44. Sibghat U, Gallinari P, Xu YZ, Goodman MF, Bloom LB, Jiricny J, Day RS 3rd. *Biochemistry.* 1996; 35:12926–12932. [PubMed: 8841138]
45. Waters TR, Swann PF. *J Biol Chem.* 1998; 273:20007–20014. [PubMed: 9685338]
46. Morgan MT, Bennett MT, Drohat AC. *J Biol Chem.* 2007; 282:27578–27586. [PubMed: 17602166]
47. Abu M, Waters TR. *J Biol Chem.* 2003; 278:8739–8744. [PubMed: 12493755]
48. Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M, Shirakawa M. *Cell.* 2001; 105:487–497. [PubMed: 11371345]
49. Drohat AC, Stivers JT. *Biochemistry.* 2000; 39:11865–11875. [PubMed: 11009598]
50. Drohat AC, Stivers JT. *J. Am. Chem. Soc.* 2000; 122:1840–1841.
51. Drohat AC, Kwon K, Krosky DJ, Stivers JT. *Nat. Struct. Biol.* 2002; 9:659–664. [PubMed: 12161745]
52. Hollis T, Ichikawa Y, Ellenberger T. *EMBO J.* 2000; 19:758–766. [PubMed: 10675345]
53. Berti PJ, McCann JA. *Chem. Rev.* 2006; 106:506–555. [PubMed: 16464017]
54. McCann JAB, Berti PJ. *J. Am. Chem. Soc.* 2008; 130:5789–5797. [PubMed: 18393424]

55. Bennett MT, Rodgers MT, Hebert AS, Ruslander LE, Eisele L, Drohat AC. J. Am. Chem. Soc. 2006; 128:12510–12519. [PubMed: 16984202]
56. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Science. 2009; 324:930–935. [PubMed: 19372391]
57. Morgan MT, Maiti A, Fitzgerald ME, Drohat AC. Nucleic Acids Res. 2011; 39:2319–2329. [PubMed: 21097883]
58. Maiti A, Drohat AC. DNA Repair. 2011; 10:545–553. [PubMed: 21474392]
59. Otwinowski Z, Minor W. Methods Enzym. 1997; 276:307–326.
60. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Acta Crystallogr D. 2005; 61:458–464. [PubMed: 15805601]
61. Emsley P, Cowtan K. Acta Crystallogr D. 2004; 60:2126–2132. [PubMed: 15572765]
62. Winn MD, Isupov MN, Murshudov GN. Acta Crystallogr D. 2001; 57:122–133. [PubMed: 11134934]

Highlights

- MBD4 initiates repair of mutagenic G·T mispairs arising from 5-methylcytosine deamination at methylated CpG sites.
- A crystal structure of DNA-bound human MBD4 glycosylase reveals the molecular basis of mismatch specificity.
- MBD4 glycosylase has little specificity for acting at CpG sites and lacks catalytic groups needed to excise 5-formylcytosine or 5-carboxylcytosine.
- The results advance our understanding of mismatch glycosylase specificity and reveal key differences between MBD4 and TDG

Fig. 1. Crystal structure of MBD4 glycosylase bound abasic DNA and alignment with other HhH glycosylases

(A) Two orientations of the MBD4·THF-DNA structure. MBD4 is blue, with the HhH motif in magenta (Y530-C549). The catalytic loops are red, and include α 2- α 3 (N467-S470), α 5- α 6 (K504-L508), and α 9- α 10 (V556-D560). Nucleotides in the target DNA strand are yellow, including the flipped THF abasic analog. Complementary DNA is light orange, and the mismatched guanine is in green. (B) Structural alignment of MBD4 and MutY (*B. stearothermophilus*, PDB ID 3G0Q) shows strong homology with regions including helices α 1- α 9 but not α 10 of MBD4 (RMS deviation = 1.33 Å for backbone atoms, 78 residues). (C) Alignment of MBD4 glycosylase and MIG (*M. thermautotrophicus*, PDB ID = 1KEA) reveals that the homology is limited to regions including helices α 1- α 3 and α 7- α 9 of MBD4 (RMS deviation = 1.48 Å for backbone atoms, 73 residues).

Fig. 2. Detailed interactions between MBD4 glycosylase and DNA product

(A) An interaction map summarizes the electrostatic and van der Waals interactions observed in the crystal structure (dashed and curvy lines, respectively). The coloring corresponds to that used in the structural figures. (B) Interactions between MBD4 glycosylase and the lesion-containing DNA strand. The color scheme is identical to that used on Fig. 1A. MBD4 residues that contact the DNA are colored cyan, and the K^+ ion is in purple. For clarity, the complementary strand is omitted, except for the mismatched guanine (green).

Fig. 3. Recognition of the mismatched Guanine and Distortion of the complementary strand
 Close-up view of the interactions MBD4 forms with the mismatched guanine (“opposing G”), including hydrogen bonds involving Gua N1H and N2H₂ and backbone oxygens of R468 and Leu506. Also shown are interactions that mediate binding and distortion of the complementary strand, including a phosphate contact with the Tyr509 side chain, and distortion mediated by L508, D510, and L511. It appears that a methyl substituent could be readily accommodated at O6 of the mismatched guanine. The coloring is the same as in Figs. 1 and 2.

Fig. 4. MBD4 glycosylase does not exhibit strong specificity for a CpG context

(A) The structure reveals no selective electrostatic interactions with the G·C pair located 3' of the flipped nucleotide (G19·C5) that would be expected to provide substantial specificity for acting on G·T mispairs in a CpG site. This figure also provides another view of the DNA distortion imposed by MBD4 glycosylase, including the role of N467, R468, and L508. (B) Bar graph shows the results of kinetics experiments performed for MBD4 glycosylase with G·T substrates in which the CpG context was altered. Also shown are previously published results for TDG⁴⁶. The key shows the DNA sequences corresponding to the CpG, TpG, GpG, and ApG nomenclature; the thymine base excised by the enzyme is underlined.

Fig. 5. MBD4 active site and glycosylase assay

(A) Close-up view of the MBD4 active site. Residues that define the active and/or were previously suggested to mediate thymine recognition are shown in cyan. The DNA backbone is orange and the flipped THF is yellow. A helix-capping hydrogen bond is shown (D560-O81 to L563-NH). A model structure with a dT nucleotide flipped into the active site (yellow-orange) was obtained by superimposing the target DNA strand from a MutY enzyme-substrate complex (PDBID: 3G0Q) with the THF strand of the MBD4 structure, and converting the flipped dA (MutY structure) to dT. This approach for modeling was used because we found that simply replacing THF with dT in the MBD4 structure results in severe steric clashes between the thymine base and Leu466 and Tyr540 (not shown). (B) Shown is an electrophoretic assay of MBD4 glycosylase activity for potential substrates, including G·T, G·fC, G·caC, and G·hmC. Activity is observed only for the G·T substrate. MBD4 glycosylase (5 μ M) was incubated with DNA (0.5 μ M) for 30 min at 22 $^{\circ}$ C prior to quenching.

Table 1

Data collection and refinement statistics

Data collection	
Space Group	P 21 21 21
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	41.0, 55.2, 122.2
α , β , γ (°)	90, 90, 90
Resolution (Å)	61.1-2.76 (2.85-2.76) *
<i>R</i> _{sym}	0.156 (0.941)
Mean <i>I</i> / σ <i>I</i>	10.6 (1.3)
Completeness (%)	99.9 (100)
Redundancy	9.8 (9.7)
Wilson B-factor (Å ²)	56.5
Refinement	
Resolution (Å)	61.1-2.76
Number of reflections	7233
<i>R</i> _{work} / <i>R</i> _{free}	23.4/25.9
Number of atoms	1604
Protein	1165
DNA	437
Water	1
Metal	1
<i>B</i> -factors	
Protein	74.5
DNA	97.6
Water	76.6
Metal	89.9
Ramachandran Plot	
Favored regions (%)	96.4
Allowed regions (%)	3.6
Outliers (%)	0
RMSD	
Bond lengths (Å)	0.012
Bond angles (°)	1.788

* Values shown in parenthesis are for highest resolution shell