Catalytic mechanism of the mismatch-specific DNA glycosylase methyl-CpG-binding domain 4

Hala Ouzon-Shubeita1, Hunmin Jung1, Michelle H. Lee, Myong-Chul Koag, Seongmin Lee
The Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712

Abstract

Thymine:guanine base pairs are major promutagenic mismatches occurring in DNA metabolism. If left unrepaired, these mispairs can cause C to T transition mutations. In humans, T:G mismatches are repaired in part by mismatch-specific DNA glycosylases such as methyl-CpG-binding domain 4 (hMBD4) and thymine DNA glycosylase. Unlike lesion-specific DNA glycosylases, T:G-mismatch-specific DNA glycosylases specifically recognize both bases of the mismatch and remove the thymine but only from mispairs with guanine. Despite the advances in biochemical and structural characterizations of hMBD4, the catalytic mechanism of hMBD4 remains elusive. Herein, we report two structures of hMBD4 processing T:G-mismatched DNA. A high-resolution crystal structure of Asp560Asn hMBD4-T:G complex suggests that hMBD4-mediated glycosidic bond cleavage occurs via a general base catalysis mechanism assisted by Asp560. A structure of wild-type hMBD4 encountering T:G-containing DNA shows the generation of an apurinic/apyrimidinic (AP) site bearing the C1´-(S)-OH. The inversion of the stereochemistry at the C1´ of the AP-site indicates that a nucleophilic water molecule approaches from the back of the thymine substrate, suggesting a bimolecular displacement mechanism (S_{N2}) for hMBD4-catalyzed thymine excision. The AP-site is stabilized by an extensive hydrogen bond network in the MBD4 catalytic site, highlighting the role of MBD4 in protecting the genotoxic AP-site.

INTRODUCTION

Thymine:guanine mispairs, the most abundant mismatches found in genomic DNA [1,2], are produced by various endogenous processes (Figure 1). Spontaneous replication errors of DNA polymerases preferentially produce T:G mismatches [2–4]. 5-Methylcytosine, which is found in the CpG islands, is susceptible to spontaneous deamination, generating the
promutagenic T:G pairs \([5,6]\). In addition, T:G mismatches at the CpG islands can be produced by active DNA demethylation processes, where activation-induced deaminase (AID) hydrolyzes the exocyclic N4 amine of 5-methylcytosine to yield thymine opposite guanine \([7]\). The resulting T:G mismatches, if left unrepaired, can induce C to T transition mutations. The promutagenic T:G mispairs can be corrected by the DNA mismatch repair pathway \([8]\), as well as T:G mismatch-specific DNA glycosylases, such as thymine DNA glycosylase (TDG) and methyl-CpG-binding domain 4 (MBD4/MED1) \([8–10]\). These mismatch-specific DNA glycosylases recognize both G and T bases and cleave thymine opposite guanine to produce an apyrimidinic/apurinic site (AP-site), which is processed further by downstream base excision DNA repair enzymes (Figure 1A) \([11]\).

MBD4 belongs to the superfamily of helix-hairpin-helix (HhH) DNA glycosylases that include 8-oxoguanine DNA glycosylase 1 (hOGG1), alkyladenine DNA glycosylase (AAG) and AlkA \([12]\). MBD4 is composed of an N-terminal methyl-CpG-binding domain and C-terminal glycosylase domain \([13]\). In addition to T:G base pairs, the enzyme has been shown to efficiently cleave uracil, 5-fluorouracil, and hydroxymethyluracil (hmU) when paired with guanine \([8,14,15]\). Inactivation of MBD4 promotes C to T mutations at CpG sites and increases tumorigenesis, suggesting MBD4 may play a role as a tumor suppressor gene \([13]\).

While previous biochemical and structural characterizations of MBD4 have provided important insights into the mismatch recognition mechanisms of the enzyme, its catalytic mechanism remains elusive. Crystal structures of the catalytic domain of hMBD4 bound to T:G mismatch-containing DNA show the thymine substrate opposite guanine is flipped out and enters the catalytic pocket of the glycosylase to engage in thymine-specific hydrogen bonds (Figure 1B) \([15–17]\). The Watson-Crick edge of the flipped thymine is recognized by Gln, Val, and Tyr residues in the catalytic site. An intercalating Arg residue (Arg468 in hMBD4 and Arg442 in mMBD4) facilitates base flipping and stabilizes the flipped-out thymine by phosphate pinching, which is similarly observed in thymine DNA glycosylase (TDG) \([18]\).

The catalytic Asp residue (Asp560 in hMBD4 and Asp534 in mMBD4) of MBD4 is critical for glycosylase activity \([14,15]\), but its role in MBD4 catalysis remains unclear. A structure of wild-type mMBD4 in complex with T:G mismatched DNA suggests that Asp534 directly attacks on an oxacarbenium ion-like intermediate to form a covalent bond between the Asp534 carboxylate oxygen and the AP-site C1´. However, the relatively low resolution (2.8 Å) of this crystal structure has prevented accurate appreciation of the catalytic mechanism. Although high-resolution structures of Asp560Ala hMBD4 in complex with T:G- and G:hmU-containing DNA are reported \([14]\), these structures have provided limited insights into the catalytic mechanism due to the incorporation of the non-conservative mutation.

To further our understanding of the catalytic mechanism of T:G-mismatch-specific DNA glycosylases, we have determined a crystal structure of Asp560Asn hMBD4 in complex with T:G mismatched DNA, which was refined to 1.6 Å resolution. In addition, to characterize the structure of the AP-site generated by hMBD4, we have solved a structure of the wild-type hMBD4 encountering T:G-mismatched DNA, which was refined to 2.0 Å resolution. These high-resolution hMBD4-T:G structures reveal detailed information.
regarding the conformations of protein, DNA, and the AP-site in the hMBD4 catalytic site, thereby providing new insights into the glycosylase mechanisms of hMBD4.

**MATERIALS AND METHODS**

**Oligonucleotides used for structural studies.**

The 12-mer oligonucleotide sequences used for crystallization were 5’-GCTGCGCGCTGG-3’ and 5’-CCACGTTGCAGC-3’, where G and T denote a T:G mismatch. Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA).

**Plasmid construction of human MBD4 glycosylase domain.**

The gene of human MBD4 glycosylase domain with residues 425–580 was amplified by PCR and inserted into pET28a vector. Using QuikChange II site-directed mutagenesis (Agilent), single mutant Asp560Asn hMBD4 was generated, which was verified by sequencing (ICMB Core Facility, The University of Texas at Austin).

**Protein expression and purification of hMBD4 glycosylase domain.**

The glycosylase domain of hMBD4 was expressed using pET28a vector. The protein was expressed in *E. coli* BL21 (DE3) cells. Cultures were grown in Luria-Bertani (LB) broth medium at 37 °C until the OD600 of 0.5, where cells were induced with 0.25 mM of isopropyl β-D-α-thiogalactopyranoside (IPTG) for four hours at 37 °C. Cultures for the hMBD4 glycosylase domain single mutant Asp560Asn hMBD4 were expressed in *E. coli* BL21 (DE3) cells, grown in LB medium at 37 °C until reaching the OD600 of 0.5, and induced by adding 0.5 mM IPTG at 28 °C for five hours. All the MBD4 glycosylase domain constructs followed the same purification steps. Pelleted cells were resuspended in buffer (50 mM sodium phosphate, pH 7.8, 500 mM NaCl, 10% glycerol, 1 mg/ml lysozyme, 0.25% NP-40, 0.25% Triton X-100, and 0.25 mM phenylmethylsulfonyl fluoride) and sonicated for 30 seconds. The lysate was centrifuged at 15000 g at 4 °C for 20 min. The MBD4 glycosylase domain protein supernatant was purified using Ni-NTA column (GE Healthcare). The imidazole-eluted fractions from the Ni-NTA column were combined and buffer-exchanged to low imidazole buffer, treated with thrombin (10 units), and incubated for 16 hours at 4 °C. This removed the 6-His tag, thereby leaving 19 extraneous N-terminal amino acids (GSHMASMTGGQQMGRGSEF). The cleaved proteins were further purified on Superdex 75 column (GE Healthcare) equilibrated with a buffer (50 mM Tris, pH 7.5, 150 mM NaCl, and 10% glycerol).

**Protein-DNA co-crystallization.**

A 12-mer crystallization DNA was annealed at a 1:1 ratio and mixed with the protein in a 1:5 molar ratio of protein to DNA in a buffer containing 5 mM Tris pH 8.0 and 25 mM NaCl. The DNA-protein mixture was incubated on ice for two to three hours. Co-crystals of Asp560Asn hMBD4-DNA complex were grown in 30% ethylene glycol. The hanging drop method was used for crystallography by mixing 1 μL of protein-DNA complex solution with 1 μL of reservoir solution. Crystals of the hMBD4-T:G DNA complex were grown at 22 °C over two weeks. The crystals were flash frozen in liquid N2 and used for data collection.
Data collection and structure determination.

Diffraction data sets for crystals of the hMBD4 complexes were collected at a wavelength of 0.97948 Å on beamline 5.0.3 at the Advanced Light Source (Berkeley, CA). Diffraction data for the hMBD4 complexes were indexed and scaled by using HKL2000. The hMBD4-DNA complex structures were solved by molecular replacement using a hMBD4-DNA complex (PDB code 4E9G) as the search model. Manual model building was carried out with COOT [19] and refined using CCP4 and REFMAC [20]. The quality and stereochemistry of the models were examined with PROCHECK. All the crystallographic figures were prepared by using PyMOL. Statistics of data processing, data quality, and the refined models are summarized in Table 1.

RESULTS AND DISCUSSION

Structure of Asp560Asn hMBD4 in complex with T:G mismatch-containing DNA.

There are two published crystal structures of T:G-mismatched DNA bound to MBD4: Asp560Ala hMBD4 and Asp534Asn mMBD4. The structure of Asp560Ala hMBD4 in complex with T:G mismatched DNA (PDB Code 4E9G) has provided critical insights into the mismatch recognition mechanism of the enzyme [14], which is described above. However, the use of a non-conservative Asp560Ala mutant in the structure has made it difficult to evaluate the thymine cleavage mechanism by hMBD4. The structure of Asp534Asn mMBD4 bound to T:G mismatched DNA (PDB Code: 4EVV) has revealed the extrahelical recognition of the thymine substrate as well as possible glycosylase mechanism of the enzyme [15]. Both structures do not show the presence of a water molecule in the catalytic site, which has been observed in other monofunctional DNA glycosylases [21–23]. In addition, the Asn534 residue of the Asp534Asn mMBD4-T:G structure is proximal to the C1’ of the flipped thymine, suggesting a direct nucleophilic attack by the catalytic Asp during the glycosidic bond cleavage. Nevertheless, the medium resolution of the structures (2.4 Å resolution for both Asp534Asn mMBD4 and Asp560Ala hMBD4 structures) has prevented the appreciation of the base excision mechanism of MBD4. To gain further insights into the glycosylase mechanism of MBD4, we have determined a high resolution (1.6 Å resolution) structure of Asp560Asn hMBD4 recognizing T:G-mismatched DNA, which was solved via molecular replacement. Refinement statistics of this structure are shown in Table 1.

The overall structure of the Asp560Asn MBD4-T:G complex is very similar to that of the published Asp560Ala hMBD4-T:G structure (RMSD = 0.444 Å) and Asp534Asn mMBD4-T:G structure (RMSD = 0.405 Å). As observed in the Asp560Ala hMBD4-T:G structure, the thymine substrate enters the open catalytic site cleft of the enzyme (Figure 2A). The Watson-Crick edge of the flipped thymine engages in hydrogen bond interactions with the side chain of Gln449 and Tyr540 and the backbone amide of Val448. The base of the substrate dT rotates around the C-N glycosidic bond relative to the conformation of the intrahelical dT in duplex DNA, which is reminiscent of the conformation of the flipped uracil and adenine in UDG and MutY structures [21,23], respectively; This base distortion induced by DNA glycosylase has been suggested to facilitate the glycosidic bond cleavage [24]. The space generated by the flipped-out thymine is filled by the Arg468 finger [18,25].
The backbone carbonyl of Arg468 forms two hydrogen bonds with the N1 and N2 of the estranged guanine, and the guanidinium moiety of Arg468 engages in “phosphate pinching” to stabilize the flipped-out state (Figures 2A and 2C).

Despite the structural similarity described above, the catalytic pocket of the Asp560Asn hMBD4-T:G complex is significantly different from that of the published Asp534Asn mMBD4-T:G structure (Figure 2B). In particular, an ordered water molecule, which is not observed in the Asp560Ala hMBD4-T:G and Asp534Asn mMBD4-T:G structures, is clearly present near the C1´ of the thymine substrate (3.7 Å) and is hydrogen bonded to the side chain of Asp560Asn, the backbone amide N-H of Lys562, and the O4´ of the thymine substrate (Figure 2C). The presence of the ordered water molecule between the C1´ of the thymine and Asp560Asn suggests the enzyme’s glycosylase mechanism may involve a catalytic water molecule that can be deprotonated by Asp560. A similar general base-catalyzed glycosylase mechanism has been proposed in several monofunctional DNA glycosylases including uracil-DNA glycosylase (UDG) and alkyladenine-DNA glycosylase (AAG) [21,22]. The Asp560Asn mutation causes a conformational shift of α9 and α10 helices relative to the positions in the Asp560Ala hMBD4 structure (Figure 2D). The Asp560Asn residue shifts ~1.7 Å away from the thymine substrate relative to the Asp560Ala, which enables the ordered water molecule to be positioned between Asp560Asn and the C1´ of the thymine (Figure 2E). An overlay of the Asp560Asn hMBD4 and Asp534Asn mMBD4 structures reveals the protein conformation in the catalytic site is essentially identical, whereas the upstream region of the thymine substrate deviates significantly (Figure 2F).

Taken together, the high resolution structure of the Asp560Asn hMBD4-T:G complex provides new insights into the thymine excision mechanism of MBD4. The observation of the ordered water molecule near the Asp560Asn residue and the C1´ of the flipped thymine suggests that MBD4 may cleave the thymine substrate using a nucleophilic water molecule activated by the catalytic Asp560. The drastic decrease in the excision activity for Asp534Asn mMBD4 and Asp560Ala hMBD4 is consistent with a potential role of Asp560 in deprotonating the nucleophilic water molecule [14,15].

**Structure of wild-type hMBD4 in complex with a T:G mismatch-containing DNA**

Cheng and colleagues have reported a crystal structure of the wild-type mMBD4 glycosylase domain processing T:G-mismatched DNA [15]. The structure of the wild-type mMBD4-T:G complex, which is refined to 2.8 Å, shows the cleavage of the thymine substrate. Interestingly, a continuous electron density between Asp534 of mMBD4 and the C16f the AP-site is observed and the density does not fit well with the AP-site containing the C1´-OH. This has raised the possibility that the structure may represent a stalled reaction intermediate possessing a covalent bond between the Asp534 carboxylate oxygen and the AP-site C1´. In agreement, DNA glycosylase-mediated formation of protein-DNA cross-links has been recently reported [26,27]. Nevertheless, the low resolution of the structure has hindered the characterization of the conformation of the AP-site as well as the nature of interactions between the AP-site and the catalytic Asp residue, thereby necessitating an MBD4-AP:G structure with a higher resolution. To gain further insights into the catalytic
mechanism of MBD4, we report a high resolution structure of the wild-type hMBD4 in complex with T:G-mismatched DNA. The structure was solved via molecular replacement and refined to 2.0 Å resolution (Table 1).

The hMBD4-T:G complex structure clearly shows the presence of the AP-site, which is generated by the excision of the thymine by MBD4 (Figure 3A). As previously observed in structures of DNA glycosylase bound to the AP-site- or AP analog-containing DNA [21,28,29], the AP-site in the wild-type hMBD4-T:G structure (hereafter hMBD4-AP:G structure) is in an extrahelical state and presented to the catalytic cleft of the enzyme. Arg468 fills the space created by the flipped-out AP-site and stabilizes the complex by making hydrogen bonding interactions with the orphaned guanine and the phosphate oxygens near the AP-site (Figure 3B).

Our hMBD4-AP:G structure reveals the strategy by which MBD4 protects the genotoxic AP-site from exposure to an aqueous environment. As similarly seen in the mMBD4-AP:G complex structure (PDB Code 4DK9), a 2F_o-F_c electron density map contoured at 1.5σ shows a continuous, strong electron density between the catalytic Asp carboxylate and the AP-site (Figure 3C). However, the electron density around the AP-site and Asp560 unambiguously indicates that the AP-site contains the C1´-OH group (Figure 3C). The Asp560 carboxylate oxygens are 3.2 Å and 4.2 Å away from the C1´ of the AP-site, which rules out the possibility of a covalent bond formation between Asp560 and the AP-site. The AP-site C1´-OH engages in bifurcated hydrogen bonds with the carboxylate oxygens of Asp560 and with an ordered water molecule, which in turn is buttressed by a hydrogen bond network created by Tyr540, Gln449, and a second ordered water molecule (Figure 3D). The extrahelical AP-site is further stabilized by Lys562, which makes polar contacts with the O4´ and O5´ of the AP-site.

Intriguingly, the AP-site generated by hMBD4 displayed an (S)-configuration at the C1´ position (Figure 3C), where the hydroxyl group points toward Asp560. The exclusive formation of the C1´-α anomer during MBD4-mediated base excision suggests the hMBD4-mediated glycosylase reaction may occur through an S_N2-like mechanism, where the activated water molecule attacks from the back side of the departing thymine (Figure 3C). The predominant formation of the α anomer could be promoted by the extensive hydrogen bond interactions between the AP-site and the MBD4 catalytic site.

A comparison of the Asp560Asn hMBD4-T:G and the hMBD4-AP:G structures reveals that the DNA and the catalytic residues of MBD4 undergo a significant conformational change after the thymine excision (Figure 3E). While the estranged G-containing strand does not make a significant conformational change, the AP-site and its 5´ nucleotide retract ~1.6 Å from the catalytic site. In response to the conformational change of the AP-site-containing strand, the guanidinium moiety of Arg468 rotates toward the orphaned G, Asp560 shifts toward the catalytic pocket, and the ε-NH_2 of Lys562 moves toward the AP-site-containing strand.

A superposition of our hMBD4-AP:G structure with the mMBD4-AP:G structure (PDB code: 4EW4) [15] shows that conformational differences primarily appear near the AP-site.
The RMSD for the hMBD4-AP:G and mMBD4-AP:G structures is 0.588 Å. Protein conformations of hMBD4 and mMBD4 are essentially indistinguishable, whereas the AP-site of the hMBD4 structure moves ~1.3 Å away from the catalytic Asp residue relative to the corresponding mMBD4 structure. In addition, the 5´ region of the AP-site slightly shifts away from the Arg finger (Arg442 in mMBD4) relative to the mMBD4 structure. Due to the difference in the resolution of the two structures, a direct comparison of the AP-sites would be difficult. A high resolution structure of the mMBD4-AP:G complex is needed to determine whether a covalent bond between the AP-site and Asp534 forms during thymine excision by mMBD4.

The overall structure of the hMBD4-AP:G complex is very similar to the published hMBD4-THF:G structure (PDB Code: 4DK9), where a tetrahydrofuran (THF) moiety is used as an AP-site analog [17] (Figure 4C and 4D). The conformations for Gln449, Arg468, Tyr540, Asp560, and Lys562 in both structures overlay very well, while the ordered water molecules are not observed in the THF:G structure, likely due to the relatively low resolution of the structure (2.8 Å). This indicates the presence/absence of the AP-site C1´-OH does not significantly affect the conformation of the catalytic site of MBD4.

Crystal structures of the Asp560Asn hMBD4-T:G and wild-type hMBD4-AP:G complexes provide new insights into the catalytic mechanism of mismatch-specific DNA glycosylase MBD4 (Figure 5). The observation of the ordered water molecule near Asp560Asn suggests that the water molecule is activated by Asp560. The extensive hydrogen bond interactions of the Watson-Crick edge of thymine with Val448, Gln449, and Tyr540 as well as the thymine base distortion would lower the activation energy for the C-N glycosidic bond cleavage. The nucleophilic water molecule activated by Asp560 would attack from the back side of the C-N glycosidic bond of the flipped dT, thereby generating the AP-site with the C1´-(S)-OH. The C1´-OH of the AP-site is protected by the extensive hydrogen bond network created by Asp560, Gln449, Tyr540, and two ordered water molecules.

While the catalytic Asp residue is found in several HhH family DNA glycosylases (e.g., AlkA, AAG, UDG, MutY, MBD4, MIG), their glycosylase mechanisms vary significantly. In a crystal structure of AlkA bound to a transition state analog [29], there is a lack of space to bind a nucleophilic water near the AP-site, raising the possibility of a direct attack of an Asp residue (Asp138 of AlkA) onto the oxacarbenium ion-like intermediate. Based on NMR studies and a crystal structure of MutY bound to a transition state analog [30], David and colleagues have proposed that Asp144 reacts with the oxacarbenium ion-like intermediate to furnish a transient covalent adduct with the AP-site, where the carboxylate oxygen of Asp144 is linked to the C1´ of the AP-site. The Asp144/AP-site adduct could undergo hydrolysis to generate the AP-site bearing the C1(βR)-OH, β anomer. A structure of wild-type hUDG bound to U:G-mismatched DNA shows the production of the AP-site with the C1´-(S)-OH moiety (Figure 6B)(PDB code 1SSP) [21], which has the same stereochemistry as the AP-site created by hMBD4. As similarly seen in the hMBD4-AP:G structure, the AP-site in the hUDG-AP:G structure participates in the extensive hydrogen bond network with the active side residues and two ordered water molecules. The inversion of the stereochemistry at the C1´ of the AP-site suggests that hUDG uses Asp145 to prime a water nucleophile.
The extensive hydrogen bonding interactions around the AP-sites of hMBD4 and hUDG would not only deter the release of the genotoxic AP-site to the aqueous environment but also prevent an aberrant break of the AP-site-containing strand by these enzymes (Figure 7). In our hMBD4-AP:G structure, the catalytic Asp560 is proximal (2.9 Å) to the C2´ of the AP-site. The presence of an Asp residue nearby the AP-site could be deleterious, as a negatively charged Asp residue could deprotonate the C2´-H of the ring-opened AP-site ($pK_a$ of the C2´-H of AP-site: 10.5 [31]) (Figure 7), potentially triggering β-elimination and a single strand break. This AP-site breakdown can be prevented by the bifurcated hydrogen bonding interactions between the AP-site C1´-OH and Asp560 (Figure 6A), which would neutralize the carboxylate. In the hUDG structure, the catalytic Asp145 is hydrogen bonded to the AP-site, an ordered water, and two amino acid residues, which would also neutralize the catalytic Asp and thus preclude a potential Asp-mediated strand break.

**Conclusions.**

The high-resolution structure of Asp560Asn hMBD4 bound to T:G-mismatched DNA reveals the presence of an ordered water molecule near the catalytic Asp560, suggesting that Asp560 activates a nucleophilic water molecule for the glycosidic bond cleavage. The crystal structure of the wild-type hMBD4 processing T:G-mismatched DNA shows the formation of the AP-site with an inversion of the stereochemistry at the C1´-OH. This indicates that the nucleophilic water molecule attacks from the back of the C1´ of the flipped dT, likely via $S_N$2-type bimolecular nucleophilic displacement. The AP-site generated by hMBD4 is greatly stabilized by several hydrogen bonds with the MBD4 active site residues, suggesting the role of MBD4 in preventing the release of the genotoxic.

**Acknowledgements:**

We are grateful to Dr. Arthur Monzingo for technical assistance. Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin.

**Funding:** This research was supported in part by the National Institutes of Health (ES 26676).

**References**

1. Arana ME and Kunkel TA (2010) Mutator phenotypes due to DNA replication infidelity. Semin. Cancer Biol. 20, 304–311. [PubMed: 20934516]
2. Kunkel TA (1985) The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. J. Biol. Chem, 260, 5787–5796. [PubMed: 3988773]
3. Bebenek K, Pedersen LC and Kunkel TA (2011) Replication infidelity via a mismatch with Watson-Crick geometry. Proc. Natl. Acad. Sci. U.S.A. 108, 1862–1867. [PubMed: 21233421]
4. Koag MC, Nam K and Lee S (2014) The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. Nucleic Acids Res. 42, 11233–11245. [PubMed: 25200079]
5. Cooper DN, Mort M, Stenson PD, Ball EV and Chuzhanova NA (2010) Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. Hum. Genomics, 4, 406–410. [PubMed: 20846930]
6. Duncan BK and Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. Nature, 287, 560–561. [PubMed: 6999365]

7. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA and Cairns BR (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell 135, 1201–1212. [PubMed: 19109892]

8. Cortellino S, Turner D, Masciullo V, Schepis F, Albino D, Daniel R, Skalka AM, Meropol NJ, Alberti C, Larue L, et al. (2003) The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. Proc. Natl. Acad. Sci. U.S.A. 100, 15071–15076. [PubMed: 14614141]

9. Wiebauer K and Jiricny J (1990) Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G·T mispairs in nuclear extracts from human cells. Proc. Natl. Acad. Sci. U.S.A. 87, 5842–5845. [PubMed: 2116008]

10. Sjolund AB, Senejani AG and Sweasy JB (2013) MBD4 and TDG: Multifaceted DNA glycosylases with ever expanding biological roles. Mutat. Res. 743–744, 12–25.

11. Jacobs AL and Schär P (2007) A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. DNA Repair 6, 695–711. [PubMed: 17337257]

12. Hendrich B, Hardeland U, Ng HH, Jiricny J and Bird A (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature, 401, 301–304. [PubMed: 10499592]

13. Moréra S, Grin I, Vigouroux A, Couvé S, Henriot V, Saparbaev M and Ishchenko AA (2012) Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. Nucleic Acids Res. 40, 9917–9926. [PubMed: 22848106]

15. Hashimoto H, Hong S, Bhagwat AS, Zhang X and Cheng X (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res. 40, 10203–10214. [PubMed: 22962365]

16. Hashimoto H, Zhang X and Cheng X (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. Nucleic Acids Res. 40, 8276–8284. [PubMed: 22740654]

17. Manvilla BA, Maiti A, Begley MC, Toth EA and Drohat AC (2012) Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. J. Mol. Biol. 420, 164–175. [PubMed: 22560993]

18. Maiti A, Morgan MT, Pozharshi E and Drohat AC (2008) Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. Proc. Natl. Acad. Sci. U.S.A. 105, 8890–8895. [PubMed: 18587051]

19. Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132. [PubMed: 15572765]

20. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242. [PubMed: 21460441]

21. Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE and Tainer JA (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J, 17, 5214–5226. [PubMed: 9724657]

22. Lau A (1998) Crystal Structure of a Human Alkylbase-DNA Repair Enzyme Complexed to DNAMechanisms for Nucleotide Flipping and Base Excision. Cell 95, 249–258. [PubMed: 9790531]

23. Lee S and Verdine GL (2009) Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. Proc. Natl. Acad. Sci. U.S.A. 106, 18497–18502. [PubMed: 19841264]
24. Mol CD, Arvai AS, Begley TJ, Cunningham RP and Tainer JA (2002) Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases. J. Mol. Biol. 315, 373–384. [PubMed: 11786018]

25. Maiti A, Michelson AZ, Armwood CJ, Lee JK and Drohat AC (2013) Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. J. Am. Chem. Soc, 135, 15813–15822. [PubMed: 24063363]

26. Ahn W-C, Aroli S, Kim J-H, Moon JH, Lee GS, Lee M-H, Sang PB, Oh B-H, Varshney U and Woo E-J (2019) Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. Nat. Chem. Biol, 15, 607–614. [PubMed: 31101917]

27. Tu J, Chen R, Yang Y, Cao W and Xie W (2019) Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. Nat. Chem. Biol, 15, 615–622. [PubMed: 31101915]

28. Lau AY, Wyatt MD, Glassner BJ, Samson LD and Ellenberger T (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. Proc. Natl. Acad. Sci. U.S.A. 97, 13573–13578. [PubMed: 11106395]

29. Hollis T, Ichikawa Y and Ellenberger T (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA. EMBO J. 19, 758–766. [PubMed: 10675345]

30. Woods RD, O’Shea VL, Chu A, Cao S, Richards JL, Horvath MP and David SS (2016) Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. Nucleic Acids Res. 44, 801–810. [PubMed: 26673696]

31. Küpfer PA and Leumann CJ (2007) The chemical stability of abasic RNA compared to abasic DNA. Nucleic Acids Res. 35, 58–68. [PubMed: 17151071]
Figure 1. Base excision repair of T:G mismatches.

(A) Formation and processing of T:G mismatches at CpG sites. DNMT: DNA methyltransferase; AID: activation induced cytidine deaminase; TDG: thymine DNA glycosylase; BER: base excision repair; TET2: Tet methylcytosine dioxygenase 2. (B) Extrahelical recognition of a T:G mismatch by hMBD4. The flipped-out thymidine is recognized by Tyr540, Gln449 and Val448 and stabilized by the Arg468-mediated phosphate pinching. The backbone carbonyl of Arg468 engages in hydrogen bonding interactions with the Watson-Crick edge of the estranged guanine.
Figure 2. Structure of Asp560Asn hMBD4 encountering a T:G-mismatched DNA.
(A) Overall structure of Asp560Asn hMBD4 in complex with T:G-mismatched DNA. (B) Close-up view of the active site of Asp560Asn hMBD4 in complex with T:G-mismatched DNA. A 2F_o−F_c electron density map contoured at 1σ around the thymine substrate and the ordered water molecule is shown. Hydrogen bonds are indicated as dashed lines. (C) The recognition of the estranged guanine and the stabilization of the flipped thymine by the Arg468 finger. Hydrogen bonding interactions of the ordered water molecule with the O4’ of dT, Asp560Asn, and Lys562 are shown. (D) Comparison of the active site of the Asp560Asn
hMBD4-T:G and Asp560Ala hMBD4-T:G (White, PDB code 4E9G) complexes. Note that Asp560Ala hMBD4-T:G structure does not show an ordered water in the catalytic site. (E) Superposition of the Asp560Asn hMBD4-T:G and Asp560Ala hMBD4-G·T (White, PDB code: 4E9G) complexes. Conformational changes of helices $\alpha_9$ and $\alpha_{10}$ are indicated. (F) Superposition of the Asp560Asn hMBD4-T:G and Asp534Asn mMBD4-T:G structures. Distances are indicated as double-headed arrows. Note that the Asp534Asn mMBD4-T:G structure does not show an ordered water in the catalytic site.
Figure 3. Structure of the wild-type hMBD4 in complex with AP:G-containing DNA
(A) Overall structure of hMBD4 bound to the AP-site-containing DNA. The AP-site is colored in magenta and the G opposite the thymine substrate is colored in black. (B) A close-up view of the active site that shows the AP-site generated by hMBD4-mediated excision. (C) A close-up view of the AP-site from a different perspective. A 2F_{o}-F_{c} electron density map contoured at 1.5σ around the AP-site, Asp560, and water is shown. (D) Hydrogen bond interactions around the AP-site. (E) Superposition of the hMBD4-AP:G
structure and Asp560Asn hMBD4-T:G structure. Shifts in amino acid residues and DNA are indicated as arrows.
Figure 4. Comparison of the hMBD4-AP:G structure with published structures of MBD4 bound to AP-site- or AP-site analog-containing DNA.
(A) Superposition of the hMBD4-AP:G (present work) and mMBD4-AP:G structures (PDB code: 4EW4). (B) Active site of mMBD4-AP:G structure (PDB code: 4EW4). (C) Superposition of the hMBD4-AP:G (present work) and hMBD4-THF:G (PDB code: 4DK9) structures. (D) Active site of the hMBD4-THF:G structure (PDB code: 4DK9).
Figure 5. Proposed glycosylase mechanism of hMBD4.
The potential water nucleophile near the flipped-out thymine is stabilized by hydrogen bonding interactions with the backbone amide of Lys562 and the O4' of the substrate thymine. Asp560 deprotonates the ordered water molecule to produce a hydroxide ion, which in turn cleaves the C-N glycosidic bond of the substrate thymine via $S_N$2-type bimolecular nucleophilic displacement. The degradation and isomerization of the resulting abasic site is prevented by the hydrogen-bonding network created by Asp560, two ordered water molecules, Tyr540, and Gln449.
Figure 6. Interactions of the AP-site in the catalytic site of DNA glycosylase bearing a catalytic carboxylate.
(A) Structure of hMBD4 in complex with AP:G-containing DNA (present work). (B) Structure of hUDG in complex with AP:G-containing DNA (PDB code: 1SSP).
Figure 7.
Possible formation of a single strand break via Aspartate-catalyzed β-elimination of the AP-site.
Table 1.

Data collection and refinement statistics.

| PDB code          | D560N MBD4-T:G (4OFA) | MBD4-AP:G (6VJW) |
|-------------------|-----------------------|------------------|
| **Data Collection** |                       |                  |
| Space Group       | P2_12_1             | P2_12_1          |
| Cell constants    |                       |                  |
| a (Å)             | 41.561               | 40.350           |
| b                 | 55.534               | 58.433           |
| c                 | 104.212              | 101.142          |
| a (°)             | 90.00                | 90.00            |
| β                 | 90.00                | 90.00            |
| γ                 | 90.00                | 90.00            |
| Resolution (Å)^d  | 29.45–1.55 (1.61–1.55) | 50–2.01 (2.04–2.01) |
| R_merg b (%)      | 0.048 (0.589)        | 0.031 (0.524)    |
| <I/σ>             | 34.3 (2.88)          | 27.9 (1.2)       |
| CC_{1/2}          | (0.740)              | (0.728)          |
| Completeness (%)  | 99.6 (98.4)          | 99.5 (97.7)      |
| Redundancy        | 6.9 (6.2)            | 9.8 (9.0)        |
| **Refinement**    |                       |                  |
| R_w (%)+R_free (%)| 13.4/18.2            | 21.9/26.3        |
| Unique reflections | 35548                | 15588            |
| Mean B factor (Å²) |                      |                  |
| Protein           | 31.9                 | 48.3             |
| Ligand            | 22.8                 | 57.2             |
| Solvent           | 43.2                 | 52.3             |
| Ramachandran plot |                      |                  |
| Most favored (%)  | 98.6                 | 98.5             |
| Additional allowed (%) | 1.4 | 0.7 |
| RMSD              |                      |                  |
| Bond lengths (Å)  | 0.017                | 0.010            |
| Bond angles (°)   | 1.67                 | 1.11             |

^aValues in parentheses are for the highest resolution shell.

^bR_merg = Σ|I-<I>|/ΣI where I is the integrated intensity of a given reflection.

^cR_w = Σ|F(obs)-F(calc)|/ΣF(obs).

^dR_free = Σ|F(obs)-F(calc)|/ΣF(obs), calculated using 5% of the data.