Insights into the substrate discrimination mechanisms of methyl-CpG-binding domain 4

Hala Ouzon-Shubeita, Lillian F. Schmaltz, Seongmin Lee*

The Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, USA

Abstract

G:T mismatches, the major mismatches generated during DNA metabolism, are repaired in part by mismatch-specific DNA glycosylases such as methyl-CpG-binding domain 4 (MBD4) and thymine DNA glycosylase (TDG). Mismatch-specific DNA glycosylases must discriminate the mismatches against million-fold excess correct base pairs. MBD4 efficiently removes thymine opposite guanine but not opposite adenine. Previous studies have revealed that the substrate thymine is flipped out and enters the catalytic site of the enzyme, while the estranged guanine is stabilized by Arg468 of MBD4. To gain further insights into mismatch discrimination mechanism of MBD4, we assessed the glycosylase activity of MBD4 toward various base pairs. In addition, we determined a crystal structure of MBD4 bound to T:O6-methylguanine-containing DNA, which suggests the O6 and N2 of purine and the O4 of pyrimidine are required to be a substrate for MBD4. To understand the role of the Arg468 finger in catalysis, we evaluated the glycosylase activity of MBD4 mutants, which revealed the guanidinium moiety of Arg468 may play an important role in catalysis. D560N/R468K MBD4 bound to T:G mismatched DNA shows that the side chain amine moiety of the Lys stabilizes the flipped-out thymine by a water-mediated phosphate pinching, while the backbone carbonyl oxygen of the Lys engages in hydrogen bonds with N2 of the estranged guanine. Comparison of various DNA glycosylase structures implies the guanidinium and amine moieties of Arg and Lys, respectively, may involve in discriminating between substrate mismatches and nonsubstrate base pairs.

INTRODUCTION

Guanine:thymine mismatches are the most abundant mismatches found in genomic DNA (Figure 1) (1–3). These mismatches are produced by various pathways. Replication errors of DNA polymerases via tautomerization preferentially generate G:T and A:C mismatches (2,4–8). The exocyclic amine of 5-methylcytosine found in the CpG repeats is susceptible to hydrolysis, undergoing spontaneous deamination of 5-methylcytosine to give G:T mismatches (9,10). In addition, G:T mismatches at the CpG islands can be produced by the deamination of 5-methylcytosine by activation-induced deaminase (AID), which initiates active DNA demethylation process (Figure 1A) (11). The G:T mismatches, if left undetected prior to replication, can promote C to T transition mutations. The promutagenic G:T
mispairs can be repaired by DNA mismatch repair (MMR) pathway (12,13) as well as base excision repair pathway that involves mismatch-specific DNA glycosylases such as thymine DNA glycosylase (TDG) and methyl-CpG-binding domain 4 (MBD4/MED1) (14–16). Both TDG and MBD4 can cleave thymine paired with guanine to produce abasic sites, which are further processed by downstream base excision DNA repair enzymes including AP endonuclease, DNA polymerase, and DNA ligase.

MBD4 is a monofunctional DNA glycosylase and a member of the superfamily of helix-hairpin-helix (HhH) DNA glycosylases including E. coli EndoIII, 8-oxoguanine DNA glycosylase 1 (hOGG1), A/G-mismatch-specific DNA glycosylase MutY, and alkyladenine DNA glycosylase (AlkA) (17). MBD4 contains N-terminal methyl-CpG-binding domain (MBD) and C-terminal glycosylase domain (15). *In vitro* studies have shown MBD4 removes a wide variety of pyrimidine lesions including uracil, 5-fluorouracil, ethenocytosine, thymine glycol, and hydroxymethyluracil opposite guanine (16,18,19). In addition, MBD4 has been shown to cleave thymine opposite O6-methylguanine (16).

Biochemical and structural characterization of MBD4 has provided important insights into the catalytic and substrate recognition mechanisms of the enzyme. Crystal structures of the catalytic domain of MBD4 bound to G:T mismatch-containing DNA show that the substrate thymine opposite guanine is flipped out and enters the catalytic site of the glycosylase to engage in thymine-specific contacts (Figure 1B) (20–22). An intercalating Arg residue (Arg468 in hMBD4 and Arg442 in mMBD4) facilitates base flipping and stabilizes the flipped-out thymine by phosphate pinching (Figure 1C). The presence of Asp residue (D560 in hMBD4 and D534 in mMBD4) within the catalytic pocket of the mismatch-specific DNA glycosylase is critical for glycosylase activity. Our recent studies show MBD4 uses Asp-assisted general base catalysis mechanism to generate abasic sites bearing the C1´-(S)-OH (22), indicating bimolecular displacement mechanism for the thymine excision (Figure 1B). The resulting abasic sites are stabilized by a hydrogen bond network, highlighting the role of MBD4 in preventing the aberrant release of the toxic abasic site.

O6-Methylguanine (O6MeG) is a highly mutagenic lesion that preferentially pairs with thymine during translesion DNA synthesis, promoting G to A transversions (23–27). Endogenous (e.g., S-adenosylmethionine) and exogenous (e.g., N-methyl-N-nitrosourea) alkylating agents attack DNA to form O6MeG as a minor adduct. The formation of the deleterious O6MeG adduct is also induced by an anticancer methylating drug temozolomide and contributes to the cytotoxicity of methylating anticancer agents (28). The genotoxic O6MeG is detoxified by methylguanine methyltransferase (MGMT), a sacrificial protein that removes the O6-methyl group by covalently modifying the catalytic cysteine residue of the protein. O6MeG:T mismatch can trigger a futile cycling of mismatch repair (MMR) (29), which can cause cell cycle arrest and cell death. *In vitro* studies show that O6MeG:T is a substrate mismatch for MBD4 and TDG (30). The base excision repair enzyme MBD4 is known to associate with mismatch repair enzymes (e.g., MLH1) and mediate alkylation damage response to methylating agents and other DNA damaging agents (16,31).

How does MBD4 recognize thymine opposite guanine or O6MeG in the presence of millions-fold excess of non-substrate base pairs? Answering this question would further our
understanding of the catalytic and substrate base pair recognition mechanisms of G:T-mismatch-specific DNA glycosylases. To this end, we conducted biochemical characterization of wild-type or mutant hMBD4 encountering mismatch-containing DNA. In addition, we report two crystal structures of human MBD4 in complex with G:T- or O6MeG:T-containing DNA, which provide new insight into the substrate recognition and glycosylase mechanisms of MBD4.

Materials and Methods

Oligonucleotides used for biochemical and structural studies.

Modified oligonucleotides were purchased from Midland Certified Reagents (Midland, TX). Oligonucleotides labeled with fluorescein amidite (FAM) were purchased from Integrated DNA Technologies (Coralville, IA). The 27-mer DNA oligonucleotide sequence used for DNA glycosylase activity assay was 5’-TCAGATCGCGCCXXGCTGCGATAAGCT-3’, where X stands for G, O6-methylguanine, hypoxanthine, 2,6-aminopurine, or 2-aminopurine. The sequence for 27-mer complementary strand containing the T mismatch in a CpG region to the above 27-mer oligonucleotides was 5’-AGCTTATCGCAGCTGGCGCGAATCTGA-FAM-3’. The 12-mer oligonucleotide sequence used for crystallization was 5’-GCTGCYCGCTGG-3’, where Y stands for G or O6MeG.

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 site-directed mutagenesis, single mutants D560N, R468K, R468Q, R468E, R468F, R468Y, and R468L and double mutant R468K/D560N, were generated. All mutant plasmids were verified by sequencing (ICMB Core Facility, University of Texas at Austin).

Protein expression and purification of MBD4 glycosylase domain.

MBD4 glycosylase domain and its mutants R468K, R468Q, R468E, R468F, R468Y, and R468L were expressed using pET28a vector (21,22). The proteins were 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 MBD4 glycosylase domain single mutant D560N and double mutant R468K/D560N were also expressed in E. coli BL21 (DE3) cells, grown in LB medium at 37 °C until 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.
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). Melting temperatures ($T_m$) of wild-type MBD4 and the MBD4 mutants were determined by UV-VIS spectroscopy to ensure proper folding of expressed proteins.

**DNA glycosylase activity assay.**

The glycosylase activities of wild-type MBD4 on various base pairs (G:T, O6MeG:T, HX:T, 2,6-AP:T, 2-AP:T) were performed using FAM-labeled oligonucleotides (see oligonucleotides above; Figure 2). The glycosylase activities of MBD4 Arg468 mutants were performed using 27-mer, 3’-FAM-labeled oligonucleotide duplex containing a G:T mismatch in a CpG region (5’-TCAGATCGCGCGCGGC TCGATAAGCT-3’ and 5’-AGCTTATCGCAGCTGGCGCGAATCTGA-3’) (see Figure 5). The standard reaction mixture (20 μl) contained 50 nM of labeled double stranded oligonucleotide and 2.5 μM of purified enzyme in 20 mM HEPES pH 7.5, 50 mM KCl, 1 mM DTT, 1 mM EDTA, and 0.1 mg/ml bovine serum albumin. The mixture was incubated at 37°C for 30 min, and the reaction was stopped by adding 4 μl of 0.5 N NaOH followed by boiling for 10 min to cleave DNA at the abasic site. After adding loading dye (5 μl, 98% formamide, 1 mM EDTA, 1 mg/ml of Bromophenol Blue, and 1 mg/ml Xylene Cyanole), the reaction sample was boiled for another 10 min and loaded immediately onto a 10 cm x 10 cm 20% denaturing urea gel. The intensities of the FAM-labeled DNA were measured by using Typhoon FLA9500 (GE Healthcare).

**Single turnover kinetics of wild-type and R468K MBD4 catalytic domain.**

Glycosylase activity assays of wild-type and R468K MBD4 were conducted using FAM-labeled 27-mer duplexes containing T:G or O6MeG:T mismatches at the central position. The mismatch-containing 27-mer duplexes (0.25 μM) and 10-fold excess of wild-type or R468K mutant MBD4 catalytic domain (2.5 μM) were mixed in reaction buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.1% BSA) and incubated for 0–30 min at room temperature. The glycosylase reactions were stopped at designated timepoints by adding 2 μL 1N NaOH and the resulting solutions were boiled for 10 min to cleave DNA at abasic sites. Loading buffer (20 μL, 98% formamide, 1 mM EDTA, 1 mg/mL Bromophenol Blue) was added to the mixture and the samples were boiled for 10 min. The resulting solutions were immediately cooled and loaded onto a 10 cm x 10 cm 20% denaturing urea PAGE gel. The intensities of FAM-labeled DNA were measured via Typhoon FLA9500 (GE Healthcare) and quantified using ImageJ (NIH). GraphPad Prism 9 (Graphpad Software Inc.) was used to fit the data by non-linear regression: $Y = (Y_0 - \text{Plateau}) \exp(−kt) + \text{Plateau}$, where $Y$ is the percent nicked yield, $k$ is the rate constant (min$^{-1}$), and Plateau is the Y value at infinite times. The kinetic values were derived from independent experiments and reported as the mean value ± standard mean error based on three independent replicates.

**DNA binding assay.**

DNA binding assays were performed by incubating D560N MBD4 with increasing concentrations (12–750 nM) and 50 nM of 3’-FAM-labeled G-T-containing DNA or O6MeG-T-containing DNA in 20 mM HEPES pH 7.5, 50 mM KCl, 1 mM DTT, 1 mM DTT, 1 mM
EDTA, and 0.1 mg/ml bovine serum albumin. The oligonucleotide sequence used in this assay was identical to the 27-mer used in the glycosylase activity assay. After incubating the mixture at room temperature for 30 min, loading buffer (4 μl, 60% TBE 0.25 and 40% glycerol) was added to the reaction mixture. The resulting mixture was loaded onto a 10 cm x 10 cm 10% native acrylamide gel and electrophoresed in 0.5 M TBE for one hour.

**Protein-DNA co-crystallization.**

A 12-mer crystallization DNA was annealed at a 1:1 ratio and mixed with the protein (1.5 mg/mL) 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 D560N MBD4-DNA complex were grown in 30% ethylene glycol. The crystals of R468K/D560N MBD4-DNA complex were obtained in 15% PEG1500, 0.1 M MES pH 6.5 and 0.2 M sodium acetate. D560N MBD4-O6MG-T and wild-type MBD4-nicked DNA complexes were crystallized in 21% 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 MBD4 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 MBD4 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 MBD4 complexes were indexed and scaled by using HKL2000. The MBD4-DNA complex structures were solved by molecular replacement using a MBD4-DNA complex (Protein Data Bank accession code 4E9G) as the search model. Manual model building was carried out with COOT (32) and refined using CCP4 (33) and REFMAC (34). The quality and stereochemistry of the models were examined with PROCHECK. All the crystallographic figures were prepared using PyMOL. Statistics of data processing, data quality, and the refined models are summarized in Table 1.

**RESULTS AND DISCUSSION**

**Characterization of substrate base pairs for MBD4.**

Previous studies have shown that MBD4 efficiently recognizes and excises T:G mismatches (15). In addition, the enzyme has been also shown to cleave thymine opposite the highly mutagenic O6-methylguanine (O6MeG) lesion (16). How does MBD4 cleave thymine opposite guanine but not thymine opposite adenine? To gain further insights into the substrate discrimination mechanism of MBD4, we assessed the glycosylase activity of MBD4 toward various base pairs including G:T, O6MeG:T, hypoxanthine (HX):T, 2,6-diaminopurine (2,6-AP):T, and 2-aminopurine (2-AP):T (Figures 2A and 2B). As observed previously (16), thymine opposite guanine was efficiently cleaved by MBD4, resulting in the formation of a fast-moving band after treatment of the reaction mixture with sodium hydroxide (Figure 2C). The glycosylase activity of wild-type MBD4 on G:T and O6MeG:T under single turnover conditions showed that rate constants ($k_{obs}$) for excision of thymine opposite O6MeG and G were 0.24 and 0.09 min$^{-1}$, respectively (Figure 3) (16), indicating that MBD4 processes G:T mismatches more efficiently than O6MeG:T mismatches. While
G:T preferentially forms a wobble base pair, O6MeG:T can form a base pair with pseudo-Watson-Crick geometry \(25,27\), which could contribute to MBD4’s reduced glycosylase activity toward O6MeG:T. Thymine opposite hypoxanthine, which forms a wobble base pair in duplex DNA (Figure 2A), was not processed by the enzyme, indicating that the hydrogen bond between the N2 of guanine and the backbone carbonyl oxygen of Arg468 is critical for substrate recognition and catalysis. When thymine is paired with 2,6-diaminopurine or 2-aminopurine, it was not processed by MBD4 (Figure 2C), suggesting base pairs that form an ideal Watson-Crick geometry are not substrates of MBD4. Taken together, these results signify that MBD4 recognizes and processes non-Watson-Crick purine:thymine base pairs that contains N2 of purine.

Structural basis for inefficient excision of O6MeG:T mismatch by MBD4.

How does MBD4 recognize and excise thymine opposite O6MeG? Why is O6MeG:T base pair less efficiently removed by MBD4 than G:T base pair? To gain structural insight into these questions, we determined a cocrystal structure of D560N MBD4 bound to O6MeG:T-containing DNA, which was refined to 2.2 Å resolution. The overall structure of the D560N MBD4-O6MeG:T complex is very similar to that of the published D560N MBD4-G:T complex structure (PDB code: 4OFA, RMSD = 0.244 Å for all heavy atoms of the MBD4-DNA complexes)\(22\). The estranged O6MeG remains intrahelical, while the substrate thymine is flipped out and fully inserted into the catalytic pocket of the enzyme (Figure 4). The flipped-out state is stabilized by Arg468, which approaches from the minor groove and intercalates between the bases 5’ and 3’ to the substrate thymine. The backbone carbonyl oxygen of Arg468 engages in hydrogen bond interaction with the N2 of the orphaned O6MeG. In addition, the guanidinium moiety of Arg468 forms polar interactions with the phosphate oxygens near the substrate thymine, stabilizing the extrahelical conformation of the thymine.

The Watson-Crick edge of the substrate thymine forms hydrogen bonds with Gln449, Tyr504, and Val448 (Figure 5), as similarly observed in the G:T mismatch structure \(22\). Unlike in the D560N MBD4-G:T structure, the carbonyl oxygen of Arg468 in the D560N MBD4-O6MeG:T structure is not hydrogen bonded to the N1 of the estranged purine. To evaluate whether the reduction in glycosylase activity is caused by MBD4’s weaker binding toward O6MeG:T-containing DNA, we compared the enzyme’s binding to G:T- and O6MeG:T-containing DNA using a non-denaturing polyacrylamide gel electrophoretic mobility shift assay (Figure 5C). The assay shows that the enzyme more tightly binds G:T-containing DNA than O6MeG:T-containing DNA, suggesting that the weaker binding of MBD4 towards O6MeG:T-containing DNA could contribute to the reduced glycosylase activity.

The role of Arg468 in MBD4-mediated excision of thymine opposite guanine

Structural studies of MBD4 have shown that, during the substrate base flipping, Arg468 penetrates DNA from the minor groove, is inserted into the void created by the flipped-out thymine, and forms hydrogen bonds with the two phosphate oxygens near the flipped thymine \(19,22\). These interactions can stabilize the flipped-out thymine in a productive state for catalysis. One question is whether Arg468 is essential for the catalytic activity of
MBD4. As DNA glycosylases use various amino acid residues (e.g., glutamine (MutY), leucine (AlkA), tyrosine (AAG), arginine (TDG, MBD4, MUG), Asn (OGG1)) as an intercalating moiety (17,35), we wanted to evaluate whether the guanidinium moiety of Arg468 plays a role in MBD4-mediated excision of substrate thymine. Such knowledge would provide insights into the role of the arginine finger in glycosylase reaction.

To assess the role of Arg468 in substrate excision, we evaluated the effect of Arg468 mutation on the glycosylase activity of MBD4. We mutated Arg468 into lysine, glutamine, phenylalanine, tyrosine, leucine, and glutamate and assessed the catalytic activity of the MBD4 mutants (Figure 6). Single turnover kinetic studies showed the R468K mutant protein was as efficient as wild type MBD4 in excising G:T mismatches (Figure 6B and Figure 3B). The R468Q MBD4 mutant also excised thymine opposite guanine, albeit to a much lesser extent (Figure 6B). In the case of R468F, Tyr, Leu, and Glu mutants, the thymine cleavage opposite guanine was not observed, revealing the identity of the intercalating moiety plays an important role in MBD4-mediated glycosylase reaction. Since the stabilization of a flipped-out substrate can be achieved by various intercalating amino acid residues, the observed loss of the excision activity is less likely to result from the lack of the phosphate pinching; a direct (e.g., Gln48 in MutY (36)) or water-mediated phosphate pinching (e.g., Tyr162 in hAAG (37), Leu125 in E. coli AlkA (38), Asn149 in hOGG1 (39)) are used in DNA glycosylases. The weak or lack of glycosylase activity of the glutamine, phenylalanine, tyrosine, leucine and glutamate mutants together with the strong activity of the lysine mutant suggests that the positively-charged side chain (NH$_3^+$ of lysine and guanidinium moiety of arginine) may play an important role in the catalysis of MBD4.

Structure of R468K/D560N MBD4 in complex with G:T-mismatched DNA

To provide the structural basis for the catalytic activity of R468K MBD4, we determined a crystal structure of R468K/D560N MBD4 mutant in complex with G:T mismatched DNA (Figure 7). The R468K/D560N MBD4 structure (PDB ID 4OFE), which was refined to 2.2 Å resolution, explains how the R468K mutant efficiently cleaves thymine opposite guanine. The R468K is similarly positioned as Arg468 in the D560N MBD4-G:T structure (Figure 5A-B) and the D560N MBD4-O6MeG:T structure (Supplementary Figure 1). Unlike Arg468, which directly hydrogen bonds to the 3' phosphate oxygen of substrate thymine nucleotide, R468K recruits a water molecule to engage in hydrogen bonding with the phosphate oxygen (Figure 7A). Unlike Arg468, R468K does not hydrogen bond with the 5' phosphate oxygen of the nucleotide adjacent to the flipped thymine, indicating that the hydrogen bond interaction with the 5' phosphate oxygen is not required for catalysis (Figure 7B). The lack of a hydrogen bond between R468K and the 5'-phosphate oxygen leads to a relatively relaxed phosphate backbone conformation. The relaxed conformation, however, does not affect the recognition of the flipped-out thymine by the active site residues. The catalytic water molecule is hydrogen bonded to D560N and proximal to the C1' of the substrate thymine.

Potential role of arginine finger in recognition of G:T mismatches.—Most DNA glycosylases use intercalating amino acid residues to stabilize the flipped substrate bases and facilitate the substrate excision. DNA glycosylases employ a wide variety of amino acid
residues—such as Gln (e.g., Gln28 of Bst. MutY; Supplementary Figure 2), Leu (e.g., Leu125 of E. coli. AlkA), Tyr (e.g., Tyr162 of hAAG), Asn (e.g., Asn149 of hOGG1), and Arg (e.g., Arg275 of hTDG)—in phosphate backbone pinching, thereby stabilizing the base-flipped conformation (Figure 8). The phosphate pinching involves either direct or water-mediated hydrogen bonds to the 5’ and/or 3’ to the flipped base.

Unlike lesion-specific DNA glycosylases that carry a wide variety of intercalating amino acid residues, U/T:G-mismatch-specific DNA glycosylases possess an arginine finger [e.g., MIG (PDB 1kea (40)), MUG (PDB 1MWI (41)), TDG (PDB 3UFJ (42))], where arginine is at the tip of an intercalating motif. The mutation of Arg275 finger of hTDG into alanine or leucine residue significantly (8-fold and 30-fold, respectively) reduces the catalytic activity (43), highlighting a role of the arginine finger in G:T-mismatch-specific glycosylase activity. As the phosphate pinching-mediated stabilization of a flipped base can be conducted by a wide variety of amino residues [e.g., Leu (nonpolar), Tyr (bulky), Gln/Asn (polar)], the requirement of the Arg finger in G:T-mismatch-specific DNA glycosylases suggests additional roles of the Arg finger in glycosylase mechanism. Fluorescence base flipping assay indicates that the guanidinium moiety of Arg275 in human TDG promotes base flipping (43,44). In particular, R275A TDG binds G:U-containing DNA with only slightly-reduced affinity relative to wild-type hTDG but it does not induce base flipping, suggesting that the guanidinium moiety participates in intrahelical-sensing and base-flipping of G:T mismatches. While other mutant forms of hMBD4 (e.g., Glu, Tyr, Phe, Asn, Leu) are devoid of glycosylase activity, R468K MBD4 mutant displays a comparable excision efficacy to wild-type MBD4 (Figure 3), suggesting that G:T mismatch recognition and base-flipping promotion by R468K residue would be similar to that by R468 of MBD4. The e-amino group and the guanidinium moiety of Lys and Arg fingers, respectively, might interact with O6 of dG and O4 of dT while dG and dT are intrahelical, thereby facilitating mismatch recognition and base flipping. Our results imply that, unlike lesion-specific glycosylases, G:T mismatch-specific DNA glycosylases rely on an intercalating residue’s ability to sense and/or disrupt intrahelical mismatches and induce base flipping.

Data Availability Statement
The atomic coordinates of MBD4-DNA complexes have been deposited in the Protein Data Bank with the following accession codes: D560N/R468K MBD4-G:T (PDB Code: 4OFE) and D560N MBD4-O6MeG:T (PDB Code: 4OFH).

Supplementary Material
Refer to Web version on PubMed Central for supplementary material.

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:
The work 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. Watt DL, Buckland RJ, Lujan SA, Kunkel TA, and Chabes A. (2016) Genome-wide analysis of the specificity and mechanisms of replication infidelity driven by imbalanced dNTP pools. Nucleic Acids Res 44, 1669–1680 [PubMed: 26609135]
4. Koag MC, Nam K, and Lee S. (2014) The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase beta. Nucleic Acids Res 42, 11233–11245
5. Wang W, Hellinga HW, and Beese LS (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc Natl Acad Sci U S A 108, 17644–17648
6. Harris VH, Smith CL, Jonathan Cummins W, Hamilton AL, Adams H, Dickman M, Hornby DP, and Williams DM (2003) The effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis. J Mol Biol 326, 1389–1401 [PubMed: 12595252]
7. Afek A, Shi H, Rangadurai A, Sahay H, Senitzki A, Xhani S, Fang M, Salinas R, Mielko Z, Pufall MA, Poon GMK, Haran TE, Schumacher MA, Al-Hashimi HM, and Gordan R. (2020) DNA mismatches reveal conformational penalties in protein-DNA recognition. Nature 587, 291–296 [PubMed: 33087930]
8. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, and Al-Hashimi HM (2015) Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. Nature 519, 315–320 [PubMed: 25762137]
9. 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]
10. Duncan BK, and Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. Nature 287, 560–561 [PubMed: 6999365]
11. 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]
12. Iyer RR, Pluciennik A, Burdett V, and Modrich PL (2006) DNA mismatch repair: functions and mechanisms. Chem Rev 106, 302–323 [PubMed: 16464007]
13. Li GM (2008) Mechanisms and functions of DNA mismatch repair. Cell Res 18, 85–98 [PubMed: 18157157]
14. 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
15. 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]
16. Cortellino S, Turner D, Masciullo V, Schepis F, Albino D, Daniel R, Skalka AM, Meropol NJ, Alberti C, Larue L, and Bellacosa A. (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
17. Jacobs AL, and Schar P. (2012) DNA glycosylases: in DNA repair and beyond. Chromosoma 121, 1–20 [PubMed: 22048164]
18. Yoon JH, Iwai S, O’Connor TR, and Pfeifer GP (2003) Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. Nucleic Acids Res 31, 5399–5404 [PubMed: 12954776]
19. 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]

20. Morera S, Grin I, Viguouroux A, Couve 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]

21. Wu P, Qiu C, Sohail A, Zhang X, Bhagwat AS, and Cheng X. (2003) Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4, J Biol Chem 278, 5285–5291 [PubMed: 12456671]

22. Ouzon-Shubeita H, Jung H, Lee MH, Koag MC, and Lee S. (2020) Catalytic mechanism of the mismatch-specific DNA glycosylase methyl-CpG-binding domain 4. Biochem J 477, 1601–1612 [PubMed: 32297632]

23. Eadie JS, Conrad M, Toorchen D, and Topal MD (1984) Mechanism of mutagenesis by O6-methylguanine. Nature 308, 201–203 [PubMed: 6322009]

24. Margison GP, Santibanez Koref MF, and Povey AC (2002) Mechanisms of carcinogenicity/chemotherapy by O6-methylguanine. Mutagenesis 17, 483–487 [PubMed: 12435845]

25. Koag MC, and Lee S. (2014) Metal-dependent conformational activation explains highly promutagenic replication across O6-methylguanine by human DNA polymerase beta. J Am Chem Soc 136, 5709–5721 [PubMed: 24694247]

26. Pence MG, Choi JY, Egli M, and Guengerich FP (2010) Structural basis for proficient incorporation of dTTP opposite O6-methylguanine by human DNA polymerase iota. J Biol Chem 285, 40666–40672

27. Warren JJ, Forsberg LJ, and Beese LS (2006) The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. Proc Natl Acad Sci U S A 103, 19701–19706 [PubMed: 17179038]

28. Fan CH, Liu WL, Cao H, Wen C, Chen L, and Jiang G. (2013) O6-methylguanine DNA methyltransferase as a promising target for the treatment of temozolomide-resistant gliomas. Cell Death Dis 4, e876

29. Rye PT, Delaney JC, Netiroijjanakul C, Sun DX, Liu IZ, and Essignmann JM (2008) Mismatch repair proteins collaborate with methyltransferases in the repair of O(6)-methylguanine. DNA Repair (Amst) 7, 170–176 [PubMed: 17951114]

30. 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]

31. Bellacosa A, Cicchillitti L, Schepis F, Riccio A, Yeung AT, Matsumoto Y, Golemis EA, Genuardi M, and Neri G. (1999) MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc Natl Acad Sci U S A 96, 3969–3974 [PubMed: 10097147]

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

33. Collaborative Computational Project N. (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760–763 [PubMed: 15299374]

34. Winn MD, Murshudov GN, and Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 374, 300–321 [PubMed: 14696379]

35. Brooks SC, Adhikary S, Rubinson EH, and Eichman BF (2013) Recent advances in the structural mechanisms of DNA glycosylases. Biochim Biophys Acta 1834, 247–271 [PubMed: 23076011]

36. 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

37. 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
38. 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]

39. Bruner SD, Norman DP, and Verdine GL (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. Nature 403, 859–866 [PubMed: 10706276]

40. 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]

41. Lau AY, Scharer OD, Samson L, Verdine GL, and Ellenberger T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. Cell 95, 249–258 [PubMed: 9790531]

42. Maiti A, Noon MS, MacKerell AD Jr., Pozharski E, and Drohat AC (2012) Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. Proc Natl Acad Sci U S A 109, 8091–8096 [PubMed: 22573813]

43. Maiti A, Morgan MT, and Drohat AC (2009) Role of two strictly conserved residues in nucleotide flipping and N-glycosyl bond cleavage by human thymine DNA glycosylase. J Biol Chem 284, 36680–36688

44. Buechner CN, Maiti A, Drohat AC, and Tessmer I. (2015) Lesion search and recognition by thymine DNA glycosylase revealed by single molecule imaging. Nucleic Acids Res 43, 2716–2729 [PubMed: 25712093]
Figure 1. Base excision repair of G:T mismatches by MBD4.
(A) Formation and processing of G:T 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 G:T mismatch by hMBD4. The flipped-out thymidine is recognized by Tyr540, Gln449 and Val448 and stabilized by Arg468-mediated phosphate pinching. The backbone carbonyl of Arg468 engages in hydrogen bonding interactions with
the Watson-Crick edge of the estranged guanine. (C) Extrahelical recognition of G:T mismatch by MBD4.
Figure 2. The glycosylase activity of MBD4 toward purine:thymine base pairs.
(A) Base pairings of thymine with purines and purine analogs. O6MeG:T can form pseudo-Watson-Crick base pair. (B) Sequence of 27-bp DNA duplex used for the glycosylase activity assay. (C) Urea PAGE gel showing the glycosylase activity of MBD4 toward purine:T base pairs. Human MBD4 was incubated with dsDNA containing T paired with G, O6MeG, hypoxanthine (HX), 2,6-aminopurine (2,6-AP), or 2-aminopurine (2-AP) and the resulting reaction mixture was treated with sodium hydroxide. The sodium hydroxide-treated products were separated on a denaturing gel.
Figure 3. Single turnover kinetics of wild-type MBD4 and R468K MBD4 on G:T and/or O6MeG:T mismatches.
(A) DNA sequence used for kinetic studies. (B) Representative denaturing PAGE gel images for wild-type MBD4 and R468K mutant toward G:T- or O6MeG:T-containing DNA. (C) The activity of MBD4 catalytic domain on G:T and O6MeG mismatches under single turnover conditions ([EMBD4] = 2.5 μM, and [SDNA] = 0.25 μM) at pH 8.0 and room temperature.

Biochem J. Author manuscript; available in PMC 2021 June 03.
Figure 4. Overall structure of D560N MBD4 in complex with O6MeG:T-containing DNA. The substrate thymine opposite O6MeG lesion enters the catalytic site of MBD4 and the flipped-out thymine and O6MeG are stabilized by polar contacts with Arg468. Hydrogen bonds by Arg468 are indicated are dotted lines.
Figure 5. Structure of D560N MBD4 in complex with O6MeG:T-containing DNA.

(A) The active site of the D560N MBD4-O6MeG:T complex. A 2Fo-Fc electron density map contoured at 1σ around the substrate dT, estranged O6MedG, Arg468 and the catalytic water is shown. Note the electron density of the catalytic water is less defined compared to the catalytic water in the published D560N MBD4-G:T structure. (B) Superposition of D560N MBD4-O6MeG:T (cyan) and D560N MBD4-G:T (white) complexes. (C) Binding of D560N MBD4 to G:T versus O6MeG:T-containing DNA. Electrophoretic mobility shift assay of D560N MBD4 with 50 nM of 27-bp-FAM-labeled G:T-containing DNA or O6MeG:T-containing DNA. DNA sequence used in this binding assay is shown in Figure 2B.
Figure 6. The Effect of Arg468 mutation on MBD4 glycosylase activity.
(A) The sequence of the FAM-labeled G:T-mismatch-containing duplex DNA used for the glycosylase activity assay. The G:T mismatch is indicated in red. (B) Glycosylase activity assay of MBD4 with G:T-containing DNA. Human MBD4, either wild type or Arg468 mutated as indicated, was incubated with dsDNA (shown above), treated with sodium hydroxide, and separated on denaturing gel. The cleaved DNA at the abasic site is labeled as product and travels faster than the longer uncleaved substrate.
Figure 7. Structure of R468K/D560N MBD4 in complex with G:T-containing DNA.
(A) Closer view of the active site of R468K/D560N MBD4-G:T complex. A $2F_{o}-F_{c}$ electron density map contoured at 1σ around the substrate dT, estranged dG, Arg468, the catalytic water, and the water bridging the hydrogen bond interactions between R468K and the phosphate backbone of the base 3’ to the flipped dT is shown. (B) Superposition of the R468K/D560N MBD4-G:T (cyan) and D560N MBD4- G:T (white) complexes.
Figure 8. Representative intercalating amino acid residues observed in DNA glycosylase structures.

(A) hMBD4 (this work, 4OFA). Arg468 directly interacts with the 5’ phosphate oxygen of the substrate thymine.

(B) Bst MutY (3G0Q). Gln48 is hydrogen bonded to the 5’ phosphate oxygen of dA opposite 8-oxoguanine.

(C) hAAG (1F4R). Tyr162 forms a water-mediated hydrogen bond with the 5’-phosphate oxygen of ethenoadenine opposite dT.

(D) E. coli AlkA (1DIZ). Leu125 from the backbone carbonyl forms a water-mediated hydrogen bond with the 5’-phosphate of the abasic site analog.

(E) hOGG1 (1EBM). Asn149 engages in a water-mediated hydrogen bond to the 5’-phosphate oxygen of 8-oxoguanine opposite dC.

(F) hTDG (3UFJ). Arg275 forms a hydrogen bond with the 5’-phosphate oxygen of dU opposite dG.
|                             | PDB code       | D560N/R468K MBD4-G·T (4OFE) | D560N MBD4-O6MeG·T (4OFH) |
|-----------------------------|----------------|-----------------------------|---------------------------|
| **Data Collection**         |                |                             |                           |
| Space Group                 | P2_1212_1     |                             | P2_1212_1                 |
| Cell constants              |                |                             |                           |
| a (Å)                       | 41.722         | 40.983                      |                           |
| b                           | 55.605         | 57.64                       |                           |
| c                           | 104.245        | 105.1                       |                           |
| α (°)                       | 90.00          | 90.00                       |                           |
| β                           | 90.00          | 90.00                       |                           |
| γ                           | 90.00          | 90.00                       |                           |
| Resolution (Å)              | 33.37–2.15     | 38.18–2.22                  |                           |
|                             | (2.23–2.15)    | (2.3–2.22)                  |                           |
| Rmerge b (%)                | 0.114          | 0.104                       |                           |
| <I/σ>                       | (0.502)        | (0.540)                     |                           |
| Completeness (%)            | 99.7 (98.2)    | 99.9 (99.3)                 |                           |
| Redundancy                  | 6.6 (5.1)      | 6.8 (6.7)                   |                           |
| **Refinement**              |                |                             |                           |
| Rwork R_free d (%)          | 14.2/20.2      | 18.6/23.0                   |                           |
| Unique reflections          | 12983          | 12160                       |                           |
| Mean B factor (Å²)          |                |                             |                           |
| Protein                     | 28.83          | 37.01                       |                           |
| Ligand                      | 22.24          | 31.81                       |                           |
| Solvent                     | 34.04          | 38.42                       |                           |
| Ramachandran plot           |                |                             |                           |
| Most favored (%)            | 99.3           | 99.3                        |                           |
| Additional allowed (%)      | 0.7            | 0.7                         |                           |
| RMSD                        |                |                             |                           |
| Bond lengths (Å)            | 0.012          | 0.013                       |                           |
| Bond angles (°)             | 1.26           | 1.45                        |                           |

a Values in parentheses are for the highest resolution shell.

b $R_{merge} = \frac{\sum|I-I'|}{\sum I}$ where I is the integrated intensity of a given reflection.

c $R_{work} = \frac{\sum|F_{obs}-F_{calc}|}{\sum F_{obs}}$.

d $R_{free} = \frac{\sum|F_{obs}-F_{calc}|}{\sum F_{obs}}$, calculated using 5% of the data.