Published online 30 July 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 19* ***9917–9926***
doi:10.1093/nar/gks714

Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA

Solange Moréra[1,*], Inga Grin[2], Armelle Vigouroux[1], Sophie Couvé[2], Véronique Henriot[1], Murat Saparbaev[2] and Alexander A. Ishchenko[2,*]

[1]Laboratoire d'Enzymologie et Biochimie Structurales (LEBS), CNRS, Gif-sur-Yvette Cedex, F-91198 and [2]Groupe ''Réparation de l'ADN'', Université Paris Sud, Laboratoire ''Stabilité Génétique et Oncogenèse'' CNRS, UMR 8200, Institut de Cancérologie Gustave Roussy, Villejuif Cedex, F-94805, France

Received May 3, 2012; Revised June 28, 2012; Accepted July 2, 2012

ABSTRACT

Active DNA demethylation in mammals occurs via hydroxylation of 5-methylcytosine to 5-hydroxymethylcytosine (5hmC) by the ten-eleven translocation family of proteins (TETs). 5hmC residues in DNA can be further oxidized by TETs to 5-carboxylcytosines and/or deaminated by the Activation Induced Deaminase/Apolipoprotein B mRNA-editing enzyme complex family proteins to 5-hydromethyluracil (5hmU). Excision and replacement of these intermediates is initiated by DNA glycosylases such as thymine-DNA glycosylase (TDG), methyl-binding domain protein 4 (MBD4) and single-strand specific monofunctional uracil-DNA glycosylase 1 in the base excision repair pathway. Here, we report detailed biochemical and structural characterization of human MBD4 which contains mismatch-specific TDG activity. Full-length as well as catalytic domain (residues 426–580) of human MBD4 ($MBD4^{cat}$) can remove 5hmU when opposite to G with good efficiency. Here, we also report six crystal structures of human $MBD4^{cat}$: an unliganded form and five binary complexes with duplex DNA containing a T•G, 5hmU•G or AP•G (apurinic/apyrimidinic) mismatch at the target base pair. These structures reveal that $MBD4^{cat}$ uses a base flipping mechanism to specifically recognize thymine and 5hmU. The recognition mechanism of flipped-out 5hmU bases in $MBD4^{cat}$ active site supports the potential role of MBD4, together with TDG, in maintenance of genome stability and active DNA demethylation in mammals.

INTRODUCTION

Post-replicative methylation of cytosine at the 5-position (5mC) in DNA provides molecular basis of the epigenetic regulation of gene expression (1). However, spontaneous hydrolytic deamination of 5mC yields a mutagenic C→T transition at the CpG methylation sites that is frequently seen in inherited diseases and in the p53 gene in cancer cells (2). In mammalian cells, both mismatch-specific thymine-DNA glycosylase (TDG) and methyl-binding domain protein 4 (MBD4/MED1) prevent mutagenic impact of 5mC deamination by excising thymine from T•G mispairs that is replaced by cytosine in the base excision repair (BER) pathway (3). In BER, a DNA glycosylase binds to the abnormal base and catalyses cleavage of the base–sugar bond, generating an abasic site, which in turn is repaired by an apurinic/apyrimidinic (AP) endonuclease (4). MBD4/MED1 is a bipartite protein that belongs to the family of methyl-CpG-binding domain (MBD) proteins and consists of an N-terminal MBD domain that is linked to a C-terminal DNA glycosylase domain (5,6). MBD4 is a nuclear protein and co-localizes to heterochromatin sites in mouse cells in DNA methylation-dependent manner (7,8). MBD4 interacts with the mismatch repair protein MLH1 (9), Fas-associated death domain protein (10) and DNA methyltransferases Dnmt1 and Dnmt3b (8,11) suggesting a potential link between post-replication repair, apoptosis and DNA methylation. Mutations of MBD4 gene were detected in tumours with defective DNA mismatch repair; however, disruption of MBD4 in mouse causes a small 2- to 3-fold increase in C→T mutations at CpG sites and did not increase mini-satellite instability suggesting that MBD4 rather act as a modifier and not as driver of tumorigenesis (12,13). The catalytic domain of $MBD4^{cat}$ excises thymines from T•G mispairs

*To whom correspondence should be addressed. Tel: +33 169823470; Fax: +33 169823129; Email: morera@lebs.cnrs-gif.fr
Correspondence may also be addressed to Alexander A. Ishchenko. Tel: +33 142115405; Fax: +33 142115008; Email: alexander.ishchenko@igr.fr

at both methylated and non-methylated CpG sequence context, uracil, 5-fluorouracil and also with low efficiency 3,N^4-ethenocytosine, particularly when these bases are opposite a guanine (5,6). It was proposed that MBD4 repairs mismatches resulting from the spontaneous and/or Activation Induced Deaminase (AID)-catalysed deamination of 5mC at CpG sites (5,6,14). MBD4 belongs to the helix-hairpin-helix (HhH) DNA glycosylase superfamily, named after a conserved structural motif involved in DNA binding (15). Among the known HhH enzymes, MBD4 has the shortest sequence following the HhH motif. Crystal structures of the mouse (PDB code 1NGN) (16) and the human (PDB code 3IHO) (17) unliganded catalytic/glycosylase domain of MBD4 ($MBD4^{cat}$ residues 426–580) have been described and are very similar. So far DNA liganded structure having a target base at the lesion site is not available for the MBD4 protein.

DNA demethylation occurs either in a passive way via inhibition of *de novo* methylation after DNA replication, or by an active process, such as direct enzymatic removal of 5mC residues from DNA. Recent advances in understanding the mechanisms of active DNA demethylation in mammals have identified the ten-eleven translocation family of proteins (TETs) as 5-methylcytosine (5mC) hydroxymethylases. TETs convert 5mC to 5-hydroxymethylcytosine (5hmC) and then further oxidize it to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), both *in vitro* and *in vivo* (18–21). Human TDG (hTDG) excises with high efficiency 5fC and 5caC residues in CpG context (20,22). In addition to TETs-dependent modifications of 5mC residues, a second mechanism was shown, in which AID catalyses conversion of 5mC to thymine via a deamination reaction resulting in a G•T mismatch base pair that is repaired by MBD4 (14). Furthermore, AID/Apolipoprotein B mRNA-editing enzyme complex (APOBEC) family of cytidine deaminases can also catalyse conversion of 5hmC to 5-hydroxymethyluracil (5hmU) residue, which is in turn excised by the TDG, MBD4 and single-strand-specific monofunctional uracil-DNA glycosylase 1 (SMUG1) (23,24). These findings suggest a new unexpected role of the mismatch-specific thymine–uracil DNA glycosylases in the control of epigenetic information via removal of oxidation and deamination products of 5mC.

In this study, we characterized and compared substrate specificities of hTDG and MBD4 proteins and obtained high-resolution crystal structures of the catalytic domain of MBD4 ($MBD4^{cat}$) in complex with duplex DNA containing T•G, 5hmU•G or AP•G base-pairs. The roles of the MBD4-initiated BER pathway in the active DNA methylation and prevention of spontaneous mutagenesis are discussed.

MATERIALS AND METHODS

Oligonucleotides and proteins

All oligodeoxyribonucleotides containing modified residues and their complementary oligonucleotides were purchased from Eurogentec (Seraing, Belgium) including the following: 30-mer d(TGACTGCATAXGCATGTAG ACGATGTGCAT) oligonucleotide for kinetic studies where X is 5hmU, 5caC, 5fC or T and 30-mer d(TGAC TGCATAXTCATGTAGACGATGTGCAT) oligonucleotide where target residue X is located in XpT-context and their 30-mer complementary regular oligonucleotides, containing dA, dG, dC or T opposite to a target residue. Unless otherwise stated, 30-mer oligonucleotides where target residues are located in XpG context were used in the DNA repair assays. Oligonucleotides were 5′-end labelled by T4 polynucleotide kinase (New England Biolabs) in the presence of [γ-^{32}P]-ATP (4500 Ci/mmol; ICN Pharmaceuticals France, S.A., Orsay, France) as recommended by the manufacturers. The 5′-[^{32}P]-labelled oligonucleotides were annealed to their appropriate complementary oligonucleotides in a buffer containing 50 mM NaCl, 17 mM HEPES–KOH pH 7.2 at 65°C for 3 min as previously described (25). The resulting duplex oligonucleotides are referred to as X•C (G, A, T), respectively, where X is a modified residue.

The 12 mer DNA duplex sequence used for crystallization assays is d(CCAGCGXGCAGC)/d(GCTGCGCGC TGG) where X is T or 5hmU. The MALDI–TOF mass spectrometry analysis of the oligonucleotides performed by the manufacturer confirmed their size and homogeneity. In addition, the purity and integrity of the oligonucleotide preparations were verified by denaturing polyacrylamide gel electrophoresis (PAGE). The 10 mM oligonucleotide solutions were mixed in equal proportions in 2 mM Tris–HCl pH 7.0 and hybridized by heating to 65°C for 3 min and cooling down to room temperature over 2 h.

Collection of the purified DNA glycosylases was from the laboratory stock (26). Human SMUG1 was purchased from New England Biolabs (Evry Cedex France).

Expression and purification of full-length human MBD4 and TDG proteins

The expression vectors pET6H-MBD4 and pET28c-hTDG were generously provided by Dr Adrian Bird (University of Edinburgh, Edinburgh, UK) and Dr Alexander Drohat (University of Maryland, Baltimore, MD, USA), respectively. *Escherichia coli* Rosetta 2 (DE3) cells transformed with a pET6H-MBD4 or pET28c-hTDG were grown at 37°C in Luria Broth medium, supplemented with appropriate antibiotics, on an orbital shaker to OD_{600nm} = 0.6–0.8. Then temperature was reduced to 30°C and the protein expression was induced by 0.2 mM isopropyl β-D-galactopyranoside (IPTG; Sigma-Aldrich Chimie S.a.r.l., Lyon, France), and the cells were further grown either for 2 h when inducing expression of hTDG or 15 h when inducing expression of the MBD4 protein. Bacteria were harvested by centrifugation and cell pellets were lysed using a French press at 18 000 psi in buffer containing 20 mM HEPES–KOH pH 7.6, 50 mM KCl supplemented with CompleteTM Protease Inhibitor Cocktail (Roche Diagnostics, Switzerland). Cell lysates were cleared by

centrifugation at 40 000*g* for 30 min at 4°C and the resulting supernatant was loaded onto HiTrap Chelating HP column (GE Healthcare, Aulnay sous Bois, France). All purification procedures were carried out at 4°C. The column was washed with buffer A (20 mM HEPES–KOH pH 7.6, 500 mM NaCl, 20 mM imidazole) and bound proteins were eluted in a 0–100% gradient of buffer B (20 mM HEPES–KOH pH 7.6, 500 mM NaCl, 500 mM imidazole). Eluted fractions were analysed by sodium dodecyl sulphate–PAGE and fractions containing the pure His-tagged MBD4 and hTDG proteins were stored at −80°C in 50% glycerol. The concentration of purified proteins was determined by the method of Bradford.

Construction, expression and purification of the MBD4^{cat} proteins

Coding sequences for MBD4^{cat} were amplified by PCR using the following primers: forward d(GGGCCCCATA TGCTTAGCCCCCCACGACGT) and reverse d(CGCC GAATTCTTAATGGTGATGGTGATGGTGAGATA GACTTAATTTTTC) and inserted into pET29b vector (Novagen, Merck4Biosciences, France) at NdeI and EcoRI sites. The MBD4^{catD560A} mutant was constructed by site-directed mutagenesis using Quick Change Kit (Stratagene, Agilent Technologies Sciences de la Vie et Analyse Chimique, Massy, France) with the following oligonucleotides: forward d(GAAGCAGGTGCACCCT GAAGCCCATTAAATAAATATA) and reverse: d(TG ATATTTATTTAATTTGTGGGCTTCAGGGTGCAC CTGCTTC). The resulting plasmids were introduced into *Escherichia coli* DH5α (Invitrogen, Life Technologies SAS, Saint Aubin, France) and the mutation was verified by sequencing (GATC Biotech SARL, Mulhouse, France).

The MBD4^{catWT}, MBD4^{catD560A} and MBD4^{catQ446A} mutant proteins were expressed in *E. coli* BL21 (DE3) (Invitrogen). Bacterial cultures were grown at 37°C in 2TY medium and the protein expression was induced by the addition of 0.5 mM IPTG at 28°C during 5 h. Pelleted cells were disrupted by sonication in a buffer containing 50 mM Tris–HCl pH 7.5, 10 mM imidazole, 500 mM NaCl, 10% glycerol and a protease inhibitor cocktail (Sigma-Aldrich). The resulting cell lysate was cleared by centrifugation at 20 000*g* for 30 min at 4°C. The MBD4^{cat} proteins were purified using nickel-affinity chromatography on Ni–NTA agarose column (GE Healthcare) followed by gel filtration chromatography on a HiLoad Superdex S200 26/60 (GE Healthcare) equilibrated in 50 mM Tris pH 7.5, 150 mM NaCl and 10% glycerol. The chromatography fractions containing highly purified MBD4^{cat} proteins were concentrated in a Vivaspin concentrator (Sartorius, Biohit France S.A.S., Dourdan, France) up to concentrations of 3 mg/ml.

Crystallization and structure determination of MBD4^{cat}

Crystallization conditions are summarized in Table 1. For the free-liganded protein and the DNA–protein complexes (140 μM protein and 700 μM 12-mer DNA), conditions were screened using the Qiagen kits Crystals and then manually optimized at 18°C in hanging drop by mixing equal volumes of the protein or protein–DNA solution with precipitant solution. Crystals were transferred to a cryoprotectant solution (paraffin oil or 20% PEG 400) and flash frozen in liquid nitrogen. Diffraction data were collected at 100K on the PROXIMA I beamline at SOLEIL synchrotron (Saint-Aubin, France) and intensities were integrated with the program XDS (27) for all crystals except that of AP•G complex which was collected on ID29 beamline at ESRF (Grenoble). Data collection and processing statistics are given in Table 1. Structure determination of all crystals was performed by molecular replacement with PHASER (28) using first the coordinates of the free-liganded structure (PDB code 3IHO) for our free-liganded structure at a better resolution and next our model for the DNA–protein structures. Refinement was performed using BUSTER (29) and electron density maps were evaluated using COOT (30). The refined models include 138 residues (from 437 to 575). Refinement details of the six structures are shown in Table 1. Molecular graphics images were generated using PYMOL (http://www.pymol.org).

DNA glycosylase activity assays

The standard reaction mixture (20 μl) contained 5 nM of 5′-[^{32}P]-labelled oligonucleotide duplex in 20 mM HEPES–KOH pH 7.6, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1 mg/ml bovine serum albumin and 50 nM of purified enzyme, unless otherwise stated. For Udg, Mug, AlkA and MutY protein reactions were performed in the buffer containing 70 mM HEPES–KOH pH 7.6, 0.5 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 1.5% glycerol, for MutY, incubation buffer was supplemented with 5 μM $ZnCl_2$. Incubations were carried out at 37°C for 30 min, the reaction was stopped by adding 0.1 M NaOH and the samples were heated at 99°C for 3 min to cleave the DNA at abasic sites and then solutions were neutralized by adding 0.1 M HCl.

The resulting samples were desalted by hand-made spin-down columns filled with Sephadex G25 (Amersham Biosciences) equilibrated in 7.5 M urea. Purified reaction products were separated by electrophoresis in denaturing 20% (w/v) polyacrylamide gels (7 M urea, 0.5× TBE). Gels were exposed to a Fuji FLA-3000 Phosphor Screen and analysed using Image Gauge V3.12 software.

Determination of the kinetic parameters of DNA glycosylase activities

To measure the kinetic parameters of DNA glycosylases-catalysed excision of modified bases, reactions were performed under single turnover conditions. For this purpose large excess of enzyme over DNA substrate were used under steady-state turnover conditions ($[E] >> [S] > K_d$). These conditions provide measurement of a rate constant (k_{obs}) that is not impacted neither by enzyme–substrate association, nor product release, nor inhibition, such that k_{obs} reflects the maximal base excision rate ($k_{obs} \approx k_{max}$) (31). The data were fitted by non-linear regression, and a one-phase exponential association model was used with the following

Table 1. Crystallographic data and refinement parameters

PDB code	WT 4E9E	W + AP•G DNA 4E9F	mutant + T•G DNA 4E9G	mutant + HMU•G (5hmU1) 4E9H	mutant + HMU•G (5hmU2) 4EA4	mutant + HMU•G (5hmU3) 4EA5
Precipitant	30% PEG 4000, 0.1 M Tris–HCl pH 8.5, 0.2 M $MgCl_2$	25% ethylene glycol	15% PEG MME 2000, 0.1 M Mes pH 6.5, 0.2 M NaAcetate	25% PEG 1500, 0.1 M HEPES pH 7.5	20% PEG 4000, 0.1 M NaCitrate pH 5.6, 20% Isoprapanol	20% Jeffamine M2070, 20% dimethylsulfoxite
Data collection						
Beamline, synchrotron	PX1 SOLEIL	ID29 ESRF	PX1, SOLEIL	PX1, SOLEIL	PX1 SOLEIL	PX1 SOLEIL
Space group	H3	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell parameters						
a, b	81. 81.	41.1 55.8	40.1 61.5	41.1 96.7	40.3 63.2	38.5 99.8
c (Å)	74.5	96.6	96.5	55.1	94.8	53.5
α, β, γ (°)	90 90 120	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90
Resolution (Å)	40–1.9 (2.02–1.9)	38–1.8 (1.9–1.8)	38–2.35 (2.49–2.35)	47–3 (3.18–3)	37.9–2 (2.12–2)	36–2.15 (2.27–2.15)
No. of observed reflections	59 217 (5383)	116 349 (18532)	29 916 (4785)	19 926 (3218)	92 323 (14 935)	81 113 (12 105)
No. of unique reflections	13 475 (1661)	21 525 (2364)	10 209 (1603)	4698 (736)	16 435 (2634)	11 828 (1783)
R_{sym} (%)	8.9 (77.6)	6.3 (64.7)	5.5 (45.5)	14.8 (71.4)	7.8 (88)	17 (79.5)
Completeness (%)	94 (72)	99.6(98.5)	98 (98)	99.2 (98.9)	96.3 (97.4)	99.2 (95.6)
I/σ	10.9 (1.9)	13.3 (2)	13.9 (2.3)	10.1 (2.15)	14.1 (1.9)	8.2 (2.4)
Refinement						
R_{cryst} (%)	19.28	21.4	20.5	17.3	22.4	19.1
R_{free} (%)	21.82	24.8	25.8	25.7	26.4	23.5
rms bond deviation (Å)	0.01	0.01	0.01	0.01	0.01	0.01
rms angle deviation (°)	0.93	1.07	1.18	1.23	1.16	1.1
Average B ($Å^2$)						
protein	35.3	38.3	43.3	41.2	32.4	26.8
DNA (C; D)		57.5; 54.7	56.2; 64.9	61.1; 66.5	44.8; 54.2	39; 35.5
Solvent	39.8	49	44.4	28.2	36.9	37.6

Values for the highest resolution shell are in parentheses.

parameters: $Y = Y_{max} \times [1 - \exp(-k_{obs} \times t)]$, where Y_{max} is the amplitude, k_{obs} is the rate constant and t is the reaction time (in min).

RESULTS

Characterization of hTDG and MBD4 DNA glycosylase activities towards 5mC derivatives

To study DNA glycosylase activities of the hTDG and MBD4 proteins, we used the 5′-[^{32}P]-labelled 30-mer 5hmU•G, 5fC•G 5caC•G and T•G duplex oligonucleotides, where modified cytosines and mismatched T were placed in a CpG context. Since mono-functional DNA glycosylases devoid of AP site-nicking activity, the oligonucleotides, after incubation with the glycosylases, were subjected to hot alkaline treatment (2 M NaOH at 90°C for 5 min) in order to cleave DNA at abasic sites generated by the excision of modified bases. As shown in Figure 1, incubation of 5hmU•G with hTDG, MBD4 and MBD4^{cat} generates 10-mer cleavage fragment indicating excision of 5hmU residue at Position 11 (Lanes 3–5). This result confirms previous observations that hTDG and MBD4 can excise 5hmU residue when present in duplex DNA and generate an AP site (23,32). As expected, hTDG but not MBD4s can excise with high-efficiency 5fC and 5caC residues (Lanes 7 and 11). In addition, hTDG and MBD4 excise 5hmU in non-CpG context with similar efficiency (Lanes 15 and 16) when compared with 5hmU positioned in CpG context (Lanes 3 and 4), whereas activity of the MBD4^{cat} protein on 5hmU exhibits some dependence on the sequence context (Lanes 5 versus 17). Comparison of protein concentration dependence and time course kinetics of hTDG, MBD4 and MBD4^{cat} on 5hmU•G showed that hTDG is more efficient than MBD4 enzyme (Supplementary Figure S1). Overall, these results show that MBD4 does not act on 5fC and 5caC residues but can excise 5hmU with good efficiency.

Next, we characterized the substrate specificity of the MBD4, MBD4^{cat} and hTDG proteins by measuring the cleavage rates of 5hmU•G and T•G duplexes in single turnover kinetics experiments which provide the maximal rate of base excision (k_{max}) for a given substrate as described in 'Materials and Methods' section (Supplementary Figure S2). As shown in Table 2, the k_{max} values for 5hmU excision by MBD4 and MBD4^{cat} are 10- and 7-fold lower when compared with that of hTDG, respectively. As for classic T•G substrate, the k_{max} values for all three enzymes were in the same order with hTDG being 2-fold more efficient when compared with MBD4s. Importantly, all enzymes tested were more efficient on 5hmU•G than on T•G duplexes (~1.5-fold for MBD4s and 6-fold for hTDG) suggesting that MBD4 and hTDG act more specifically on 5hmU bases. As expected, both MBD4 and MBD4^{cat} have very similar substrate specificity (Table 2) confirming that the MBD domain of MBD4 has no catalytic function (33). These results together with previously published data (23,32) suggest that *in vivo* both TDG and MBD4 play a role in the removal of deaminated 5hmC residues.

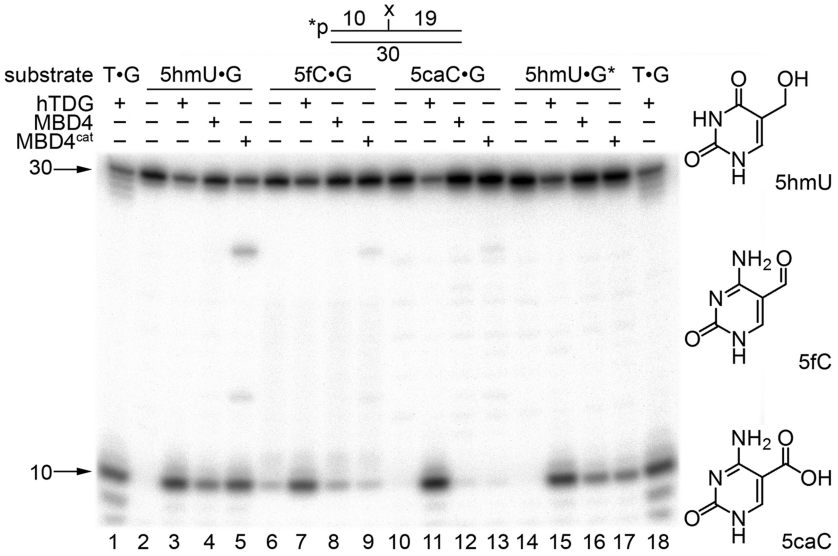

Figure 1. Substrate specificity of the hTDG, MBD4, MBD4^{cat} proteins. Duplex oligonucleotides 5hmU•G, 5caC•G, 5fC•G containing 5mC-derivatives in CpG-context and 5hmU•G* containing 5hmU in CpT-context were used as DNA substrates. The 5′-[^{32}P]-labelled 30-mer oligonucleotide DNA (5 nM) was incubated with a large excess of the designated DNA glycosylase (50 nM) at 37°C for 30 min. The reaction products were analysed as described in 'Materials and Methods' section.

Table 2. Kinetic parameters for hTDG, MBD4 and MBD4^{cat} measured under single turnover condition for the removal of 5hmU and T residues when present in duplex DNA

Enzyme	$k_{obs} \approx k_{max}$, min^{-1}	
	5hmU•G	T•G
hTDG	7.95 ± 0.26	1.25 ± 0.04
MBD4^{cat}	1.10 ± 0.06	0.62 ± 0.02
MBD4	0.78 ± 0.03	0.55 ± 0.03

Activity of bacterial and human DNA glycosylases on oligonucleotides containing oxidized and deaminated derivatives of 5mC

We investigated whether 5hmU, 5caC and 5fC residues are also substrates for the previously characterized bacterial and human DNA glycosylases. The 5′-[^{32}P]-labelled single-stranded and duplex oligonucleotides containing 5hmU•G, 5caC•G or 5fC•G were challenged with a variety of highly purified DNA glycosylases. When using the mono-functional DNA glycosylases, the samples after incubation were subjected to hot alkaline treatment. The *E. coli* DNA glycosylase Mug (a homologue of hTDG) excises 5hmU with good efficiency in duplex DNA, but not in single-stranded form, while Nth and Nei exhibit the same activity but with much lower efficiency (Figure 2A). In agreement with previous observations all three human DNA glycosylases: TDG, MBD4 and SMUG1 excise 5hmU with good efficiency when it is present in duplex DNA but only SMUG1 was able to excise 5hmU in a single-stranded form (Figure 2B and Supplementary Figure S3). In addition, the human homolog of bacterial Nei, NEIL1 can excise, although with weak efficiency, 5hmU residue in duplex DNA. No detectable activity on 5hmU-containing DNA substrates was observed for hUNG2, hNTH1, ANPG70, hOGG1 or NEIL2 glycosylases. Next, we examined the repair of 5caC and 5fC residues by bacterial and human enzymes. Interestingly, the *E. coli* DNA glycosylase Mug can excise with good efficiency 5caC and 5fC residues when present in both single-stranded and duplex DNA (Figure 3A and Supplementary Figure S4A). Despite being used in molar excess, none of the others *E. coli* DNA glycosylases tested were able to excise 5caC and 5fC residues in DNA (Figure 3A and Supplementary Figure S4A). When incubating 5caC and 5fC containing DNA substrates with human enzymes, we showed that hTDG excises with high efficiency 5caC and 5fC residues not only when it present in duplex DNA but also in single-stranded form (Figure 3B and Supplementary Figure S4B). Importantly, DNA repair activities of bacterial and human enzymes on 5fC containing oligonucleotides mimic those on 5caC (Figure 3 and Supplementary Figure S4). These results indicate that hTDG is a main human enzyme removing carboxylated and formylated cytosines and confirms previous findings by other laboratories (20,22). The hUNG2, hNTH1, ANPG70, hOGG1, NEIL2 and hSMUG1 glycosylases have no detectable activity on 5caC and 5fC substrates, whereas MBD4 and NEIL1 exhibited very weak activity on 5caC substrates (Figure 3 and Supplementary Figures S3 and S4).

Crystal structures of ligand-free and substrate-bound MBD4^{cat}

In order to get insight into the structural bases of substrate specificity and catalytic mechanism of human MBD4, we performed crystallographic studies of MBD4^{cat} complexed with its DNA substrates. For this purpose, a catalytically

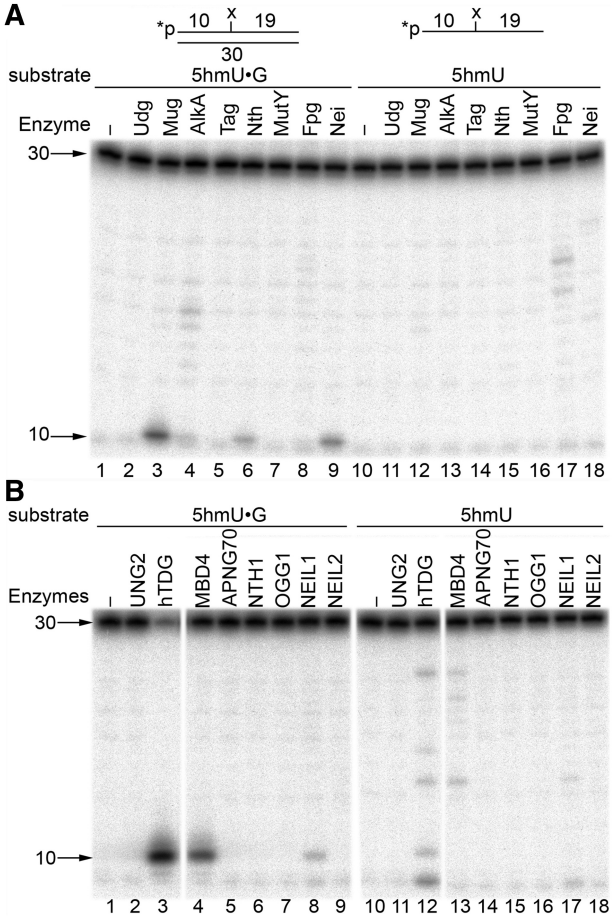

Figure 2. Enzymatic activity of various *E. coli* and human DNA glycosylases on dsDNA and ssDNA containing single 5hmU residue. (**A**) DNA glycosylase activities of the *E. coli* Ung, Mug, AlkA, Tag, Nth, MutY, Fpg and Nei proteins. (**B**) DNA glycosylase activities of the human UNG2, hTDG, MBD4, APNG70, NTH, OGG1, NEIL1 and NEIL2 proteins. The 5′-[^{32}P]-labelled 5hmU•G and 5hmU (5 nM) were incubated with a large excess of the designated DNA glycosylase (50 nM) at 37°C for 30 min. The reaction products were analysed as described in 'Materials and Methods' section.

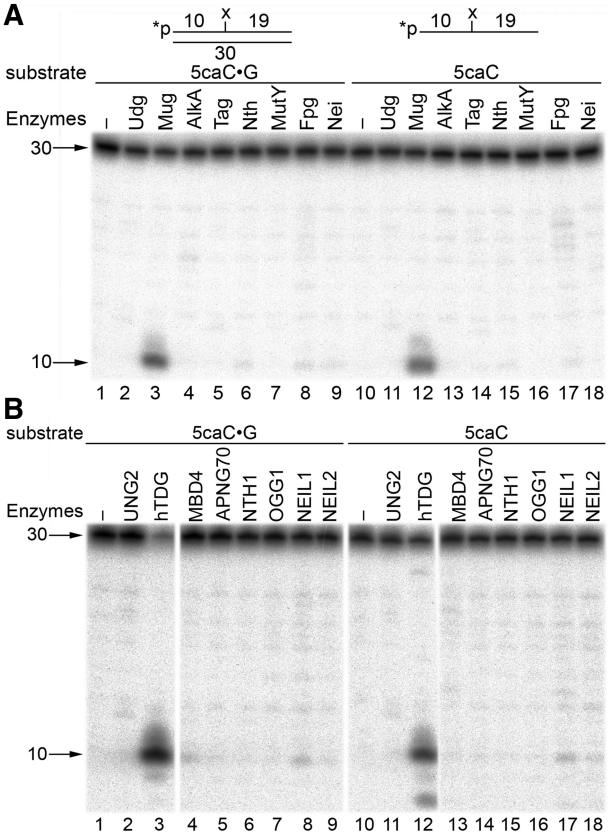

Figure 3. Enzymatic activity of various *E. coli* and human DNA glycosylases on dsDNA and ssDNA containing single 5caC residue. (**A**) DNA glycosylase activities of the *E. coli* Ung, Mug, AlkA, Tag, Nth, MutY, Fpg and Nei proteins. (**B**) DNA glycosylase activities of the human UNG2, hTDG, MBD4, APNG70, NTH, OGG1, NEIL1 and NEIL2 proteins. The 5′-[^{32}P]-labelled 5caC•G and 5caC (5 nM) oligonucleotide were incubated with a large excess of the designated DNA glycosylase (50 nM) at 37°C for 30 min. The reaction products were analysed as described in 'Materials and Methods' section.

inactive MBD4^{cat} mutant has been generated. Previous studies of the crystal structure of the C-terminal domain of murine MBD4 revealed that it is a member of the HhH DNA glycosylase superfamily similar to AlkA and OGG1 and that the conserved amino acid D534 is a putative catalytic residue (16). To examine the role of the corresponding catalytic D560 residue in human MBD4 protein, we obtained mutant MBD4^{catD560A} and characterized its DNA glycosylase activity. As expected, MBD4^{catD560A} exhibits a drastic, >50-fold, decrease in excision rate of 5hmU residue in 5hmU•G duplex when compared with the wild-type MBD4^{cat} protein (Supplementary Figure S1) suggesting that D560 is indeed essential for catalytic activity and that this inactive mutant protein can be used to obtain complex with DNA substrate.

We determined five high-resolution X-ray structures of binary complexes with a 12-mer DNA containing either an 5hmU•G (3 Å, 2 Å and 2.15 Å resolution named 5hmU1 (5hmU is in a productive state), 5hmU2 (in a non-productive state) and 5hmU3 (in a disordered state), respectively, in Table 1) or a T•G mismatch (2.35 Å resolution) or AP•G (1.8 Å resolution) at Position 7. We also determined an unliganded structure at higher resolution (1.9 Å) than that deposited in the PDB (3IHO, 2.7 Å) (17) which was used first as a search model for phasing determination by molecular replacement.

The structures of all DNA complexes share neighbouring crystal packing with an asymmetric unit containing an MBD4^{cat} monomer bound to the 12-mer DNA. All MBD4^{cat} molecules are very similar with an average root-mean-square deviation (RMSD) value of 0.4 Å between 137 Cα atoms. The unbound MBD4^{cat} also resemble the DNA-bound MBD4^{cat} molecules with an average RMSD value of 0.6 Å between all defined Cα atoms. Only two loop regions (residues 466–471 between helices α2 and α3 and residues 503–508 between helices α5 and α6) can move up to 2 Å to accommodate a bound DNA. Thus, MBD4^{cat} does not require significant conformational changes for DNA binding. All 12-mer DNA bound to MBD4^{cat} show the same large distortion at the target base (Figure 4A) similarly to what has been

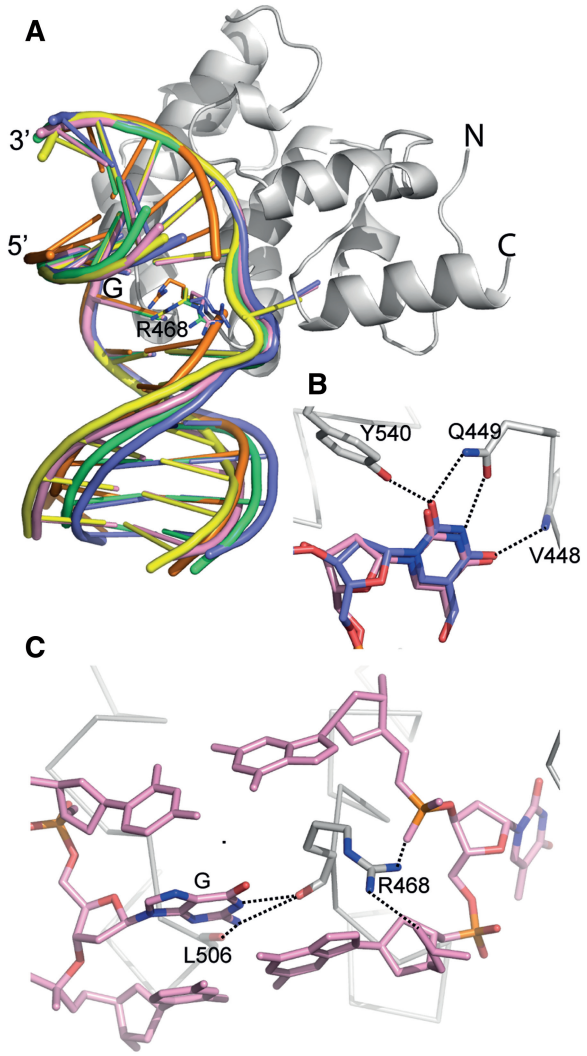

Figure 4. MBD4^{cat}-DNA binding. (**A**) Ribbon representation showing MBD4^{cat} (in grey) and the superposition of the whole DNA fragments. The mismatched thymine, AP site and 5hmU bases in productive, non-productive binding and mobile state are coloured pink, green, slate, yellow and orange, respectively. N- and C-termini are indicated. (**B**) Superposition of MBD4^{catD560A} 5hmU•G and T•G complexes, flipped-out 5hmU in a productive binding and thymine are shown in slate and pink respectively. Residues involved in the interactions are labelled and shown as sticks and hydrogen bonds (distances <3.2 Å) are shown as black dashes. (**C**) Residues (in grey) involved in the interactions with the orphan guanine (pink and atom colours) are labelled and shown as sticks. Hydrogen bonds (distances <3.2 Å) are shown as black dashes.

observed for HhH DNA glycosylases. MBD4^{cat} bends the DNA 55°, which is roughly the same as the bend induced by MutY and EndoIII (34). The flipped-out abasic site, thymine and 5hmU from 5hmU1 structure are well defined in electron density map into the enzyme active site pocket defined by residues 447–449, 560–562, Leu466, Gly471 and Tyr540 (Supplementary Figure S5). The bases superimpose well and make the same protein interactions (Figure 4B). Their O4 and O2 atoms interact with the main chain amino group of Val448 and the Tyr540 side chain, respectively. Both N3 and O2 interact with the side chain of Gln449. The mutant MBD4^{catQ449A} is completely inactive towards all DNA substrates tested indicating the essential role of Gln449 in substrate recognition and stabilization (Supplementary Figure S6). The 5-hydroxymethyl group of 5hmU does not make protein interaction.

MBD4^{cat} binds DNA via three regions (the loops 466–471 and 503–511 and the Gly-rich hairpin loop of HhH motif 534–541) and interacts mostly with the strand containing the substrate base. The loop 466–471 penetrates into the DNA duplex through the minor groove and Arg468 fills the space in the DNA duplex vacated by the flipped nucleotide or AP site. In T•G, AP•G and 5hmU1•G structures, its guanidinium group interacts with both the phosphate groups 5′ of two bases upstream and 3′ to the productive target base or AP site (Figure 4C). The main chain carbonyl groups of Arg468 and Leu506 pack against the opposite guanine and provide specific hydrogen bonds to its N1 and N2 atoms (Figure 4C).

Notably, Arg468 seems to have a key role in locking the flipped-out base in a productive binding for catalysis. Indeed, this mobile residue can move more the 5 Å as observed between the productive and non-productive complex (Supplementary Figure S7). In both 5hmU2 (5hmU is in a non-productive state) and 5hmU3 (5hmU is disordered) structures, Arg468 interacts with the O6 atom of the unpaired G. The 5hmU2 structure reveals a flipped-out 5hmU located at the entrance of the active site pocket in a position incompatible with the presence of the catalytic residue Asp560. The C1′ atom is 2.1 Å away from its position observed in the productive complex (Supplementary Figure S7). The 5hmU3 structure shows a disordered 5hmU base despite the presence of a distorted bound DNA. However, a symmetric cytosine of the disrupted terminal C•G base pair is held into the active site pocket of the enzyme preventing any target base penetration for a productive binding (Figure 5). This bound cytosine acts as an inhibitor by interacting with the same protein residues as those involved in the flipped-out base recognition. Indeed, Tyr540 and Gln449 make hydrogen bonds with the NH_2 group and the amino group of Val448 interacts with O2 and N3 atoms.

DISCUSSION

While this article was submitted for publication, Manvilla *et al.* (35) reported the crystal structure of MBD4 in complex with a 11-mer DNA containing an abasic site. Despite lower resolution of the crystal structure (2.76 Å) when compared with our MBD4-AP•G DNA structure (1.8 Å) and differences in the primary sequence of the DNA fragment used, both structures are very similar with an average RMSD value of 0.4 Å between all defined Cα atoms (Supplementary Figure S8). Both MBD4-AP•G DNA structures exhibit the same DNA–protein interactions; however, owing to the lack of a base in the enzyme active site pocket in the published structure, functionally important residues that participate in base recognition by MBD4 have not been demonstrated (35).

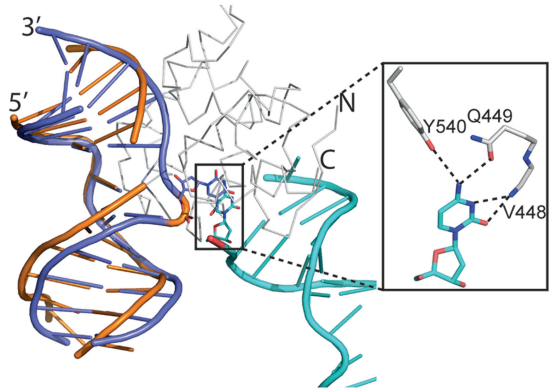

Figure 5. Superposition of 5hmU•G (disordered 5hmU base in orange) and 5hmU•G (productive complex in slate) complexes with the flipped-out thymine shown in sticks. A crystal symmetric terminal cytosine (in cyan) is held in the active site pocket. N- and C-termini are indicated. A close-up view of the polar interactions (distances< 3.2 Å) between the symmetric cytosine (in cyan) and MBD4 residues (in grey) is shown in the box. The bound cytosine acts as an inhibitor.

Our work describes the first crystal structures of the catalytic domain of MBD4 in complex with mismatched bases located at the centre of a 12-mer DNA duplex. The thymine and 5hmU mispaired with guanine is extruded from the DNA helix and located in the enzyme active site. The structures revealed that MBD4 specifically recognizes thymine and 5hmU opposite a guanine (Figure 4). Interestingly, a group such as 5-hydroxymethyl on C5 would have no effect on MBD4 ligand binding as there is no interaction between it and the enzyme. Importantly, it appeared unlikely for a cytosine and oxidized 5mC bases to be trapped in the active site pocket of MBD4^{cat} due to the unfavourable environment of the main chain amino group of Val448 which would create a repulsive force directly towards their NH_2 group. This structural feature of the active site pocket is consistent with the absence of MBD4 activity on 5hmC, 5caC and 5fC, indicating that the main biological function of MBD4 is to repair mismatched/deaminated cytosine residues. This makes a major difference with mammalian TDG which has broader substrate specificity and can recognize 5caC and 5fC in both duplex and single-stranded DNA (Figures 1 and 3 and Supplementary Figure S4). Previous studies have also shown that 5caC and 5fC residues are substrates only for mammalian TDG proteins (22). Repair of 5caC and 5fC residues in other than mammal organisms was unknown. In this study, for the first time, we demonstrated that *E. coli* contains the DNA glycosylase Mug that can efficiently remove 5hmU, 5caC and 5fC from duplex DNA suggesting high evolutionary conservation of the broad substrate specificity of TDG family enzymes. Moreover, Mug and hTDG can efficiently remove 5caC and 5fC from single-stranded DNA. At present, biological role of Mug-catalysed removal of 5mC derivatives is not clear, since bacteria lack genome-wide methylation and TETs enzymes. We also showed that in human cell, NEIL1 can weakly excise 5hmU in addition to TDG, SMUG1 and MBD4 thus confirming previous observation (36). Although previous studies have shown that 5hmU residues are substrates for the mammalian mismatch-specific uracil-DNA glycosylase family enzymes SMUG1, TDG and MBD4 (23,24), up to now, no detailed characterization of the substrate specificity of human MBD4 has been performed. Here, we examined the substrate specificity of the full-length human MBD4 protein and MBD4^{cat} towards 5hmU and other oxidized derivatives of 5mC in order to further define the biological relevance of these DNA glycosylases. Detailed characterization of the substrate specificity of the human full-length and catalytic domain MBD4 proteins confirmed previous data showing that MBD4 excises 5hmU residues from 5hmU•G duplex with good efficiency (Figures 1 and 2). Interestingly, MBD4 was somewhat less efficient on 5hmU when compared with hTDG suggesting that later enzyme together with hSMUG1 are major human DNA glycosylases removing 5hmU *in vivo* (Table 2 and Supplementary Figure S3). Nevertheless, our biochemical data provide evidence for the role of MBD4 as an efficient back-up enzyme which can specifically act in densely methylated CpG regions of chromosomal DNA, where deamination of 5mC and 5hmC is expected to be more frequent. Indeed, the MBD domain may facilitate the localization of MBD4 to methyl-CpG-rich regions of the genome *in vivo*, reflecting a specific role of MBD4 in the repair of 5hmU•G, U•G and T•G mismatches in heterochromatic regions (37). Although TDG is endowed with wider substrate specificity when compared with MBD4, it lacks a MBD domain and tends to associate with transcriptionally active euchromatin (38) and non-methylated CpG islands to protect them from aberrant DNA methylation (23,39).

Finally, the current crystal structures (especially the 5hmU3 structure) can be used as a template to develop inhibitors of MBD4^{cat} in the context of the active DNA demethylation process in human cells. Increased rate of C→T mutations at CpG sites in *Mbd4*$^{-/-}$ mice (12,13) strongly suggest that MBD4 plays an important role in prevention of spontaneous mutation which could be due to either spontaneous or enzymatically induced deamination of 5mC and 5hmC to T and 5hmU residues, respectively. The fact that 5hmU residues are intermediates of active DNA demethylation that is produced enzymatically by combined action of TETs and AID/APOBEC proteins implies that 5hmU may be generated at significant amount in the cells during reprogramming and cancerogenesis. If not, repaired 5hmU can lead to mutation, therefore human cells hold three DNA glycosylases to ensure efficient repair of this extremely mutagenic derivative of 5mC residue. A recent study on the embryonic lethal phenotype of TDG knockout mice demonstrated that mutation frequencies in a Big Blue transgenic MEFs *Tdg*$^{-/-}$ were similar to that of MEFs *WT* suggesting that the biological role of TDG in the repair of oxidized and/or deaminated cytosine damage may be rather minor (39). Furthermore, *Smug1*-knockout mice show no obvious cancer predisposition phenotype possibly implying no increase in spontaneous mutation rate (40). These observations point to prevalent role of MBD4 in spontaneous mutation prevention *in vivo*. Indeed, it was shown that TDG is associated

with transcriptionally active euchromatin (38), whereas MBD4 rather localize in heterochromatin regions which is in general heavily methylated (7,8). It is tempting to speculate that MBD4 may function to remove 5hmU in heterochromatin regions that arise due to active DNA demethylation processes in order to prevent mutation and also to serve as back up enzyme for TDG and SMUG1.

ACCESSION NUMBERS

The atomic coordinates and structure factors of the unliganded, AP•G, T•G, 5hmU1, 5hmU2 and 5hmU3 $MBD4^{cat}$ structures have been deposited in the protein data bank (http://www.rcsb.org) under accession 4E9E, 4E9F, 4E9G, 4E9H, 4EA4 and 4EA5, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–8.

ACKNOWLEDGEMENTS

The authors are grateful to Beatriz Guimaraes for help in data collection on PROXIMA I at SOLEIL. The authors thank the staff of ESRF (Grenoble) for making station ID29 available. We thank Laurent Larivière for help in preparing figures.

FUNDING

Centre National de la Recherche Scientifique, CNRS (to S.M.); Fondation de France [#00012091 and #2012 00029161 to A.A.I.]; CNRS [PICS N5479-Russie to M.S.]; Electricité de France [EDF RB2011 to M.S.]; Agence Nationale de la Recherche [ANR-09-GENO-000 to M.S.]; [RFBR 11-04-00807-a and MK-2703.2011.4 to I.G.]; Fondation pour la Recherche Médicale [Equipe FRM 2007] and Association pour la Recherche sur le Cancer [ARC PDF20101202141 to S.C. and I.G.], respectively. The cloning and crystallization work has benefited from the LEBS facilities of the IMAGIF Structural Biology and Proteomic Unit in the 'Centre de recherche de Gif' (www.imagif.cnrs.fr). Funding for open access charge: CNRS, ANR.

Conflict of interest statement. None declared.

REFERENCES

1. Bird,A. (2011) Putting the DNA back into DNA methylation. *Nat. Genet.*, **43**, 1050–1051.
2. Pfeifer,G.P. and Besaratinia,A. (2009) Mutational spectra of human cancer. *Hum. Genet.*, **125**, 493–506.
3. Cortazar,D., Kunz,C., Saito,Y., Steinacher,R. and Schar,P. (2007) The enigmatic thymine DNA glycosylase. *DNA Repair*, **6**, 489–504.
4. Hitomi,K., Iwai,S. and Tainer,J.A. (2007) The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair*, **6**, 410–428.
5. Hendrich,B., Hardeland,U., Ng,H.H., 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.
6. Petronzelli,F., Riccio,A., Markham,G.D., Seeholzer,S.H., Genuardi,M., Karbowski,M., Yeung,A.T., Matsumoto,Y. and Bellacosa,A. (2000) Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J. Cell. Physiol.*, **185**, 473–480.
7. Hendrich,B. and Bird,A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.
8. Ruzov,A., Shorning,B., Mortusewicz,O., Dunican,D.S., Leonhardt,H. and Meehan,R.R. (2009) MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development*, **136**, 2277–2286.
9. Bellacosa,A., Cicchillitti,L., Schepis,F., Riccio,A., Yeung,A.T., Matsumoto,Y., Golemis,E.A., 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. USA*, **96**, 3969–3974.
10. Screaton,R.A., Kiessling,S., Sansom,O.J., Millar,C.B., Maddison,K., Bird,A., Clarke,A.R. and Frisch,S.M. (2003) Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. *Proc. Natl Acad. Sci. USA*, **100**, 5211–5216.
11. Boland,M.J. and Christman,J.K. (2008) Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J. Mol. Biol.*, **379**, 492–504.
12. Millar,C.B., Guy,J., Sansom,O.J., Selfridge,J., MacDougall,E., Hendrich,B., Keightley,P.D., Bishop,S.M., Clarke,A.R. and Bird,A. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*, **297**, 403–405.
13. Wong,E., Yang,K., Kuraguchi,M., Werling,U., Avdievich,E., Fan,K., Fazzari,M., Jin,B., Brown,A.M., Lipkin,M. *et al.* (2002) Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc. Natl Acad. Sci. USA*, **99**, 14937–14942.
14. Rai,K., Huggins,I.J., James,S.R., Karpf,A.R., Jones,D.A. and Cairns,B.R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, **135**, 1201–1212.
15. Thayer,M.M., Ahern,H., Xing,D., Cunningham,R.P. and Tainer,J.A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.*, **14**, 4108–4120.
16. Wu,P., Qiu,C., Sohail,A., Zhang,X., Bhagwat,A.S. 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.
17. Zhang,W., Liu,Z., Crombet,L., Amaya,M.F., Liu,Y., Zhang,X., Kuang,W., Ma,P., Niu,L. and Qi,C. (2011) Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. *Biochem. Biophys. Res. Commun.*, **412**, 425–428.
18. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
19. Ito,S., D'Alessio,A.C., Taranova,O.V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
20. He,Y.F., Li,B.Z., Li,Z., Liu,P., Wang,Y., Tang,Q., Ding,J., Jia,Y., Chen,Z., Li,L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
21. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert

5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
22. Maiti,A. and Drohat,A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.*, **286**, 35334–35338.
23. Cortellino,S., Xu,J., Sannai,M., Moore,R., Caretti,E., Cigliano,A., Le Coz,M., Devarajan,K., Wessels,A., Soprano,D. *et al.* (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, **146**, 67–79.
24. Guo,J.U., Su,Y., Zhong,C., Ming,G.L. and Song,H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423–434.
25. Saparbaev,M. and Laval,J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. USA*, **91**, 5873–5877.
26. Redrejo-Rodriguez,M., Saint-Pierre,C., Couve,S., Mazouzi,A., Ishchenko,A.A., Gasparutto,D. and Saparbaev,M. (2011) New insights in the removal of the hydantoins, oxidation product of pyrimidines, via the base excision and nucleotide incision repair pathways. *PLoS One*, **6**, e21039.
27. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, **26**, 795–800.
28. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
29. Blanc,E., Roversi,P., Vonrhein,C., Flensburg,C., Lea,S.M. and Bricogne,G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2210–2221.
30. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
31. Morgan,M.T., Bennett,M.T. and Drohat,A.C. (2007) Excision of 5-halogenated uracils by human thymine DNA glycosylase. Robust activity for DNA contexts other than CpG. *J. Biol. Chem.*, **282**, 27578–27586.
32. Hashimoto,H., Liu,Y., Upadhyay,A.K., Chang,Y., Howerton,S.B., Vertino,P.M., Zhang,X. and Cheng,X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res*, **40**, 4841–4849.
33. Petronzelli,F., Riccio,A., Markham,G.D., Seeholzer,S.H., Stoerker,J., Genuardi,M., Yeung,A.T., Matsumoto,Y. and Bellacosa,A. (2000) Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. *J. Biol. Chem.*, **275**, 32422–32429.
34. Fromme,J.C., Banerjee,A. and Verdine,G.L. (2004) DNA glycosylase recognition and catalysis. *Curr. Opin. Struct. Biol.*, **14**, 43–49.
35. Manvilla,B.A., Maiti,A., Begley,M.C., Toth,E.A. and Drohat,A.C. (2012) Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. *J. Mol. Biol.*, **420**, 164–175.
36. Zhang,Q.M., Yonekura,S., Takao,M., Yasui,A., Sugiyama,H. and Yonei,S. (2005) DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair*, **4**, 71–79.
37. Hendrich,B. and Bird,A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.
38. Tini,M., Benecke,A., Um,S.J., Torchia,J., Evans,R.M. and Chambon,P. (2002) Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell.*, **9**, 265–277.
39. Cortazar,D., Kunz,C., Selfridge,J., Lettieri,T., Saito,Y., MacDougall,E., Wirz,A., Schuermann,D., Jacobs,A.L., Siegrist,F. *et al.* (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, **470**, 419–423.
40. Kemmerich,K., Dingler,F.A., Rada,C. and Neuberger,M.S. (2012) Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res*, **40**, 6016–6025.