Published in final edited form as:

DNA Repair (Amst). 2013 July ; 12(7): 535–540. doi:10.1016/j.dnarep.2013.04.003.

Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH

Hideharu Hashimoto, Xing Zhang, and Xiaodong Cheng*

Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA

Hideharu Hashimoto: hhashi3@emory.edu; Xing Zhang: xzhan02@emory.edu; Xiaodong Cheng: xcheng@emory.edu

Abstract

The mammalian thymine DNA glycosylase (TDG) excises 5-carboxylcytosine (5caC) when paired with a guanine in a CpG sequence, in addition to mismatched bases. Here we present a complex structure of the human TDG catalytic mutant, asparagine 140 to alanine (N140A), with a 28-base pair DNA containing a G:5caC pair at pH 4.6. TDG interacts with the carboxylate moiety of target nucleotide 5caC using the side chain of asparagine 230 (N230), instead of asparagine 157 (N157) as previously reported. Mutation of either N157 or N230 residues to aspartate has minimal effect on G:5caC activity while significantly reducing activity on G:U substrate. Combination of both the asparagine-to-aspartate mutations (N157D/N230D) resulted in complete loss of activity on G:5caC while retaining measurable activity on G:U, implying that 5caC can adopt alternative conformations (either N157-interacting or N230-interacting) in the TDG active site to interact with either of the two asparagine side chain for 5caC excision.

Keywords

5-Carboxylcytosine; Thymine DNA glycosylase; DNA modification; DNA 5mC oxidation; Epigenetic regulation

1. Introduction

The control of gene expression in mammals relies in part on the modification status of DNA cytosine residues, which exist in at least five forms, cytosine (C), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [1–6]. Three DNA methyltransferases, Dnmt1, Dnmt3a and Dnmt3b, methylate cytosine in the context of CpG dinucleotides, generating 5mC in the genome [7,8]. Three ten-eleven translocation (Tet) proteins convert 5mC to 5hmC, 5fC and 5caC in three consecutive oxidation reactions [9–12]. While 5mC and 5hmC occur in almost all human tissues and cell types examined [2], 5hmC is relatively enriched in embryonic stem (ES) cells [9] and brain [1,13]. The genomic levels of 5fC and 5caC may also be similarly enriched in ES cells and brain, because both are derived from 5hmC oxidation. The exact functions of these oxidative cytosine bases are under investigation.

The mammalian thymine DNA glycosylase (TDG) is involved in active DNA demethylation through the removal of either deamination products of 5mC and 5hmC (G:T and G:5hmU mismatches) or its oxidized derivatives (G:5fC and G:5caC) via the base excision repair pathway [12,14–16]. Human TDG catalytic domain (residues 111–308) has been crystallized in complex with DNA containing various mismatches [17–19], including an A:5caC mismatch or a modified 5caC (with a 2'-fluoro substitution on the deoxyribose of 5caC) paired with G [20]. The 2'-fluoro substitution and the 5caC base pairing nucleotide (A or G) have little effect on the 5caC base conformation in the active site pocket of TDG [20]. In the structure of TDG with a 22-base pair (bp) oligonucleotide containing A:5caC crystallized under the condition of pH 7 [20], the 5-carboxyl moiety of 5caC forms a weak hydrogen bond with the side chain amino group of Asn157. We refer to this conformation as the N157-interacting conformation. Surprisingly, the asparagine-to-alanine (N157A) mutant only slightly reduces activities on G:5caC and G:U, whereas the asparagine-to-aspartate (N157D) mutant retains wild type activity on 5caC while abolishing excision on U, becoming highly selective for 5caC substrate [21]. Furthermore, both wild type (WT) TDG and N157D mutant exhibited higher activities for 5caC at lower pH values (5.5–6.0) [19,21], suggesting that increased protonation of the carboxylate of 5caC and the aspartate facilitates base excision.

Previously, we reported a post-reactive complex structure of WT TDG catalytic domain with a 28-bp DNA containing a G:5hmU (5-hydroxymethyluracil) mismatch with the cleaved 5hmU base remaining in the active-site binding pocket [19]. Here we report a structure of N140A mutant (nearly catalytically inert) complexed with the same 28-bp oligonucleotide containing a G:5caC under the condition of pH 4.6 (Table 1). The 5caC, which is flipped out from the double-stranded DNA, is not in the N157-interacting conformation as previously described [20], rather it interacts with Asn230 in the active site. We refer to this second conformation as the N230-interacting conformation. Similar to N157D mutant, the Asn230-to-aspartate (N230D) mutation has minimal effect on G:5caC activity but causes much greater reduction on G:U activity. In contrast, double mutations of N157D/N230D abolishes the activity on G:5caC, while retaining measurable activity on G:U. We suggest that 5caC can adopt two conformations interacting with either Asn157 or Asn230 for efficient base excision.

2. Materials and methods

2.1. Protein purification

Human TDG catalytic domain residues 111–308 (pXC1056) and its mutants N140A (pXC1057), N140D (pXC1105), N230D (pXC1113), N157A (pXC1155), N157D (pXC1138), N157D/N230D (pXC1154) and N140D/N157D (pXC1229) were prepared as previously described [19].

2.2. Crystallography

TDG N140A mutant protein (0.35 mM) was mixed with 0.2 mM of annealed 28-bp oligonucleotide (synthesized by the New England Biolabs, Inc.): 5'-CAG CTC TGT ACG TGA GCG ATG GAC AGC T-3' and 5'-AGC TGT CCA TCG CTC AXG TAC AGA GCT G-3' where X is 5caC. The 5-carboxy-deoxyC phosphoramidites used for DNA synthesis were purchased from Glen Research. Small crystals appeared within 24 h under the conditions of 30% polyethylene glycol (PEG) 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 4.6. After 3 days, crystals were flash-frozen in liquid nitrogen after soaking in mother liquor supplemented with 20% ethylene glycol. X-ray diffraction datasets at the wavelength of 1.0 Å were collected at the SER-CAT beamline (22ID-D) of the Advanced Photon Source, Argonne National Laboratory and processed using HKL2000 [22]. Programs

Coot [23] and PHENIX [24] were used for model building and refinement. The R_{free} and R_{work} values were calculated for 5% (randomly selected) and 95%, respectively, of observed reflections (Table 1).

2.3. DNA glycosylase activity assay

The excision reaction was monitored, as previously described [19], by denaturing gel electrophoresis following NaOH hydrolysis of the a basic site under the single turnover condition ($[S_{\text{DNA}}] = 0.25 \mu\text{M}$ and $[E_{\text{TDG}}] = 2.5 \mu\text{M}$) in 0.1% BSA, 1 mM EDTA, 100 mM NaCl and 10 mM BisTris HCl, pH 6.0, with various reaction times and temperatures (see below). Various 32 bp oligonucleotides labeled with 6-carboxy-fluorescein (FAM) were used as substrates: (FAM)-5'-TCG GAT GTT GTG GGT CAG XGC ATG ATA GTG TA-3' (where X = C, 5mC, 5hmC, 5fC, 5caC, U, T and 5hmU) and 5'-TAC ACT ATC ATG CGC TGA CCC ACA ACA TCC GA-3'.

For the substrate profiling of eight oligonucleotides (left panels of Fig. 2), 10 min reaction time at 30 °C was used. For activities on G:U substrate (middle panels of Fig. 2), reaction temperature 16 °C were chosen because the fast reaction rate of WT TDG at 30 °C does not allow us to accurately measure the reaction rate. The time courses (0–300 s for WT, 0–30 min for N157A mutant, and 0–8 h for the other mutants) were performed. Additional time points (24–72 h) were also taken for N140A, N157D and N140D/N157D mutants. For G:5caC substrate (right panels of Fig. 2), the time courses (0–30 min) at 30°C were completed. The intensities of the FAM labeled DNA were determined by Typhoon Trio+, and quantified by the image-processing program ImageJ. The data were fitted to non-linear regression using software GraphPad PRISM 5.0d (GraphPad Software Inc.): $[\text{Product}] = P_{\text{max}}(1 - e^{-kt})$, where P_{max} is the product plateau level, k (h^{-1}) is the observed rate constant, and t is the reaction time.

3. Results

3.1. Structure of DNA containing 5caC bound to TDG

TDG belongs to a superfamily of monofunctional DNA glycosylases including uracil-specific DNA glycosylase (UDG), mismatch-specific uracil glycosylase (MUG), and single-strand specific monofunctional uracil glycosylase (SMUG1) (reviewed in Refs. [25–27]). While UDG uses an aspartate (Asp145 in human UDG and Asp64 in *E. coli* UDG) as the catalytic residue [28], human TDG (Asn140) [17], *E. coli* MUG (Asn18) [29], and *Xenopus laevis* SMUG1 (Asn96) [30] use an asparagine in the corresponding position. We used a catalytic inert mutant Asn140-to-Ala (N140A) of TDG [19,31,32] crystallized with a 28-bp oligonucleotide containing a G:5caC within a CpG site (Fig. 1a). We determined the structure to the resolution of 2.6 Å (Table 1).

The structure was solved by molecular replacement using the coordinates of the post-reactive complex formed by WT TDG and G:5hmU DNA (PDB 4FNC) as the search model. We deleted the coordinates of the side chain of Asn140 to be an alanine and removed the cleaved 5hmU base in the search model. The resulting structure includes TDG residues 111–305 and all 28 base pairs of DNA (Fig. 1a). The extra density in the omit 2Fo-Fc and Fo-Fc difference Fourier maps (without the 5caC base in the Fc calculation) in the active site accounts for the 5caC base (Fig. 1b). Although we expected an intact 5caC nucleotide to be present in the structure as N140A was thought to be catalytically inactive, we were quite surprised to find the partially broken difference electron density for the *N*-glycosidic bond connecting the sugar C1' atom and the base N1 atom. In addition, there is extra electron density near the sugar C1' atom indicating the possible presence of an attached hydroxyl oxygen (Fig. 1b), similar to the abasic sugar conformation after the cleavage of 5hmU base (Fig. 1c) [19]. Both observations suggest that partial base excision occurred in the crystal

after a considerably long time in the crystal environment. Unfortunately, we were unable to position one single 5caC conformation accurately into the density, probably due to a mixture of substrates and products for which the resolution of the current structure does not permit us to distinguish. What depicted in Fig. 1b represents an averaged conformation of TDG-DNA substrate and product complexes in the crystal.

A previous study indicated that N140A mutant has no detectable base excision activity for a G:T substrate, and its excision rate is vastly diminished (by approximately $10^{4.4}$ -fold) for other mismatches (including G:U) at pH 7.5 [32]. We were able to detect residual activity of N140A at pH 6.0 after 24 h for G:U and 72 h for G:5caC (Fig. 2b) at 16°C (the temperature and time used for crystallization), consistent with the finding that the N140A mutant can slowly react on 5caC at an even lower pH (4.6) of the crystallization condition, resulting in a mixture of substrates and products.

In a pre-reactive WT TDG complex containing 2'-deoxy-2'-fluoroarabinouridine, a mimic of deoxyuridine that is non cleavable by TDG [18], a putative nucleophilic water molecule is held in position by the side-chain carbonyl oxygen atom of Asn140 and the backbone carbonyl oxygen of Thr197 (Fig. 1d). Furthermore, a network of hydrogen bonds involving the side chain hydroxyl oxygen of Thr197 and the backbone carbonyl oxygen of Arg195 coordinates the side chain amide nitrogen atom of Asn140 (Fig. 1d). No such water molecule was found in the corresponding position of the current structure where the C1' hydroxylation might have already occurred (Fig. 1d). Mutations that disrupt the hydrogen bonding network and thus affect the positioning of the nucleophilic water resulted in significant loss of activity: the T197A mutant is approximately 30-fold less active on a G:T substrate [18] and the N140D mutant has approximately a 100 fold reduction of activity on a G:U substrate and no measurable activity on G:T (Fig. 2c). The absence of the Asn140 side chain and the resulting movement of the side chain of Thr197 (Fig. 1d) render the N140A mutant unable to properly position the water nucleophile for attacking the C1' carbon, thus accounting for its extremely diminished activity.

Two general reaction mechanisms for *N*-glycoside hydrolysis have been proposed [26] (Fig. 1e). In a S_N2 mechanism, nucleophile approach and leaving group departure are precisely synchronous. In a S_N1 mechanism, reactions could proceed through a discrete oxacarbenium ion intermediate formed after leaving group departure and before nucleophile approach, as in the case of *E. coli* UDG [33]. In *E. coli* UDG, His187 forms a hydrogen bond to the uracil O2 atom [34], which facilitates the departure of the leaving uracil base, and Asp64 stabilizes the cationic sugar intermediate electrostatically [33]. Neither His187 nor Asp64 of *E. coli* UDG is conserved in TDG. Moreover, replacing Asn140 of TDG to aspartate (N140D) – that could potentially better stabilize the positively charged C1' oxacarbenium ion intermediate in S_N1 mechanism – caused significant reduction of activity instead (Fig. 2c), suggesting that TDG employs a different reaction mechanism from that of *E. coli* UDG. We also note that the electron density does not fully mask the exocyclic O2 and N4 atoms of 5caC (Fig. 1b), probably due to the heterogeneous nature of the base-sugar bond undergoing various stages of breakage (i.e., resulting from a mixture of substrates and products).

3.2. N230 interacts with the carboxylate groups of 5caC

One interesting finding is that the everted and “partially” cleaved 5caC ring is not in the N157-interacting conformation as previously described [20]. It wedges into the space between Gly138-Ile139-N140A on one side and Tyr152 on the other side and no interaction with Asn157 is observed (Fig. 1b and f). The carboxylate group at the C5 position is within hydrogen bonding distance to the side chains of Ser271 and Asn230 (Fig. 1g). We refer to the current conformation as the N230-interacting conformation. The crystallographic thermal B-factor of 5caC ($54 \text{ \AA}^2$ for the deoxyribose and $60 \text{ \AA}^2$ for the base) is comparable

to that of the neighboring nucleotides ($50 \text{ \AA}^2$ for the 5' adenine and $46 \text{ \AA}^2$ for the 3' guanine) and the interacting protein residues ($45 \text{ \AA}^2$ for Gly138, $45 \text{ \AA}^2$ for Ile139, $52 \text{ \AA}^2$ for N140A, $51 \text{ \AA}^2$ for Tyr152 and $51 \text{ \AA}^2$ for Asn157).

Structure-based sequence alignments previously suggested that Ser271 of TDG occupies an equivalent position of the uracil-O2 interacting histidine residue (His187 of *E. coli* UDG or His268 of human UDG) [19,26]. However, mutating Ser271 to histidine (S271H) or alanine (S271A) does not affect substrate specificity [19]. We thus mutated Asn230 to alanine (N230A) or aspartate (N230D). N230A mutant protein is not stable, perhaps due to the loss of an intra-molecular hydrogen bonding network involving Ser272-Asn230-Tyr152 [19], thus is not included in further analysis. We compared the glycosylase activity of N230D (Fig. 2d) to that of wild type (WT) (Fig. 2a), single-point mutations of N140A and N140D (Fig. 2b and c), N157A and N157D (Fig. 2e and f), as well as two double mutants of N157D/N230D and N140D/N157D (Fig. 2g and h). Various 32-base-pair (bp) DNA oligonucleotides were used, each containing a single modified base X (=C, 5mC, 5hmC, 5fC, 5caC, U, T and 5hmU) within a G:X pair in a CpG sequence (Fig. 2, left panels). All mutations abolished G:T and G:5hmU activities under the assay condition of 10 min at 30 °C, probably because a G:T mismatch is a weak substrate even for WT TDG [35], but the effects of the mutations on G:U and G:5caC substrates are quite varied among the mutants. We therefore carried out single turnover kinetic experiments on G:U (Fig. 2, middle panels) and G:5caC substrates (Fig. 2, right panels) and compared the k_{obs} of the mutants to that of the WT (Fig. 2i).

We have previously described N157D as a mutant that selectively excises 5caC [21]. The selectivity is the result of drastic reduction of activity (by a factor of 1350) on G:U (to the same extent as the catalytic N140A mutant) while maintaining nearly full activity on 5caC (Fig. 2f). Similarly, N230D mutation also causes more severe impairment on G:U activity (by a factor of 32) than on G:5caC (by a factor of 3). Interestingly, mutating both asparagines (N157D/N230D) resulted in no detectable activity on G:5caC while regaining some of the residual activity on G:U (Fig. 2g) compared with N157D mutation alone (Fig. 2f). Thus, it appears that the carboxylate group of 5caC base must interact with either N157 or N230 for the excision to occur. Superimposition of the structure of N140A complexed with A:5caC (PDB 3UO7) or a modified 5caC paired with G [20] with the current N140A structure with G:5caC (this study) suggests that 5caC base can adopt two conformations (either interacting with Asn157 or Asn230) by simply rotating approximately 120° along the *N*-glycoside bond (Fig. 1h). Such alternative 5caC conformations in two different “environments” of the TDG active site would explain the fact that single mutation of N157D or N230D has little effect on 5caC excision, while N157D/N230D double mutations resulted in complete loss of activity.

4. Discussion

Among the human TDG structures examined so far, the same catalytic domain (residues 111–308) was used in complex with two different lengths of DNA oligonucleotides. One is a 22-bp DNA with one 3'-overhanging adenine or thymine, and the other is a 28-bp blunt-ended DNA. The 22-bp DNA was crystallized with two protein molecules (1:2 complex) in pH 7–7.5 [17,18,20], whereas the 28-bp DNA was crystallized with one protein molecule (1:1 complex) in pH 4.6 [19] (this study). The diffraction quality of the 1:2 complex crystals varied significantly and required screening of many crystals to achieve $\sim 3 \text{ \AA}$ resolution [17]. For example, crystals with the 22-bp DNA containing either an A:5caC mismatch or a modified 5caC paired with G diffracted asymmetrically to $3 \text{ \AA}$ along the *a* and *b* axes and $4 \text{ \AA}$ along the *c* axis [20]. However, “a clear electron density map of 5caC in the pocket” was observed [20]. In the case of the 1:1 complex with the 28-bp DNA, crystals consistently

reached the modest resolutions of 2.3–2.6 Å [19] (this study). Ironically, while the electron density map is clear for the protein residues (particularly in the active-site pocket) and all of the 28-bp DNA (Fig. 1a), the electron density is unclear for the 5caC (Fig. 1b and f). We attribute this observation to the slow ongoing catalysis of N140A (Fig. 2b) in the crystal environment of pH 4.6.

Several factors could contribute to the alternative 5caC conformations described here. First, the lengths of DNA (22-bp versus 28-bp) in the complexes are different. However, the ten nucleotides flanking either side of the 5caCpG site are identical for both DNA oligonucleotides. TDG makes phosphate contacts spanning only five base pairs surrounding the 5caC (two 5'- and three 3'-phosphate groups), thus it is unlikely for the DNA length difference beyond the contacted nucleotides to affect the 5caC conformation. Second, the two previously described 5caC structures [20] have either a 2'-fluoro substitution on the deoxyribose or an A:5caC mismatch compared to this current study, yet both structures have similar 5caC base conformations (the N157-interacting) that are distinct from what we observed here (the N230-interacting conformation). Very unlikely that the lack of both 2'-fluoro substitution and A:5caC mismatch could result in the alternative 5caC conformation. Third, we believe the difference in crystallization conditions (particularly pH 4.6 versus pH 7 or above) is the main reason for alternative 5caC conformations. The pH values correlate inversely with the activity of TDG on 5caC substrate, explaining the partial cleavage of the base-sugar bond in the crystal as well as the measurable activity of N140A in solution at acidic pH (Fig. 2b). As mentioned above, the alternative conformation described here only represents an average conformation of the mixture of products and substrates in the crystal. Nevertheless, this N230-interacting conformation is clearly distinct from the N157-interacting conformation as previously described [20]. Whether it represents the reactive form in general or only at acidic pH is not clear.

UDGs and the family of Helix-hairpin-Helix (HhH) DNA glycosylases including MBD4 require an aspartate as the catalytic residue [25,27,36,37]. Interestingly there is no acidic residue in the active site of TDG; instead it contains four asparagine residues (140,157, 191 and 230). We were surprised to find that replacing the catalytic Asn140 with aspartate (N140D) has similar effect to that of N157D and N230D substitution, in that it causes significantly more severe reduction of G:U activity (by a factor of 96) than G:5caC activity (by a factor of 5) (Fig. 2c). The double mutation of N140D/N157D completely abolished activities on both G:U and G:5caC even after 3 days of incubation at 16 °C (Fig. 2h), and is even more inert than the catalytic mutant N140A (Fig. 2b). It is still unclear, however, why single mutations, particularly N157D, would cause such drastic reduction of G:U activity, as a uracil base can be accommodated, just as 5caC, in both conformations. Furthermore, it is also puzzling why N157A mutant only has relatively minor effects on activities of G:5caC (3-fold reduction) and G:U (6-fold reduction), respectively (Fig. 2i). Additional studies are needed to understand the function of these asparagine residues unique to the TDG subfamily of glycosylases.

Acknowledgments

We thank sincerely Brenda Baker of New England Biolabs for DNA oligo synthesis, and Dr. John R. Horton for comments on the manuscript. H.H. performed all experiments, X. Z. and X.C. organized and designed the scope of the study, and all were involved in analyzing data and preparing the manuscript. U.S. National Institutes of Health (GM049245-19) funded the study. X.C. is a Georgia Research Alliance Eminent Scholar.

References

1. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009; 324:929–930. [PubMed: 19372393]

2. Globisch D, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*. 2010; 5:e15367. [PubMed: 21203455]
3. Booth MJ, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*. 2012; 336:934–937. [PubMed: 22539555]
4. Yu M, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 2012; 149:1368–1380. [PubMed: 22608086]
5. Raiber EA, et al. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol*. 2012; 13:R69. [PubMed: 22902005]
6. Sun Z, et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep*. 2013; 3:567–576. [PubMed: 23352666]
7. Bestor T, et al. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol*. 1988; 203:971–983. [PubMed: 3210246]
8. Okano M, et al. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*. 1998; 19:219–220. [PubMed: 9662389]
9. Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324:930–935. [PubMed: 19372391]
10. Ito S, et al. Role of Tet proteins in 5mC to 5hmC conversion ES-cell self-renewal and inner cell mass specification. *Nature*. 2010; 466:1129–1133. [PubMed: 20639862]
11. Ito S, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 2011; 333:1300–1303. [PubMed: 21778364]
12. He YF, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011; 333:1303–1307. [PubMed: 21817016]
13. Khare T, et al. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat Struct Mol Biol*. 2012; 19:1037–1043. [PubMed: 22961382]
14. Hajkova P, et al. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science*. 2010; 329:78–82. [PubMed: 20595612]
15. Guo JU, et al. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*. 2011; 145:423–434. [PubMed: 21496894]
16. Cortellino S, et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*. 2011; 146:67–79. [PubMed: 21722948]
17. Maiti A, et al. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci USA*. 2008; 105:8890–8895. [PubMed: 18587051]
18. Maiti A, et al. Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc Natl Acad Sci USA*. 2012; 109:8091–8096. [PubMed: 22573813]
19. Hashimoto H, et al. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res*. 2012; 40:10203–10214. [PubMed: 22962365]
20. Zhang L, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol*. 2012; 8:328–330. [PubMed: 22327402]
21. Hashimoto H, et al. Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *J Mol Biol*. 2013; 425:971–976. [PubMed: 23337108]
22. Otwinowski Z, et al. Multiparametric scaling of diffraction intensities. *Acta Crystallogr A*. 2003; 59:228–234. [PubMed: 12714773]
23. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D: Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
24. Adams PD, et al. PHENIX: a comprehensive Python-based system for macro-molecular structure solution. *Acta Crystallogr D: Biol Crystallogr*. 2010; 66:213–221. [PubMed: 20124702]

25. Hitomi K, et al. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst)*. 2007; 6:410–428. [PubMed: 17208522]
26. Berti PJ, McCann JA. Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chem Rev*. 2006; 106:506–555. [PubMed: 16464017]
27. Huffman JL, et al. DNA base damage recognition and removal: new twists and grooves. *Mutat Res*. 2005; 577:55–76. [PubMed: 15941573]
28. Slupphaug G, et al. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*. 1996; 384:87–92. [PubMed: 8900285]
29. Barrett TE, et al. Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*. 1998; 92:117–129. [PubMed: 9489705]
30. Wibley JE, et al. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol Cell*. 2003; 11:1647–1659. [PubMed: 12820976]
31. Hardeland U, et al. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem*. 2000; 275:33449–33456. [PubMed: 10938281]
32. Maiti A, et al. Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase. *J Biol Chem*. 2009; 284:36680–36688. [PubMed: 19880517]
33. Werner RM, Stivers JT. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: evidence for an oxocarbenium ion-uracil anion intermediate. *Biochemistry*. 2000; 39:14054–14064. [PubMed: 11087352]
34. Drohat AC, Stivers JT. Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry*. 2000; 39:11865–11875. [PubMed: 11009598]
35. Drohat AC, et al. How a mismatch repair enzyme balances the needs for efficient lesion processing and minimal action on undamaged DNA. *Cell Cycle*. 2012; 11:3345–3346. [PubMed: 22918241]
36. Hashimoto H, et al. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res*. 2012; 40:8276–8284. [PubMed: 22740654]
37. Morera S, et al. Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Res*. 2012; 40:9917–9926. [PubMed: 22848106]

Abbreviations

TDG	thymine DNA glycosylase
5caC	5-carboxylcytosine

Fig. 1.

Structure of TDG N140A mutant in complex with G:5caC containing DNA: (a) 2Fo-Fc electron density, contoured at 1σ above the mean, for the entire 28-bp DNA (blue) and human TDG glycosylase domain (green) used in the structure determination; (b) an enlarged active site with the 5caC base in the middle and TDG active site residues in green. The omit electron densities, after omitting the 5caC base, are shown as 2Fo-Fc in 1σ (gray) and Fo-Fc in 5σ (blue). The Fo-Fc in 3σ shows a similar 5caC envelope as that of 2Fo-Fc (not shown). The modeled 5caC base conformation represents an averaged conformation of TDG-DNA substrate and product complexes in the crystal, with the distance between the C1' and N1 atoms at approximately 2.3 Å, obviously longer than the *N*-glycosidic bond distance of 1.5 Å; (c) superimposition of a post-reactive complex with an a basic sugar with a hydrolyzed C1' (colored in cyan; PDB 4FNC) [19]; (d) superimposition of a pre-reactive complex (colored in salmon pink; PDB 3UFJ) [18] shows a putative nucleophilic water molecule (in black), held in position by the side chain carbonyl oxygen atom of Asn140, attacking the C1' from the opposite side of the leaving base. The numerical numbers indicate relative movement (in Angstroms) of residue Thr197 and the water molecule. The omit Fo-Fc electron density is contoured in 5σ (blue); (e) two possible catalytic mechanisms for TDG: a concerted, synchronous S_N2 mechanism (top) and a dissociated (asynchronous) S_N1 mechanism proceeding through a discrete transition state (bottom). The reactant (left) and the product (right) are shown. The structural feature for the reactant is adopted from a pre-reactive complex of WT TDG with a uracil analog (2'-deoxy-2'-fluoroarabinouridine) (PDB 3UFJ) [18]. The structural feature for the product is adopted from a post-reactive complex of WT TDG with a G:5hmU mismatch (PDB 4FNC) that generated a C1' hydrolyzed a basic sugar [19]. *E. coli* UDG follows a discrete oxocarbenium ion intermediate (bottom) [33]. The question mark signals a different reaction mechanism employed by TDG; (f) hydrophobic interactions between the 5caC base and Gly138 and Ile139; (g) the leaving 5caC base interacts with side chains of Ser271, Asn230 and Tyr152. The numerical numbers indicate hydrogen bond distances in Angstroms; (h) superimposition of N140A in complex with a 5caC-containing 22-bp DNA crystallized in pH 7 (colored in yellow; PDB 3UO7) [20] and 5caC-containing 28-bp DNA crystallized in pH 4.6 (colored in

green and gray; this study) suggests a base rotation of approximately 120° around the glycoside bond.

Fig. 2. Mutations of three asparagine residues in the active site of TDG; the activities of TDG wild type (WT) (a), mutants N140A (b), N140D (c), N230D (d), N157A (e), N157D (f), N157D/N230D (g), and N140D/N157D (h). (i) Summary of relative activities (by fold reduction from that of WT) of mutants for G:U and G:5caC substrates (NA: no activity observed).

Table 1

Data collection and refinement statistics.

Protein	Human TDG (111–308) N140A mutant
DNA	28-bp G:5caC
Data collection	
Space group	C2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	91.5, 53.6, 81.9
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 95.1, 90
Resolution (Å)	45.56–2.59 (2.68–2.59)
<i>R</i> _{merge}	0.077 (0.365)
$\langle I \rangle / \sigma(I)$	15.7 (2.0)
Completeness (%)	99.9 (97.4)
Redundancy	4.0 (2.8)
Unique reflections	11,604 (922)
Observed reflections	46,994
Refinement	
Resolution (Å)	2.59
No. reflections	11,568
<i>R</i> _{work} / <i>R</i> _{free}	0.229/0.270
No. atoms	
Protein	1535
DNA	1144
Water	29
<i>B</i> factors (Å ²)	
Protein	63.5
DNA	65.6
5caC	57.4
Water	51.1
r.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.8

Values in parentheses are for the highest-resolution shell.