Published in final edited form as:

J Mol Biol. 2013 March 25; 425(6): 971–976. doi:10.1016/j.jmb.2013.01.013.

Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant

Hideharu Hashimoto, Xing Zhang, and Xiaodong Cheng

Departments 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 the mismatched base, uracil, thymine, or 5-hydroxymethyluracil (5hmU), as well as removes 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) when paired with a guanine. In the previously solved structure of TDG in complex with DNA containing 5caC, the side chain of asparagine 157 (N157) contacts the 5-carboxyl moiety of 5caC via a weak hydrogen bond. We examined the role of N157 in recognition of 5caC by mutagenesis. The asparagine-to-alanine (N157A) mutant has no detectable base excision activity for a G:T mismatch, and its excision activity is reduced for other substrates including G:5caC. Unexpectedly, the asparagine-to-aspartate (N157D) mutant has a comparable base excision rate for G:5caC substrate to that of wild type, but it only has residual activity for G:U and no detectable activity for other substrates. We further show that the N157D mutant has higher activity for 5caC at a lower pH (6.0), suggesting that increased protonation of the carboxylate of 5caC and the aspartate facilitates base excision. The N157D mutant remains highly specific for 5caC even in the presence of large excess of genomic DNA, a property that can potentially be used for mapping the very low amount of 5caC in genomes.

Keywords

5-carboxylcytosine; thymine DNA glycosylase; DNA modification; DNA 5mC oxidation; epigenetic regulation

Mammalian DNA cytosine modification is a dynamic process and occurs by converting cytosine (C) to 5-methylcytosine (5mC), established by specific DNA methyltransferases, and then to 5-hydroxymethylcytosine (5hmC) by ten eleven translocation (Tet) proteins. Tet proteins can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The mammalian thymine DNA glycosylase (TDG) involves in active DNA demethylation through the removal of deamination products of 5mC (G:T mismatches) or its oxidized derivatives (G:5hmU mismatches) via the base excision repair pathway. Of particular interest are recent reports that TDG can excise 5fC and 5caC (but not 5hmC and 5mC) from DNA. This new specificity of TDG suggests a deamination-independent active DNA demethylation pathway through Tet-mediated oxidation of 5hmC.

© 2012 Elsevier Ltd. All rights reserved.

Correspondence should be addressed to: Xiaodong Cheng (xcheng@emory.edu; phone: 404-727-8491; fax: 404-727-3746).

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

Human TDG catalytic domain (residues 111–308) has been crystallized in complex with DNA containing an abasic analog (tetrahydrofuran)¹¹, a uracil analog (2'-deoxy-2'-fluoroarabinouridine)¹², an A:5caC mismatch or a modified 5caC (with a 2'-fluoro substitution on the deoxyribose of 5caC) paired with G¹³, and a post-reactive complex containing a G:5hmU mismatch with the cleaved 5hmU base remaining in a binding pocket of the enzyme¹⁴. Among the structures of TDG examined so far, the side chain of Asn157 interacts with the 5' phosphate of the target nucleotide (Fig. 1a-c), which is flipped out from the double stranded DNA. In the structure of TDG with A:5caC¹³, the 5-carboxyl moiety of 5caC forms a weak hydrogen bond with the side chain amino group of Asn157 (Fig. 1a). Here we study the role of Asn157 in recognition of 5caC by mutagenesis.

N157A does not affect substrate specificity

We first mutated Asn157 to alanine (N157A) and compared its glycosylase activity to that of wild type (WT) using various 32-base-pair (bp) DNA oligonucleotides, 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. As expected, glycosylase activity was not observed with TDG on oligonucleotides bearing the 'natural' G:C, G:5mC and G:5hmC base pairs, while substrates bearing G:T, G:U and G:5hmU mismatches and G:5fC and G:5caC 'wobble' pairs¹⁴ were efficiently cleaved by the WT enzyme (Fig. 1d). Instead of losing activity on G:5caC as we had expected, N157A retains reduced but measurable activity on G:U, G:5hmU, G:5fC and G:5caC and no detectable activity on G:T substrate (Fig. 1e). G:T mismatch is known to be a weak substrate even for WT TDG¹⁵ (see Fig. 1d, lane T). Further reduction of this weak activity could render the N157A inactive on G:T substrate.

N157D selectively excises 5caC

Mutation of Asn157 to aspartate (N157D) resulted in loss of detectable activity on G:T and G:5hmU and only residual activity on G:U and G:5fC (Fig. 1f). Surprisingly, the N157D mutant remains fully active on G:5caC substrate (Fig. 1f and 2a). The activity of N157D mutant on G:5caC substrate is very sensitive to pH (Fig. 2b) and salt (Fig. 2c).

We measured the mutant N157D activities on G:5caC substrate under single turnover conditions (i.e. $[E_{TDG}] \gg [S_{DNA}]$) at a pH range between 6.0 and 7.2 with an interval of 0.4 units. We found that the mutant TDG has higher activity by a factor of approximately 2 for the G:5caC at pH 6.0 ($k_{obs}=0.85 \text{ min}^{-1}$) as compared to the activity at pH 6.4 ($k_{obs}=0.4 \text{ min}^{-1}$) (Fig. 2b). Changing the pH to 6.8 weakened the activity by another factor of 2.5 with $k_{obs}=0.16 \text{ min}^{-1}$. Further increasing the pH to 7.2 resulted in accumulative loss of activity by a factor of 17 ($k_{obs}=0.05 \text{ min}^{-1}$), as compared to the activity at pH 6.0. Like WT TDG¹⁴, the continuing enhancement of activity on G:5caC at lower pH values suggests that the chemical nature of the target nucleotide and the aspartate side chain can contribute to the mutant being active on 5caC. We reasoned that the protonated forms of 5caC and/or the aspartate side chain at lower pH could maintain the interaction between the DNA phosphate and N157D as well as the interaction between N157D and 5caC.

We note that TDG activity on the G:U substrate is relatively constant between pH 5.5–9¹⁶, with a slope of 0.1 in a log plot of activity against pH (Fig. 2d, line 1). The enhancement of WT activity on G:5caC at lower pH suggests that the increasing protonation of the carboxylate group of the target nucleotide, as indicated by the linear fits under two different buffer systems (Fig. 2d, lines 2 and 3), can contribute to catalysis. The slope is even steeper for N157D mutant (1.0 vs. 0.5 for WT; Fig. 2d, line 4), suggesting that the protonation of the target nucleotide 5caC and the aspartate side chain both contribute to catalysis. However, the exact values of the slopes do not exactly match the proposed number of protonation events (one for WT and two for N157D). This could be complicated by other residues in the active

site, particularly His151¹², whose protonation might negatively impact catalysis. Additional study will be required to settle this point.

N157D maintains its specificity for 5caC in the presence of large excess of genomic DNA

Genomic 5fC and 5caC contents are very low, approximately 2×10^{-5} and 3×10^{-6} per cytosine, respectively, compared to 1.3×10^{-3} of 5hmC, in embryonic stem cells⁵. A separate study estimated the ratio of 5fC to be about 10 fold higher (2×10^{-4}), although the amount of 5caC was not measured¹⁷. One potential application of the N157D mutant is in mapping genomic 5caC sites. The product of TDG reaction is abasic nucleotide, which can be readily selected and analyzed from the rest of the DNA using biotin containing Aldehyde Reactive Probe (ARP)¹⁸.

We tested whether N157D protein can maintain its specificity for 5caC in the presence of large excess of genomic DNA. An excision reaction (200 μ l) was performed with 2.5 nM of 5caC in FAM-labeled oligonucleotide in the presence of 100 μ g of salmon sperm DNA (with the ratio of 5caC to C to be $\sim 7 \times 10^{-6}$). When 25 μ M of N157D was used, 5caC cleavage was near completion after 10 min, while little cleavage of U had occurred (Fig. 3a), confirming that the specificity of N157D was not affected in the presence of large amount of genomic DNA. We noted, however, large excess of N157D enzyme was required to rapidly cleave 5caC (Fig. 3b), likely due to nonspecific DNA binding by N157D while searching for its target.

Discussion

Of the three oxidized forms of 5mC generated by Tet proteins, 5fC and 5caC are present in much lower amounts than that of 5hmC. The low levels of 5fC and 5caC are likely due to the observation that both bases can be excised by TDG and repaired by a base excision process. However, a TDG-knockout shows only a minor increase of 5caC in mouse embryonic stem cells⁶. Another scenario is that only a subset of 5hmC is oxidized to 5fC and even a smaller amount of 5fC becomes 5caC, due to a much slower rate of conversion from 5hmC to 5fC and 5fC to 5caC than that of 5mC to 5hmC. *In vitro* analysis of Tet2 suggested that this is indeed the situation⁵. To understand the function and regulation mechanism that control the levels and genomic distribution of 5fC and 5caC, we need to know the genomic locations of 5fC and 5caC accurately, preferably at single-base resolution.

Asn157 of TDG was initially identified as the residue making a specific weak interaction with 5caC base from a low-resolution structure, which was determined by X-ray diffraction data to 3 Å resolution along the *a* and *b* axes and asymmetrically to 4 Å along the *c* axis¹³. Unexpectedly, the N157D mutant became specific for 5caC among the substrates examined, a property that can potentially be used for mapping the very low amount of 5caC in genomes.

To successfully map 5caC, genomic DNA should be first treated with uracil-specific DNA glycosylase (UDG) to remove any G:U mismatches that can be substrate for the TDG N157D mutant, albeit very slowly. Subsequently, the DNA must be treated with NaBH₄ (0.1%) to eliminate traces of potentially reactive aldehyde including endogenous abasic sites and product of UDG, before using N157D protein to convert 5caC to abasic sites. DNA fragments containing abasic sites will bind to Aldehyde Reactive Probe (ARP), be selected by avidin bead¹⁸ and subjected to sequencing analysis. One must note that abasic site will somewhat hinder amplification by DNA polymerase, resulting in incorporation of dATP at

the abasic site and occasionally cause one or two base pairs deletion¹⁹, both of which can be used as an indication of 5caC presence in the original DNA. Using wild type TDG by the same method, both 5fC and 5caC sites could be mapped.

TDG belongs to the family of DNA glycosylases with similar architectures including UDG, mismatch-specific uracil glycosylase (MUG), and single-strand specific monofunctional uracil DNA glycosylase (SMUG1) (reviewed in). Like TDG, *Escherichia coli* MUG possesses an Asn35, which corresponds to Asn157 of TDG¹⁴, and is capable of excising G:5caC in addition to G:U mismatch. In addition, SMUG1 excises U and 5hmU²³, but like UDG and MUG is inactive against thymine. This difference of substrate specificity within the family members has yet to be studied structurally, (bio)chemically and thermodynamically.

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. The authors thank an anonymous reviewer for the suggestion of a log plot of activity against pH.

References

1. Globisch D, Munzel M, Muller M, Michalakakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*. 2010; 5:e15367. [PubMed: 21203455]
2. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5- methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324:930–935. [PubMed: 19372391]
3. 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]
4. Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, Reik W, Balasubramanian S. Quantitative sequencing of 5-methylcytosine and 5- hydroxymethylcytosine at single-base resolution. *Science*. 2012; 336:934–937. [PubMed: 22539555]
5. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5- carboxylcytosine. *Science*. 2011; 333:1300–1303. [PubMed: 21778364]
6. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011; 333:1303–1307. [PubMed: 21817016]
7. Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science*. 2010; 329:78–82. [PubMed: 20595612]
8. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5- methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*. 2011; 145:423–434. [PubMed: 21496894]
9. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*. 2011; 146:67–79. [PubMed: 21722948]
10. Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5- formylcytosine and 5- carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*. 2011; 286:35334–35338. [PubMed: 21862836]

11. Maiti A, Morgan MT, Pozharski E, Drohat AC. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci U S A*. 2008; 105:8890–8895. [PubMed: 18587051]
12. Maiti A, Noon MS, Mackerell AD Jr, Pozharski E, Drohat AC. 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*. 2012; 109:8091–8096. [PubMed: 22573813]
13. Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol*. 2012; 8:328–330. [PubMed: 22327402]
14. Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X. 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]
15. Drohat AC, Pozharski E, Maiti A. 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]
16. Maiti A, Drohat AC. Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and processing of G.T mismatches. *DNA Repair (Amst)*. 2011; 10:545–553. [PubMed: 21474392]
17. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl*. 2011; 50:7008–7012. [PubMed: 21721093]
18. Kubo K, Ide H, Wallace SS, Kow YW. A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion. *Biochemistry*. 1992; 31:3703–3708. [PubMed: 1567824]
19. Takeshita M, Chang CN, Johnson F, Will S, Grollman AP. Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J Biol Chem*. 1987; 262:10171–10179. [PubMed: 2440861]
20. Huffman JL, Sundheim O, Tainer JA. DNA base damage recognition and removal: new twists and grooves. *Mutat Res*. 2005; 577:55–76. [PubMed: 15941573]
21. Hitomi K, Iwai S, Tainer JA. 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]
22. Morera S, Grin I, Vigouroux A, Couve S, Henriot V, Saparbaev M, Ishchenko AA. 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]
23. Wibley JE, Waters TR, Haushalter K, Verdine GL, Pearl LH. Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol Cell*. 2003; 11:1647–1659. [PubMed: 12820976]
24. Bucciarelli G, Bernardi G. An ultracentrifugation analysis of two hundred fish genomes. *Gene*. 2002; 295:153–162. [PubMed: 12354649]

Highlights

- Human TDG N157D mutant has a selective base excision for G:5caC substrate
- N157D mutant has higher activity for 5caC at a lower pH
- N157D remains specific for 5caC under the ratio of 5caC/C of $\sim 7 \times 10^{-6}$

Figure 1. N157 of human TDG is involved in DNA binding

(a) Structure of TDG in complex with an A:5caC mismatch (PDB 3UO7)¹³. The side chain amino group of Asn157 forms a weak hydrogen bond with the 5-carboxyl moiety of 5caC, which is flipped out from the double stranded DNA. (b) Structure of TDG in complex with a uracil analog (2'-deoxy-2'-fluoroarabinouridine) (PDB 3UFJ)¹². (c) A post-reactive complex structure of TDG containing a G:5hmU mismatch with the cleaved 5hmU base remaining in the active site pocket (PDB 4FNC)¹⁴. The side chain of Asn157 interacts with the 5' phosphate of the target nucleotide. (d-f) The activities of TDG wild type (WT) (panel d), mutants N157A (panel e) and N157D (panel f) on G:X under the single turnover condition ($[S_{DNA}] = 0.25 \mu M$ and $[E_{TDG}] = 2.5 \mu M$) after 10 min reaction at 30°C in 0.1% BSA, 1 mM EDTA, 100 mM NaCl and 10 mM BisTris HCl, pH 6.0. 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'. Human TDG residues 111-308 (pXC1056) and its mutants N157A

(pXC1155) and N157D (pXC1138) were prepared as described¹⁴. The excision reaction was monitored, as described¹⁴, by denaturing gel electrophoresis following NaOH hydrolysis of the abasic site.

Figure 2. N157D mutant selectively cleaves 5caC

(a) The time course (0–60 min) of N157D mutant reaction on seven oligonucleotides with various modifications under the single turnover condition at 30°C, pH 6.0 and 100 mM NaCl. For G:5caC substrate, more time points between 0 to 10 min were also taken. (b) N157D has enhanced activity on G:5caC at lower pH. Four different pH values between 6.0 and 7.2 with 0.4 increments were used at 10 mM BisTris HCl, 100 mM NaCl and 30°C. 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.): $[Product] = P_{max}(1 - e^{-kt})$, where P_{max} is the product plateau level, k (min⁻¹) is the observed rate constant, and t is the reaction time. (c) The activities of N157D on G:5caC at 30°C were measured at pH 6.0 with varying salt concentration between 0–250 mM NaCl with 25 or 50 mM increments. The K_{obs} values (min⁻¹) were calculated for reactions between 0–10 min. Plot of activity as a function of

NaCl concentration after 2.5 min reactions. **(d)** The log plots of activities of WT TDG (lines 1, 2 and 3) and N157D mutant (line 4) as a function of pH. For lines 1 and 2, the data were measured from a mixture of three buffers ¹⁴, whereas lines 3 and 4 were measured from 10 mM BisTrisHCl as indicated in the legend of panel b.

Figure 3. Detection of 5caC in the presence of large excess of genomic DNA

(a) Specificity of N157D (5caC vs. U) in the presence of large excess of genomic DNA. The reaction (200 μl) contains 25 μM of N157D, 2.5 nM of 32-bp FAM-labeled G:X (X=5caC or U) oligonucleotides and 100 μg (or 0.5 mg/ml) of salmon sperm genome DNA (Rockland Inc.). The G+C contents of salmon DNA are approximately 44.4%²⁴, and thus the salmon DNA used in the reaction contains 757 μM of base pair and 360 μM G:C pairs. The 5caC/C ratio (2.5 nM/360 μM) was estimated to be 7x10⁻⁶. (b) Excision reactions (20 μl) contain 25 μM (top panel) or 2.5 μM (bottom panel) of N157D, 25 nM of 32-bp FAM-labeled G:5caC oligonucleotides and 10 μg of salmon sperm genome DNA.