Supporting Information

Excision of 5-carboxylcytosine by Thymine DNA Glycosylase

Lakshmi S. Pidugu,[†] Qing Dai,[‡] Shuja S. Malik,[†] Edwin Pozharski,[†,§,*] and Alexander C. Drohat[†,*]

[†]Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD 21201, United States [‡]Department of Chemistry, The University of Chicago, Chicago, IL 60637, United States [§]Center for Biomolecular Therapeutics, Institute for Bioscience and Biotechnology Research, Rockville, MD 20850, United States

Table S1. Data collection and refinement statistics

	TDG G·caC^F (PDB ID: 6U17)	**N140A-TDG G·caC^F** (PDB ID: 6U15)	**N140A-TDG G·caC** (PDB ID: 6U16)
Data collection			
Space Group	C2	C2	C2
Cell Dimensions			
a, b, c (Å)	96.9, 52.8, 81.6	93.2, 53.5, 82.5	94.6, 53.2, 81.3
β (°)	95.3	95.4	94.8
Resolution (Å)	40.25-1.55 (1.58-1.55)	43.37-2.40 (2.49-2.40)	81.03-1.60 (1.63-1.60)
R_{pim}	0.030	0.162	0.047
Mean $I/\sigma I$	9.4 (0.8)	4.0 (0.5)	6.7 (0.8)
$CC_{1/2}$	0.999 (0.343)	0.975 (0.214)	0.997 (0.288)
Completeness (%)	95.5 (86.7)	95.0 (97.3)	97.3 (98.2)
Redundancy	21.7 (17.3)	5.8 (5.8)	6.2 (6.2)
Wilson B-factor (Å^2)	32.5	51.3	31.9
Refinement			
Program	Buster 2.10.3	Buster 2.10.3	Buster 2.10.3
Resolution (Å)	43.25-1.55	42.13-2.40	42.31-1.6
No. of reflections	56810	15179	51710
R_{work}/R_{free}	0.189/0.224	0.199/0.213	0.186/0.210
Number of atoms			
Protein	1520	1479	1533
DNA	1123	1123	1123
Water	308	91	308
Other	34	30	41
B-factors (Å^2)			
Protein	43.13	56.35	36.30
DNA	61.12	78.15	52.01
Water	60.65	60.10	52.56
Acetate	-	-	57.95
Ramachandran Plot			
Favoured (%)	96.9	94.8	97.9
Allowed (%)	3.1	5.2	2.1
Outliers (%)	0.0	0.0	0.0
RMSD from ideal			
Bond lengths (Å)	0.010	0.010	0.010
Bond angles (°)	0.98	1.03	0.98

Values shown in parenthesis are for highest resolution shell. The Ramachandran analysis was performed using Molprobity.[1] Wilson B-factor estimated by phenix.xtriage. Number of atoms includes all atom records explicitly included in the model, including alternate positions.

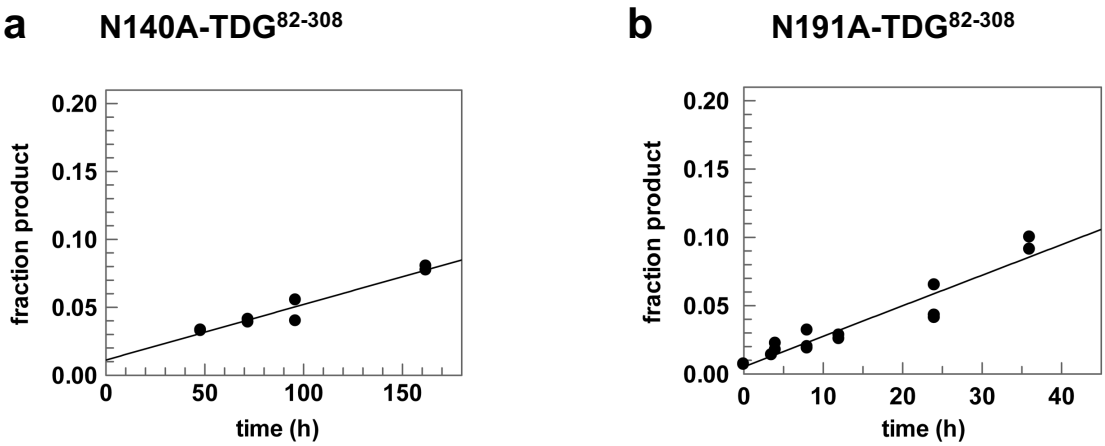

Figure S1. Single turnover kinetics for TDG$^{82\text{-}308}$ variants acting on a G·caC substrate at 23°C. Due to low activity the data were fitted by linear regression. (**a**) Fitting of data for of N140A-TDG$^{82\text{-}308}$ gives rate constant of k_{obs} = (6.8 ± 0.7) × 10^{-6} min^{-1} with a corresponding reaction half-life of $t_{1/2}$ = 70 d. Experiments were collected with 0.5 μM substrate and 1.5 μM enzyme. Given the tight binding of N140A-TDG to G·caC DNA (K_d = ~2 nM),[2] this provides saturating enzyme conditions, such that k_{obs} represents the maximal rate of product formation ($k_{obs} \approx k_{max}$). (**b**) Data fitting for N191A-TDG$^{82\text{-}308}$ gives a rate constant of k_{obs} = (3.7 ± 0.3) × 10^{-5} min^{-1} and $t_{1/2}$ = 12.8 d. Experiments were collected with 0.5 μM substrate and 1.5 μM enzyme. Given that N191-TDG binds tightly to G·caC DNA (K_d = ~10 nM),[3] this provides saturating enzyme conditions such that $k_{obs} \approx k_{max}$.

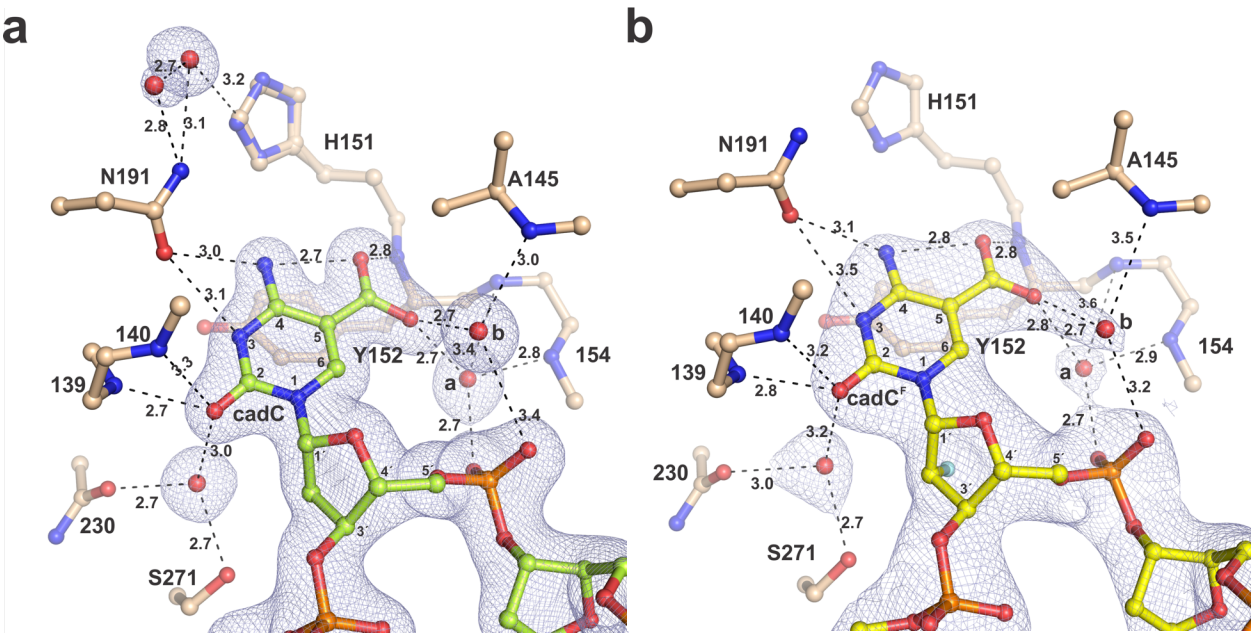

Figure S2. The N140A mutation does not substantially alter TDG interactions with the caC base. (**a**) Structure of N140A-TDG$^{82\text{-}308}$ bound to DNA with cadC (non-substituted) flipped into its active site (PDB ID: 6U16). TDG residues are in tan and cadC DNA is lime. (**b**) Structure of N140A-TDG$^{82\text{-}308}$ bound to DNA with cadC^{F} (2′-F-cadC) in its active site (PDB ID: 6U15), solved at substantially reduced resolution of 2.40 Å. TDG residues are tan and cadC^{F} DNA is yellow. For both panels, red spheres are water molecules (some omitted for clarity), and the $2F_o$–F_C electron density map, contoured at 1.0σ, is shown for DNA and water molecules. Dashed lines represent hydrogen bonds, with interatomic distances (Å).

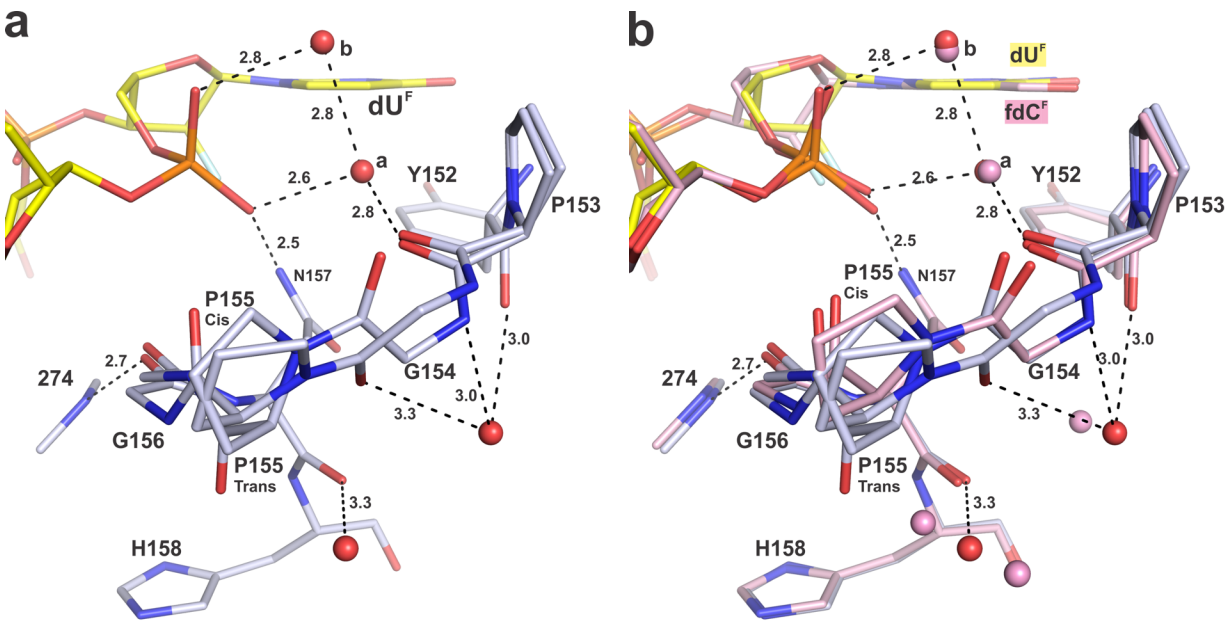

Figure S3. Conformation of the α1-α2 loop for TDG$^{82\text{-}308}$ bound to DNA containing dU^F or fdC^F. (a) Structure TDG$^{82\text{-}308}$ bound to DNA with dU^F (2´-F-dU) flipped into in its active site (PDB ID: 5HF7). TDG residues are grey and dU^F DNA is yellow, and red spheres are water molecules. Note that both cis and trans isomers of P155 are observed. (b) Alignment of the structure in panel a (TDG-dU^F) with that of TDG$^{82\text{-}308}$ bound to DNA containing fdC^F (2´-F-fdC) (PDB ID: 5T2W), with all elements (TDG, DNA, waters) shown in pink. The cis isomer of P155 is exclusively observed for TDG bound to fdC^F DNA. Note that, regardless of P155 isomerization for TDG bound to dU^F DNA, the backbone oxygen of P153 receives a hydrogen bond from a water that provides a hydrogen bond to the 5´-phosphate, which is exactly as observed for TDG bound to dfC DNA, but different from the hydrogen bonding involving the corresponding water for TDG bound to dcaC DNA (Figure 6, main text).

METHODS

Glycosylase Assays

Glycosylase activity was determined for TDG^{82-308} variants acting on DNA containing a G·caC pair using single turnover kinetics experiments performed with saturating enzyme conditions.[4-5] Reactions were initiated by adding enzyme to G·caC substrate (0.5 uM) in HEN.1 buffer (0.02 M HEPES pH 7.5, 0.1 M NaCl, 0.2 mM EDTA) and were quenched with 50% (v:v) 0.3 M NaOH, 0.03 M EDTA, and heated (15 m, 85 °C) to quantitatively cleave the DNA backbone at abasic sites. The resulting DNA fragments were resolved by UHPLC using a DNAPac PA200 RS ion exchange column (Thermo Scientific) under denaturing (pH 12) conditions,[6] and peak areas were used to determine the fraction product for a given sample. Progress curves (fraction product versus time) were fitted by non-linear regression to eq. 1:

$$\text{fraction product} = A(1 - \exp(-k_{obs}t)) \quad (1)$$

where A is the amplitude, k_{obs} is the rate constant, and t is the reaction time. Because the reactions for N140A- and N191A-TDG^{82-308} were very slow the data were fitted by linear regression using time points in the initial linear range of the reaction (fraction product $\leq$ 0.1). Experiments were performed with saturating enzyme ([E] >> K_d; [E] > [S]) such that the rate constant reflects the maximal rate of product formation ($k_{obs} \approx k_{max}$) and is not influenced by enzyme-substrate association or by product release or product inhibition [4]. TDG binds G·caC DNA with a K_d of 0.002 μM [2] and TDG^{82-308} is expected to bind with about the same affinity,[5] and the kinetics experiments were performed with a TDG^{82-308} concentration of 1.5 μM, giving saturating enzyme conditions.

X-ray Crystallography

The structures include multiple conformations and partial oxidation of the C276 side chain, as indicated by electron density. By contrast, for our previous high-quality structures of TDG^{82-308} bound to other DNA substrates (G·fC, G·U) or abasic DNA, the electron density indicated a single conformation and reduced state for Cys276.[5,7] However, the C276A mutation does not alter TDG^{82-308} activity for a G·caC substrate (not shown), indicating the multiple C276 conformations observed in the current structures is not likely relevant to glycosylase activity.

REFERENCES

(1) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 12-21.

(2) Coey, C. T.; Drohat, A. C. Defining the impact of sumoylation on substrate binding and catalysis by thymine DNA glycosylase. *Nucleic Acids Res* **2018**, *46*, 5159-5170.

(3) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J Am Chem Soc* **2013**, *135*, 15813-15822.

(4) Coey, C. T.; Drohat, A. C. Kinetic Methods for Studying DNA Glycosylases Functioning in Base Excision Repair. *Methods Enzymol* **2017**, *592*, 357-376.

(5) Coey, C. T.; Malik, S. S.; Pidugu, L. S.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res* **2016**, *44*, 10248-10258.

(6) Morgan, M. T.; Bennett, M. T.; Drohat, A. C. Excision of 5-halogenated uracils by human thymine DNA glycosylase: Robust activity for DNA contexts other than CpG. *J Biol Chem* **2007**, *282*, 27578-27586.

(7) Pidugu, L. S.; Flowers, J. W.; Coey, C. T.; Pozharski, E.; Greenberg, M. M.; Drohat, A. C. Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase. *Biochemistry* **2016**, *55*, 6205-6208.