# Supplementary Data for:

# Structural basis of damage recognition by Thymine DNA Glycosylase

Christopher T. Coey[1,§], Shuja S. Malik[1,§], Lakshimi S. Pidugu[1], Kristen Varney[1,2,3], Edwin Pozharski[1,2,3,*], and Alexander C. Drohat[1,2,*]

[1]Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD 21201, USA, [2]University of Maryland Marlene and Stewart Greenebaum Cancer Center, Baltimore, MD 21201, USA, [3]Center for Biomolecular Therapeutics, Institute for Bioscience and Biotechnology Research, Rockville, MD 20850, USA

* To whom correspondence should be addressed. Tel. 410-706-8118; Email: adrohat@som.umaryland.edu. Tel: 240-314-6255; Email: EPozharskiy@som.umaryland.edu

**Supplementary Table S1. Data collection and refinement statistics**

| | E·S TDG$^{82\text{-}308}$ G·U$^{F}$ (PDBID: 5HF7) | E·P TDG$^{82\text{-}308}$ (PDBID: 5FF8) | E·S TDG$^{111\text{-}308}$ G·U$^{F}$ (PDBID: 5JXY) |
|---|---|---|---|
| **Data collection** | | | |
| Space Group | C2 | C2 | C2 |
| Cell Dimensions | | | |
| *a, b, c* (Å) | 97.0, 52.8, 81.2 | 96.9, 52.5, 81.3 | 88.5, 53.3, 82.6 |
| β (°) | 95.2 | 95.2 | 95.6 |
| Resolution (Å) | 32.52-1.54 (1.57-1.54) | 31.08-1.70 (1.73-1.70) | 39.03-1.71 (1.74-1.71) |
| $R_{pim}$ | 0.035 | 0.057 | 0.052 |
| Mean *I*/σ*I* | 9.3 (0.6) | 7.9 (0.7) | 12.9 (0.5) |
| $CC_{1/2}$ | 0.998 (0.35) | 0.998 (0.50) | 0.999 (0.224) |
| Completeness (%) | 100.0 (100.0) | 100.0 (99.9) | 95.7 (37.0) |
| Redundancy | 5.0 (4.7) | 7.5 (7.4) | 19.8 (7.3) |
| Wilson B-factor (Å$^2$) | 28.2 | 29.0 | 24.6 |
| **Refinement** | | | |
| Program | BUSTER-TNT | BUSTER-TNT | BUSTER-TNT |
| Resolution (Å) | 32.51-1.54 | 32.48-1.70 | 24.97-1.71 |
| No. of reflections | 59281 | 44159 | 39815 |
| $R_{work}/R_{free}$ | 0.186/0.216 | 0.194/0.217 | 0.197/0.231 |
| Number of atoms | | | |
| Protein | 1603 | 1596 | 1544 |
| DNA | 1123 | 1123 | 1164 |
| Water | 330 | 235 | 228 |
| B-factors (Å$^2$) | | | |
| Protein | 36.8 | 37.4 | 27.6 |
| DNA | 53.2 | 55.3 | 43.8 |
| Water | 50.2 | 43.3 | 40.6 |
| Ramachandran Plot | | | |
| Favoured (%) | 98.1 | 97.5 | 96.4 |
| Allowed (%) | 100 | 100 | 100 |
| Outliers (%) | 0 | 0 | 0 |
| RMSD from ideal | | | |
| Bond lengths (Å) | 0.010 | 0.010 | 0.008 |
| Bond angles (°) | 1.0 | 0.98 | 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.

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

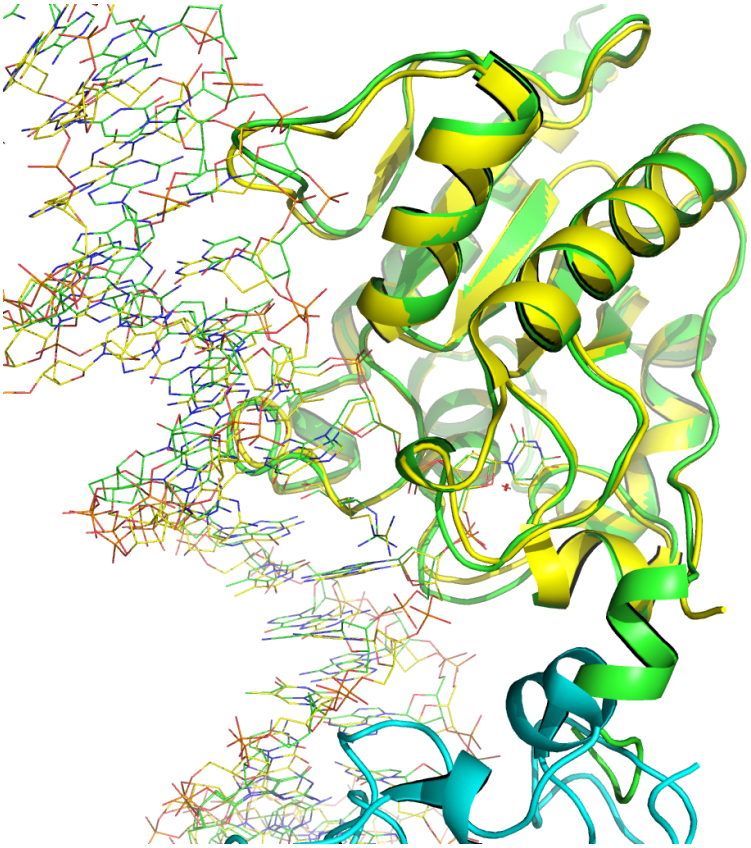

**Figure S1**. Alignments show that the previous TDG$^{111\text{-}308}$-G·U$^{F}$ structure (2.97 Å, yellow) differs significantly from the new TDG$^{111\text{-}308}$-G·U$^{F}$ structure (green), with a percentile-based spread (p.b.s.) of 0.51 Å (backbone; see main text). The previous TDG$^{111\text{-}308}$-G·U$^{F}$ structure (PDBID: 3UFJ) was obtained from crystals in which TDG$^{111\text{-}308}$ binds DNA with 2:1 stoichiometry, one subunit (yellow) at the G·U$^{F}$ site and the other subunit (cyan) at a nonspecific site. The new TDG$^{111\text{-}308}$-G·U$^{F}$ structure reported here features 1:1 stoichiometry. The N-terminal helix observed for structures with 1:1 binding (green helix with no yellow counterpart) is likely destabilized by the TDG:TDG dimer interface for complexes featuring 2:1 (TDG:DNA) binding.

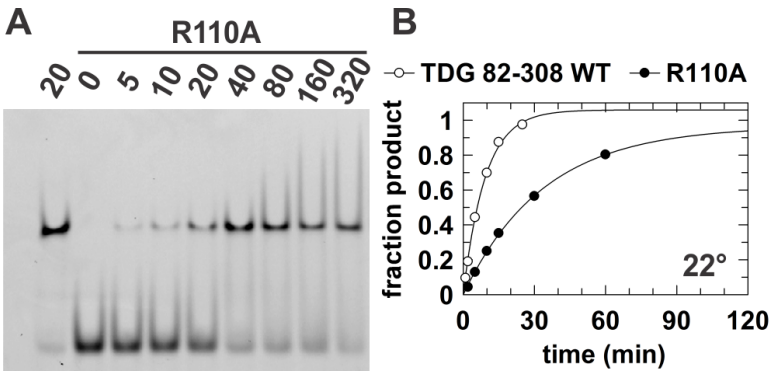


**Figure S2**. Biochemical studies of TDG$^{82\text{-}308}$ and its R110A variant. (**A**) Equilibrium binding of TDG$^{82\text{-}308}$ to DNA (10 nM) containing a G·T$^{F}$ mismatch, monitored by EMSA. For reference, the far left lane shows binding of native TDG$^{82\text{-}308}$ (20 nM) to G·T$^{F}$ DNA. (**B**) Single-turnover kinetics for excision of T from a G·T DNA substrate by wild-type TDG$^{82\text{-}308}$ and the R110A mutant at 22 °C (R110A-TDG$^{82\text{-}308}$ is unstable at 37 °C). The rate constants are $k_{max}$ = 0.108 ± 0.006 min$^{-1}$ for TDG$^{82\text{-}308}$ and $k_{max}$ = 0.030 ± 0.001 min$^{-1}$ for R110A-TDG$^{82\text{-}308}$, giving a 3.7-fold mutational affect.

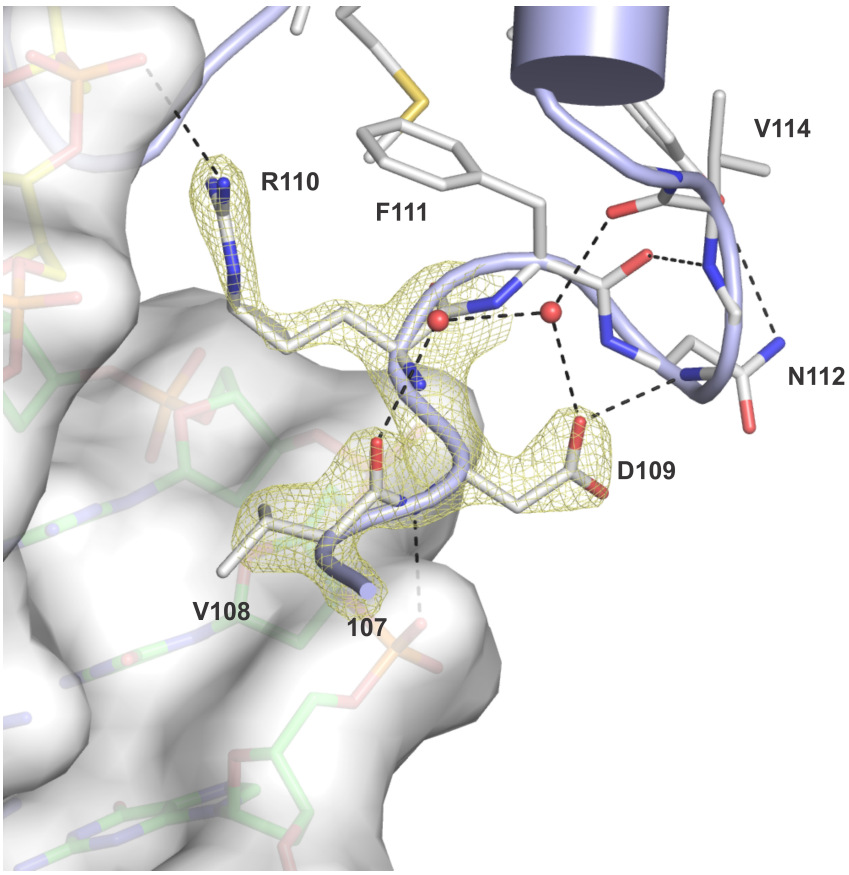


**Figure S3**. Crystal structure of an E·S complex for $TDG^{82\text{-}308}$ bound to $G{\cdot}U^{F}$ DNA (PDBID: 5HF7), solved at 1.54 Å resolution, focusing on N-terminal residues. TDG is in cartoon format (blue) with some residues in stick format (C, white; N, blue; O, red). The $2F_o\text{-}F_c$ electron density map, contoured at 1.0 σ, is shown in yellow for residues 107 to 110 of TDG. DNA is shown in space and stick formats, with the dUrd-containing strand yellow and the complementary strand green. Water molecules are shown as red spheres and dashed lines represent hydrogen bonds.

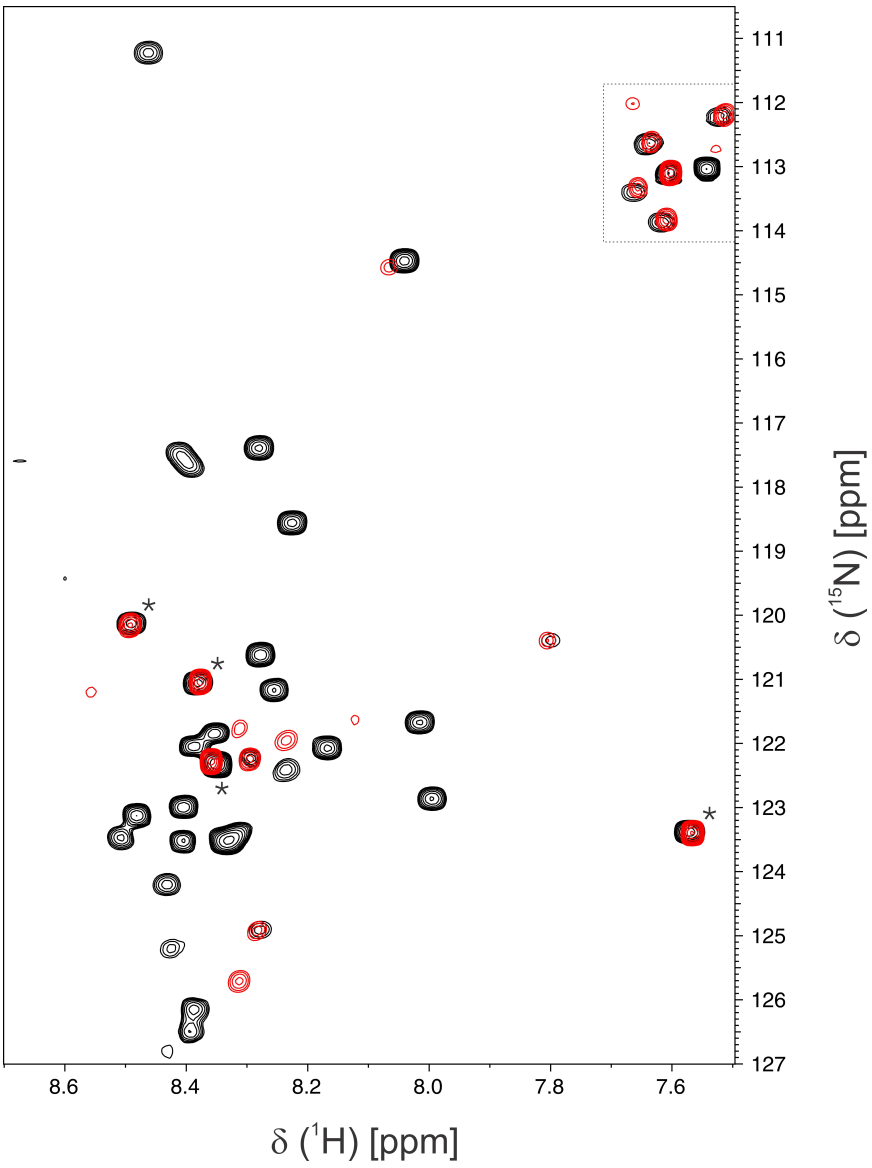


**Figure S4**. $^{15}$N-HSQC NMR spectra for TDG$^{82\text{-}308}$ (black peaks) and TDG$^{111\text{-}308}$ (red). Four resonances that are nearly identical in both spectra (marked with *) likely represent disordered C-terminal residues (see main text). Resonances in the upper right (box with dotted lines) are from side chain amino groups. The NMR samples contained TDG$^{82\text{-}308}$ or TDG$^{111\text{-}308}$ in 0.02 M sodium phosphate pH 6.5, 0.15 M NaCl, 0.2 mM EDTA, 0.2 mM DTT, 7% $D_2O$. Experiments were collected at 18 °C on an 800 MHz NMR spectrometer.

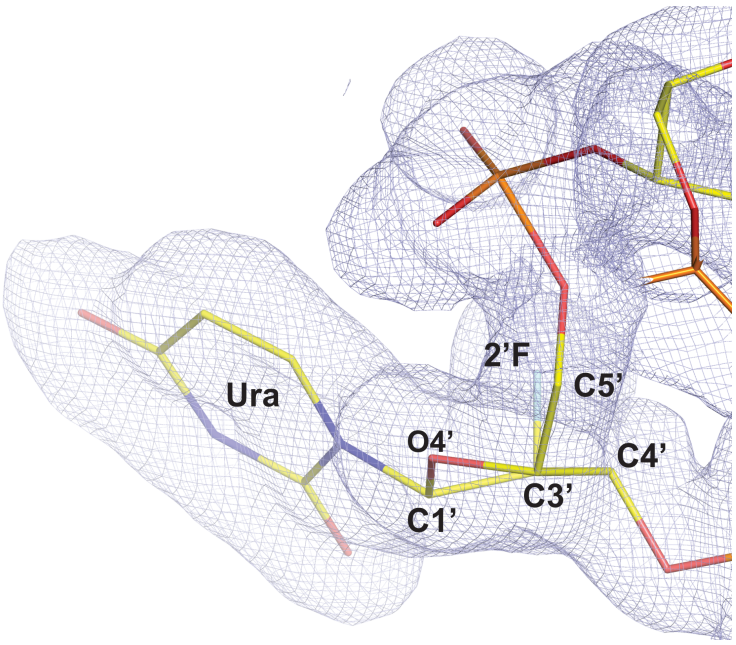


**Figure S5**. The 2´-fluoroarabino-dUrd analog flipped into the TDG$^{82\text{-}308}$ active site exhibits a C1´-exo-O4´-endo sugar pucker. The DNA is shown in stick format colored yellow (cyan for 2´-F), with a $2F_o$-$F_c$ map contoured at 1.0 σ. Note that the sugar is oriented such that C2´ and C3´ are aligned, with the bond vector pointing toward the viewer and C3´ in the foreground.

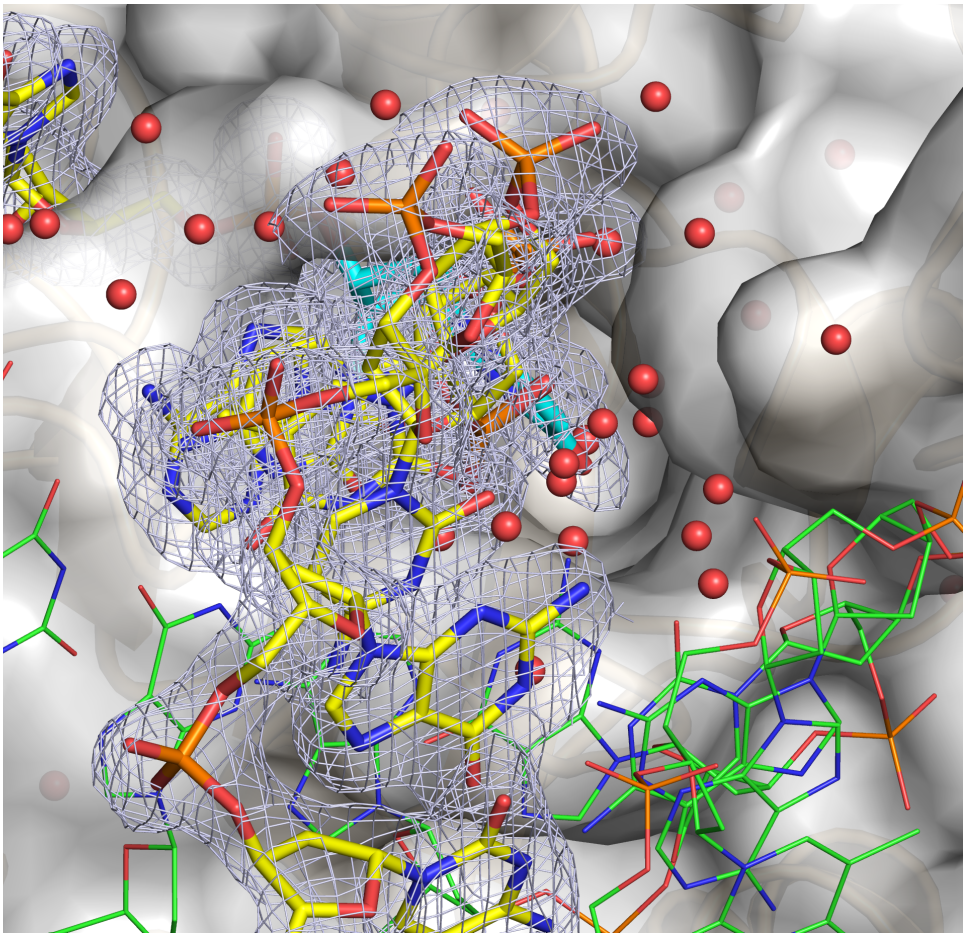

**Figure S6**. Structure of an E·S complex for $TDG^{82\text{-}308}$ bound to $G{\cdot}U^{F}$ features a solvent-filled channel from the active site to the enzyme surface, along the target DNA strand. $TDG^{82\text{-}308}$ is shown in space-filling and cartoon modes, the DNA is in stick format, with the target strand in yellow, the dUrd nucleotide cyan, and the complementary DNA strand in green. The $2F_o\text{-}F_c$ map, contoured at 1.0 σ, is shown in blue for the target DNA strand. Water molecules are shown as red spheres.

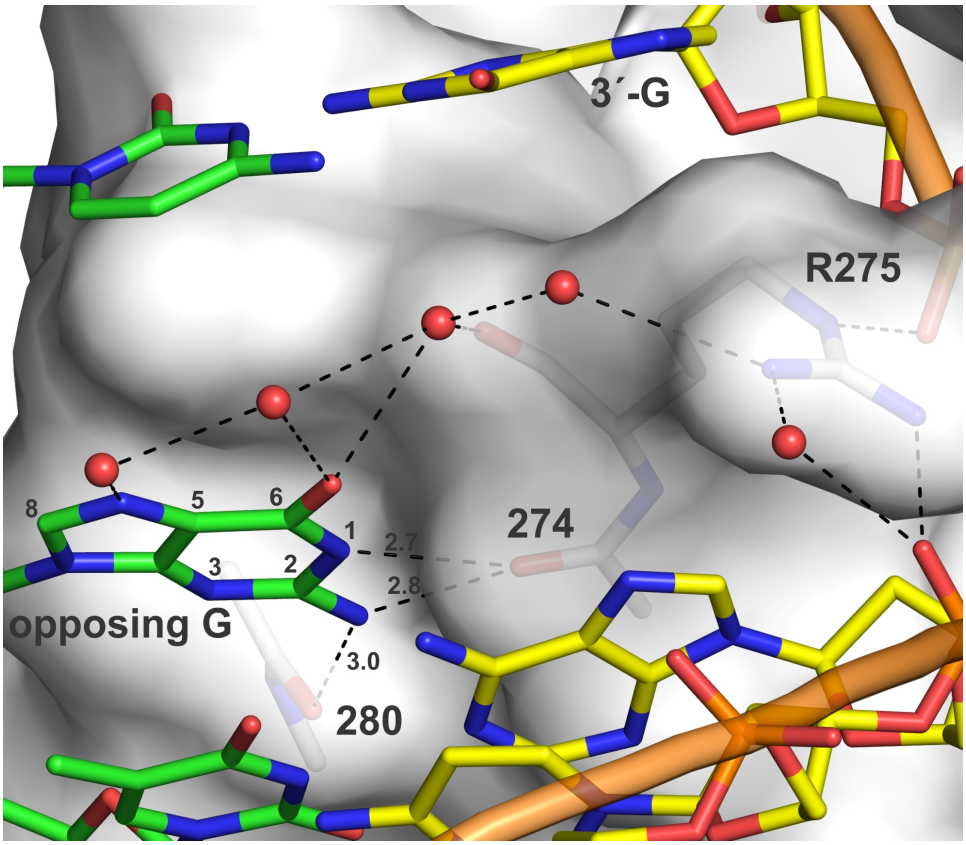


**Figure S7**. TDG contacts the Gua ("opposing G") of the G·U mispair. $TDG^{82-308}$ is shown in surface representation with residues that contact the opposing G shown as sticks. The Ura-containing DNA strand is yellow and the complementary strand green, and water molecules are red spheres. Dashed lines represent hydrogen bonds, with interatomic distances (Å). The backbone oxygen of $Ala^{274}$ contacts the opposing Gua at both N1H (2.7 Å) and $N2H_2$ (2.8 Å), and the Gua is also contacted at $N2H_2$ by the backbone oxygen of $Pro^{280}$ (3.0 Å).