Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues

Christopher T. Coey¹,†, Shuja S. Malik¹,†, Lakshmi S. Pidugu¹, Kristen M. Varney¹,2,3, Edwin Pozharski¹,2,3,* and Alexander C. Drohat¹,2,*

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

Received July 19, 2016; Revised August 20, 2016; Accepted August 22, 2016

ABSTRACT

Thymine DNA Glycosylase (TDG) is a base excision repair enzyme functioning in DNA repair and epigenetic regulation. TDG removes thymine from mutagenic G-T mispairs arising from deamination of 5-methylcytosine (mC), and it processes other deamination-derived lesions including uracil (U). Essential for DNA demethylation, TDG excises 5-formylcytosine and 5-carboxylcytosine, derivatives of mC generated by Tet (ten-eleven translocation) enzymes. Here, we report structural and functional studies of TDG82-308, a new construct containing 29 more N-terminal residues than TDG111-308, the construct used for previous structures of DNA-bound TDG. Crystal structures and NMR experiments demonstrate that most of these N-terminal residues are disordered, for substrate- or product-bound TDG. Nevertheless, G-T substrate affinity and glycosylase activity of TDG82-308 greatly exceeds that of TDG111-308 and is equivalent to full-length TDG. We report the first high-resolution structures of TDG in an enzyme-substrate complex, for G-U bound to TDG82-308 (1.54 Å) and TDG111-308 (1.71 Å), revealing new enzyme-substrate contacts, direct and water-mediated. We also report a structure of the TDG82-308 product complex (1.70 Å). TDG82-308 forms unique enzyme–DNA interactions, supporting its value for structure-function studies. The results advance understanding of how TDG recognizes and removes modified bases from DNA, particularly those resulting from deamination.

INTRODUCTION

Thymine DNA glycosylase (TDG) is an enzyme that initiates base excision repair by removing modified forms of 5-methylcytosine (mC) that are generated by deamination or oxidation (1). TDG excises thymine from G-T mispairs, thereby protecting against C→T transition mutations that arise via deamination of mC to T (2,3). TDG is also essential for active DNA demethylation, which likely accounts for findings that its depletion in mice leads to embryonic lethality (4,5). An established pathway for active DNA demethylation includes TDG excision of 5-formylcytosine or 5-carboxylcytosine (6,7), epigenetic bases that are generated via oxidation of mC by one of three ten-eleven translocation enzymes (7–11). TDG also removes many other bases (in vitro), including uracil, 5-halogenated uracils (5FU, 5ClU, 5BrU, 5IU) and 5-hydroxymethyl-U (hmU), among others (12,13). Human TDG (410 residues) contains a central catalytic domain of about 195 residues flanked by N-terminal and C-terminal regions of roughly equivalent size that are disordered and yet important for certain functions, interactions with other proteins and regulation by post-translational modifications such as acetylation, phosphorylation and SUMO conjugation (14–21).

Here, we report structural and functional studies of TDG82-308, a new construct of human TDG comprised of residues 82–308 (410 total) that includes the catalytic domain and an N-terminal region that contains amino acid residues involved in regulation of TDG via protein interactions or post-translational modifications. All previous structural studies of TDG have used a construct referred to as the catalytic domain, TDGcat, comprising residues 111–308 (22–25) or SUMO-conjugated TDG, which included residues 117–332 (26). TDG82-308 contains 29 additional N-terminal residues compared to TDG111-308 (Figure 1). These include the PIP degron (residues 95–106), which mediates TDG depletion during S phase of the cell cycle via interac-
Importantly, all residues of the new TDG82-308 construct interact with other proteins, including CBP. Putative sites for post-translational modification are indicated for TDG82-308, including acetylation (A) and phosphorylation (P). Also shown is the PIP degron and an α-helix that has not been observed in previous structures of E-S complexes. Shown in red for TDG111-308 are the six non-native residues that remain after removal of the poly-His tag. All residues of TDG82-308 are native to TDG.

**MATERIALS AND METHODS**

**Materials**

Full length human TDG (410 residues) and TDG111-308 (also referred to as TDGcat) were expressed in *Escherichia coli* and purified essentially as previously described (23,35). A vector (pJ401) for expressing TDG82-308, a new construct containing residues Ser82-Val308 of human TDG (Figure 1) plus an N-terminal poly-His tag, was obtained from DNA 2.0 (Newark, CA, USA) and transformed into *E. coli* BL21(DE3) cells. TDG82-308 was expressed (at 15°C) and purified essentially as described for TDG111-308, using Ni-affinity, ion-exchange (SP Sepharose) and size exclusion chromatography (23,35). The poly-His tag was removed (after Ni-affinity) using the tobacco etch virus (TEV) protease (36), which cleaves on the carboxyl end of its recognition site. As such, following TEV cleavage, all residues of TDG82-308 are native to TDG. By contrast, TDG111-308 contains six non-native N-terminal residues (GSHMAS) that remain after thrombin cleavage of the N-terminal poly-His tag (Figure 1) (22,23).

We investigated the activity of TDG82-308, TDG111-308 and full-length TDG for acting on G-T mismatches in DNA. We find that substrate binding and catalysis for TDG82-308 is equivalent to that of full-length TDG, and much greater than that of TDG111-308. Using improved crystallization methods (33), we solved high-resolution structures of the enzyme bound to a stable G-U substrate analog, for both TDG82-308 (1.54 Å) and TDG111-308 (1.71 Å). These new structures are of much higher resolution than previously reported structures of TDG in an enzyme-substrate (E-S) complex, which were solved at 3.0 Å resolution (24,34). We also solved a high-resolution (1.70 Å) enzyme-product (E-P) complex using the new TDG82-308 construct. The structures reveal important enzyme-DNA contacts that have not previously been observed, many of which are mediated by ordered water molecules (absent from previous structures). The structures also confirm many key enzyme-substrate interactions that were suggested by the previous lower-resolution structures. Notably, several key enzyme-DNA interactions are observed only for the new construct, TDG82-308, suggesting its value for future structure-function studies. Remarkably, crystal structures and solution studies using NMR spectroscopy show that most of the 29 N-terminal residues of TDG82-308 are disordered, even when the enzyme is tightly bound to G-U or G-T substrate DNA. Together, our results advance the understanding of how TDG recognizes and excises modified bases in DNA, particularly those arising from deamination.

**Figure 1.** Amino acid sequence for the N-terminal regions of TDG82-308 and TDG111-308 and the initial residues of the bacterial homologue MUG (*E. coli*). Putative sites for post-translational modification are indicated for TDG82-308, including acetylation (A) and phosphorylation (P). Also shown is the PIP degron and an α-helix that has not been observed in previous structures of E-S complexes. Shown in red for TDG111-308 are the six non-native residues that remain after removal of the poly-His tag. All residues of TDG82-308 are native to TDG.
described (35). ODNs containing the 2′-fluoroarabinobase analogues of deoxyuridine or deoxythymidine, referred to as UF and TF, respectively, were synthesized at the Yale facility using phosphoramidites obtained from Glen Research (UF) or Link Technologies (TF) (39). TDG binds productively to DNA containing UF or TF but it cannot hydrolyze the N6-glycosyl bond because the single-atom fluorine substitution destabilizes the chemical transition-state (38,39,42,43). The duplex included a 28mer target strand, 5'-AGCTGTCCATTCGCTAxTGATCAAGAGTC, where x is T, UF or TF and its complement, 5'-CAGCTCTGTACGTAGCGATGGACAGCT, such that the target base (x) is paired with G and located in a CpG dinucleotide context (underlined), consistent with TDG specificity (35,44). The same 28 bp DNA construct was used for glycosylase assays (x = U, T), NMR experiments (x = T/F) and equilibrium binding studies (x = T/F; with 3′ 6-FAM on the complementary strand).

X-ray crystallography

Samples used for crystallization contained 0.35 mM enzyme (TDG82-308 or TDG111-308) and 0.42 mM DNA in a buffer of 5 mM Tris-HCl pH 7.5, 0.13 M NaCl, 0.2 mM DTT, 0.2 mM EDTA. The enzyme–product complex was produced by incubating TDG82-308 with G-U DNA substrate for a sufficient time to ensure full conversion to product, as confirmed by HPLC (12). Crystals were grown at room temperature (~22°C) by sitting drop vapor diffusion, using 1 μl of the TDG–DNA sample and 1 or 2 μl of mother liquor, which was 30% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 6.0. Crystals typically appeared within a few days. Crystals were cryo-protected using mother liquor supplemented with 18% ethylene glycol and flash cooled in liquid nitrogen.

X-ray diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL beamlines 11–1) and at the Advanced Light Source (ALS beamline 8.2.2). The images were processed using XDS (45) and scaled with Aimless (46) from the CCP4 program suite (47) with the help of the autodxs script developed by Ana Gonzalez and Yingssu Tsai (http://smb.slac.stanford.edu/facilities/software/xds). For the TDG111-308 E-S complex 3 data sets were collected from separate section of a single crystal and merged to increase resolution. Resolution cutoff was determined based on CC1/2 values (48). Structures were solved by molecular replacement using Phaser (49), and a previously reported structure of DNA-bound TDG111-308 as the search model (PDBID: 4Z47). Refinement was performed using BUSTER-TNT (50) or REFMAC5 (51), and model building was performed using Coot (52). TLS refinement protocol utilized TLSMD server (53,54) as described (33). The structural figures were made with PyMOL (http://www.pymol.org).

NMR spectroscopy

15N-HSQC experiments were collected on an 800 MHz Bruker Avance III NMR spectrometer, and the data were processed and analyzed using NMRPipe and NMRDraw (55). Sample conditions are provided in the relevant figure legends.

Equilibrium binding assays

Equilibrium binding of enzyme (TDG, TDG82-308 or TDG111-308) to a G-TF substrate analog was analyzed using electrophoretic mobility shift assays (EMSAs), performed essentially as described (20), where TF is the 2′-fluoroarabinobase analogue of T (described above). Samples contained a 10 nM concentration of G-TF DNA and varying concentrations of enzyme. Binding reactions were incubated at room temperature for 30 min, loaded onto a 6% native denaturing polyacrylamide gel (Invitrogen) and run at 4°C for 2 h at 50 V. Gels were imaged using a Typhoon 9400 variable mode imager (GE Healthcare) as described (56).

Glycosylase assays

Single turnover kinetics reactions were initiated by adding enzyme (TDG, TDG82-308 or TDG111-308) to G-T substrate (0.5 μM) in HEMN.1 buffer (0.02 M HEPES pH 7.5, 0.1 M NaCl, 0.2 mM EDTA, 2.5 mM MgCl2). Aliquots were removed at desired time points, quenched with 50% (v:v) 0.3 M NaOH, 0.03 M EDTA, and heated (15 min, 85°C) to quantitatively cleave the DNA backbone at TDG-generated abasic sites. The resulting DNA fragments were resolved by HPLC (35) and peak areas were used to determine fraction product. Progress curves (fraction product versus time) were fitted by non-linear regression to eq. 1:

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

where \( A \) is the amplitude, \( k_{\text{obs}} \) is the rate constant, and \( t \) is the reaction time. Experiments were performed with saturating enzyme ([E] >> \( K_d \); [E] >> [S]) such that the observed rate constant reflects the maximal rate of product formation (\( k_{\text{obs}} \approx k_{\text{max}} \)) and is not influenced by enzyme–substrate association or by product release or product inhibition (39).

Previous studies show that TDG binds G-T DNA with a \( K_d \) of 0.02 μM (38), while TDG111-308 binds G-T DNA with \( K_d \) of roughly 1.3 μM (24). Findings here indicate similar values and show that TDG82-308 binds with tighter affinity than TDG. Thus, kinetics experiments were performed with an enzyme concentration of 5 μM for TDG or TDG82-308, and 32 μM for TDG111-308. Saturating conditions were confirmed by observation of identical rate constants for experiments performed at other enzyme concentrations (not shown).

RESULTS AND DISCUSSION

Residues 82–110 of TDG confer tight substrate binding and full glycosylase activity for G-T mispairs

To compare the activity of TDG82-308 with the smaller construct, TDG111-308 and with full-length TDG, we examined binding affinity and glycosylase (base excision) activity for a G-T substrate. This provides a stringent test for the role of N-terminal residues 82–110 (absent on TDG111-308), because substrate binding and base excision are weak for G-T
First high-resolution structures of enzyme-substrate complexes for TDG

Previous structures of enzyme-substrate (E-S) complexes for TDG111-308, with either G-U or G-caC, revealed important features of substrate recognition and catalysis, but were solved at moderate (3.0 Å) resolution (24,34). Moreover, the crystals used for these previous structures were obtained under conditions that give 2:1 binding (TDG:DNA), involving substantial interactions between the two TDG subunits; 2:1 binding is likely an artifact of crystallization and 1:1 binding appears to be more physiologically relevant (23,38,58). Structures of enzyme-product (E-P) complexes have been determined using crystals generated under conditions that give 1:1 binding (TDG:DNA) (25), including our recent structures solved at high resolution (up to 1.45 Å) (33). Following this approach, we solved high-resolution structures of E-S complexes, for G-U^F DNA bound to TDG82-308 (1.54 Å) or TDG111-308 (1.77 Å) (Supplementary Table S1). These structures are of much higher quality than previous structures of E-S complexes, revealing new enzyme–DNA interactions. Moreover, the new structures feature hundreds of water molecules, some mediating key enzyme-DNA interactions. By contrast, no water molecules were observed in previous structures of E-S complexes, except for the putative nucleophile in the structure of TDG111-308 bound to a G-U^F mispair (34). Structural comparisons using the percentile-based spread (p.b.s.) approach (59) reveal that the previous TDG111-308-G-U^F structure (2.97 Å resolution) differs substantially from the new structures reported here, with a backbone p.b.s. of 0.51 Å for TDG111-308-G-U^F and 0.58 Å for TDG82-308-G-U^F (Supplementary Figure S1). By contrast, the two new E-S structures reported here are quite similar, with a backbone p.b.s. of 0.23 Å, even though they feature different TDG constructs and crystal cell parameters.

N-terminal residues in crystal structures of TDG82-308

While the results above show that the 29 N-terminal residues of TDG82-308 confer tight binding to G-T DNA (Figure 2), only four of these residues (107–110) are observed in the new structures, for both the E-S complex (Figure 3) and the E-P complex (not shown), indicating that the 25 N-terminal residues (82–106) are disordered. This conclusion is supported by NMR studies, as discussed below. Nevertheless, the structures reported here provide new information regarding the N-terminal residues that are observed. First, it is important to note that many more N-terminal residues are observed in structures that feature 1:1 versus 2:1 binding stoichiometry (TDG:DNA), as shown by ourselves and others (25,33). These previous studies showed that, depending on the DNA construct, TDG111-308 can bind DNA with either 1:1 or 2:1 stoichiometry (25,33). Importantly, all previous structures of enzyme–substrate complexes for TDG were obtained from crystals generated using DNA that yielded 2:1 binding (TDG:DNA), and therefore lack structural information for residues 111 to 122 (23,24,34). By contrast, the two enzyme–substrate structures reported here were generated from DNA that gives 1:1 binding and include N-terminal residues beginning at 107 for TDG82-308 and at 111 for TDG111-308. These N-terminal

---

**Figure 2.** Biochemical studies of full length TDG, TDG82-308 and TDG111-308. (A) Equilibrium binding of a given TDG construct to DNA (10 nM) containing a G-T^F mismatch, where T^F is a non-cleavable Thd analog, monitored by electrophoretic mobility shift assays (EMSA). The concentration of enzyme (nM) is indicated. For TDG111-308, the left lane marked ‘20*’ indicates binding to abasic DNA product (10 nM, 28 bp), to show the mobility of a tight 1:1 complex, given that such a complex is not clearly observed for G-T^F DNA in the gel. (B) Single-turnover kinetics for excision of T from a G-T substrate binding and catalytic events reported here were generated from DNA that gives 1:1 binding and include N-terminal residues beginning at 107 for TDG82-308 and at 111 for TDG111-308. These N-terminal
residues interact with other regions of the TDG catalytic domain, and feature an α-helix (α0; Glu16-Thr121) (Figure 3). Notably, residues that participate in the TDG:TDG dimer interface for the 2:1 complex (Leu143, Tyr147, Thr196 and Pro198) form contacts with the N-terminal helix (α0) and other N-terminal residues in the 1:1 complexes, which could explain why more N-terminal residues are disordered in structures featuring 2:1 versus 1:1 binding (Supplementary Figure S1).

The E-S and E-P structures for TDG82-308 reveal that the Arg110 side chain contacts the phosphate of the nucleotide located immediately 5′ of the flipped nucleotide (Figure 3). Mutational studies show that substrate binding is about 2-fold weaker and the maximal rate of base-excision (k_{max}) is 4-fold slower for R110A relative to wild-type TDG82-308 (Supplementary Figure S2). Thus, Arg110 contributes significantly to substrate binding and glycosylase activity. The structures also reveal that the backbone N-H of Asp109 contacts a phosphate in the complementary DNA strand. Structures of TDG82-308 exhibit reasonable electron density for the N-terminal residues Lys107-Arg110 (Supplementary Figure S3). The residues that contact DNA backbone phosphates, Asp109 and Arg110, have B-factors of 49.5 Å² and 44.5 Å² in the E-S complex and 46.4 Å² and 44.7 Å² in the E-P complex, respectively. Together, the structural and biochemical findings suggest that the minimal catalytic domain of TDG should be redefined to include Asp109 and Arg110.

Interestingly, the presence of residues 82–110 (or some fraction thereof) has a greater effect on overall protein structure than does a transition from the E-S to the E-P complex (for a given TDG construct). More specifically, the percent-based spread (p.b.s.) (59) between the E-S and E-P complexes is 0.18 Å for TDG82-308 and 0.11 Å for TDG111-308. However, when comparing corresponding structures for TDG82-308 versus TDG111-308, the p.b.s is 0.23 Å for the E-S complexes and 0.29 Å for E-P complexes.

High-resolution structure of an E-P complex for TDG82-308

We also solved a crystal structure of the TDG82-308 E-P complex at 1.70 Å, using crystals obtained by incubating the enzyme with a G-U substrate (Supplementary Table S1). As observed for the TDG82-308 E-S complex, Arg110 contacts the DNA phosphate 5′ of the flipped nucleotide, Asp109 (backbone N-H) contacts a phosphate in the complementary strand, but electron density is not observed for residues 82–106. The new TDG82-308 E-P complex is very similar to those we reported recently for TDG111-308 (33), as indicated by the backbone p.b.s. of 0.29 Å (noted above). As observed for our high-resolution TDG111-308 product complexes (33), the TDG82-308 structure shows unambiguously that the excised base (Ura) is absent from the E-P complex. Similarly, the abasic sugar adopts a roughly even mix of the α and β anomers (not shown). Although the product complexes of TDG111-308 and TDG82-308 exhibit very similar overall structures, the DNA conformation differs somewhat, particularly for several nucleotides 5′ of the flipped site (not shown), which may reflect DNA contacts involving Asp109 and Arg110 (lacking for TDG111-308). Additional differences are also noted in relevant sections below.

NMR studies of TDG82-308 N-terminal residues

The crystal structures indicate that the 26 N-terminal residues of TDG82-308 are disordered, even when the enzyme is bound to DNA in a tight E-S or E-P complex. Nevertheless, these N-terminal residues greatly enhance G-T substrate binding for TDG82-308 relative to TDG111-308 (Figure 2). We sought to further explore these findings using NMR chemical-shift-perturbation experiments, a powerful and widely used approach for monitoring protein–ligand interactions, particularly for disordered protein regions such as those of TDG (19,60–62). We compared the backbone 1H-15N chemical shifts for disordered residues of TDG82-308, in the presence and absence of DNA containing a G-TF substrate analog (same DNA used for binding studies, Figure 2). The 15N-HSQC spectrum for DNA-free TDG82-308 reveals about 30 resonances that are likely from disordered residues, as indicated by their relatively high intensity and chemical shifts similar to that expected for random coil (Figure 4, black peaks). Notably, nearly all of these peaks are absent in the 15N-HSQC of free TDG111-308 (Supplementary Figure S4), indicating that they likely reflect disordered N-terminal residues of TDG82-308. The same NMR
NMR studies indicate the N-terminal residues of TDG82-308 are disordered, when the enzyme is free or bound to G-TF DNA. Shown are 15N-HSQC spectra for TDG82-308 (0.13 mM) in the absence of DNA (black peaks) and with a saturating concentration (0.20 mM) of G-TF DNA (red peaks). Samples were in 0.02 M sodium phosphate pH 6.5, 0.15 M NaCl, 0.2 mM EDTA, 0.2 mM DTT, 7% D2O. Resonances in the upper right (within dotted lines) are side chain amino groups. NMR spectra were collected at 18°C using an 800 MHz NMR spectrometer. Four resonances are equivalent for free and DNA-bound TDG82-308 (*) and likely correspond to C-terminal residues (305NMDV308), which are far removed from the DNA-binding surface and not seen in crystal structures of TDG111-308 or TDG82-308 (23,33,34), including those reported here. This assignment is supported by observation that the same four resonances appear in NMR spectra for DNA-free TDG111-308 (Supplementary Figure S3).

experiment was collected for a sample containing TDG82-308 with a saturating concentration of G-TF DNA (Figure 4, red peaks). The overlaid spectra for free and G-TF-bound TDG82-308 reveal substantial chemical shift perturbations for most of the disordered N-terminal residues, indicating a change in the conformational ensemble upon binding G-TF DNA. Together, the crystallographic and NMR results indicate that the dramatically enhanced DNA binding affinity and glycosylase activity afforded by the TDG82-308 N-terminal residues is attained without adopting an ordered structure, suggesting non-specific interactions between the numerous cationic side chains (Lys, Arg) of TDG82-308 (Figure 1) and the anionic DNA phosphates.

Figure 5. Interactions with the flipped Ura base in the enzyme–substrate complex of TDG82-308 bound to G-UfF DNA (PDBID: 5HF7; 1.54 Å). TDG residues are in stick format (white with nitrogen blue and oxygen red), the Ura-containing DNA is yellow (with 2′-F colored cyan) and water molecules are red spheres. The 2Fo–Fc electron density map, contoured at 1.0 σ, is shown for the DNA and water molecules, but not the enzyme residues (for clarity). Dashed lines represent hydrogen bonds, with interatomic distances (Å).

Interactions with the flipped dUrd nucleotide in the enzyme–substrate complex

The new E·S structures of TDG82-308 and TDG111-308 bound to a G-UfF mismatch reveal detailed interactions with the flipped Ura base, some involving ordered water molecules that have not previously been observed in any structure of TDG or the related bacterial MUG enzymes (Figure 5). Notably, the E·S interactions shown in Figure 5 for TDG82-308 are also observed for TDG111-308 (not shown). The O2 of Ura forms two hydrogen bonds with TDG backbone N–H groups (Ile139, Asn140), as noted previously (34), and one of these contacts is relatively short (2.8 Å) in the new structures. Ura O2 forms a weak interaction with a water molecule, which is itself tightly bound by a backbone N–H and the Ser271 side chain. Notably, Ser271 is structurally analogous to the catalytic His residue of the related enzyme, uracil DNA glycosylase (UNG), which uses the catalytic His to contact Ura O2 directly (63). The O4 of Ura contacts a backbone amide (residue 152) of TDG and forms a close contact with a water molecule that is coordinated by the side chains of His151 and Asn191 and another ordered water molecule. The imino N3-H of Ura contacts the side chain oxygen of Asn191, confirming a previous observation (34). For uracil and its analogues (including thymine), these interactions are likely important for stabilizing the flipped
structures of product complexes for TDG111-308 (33). The filled channel was recently observed in high-resolution to the active site (Supplementary Figure S6). The solvent-...The Asn140 side chain is essential for TDG activity...The Asn140 side chain is essential for TDG activity...The Asn140 side chain is positioned by contactsto the Thr197 hydroxyl and a backbone oxygen (195). Given the substantial homology between TDG and MUG enzymes (34,71), this nucleophile-binding mechanism for TDG likely applies to MUG enzymes. This is significant because no putative nucleophile is observed the structure of an E-S complex for TDG82-308 bound to G-U\textsuperscript{F} DNA. TDG is shown in cartoon format with residues of interest in stick format (white with nitrogen blue and oxygen red), the flipped dUrd nucleotide is in yellow stick format, and the nucleophile water molecule is a red sphere (other waters not shown, for clarity). The 2\textsubscript{F}-F\textsubscript{2} map, contoured at 1.0 \sigma, is shown for the nucleophile water. Dashed lines represent hydrogen bonds, with interatomic distances (Å). The distance between the nucleophile water molecule and C1' of the flipped dUrd nucleotide (4.0 Å) is indicated by a thin dashed line.

**Figure 6.** Binding of the nucleophilic water molecule in the E-S complex of TDG82-308 bound to G-U\textsuperscript{F} DNA. TDG is shown in cartoon format with residues of interest in stick format (white with nitrogen blue and oxygen red), the flipped dUrd nucleotide is in yellow stick format, and the nucleophile water molecule is a red sphere (other waters not shown, for clarity). The 2\textsubscript{F}-F\textsubscript{2} map, contoured at 1.0 \sigma, is shown for the nucleophile water. Dashed lines represent hydrogen bonds, with interatomic distances (Å). The distance between the nucleophile water molecule and C1' of the flipped dUrd nucleotide (4.0 Å) is indicated by a thin dashed line.

reported here is excellent (Figure 6), defining clearly how TDG coordinates the water nucleophile and confirming our findings in the lower-resolution structure (34). The Asn\textsuperscript{140} side chain oxygen forms a short (2.7 Å) hydrogen bond to the putative nucleophilic water molecule. Loss of this interaction likely accounts, at least in part, for findings that the N140A mutation completely depletes G-T glycosylase activity and causes a huge (27 000-fold) loss in G-U activity (34). As shown in Figure 6, the nucleophile is also contacted by the backbone oxygen and side chain hydroxyl of Thr\textsuperscript{197}, a residue that is also strictly conserved and important for base excision, as indicated by findings that the T197A mutation causes a 32-fold reduction in G-T glycosylase activity (34). Previous findings that the N140A mutant retains a low level of glycosylase activity for some substrates (G-U, G-5FU) (39,72) is likely explained by contacts to the nucleophile from Thr\textsuperscript{197} and perhaps other enzyme groups. Notably, the Asn\textsuperscript{140} side chain is positioned by contacts to the Thr\textsuperscript{197} hydroxyl and a backbone oxygen (195). Given the substantial homology between TDG and MUG enzymes (34,71), this nucleophile-binding mechanism for TDG likely applies to MUG enzymes. This is significant because no putative nucleophile is observed the structure of an E-S complex for MUG bound to G-U\textsuperscript{F} DNA (43).

**Coordination of the water nucleophile**

The Asn\textsuperscript{140} side chain is essential for TDG activity (39,72), likely because it coordinates the nucleophilic water molecule. The previous structure of TDG\textsuperscript{111-308} bound to G-U\textsuperscript{F} suggested how TDG and MUG enzymes coordinate the water nucleophile (34). However, the putative nucleophile was the sole water molecule observed in the relatively low-resolution (2.97 Å) structure, and confidence in its assignment and placement was therefore somewhat limited. By contrast, the electron density for the two structures

conformation of the target nucleotide. Moreover, the contacts with O2 and O4 likely stabilize the anionic leaving group that is generated upon cleavage of the N-glycosyl bond (64).

We note that the sugar puckers of the flipped 2\textsuperscript{F}-dUrd nucleotide, C1'-exo-O4'-endo, is well defined by the electron density for E-S structures of TDG\textsuperscript{82-308} and TDG\textsuperscript{111-308} (Supplementary Figure S5). High-resolution structures of other 2\textsuperscript{F}-fluoroarabino deoxyribonucleotides, in free DNA or flipped into a glycosylase active site, reveal either the same or a similar sugar pucker, including O4'-endo and C2'-endo (65–68). Each of these conformations is observed in high-resolution structures of B-DNA, where pyrimidines often exhibit a pucker other than C2'-endo (69). A previous structure of the E-S complex for TDG\textsuperscript{111-308}_UF indicated a...
strictly conserved Arg 275 sidechain penetrate the DNA-
medi ted contacts for Arg 275 (Figure 7), most of which have
(nucleophile to C1′ the nucleophile–electrophile distance is reduced to 2.8 ˚A,
analog of the glycosyl cation intermediate reveals that
via nucleophile migration (73).

The cationic Arg 275 side chain directly contacts each of the
generating cleavage of the N-glycosyl bond and migration of the
electrophile (Cl′ of an oxacarbenium ion intermediate) and possibly the nucleophile (73). As we noted previously (34),
the proximity and relative position of the nucleophile and
electrophile (Cl′) observed for the E-S complex of TDG is
nearly identical to that observed in a high-resolution (1.8 Å)
structure of an E-S complex for the related enzyme UNG
(nucleophile to Cl′ distance of 3.5 Å) (63). Notably, a structure
(1.9 Å resolution) of UNG bound to DNA containing
analog of the glycosyl cation intermediate reveals that the nucleophile–electrophile distance is reduced to 2.8 ˚A,
via nucleophile migration (73).

Arg ‘plug’ residue

Previous structures of DNA-bound TDG show that the
strictly conserved Arg275 side chain penetrates the DNA mi-
nor groove and fills a void generated by flipping of the dUrd nucleotide. TDG
is shown in cartoon format with key residues in stick format (white with
nitrogen blue and oxygen red). The Ura-containing DNA is yellow and
the complementary strand green; water molecules are red spheres. Dashed
lines represent hydrogen bonds, with interatomic distances (Å) shown.
The ‘opposing G’ is the Gua of the G-U·F mispair. The contact involving
Arg110, unique to TDG82-308, is also shown.

Arg 275 ‘plug’ conformation of Arg275, and thereby stabilize nu-
ucleotide flipping. Moreover, Pro198 is in a loop that would
block reverse flipping of the target nucleotide out of the ac-
tive site and back into the DNA duplex. Together, these ef-
tects might contribute to the much tighter binding to G-T
DNA observed for TDG82-308 relative to TDG111-308 (Figure 2).
Because Pro198 flanks Thr197, which helps coordinate the
water nucleophile, the Arg275-Pro198 contact might also
serve to couple nucleotide flipping with the chemical step
or help to properly position the nucleophilic water. Such a
mechanism might account in part for findings that the max-
imal enzymatic rate (kmax) for G-T activity is 4-fold higher
for TDG82-308 relative to TDG111-308 (Figure 2).

Notably, the closely related bacterial MUG enzymes have a
Leu rather than an Arg residue in the corresponding ‘plug’
position (43,74), which cannot form any of the direct or
water-mediated contacts observed here for TDG. MUG
acts on G-U and other substrates but not on G-T mispairs.
An Arg residue serves as the ‘plug’ in MBD4 and MIG
enzymes (75–78), which act on G-T (and G-U) mispairs
but are unrelated to TDG. These observations suggest that
some of the detailed electrostatic contacts observed here
for the Arg plug of TDG may be generally important for
G-T glycosylase activity and may be conserved for MBD4
and MIG. Indeed, for MBD4, the Arg ‘plug’ forms similar
contacts with the two phosphates flanking the flipped
nucleotide as observed for TDG (75,77). No DNA-bound
structures have been reported for MIG.

Interactions that may confer sequence context

TDG exhibits a strong preference for excising thymine,
uracil and 5-substituted uracil analogs when these bases are
flanked by a 3′ Gua and base-paired with Gua rather than
Ade (35,44,79). Structures of the two G-U·F complexes re-
ported here (TDG82-308 and TDG111-308) validate the con-
tacts to these two Gua bases that were suggested by a pre-
vious structure of TDG111-308 with G-U·F solved at moderate
(2.97 Å) resolution (34). In the previous structure, poten-
tial contacts to the opposing Gua from backbone oxygens
were suggested, but the interatomic distances were rather
long (d ≥ 3.6 Å), indicating weak interactions. By contrast,
the two high-resolution E-S structures reported here reveal
three clear contacts to the opposing Gua (d ≤ 3.0 Å) (Sup-
plementary Figure S7). The backbone oxygens that contact
the opposing Gua would likely present a repulsive environ-
ment to the corresponding regions of Ade (N1, C2), sug-
uggesting that specificity against canonical A-T pairs might
involve repulsion of Ade as the pairing partner for Thy or
other uracil analogs as the target base.

The new structures also demonstrate that the Gua located
3′ of the flipped Ura is contacted (at its N2H2) by the side
chain of Gln278 and the backbone nitrogen of Ala277 (Fig-
ure 8). These contacts, which were not seen in the previous
structure of the E-S complex (TDG111-308 with G-U·F), offer
an explanation for findings that TDG excision of Thy, and
Ura analogues, is most efficient when the 3′ base is Gua, that
Figure 8. TDG contacts two bases on the 3′ side of the flipped nucleotide, including the 3′-Gua. TDG82-308 is shown in surface representation with residues that contact the 3′-Gua shown as sticks. The dUrd-containing DNA strand is yellow. For clarity, the complementary strand is not shown except for the opposing Gua (green). Water molecules are red spheres. Dashed lines represent hydrogen bonds with interatomic distances (Å).

is, a CpG (or CG) sequence context (relative activity: XpG >> XpA > XpC > XpT; where X is Thy or a Ura analog) (35). The structures also reveal that the side chain of Glu278 and potentially Lys201 can contact the second base located 3′ to the flipped Ura (Figure 8). While Lys201 does not contact the Cyt base located two nucleotides away from dUrd for TDG82-308, it does so for the TDG111-308 structure reported here (not shown). This could be significant because cytosine methylation (mC) is found in non-CG sites, with mCAB and mCAG observed most frequently (80,81). Our structures suggest TDG would not form specific interactions with a 3′ Ade (relative to the flipped nucleotide) but can potentially contact Cyt and perhaps Gua in the second position located 3′ of the flipped nucleotide.

CONCLUSION

Our results indicate that the 29 N-terminal residues of TDG82-308 confer enhanced substrate binding and faster base excision such that the activity of TDG82-308 is equivalent to that of intact TDG and much greater than TDG111-308. At the same time, crystal structures and NMR studies indicate that most (25) of these N-terminal residues are disordered, even when TDG82-308 is tightly bound to G·Tor G·U substrate DNA. Thus, the enhanced biochemical activity provided by the N-terminal residues is attained without a gain in ordered structure. It will be of interest to investigate the nature of the disordered state of these N-terminal residues in future studies. One possibility might be that they participate in the search for specific sites and that this function might lead to the observed disorder through exchange between search and recognition conformations. Another possibility is that these residues, many of which are cationic (Lys, Arg), form non-specific and transient and interactions with the DNA backbone, thereby enhancing the binding affinity of TDG for specific and non-specific sites. In addition, transient DNA interactions for the N-terminal residues might enable TDG to bind DNA and allow residues 95–106, the PIP degron, to interact with PCNA and the E3 ligase CRL4^Cdt2 to mediate degradation of TDG in S phase (27–29). The new high-resolution structures of TDG82-308 reported here reveal new and important enzyme–DNA contacts, some of which are not observed in corresponding structures for the smaller TDG111-308 construct. Together, our findings indicate that TDG82-308 is superior to TDG111-308 as a model for studying the structure and function of TDG. Our findings also suggest that the catalytic domain of TDG should be redefined to include Asp109 and Arg110, which contact the DNA substrate but are not found in TDG111-308.

ACCESSION NUMBERS

Coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org/) with accession numbers 5HF7, 5FF8 and 5JXY.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office (DOE) by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health (NIH), National Institute of General Medical Sciences (NIGMS; including P41GM103393) and the National Center for Research Resources (NCRR; P41RR001209). The Berkeley Center for Structural Biology is supported in part by the NIH, NIGMS and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the DOE under Contract No. DE-AC02-05CH11231. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS, NCRR or NIH.

FUNDING

National Institutes of Health [GM072711 to A.C.D. in part]; National Institutes of Health [S01-OD011969 to Support for procuring the imaging system (GE Typhoon FLA 9500)]. Funding for open access charge: National Institutes of Health (NIH) [grant GM072711 to A.C.D.].

Conflict of interest statement. None declared.
REFERENCES

1. Bellacosa, A. and Drohat, A.C. (2015) Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair (Amst), 32, 33–42.

2. Neddermann, P. and Jiricny, J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. J. Biol. Chem., 268, 21218–21224.

3. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiedberg, K. and Jiricny, J. (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. J. Biol. Chem., 271, 12757–12774.

4. 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.

5. 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.

6. 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.

7. 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.

8. Ito, S., Shen, L., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, X. (2011) Tet proteins can convert 5-Methylcytosine to 5-Carboxylcytosine and its excision by TDG in mammalian DNA. Science, 333, 1300–1303.

9. Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deilmann, I., Waugh, D.S. and Waugh, D.S. (2008) 5-Formylcytosine and 5-Carboxylcytosine. J. Biol. Chem., 283, 319–326.

10. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D’Alessio, A.C., Fung, H.L., Zhang, K. and Zhang, Y. (2013) Genomewide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell, 153, 692–706.

11. Bennett, M.T., Rodgers, M.T., Hebert, A.S., Ruslander, L.E., Eisele, L. and Drohat, A.C. (2006) Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. J. Am. Chem. Soc., 128, 12510–12519.

12. Cortazar, D., Kunz, C., Saito, Y., Steinacher, R. and Schar, P. (2007) The enigmatic thymine DNA glycosylase. DNA Repair (Amst), 6, 489–504.

13. Tini, M., Benecke, A., Um, S.J., Torcia, 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.

14. Maiti, A., Noon, M.S., Mackerell, A.D., Jr, Pozhariski, E. and Drohat, A.C. (2012) 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., 109, 8091–8096.

15. 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. Nucleic Acids Res., 33, 4383–4393.

16. Tropea, J.E., Cherry, S. and Waugh, D.S. (2009) Expression and purification of soluble His(6)-tagged TEV protease. Methods Mol. Biol., 498, 297–307.

17. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem., 182, 319–326.

18. Morgan, M.T., Maiti, A., Drohat, A.C. and Drohat, A.C. (2011) Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. Nucleic Acids Res., 39, 2319–2329.

19. Maiti, A. and Drohat, A.C. (2008) Crystal structure of mammalian DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. Proc. Natl. Acad. Sci. U.S.A., 105, 8890–8895.

20. Hashimoto, H., Hong, S., Bhagwat, A.S., Zhang, Y. and Walter, J.C. (2014) Thymine DNA glycosylase is a CRL4Cdt2 substrate. J. Biol. Chem., 289, 23043–23055.

21. Baba, D., Maiti, A.,$n$, Lee, G., Uchiuma, Y., Saitoh, H., Sugawa, K., Hanaoka, F., Tochio, H., Hiroaki, H. and Shirakawa, M. (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. Nature, 435, 979–982.

22. Parikh, A., Filipov, D., Sato, T., Goto, S. and Drohat, A.C. (2007) Thymine DNA glycosylase is a CRL4Cdt2 substrate. J. Biol. Chem., 282, 23043–23054.

23. Hsuan, J.J., Wiebauer, K. and Jiricny, J. (1996) Cloning and expression of the apparently redundant uracil-DNA glycosylase from human DNA glycosylase. EMBO J., 15, 6207–6218.

24. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem., 182, 319–326.

25. Maiti, A. and Drohat, A.C. (2008) Crystal structure of mammalian DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. Proc. Natl. Acad. Sci. U.S.A., 105, 8890–8895.

26. Tropea, J.E., Cherry, S. and Waugh, D.S. (2009) Expression and purification of soluble His(6)-tagged TEV protease. Methods Mol. Biol., 498, 297–307.

27. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem., 182, 319–326.

28. Morgan, M.T., Maiti, A., Drohat, A.C. and Drohat, A.C. (2011) Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. Nucleic Acids Res., 39, 2319–2329.

29. Maiti, A. and Drohat, A.C. (2008) Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair (Amst), 32, 33–42.

30. McLaughlin, D., Coey, C.T., Yang, W.C., Drohat, A.C. and Matunis, M.J. (2016) Characterizing requirements for SUMO modification and binding on base excision repair activity of thymine DNA glycosylase in vivo. J. Biol. Chem., 291, 9014–9024.

31. Steinacher, R. and Schar, P. (2005) Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. Cell, 15, 616–623.

32. Maiti, A., Morgan, M.T., Pozhariski, E. and Drohat, A.C. (2008) Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. Proc. Natl. Acad. Sci. U.S.A., 105, 8890–8895.
40. Manvilla,B.A., Varney,K. M. and Drohat,A.C. (2009) Chemical shift assignments for human apurinic/apyrimidinic endonuclease 1. *Biomol. NMR Assign.*, 4, 5–8.
41. Amburgey,J.C., Abildgaard,F., Starich,M.R., Shah,S., Hilt,D.C. and Weber,D.J. (1995) 1H, 13C and 15N NMR assignments and solution secondary structure of rat Apo-S100 beta. *Biochemistry, 6*, 171–179.
42. Maiti,A., Drohat,A.C. (1997) Investigation of the mechanisms of DNA binding of the human G/T glycosylase using designed inhibitors. *Proct. Natl. Acad. Sci. U.S.A., 94*, 4878–4883.
43. Buechner,C.N., Maiti,A., Drohat,A.C. and Tessmer,I. (2015) Lesion structures of B-DNA with incorporated 2-deoxy-2-fluoro-arabinono-furanosyl thymines: implications of conformational preorganization for duplex stability. *Nucleic Acids Res.*, 26, 2473–2480.
44. Lee,S. and Verdin,G.L. (2009) Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A., 106*, 18497–18502.
45. Bowman,B.R., Lee,S.M., Wang,S.Y. and Verdin,G.L. (2008) Structure of the E.coli DNA glycosylase AlkA bound to the ends of duplex DNA: A system for the structure determination of lesion-containing DNA. *Structure, 16*, 1166–1174.
46. Evans,P.R. (2011) An introduction to data reduction: Space-group search and recognition by thymine DNA glycosylase revealed by crystallographic model and data quality. *Science, 336*, 1030–1033.
47. Mann,P., Emsley,P. and Cowtan,K.D. (2011). Global Phasing Ltd., Cambridge.
48. Winn,M.D., Ballard,C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, 67, 235–242.
49. Berti,P.J. and McCann,J.A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycosidase hydrolysis and N-glycoside transfer. *Chem. Rev.*, 106, 595–585.
50. Berti,P.J. and McCann,J.A. (2006) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem., 275*, 33449–33456.
51. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.*, 61, 458–464.
52. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.*, 57, 122–133.
53. Bercovier,W. and Cowtan,K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2126–2132.
54. Manilla,B.A., Varney,K.M. and Drohat,A.C. (2002) Structure and activity of a thermostable thymine-DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A., 99*, 26786–26792.
55. Maiti,A., Drohat,A.C. (2011) 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.), 10*, 545–558.
56. Maiti,A., Michelson,A.Z., Armwood,C.J., Lee,J.K. and Drohat,A.C. (2014) Mechanisms for enzymatic cleavage and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: evidence for base twisting to remove mismatched normal base pairs. *Acta Crystallogr., 13*, 15813–15822.
57. Maich,A. and Drohat,A.C. (2011) 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.), 10*, 545–558.
58. Maiti,A. and Drohat,A.C. (2013) Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxycytosine from DNA. *J. Am. Chem. Soc., 135*, 15813–15822.
59. Maiti,A. and Drohat,A.C. (2011) 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.), 10*, 545–558.