5-2023

Biochemical Analysis of DNA Glycosylase in Dragonfish (Scleropages Formosus, Sfo) TDG

Jenna Perry

Biochemical Analysis of DNA Glycosylase in Dragonfish (Scleropages Formosus, Sfo) TDG

Jenna Perry*, Dr. Chenyan Chang, Dr. Weiguo Cao

April 28th, 2023

Abstract

TDG, which is a member of the uracil DNA glycosylase superfamily, plays a critical role in the active demethylation process in mammals by removing oxidized derivatives of 5-methylcytosine (mC). While human TDG has been extensively studied and found to possess 5-formylcytosine and 5-carboxylcytosine DNA glycosylase activity, little is known regarding this demethylation process in vertebrates other than the human species. Previous research proposed a catalytic sequence in motif 1 of zebrafish TDG1 and TDG2 that modulates DNA glycosylase activity based on mutational, kinetic, and modeling analyses of these enzymes. Building on this work, researchers discovered that the Sfo TDG is an exclusive 5-carboxylcytosine-glycosylase and that a strong correlation exists between the -IGHHYPN- of motif 1 and this activity. By converting I190-H192 within -IGHHYPN- to K190-R192, researchers were able to oppose the decrease in 5-carboxylcytosine activity from the single mutagenesis of I190K and H192R, reaffirming the heavy correlation existing between these residues. Since I190K-H192R maintain such a strong correlation, the Sfo structure was predicted and displayed the role of this double positive mutation in positioning the adjacent H193 to stabilize the leaving group in the transition state of catalysis.The study expanded our knowledge of TDG catalyzation in other species besides human and has important implications in studying protein structural determinants and their functions.

Introduction

The Uracil DNA Glycosylase superfamily, or UDG superfamily, serve an important function as DNA repair enzymes. The main role of these enzymes include recognizing and

cleaving base damage in the DNA and then initiating the base excision repair pathway (BER). It is important to note that DNA is only meant to contain thymine, adenine, cytosine, and guanine.The deamination of cytosine to uracil occurs spontaneously. This deamination process is significant in the communication between immune cells and pathogens [1]. UDG enzymes specifically excise uracil from DNA via cutting the N-glycosidic bond between the uracil and the deoxyribose sugar on the DNA backbone [2]. After the UDG enzyme excises the uracil by cutting it out of the DNA, an AP site remains in the helix. Uracil will be replaced by cytosine in the BER pathway ultimately.

The UDG superfamily can be distinctly separated into at least 6 families on the basis of motifs 1, 2 and 3, three highly conserved sequences in different organisms. This paper focuses on a member of this UDG superfamily, the thymine-DNA glycosylase in family 2. The first TDG was found in humans and could recognize 5-formylcytosine (fC) and caC (carboxylcytosine), both of which can result from the demethylation process or DNA damage. DNA methylation is important in epigenetics, such as epigenetic silencing. It can silence genes that are important in transcription, cell proliferation, cell cycle control, etc. Active demethylation is a complex process. First, a TET enzyme oxidizes a 5-methylcytosine (mC) to a 5-hydroxymethylcytosine (hmC). This 5hmC is further oxidized to fC and 5caC by TET. It is then recognized by the TDG which cuts fC and 5caC out of the DNA, leaving an AP site. Finally, through the BER pathway, it is converted back to an unmethylated cytosine.

The *Scleropages formosus* (Sfo) has only one TDG and acts only on caC, at low levels, known from previous research on the TDG in fish species. However, there are duplicated TDGs found in other fish, for example, zebrafish. TDG1 from the zebrafish is similar to the human TDG, having a broad substrate preference. But for TDG2, comparatively, it acts exclusively at

high levels for caC. It is possible that the sequence responsible for this catalytic difference lies in the motif1 differences between TDG1 and TDG2 from previous studies **[ ]**. In this research, we want to investigate the catalytic activity mechanism of the Sfo TDG and why it maintains this exclusive activity for caC. Sfo TDG will be subjected to mutagenesis, enzyme activity analyses, and structural modeling. This study expands our understanding of TDG in other species besides mammals.

Materials and Methods

Reagents and media

All routine chemical reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), or VWR (Suwanee, GA), and all buffers were prepared in high-quality deionized water from a Thermo Scientific Nanopure Water System (Suwanee, GA) with a resistivity greater than 18.2 MΩ.cm. Restriction enzymes, Phusion High-Fidelity DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). HisTrap FF and HiTrap SP FF columns were purchased from GE Healthcare Life Sciences (Piscataway, NJ). Hi-Di Formamide and GeneScan 500 LIZ dye Size Standard for ABI 3130xl were purchased from Applied Biosystems. Sonication buffer consisted of 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.1% Triton X-100, and 40 mM imidazole with freshly added 1 mM dithiothreitol (DTT) and 0.15 mM phenylmethylsulfonyl fluoride (PMSF). Buffer A consisted of 50 mM Tris-HCl (pH 7.5), 400 mM NaCl, and 10% glycerol. Buffer B consisted of 20 mM Tris-HCl (pH 7.5), 400 mM NaCl, 10% glycerol and 500 mM imidazole. 1x Protein storage buffer

consisted of 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 1 mM DTT, 1 mM EDTA, and 50% Glycerol.

Oligodeoxynucleotide substrates

The substrate nucleobases and sequences are highlighted in **Fig 4.1**.The oligonucleotides used for the DNA glycosylase activity assay were prepared as previously described [16]. Briefly, the oligodeoxynuleotide containing uracil (U), cytosine (C), thymine (T), 5-mythlcytosine (5mC), 5-hydroxymethylcytosines (5hmC), 5-formylcytosine (5fC), 5-carboxylcytosine (5caC), 5-hydroxymethyluracil (5hmU) strand (10 μM) was mixed with the unlabeled strand in 1.5-fold molar excess, incubated at 85°C for 3 min, and allowed to form duplex DNA substrates at room temperature for more than 30 min.

Cloning, expression and purification of Sfo TDG with site-directed mutagenesis

The DNA sequence of Sfo TDG was synthesized by Eurofins Genomics LLC. TDG from Scleropages formosus (GenBank accession number: XP_018585620) was amplified with the forward primer SfoTDG_F (5'- CCGGAATTC CGCTTCAATGGCATG -3'; the EcoRI site is underlined) and the reverse primer SfoTDG_R (5'- CCCAAGCTT CACGCCTCTTAGC -3'; the HindIII site is underlined).

The PCR procedure included a pre-denaturation step at 98 °C for 30 s; 30 cycles of three-step

amplification with each cycle consisting of denaturation at 98 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72 °C for 20 s; and a final extension step at 72 °C for 10 min.

The PCR product was purified and cloned into pET21a vector. The recombinant plasmid was confirmed by DNA sequencing. Briefly speaking, purified PCR products and plasmid pET21a(+) were digested with appropriate restriction enzymes, and ligated using T4 ligase. Ligation mixtures were transformed into E. coli DH5α competent cells. The resulting sequences encoding a C-terminal His-6-tag following protein of interest were confirmed by DNA sequencing.

The resulting plasmid with wild-type Sfo TDG was then used as the template plasmid for all Sfo TDG mutants. The PCR procedure included a pre-denaturation step at 98°C for 2 min; 16 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 30 s, annealing at 65°C for 30 s, and extension at 72°C for 7 min, and a final extension step at 72°C for 15 min. After treatment with 2 units of DpnI for 3 hours at 37°C, 5 μL of PCR products were transformed into E. coli DH5α competent cells. Successful mutation in the resultant clones were confirmed by DNA sequencing.

Table 1. Oligonucleotide primer sequences used in site-directed mutagenesis. Mutated sites are underlined.

Primer	Sequence
M186L_F	GAATTAACCCCGGACTGTTGTCAGCCTATATAGGA
M186L_R	TCCTATATAGGCTGACAACAGTCCGGGGTTAATTC
I190K_F	CGGACTGATGTCAGCCTATAAAGGACATCATTACC
I190K_R	GGTAATGATGTCCTTTATAGGCTGACATCAGTCCG
H192R_F	TGTCAGCCTATATAGGACGTCATTACCCAAATCCAGG
H192R_R	CCTGGATTTGGGTAATGACGTCCTATATAGGCTGACA
H193W_F	CGGACTGATGTCAGCCTATATAGGACATTGGTACCCAAATCCAGGA
H193W_R	TCCTGGATTTGGGTACCAATGTCCTATATAGGCTGACATCAGTCCG
N196G_F	TGTCAGCCTATATAGGACATCATTACCCAGGTCCAGGAAACCATTT
N196G_R	AAATGGTTTCCTGGACCTGGGTAATGATGTCCTATATAGGCTGACA
I190K-H192R_F	CGGACTGATGTCAGCCTATAAAGGACGTCATTAC
I190K-H192R_R	GTAATGACGTCCTTTATAGGCTGACATCAGTCCG
H192R-H193W_r	CTGATGTCAGCCTATATAGGACGTTGGTACCCAAATCCAGGAAACCATTTT
H192R-H193W_R	AAAATGGTTTCCTGGATTTGGGTACCAACGTCCTATATAGGCTGACATCAG
H193W-N196G_R	CGGACTGATGTCAGCCTATATAGGACATTGGTACCCAGGTCCAGGA
H193W-N196G_R	TCCTGGACCTGGGTACCAATGTCCTATATAGGCTGACATCAGTCCG

To express the C-terminal His6-tagged wild-type Sfo TDG and mutant SfoTDG, the recombinant plasmids were then transformed into E. coli strain BH214 (ung - , mug -) by the standard protocol to express the C-terminal His-6-tagged proteins. An overnight E. coli culture from a single colony transformed with recombinant plasmid was diluted 100-fold into LB medium (500 ml) supplemented with 100 mg/ml ampicillin and grown at 37°C with shaking at

250 rpm until the optical density at 600 nm reached approximately 0.4. After adding isopropyl-1-thio-α-d-galactopyranoside (IPTG) to a final concentration of 1 mM, the culture was grown at 20°C for an additional 18 h. The cells were collected by centrifugation at 5000 rpm with a JLA-81000 rotor at 4°C. To purify the target protein, the cell pellet from a 500 ml culture was suspended in 7 ml of sonication buffer and sonicated at output 5 for 3 × 1 min with 5 min rest on ice between intervals using Qsonica model Q125. The sonicated solution was clarified by centrifugation at 12 000 rpm with a JLA-16.250 rotor at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded onto a 1 ml HisTrap FF column. The column was washed with 20 ml of Buffer A. The bound protein in the column was eluted with a linear gradient of 0–100% Buffer B. Fractions containing the target protein were identified by SDS-PAGE, pooled, diluted 3-fold, and applied to a 1 ml HiTrap SP FF column, which was pre-equilibrated withBuffer A without imidazole (20 mM Tris-HCl (pH 7.5), 300 mM NaCl). By linear gradient elution (100 mM–1000 mM NaCl), fractions containing target protein were identified by SDS-PAGE, pooled, concentrated, and exchanged with a storage buffer through Microcon YM 10 (Millipore). The protein concentration was quantified by the Bradford method using bovine serum albumin as a standard. The purified proteins were stored in aliquots at −20°C. The final purified proteins were evaluated by SDS-PAGE and Coomassie Brilliant Blue staining (Fig. 3.2).

DNA glycosylase assay

DNA glycosylase assay for SfoTDG and mutants were performed at 37°C for 60 min in a 10 µl reaction mixture containing 10 nM oligonucleotide substrate, 1000 nM enzyme, 20 mM

Tris-HCl buffer (pH 7.5), 350 mM KCl, 1mM DTT, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95°C for 5 min after adding 1 µl of 1 M NaOH. The reaction mixtures (2 µl) were mixed with 7.8 µl Hi-D formamide and 0.2 µl GeneScan 500 LI Size Standard and analyzed by Applied Biosystems 3130xl genetic analyzer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

Results

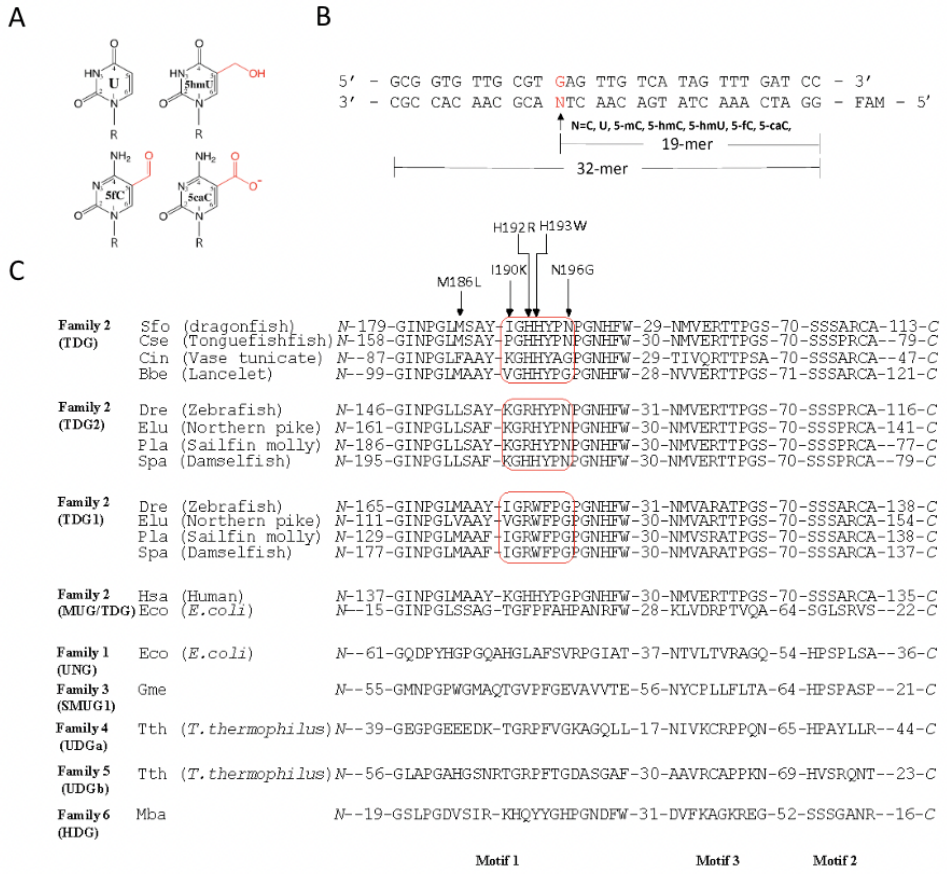

Figure 1. Sequence alignment of fish TDG with other known glycosylase superfamily Enzymes.

GenBank accession numbers are shown after the species names. Family 2 TDG: Sfo, *Scleropages formosus*, XP_018585620; Cse, *Cynoglossus semilaevis*, XP_008309716; Cin, *Ciona intestinalis*, XP_009861570; Bbe, *Branchiostoma belcheri*, XP_019614398. Family 2 TDG2: Dre, *Danio rerio*, XM_009297976; Elu, *Esox Lucius*, XP_010867260.1; Pla, *Poecilia latipinna*, XP_014876710.1; Spa, *Stegastes partitus*, XP_008292365.1. Family 2 TDG1: Dre, *Danio rerio*, NM_001020751; Elu, *Esox Lucius*, XP_010867259.1; Pla, *Poecilia latipinna*, XP_014876717.1; Spa, *Stegastes partitus*, XP_008292409. Family 2 MUG/ TDG: Hsa, *Homo Sapiens*, NP_003202; Eco, *Escherichia coli*, P0A9H1. Family 1 UNG: Eco, *Escherichia coli*, NP_289138; Family 3 SMUG1: Gme, *G. metallireducens* GS-15, YP_383069. Family 4 UDGa: Tth, *Thermus thermophilus* HB27, YP_004341.1. Family 5 UDGb: Tth, *Thermus thermophilus* HB8, YP_144415.1; Family 6 (HDG): Mba, *Methanosarcina barkeri str*. Fusaro, YP_304295.1.

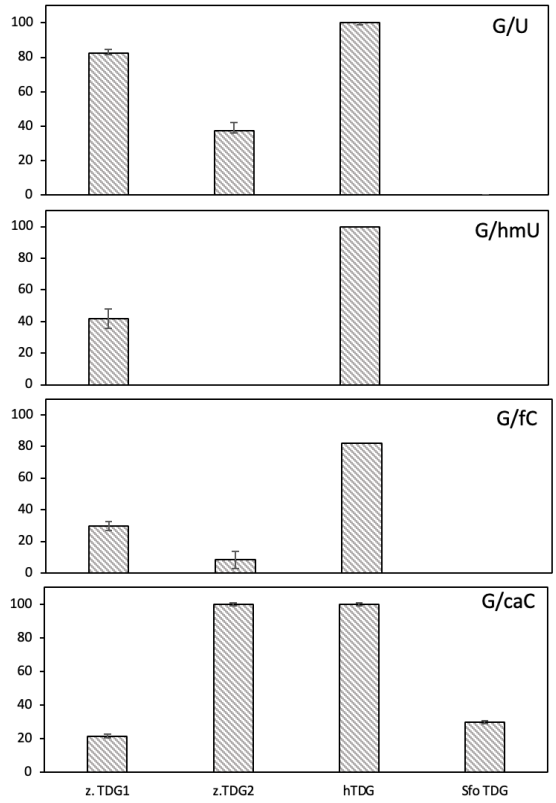

Figure 2. General glycosylase activity comparison of Sfo TDG to other TDG enzymes in Family 2. Enzyme activity screening on G/fC, G/caC, G/U and G/hmU. TDG enzyme activity screenings were performed at 37°C for 1 hour in a reaction mixture (10 µl) containing 1000 nM UDG superfamily protein, 10 nM C, T, mC, hmC, CaC, fC, U and hmUcontaining dsDNA substrate, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 1 mM DTT. Reactions were quenched by addition of 1 µl of 1 M NaOH. The reaction mixtures (2 µl) were mixed with 7.8 µl Hi-D formamide and 0.2 µl GeneScan 500 LI Size Standard and analyzed by Applied Biosystems 3130x1 genetic analyzer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software. All experiments were performed in triplicate, and all data are expressed as the mean ± standard deviation.

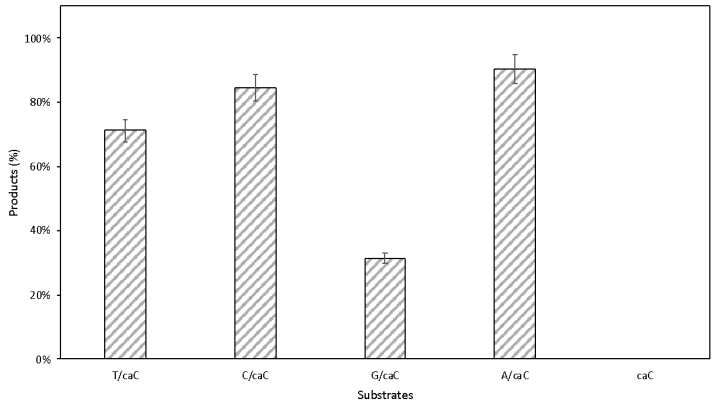

Figure 3. DNA glycosylase activity from Sfo TDG on caC-containing DNA substrates with different opposite bases. DNA glycosylase cleavage assays were performed at 37°C for 60 min in a reaction mixture (10 µl) containing 1000 nM Phe protein, 10 nM DNA substrate, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 1 mM 2-mercaptoethanol. Reactions were quenched by addition of 1 µl of 1 M NaOH. Samples (3.5 µl) were loaded into Applied Biosystems 3130x1 genetic analyzer and analyzed by GeneMapper software as previously described. All experiments were performed in triplicate, and all data are expressed as the mean ± standard deviation.

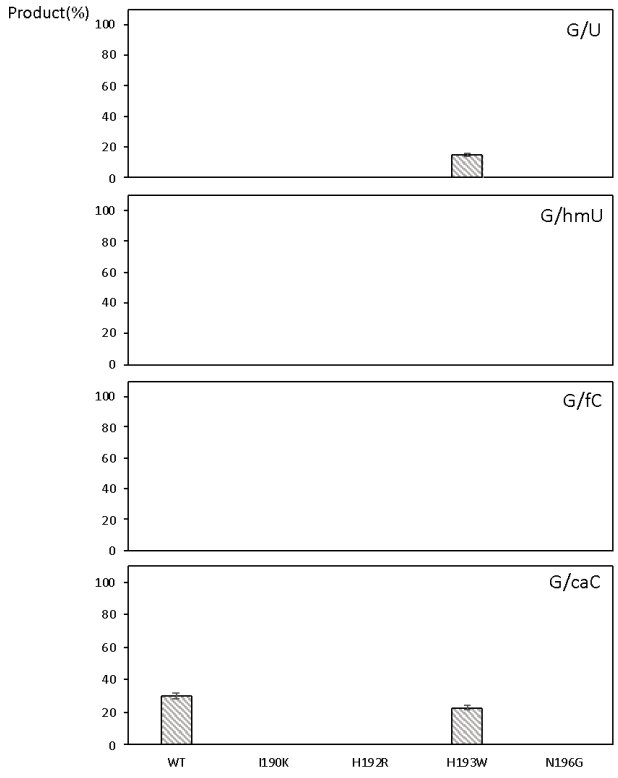

Figure 4. Sfo TDG mutants enzyme activity on G/U, G/hmU, G/fC, G/caC substrates. DNA glycosylase cleavage assays were performed at 37°C for 60 min in a reaction mixture (10 µl) containing 1000 nM Sfo protein, 10 nM DNA substrate, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 1 mM 2-mercaptoethanol.

Reactions were quenched by addition of 1 µl of 1 M NaOH. The reaction mixtures (2 µl) were mixed with 7.8 µl Hi-D formamide and 0.2 µl GeneScan 500 LI Size Standard and analyzed by Applied Biosystems 3130x1 genetic analyzer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software. All experiments were performed in triplicate, and all data are expressed as the mean ± standard deviation.

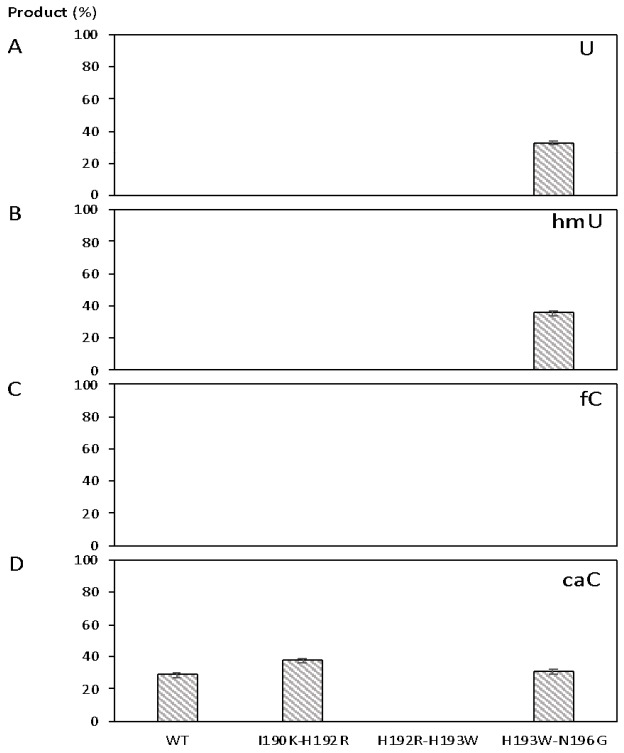

Figure 5. Sfo TDG motif 1 double mutants enzyme activity on G/U, G/hmU, G/fC, G/caC substrates. DNA glycosylase cleavage assays were performed using the same methods and analysis as described in single mutational analysis. All experiments were performed in triplicate, and all data are expressed as the mean ± standard deviation.

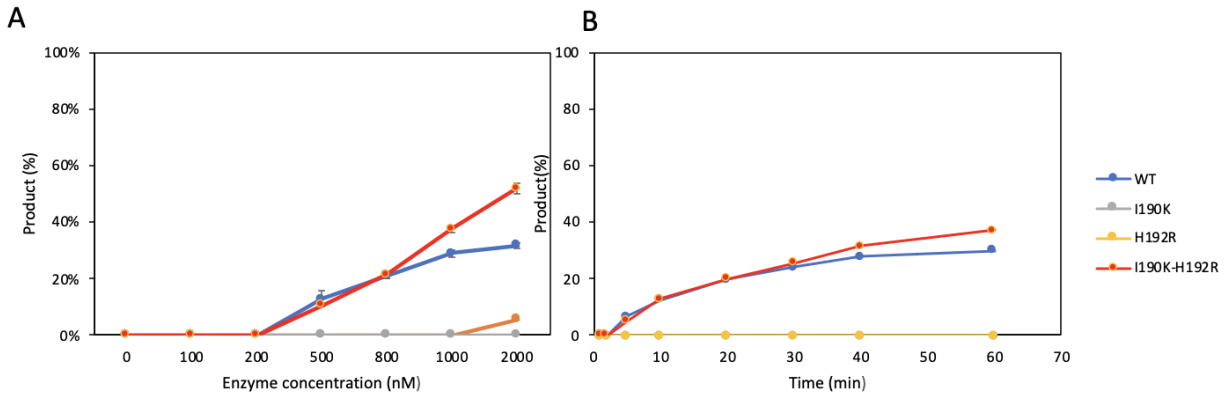

Figure 6. Catalytic analysis with increasing enzyme concentration and time course analysis of Sfo TDG wild type, I190K mutant, H192R mutant and I190K-H192R double mutant enzymes on 5caC. A. Titration study of enzymes on caC activity. The reaction were conducted at different concentration of enzymes range from 100 nM to 2000nM at same condition as of previous analysis. Samples were collected at 60min manually. All tests were performed in triplicate, and all data are expressed as the mean ± standard deviation. **B.** Reaction curve of enzymes on caC-containing substrate. The reaction were conducted at enzyme concentrations of 1000nM with an initial substrate concentration of 10 nM. Samples of wild type Sfo TDG reaction were collected at 5 min, 10 min, 20 min 30 min, 40 min and 60 min manually. Samples of Sfo TDG mutants reaction were collected at 1 min, 2 min, 5 min, 10 min, 20 min, 30 min, 40 min and 60 min manually.

Discussion and Conclusion

The first step in our experiment was the determination of Sfo glycosylase activity. The three significant variations in motif **1** the three TDG enzymes were compared using sequence alignment: -IGHHYPN- in Sfo TDG, -KGRHYPN- in zebrafish TDG2 and -IGRWFPG- in zebrafish TDG1 (Figure 1). There were single mutations in four sites of motif 1. In Sfo TDG, lysine replaced isoleucine (I190K). This single mutant produced catalytic activity most alike to zTDG. No enzymatic activity was observed on all substrates in the replacement of isoleucine to lysine (K157I) and to alanine (K157A) in zTDG. These results demonstrated the necessity of this specific lysine residue in the activity of zTDG. Additionally, the Sfo-TDG-I190K displayed the absence of activity on all of the substrates (Figure 4**)**. These results imply that I190K could be both structurally and functionally incompatible next to the other adjacent residues. Next,

arginine, a conserved site in zTDG1 and zTDG2, replaced histidine, a conserved site in Sfo TDg and hTDG (H192R). There was also no activity on all of the substrates from this mutation (Figure 4). Tryptophan replaced histidine (H193W) which altered Sfo TDg activity to be more alike to zTDG1. The single mutation of H160W in zTDG2 ultimately altered the catalytic characteristics of zTDG, and demonstrated similarity to zTDg1 (Figure 4). The single mutation H193W in Sfo TDG demonstrated similar activity on the caC substrate and also showed increased activity on the U substrate(Figure 4) So this mutation enhanced activity on different substrates as seen in zTDG2-H160W (Figure 4). Finally, glycine was replaced by asparagine (N196G). In zTDG1 and zTDG2, this particular residue is different compared to human TDG but contains the same asparagine residue as Sfo TDG. The single mutation N193G altered Sfo activity to be more similar to zTD1. Similarly to zTDG2, this single mutation yielded no enzyme activity on the entirety of substrate samples. These results imply that the N193 residue in Sfo TDG is essential in the role of catalysis specifically on the caC containing substrate (Figure 4).

After substitution of single amino acids in different sites (Figure 4), no significant increase in activity was observed in certain single mutants. So in the next experiment implemented, double mutants were constructed to investigate the correlated mutation in motif 1 of Sfo TDG. Compared with various substrates, both Sfo TDg and zTDG2 displayed increased activity on the caC substrate. Many essential residues in Sfo TDG are very highly conserved when comparing motif 1 to zTDG2. A significant difference is that the residues K157-R159 in zebrafish TDG correspond to I190-H192 in Sfo TDG (Figure 5). The various combinations of certain residues could account for differences in enzymatic characteristics. Individual substitutions in I190 and H192 erase all catalytic activity of the caC substrate in Sfo TDG. The doublet I190K-H192R mutation showed increased activity on only caC. The difference in these

sites in both zTDG2 and Sfo TDG enzymes imply that a correlation might exist between I190 and H192. The H192-H193 and H193-N196 doublet in motif 1 of Sfo TDG also are notable. Additionally, when histidine is changed to tryptophan along with asparagine changed to glycine, there is similar caC activity and increased activity on uracil. Interestingly, this double mutation shows increased activity on hmU. Along with the face that there was increased activity on uracil, we suspect a possible correlation in neighboring residues. When histidine is changed to tryptophan, this gives the enzyme broader specificity. Overall, compared to single mutants H193W and N196G, the double mutants displayed increased enzyme activity over various substrates than H193W and showed limited enzyme activity on hmU-containing substrates.

A reaction curve was created for Sfo TDG to compare the enzyme concentration vs. product (%) and time vs. product (%). The double mutant I190K-H192R displayed the highest product yield when plotted against both enzyme concentration and time (Figure 6).

To further understand this correlation, structural analysis was performed in AlphaFold (Figure 7). Pymol was utilized to obtain a complex with caC which gave 5 different predictions. The most reliable prediction, the one most similar to the homolog, was chosen. The distance between the residues and substrates in the double mutants were measured and compared. It was observed that when histidine is closer to the substrate, this position enhances the catalysis of the enzyme due to the strength of the hydrogen bonds.

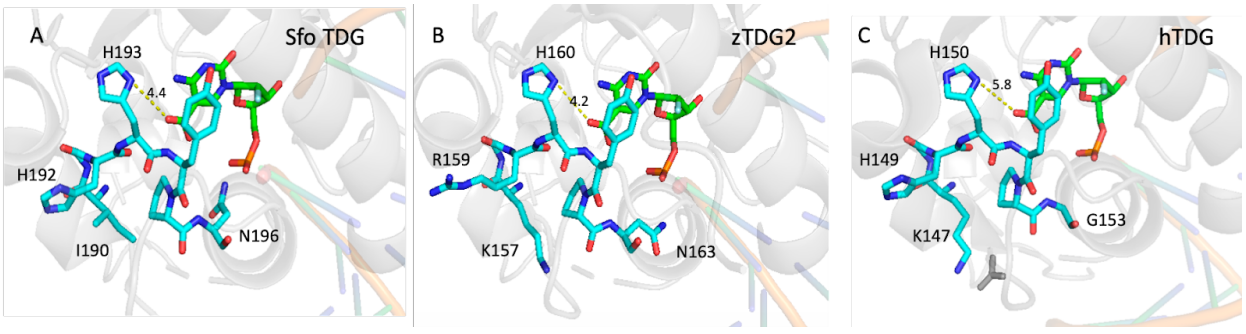

Figure 7. Close-up view of important residues in predicted Sfo TDG, zTDG2, and hTDG structures. caC is shown as licorice and colored by atom type. **A.** Predicted structure of Sfo TDG with caC complex. **B.** Predicted structure of zTDG2 with caC complex. **C.** Structure of hTDG with caC complex.

In conclusion, if one single amino acid was changed, no visible activity was observed. When we changed two amino acids together, the product increased activity on only caC. So, zebrafish TDG 2 has very robust activity on caC and human TDG shows broad specificity towards different substrates. But, Sfo TDG has narrow specificity as a DNA glycosylase. Additionally, double mutagenesis shows some residues are co-related. Structural analysis using Pymol reaffirmed roles of residues in determining substrate specificity. Overall, these 4 sites in motif 1 for mutational analysis underwent substitution mutations which changed some of the catalytic patterns of these enzymes and their glycosylase activity. In future research, more investigation is needed to fully understand specificity and catalysis in TDG enzymes.