BIOCHEMICAL ANALYSES OF UDGX-A CROSSLINKING URACIL-DNA GLYCOSYLASE

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biochemistry and Molecular Biology

---

by
Chuan Liang
December 2023

Accepted by:
Dr. Weiguo Cao, Committee Chair
Dr. Cheryl Ingram-Smith
Dr. Haiying Liang
Dr. Tzuen-Rong Tzeng

ABSTRACT

DNA base damage is common due to exposure to various endogenous and exogenous factors. To repair the base lesions, such as uracil from cytosine deamination, enzymes from the uracil-DNA glycosylase (UDG) superfamily are critical, which can recognize the damaged base and initiate the base excision repair (BER) pathway. There used to be six families of proteins identified in the UDG superfamily until a new member, UDGX, was found in *Mycobacterium smegmatis*, which is a unique DNA-crosslinking UDG. In this dissertation work, a series of biochemical analyses of the newly found UDGX are performed, including the analyses of structures, functions, catalytic mechanism, distribution in species and even potential evolutionary trends. Chapter 1 is a general introduction of common DNA base damage and enzymes from different families in the UDG superfamily. Chapter 2 presents the structural and functional analyses of UDGX by mutational and kinetics studies. It possesses multiple inter-motif and intra-motif interactions of residues in the structure of UDGX, which results in the coupling of uracil excision and DNA crosslinking functions of UDGX. Based on the understanding of the structural and functional coupling, a catalytic mechanism of UDGX on uracil-DNA substrates is also proposed. In Chapter 3, the DNA crosslinking site, the residue H109 of UDGX is investigated. By mutational and kinetics analyses, the function coupling of UDGX is further confirmed; additionally, a unique mutant, H109E was also found possessing DNA crosslinking capability via an ester bond with the abasic site instead. In Chapter 4, the DNA crosslinking activities of six putative UDGX homologs from different species were screened. It has been found that there are probably some other

interactions of residues outside motifs essential to the activities of UDGX. For example, we found a salt bridge between R158 and D200 that is critical when we studied two UDGX homologs which have extremely similar sequences to the Msm UDGX. In conclusion, with these biochemical analyses of UDGX, we have achieved a better understanding of this unique bifunctional DNA-crosslinking UDG, which expands the biological significance and application of enzymes in UDG superfamily.

DEDICATION

This dissertation is dedicated to my parents and my family.

ACKNOWLEDGEMENTS

First, I'd like to give my sincere appreciation to my advisor, Dr. Weiguo Cao. With his professional advice and support, I became a researcher successfully from a clinical surgeon and completed a series of studies to pursue my Ph. D degree. His extensive knowledge and experience, patience and encouragement, helped me a lot in my work and life, and I believe these will also have a good effect on my future career.

Then, I'd like to thank my committee members, Dr. Cheryl Ingram-Smith, Dr. Haiying Liang and Dr. Tzuen-Rong Tzeng. Their suggestions and support helped me to polish my research and become a better researcher with scientific thinking and accumulated practices.

Next, I'd also like to thank all the previous members in Cao's Lab (Dr. Ye Yang, Dr. Ping Ning and Dr. Chenyan Chang). Without their help, it would be much harder to complete all my research work. Their suggestions and support helped me to come up with some good ideas in my studies and helped me to overcome the challenges I met in the research.

Last, I'd like to say thank-you to my parents, my family and my friends. With the understanding and support from my parents and my family, I could make the decision to come to the U.S. to pursue a Ph. D degree and have a carefree life during the period. With the companionship of my friends, I could spend years contentedly here until the coming day of my graduation.

TABLE OF CONTENTS

Page

TITLE PAGE i

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

ABBREVIATIONS xi

CHAPTER

1. BASE EXCISION REPAIR OF DNA BASE DAMAGE AND ENZYMES FROM URACIL-DNA GLYCOSYLASE SUPERFAMILY 1

    I. Introduction 1
    II. DNA base damage and base excision repair 3
    III. Uracil-DNA glycosylase superfamily 10
    IV. Reference 24

2. STRUCTURAL AND FUNCTIONAL COUPLING IN CROSSLINKING URACIL-DNA GLYCOSYLASE UDGX 46

    I. Abstract 46
    II. Introduction 47
    III. Materials and Methods 48
    IV. Results 57
    V. Discussion 68
    VI. Reference 75

3. BIOCHEMICAL PROPERTIES OF THE CROSSLINKING SITE OF UDGX .. ……83

    I. Abstract 83

II. Introduction 84
III. Materials and Methods 85
IV. Results 91
V. Discussion 97
VI. Reference 99

4. URACIL-DNA CROSSLINKING ACTIVITY IN PUTATIVE UDGX HOMOLOGS 103

I. Abstract 103
II. Introduction 104
III. Materials and Methods 105
IV. Results and Discussion 110
V. Reference 118

5. RESEARCH SIGNIFICANCE AND CONCLUDING REMARKS 125

Reference 128

LIST OF TABLES

Table Page

Table 2.1 Primers for mutagenesis.....52

Table 2.2 Kinetic parameters of UDGX with G/U and A/U base pairs.....61

Table 3.1 Primers for mutagenesis.....87

Table 3.2 Kinetic parameters of UDGX mutants with G/U and A/U base pairs.....93

Table 4.1 Primers to amplify cDNA sequences of UDGX homologs by PCR.....107

Table 4.2 Primers for mutagenesis.....108

LIST OF FIGURES

Figure Page

Figure 1.1 Common DNA base damage 2

Figure 1.2 DNA base deamination 6

Figure 1.3 Representative scheme of the uracil DNA repair by BER 8

Figure 1.4 Phylogenetic tree of UDG superfamily 10

Figure 1.5 Sequence alignment and structures of proteins in UDG superfamily 12

Figure 1.6 Sequence alignment of family 3 SMUG1, SMUG1-like and SMUG2 proteins 16

Figure 1.7 Crosslinking between UDGX and AP site 21

Figure 2.1 Representative kinetics analysis of the wild-type and mutant UDGX 57

Figure 2.2 Overview of UDG superfamily in selected bacterial genomes 58

Figure 2.3 Catalytic scheme and biochemical analyses of UDGX 60

Figure 2.4 Sequence alignment of UDG enzymes and mutagenesis sites 62

Figure 2.5 DNA crosslinking analysis of UDGX-E52A and Q53A mutants 63

Figure 2.6 DNA crosslinking analysis of UDGX-F65A, N91A, K94A and H178A mutants 64

Figure 2.7 Tripartite interactions by D56, D59 and R107 65

Figure 2.8 DNA crosslinking and uracil excision analyses of UDGX-H109A mutant 66

Figure 2.9 Uracil excision and AP site crosslinking coupling of UDGX 67

Figure 2.10 Inter-motif and intra-motif interactions in UDGX structure 70

Figure 2.11 Proposed catalytic mechanism of uracil excision and AP site crosslinking by UDGX……..73

Figure 3.1 Crosslinking between UDGX H109E mutant and AP site.....85

Figure 3.2 Activity determination of UDGX mutants 91

Figure 3.3 Alkaline hydrolysis of H109E-DNA crosslinking product 94

Figure 3.4 Uracil excision and DNA crosslinking coupling of UDGX H109E mutant 94

Figure 3.5 Effects of pH on uracil excision and DNA crosslinking of the H109E mutant ……..96

Figure 3.6 Catalytic scheme of UDGX H109E mutant 98

Figure 4.1 Substrates and sequence alignment 111

Figure 4.2 DNA crosslinking analysis of Sma UDGX, Nji UDGX and Svi UDGX...113

Figure 4.3 DNA crosslinking analysis of Bgl UDGX wildtype and mutants.....114

Figure 4.4 Whole sequence alignment of MsmUDGX, Mgo UDGX1 and Mgo UDGX2 ……115

Figure 4.5 DNA crosslinking analysis of Mgo UDGX1 and Mgo UDGX2.....116

Figure 4.6 Role of R158 and D200 in Msm UDGX 117

ABBREVIATIONS

5-BrU, 5-bromouracil

5-caC, 5-carboxylcytosine

5'-dRP, 5'-deoxyribose-5-phosphate

5-fC, 5-formylcytosine

5-foU, 5-formyluracil

5-FU, 5-fluorouracil

5-hmU, 5-hydroxymethyluracil

5-mC, 5-methylcytosine

5-OHC, 5-hydroxycytosine

5-OHU, 5-hydroxyuracil

64PPs, Pyrimidine(6-4)pyrimidone photoproducts

8-OHG, 8-hydroxyguanine

8-oxo-G, 7,8-dihydro-8-oxoguanine

AGT/MGMT, O^6-alkylguanine-DNA alkyltransferase/O^6-methylguanine-DNA methyltransferase

AID, Activation-induced cytidine deaminase

AlkA, 3-methyladenine DNA glycosylase II

AlkB, AlkB (alkylation B)-related α-ketoglutarate-dependent dioxygenase

APOBEC, Apolipoprotein B editing complex

AP site, Apurinic/apyrimidinic site

BER, Base excision repair

CBP, CREB (cyclic adenosine monophosphate response element binding protein) binding protein

CPDs, Cyclobutane pyrimidine dimers

DKC1, Dyskerin pseudouridine synthase

dL, 2-deoxyribonolactone

εC, 3,N^4-ethenocytosine

ERα, Estrogen receptor α

FaPy-A, 4,6-diamino-5-formamidopyrimidine

FaPy-G, 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine

Fpg, Formamidopyrimidine-DNA glycosylase

HDG, Hypoxanthine-DNA glycosylase

HR, Homologous recombination

IDLs, Insertion-deletion loops

LP-BER, Long-patch base excision repair

MMR, Mismatch repair

MUG, Mismatch-specific uracil-DNA glycosylase

NCP, Nucleosome core particles

Nei, Endonuclease VIII

NEIL, Endonuclease VIII-like

NER, Nucleotide excision repair

NHEJ, Non-homologous end joining

Nth, Endonuclease III

OGG1, 8-oxoguanine glycosylase

PCNA, Proliferating cell nuclear antigen

Polβ, Polymerase β

RAR, Retinoic acid receptor

RAREs, Retinoic acid response elements

RXR, Retinoid X receptor

SAM, S-adenosyl-L-methionine

SMUG, Single strand selective monofunctional uracil-DNA glycosylase

TDG, Thymine-DNA glycosylase

TERC, Telomeric RNA component

TOP1, Topoisomerase 1

TOP-DPC, Topoisomerase DNA-protein crosslinks

TTF-1, Thyroid transcription factor-1

UDG, Uracil-DNA glycosylase

UNG, Uracil-*N*-glycosylase

CHAPTER ONE

BASE EXCISION REPAIR OF DNA BASE DAMAGE AND ENZYMES FROM URACIL-DNA GLYCOSYLASE SUPERFAMILY

I. Introduction

Deoxyribonucleic acid (DNA) is the major macromolecule carrying the essential genetic information in organisms. DNA is a double-helix polymer of two chains of deoxyribonucleotides, which are composed of phosphate groups, deoxyribose and four types of bases. The deoxyribose and the phosphate groups form the backbone of the DNA molecules by phosphodiester bonds, and the bases attached to the backbone by glycosidic bonds are the carriers of the genetic information. There are four types of bases in DNA, including adenine (A), cytosine (C), thymine (T) and guanine (G) (Figure 1.1 A), and the genetic information is carried in the arranging sequence of these four bases. Any damage or change to the DNA, even just one base, could cause severe consequences, such as cancer and neurodegenerative diseases [1-6]. The occurrence of DNA base damage is common due to various environmental exposures, e.g., UV [7], and endogenous oxyradicals [1, 3]. To repair the damage, multiple DNA repair pathways could be conducted, such as nucleotide excision repair (NER), mismatch repair (MMR), base excision repair (BER) and recombination repair, which includes non-homologous end joining (NHEJ) and homologous recombination (HR) [8]. As an important DNA repair pathway, BER can remove the damaged base directly and complete the repair via multiple steps with a series of enzymes involved [9]. In this chapter, an overview of DNA base damage and base

excision repair, especially base deamination and enzymes from uracil-DNA glycosylase (UDG) superfamily, which are involved in the BER, will be provided.

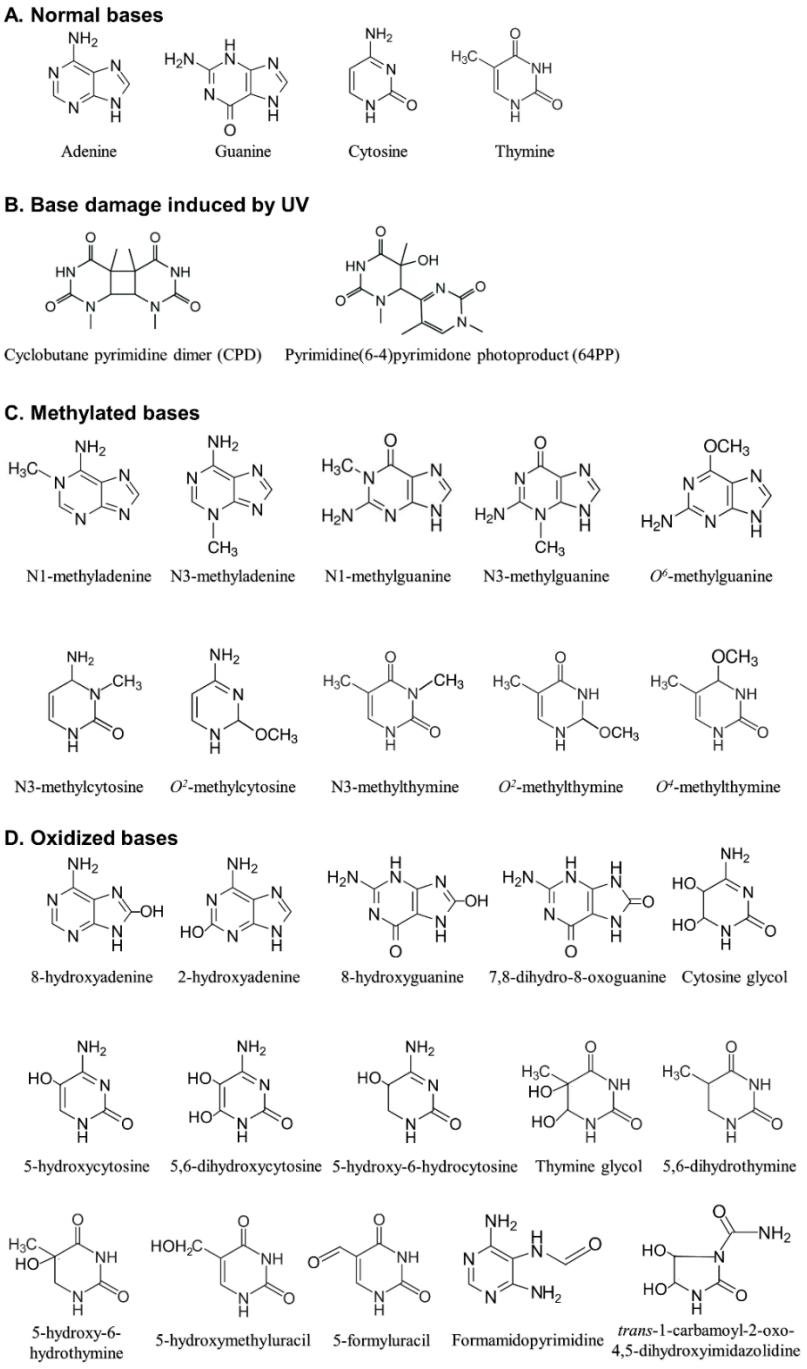

Figure 1.1 Common DNA base damage. A. Normal bases. **B.** UV radiation-induced base damage. **C.** Methylated base damage. **D.** Oxidative base damage.

II. DNA base damage and base excision repair

A. DNA base damage

The cause of DNA base damage can be exogenous and endogenous. Exogenous DNA damage causes include environment, physical factors and chemical agents. For example, ultraviolet (UV) light can produce specific photochemical DNA base damage, such as cyclobutane pyrimidine dimers (CPDs) and pyrimidine(6-4)pyrimidone photoproducts (64PPs) at dipyrimidine sites, where two adjacent pyrimidine bases are juxtaposed in tandem in the DNA strand [7] (Figure 1.1 B). Ionizing radiation could directly or indirectly damage DNA bases, as a result, abnormal bases, such as 8-oxo-guanaine, thymine glycol and formamidopyrimidines could be formed in DNA [10]. Chemical agents, for instance, alkylating agents can react with nucleophilic base ring nitrogens, especially the N7 of guanine and N3 of adenine, forming alkylated bases, such as N7-methylguanine and N3-methyladenine [10]. In addition, aromatic amines, polycyclic aromatic hydrocarbon and other reactive electrophiles can also cause DNA base damage [10].

Endogenous DNA base damage can come from replication errors, base deamination, DNA methylation and oxidative DNA damage [10]. Occasionally, DNA polymerase could incorrectly incorporate uridine in the DNA during replication [11, 12]. Though DNA alkylation is a natural chemical modification of DNA and plays an important role in epigenetics [13], it can also cause mutation and other harmful consequences [14, 15]. For example, O^6-methylguanine is well known for its ability to induce G:C to A:T mutations because it can mispair with thymine [16]. In addition, O^4-methylthymine is also a cause of mutations in human cells [17]. Besides mutagenesis, lesions like N3-methyladenine could

even block DNA polymerase during replication, which may finally result in cell death [18]. Some common methylated base lesions are shown in Figure 1.1 C.

Although oxidative DNA damage could be concomitant with exogenous damage, e.g. ionizing radiation, spontaneous DNA damage caused by normal cellular metabolism also widely exists. The endogenous oxidative DNA damage is typically induced by reactive oxygen species (ROS) [1, 10, 19]. The most common ROS species are the superoxide radicals ($\bullet O_2^-$), hydrogen peroxide (H_2O_2), the hydroxyl radical (•OH), alkoxyl radicals (•OR) and alkylperoxyl radicals (•OOR) [10, 19]. There are more than 100 oxidative base lesions identified in mammalian genomes [20], some representative oxidized bases are shown in Figure 1.1 D, including 7,8-dihydro-8-oxoguanine (8-oxo-G), 8-hydroxyguanine (8-OHG), 5-formyluracil (5-foU) and so on [1]. Because guanine is the most sensitive to ROS among the four types of bases, oxidized guanine is the most abundant oxidative base damage, e.g. 8-oxo-G, which is an oxidized product by alkylperoxyl radicals [19]. Thus, oxidized guanine lesion, such as 8-oxo-G, has been widely used as a biomarker of oxidative DNA damage [21, 22]. The damaged base 8-oxo-G could cause G:C to A:T and G:C to C:G mutations because of the A:8-oxo-G and G:8-oxo-G mispairs [20, 23]. Furthermore, accumulations of these mutations have been proven to be related to tumorigenesis [2-4]. Not only cancer, oxidative base damage, such as 8-oxo-G, has also been found related to neurodegenerative diseases and aging [5, 6, 24, 25]. Another mutagenic oxidative base damage is 5-foU, which is the oxidized product of thymine. The 5-foU base in DNA could induce A:T to T:A transversions, T:A to C:G transitions and -1 deletions [26].

B. DNA base deamination

Besides replication errors, DNA methylation and oxidation, DNA base deamination is also one of the major causes of endogenous DNA damage. Base deamination refers to the loss of exocyclic amino groups of cytosine (C), adenine (A), guanine (G) and 5-methylcytosine (5-mC), becoming uracil (U), hypoxanthine, xanthine and thymine (T), respectively [10] (Figure 1.2). As a spontaneous modification of DNA, base deamination universally occurs in genomes. For example, the deamination of cytosine to uracil is estimated around 100-500 residues per mammalian genome per day [27-30]. Because the deamination of cytosine in single-stranded DNA is 200-300 times faster than in double-stranded DNA [27-31], it can occur in the single-stranded regions of replication fork or in the single-stranded DNA of transcription bubbles, causing G:C to A:T transition mutations [10, 31, 32]. The deamination of the methylated product of cytosine, 5-mC, is even more frequent than the deamination of cytosine, resulting in G:T mispairs [10]. As a consequence, the G:C to A:T transitions caused by the deamination of C or 5-mC at the CpG sequences account for one-third of the single site mutations responsible for human hereditary diseases [10, 33]. Despite the lower frequency than cytosine, the deamination of adenine could also induce A:T to G:C transition mutations because the resulted hypoxanthine can pair with cytosine [34]. Although the consequence of xanthine in DNA, the deaminated product of guanine, is still not completely revealed, there are some proofs indicating the mutagenic potential of it [35].

Though the deamination of cytosine is a source of DNA damage and subsequent consequences, it is also a normal chemical modification of DNA, especially for

immunological purposes [36, 37]. This deamination is induced by activation-induced cytidine deaminase (AID) and apolipoprotein B editing complex (APOBEC), which aims to broaden the antibody diversification via somatic hypermutation and class-switch recombination to help the host to defend against various pathogens [36-39]. For example, AID introduces uracil into the immunoglobulin gene by deaminating cytosine [40]. Then, the uracil will be repaired, which however is an error-prone repair pathway, and hypermutation of antibody will be induced as a result [36]. Another case is that AID targets cytosine residues on opposite strands in the switch regions upstream of various immunoglobulin heavy chain loci. The following recombination occurs after the double-stranded DNA breaks caused by clustered deamination to achieve isotype switching [36].

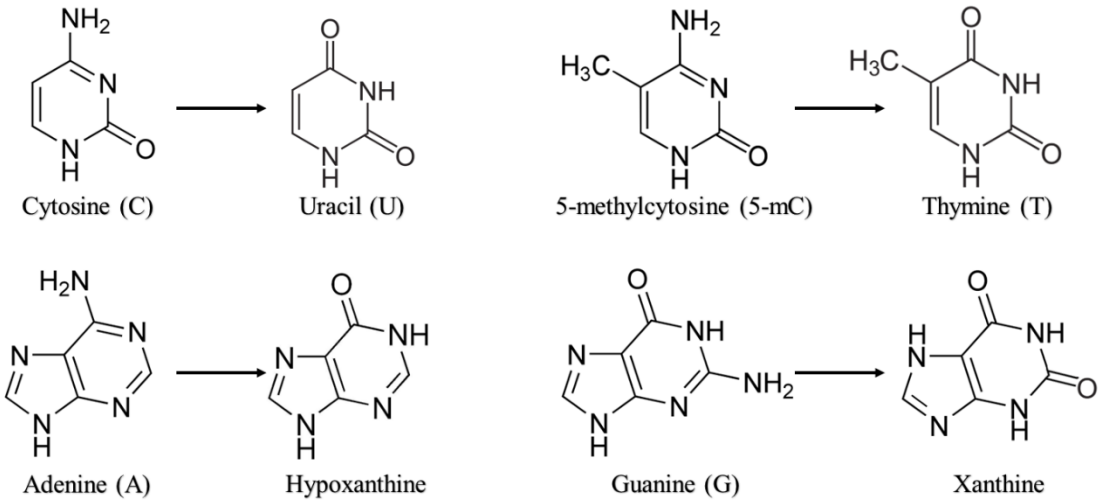

Figure 1.2 DNA base deamination.

C. Base excision repair

Despite of high frequency of DNA base damage, cells still survive and maintain normal function and metabolism, which is a result of a series of DNA repair pathways, including direct reversal of damaged DNA base, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and so on [10]. There are two classes of enzymes that can reverse damaged bases directly [10]. One is O^6-alkylguanine-DNA alkyltransferase (AGT/MGMT), which reverses O-alkylated DNA base lesions, such as O^6-methyl, ethyl, 2-chloroethyl, benzyl and aliphatic groups, the pyridyloxobutyl adducts of guanine, and even the O^6–G-alkyl-O^6- G interstrand crosslinks [8, 10, 41, 42, 43]; the other is AlkB-related α-ketoglutarate-dependent dioxygenase (AlkB), which reverses N-alkylated base lesions [10, 44]. NER is a pathway to repair bulky lesions, which can distort the DNA helix structure, such as CPDs and 64PPs induced by UV radiation [8, 10, 45]. Whereas MMR is a conserved DNA repair pathway to remove base mismatches during replication and the insertion-deletion loops (IDLs) produced by strand slippage events[8, 10, 46].

As another essential DNA repair pathway, BER can be conducted to repair oxidative, deamination and alkylation single base damage as well as abasic damage without significant distortion of DNA helix structure, e.g., uracil residue in DNA strands [9, 10, 47]. In general, BER is initiated with excision of the damaged base by DNA glycosylases, which is a group of enzymes catalyzing hydrolysis of the *N*-glycosidic bond between the damaged base and deoxyribose. After the removal of the damaged base, an apurinic/apyrimidinic site (AP site) is formed. With AP endonucleases and

phosphodiesterases, the abasic deoxyribose phosphate is then removed, leaving a single-nucleotide gap in the DNA strand. Finally, DNA polymerases add a correct nucleotide to the 3'-terminus of the DNA strand in the gap and DNA ligases ligate the newly added nucleotide to the 5'-terminus of the strand on the other end of the gap to complete the repair (Figure 1.3).

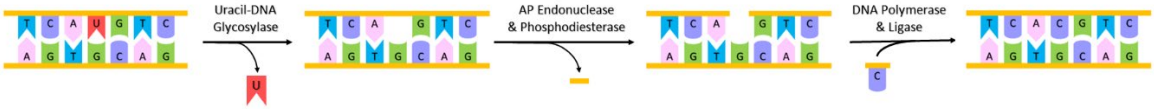

Figure 1.3 Representative scheme of the uracil DNA repair by BER.

Although, the general repair process is described as above, there are still some different mechanisms of BER depending on the types of enzymes involved, especially the DNA glycosylases to initiate the repair [9]. For instance, monofunctional DNA glycosylases, such as uracil-DNA glycosylase UNG, break *N*-glycosidic bond only, forming an AP site. Subsequent AP endonucleases hydrolyze the 5' phosphodiester bond at the AP site immediately, producing a 3'-terminus which can be recognized by DNA polymerase, and then a new nucleotide can be added to this position. However, the hanging 5' abasic deoxyribose phosphate (5'-deoxyribose-5-phosphate, 5'-dRP) is still linked to the DNA strand, so it requires the dRP lyases to excise it from the DNA strand via β-elimination and produce a 5'-terminus. After the 5'-dRP is removed, a new phosphodiester bond is formed by ligases between the hydroxyl group at the 3'-terminus of the new nucleotide and the phosphate group at the 5'-terminus produced by dRP lyases. Differently, if the base excision is induced by a bifunctional DNA glycosylase,

which can not only hydrolyze *N*-glycosidic bond but also perform β-elimination at 3' of the damaged nucleotide, e.g., 8-oxoguanine glycosylase (OGG1), dRP lyases are no more needed. Some bifunctional DNA glycosylases, such as endonuclease VIII-like 1 (NEIL1), can even conduct two consecutive elimination steps (β,δ-elimination) besides base excision, producing a single-nucleotide gap directly but with a 3'-terminus phosphate group. After one more step of dephosphorylation, the gap can be filled by DNA polymerases and ligases at last [8, 9]. Besides the classic single-nucleotide BER pathway, there is, in fact, an alternative pathway for BER, long-patch BER (LP-BER), which is a strand displacement repair under specific circumstances, such as adenosine triphosphate (ATP) concentrations are low or during the S phase of the cell cycle [48].

There are three main superfamilies of DNA glycosylases to initiate BER pathway, uracil-DNA glycosylase superfamily, Fpg/Nei superfamily and Nth superfamily [9]. Fpg/Nei superfamily consists of two main members, formamidopyrimidine-DNA glycosylase (Fpg) and endonuclease VIII (Nei) [49-51]. Fpg proteins characteristically recognize and excise oxidized purines from DNA, such as 8-oxo-G, 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine (FaPy-G) and 4,6-diamino-5-formamidopyrimidine (FaPy-A) [49]; whereas Nei proteins and a series of endonuclease VIII-like (NEIL) proteins recognize and excise oxidized pyrimidines, such as 5-hydroxycytosine (5-OHC) and 5-hydroxyuracil (5-OHU) [50]. Nth superfamily is a diverse group of endonuclease III (Nth) related DNA glycosylases, which catalyze the excision of oxidized pyrimidines from DNA, such as thymine glycol and 5-hydroxy-6-hydrothymine [52].

III. Uracil-DNA glycosylase superfamily

As one of the main DNA glycosylase superfamilies, enzymes from uracil-DNA glycosylase (UDG) superfamily play an important role in BER. There have been at least six families of UDG proteins identified for now (Figure 1.4). In addition, a unique member, UDGX, has been added to this superfamily, which broadens the structural and functional diversity of proteins in the superfamily.

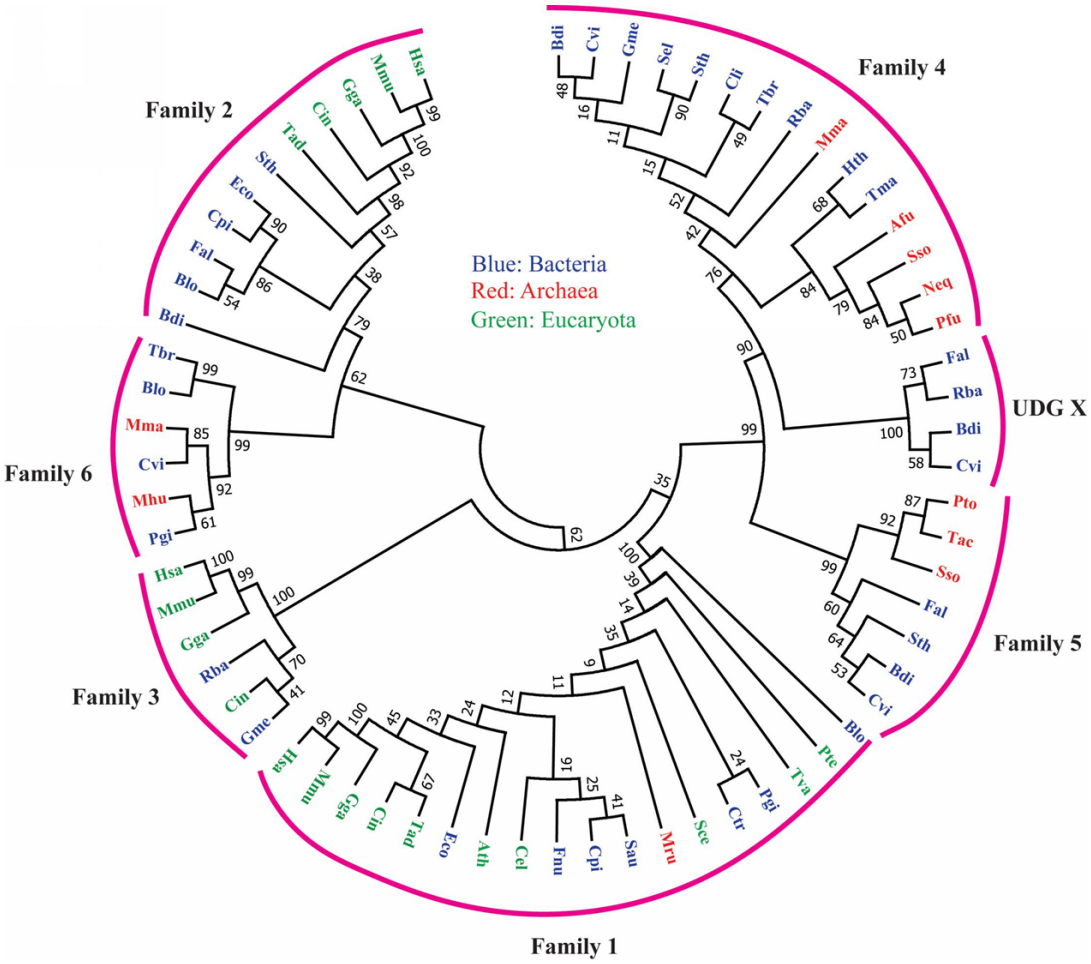

Figure 1.4 Phylogenetic tree of UDG superfamily. Proteins from family 1-6 and UDGX are analyzed and shown in branches. The genomic sources of proteins are distinguished by colors (blue: bacterial proteins; red: archaeal proteins; green: eukaryotic proteins).

A. Family 1 UNG

Family 1 uracil-*N*-glycosylase (UNG) from *Escherichia coli* was first discovered and characterized in 1977 [53]. Since then, numerous studies have been conducted to clarify the properties of this conventional uracil-DNA glycosylase. UNG proteins and conserved UNG-like sequences have been found in almost all cellular organisms, even in pox- and herpesviruses [54]. In humans, based on the locations where the enzyme functions, it can be further categorized to UNG1 and UNG2 (in mitochondria and nuclei, respectively). However, these two proteins share the same biological function and genomic source, the *ung* gene, the only difference is their N-terminal regions for cellular targeting, which is a result of alternative splicing of the *ung* gene [55, 56]. Such a wide distribution among species implies its critical role in maintaining the DNA integrity of organisms. UNG enzymes specifically excise uracil from both double-stranded and single-stranded DNA, but no activity is detected on normal DNA bases or uracil in RNA [53, 57, 58, 59]. Besides uracil, UNG enzymes have been reported to catalyze the excision of a uracil derivative, 5-fluorouracil (5-FU), which is a commonly used anti-tumor agent [60], and the excision of some products of cytosine oxidation, such as 5-hydroxyuracil (5-OHU), isodialuric acid, and alloxan [61, 62].

The general structural fold of UNG proteins is four β-sheets sandwiched between two pairs of α-helices [63]. On the surface of the proteins, there is a shallow and narrow positively charged groove, where the uracil-containing DNA binds, and in the groove, a deep pocket which contains the enzyme's active sites, accommodates and reacts with the

uracil residue in the DNA strand [64]. The active sites comprise three catalytic motifs, which are highly conserved among different families of UDG enzymes (Figure 1.5).

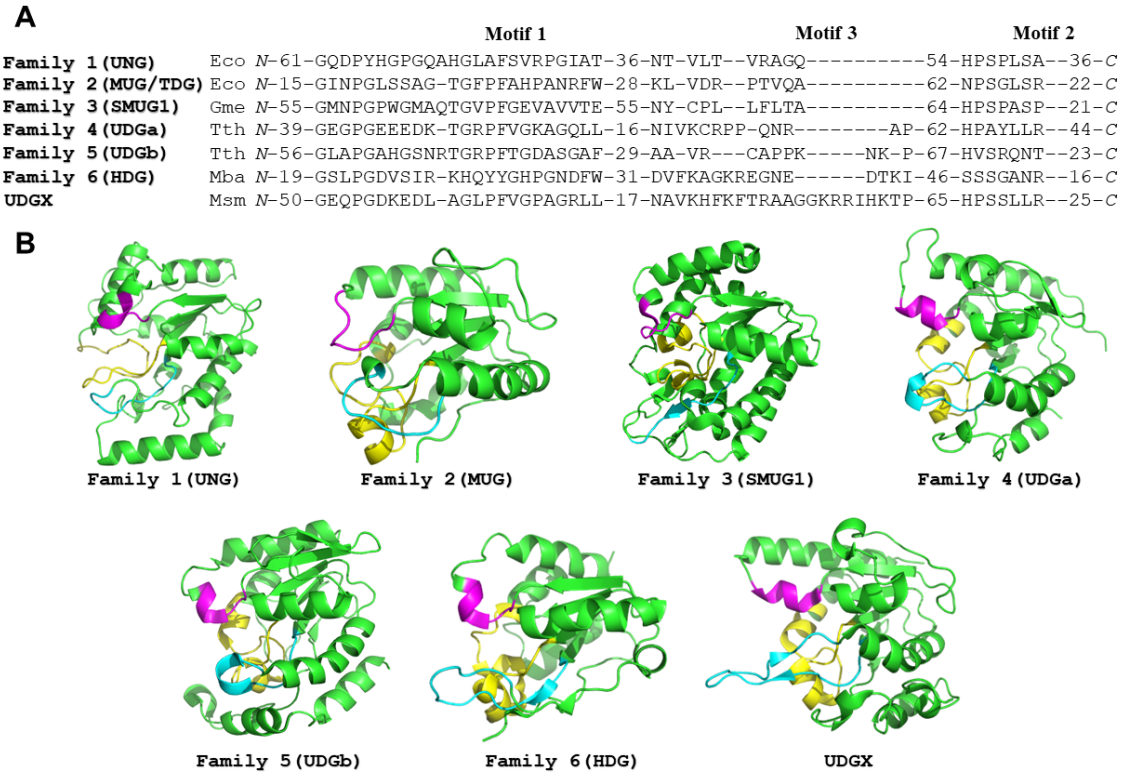

Figure 1.5 Sequence alignment and structures of proteins in UDG superfamily. A. Sequence alignment of proteins from 6 families and UDGX. Family 1 (UNG): Eco, *Escherichia coli* O157:H7 str. EDL933, NP_289138.1; Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1.1; Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069.1; Family 4 (UDGa): Tth, *Thermus thermophilus*, WP_011172816.1; Family 5 (UDGb): Tth, *Thermus thermophilus*, WP_011173217.1; Family 6 (HDG): Mba, *Methanosarcina barkeri*, WP_011305765.1; UDGX: Msm, *Mycobacterium smegmatis*, WP_011726794.1 (GenBank accession numbers are shown after the specie names). **B.** Structures of proteins from 6 families and UDGX. Family 1 (UNG): PDB code 1SSP; Family 2 (MUG): PDB code 1MUG; Family 3 (SMUG1): PDB code 1OE4; Family 4 (UDGa): PDB code 1UI0; Family 5 (UDGb): PDB code 2DDG; Family 6 (HDG): PDB code 2L3F; UDGX: PDB code 6IO9. Catalytic motifs are highlighted by colors (motif 1: yellow; motif 2: magenta; motif 3: cyan).

Because of the specific uracil excision activity of UNG, it has been widely used as a tool to conduct ligase-free and cDNA cloning in polymerase chain reaction (PCR) [65],

[66]. When a U-containing sequence is added to the 5' end of PCR primers, all the amplified DNA products will also have U residues placed in their 5' ends. After selective excision of these U residues by UNG, the 5' end of products will be unpaired because of the existence of multiple AP sites, which could be annealed to a plasmid vector later without DNA ligase [65]. In addition, the substitution of U residues for all the T residues in PCR primers, combined with UNG digestion, can eliminate "primer-dimers" and increase the yield of product DNA [65]; what's more, the UNG digested PCR products can be sequenced directly with other primers without cleaning up [67]. Some heat-labile UNG proteins have been found and used in PCR to eliminate contamination in products [68], [69]. Because they can be inactivated during the heating cycles, no more steps to remove or inactivate the UNG proteins are needed. As the deaminated product of cytosine, uracil in DNA could induce mutations. Since UNG could recognize and excise uracil specifically, it has also been developed as a biomarker to detect mutations in DNA [70], [71].

B. Family 2 MUG/TDG

Family 2 UDG includes a group of enzymes named mismatch-specific uracil-DNA glycosylase (MUG) or thymine-DNA glycosylase (TDG). Human TDG (hTDG) was the first mismatch-specific DNA glycosylase found to repair G:T mismatches in DNA, which is a result of deamination of 5-mC [72-75]. MUG was first identified as a homolog of TDG in *Escherichia coli* [75]. With the identification of TDG and MUG homologs from various species successively, the distribution of family 2 MUG/TDG proteins has been broadened. Upon the discovery of hTDG, not only G:T mismatches, G:U mismatches are

also found as a substrate of it [76]. Through the efforts of researchers over the last decades, MUG/TDG has been reported to have efficient excision activities on broad substrates, especially deamination and/or oxidation products of cytosine in double-stranded DNA, including modification or substitution derivatives at the 5-carbon position of U, such as 5-OHU, 5-hydroxymethyluracil (5-hmU), 5-FU and 5-bromouracil (5-BrU); ethenoadducts, such as 1, N^2-ethenoguanine and 3, N^4-ethenocytosine; deaminated purines, such as hypoxanthine and xanthine; oxidized thymine, e.g. thymine glycol; and even normal DNA bases, such as 5-mC [77-83]. Interestingly, despite of the broad substrate spectra, hTDG only excises bases in double-stranded DNA in a mismatch with G, whereas TDG proteins from lower eukaryotes, e.g. fission yeast, and most of bacterial MUG proteins can excise bases from a mismatch with A or even in single-stranded DNA [77, 78, 84, 85]. Thus, compared to the main function of avoiding mutations induced by G:T mismatches of mammalian TDG, MUG/TDG proteins in lower organisms seem to serve as a protection against DNA base damage by deamination (e.g. hypoxanthine), oxidation (e.g. 5-OHU) and modification by products of lipid peroxidation (e.g. 3, N^4-ethenocytosine) in addition to their anti-mutagenesis function [77].

The structures of proteins from family 2 MUG/TDG are similar to the structures of family 1 UNG proteins [54, 75]. However, unlike UNG, which is an efficient DNA repair enzyme, family 2 MUG/TDG proteins seem to also have other biological functions. For example, since the interactions between TDG and transcription factors were uncovered, TDG has been posited to be related to the regulation of gene expression. TDG was reported to be able to interact with the retinoic acid receptor (RAR) and the retinoid X

receptor (RXR) [86], which form dimeric complexes binding retinoic acid response elements (RAREs) to regulate gene expression [87]. The interaction can enhance the binding of receptor complexes to RARE elements [86]. Another case is the interaction between TDG and estrogen receptor α (ERα), which is a nuclear receptor regulating gene transcription to response to estrogen [88]. By this interaction, TDG becomes a coactivator for ERα [88]. Other interactions have been discovered successively, such as interacting with CREB binding protein (CBP) and its paralog p300 to stimulate transcription [89], and interacting with thyroid transcription factor-1 (TTF-1) to repress transcription [90]. Based on the excision activity of TDG on 5-mC [83, 91] as well as 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) [92], family 2 MUG/TDG proteins have been considered as part of DNA demethylation pathways.

C. Family 3 SMUG1/SMUG2

Family 3 UDG proteins are a group of monofunctional DNA glycosylases that were initially recognized as enzymes excising uracil from single-stranded DNA, so they were given the name, single-strand selective monofunctional UDG (SMUG). This family contains three subfamilies, SMUG1, SMUG2 and SMUG1-like. SMUG1 was first discovered as an analogue of UDG in *Xenopus laevis* [93]. SMUG2 was identified later from *Pedobacter heparinus*, which has a close relationship to SMUG1 but belongs to a distinct branch of the family by phylogenetic analyses [94]. SMUG1 has been reported to have excision activities on broad substrates. It can excise uracil in A:U and G:U mismatches as well as single-stranded DNA, but no activity on G:T mismatches [95]. In addition, 5-hmU, 3, N^4-ethenocytosine (εC) and 5-FU are also substrates for SMUG1,

with a preference of U > 5-hmU >> εC > 5-FU [96-98]. By phylogenetic study, a series of SMUG1-like proteins have been found, which have similar catalytic motifs with SMUG1 enzymes in sequences but a distinct doublet SS in motif 1, where the residues are usually MN in SMUG1 [99] (Figure 1.6). Surprisingly, a representative SMUG1-like protein from *Listeria innocua* didn't show any activities on 5-hmU, xanthine and other substrates except for uracil [99]. With the discovery of SMUG2 proteins, substrate spectra and biological function of family 3 UDG have been further broadened. SMUG2 enzymes have glycosylase activities on deaminated base lesions, such as uracil, hypoxanthine and xanthine, as well as 5-fC and 5-caC [94, 100].

Figure 1.6 Sequence alignment of family 3 SMUG1, SMUG1-like and SMUG2 proteins. Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069.1; Rba, *Rhodopirellula baltica* SH 1, NP_869403; Hsa, *Homo sapiens*, NP_055126; Xla, *X. laevis*, AAD17300. Family 3 (SMUG1 Like): Lin, *Listeria innocua*, WP_010991469.1; Liv, *Listeria ivanovii*, WP_025279932.1; Lpa, *Lactobacillus paracasei*, WP_016381167.1; Ssp, *Streptomyces sp.* NRRL WC-3626, WP_030213602.1. Family 3 (SMUG2): Phe, *Pedobacter heparinus* DSM 2366, WP_012780920.1; Cth, *Chloroherpeton thalassium*, WP_012500224.1; Csp, *Chlorobium sp.* GBChlB, KER10671.1; Ace, *Arenibacter certesii*, WP_026813179.1 (GenBank accession numbers are shown after the specie names).

The structures of family 3 SMUG1/SMUG2 proteins are generally close to family 1 UNG and family 2 MUG/TDG proteins, but some differences exist, unambiguously [101]. The different structure and substrate specificity of SMUG1/SMUG2 from UNG

and MUG/TDG suggest its distinct biological functions. As a DNA repair enzyme, the uracil excision activity of SMUG1 on uracil-DNA packaged into nucleosome core particles (NCP) is severely inhibited, whereas UNG and TDG still retain significant activities, suggesting that repairing damaged bases in highly-packaged DNA strands in chromosomes is probably not the main function of SMUG1 or at the least some cofactors are required to assist the DNA repair in chromosomes by SMUG1 [102-104]. Several studies have found SMUG1 presumably has protective roles in cells against mutagenic and lethal conditions, such as ionizing radiation and therapeutic in chemotherapy of cancer [105-107].

In addition to the DNA repair function in BER, SMUG1 has also been found to play important roles in RNA maturation and RNA quality control. For example, SMUG1 can interact with the pseudouridine synthase Dyskerin (DKC1) and excise 5-hmU from single-stranded RNA but not pseudouridine, which is the nucleoside resulting from isomerization of uridine by DKC1 [108]. This process is essential to the maturation of ribosomal RNA (rRNA) and the regulation of 5-hmU levels in RNA [108]. Another case is that SMUG1 interacts with the telomeric RNA component (hTERC), a non-coding RNA subunit in telomerase holoenzyme, via regulating base modifications in a region between the CR4/CR5 domain and the H box of hTERC [109]. When SMUG1 is knocked out, base modifications accumulate in hTERC, resulting in reduced binding of DKC1, another subunit in telomerase, which finally leads to the degradation of hTERC [109]. Thus, SMUG1 is considered essential to the hTERC stability in telomerase biogenesis [109].

D. Family 4 UDGa

Family 4 UDGa is a group of UDG enzymes found in thermophilic bacteria and archaea. Although UDG activities have been detected in some hyperthermophilic micro-organisms [110], the enzymes responsible for the activities were finally identified until the first family 4 UDGa was discovered from *Thermotoga maritima* [111]. The first crystal structure of family 4 UDGa from *Thermus thermophilus* (Tth UDGa) was then reported [112]. Despite the low sequence identity with family 1 UNG and family 2 MUG/TDG proteins, UDGa still has similar topology and order of secondary structures to them [112]. In addition, a 4Fe-4S cluster is found in the structure of UDGa, which is inferred to be necessary to the structural stability instead of the catalytic activity [112], [113], not like other iron-sulfur cofactor possessing enzymes, such as radical S -adenosyl-L-methionine (SAM) enzymes, which require the participation of the 4Fe-4S cluster in the catalytic reactions [114]. Another significant feature of UDGa is its thermostability, which is probably a result of a series of salt bridges and ion pairs on the molecular surface and the presence of proline on loops and turns in the structure [112].

Uracil in double-stranded and single-stranded DNA is the only substrate found for UDGa [115, 116]. The excision of uracil by Tth UDGa is conducted via the residues His155 of motif 2, which interacts with O2 of uracil, and Asn89 of motif 3, which coordinates with a water molecule to attack C1’ carbon of deoxyribose [115]. The biological function of UDGa seems to be excising uracil in DNA to protect against uracil-induced mutations in thermophilic micro-organisms, since deamination of cytosine to produce uracil is more frequent at high temperatures [117]. Lack of UDGa could lead to

the increase of G:C to A:T mutations in cells [117]. Another interesting finding about UDGa is that the C-terminus of the protein can interact with the proliferating cell nuclear antigen (PCNA) [118], which is a sliding clamp tethering a series of DNA replication and repair factors to DNA in the replication fork [119]. A similar interaction also exists in eukaryotes, which is between UNG2 and PCNA to perform post-replicative base excision repair in the replication fork [120]. As an analog of UNG2, UDGa in thermophilic archaea perhaps is responsible for the PCNA-dependent post-replicative removal of misincorporated uracil in DNA, especially at high temperatures based on the thermostability of UDGa-PCNA complex [118, 121].

E. Family 5 UDGb

Family 5 UDGb is another group of UDG enzymes found from thermophilic archaea and bacteria. The first identified UDGb was reported in 2002 from the thermophilic eubacterium *Thermus thermophilus* (Tth UDGb) [122]; one month later, another UDGb was identified from the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* (Pae UDGb) [123]. Compared with UDGa, which only has the activity on uracil, UDGb exhibits a much broader substrate specificity, besides uracil, it can also excise other base lesions in DNA, such as 5-hmU, 5-FU, εC, hypoxanthine and xanthine [123, 124]. With the determination of the crystal structure complexed with AP site containing DNA of Tth UDGb, it suggests both steric force and water activation contribute to the base excision activity of UDGb [125]. Like UDGa, the first histidine of motif 2 of Tth UDGb, His190, plays an important role in the excision of both pyrimidine and purine base lesions by hydrogen bonding [115, 124]. The Asp75 of motif 1interacts with the damaged bases via

an activated water molecule, and the Asn120 of motif 3 activates another water molecule to attack the glycosidic bond at C1' carbon of deoxyribose, which plays a similar role of Asn89 of Tth UDGa [115, 124]. Based on the phylogenetic distribution among species and substrate spectra of UDGb, it seems to be a homological and functional continuation and extension of UDGa.

F. Family 6 HDG

Family 6 HDG is described as a group of hypoxanthine-DNA glycosylases (HDG). Not like UDGs from other families, enzymes from family 6 HDG predominantly have a preference for hypoxanthine residues in DNA [126]. For now, it has been illustrated that family 2 MUG/TDG, family 3 SMUG2, family 5 UDGb, as well as family 6 HDG in UDG superfamily all show their activities on hypoxanthine, but the high selectivity of HDG may indicate it's an evolutionary trend to be a selective hypoxanthine-DNA glycosylase.

G. UDGX

Recently, a novel member of UDG superfamily has been introduced. Given the specific DNA crosslinking activity of this type of UDG, it's named UDGX, as a new family in the superfamily [127]. According to the sequence alignment of all types of UDG enzymes (Figure 1.5), though the sequence of UDGX shows much similarity to family 4 UDGa [127], there is a significantly extended loop in the motif 3 of UDGX, separating this type of enzymes from the other UDG families. Structurally, UDGX also possesses similar structure to family 4 UDGa, including the 4Fe-4S cluster in the structure, except for the signature loop in motif 3 [128, 129]. Functionally, UDGX

proteins can covalently crosslink to the AP site formed after the excision of uracil in DNA strand, and this crosslinking is performed by His109 of motif 3 [127-129] (Figure 1.7). Although uracil is the only reported substrate for UDGX [127-129], this bifunctional UDG still has some unique features, and probably play some distinct roles physiologically.

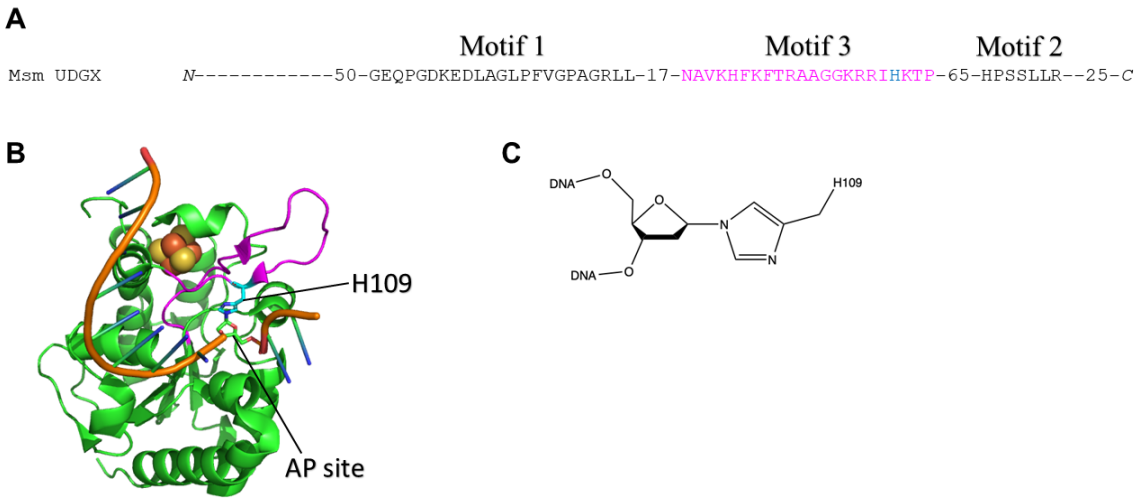

Figure 1.7 Crosslinking between UDGX and AP site. A. Sequence of catalytic motifs of UDGX from *Mycobacterium smegmatis* (Genebank accession number: WP_011726794.1). Motif 3 is highlighted in magenta, His109 is highlighted in blue. **B.** Structure of crosslinked complex of UDGX and DNA (PDB code: 6IOD). Motif 3 is highlighted in magenta. **C.** Chemical structure of C-N bond formed by His109 and AP site.

The first feature of UDGX is its capability to crosslink DNA covalently. In fact, there have been numbers of DNA binding proteins reported, including some enzymes involved in DNA replication and repair [130]. For example, topoisomerase 1 (TOP1) enzymes, which are responsible for torsional stress relief during DNA transcription and replication [131], can covalently bind to the 3' end of the single strand break inserted to allow DNA strand rotation via a tyrosine at the active site until the single strand break is religated and

the enzymes are released from DNA quickly [132]. Some known DNA glycosylases also exhibit the DNA crosslinking ability, such as mismatch-specific adenine glycosylase MutY, which excises adenine from 8-oxo-G:A mismatches to avoid mutations [52]. MutY enzymes can covalently crosslink to the C1' carbon of deoxyribose via Asp144 immediately after the excision of adenine, albeit the covalent bond is broken soon [133]. Although most of the covalent binding between proteins and DNA is transient, there are still some examples of stable protein-DNA crosslinks. For instance, the TOP2-like enzyme SPO11, which induces double strand breaks during meiosis [134], remains covalently crosslinked to the 5' ends of the break steadily until the break gets repaired [135].

The examples above are enzymes that bind to DNA in purpose to conduct their biological functions. Whereas some crosslinking between enzymes and DNA is an unexpected result, which could even be cytotoxic. For example, 2-deoxyribonolactone (dL), which is an oxidative DNA lesion produced by hydroxyl radical attack on the C1' carbon of a nucleotide [136], can trap DNA polymerase β (Polβ) on the DNA strand [137]. Polβ attacks the C1' lactone by its Lys72 residue and forms an amide bond, trapping the protein on the DNA strand [137]. DNA lesions like dL as well as other oxidative lesions can also covalently trap DNA glycosylases [136]. For instance, bacterial Fpg, Nei, 3-methyladenine DNA glycosylase II (AlkA) proteins, and eukaryotic OGG1 can be trapped by oxanine, an oxidized product of guanine by nitric oxide [138]. Other cases, such as Fpg, Nei and mammalian NEIL1 covalently trapped by 5-hydroxy-5-methylhydantoin have also been reported [139].

Such unexpected DNA-protein crosslinks not only disrupt the process of replication and transcription, but also cause the enzymes to be “inactivated”, which means they cannot react with other substrates anymore. This is also the second feature of UDGX, suicidal self-inactivation. It has also been found that the repair of those unexpected DNA-protein crosslinks relies on protease to trim trapped proteins or recombination repair [140, 141]. However, not like the enzymes trapped on the DNA by accident, the crosslinking of UDGX is along with its uracil-DNA glycosylase activity, which suggests that UDGX crosslinks to AP site in purpose, but the way to release the proteins remains unknown.

Although this kind of suicidal catalytic reaction has never been found in the other families of UDG, there exist such suicidal enzymes in nature. For example, the O^6-methylguanine-DNA methyltransferase (MGMT), the function of which is to convert O^6-methylguanine to normal guanine by transferring the methyl group directly from guanine to a cysteine residue of the enzyme, loses its activity with the formation of the S-methylcysteine in the active site [142]. Despite of the uracil excision activity of UDGX, the resulting AP site is occupied by UDGX, disrupting the BER, which leaves the biological function of UDGX still mysterious.

As a new member in the UDG superfamily, albeit it’s not well understood, UDGX has already been developed as a tool to detect uracil specifically and sensitively based on its features. For example, tagged UDGX can be used as a probe to detect uracil in DNA and visualized by fluorescent tags directly or fluorescence labelled antibodies [143],

[144]. In addition, UDGX has also been used in uracil sequencing, which improves the resolution compared with the conventional sequencing method using UNG [145].

IV. Reference

[1] M. S. Cooke et al, "Oxidative DNA damage: mechanisms, mutation, and disease," The FASEB Journal, vol. 17, (10), pp. 1195-1214, 2003. DOI: 10.1096/fj.02-0752rev.

[2] C. Li et al, "The role of 8-oxoG repair systems in tumorigenesis and cancer therapy," Cells, vol. 11, (23), pp. 3798, 2022. DOI: 10.3390/cells11233798.

[3] L. J. Marnett, "Oxyradicals and DNA damage," Carcinogenesis, vol. 21, (3), pp. 361-370, 2000. DOI: 10.1093/carcin/21.3.361.

[4] S. Ohnishi et al, "DNA damage in inflammation-related carcinogenesis and cancer stem cells," Oxidative Medicine and Cellular Longevity, vol. 2013, pp. 387014-9, 2013. DOI: 10.1155/2013/387014.

[5] Z. I. Alam et al, "Oxidative DNA damage in the parkinsonian brain: an apparent selective increase in 8-hydroxyguanine levels in substantia nigra," Journal of Neurochemistry, vol. 69, (3), pp. 1196-1203, 1997. DOI: 10.1046/j.1471-4159.1997.69031196.x.

[6] M. A. Lovell and W. R. Markesbery, "Oxidative DNA damage in mild cognitive impairment and late-stage Alzheimer's disease," Nucleic Acids Research, vol. 35, (22), pp. 7497-7504, 2007. DOI: 10.1093/nar/gkm821.

[7] H. Ikehata and T. Ono, "The mechanisms of UV mutagenesis," Journal of Radiation Research, vol. 52, (2), pp. 115-125, 2011. DOI: 10.1269/jrr.10175.

[8] T. Iyama and D. M. Wilson, "DNA repair mechanisms in dividing and non-dividing cells," DNA Repair, vol. 12, (8), pp. 620-636, 2013. DOI: 10.1016/j.dnarep.2013.04.015.

[9] D. O. Zharkov, "Base excision DNA repair," Cell. Mol. Life Sci, vol. 65, (10), pp. 1544-1565, 2008. DOI: 10.1007/s00018-008-7543-2.

[10] N. Chatterjee and G. C. Walker, "Mechanisms of DNA damage, repair, and mutagenesis," Environmental and Molecular Mutagenesis, vol. 58, (5), pp. 235-263, 2017. DOI: 10.1002/em.22087.

[11] S. A. Nick McElhinny et al, "Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases," Proceedings of the National Academy of Sciences - PNAS, vol. 107, (11), pp. 4949-4954, 2010. DOI: 10.1073/pnas.0914857107.

[12] S. Andersen et al, "Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts," Carcinogenesis, vol. 26, (3), pp. 547-555, 2005. DOI: 10.1093/carcin/bgh347.

[13] S. Kumar, V. Chinnusamy and T. Mohapatra, "Epigenetics of modified DNA bases: 5-methylcytosine and beyond," Frontiers in Genetics, vol. 9, pp. 640, 2018. DOI: 10.3389/fgene.2018.00640.

[14] B. Tudek, S. Boiteux and J. Laval, "Biological properties of imidazole ring-opened N7-methylguanine in M13mp18 phage DNA," Nucleic Acids Research, vol. 20, (12), pp. 3079-3084, 1992. DOI: 10.1093/nar/20.12.3079.

[15] M. D. Wyatt and D. L. Pittman, "Methylating agents and DNA repair responses: methylated bases and sources of strand breaks," Chemical Research in Toxicology, vol. 19, (12), pp. 1580-1594, 2006. DOI: 10.1021/tx060164e.

[16] E. L. Loechler, C. L. Green and J. M. Essigmann, "In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome," Proceedings of the National Academy of Sciences - PNAS, vol. 81, (20), pp. 6271-6275, 1984. DOI: 10.1073/pnas.81.20.6271.

[17] G. T. Pauly and R. C. Moschel, "Mutagenesis by O(6)-methyl-, O(6)-ethyl-, and O(6)-benzylguanine and O(4)-methylthymine in human cells: effects of O(6)-alkylguanine-DNA alkyltransferase and mismatch repair," Chemical Research in Toxicology, vol. 14, (7), pp. 894-900, 2001. DOI: 10.1021/tx010032f.

[18] G. Fronza and B. Gold, "The biological effects of N3-methyladenine," Journal of Cellular Biochemistry, vol. 91, (2), pp. 250-257, 2004. DOI: 10.1002/jcb.10698.

[19] C. J. Burrows and J. G. Muller, "Oxidative nucleobase modifications leading to strand scission," Chemical Reviews, vol. 98, (3), pp. 1109-1152, 1998. DOI: 10.1021/cr960421s.

[20] B. van Loon, E. Markkanen and U. Hübscher, "Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine," DNA Repair, vol. 9, (6), pp. 604-616, 2010. DOI: 10.1016/j.dnarep.2010.03.004.

[21] D. Mangal et al, "Analysis of 7,8-dihydro-8-oxo-2′-deoxyguanosine in cellular DNA during oxidative stress," Chemical Research in Toxicology, vol. 22, (5), pp. 788-797, 2009. DOI: 10.1021/tx800343c.

[22] A. R. Collins et al, "Oxidative damage to DNA: do we have a reliable biomarker?" Environmental Health Perspectives, vol. 104, (suppl 3), pp. 465-469, 1996. DOI: 10.1289/ehp.96104s3465.

[23] Q. Zhang et al, "Escherichia coli MutY protein has a guanine-DNA glycosylase that acts on 7,8-dihydro-8-oxoguanine:guanine mispair to prevent spontaneous G:C→C:G transversions," Nucleic Acids Research, vol. 26, (20), pp. 4669-4675, 1998. DOI: 10.1093/nar/26.20.4669.

[24] M. M. L. de Sousa et al, "Impact of oxidative DNA damage and the role of DNA glycosylases in neurological dysfunction," Int J Mol Sci, vol. 22, (23), pp. 12924, 2021. DOI: 10.3390/ijms222312924.

[25] S. Maynard et al, "Base excision repair of oxidative DNA damage and association with cancer and aging," Carcinogenesis, vol. 30, (1), pp. 2-10, 2009. DOI: 10.1093/carcin/bgn250.

[26] Q. M. Zhang, "Role of the Escherichia coli and human DNA glycosylases that remove 5-formyluracil from DNA in the prevention of mutations," Journal of Radiation Research, vol. 42, (1), pp. 11-19, 2001. DOI: 10.1269/jrr.42.11.

[27] H. E. Krokan, R. Standal and G. Slupphaug, "DNA glycosylases in the base excision repair of DNA," Biochemical Journal, vol. 325 (Pt 1), (1), pp. 1-16, 1997. DOI: 10.1042/bj3250001.

[28] H. E. Krokan, F. Drabløs and G. Slupphaug, "Uracil in DNA --occurrence, consequences and repair," Oncogene, vol. 21, (58), pp. 8935-8948, 2002. DOI: 10.1038/sj.onc.1205996.

[29] D. E. Barnes and T. Lindahl, " Repair and genetic consequences of endogenous DNA base damage in mammalian cells," Annual Review of Genetics, vol. 38, (1), pp. 445-476, 2004. DOI: 10.1146/annurev.genet.38.072902.092448.

[30] T. Lindahl, "Instability and decay of the primary structure of DNA," Nature, vol. 362, (6422), pp. 709-715, 1993. DOI: 10.1038/362709a0.

[31] S. Yonekura et al, "Generation, biological consequences and repair mechanisms of cytosine deamination in DNA," Journal of Radiation Research, vol. 50, (1), pp. 19-26, 2009. DOI: 10.1269/jrr.08080.

[32] B. K. Duncan and J. H. Miller, "Mutagenic deamination of cytosine residues in DNA," Nature, vol. 287, (5782), pp. 560-561, 1980. DOI: 10.1038/287560a0.

[33] R. De Bont and N. van Larebeke, "Endogenous DNA damage in humans: a review of quantitative data," Mutagenesis, vol. 19, (3), pp. 169-185, 2004. DOI: 10.1093/mutage/geh025.

[34] T. E. Parry, "On the mutagenic action of adenine," Leukemia Research, vol. 31, (12), pp. 1621-1624, 2007. DOI: 10.1016/j.leukres.2007.05.002.

[35] G. E. Wuenschell, T. R. O'Connor and J. Termini, "Stability, miscoding potential, and repair of 2'-deoxyxanthosine in DNA: implications for nitric oxide-induced mutagenesis," Biochemistry, vol. 42, (12), pp. 3608-3616, 2003. DOI: 10.1021/bi0205597.

[36] C. S. Nabel, S. A. Manning and R. M. Kohli, "The curious chemical biology of cytosine: deamination, methylation,and oxidation as modulators of genomic potential," ACS Chemical Biology, vol. 7, (1), pp. 20-30, 2012. DOI: 10.1021/cb2002895.

[37] V. Chandra, A. Bortnick and C. Murre, "AID targeting: old mysteries and new challenges," Trends in Immunology, vol. 36, (9), pp. 527-535, 2015. DOI: 10.1016/j.it.2015.07.003.

[38] S. G. Conticello, "The AID/APOBEC family of nucleic acid mutators," Genome Biology, vol. 9, (6), pp. 229, 2008. DOI: 10.1186/gb-2008-9-6-229.

[39] R. Pecori et al, "Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination," Nature Reviews Genetics, vol. 23, (8), pp. 505-518, 2022. DOI: 10.1038/s41576-022-00459-8.

[40] P. J. Gearhart et al, "Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions," Nature Immunology, vol. 12, (1), pp. 70-76, 2011. DOI: 10.1038/ni.1970.

[41] J. L. Tubbs, A. E. Pegg and J. A. Tainer, "DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy," DNA Repair, vol. 6, (8), pp. 1100-1115, 2007. DOI: 10.1016/j.dnarep.2007.03.011.

[42] Q. Fang et al, "Repair of O6-G-alkyl-O6-G interstrand cross-links by human O6-alkylguanine-DNA alkyltransferase," Biochemistry, vol. 47, (41), pp. 10892-10903, 2008. DOI: 10.1021/bi8008664.

[43] A. E. Pegg, "Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools," Chemical Research in Toxicology, vol. 24, (5), pp. 618-639, 2011. DOI: 10.1021/tx200031q.

[44] C. Yi and C. He, "DNA repair by reversal of DNA damage," Cold Spring Harbor Perspectives in Biology, vol. 5, (1), pp. a012575, 2013. DOI: 10.1101/cshperspect.a012575.

[45] L. C. J. Gillet and O. D. Schärer, "Molecular mechanisms of mammalian global genome nucleotide excision repair," Chemical Reviews, vol. 106, (2), pp. 253-276, 2006. DOI: 10.1021/cr040483f.

[46] J. Jiricny, "The multifaceted mismatch-repair system," Nature Reviews Molecular Cell Biology, vol. 7, (5), pp. 335-346, 2006. DOI: 10.1038/nrm1907.

[47] H. E. Krokan and M. Bjørås, "Base excision repair," Cold Spring Harbor Perspectives in Biology, vol. 5, (4), pp. a012583, 2013. DOI: 10.1101/cshperspect.a012583.

[48] R. Gary et al, "Proliferating cell nuclear antigen facilitates excision in long-patch base excision repair," The Journal of Biological Chemistry, vol. 274, (7), pp. 4354-4363, 1999. DOI: 10.1074/jbc.274.7.4354.

[49] D. O. Zharkov, G. Shoham and A. P. Grollman, "Structural characterization of the Fpg family of DNA glycosylases," DNA Repair, vol. 2, (8), pp. 839-862, 2003. DOI: 10.1016/s1568-7864(03)00084-3.

[50] D. Jiang et al, "Characterization of Escherichia coli endonuclease VIII," The Journal of Biological Chemistry, vol. 272, (51), pp. 32230-32239, 1997. DOI: 10.1074/jbc.272.51.32230.

[51] A. Prakash, S. Doublié and S. S. Wallace, "The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage," Progress in Molecular

Biology and Translational Science, vol. 110, pp. 71-91, 2012. DOI: 10.1016/b978-0-12-387665-2.00004-3.

[52] S. S. David and S. D. Williams, "Chemistry of glycosylases and endonucleases involved in base-excision repair," Chemical Reviews, vol. 98, (3), pp. 1221-1262, 1998. DOI: 10.1021/cr980321h.

[53] T. Lindahl et al, "DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli," The Journal of Biological Chemistry, vol. 252, (10), pp. 3286-3294, 1977. DOI: 10.1016/S0021-9258(17)40386-3.

[54] L. Aravind and E. V. Koonin, "The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates," Genome Biology, vol. 1, (4), pp. RESEARCH0007, 2000. DOI: 10.1186/gb-2000-1-4-research0007.

[55] H. Nilsen et al, "Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene," Nucleic Acids Research, vol. 25, (4), pp. 750-755, 1997. DOI: 10.1093/nar/25.4.750.

[56] M. Akbari et al, "Different organization of base excision repair of uracil in DNA in nuclei and mitochondria and selective upregulation of mitochondrial uracil-DNA glycosylase after oxidative stress," Neuroscience, vol. 145, (4), pp. 1201-1212, 2007. DOI: 10.1016/j.neuroscience.2006.10.010.

[57] L. H. Pearl, "Structure and function in the uracil-DNA glycosylase superfamily," Mutation Research, vol. 460, (3-4), pp. 165-181, 2000. DOI: 10.1016/s0921-8777(00)00025-2.

[58] S. S. Parikh, C. D. Putnam and J. A. Tainer, "Lessons learned from structural results on uracil-DNA glycosylase," Mutation Research, vol. 460, (3-4), pp. 183-199, 2000. DOI: 10.1016/s0921-8777(00)00026-4.

[59] D. O. Zharkov, G. V. Mechetin and G. A. Nevinsky, "Uracil-DNA glycosylase: structural, thermodynamic and kinetic aspects of lesion search and recognition," Mutation Research, vol. 685, (1), pp. 11-20, 2010. DOI: 10.1016/j.mrfmmm.2009.10.017.

[60] H. R. Warner and P. A. Rockstroh, "Incorporation and excision of 5-fluorouracil from deoxyribonucleic acid in Escherichia coli," Journal of Bacteriology, vol. 141, (2), pp. 680-686, 1980. DOI: 10.1128/JB.141.2.680-686.1980.

[61] Z. Hatahet et al, "New substrates for old enzymes. 5-Hydroxy-2'-deoxycytidine and 5-hydroxy-2'-deoxyuridine are substrates for Escherichia coli endonuclease III and formamidopyrimidine DNA N-glycosylase, while 5-hydroxy-2'-deoxyuridine is a substrate for uracil DNA N-glycosylase," The Journal of Biological Chemistry, vol. 269, (29), pp. 18814-18820, 1994.

[62] M. Dizdaroglu et al, "Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage," Nucleic Acids Research, vol. 24, (3), pp. 418-422, 1996. DOI: 10.1093/nar/24.3.418.

[63] C. D. Mol et al, "Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis," Cell, vol. 80, (6), pp. 869-878, 1995. DOI: 10.1016/0092-8674(95)90290-2.

[64] S. S. Parikh et al, "Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects,"

Proceedings of the National Academy of Sciences - PNAS, vol. 97, (10), pp. 5083-5088, 2000. DOI: 10.1073/pnas.97.10.5083.

[65] A. Rashtchian et al, " Uracil DNA glycosylase-mediated cloning of polymerase chain reaction-amplified DNA: application to genomic and cDNA cloning," Analytical Biochemistry, vol. 206, (1), pp. 91-97, 1992. DOI: 10.1016/s0003-2697(05)80015-6.

[66] S. K. Poddar, M. H. Sawyer and J. D. Connor, "Optimized PCR amplification of influenza A virus RNA using Tth DNA polymerase, incorporating uracil N glycosylase (UNG) in a single tube reaction," Journal of Clinical Laboratory Analysis, vol. 11, (6), pp. 323-327, 1997. DOI: 10.1002/(sici)1098-2825(1997)11:6<323::aid-jcla2>3.0.co;2-6.

[67] J. K. Ball and U. Desselberger, "The use of uracil-N-glycosylase in the preparation of PCR products of direct sequencing," Nucleic Acids Research, vol. 20, (12), pp. 3255, 1992. DOI: 10.1093/nar/20.12.3255.

[68] E. W. Taggart et al, "Use of heat labile UNG in an RT-PCR assay for enterovirus detection," Journal of Virological Methods, vol. 105, (1), pp. 57-65, 2002. DOI: 10.1016/S0166-0934(02)00080-0.

[69] G. A. Kim et al, "Characterization of cold-active uracil-DNA glycosylase from Bacillus sp. HJ171 and its use for contamination control in PCR," Appl Microbiol Biotechnol, vol. 80, (5), pp. 785-794, 2008. DOI: 10.1007/s00253-008-1585-0.

[70] P. Vaughan and T. V. McCarthy, "A novel process for mutation detection using uracil DNA-glycosylase," Nucleic Acids Research, vol. 26, (3), pp. 810-815, 1998. DOI: 10.1093/nar/26.3.810.

[71] J. Park et al, "High expression of uracil DNA glycosylase determines C to T substitution in human pluripotent stem cells," Molecular Therapy Nucleic Acids, vol. 27, pp. 175-183, 2022. DOI: 10.1016/j.omtn.2021.11.023.

[72] K. Wiebauer and J. Jiricny, "In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells," Nature, vol. 339, (6221), pp. 234-236, 1989. DOI: 10.1038/339234a0.

[73] P. Neddermann and J. Jiricny, "The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells," The Journal of Biological Chemistry, vol. 268, (28), pp. 21218-21224, 1993.

[74] P. Neddermann et al, "Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase," The Journal of Biological Chemistry, vol. 271, (22), pp. 12767-12774, 1996. DOI: 10.1074/jbc.271.22.12767.

[75] P. Gallinari and J. Jiricny, "A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase," Nature, vol. 383, (6602), pp. 735-738, 1996. DOI: 10.1038/383735a0.

[76] P. Neddermann and J. Jiricny, "Efficient removal of uracil from G.U mispairs by themismatch-specific thymine DNA glycosylase from HeLa cells," Proceedings of the National Academy of Sciences - PNAS, vol. 91, (5), pp. 1642-1646, 1994. DOI: 10.1073/pnas.91.5.1642.

[77] D. Cortázar et al, "The enigmatic thymine DNA glycosylase," DNA Repair, vol. 6, (4), pp. 489-504, 2007. DOI: 10.1016/j.dnarep.2006.10.013.

[78] U. Hardeland et al, "The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs," Nucleic Acids Research, vol. 31, (9), pp. 2261-2271, 2003. DOI: 10.1093/nar/gkg344.

[79] M. Saparbaev et al, "1,N(2)-ethenoguanine, a mutagenic DNA adduct, is a primary substrate of Escherichia coli mismatch-specific uracil-DNA glycosylase and human alkylpurine-DNA-N-glycosylase," The Journal of Biological Chemistry, vol. 277, (30), pp. 26987-26993, 2002. DOI: 10.1074/jbc.M111100200.

[80] E. Borys-Brzywczy et al, "Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts," Acta Biochimica Polonica, vol. 52, (1), pp. 149-165, 2005.

[81] H. Lee et al, "Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase," The Journal of Biological Chemistry, vol. 285, (53), pp. 41483-41490, 2010. DOI: 10.1074/jbc.M110.150003.

[82] J. Yoon et al, "Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair," Nucleic Acids Research, vol. 31, (18), pp. 5399-5404, 2003. DOI: 10.1093/nar/gkg730.

[83] B. Zhu et al, "5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex," Proceedings of the National Academy of Sciences - PNAS, vol. 97, (10), pp. 5135-5139, 2000. DOI: 10.1073/pnas.100107597.

[84] L. Dong et al, "Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase," DNA Repair, vol. 7, (12), pp. 1962-1972, 2008. DOI: 10.1016/j.dnarep.2008.08.006.

[85] E. Moe et al, "The crystal structure of mismatch-specific uracil-DNA glycosylase (MUG) from Deinococcus radiodurans reveals a novel catalytic residue and broad substrate specificity," The Journal of Biological Chemistry, vol. 281, (1), pp. 569-577, 2006. DOI: 10.1074/jbc.M508032200.

[86] S. Um et al, "Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase," The Journal of Biological Chemistry, vol. 273, (33), pp. 20728-20736, 1998. DOI: 10.1074/jbc.273.33.20728.

[87] J. Bastien and C. Rochette-Egly, "Nuclear retinoid receptors and the transcription of retinoid-target genes," Gene, vol. 328, pp. 1-16, 2004. DOI: 10.1016/j.gene.2003.12.005.

[88] D. Chen et al, "T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor α," The Journal of Biological Chemistry, vol. 278, (40), pp. 38586-38592, 2003. DOI: 10.1074/jbc.M304286200.

[89] M. Tini et al, "Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription," Molecular Cell, vol. 9, (2), pp. 265-277, 2002. DOI: 10.1016/S1097-2765(02)00453-7.

[90] C. Missero et al, "The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription," The Journal

of Biological Chemistry, vol. 276, (36), pp. 33569-33575, 2001. DOI: 10.1074/jbc.M104963200.

[91] J. P. Jost et al, "Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase," The Journal of Biological Chemistry, vol. 270, (17), pp. 9734-9739, 1995. DOI: 10.1074/jbc.270.17.9734.

[92] A. Maiti and A. C. Drohat, "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites," The Journal of Biological Chemistry, vol. 286, (41), pp. 35334-35338, 2011. DOI: 10.1074/jbc.C111.284620.

[93] K. A. Haushalter et al, "Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors," Current Biology, vol. 9, (4), pp. 174-185, 1999. DOI: 10.1016/s0960-9822(99)80087-6.

[94] P. Pang et al, "SMUG2 DNA glycosylase from Pedobacter heparinus as a new subfamily of the UDG superfamily," Biochemical Journal, vol. 474, (6), pp. 923-938, 2017. DOI: 10.1042/BCJ20160934.

[95] H. Nilsen et al, "Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase," The EMBO Journal, vol. 20, (15), pp. 4278-4286, 2001. DOI: 10.1093/emboj/20.15.4278.

[96] R. J. Boorstein et al, "Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1," The Journal of Biological Chemistry, vol. 276, (45), pp. 41991-41997, 2001. DOI: 10.1074/jbc.M106953200.

[97] B. Kavli et al, "hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup," The Journal of Biological Chemistry, vol. 277, (42), pp. 39926-39936, 2002. DOI: 10.1074/jbc.M207107200.

[98] R. Mi et al, "Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG," Journal of Molecular Biology, vol. 385, (3), pp. 761-778, 2009. DOI: 10.1016/j.jmb.2008.09.038.

[99] J. Li et al, "Identification of a prototypical single-stranded uracil DNA glycosylase from Listeria innocua," DNA Repair, vol. 57, pp. 107-115, 2017. DOI: 10.1016/j.dnarep.2017.07.001.

[100] C. Chang et al, "Screening of glycosylase activity on oxidative derivatives of methylcytosine: Pedobacter heparinus SMUG2 as a formylcytosine- and carboxylcytosine-DNA glycosylase," DNA Repair, vol. 119, pp. 103408, 2022. DOI: 10.1016/j.dnarep.2022.103408.

[101] J. E. A. Wibley et al, "Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1," Molecular Cell, vol. 11, (6), pp. 1647-1659, 2003. DOI: 10.1016/S1097-2765(03)00235-1.

[102] H. Nilsen, T. Lindahl and A. Verreault, "DNA base excision repair of uracil residues in reconstituted nucleosome core particles," The EMBO Journal, vol. 21, (21), pp. 5943-5952, 2002. DOI: 10.1093/emboj/cdf581.

[103] M. E. Tarantino et al, "Nucleosomes and the three glycosylases: high, medium, and low levels of excision by the uracil DNA glycosylase superfamily," DNA Repair, vol. 72, pp. 56-63, 2018. DOI: 10.1016/j.dnarep.2018.09.008.

[104] S. Raja and B. Van Houten, "The multiple cellular roles of SMUG1 in genome maintenance and cancer," Int J Mol Sci, vol. 22, (4), pp. 1981, 2021. DOI: 10.3390/ijms22041981.

[105] Q. An et al, "C → T mutagenesis and γ-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases," The EMBO Journal, vol. 24, (12), pp. 2205-2213, 2005. DOI: 10.1038/sj.emboj.7600689.

[106] Q. An et al, "5-Fluorouracil incorporated into DNA is excised by the Smug1 DNA glycosylase to reduce drug cytotoxicity," Cancer Research, vol. 67, (3), pp. 940-945, 2007. DOI: 10.1158/0008-5472.CAN-06-2960.

[107] J. An et al, "A preliminary study of genes related to concomitant chemoradiotherapy resistance in advanced uterine cervical squamous cell carcinoma," Chinese Medical Journal, vol. 126, (21), pp. 4109-4115, 2013. DOI: 10.3760/cma.j.issn.0366-6999.20131320.

[108] L. Jobert et al, "The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control," Molecular Cell, vol. 49, (2), pp. 339-345, 2013. DOI: 10.1016/j.molcel.2012.11.010.

[109] P. Kroustallaki et al, "SMUG1 promotes telomere maintenance through telomerase RNA processing," Cell Reports, vol. 28, (7), pp. 1690-1702.e10, 2019. DOI: 10.1016/j.celrep.2019.07.040.

[110] A. Koulis et al, "Uracil-DNA glycosylase activities in hyperthermophilic micro-organisms," FEMS Microbiology Letters, vol. 143, (2-3), pp. 267-271, 1996. DOI: 10.1111/j.1574-6968.1996.tb08491.x.

[111] M. Sandigursky and W. A. Franklin, "Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes," Current Biology, vol. 9, (10), pp. 531-534, 1999. DOI: 10.1016/s0960-9822(99)80237-1.

[112] J. Hoseki et al, "Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8," Journal of Molecular Biology, vol. 333, (3), pp. 515-526, 2003. DOI: 10.1016/j.jmb.2003.08.030.

[113] J. A. Hinks et al, "An iron-sulfur cluster in the family 4 uracil-DNA glycosylases," The Journal of Biological Chemistry, vol. 277, (19), pp. 16936-16940, 2002. DOI: 10.1074/jbc.M200668200.

[114] N. D. Lanz and S. J. Booker, "Auxiliary iron–sulfur cofactors in radical SAM enzymes," Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research, vol. 1853, (6), pp. 1316-1334, 2015. DOI: 10.1016/j.bbamcr.2015.01.002.

[115] B. Xia et al, "Correlated mutation in the evolution of catalysis in uracil DNA glycosylase superfamily," Scientific Reports, vol. 7, (1), pp. 45978, 2017. DOI: 10.1038/srep45978.

[116] Q. Gan et al, "Characterization of a Family IV uracil DNA glycosylase from the hyperthermophilic euryarchaeon Thermococcus barophilus Ch5," International Journal of Biological Macromolecules, vol. 146, pp. 475-481, 2020. DOI: 10.1016/j.ijbiomac.2019.12.202.

[117] T. Sakai et al, "Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus," DNA Repair, vol. 7, (4), pp. 663-669, 2008. DOI: 10.1016/j.dnarep.2008.01.006.

[118] I. Dionne and S. D. Bell, "Characterization of an archaeal family 4 uracil DNA glycosylase and its interaction with PCNA and chromatin proteins," Biochemical Journal, vol. 387, (Pt 3), pp. 859-863, 2005. DOI: 10.1042/BJ20041661.

[119] J. B. Vivona and Z. Kelman, "The diverse spectrum of sliding clamp interacting proteins," FEBS Letters, vol. 546, (2-3), pp. 167, 2003. DOI: 10.1016/s0014-5793(03)00622-7.

[120] M. Otterlei et al, "Post-replicative base excision repair in replication foci," The EMBO Journal, vol. 18, (13), pp. 3834-3844, 1999. DOI: 10.1093/emboj/18.13.3834.

[121] H. Yang et al, "Direct interaction between uracil-DNA glycosylase and a proliferating cell nuclear antigen homolog in the crenarchaeon Pyrobaculum aerophilum," The Journal of Biological Chemistry, vol. 277, (25), pp. 22271-22278, 2002. DOI: 10.1074/jbc.M201820200.

[122] V. Starkuviene and H. Fritz, "A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus," Nucleic Acids Research, vol. 30, (10), pp. 2097-2102, 2002. DOI: 10.1093/nar/30.10.2097.

[123] A. A. Sartori et al, "A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site," The EMBO Journal, vol. 21, (12), pp. 3182-3191, 2002. DOI: 10.1093/emboj/cdf309.

[124] B. Xia et al, "Specificity and catalytic mechanism in family 5 uracil DNA glycosylase," The Journal of Biological Chemistry, vol. 289, (26), pp. 18413-18426, 2014. DOI: 10.1074/jbc.M114.567354.

[125] H. Kosaka et al, "Crystal structure of family 5 uracil-DNA glycosylase bound to DNA," Journal of Molecular Biology, vol. 373, (4), pp. 839-850, 2007. DOI: 10.1016/j.jmb.2007.08.022.

[126] H. Lee, B. N. Dominy and W. Cao, "New family of deamination repair enzymes in uracil-DNA glycosylase superfamily," The Journal of Biological Chemistry, vol. 286, (36), pp. 31282-31287, 2011. DOI: 10.1074/jbc.M111.249524.

[127] P. B. Sang et al, "A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, vol. 43, (17), pp. 8452-8463, 2015. DOI: 10.1093/nar/gkv854.

[128] W. Ahn et al, "Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision," Nature Chemical Biology, vol. 15, (6), pp. 607-614, 2019. DOI: 10.1038/s41589-019-0289-3.

[129] J. Tu et al, "Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation," Nature Chemical Biology, vol. 15, (6), pp. 615-622, 2019. DOI: 10.1038/s41589-019-0290-x.

[130] P. Weickert and J. Stingele, "DNA-protein crosslinks and their resolution," Annual Review of Biochemistry, vol. 91, (1), pp. 157-181, 2022. DOI: 10.1146/annurev-biochem-032620-105820.

[131] Y. Pommier et al, "Roles of eukaryotic topoisomerases in transcription, replication and genomic stability," Nature Reviews Molecular Cell Biology, vol. 17, (11), pp. 703-721, 2016. DOI: 10.1038/nrm.2016.111.

[132] J. J. Champoux, "DNA topoisomerases: structure, function, and mechanism," Annual Review of Biochemistry, vol. 70, (1), pp. 369-413, 2001. DOI: 10.1146/annurev.biochem.70.1.369.

[133] R. D. Woods et al, "Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases," Nucleic Acids Research, vol. 44, (2), pp. 801-810, 2016. DOI: 10.1093/nar/gkv1469.

[134] S. Keeney, C. N. Giroux and N. Kleckner, "Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family," Cell, vol. 88, (3), pp. 375-384, 1997. DOI: 10.1016/S0092-8674(00)81876-0.

[135] S. Keeney, M. J. Neale and J. Pan, "Endonucleolytic processing of covalent protein-linked DNA double-strand breaks," Nature, vol. 436, (7053), pp. 1053-1057, 2005. DOI: 10.1038/nature03872.

[136] R. A. Caston and B. Demple, "Risky repair: DNA-protein crosslinks formed by mitochondrial base excision DNA repair enzymes acting on free radical lesions," Free Radical Biology & Medicine, vol. 107, pp. 146-150, 2017. DOI: 10.1016/j.freeradbiomed.2016.11.025.

[137] M. S. DeMott et al, "Covalent trapping of human DNA polymerase beta by the oxidative DNA lesion 2-deoxyribonolactone," The Journal of Biological Chemistry, vol. 277, (10), pp. 7637-7640, 2002. DOI: 10.1074/jbc.C100577200.

[138] T. Nakano et al, "DNA-protein cross-link formation mediated by oxanine. A novel genotoxic mechanism of nitric oxide-induced DNA damage," The Journal of Biological Chemistry, vol. 278, (27), pp. 25264-25272, 2003. DOI: 10.1074/jbc.M212847200.

[139] Y. Le Bihan et al, "5-Hydroxy-5-methylhydantoin DNA lesion, a molecular trap for DNA glycosylases," Nucleic Acids Research, vol. 39, (14), pp. 6277-6290, 2011. DOI: 10.1093/nar/gkr215.

[140] J. L. Quiñones et al, "Enzyme mechanism-based, oxidative DNA–protein cross-links formed with DNA polymerase β in vivo," Proc. Natl. Acad. Sci. U. S. A., vol. 112, (28), pp. 8602, 2015. DOI: 10.1073/pnas.1501101112.

[141] J. Stingele and S. Jentsch, "DNA–protein crosslink repair," Nature Reviews Molecular Cell Biology, vol. 16, (8), pp. 455-460, 2015. DOI: 10.1038/nrm4015.

[142] T. Lindahl, B. Demple and P. Robins, "Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase," The EMBO Journal, vol. 1, (11), pp. 1359-1363, 1982. DOI: 10.1002/j.1460-2075.1982.tb01323.x.

[143] M. Datta et al, "Development of mCherry tagged UdgX as a highly sensitive molecular probe for specific detection of uracils in DNA," Biochemical and Biophysical Research Communications, vol. 518, (1), pp. 38-43, 2019. DOI: 10.1016/j.bbrc.2019.08.005.

[144] J. A. Stewart, G. Schauer and A. S. Bhagwat, "Visualization of uracils created by APOBEC3A using UdgX shows colocalization with RPA at stalled replication forks," Nucleic Acids Research, vol. 48, (20), pp. e118, 2020. DOI: 10.1093/nar/gkaa845.

[145] L. Jiang et al, "UdgX-mediated uracil sequencing at single-nucleotide resolution," J. Am. Chem. Soc., vol. 144, (3), pp. 1323, 2022. DOI: 10.1021/jacs.1c11269.

CHAPTER TWO

STRUCTURAL AND FUNCTIONAL COUPLING IN CROSSLINKING URACIL-DNA GLYCOSYLASE UDGX

I. Abstract

Enzymes in uracil-DNA glycosylase (UDG) superfamily are involved in the removal of deaminated nucleobases such as uracil, methylcytosine derivatives such as formylcytosine and carboxylcytosine, and other base damage in DNA repair. UDGX is the latest addition of a new class to the UDG superfamily with a sporadic distribution in bacteria. UDGX-type enzymes have a distinct biochemical property of crosslinking themselves to the resulting AP site after uracil removal. Built on previous biochemical and structural analyses, this work comprehensively investigated the kinetic and enzymatic properties of *Mycobacterium smegmatis* UDGX. Kinetics and mutational analyses, coupled with structural information, defined the roles of E52, D56, D59, F65 of motif 1, H178 of motif 2 and N91, K94, R107 and H109 of motif 3 play in uracil excision and crosslinking. More importantly, a series of quantitative analyses underscored the structural coupling through inter-motif and intra-motif interactions and subsequent functional coupling of the uracil excision and crosslinking reactions. A catalytic model is proposed, which underlies this catalytic feature unique to UDGX type enzymes. This study offers new insight on the catalytic mechanism of UDGX and provides a unique example of enzyme evolution.

II. Introduction

Uracil generated from cytosine deamination is a common base damage in DNA. Enzymes in Uracil-DNA Glycosylase (UDG) superfamily are monofunctional glycosylases involved in the removal of uracil and other types of base modifications from DNA. UDG superfamily is classified into at least six families based on the sequence conservation of three catalytic motifs. Family 1 UNG (uracil-*N*-glycosylase) is a group of narrow specificity but highly efficient enzymes, represented by *E. coli* UNG as the first DNA glycosylase discovered [1-3]. Family 2 TDG/MUG (thymine-DNA glycosylase/mismatch-specific uracil-DNA glycosylase) is a group of UDGs with broad substrate specificity, excising uracil and other base lesions [4-11]. Human TDG, although previously discovered as a thymine DNA glycosylase removing thymine from G/T mispairs, is now recognized as a demethylase to remove formylcytosine (fC) and carboxylcytosine (caC) generated by TET-mediated oxidation of methylcytosine (mC) during the demethylation process [12, 13]. Family 3 SMUG1 (single strand selective monofunctional uracil-DNA glycosylase) contains several subfamilies with different substrate specificities [14-20]. Family 4 UDGa was found as a group of narrow specificity but highly efficient enzymes in bacterial organisms [21-24]. Family 5 UDGb enzymes were found in bacteria and archaea with relatively broad substrate specificities and moderate catalytic efficiencies [25-27]. Family 6 HDG (hypoxanthine-DNA glycosylase) is predominantly a hypoxanthine DNA glycosylase [28].

In recent years, a group of highly unusual UDG (UDGX) was found in certain bacterial species, in which the enzyme forms a covalent bond with an AP site after excision of a

uracil base, i.e., the enzyme and the DNA become a crosslinked protein-DNA complex after the first catalytic step to remove a uracil [29]. This unique enzymatic property has been explored to develop tools for detection or visualization of uracil in DNA [30-32]. Structural and biochemical studies have identified a histidine residue (H109) in the extended loop of motif 3 in UDGX as the site of crosslinking [33, 34]. This extended loop has been recognized as a significant structural feature different from the other families in UDG superfamily. Despite of the availability of crystal structures of UDGX, the basic kinetic information is missing and the roles of some key residues in uracil excision and AP site crosslinking are not well defined. The unique enzymatic function of UDGX in relationship to its catalytic motifs as well as structure is not completely understood. In this work, a combination of quantitative and biochemical studies was conducted to define the roles of a series of amino acid residues of motifs 1, 2 and 3 in uracil excision and protein-DNA crosslinking. Furthermore, a catalytic mechanism underlying the unique enzymatic properties of UDGX enzymes was proposed.

III. Materials and Methods

A. Reagents, media and strains

All routine reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), VWR (Suwanee, GA), ThermoFisher Scientific (Waltham, MA) or Gel Company (San Francisco, CA) 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. Plasmid miniprep kits and DNA gel extraction kits

were purchased from New England Biolabs (Ipswich, MA) and ThermoFisher Scientific (Waltham, MA). Restriction enzymes, Phusion DNA polymerase, T4 DNA ligase and dNTP were purchased from ThermoFisher Scientific (Waltham, MA). Cod UNG (heat-labile UDG from *Gadus morhua*) was purchased from ArcticZymes (Tromsø, Norway). HisTrap FF (1 mL), HiTrap Q FF (1 mL) and HiTrap SP FF (1 mL) 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. Gene strands and oligonucleotide primers for PCR were synthesized from Eurofins Genomics (Huntsville, AL). Oligodeoxynucleotide substrates with carboxyfluorescein (FAM) fluorescence label were ordered from Integrated DNA Technologies Inc. (Coralville, IA). The LB medium was prepared according to standard recipes. The sonication buffer consisted of 50 mM Tris-HCl (pH 7.5), 300 mM NaCl and 40 mM imidazole. Buffer A for HisTrap FF columns 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, 500 mM imidazole and 10% glycerol. Buffer A for HiTrap Q/SP FF columns consisted of 50 mM Tris-HCl (pH 7.5) and buffer B consisted of 50 mM Tris-HCl (pH 7.5) and 1 M NaCl. *E. coli* strain DH5α was purchased from ThermoFisher Scientific (Waltham, MA) and *E. coli* strain Rosetta was purchased from VWR (Suwanee, GA).

B. Identification of UDG genes in sequenced genomes

A tree of life based on 81 bacterial species was obtained from TimeTree [35] and visualized by iTOL [36]. UDG gene distribution of each bacterial species was superimposed into the tree of life. The amino acid sequences of UDG families and UDGX

genes were obtained by searching non-redundant protein sequences (nr) database within NCBI using BLASTP [37]. For family 1 UNG, family 2 TDG/MUG, family 3 SMUG1, family 4 UDGa, family 5 UDGb, family 6 HDG and UDGX genes, *Escherichia coli* UNG (Genbank accession number NP_289138.1), *E. coli* MUG (Genbank accession number P0A9H1.1), *Geobacter metallireducens* SMUG1 (Genbank accession number YP_383069.1), *Thermus thermophiles* UDGa (Genbank accession number WP_011228142.1), UDGb (Genbank accession number WP_011173217.1), *Methanosarcina barkeri* HDG (Genbank accession number WP_011305765.1), *Mycobacterium smegmatis* UDGX (GenBank accession number: WP_011726794.11) were used as a query.

C. Cloning, expression and purification of UDGX

The UDGX family uracil-DNA binding protein gene (*ugdx*) from *Mycobacterium smegmatis* str. MC2 155 (GenBank accession number: CP009494.1) was amplified by PCR using the forward primer 5'-TGCATATGGCGGGTGCGCAAGAT-3' (Nde I) and the reverse primer 5'- CAAGCTTGCAGATGGGCTCCATC-3' (Hind III). The PCR reaction mixture (50 μL) consisted of 25 ng genomic DNA template, 100 nM forward and reverse primers, 1x Phusion DNA polymerase buffer, 200 μM each dNTP, and 0.5 Unit of Phusion DNA polymerase (ThermoFisher Scientific). The PCR procedure included a pre-denaturation step at 98°C for 1 min; 35 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 10 s, annealing at 60°C for 20 s, and extension at 72°C for 90 s; and a final extension step at 72°C for 7 min. The PCR product was purified and digested with Nde I and Hind III. After digestion, the PCR product was ligated with

Nde I/Hind III digested pET28b or pET21a vector with T4 DNA ligase. The ligation mixture was then transformed into *E. coli* strain DH5α competent cells. The recombinant plasmid was finally verified by DNA sequencing with T7 promoter and T7 terminator primers.

PCR-based site-directed mutagenesis was performed using the pET28b-udgx or pET21a-udgx recombinant plasmid as the template as previously described [38]. The PCR reaction mixture (40 μL) consisted of 20 ng DNA template, 50 nM of each primer pair carrying the desired mutations (Table 2.1), 200 μM each dNTP, 1x Phusion DNA polymerase buffer and 1 Unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation step at 98°C for 2 min; 25 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 10 s, annealing at 65°C for 20 s, and extension at 72°C for 7.5 min; and a final extension step at 72°C for 5 min. The PCR product was then treated with Dpn I and transformed into *E. coli* strain DH5α competent cells. Successfully mutated plasmid was confirmed by DNA sequencing. The double mutants of UDGX were generated by two rounds of PCR-based site-directed mutagenesis. Taking the mutant Q53A-H109A as an example: the mutant Q53A was first generated by PCR using the pET28b-udgx recombinant plasmid as the template and the primer pair carrying Q53A mutation; then using the pET28b-udgx-Q53A plasmid as the template, the double mutant Q53A-H109A was generated by PCR with the primer pair carrying H109A mutation.

Mutant	Primers
E52A	5'-GTCACCGGGCTGCGCGCCGATCATCAT-3' 5'-ATGATGATCGGCGCGCAGCCCGGTGAC-3'
Q53A	5'-CTCTTTGTCACCGGGCGCCTCGCCGATCATCATG-3' 5'-CATGATGATCGGCGAGGCGCCCGGTGACAAAGAG-3'
D56A	5'-CAGGTCCTCTTTGGCACCGGGCTGCTC-3' 5'-GAGCAGCCCGGTGCCAAAGAGGACCTG-3'
D59A	5'-CAGGCCCGCCAGGGCCTCTTTGTCACC-3' 5'-GGTGACAAAGAGGCCCTGGCGGGCCTG-3'
F65A	5'-GCGGGACCCACGGCGGGCAGGCCCGC-3' 5'-GCGGGCCTGCCCGCCGTGGGTCCCGC-3'
N91A	5'-AAGTGCTTGACCGCGGCGGTGACGTAGAGGGC-3' 5'-GCCCTCTACGTCACCGCCGCGGTCAAGCACTT-3'
K94A	5'-CGTGAACTTGAAGTGCGCGACCGCGTTGGTGACG-3' 5'-CGTCACCAACGCGGTCGCGCACTTCAAGTTCACG-3'
R107A	5'-GGGTCTTGTGGATGGCTCGTTTGCCTCCCG-3' 5'-CGGGAGGCAAACGAGCCATCCACAAGACCC-3'
H109A	5'-CGACTGGGGGTCTTGGCGATGCGTCGTTTGCC-3' 5'-GGCAAACGACGCATCGCCAAGACCCCCAGTCG-3'
H178A	5'-GAGACGACGGGGCAACGGTGGCGACCAG-3' 5'-CTGGTCGCCACCGTTGCCCCGTCGTCTC-3'

Table 2.1 Primers for mutagenesis.

To express the wild-type and mutants of UDGX, the recombinant plasmids were transformed into *E. coli* strain Rosetta competent cells. Culture incubation, IPTG-induced protein expression and purification were performed as described below. A single *E. coli* colony transformed with recombinant plasmid was cultured in 500 mL LB medium supplemented with 50 μg/mL Kanamycin or 100 μg/mL Ampicillin at 37°C with shaking at 250 rpm until the optical density at 600 nm was over 0.6. After adding IPTG to a final concentration of 0.8 mM, the culture was grown at 22°C overnight. The *E. coli* cells were

harvested at 5000 rpm with a JLA8.1000 rotor (Beckman Coulter) at 4°C for 20 min. The cell pellet was resuspended in 7 mL sonication buffer and then sonicated at 5 s on/ 5 s off cycle for 10 min on ice using Qsonica model Q125. The sonicated solution was then centrifuged at 12000 rpm with a JLA16.250 rotor (Beckman Coulter) at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded onto a 1 mL HisTrap FF column. After the sample loading, the column was washed with 15 mL of buffer A. The bound protein in the column was eluted with a linear gradient of 0-100% buffer B and collected in 1 mL fractions. The fractions identified by UV280 and SDS-PAGE were pooled and diluted 5-fold with 50 mM Tris-HCl (pH 7.5) buffer. The diluted solution was loaded onto a 1 mL HiTrap Q FF or HiTrap SP FF column. After the column was washed with 15 mL of buffer A, the bound protein in the column was eluted with a linear gradient of 10-100% buffer B and collected in 1 mL fractions. After identification, the fractions containing wild-type or mutants of UDGX were pooled and concentrated. The protein concentration was determined by Nanodrop One (ThermoFisher Scientific). The protein solution was stored at -20°C in 50% glycerol.

D. Oligodeoxynucleotide substrates

Oligodeoxynucleotide substrates containing uracil were prepared as previously described [16]. Carboxyfluorescein (FAM) fluorescence-labelled single-stranded oligodeoxynucleotides containing deoxyuridine (U) and 1.5-fold molar excess complementary single-stranded oligodeoxynucleotides were mixed and incubated at 85°C for 3 min, followed by annealing to form duplex DNA substrates at room temperature for more than 30 min.

An apurinic/apyrimidinic site (AP site) containing single-stranded oligodeoxynucleotide substrate was produced by incubating 5 μM of uracil-containing single-stranded oligodeoxynucleotides with 1 unit of Cod UNG at 37°C for 30 min. After incubation, the Cod UNG was inactivated completely and irreversibly by heating at 55°C for 20 min [39].

E. DNA crosslinking assay

DNA crosslinking assays for wild-type and mutants of UDGX were performed at 37°C for 30 min in 20 μL reaction mixtures containing 100 nM uracil-containing DNA substrate, 100 nM ([E]:[S]=1:1) or 10000 nM enzyme [E]:[S]=100:1), 25 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA and 1 mM DTT. The mixtures were heated at 95°C for 5 min after adding 5 μL SDS-PAGE loading buffer, followed by loading into 15% SDS-PAGE gel. Electrophoresis was conducted at 300 V for 30 min. Fluorescence labelled oligodeoxynucleotides in the gel were visualized by a Typhoon FLA 7000 imager (GE Healthcare).

F. Uracil excision assay

Uracil excision assays for UDGX-H109A, E52A-H109A and Q53A-H109A mutants were performed at 37°C for 30 min in 10 μL reaction mixtures containing 10 nM oligodeoxynucleotide substrate, 1000 nM enzyme, 25 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA and 1 mM DTT. The resulting abasic sites were cleaved by heating at 95°C for 5 min after adding 1 μL of 1 M NaOH to stop the reaction. The mixtures (2 μL) were then mixed with 7.8 μL Hi-Di formamide and 0.2 μL GeneScan 500 LIZ dye size standard and

analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

G. Enzyme kinetics analysis

Excess enzymes (range from 200 nM to 2500 nM) of wild-type and mutants of UDGX were incubated with 20 nM of A/U or G/U base pair containing double-stranded DNA substrates, 25 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA and 1 mM DTT at 37°C in 10 μL reaction mixtures. Samples were collected at 10 s, 30 s, 1 min, 1.5 min, 2.5 min, 5 min, 7.5 min, 10 min and 15 min for the wild-type UDGX; at 5 min, 10 min, 15 min, 20 min, 30 min, 40 min, 55 min, 70 min and 90 min for mutants by adding 10 μL of 100 mM NaOH to terminate the reactions. After heating at 95°C for 5 min, the samples were supplemented with 5 μL of 50% glycerol loading buffer and electrophoresed at 130 V for 100 min on a 15% urea denaturing polyacrylamide gel in 0.5 x TB buffer (44.6 mM Tris base and 44.6 mM boric acid) supplemented with 2.5 mM EDTA. The intensities of the fluorescence signals of the crosslinked product and free DNA species were quantified using a Typhoon 7000 FLA imager and ImageQuant TL software (GE Healthcare).

For H109A and Q53A-H109A mutants, excess enzymes ranging from 200 nM to 2000 nM were incubated with A/U or G/U DNA substrates at 37°C in 5 μL reaction mixtures supplemented with 25 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA and 1 mM DTT. Samples were collected at 5 min, 10 min, 15 min, 20 min, 30 min, 40 min, 55 min and 70 min by adding 5 μL of 100 mM NaOH to terminate the reactions. After incubation at 95 °C for 5 min, 2 μL of reaction mixtures were mixed with 7.8 μl Hi-Di formamide and 0.2 μl GeneScan 500 LIZ dye size standard. Samples were then analyzed by ABI 3130xl with

a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

The apparent rate constants (k_{obs}) for each concentration of the wild-type and mutants of UDGX were determined by nonlinear fitting using the integrated first-order rate equation (1):

$$P=P_{max}(1-e^{-k_{obs}t}) \quad (1)$$

where P is the product yield, P_{max} is the maximal yield, t is time and k_{obs} is the apparent rate constant.

The kinetic parameters k_2 and K_m were obtained from plots of k_{obs} against the total enzyme concentration ($[E_0]$) using a standard hyperbolic kinetic expression with the program GraphPad Prism 9 following the equation (2) (Figure 2.1B and D) [24, 40, 41]:

$$k_{obs} = \frac{k_2[E_0]}{K_m+[E_0]} \quad (2)$$

For the wild-type and some mutants, because of a large K_m, in which $K_m \gg [E_0]$, the kinetic parameter k_2/K_m values were obtained from plots of k_{obs} against the total enzyme concentration ($[E_0]$) using a linear regression with the program GraphPad Prism 9 following the equation (3) (Figure 2.1A and C) [24, 40, 41]:

$$k_{obs} = \frac{k_2[E_0]}{K_m} \quad (3)$$

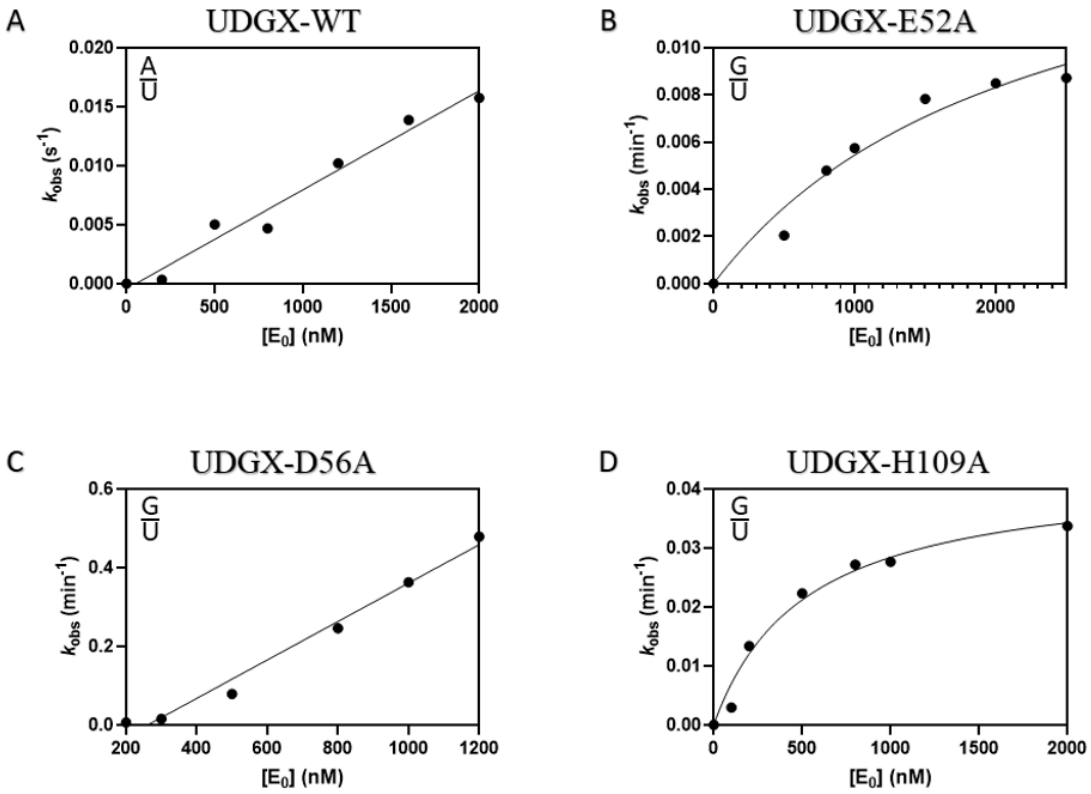

Figure 2.1 Representative kinetics analysis of the wild-type and mutant UDGX. See Enzyme kinetics analysis in Materials and Methods for details. **A.** UDGX wild-type on A/U base pair DNA substrate. **B.** UDGX-E52A mutant on G/U base pair DNA substrate. **C.** UDGX-D56A mutant on G/U base pair DNA substrate. **D.** UDGX-H109A mutant on G/U base pair DNA substrate.

IV. Results

UDGX was first discovered in *Mycobacterium smegmatis* [29]. We searched sequenced genomes for the existence of *udgx* genes in other species. So far, the *udgx* genes were only detected in bacteria, no *udgx* genes were found in archaea or eukaryotic organisms (Figure 2.2). To gain a better understanding of the distribution of *udgx* genes in bacteria, we superimposed *udgx* genes to a bacterial tree of life. Within the 81 bacterial species shown in Figure 2.2, we found limited and scattered distribution in a variety of

bacteria ranging from *Mycobacterium smegmatis* in actinobacteria to *Erythrobacter litoralis* in alphaproteobacteria to *Caballeronia arationis* in betaproteobacteria. Interestingly, some genomes contain more than one *udgx* genes within a family. For example, *Rhizobium leguminosarum* contains three *udgx* genes in its genome and there is a cluster of betaproteobacteria with multiple *udgx* genes in their genomes (Figure 2.2). UDGX from *Mycobacterium smegmatis* has served as a prototype to elucidate the unique biochemistry and catalytic mechanism of UDGX enzymes.

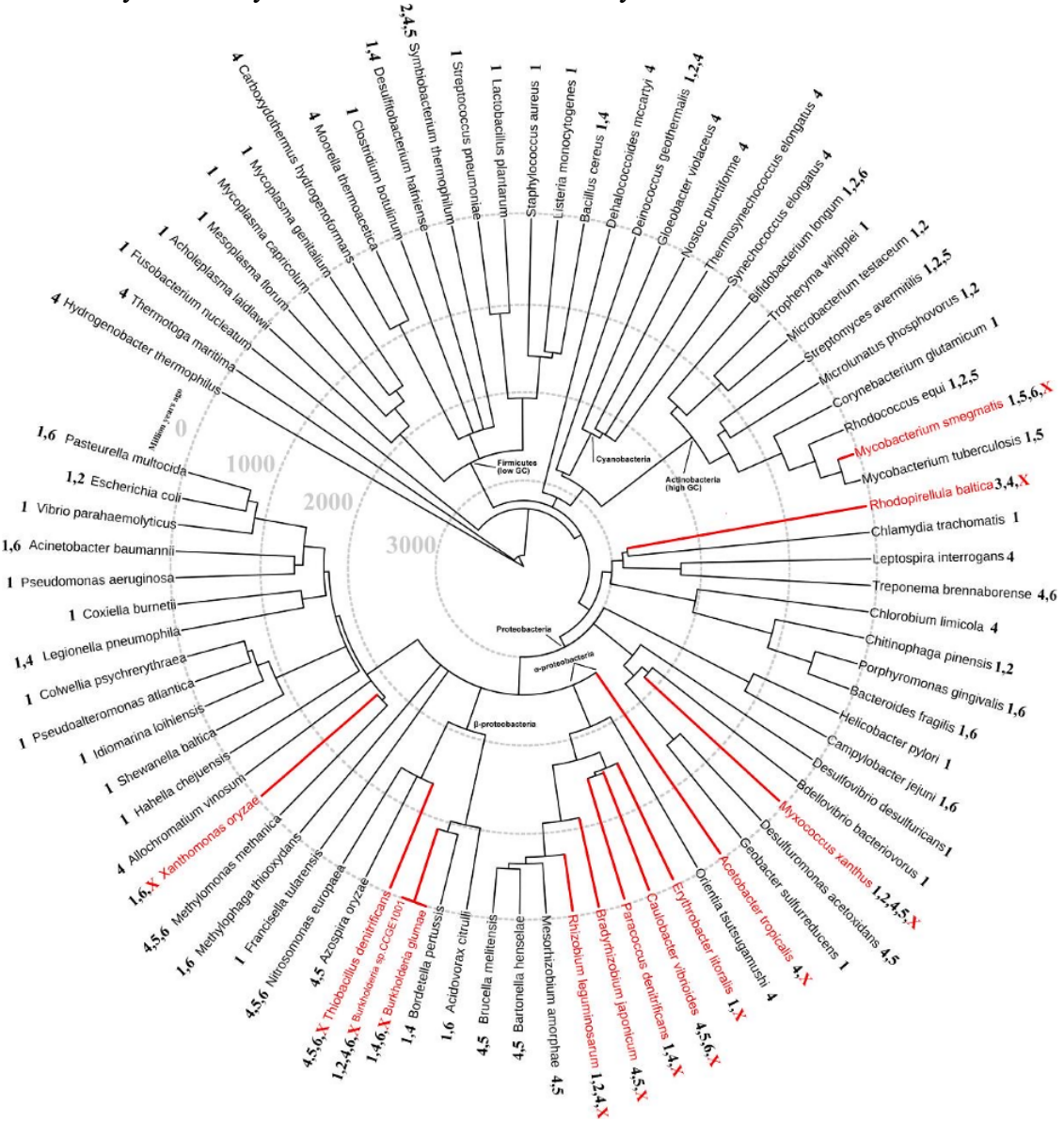

Figure 2.2 Overview of UDG superfamily in selected bacterial genomes. The species tree was obtained from TimeTree [35] and visualized by iTOL [36]. Numbers after species names mean different UDG superfamily genes in bacterial genomes, 1, family 1 UNG; 2, family 2 TDG/MUG; 3, family 3 SMUG1; 4, family 4 UDGa; 5, family 5 UDGb; 6, family 6 HDG; and X, UDGX. Multiple genes within a family are shown as a superscript.

A general reaction scheme of UDGX is shown in Figure 2.3A. Like other UDG enzymes, UDGX searches for uracil in DNA and forms an ES complex and then removes the uracil base from DNA and leaves an AP site (Figure 2.3A, uracil excision step). Unlike conventional UDG enzymes, histidine 109 of motif 3 in UDGX forms a covalently crosslinked product with the remaining AP site (Figure 2.3A, crosslinking step). A representative SDS-PAGE analysis of UDGX-catalyzed reaction is shown in Figure 2.3C. Under the assay conditions in which the enzyme:substrate (E:S) ratio was 1:1, we did not detect any crosslinked product. When the E:S ratio was 100:1, we observed complete conversion of substrate to the crosslinked product for all basepairs (A/U, C/U, T/U, G/U) and single-stranded uracil-containing DNA (Figure 2.3B and C). A time-course analysis of the reactions with all five substrates (E:S = 50:1) is shown in Figure 2.3D. Within 60 seconds, over eighty percent of the reactions were completed (Figure 2.3D, inset) and within 180 seconds, all the substrates were converted to crosslinked products (Figure 2.3D). The reaction with the A/U base pair was slower than other substrates, consistent with the notion that A/U forms a normal Watson-Crick base pair while C/U, T/U and G/U form mismatch base pairs. To further quantitatively characterize the kinetic behavior of the UDGX enzyme, we measured the kinetic parameters using a previously established approach [24, 40, 41]. Because the K_m was too large, we were unable to determine the K_m values alone (Figure 2.1), instead, we obtained k_2/K_m values for G/U and A/U basepairs

(Table 2.2). Consistent with the time-course analysis shown in Figure 2.3D, the reaction with A/U was two-fold slower than with G/U as judged by the k_2/K_m values.

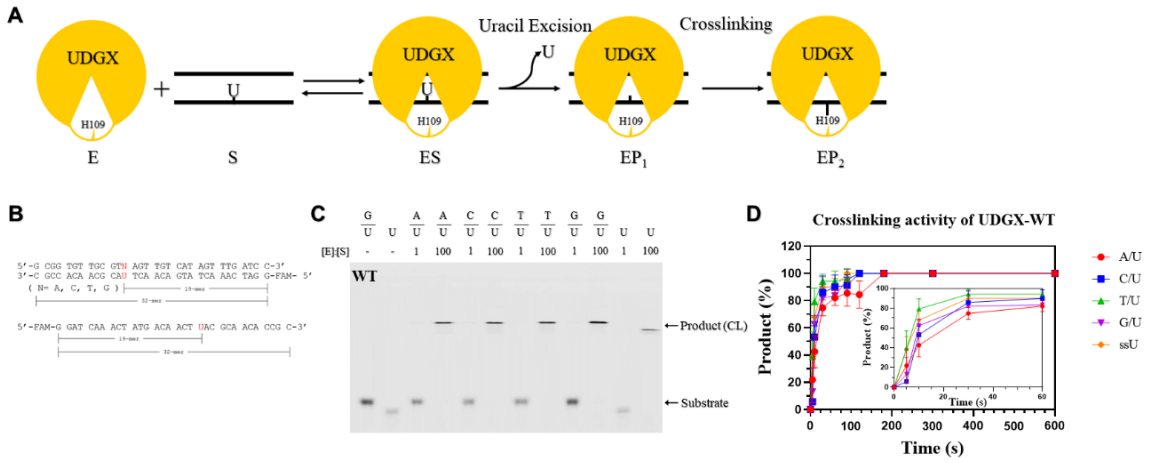

Figure 2.3 Catalytic scheme and biochemical analyses of UDGX. A. Catalytic scheme of UDGX on U-containing DNA substrate. E: Enzyme; S: Substrate; ES: Enzyme-Substrate complex; EP_1: Complex of enzyme and uracil excised DNA; EP_2: DNA-UDGX crosslinking complex. **B.** Sequences of uracil-containing DNA substrates (A/U, C/U, T/U and G/U base pair containing double-stranded DNA and single-stranded U DNA). **C.** Crosslinking analysis of wild-type UDGX. DNA crosslinking assays were performed as described in Materials and Methods. Product (CL) indicates the DNA-UDGX crosslinking product. **D.** Time course analysis of DNA crosslinking activity of wild-type UDGX on U-containing DNA substrates. (●) A/U; (■) C/U; (▲) T/U; (▼) G/U; (◆) single-stranded U (ssU). The assays were performed as described in Material and methods under Enzyme kinetics analysis with modification. Five types of uracil DNA substrates (20 nM) were incubated with 1000 nM wild-type UDGX enzyme, and samples were collected at 5 s, 10 s, 30 s, 1 min, 1.5 min, 2 min, 3 min, 5 min, 10 min. Data are shown as the average of three independent experiments and the standard deviation for each point is shown.

Enzyme	G/U			A/U		
	K_m (M)	k_2 (s^{-1})	k_2/K_m ($M^{-1}s^{-1}$)	K_m (M)	k_2 (s^{-1})	k_2/K_m ($M^{-1}s^{-1}$)
Wild-Type	N.D.[b]	N.D.	$1.4\ (0.09) \times 10^4$	N.D.	N.D.	$7.4\ (1.8) \times 10^3$
E52A	$2.2\ (0.4) \times 10^{-6}$	$2.8\ (0.1) \times 10^{-4}$	$1.3\ (0.2) \times 10^2$	$1.1\ (0.2) \times 10^{-6}$	$1.6\ (0.2) \times 10^{-4}$	$1.5\ (0.1) \times 10^2$
Q53A	N.D.	N.D.	$8.4\ (1.2) \times 10^3$	N.D.	N.D.	$8.1\ (1.4) \times 10^3$
D56A	N.D.	N.D.	$6.0\ (2.0) \times 10^3$	$2.4\ (1.2) \times 10^{-6}$	$4.5\ (1.3) \times 10^{-3}$	$2.0\ (0.3) \times 10^3$
D59A	N.A.[c]	N.A.	N.A.	N.A.	N.A.	N.A.
F65A	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
N91A	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
K94A	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
R107A	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
H109A[d]	$3.9\ (1.0) \times 10^{-7}$	$6.4\ (0.7) \times 10^{-4}$	$1.7\ (0.3) \times 10^3$	$2.3\ (0.9) \times 10^{-7}$	$1.8\ (0.04) \times 10^{-4}$	$8.8\ (3.6) \times 10^2$
H178A	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Q53A-H109A[d]	N.D.	N.D.	$2.6\ (0.3) \times 10^3$	N.D.	N.D.	$1.5\ (0.1) \times 10^3$
E52A-H109A	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Table 2.2 Kinetic parameters of UDGX with G/U and A/U base pairs[a]

[a]: Enzyme kinetics analysis was performed as described in Materials and Methods. For the Q53A-H109A mutant on A/U DNA substrate, samples were collected at 10 min, 20 min, 30 min, 40 min, 55 min, 70 min, 85 min and 100 min due to the slow reaction rate. Data are shown as the average of three independent experiments. S.D. values are shown in parentheses.

[b]: N.D. Not determined due to large K_m.

[c]: N.A. No activity detected under assay condition.

[d]: For H109A and Q53A-H109A mutants, k_2 values reflect uracil excision only, whereas k_2 values for wild-type and the other mutants reflect uracil excision and subsequent crosslinking process together.

A comparison of motifs 1, 2 and 3 between representative UDGX enzymes with UDGs from other families is shown in Figure 2.4. UDGX is most closely related to family 4 UDGa. For example, Msm UDGX shares 37.20% sequence identity with *Thermus thermophilus* UDGa. Some sequence conservation within the motifs is also evident, as most UDGX enzymes contain GEQP, N and HP at the start of each motif while most of family 4 UDGa enzymes contain GEGP, N and HP, respectively (Figure 2.4). Despite of these similarities, notable differences exist between UDGX and family 4 UDGa, especially in motif 1 and motif 3. In addition to the extra R-loop at the end of motif 3, UDGX has two conserved negatively charged aspartate residues (D56, D59) in motif 1 and a conserved

positively charged residue (R107) in motif 3, which are missing in other UDG families (Figure 2.4). Another residue, K94, is also highly conserved in UDGX enzymes, which is not observed in other UDG families except for family 4 UDGa.

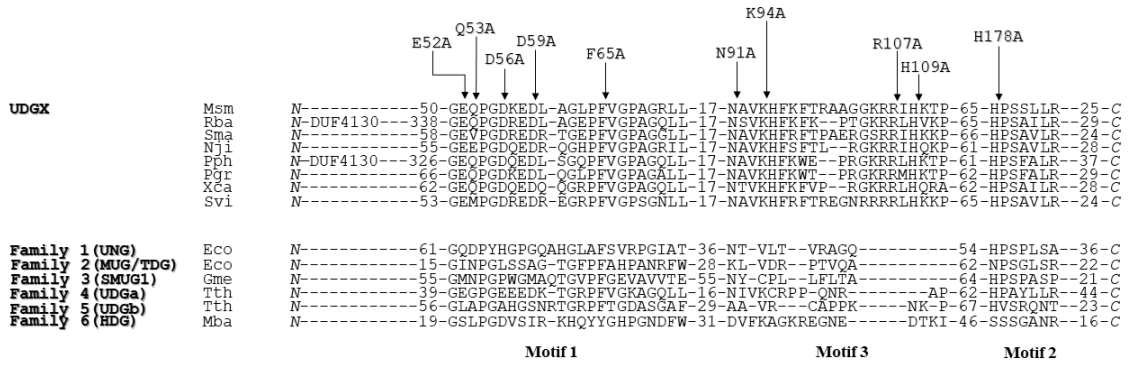

Figure 2.4 Sequence alignment of UDG enzymes and mutagenesis sites. GenBank accession numbers are shown after the species names. UDGX: Msm, *Mycobacterium smegmatis*, WP_011726794.1; Rba, *Rhodopirellula baltica*, WP_011119624.1; Sma, *Saccharomonospora marina*, WP_009153389.1; Nji, *Nonomuraea jiangxiensis*, WP_090946880.1; Pph, *Paraburkholderia phytofirmans*, WP_012433335.1; Pgr, *Paraburkholderia graminis* C4D1M, EDT13144.1; Xca, *Xanthomonas campestris*, WP_011038956.1; Svi, *Saccharomonospora viridis*, WP_015785655.1. Family 1 (UNG): Eco, *Escherichia coli* O157:H7 str. EDL933, NP_289138.1; Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1.1; Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15, YP_383069.1; Family 4 (UDGa): Tth, *Thermus thermophilus*, WP_011228142.1; Family 5 (UDGb): Tth, *Thermus thermophilus*, WP_011173217.1; Family 6 (HDG): Mba, *Methanosarcina barkeri*, WP_011305765.1.

Based on the sequence alignment and previous structural studies, we chose ten sites among the three motifs for mutational studies and constructed ten single mutants (Figure 2.4). As with the wild-type enzyme, the mutant proteins were subjected to DNA crosslinking assay first to test their ability to form a covalent bond after uracil excision. At E52 and Q53 positions, we constructed E52A and Q53A mutants. As shown in Figure 2.5, E52A and Q53A still retained their ability to crosslink to the AP site after uracil excision with all five substrates (A/U, C/U, T/U, G/U and single-stranded U). The time-course analyses indicated that both E52A and Q53A mutants, especially the former, were less

efficient in forming the crosslinked product in comparison with the wild-type UDGX enzyme (Figure 2.3 and Figure 2.5). To more quantitatively define the kinetic properties, we measured the K_m, k_2 and k_2/K_m values. E52A showed a K_m value in the micromolar range for both G/U and A/U base pairs and k_2 of 2.8 x 10^{-4}/sec for G/U and 1.6 x 10^{-4}/sec for A/U (Table 2.2). Judging by the k_2/K_m values, E52A was two-orders of magnitude slower than the wild-type enzyme for the G/U base pair and one-order of magnitude for the A/U base pair (Table 2.2). The kinetic behavior of Q53A was similar to the wild-type enzyme but the catalytic efficiency was about two-fold lower for the G/U and comparable for the A/U (Table 2.2).

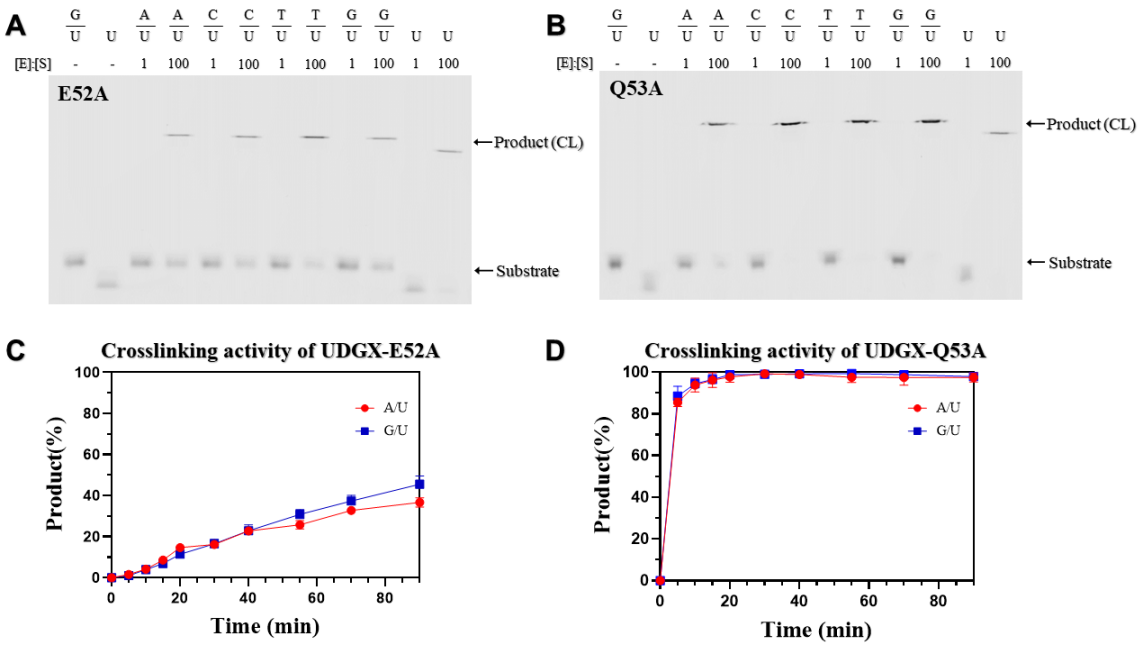

Figure 2.5 DNA crosslinking analysis of UDGX-E52A and Q53A mutants. DNA crosslinking assays were performed as described in Material and methods. Product (CL) indicates the DNA-UDGX crosslinking product. Time course analysis assays were performed as described in Material and methods under Enzyme kinetics analysis with modification. Uracil DNA substrates were incubated with 1000 nM of UDGX-E52A or Q53A mutant. Data are shown as the average of three independent experiments and the standard deviation for each point is shown. **A.** Crosslinking analysis of UDGX-E52A mutant. **B.** Crosslinking analysis of UDGX-Q53A mutant. **C.** Time

course analysis of DNA crosslinking activity of UDGX-E52A mutant. (●) G/U; (■) A/U. **D.** Time course analysis of DNA crosslinking activity of UDGX-Q53A mutant. (●) G/U; (■) A/U.

F65 position is highly conserved in UDG enzymes being either phenylalanine or tyrosine. In UDG structures, the aromatic sidechain from Phe or Tyr stacks on the uracil base. Consistent with previous studies, F65A lost its catalytic activity on all substrates tested due to the loss of the aromatic ring in the amino acid sidechain (Figure 2.6A). Mutations at N91 position (N91A) and K94 position (K94A) in motif 3 inactivated the enzyme, resulting in loss of crosslinking activity (Figure 2.6B and C). Uracil excision assays also confirmed the loss of uracil excision activity of these mutants. The first residue in motif 2, H178, is known as an important catalytic residue in some families of UDG enzymes. As expected, H178A mutant completely lost its catalytic activity (Figure 2.6D).

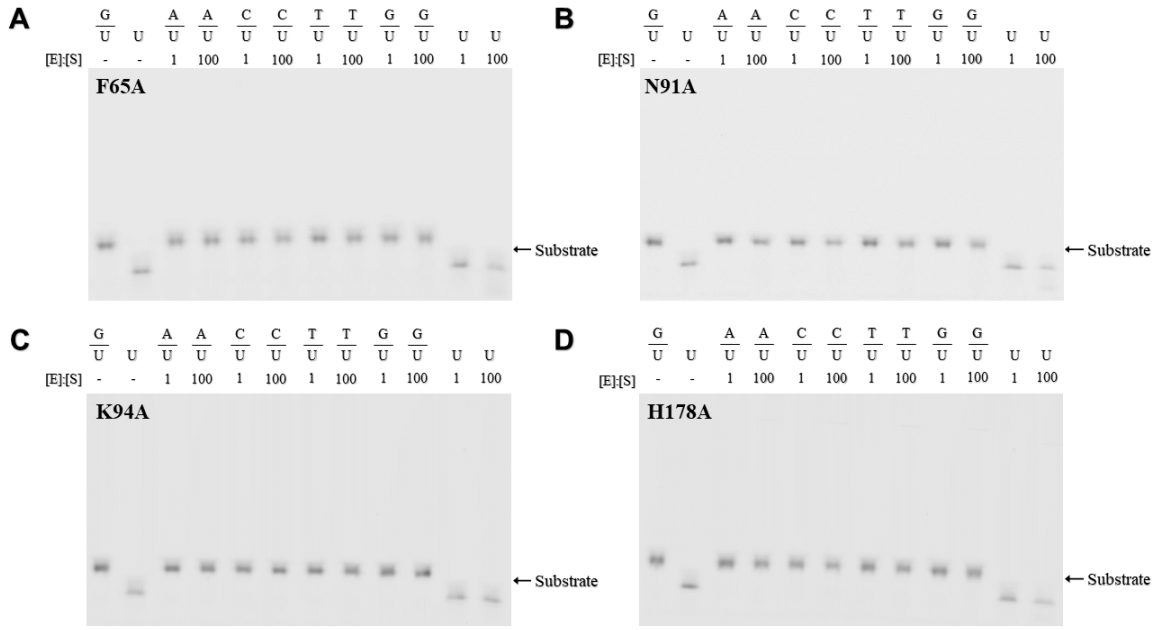

Figure 2.6 DNA crosslinking analysis of UDGX-F65A, N91A, K94A and H178A mutants. DNA crosslinking assays were performed as described in Materials and Methods. **A.** UDGX-F65A mutant. **B.** UDGX-N91A mutant. **C.** UDGX-K94A mutant. **D.** UDGX-H178A mutant.

Previous structural analysis indicated the possibility of salt bridges among D56 and D59 of motif 1 and R107 of motif 3 in UDGX [42]. To definitively determine the interactions among the three residues, we investigated the catalytic activities of D56A, D59A and R107A mutants. D56A exhibited crosslinking activity with all five substrates (A/U, C/U, T/U. G/U and single-stranded U) (Figure 2.7A). Judging by the k_2/K_m values, D56A is two-fold slower for the G/U basepair and around four-fold slower for the A/U base pair (Table 2.2). D59A and R107A completely lost their catalytic activities (Figure 2.7B and C, Table 2.2), suggesting that D59 and R107 play an essential role in establishing the interactions between motif 1 and motif 3. Structurally, the guanidino sidechain of R107 of motif 3 forms bidentate salt bridges with the carboxyl sidechain of D59 of motif 1 within 3 Å distance and interacts with D56 through a hydrogen bond with the mainchain amide group and the carboxyl sidechain through a salt bridge (Figure 2.7D).

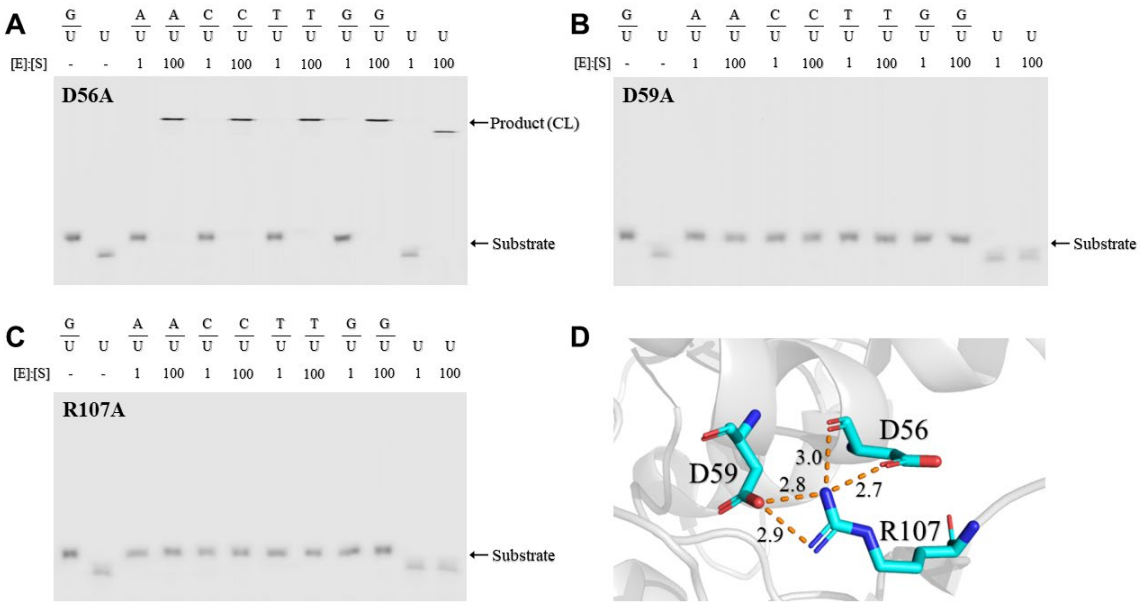

Figure 2.7 Tripartite interactions by D56, D59 and R107. DNA crosslinking assays were performed as described in Materials and Methods. **A.** Crosslinking analysis of UDGX-D56A

mutant. Product (CL) indicates the DNA-UDGX crosslinking product. **B.** Crosslinking analysis of UDGX-D59A mutant. **C.** Crosslinking analysis of UDGX-R107A mutant. **D.** Close-up view of interactions by D56, D59 and R107 in crystal structure (PDB code 6IO9). Distances are shown in Å.

H109 of motif 3 has been identified as the site of crosslinking and substitution at this position renders the enzyme inactive in crosslinking [29, 33, 42, 43]. Consistent with a previous report [29, 43], though H109A did not exhibit crosslinking activity in SDS-PAGE analysis (Figure 2.8A), it still showed uracil excision activity as illustrated by urea-denaturing electrophoresis analysis on all five substrates tested (A/U, C/U, T/U, G/U, single-stranded U) (Figure 2.8B). A time-course analysis indicated that the uracil excision activity was much lower than the wild-type UDGX with the activity on A/U as the lowest (Figure 2.8C). As judged by the k_2/K_m values, the uracil excision by H109A was one order of magnitude slower than the crosslinking by the wild-type UDGX for the G/U and A/U base pairs, respectively (Table 2.2). When H109A mutation was coupled with E52A mutation, the double mutant, E52A-H109A totally lost its catalytic activity (Table 2.2), which is consistent with a previous study showing significantly decreased activity of a E52N-H109S double mutant [43].

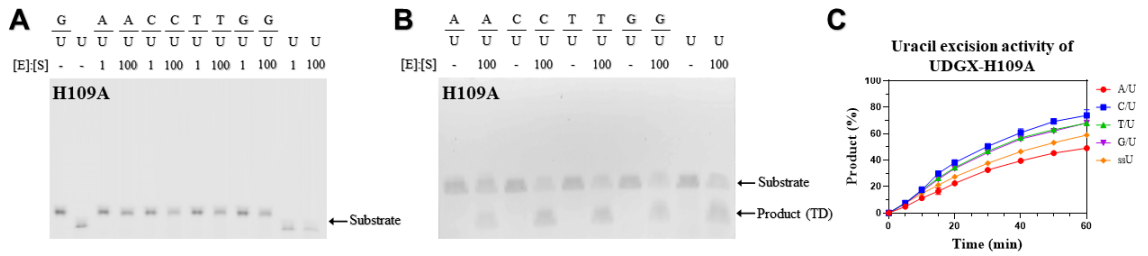

Figure 2.8 DNA crosslinking and uracil excision analyses of UDGX-H109A mutant. A. DNA crosslinking analysis of UDGX-H109A mutant. Assays were performed as described in Material and methods. **B.** Uracil excision analysis of UDGX-H109A mutant. Assays were performed as described in Material and methods but visualized by electrophoresis on a 15% urea denaturing polyacrylamide gel. Product (TD) indicates the truncated DNA product after NaOH/heat treatment. **C.** Time course analysis of uracil excision activity of UDGX-H109A mutant. (●) A/U; (■) C/U;

(▲) T/U; (▼) G/U; (◆) single-stranded U (ssU). The assays were performed as described in Material and methods under Enzyme kinetics analysis with modification. Uracil DNA substrates were incubated with 1000 nM UDGX-H109A enzyme. Samples were collected at 5 min, 10 min, 15 min, 20 min, 30 min, 40 min, 50 min and 60 min. Data are shown as the average of three independent experiments and the standard deviation for each point is shown.

A hallmark of UDGX is its ability to crosslink to the AP site generated after uracil excision. To understand how UDGX reacts with an AP site, we prepared an AP site substrate after removal of uracil by a conventional UDG enzyme. As shown in Figure 2.9, UDGX was able to convert a single-stranded uracil-containing DNA substrate to the crosslinked product (Figure 2.9A). On the contrary, UDGX was incapable of converting the AP site substrate to a crosslinked product even at the highest enzyme concentration tested (Figure 2.9B). Furthermore, no truncated DNA fragment (as a result of forming AP site) was detected with the A/U (Figure 2.9C) or the G/U (Figure 2.9D) substrates. The implication of this finding will be discussed in detail later.

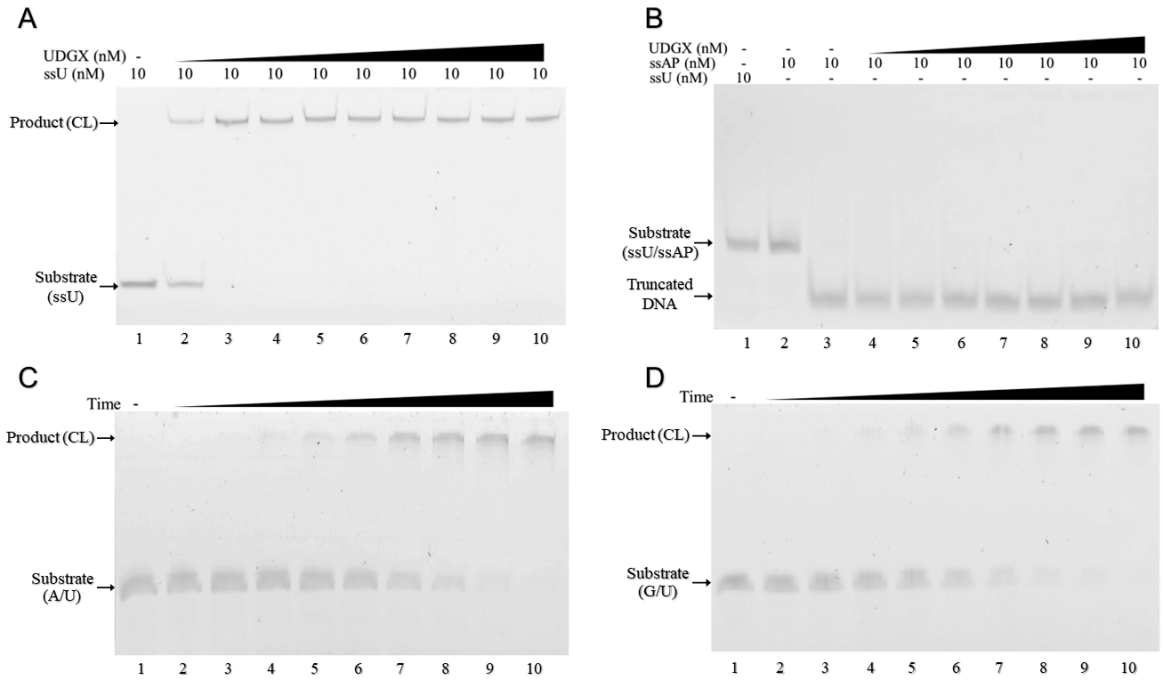

Figure 2.9 Uracil excision and AP site crosslinking coupling of UDGX. A. Crosslinking analysis of UDGX on single-stranded U DNA substrate (ssU). The assays were performed as described in Material and methods under DNA crosslinking assay with following modifications. 10 nM of ssU DNA substrates were incubated with increasing concentration (ranging from 10 to 1000 nM) of UDGX at 37°C for 60 min. Samples were electrophoresed for 85 min on a 15% urea denaturing polyacrylamide gel and visualized by fluorescence scanning. Product (CL) indicates the DNA-UDGX crosslinking product. **B.** Crosslinking analysis of UDGX on single-stranded AP site DNA substrate (ssAP). The assays were performed similarly as described for the assays on ssU substrates in Fig. 8A. Except for samples in lane 1 and 2, samples in lane 3-10 were treated with NaOH and heat. The AP sites not crosslinked by UDGX were cleaved and the truncated DNA products are shown in the gel. **C.** Crosslinking analysis of UDGX on A/U base pair DNA substrate. The assays were performed as described in Materials and Methods under Enzyme kinetics analysis with 500 nM of UDGX. The assays were stopped by adding 10 μL of 100 mM NaOH at 10 s, 30 s, 1 min, 1.5 min, 2.5 min, 5 min, 7.5 min, 10 min and 15 min. **D.** Crosslinking analysis of UDGX on G/U base pair DNA substrate. The assays were performed as described in Materials and Methods under Enzyme kinetics analysis with 500 nM of UDGX. The assays were stopped by adding 10 μL of 100 mM NaOH at 10 s, 30 s, 1 min, 1.5 min, 2.5 min, 5 min, 7.5 min, 10 min and 15 min.

V. Discussion

Since the discovery of the first DNA glycosylase, family 1 UNG, in *E. coli* in 1974, UDG enzymes have been discovered in many bacteria, archaea and eukaryotes. Almost all living organisms possess at least one UDG gene in their genomes and many have multiples, forming a UDG superfamily with families of a variety of catalytic functions, going far beyond as a narrow specificity family 1 UNG-type DNA repair enzyme. Unlike many other enzyme superfamilies, a rather unique feature of UDG superfamily is that none of the sites is completely conserved in all UDG enzymes [44], although sequence or motif conservation is evident within UDG families. This phenomenon suggests a broad catalytic diversity in UDG superfamily. UDGX represents a class of UDG enzymes that not only possess uracil excision ability, but also acquired an unprecedented crosslinking function upon uracil removal. Taking advantage of the solved crystal structures and building on previous studies, this work presents a comprehensive biochemical and enzymological

analysis of UDGX and provides catalytic insights to understand the inner working of this bi-functional enzyme.

A. Structural coupling

Over the years, a large number of crystal structures of UDG enzymes have been reported. The solved UDGX structures reveal a similar structural fold like other UDG enzymes with known structures. The extended loop located in motif 3 contains the critical crosslinking site (H109 in Msm UDGX) that differentiates UDGX from other conventional UDG enzymes. Besides the obvious importance of the addition of the extended loop to motif 3, several structural features in UDGX are quite distinct. The role of E52 is demonstrated by E52A mutant, which reduced the catalytic activity significantly (Figure 2.5 and Table 2.2). Structurally, the mainchain of E52 interacts with O2 of uracil in the active site, facilitating the removal of uracil in a manner similar to previous observation on family 4 UDGa and other UDG enzymes [2, 4, 24]. In addition, a unique interaction between E52 and the imidazole sidechain of H109 becomes feasible as H109 in the extended loop is located nearby (Figure 2.10A), which was also reported in previous studies [33, 43]. Because the interaction is between the carboxyl sidechain of E52 and the ε^2 nitrogen of the imidazole sidechain of H109 (the nitrogen that crosslinks to the AP site), we surmise that this interaction may position H109 for the crosslinking step and may help stabilize the ε^2 nitrogen of the imidazole and C1' carbon of deoxyribose interaction during the transition state (Figure 2.11). While E52A retained partial crosslinking activity and H109A retained reduced uracil excision activity, the E52A-H109A double mutant was inactive (Table 2.2). The implication is that when H109A mutation is added to E52A

mutation, it would further comprise the catalytic function of the latter, suggesting that the uracil excision and the crosslinking are connected. This is the first case of inter-motif interaction in UDGX.

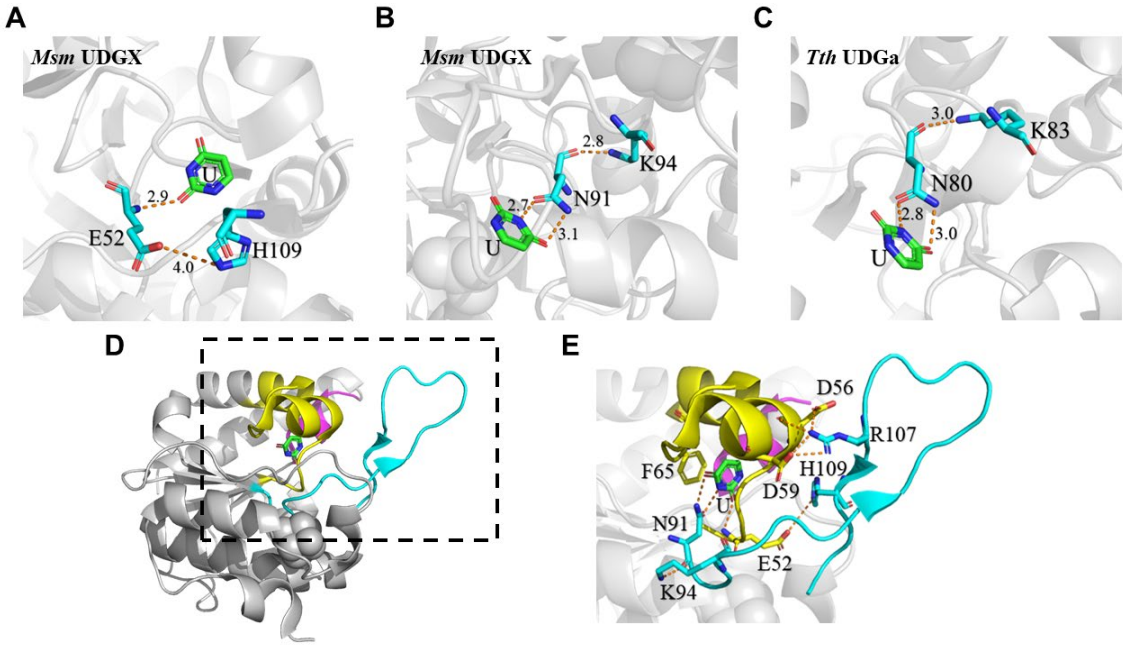

Figure 2.10 Inter-motif and intra-motif interactions in UDGX structure. A. Interaction between E52 of motif 1 and H109 of motif 3 of UDGX (PDB code 6IOA)**. B.** Interaction between N91 and K94 of motif 3 of UDGX (PDB code 6IOA). **C.** Interaction between N80 and K83 of motif 3 of Tth UDGa (PDB code 1UI0). Distances between residues were shown in Å. **D.** Relative positions of motifs 1, 2 and 3 in UDGX. Motifs 1, 2 and 3 were highlighted by yellow, magenta, and cyan respectively (PDB code 6IOA). **E.** Close-up view of relative positions of motif 1 and 3 driven by residue interactions (dashed rectangle area in **Figure 2.10D**).

The second case of the inter-motif interaction is illustrated by the D59-R107-D56 interaction between motif 1 and motif 3, in which D59 and D56 located in an α helix anchor R107 located in the extended loop (Figure 2.7D, Figure 2.10E) [42]. The kinetics analysis and available structural information allow us to define the tripartite interactions. We propose that the guanidino sidechain of R107 of motif 3 and the carboxyl sidechain of D59 of motif 1 form bidentate salt bridges, and this inter-motif interaction is further supported

by bidentate interactions between the sidechain of R107 and the mainchain and the sidechain of D56, respectively (Figure 2.7D). The essential role of the inter-motif tripartite interactions is to anchor the H109-containing extended loop to the proximity of the active site to ready it for crosslinking reaction.

In addition to the inter-motif interactions, the mutational and kinetics analyses also uncover an intra-motif hydrogen bond between the mainchain of N91 and the sidechain of K94 (Figure 2.10B). The first residue in motif 3 (N91 in UDGX) plays an important catalytic role in some UDG enzymes by forming bidentate hydrogen bonds with the uracil base. In *E. coli* family 2 MUG enzyme, conversion of K68 (equivalent to N91 in UDGX) to Asn leads to greatly enhanced glycosylase activity towards all mismatched T/U, G/U and C/U base pairs and acquisition of activity on A/U base pairs [6]. Msm UDGX is most homologous to family 4 UDGa such as Tth UDGa (Figure 2.4). Interestingly, we found potentially a similar interaction between the mainchain of N80 and the sidechain of K83 in Tth UDGa (Figure 2.10C). In UDGX, the bidentate interactions of N91 with the departing uracil base and the hydrogen bond interaction with K94 are essential for catalysis.

B. Functional coupling

The structural coupling described above provides a basis for discussing functional coupling detailed below. Unlike a conventional UDG, which will dissociate from the AP site generated by uracil excision or displaced by a downstream AP endonuclease during Base Excision Repair (BER). UDGX performs crosslinking reaction upon uracil excision. The discussion above on E52 and H109 already suggests that uracil excision and crosslinking are connected. A comparison of H109A mutant with the wild-type UDGX

offers another clue about the coupling of uracil excision and crosslinking. As shown in Figure 2.8B, H109A was able to generate the AP site product, which was hydrolyzed to smaller fragments after alkaline treatment. However, we consistently failed to observe any AP site product when an uracil-containing substrate was treated with the wild-type UDGX (Figure 2.9A, C and D). These results suggest that once an uracil base is excised, the resulting AP site intermediate is likely immediately crosslinked by UDGX, indicating a coupling of the two catalytic steps. Lastly, we let the wild-type UDGX react with the AP site substrate but could not detect any crosslinking product (Figure 2.9B). These results suggest that UDGX could not crosslink to an existing AP site without coupling to a preceding uracil excision step. For all the evidence described above, we conclude that in the UDGX-catalyzed reaction, the uracil excision and crosslinking steps are tightly coupled.

C. Catalytic mechanism

The catalytic mechanisms in various families of enzymes in UDG superfamily are extensively investigated [2, 4, 45, 46]. Key catalytic residues in motifs 1, 2 and 3 have been identified and studied. As a member of the UDG superfamily, UDGX maintains a catalytic mechanism for uracil excision in a manner similar to what is proposed for its close homolog in family 4 UDGa [24]. Here, we propose a catalytic mechanism for the UDGX-catalyzed uracil excision and crosslinking to the AP site, integrating the conventional mechanism with the structural and functional coupling information described above (Figure 2.11). Similar to family 4 UDGa, the uracil base in the active site is recognized by the bidentate hydrogen bond between O3 and N4 of uracil and the N91 of motif 3, stacked by the

aromatic ring of the F65 of motif 1. Likewise, the O2 of uracil is contacted by a mainchain interaction with E52 of motif 1 and the imidazole ring of H178 of motif 2.

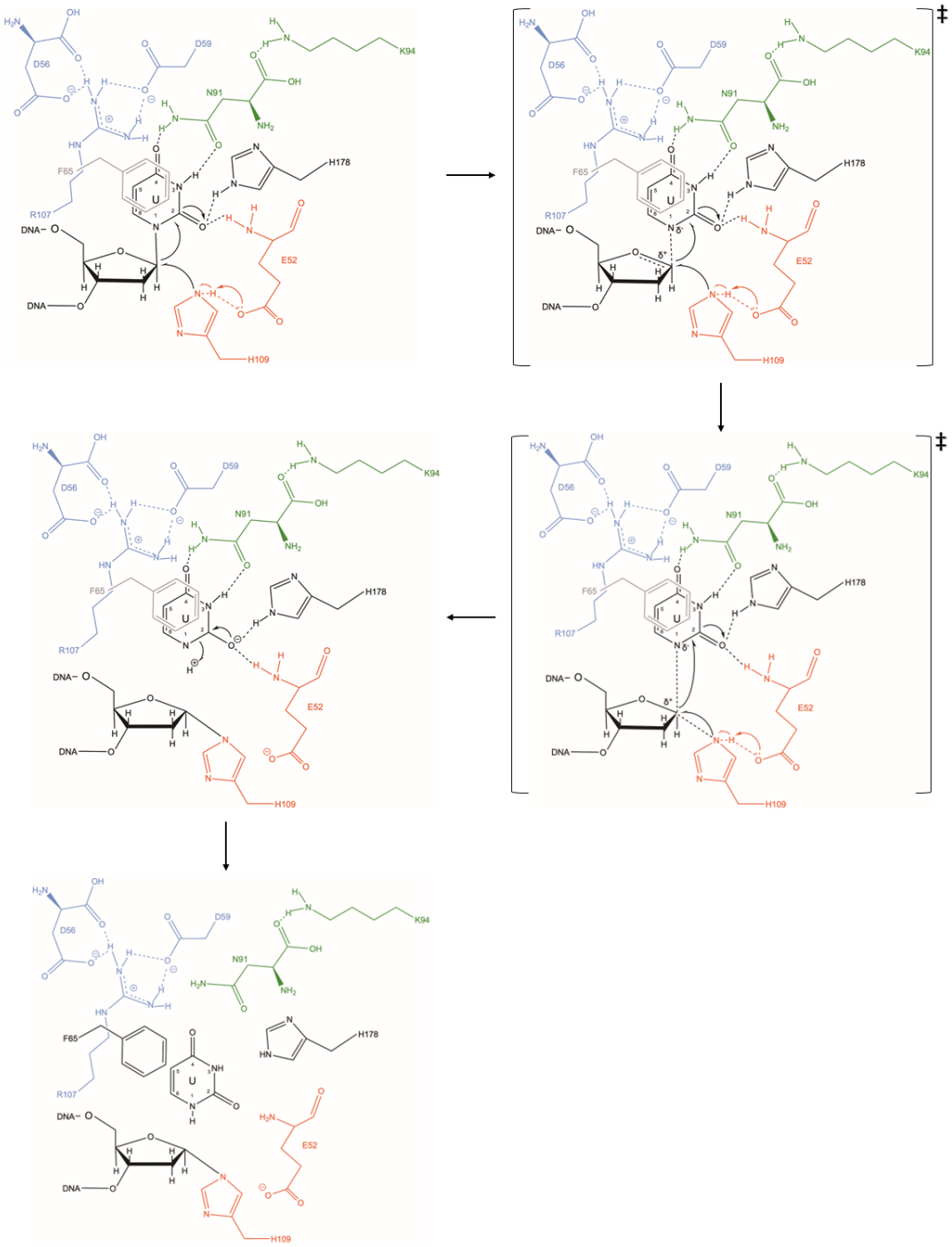

Figure 2.11 Proposed catalytic mechanism of uracil excision and AP site crosslinking by UDGX. Supported by the inter-motif interactions between E52 and H109, D56-D59-R107 tripartite interactions and intra-motif interaction between K94 and N91 (The interactions by D56, D59 and R107 are highlighted in blue; K94-N91 interaction is highlighted in green; and the E52-H109 interaction is highlighted in red), E52 and H178 interact with O2 of uracil, and N91 forms bidentate interactions with O3 and N4 of uracil, promoting the breakage of N-glycosidic bond and the departure of uracil, meanwhile, H109 initiates attack at the C1' position of the deoxyribose to form a new covalent C-N bond and to further push the departure of the uracil base, as shown in the transition states (square brackets with ‡). After the new covalent C-N bond is formed and the N-glycosidic bond is broken, the negatively charged uracil base is protonated and released.

What sets UDGX apart from family 4 UDGa and other families in the UDG superfamily is the structural coupling and consequent functional coupling. The extended loop of motif 3 is brought to the proximity of the scissile bond and crosslinking site by the E52-H109 interaction and D59-R107-D56 tripartite interactions. As a consequence, the interaction between the mainchain of E52 and O2 of uracil becomes dependent on E52-H109 interaction. Likewise, the bidentate interaction between the amide sidechain of N91 and O3 and N4 of uracil relies on the support of N91-K94 interaction. These interdependent inter-motif and intra-motif interactions lead to concerted actions in catalysis. According to the results and functional coupling discussed above, the catalytic mechanism bears similarities to previously proposed models for UDGX and other UDG enzymes but with the distinct feature of coupling the two catalytic steps [33, 42, 43, 45].

During the uracil excision step, while several interactions including E52, N91, H178 contribute to the promotion of uracil departure, the reaction is immediately coupled to the second catalytic step, in which H109 initiates the attack at C1' carbon to form the C-N bond. The breakage of N-glycosidic bond and the departure of uracil are likely promoted by E52-O2 of uracil and H178-O2 of uracil interactions that stabilize the negative charge

developed on uracil during the catalysis. Meanwhile, guided by the well knitted network of interactions, H109 is poised to initiate attack at the C1' position of the deoxyribose to form a new covalent C-N bond and in the meantime to further push the departure of the uracil base, which therefore is proposed that the breaking of N-glycosidic bond and the forming of new covalent C-N bond are completed almost simultaneously (Figure 2.11, transition states). Thus, evolution has allowed UDGX to build an elegant network of interactions to enable functional coupling to achieve concerted completion of uracil excision and AP site crosslinking in a single catalytic reaction, providing an example of the interconnectivity of motifs to perform multi-facet catalytic function.

VI. Reference

[1] T. Lindahl et al, "DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli," The Journal of Biological Chemistry, vol. 252, (10), pp. 3286-3294, 1977. DOI: 10.1016/S0021-9258(17)40386-3.

[2] S. S. Parikh, C. D. Putnam and J. A. Tainer, "Lessons learned from structural results on uracil-DNA glycosylase," Mutation Research/DNA Repair, vol. 460, (3-4), pp. 183-199, 2000. DOI: 10.1016/s0921-8777(00)00026-4.

[3] J. Li et al, "An unconventional family 1 uracil DNA glycosylase in Nitratifractor salsuginis," The FEBS Journal, vol. 284, (23), pp. 4017-4034, 2017. DOI: 10.1111/febs.14285.

[4] L. H. Pearl, "Structure and function in the uracil-DNA glycosylase superfamily," Mutation Research/DNA Repair, vol. 460, (3-4), pp. 165-181, 2000. DOI: 10.1016/s0921-8777(00)00025-2.

[5] U. Hardeland et al, "The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs," Nucleic Acids Research, vol. 31, (9), pp. 2261-2271, 2003. DOI: 10.1093/nar/gkg344.

[6] D. Lee et al, "A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs," Nucleic Acids Research, vol. 43, (2), pp. 1081-1089, 2015. DOI: 10.1093/nar/gku1332.

[7] H. Lee et al, "Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase," The Journal of Biological Chemistry, vol. 285, (53), pp. 41483-41490, 2010. DOI: 10.1074/jbc.M110.150003.

[8] L. Dong et al, "Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase," DNA Repair, vol. 7, (12), pp. 1962-1972, 2008. DOI: 10.1016/j.dnarep.2008.08.006.

[9] E. Lutsenko and A. S. Bhagwat, "The role of the Escherichia coli mug protein in the removal of uracil and 3,N(4)-ethenocytosine from DNA," The Journal of Biological Chemistry, vol. 274, (43), pp. 31034-31038, 1999. DOI: 10.1074/jbc.274.43.31034.

[10] M. Saparbaev and J. Laval, "3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase," Proceedings of the National

Academy of Sciences - PNAS, vol. 95, (15), pp. 8508-8513, 1998. DOI: 10.1073/pnas.95.15.8508.

[11] M. Saparbaev, J. C. Mani and J. Laval, "Interactions of the human, rat, Saccharomyces cerevisiae and Escherichia coli 3-methyladenine-DNA glycosylases with DNA containing dIMP residues," Nucleic Acids Research, vol. 28, (6), pp. 1332-1339, 2000. DOI: 10.1093/nar/28.6.1332.

[12] Y. He et al, "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA," Science, vol. 333, (6047), pp. 1303-1307, 2011. DOI: 10.1126/science.1210944.

[13] A. Maiti and A. C. Drohat, "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites," The Journal of Biological Chemistry, vol. 286, (41), pp. 35334-35338, 2011. DOI: 10.1074/jbc.c111.284620.

[14] K. A. Haushalter et al, "Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors," Current Biology, vol. 9, (4), pp. 174-185, 1999. DOI: 10.1016/s0960-9822(99)80087-6.

[15] B. Kavli et al, "hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup," The Journal of Biological Chemistry, vol. 277, (42), pp. 39926-39936, 2002. DOI: 10.1074/jbc.M207107200.

[16] R. Mi et al, "Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG," Journal of Molecular Biology, vol. 385, (3), pp. 761-778, 2009. DOI: 10.1016/j.jmb.2008.09.038.

[17] P. Pang et al, "SMUG2 DNA glycosylase from Pedobacter heparinus as a new subfamily of the UDG superfamily," Biochemical Journal, vol. 474, (6), pp. 923-938, 2017. DOI: 10.1042/BCJ20160934.

[18] Z. Zhang et al, "Structural basis of substrate specificity in Geobacter metallireducens SMUG1," ACS Chemical Biology, vol. 11, (6), pp. 1729-1736, 2016. DOI: 10.1021/acschembio.6b00164.

[19] M. Matsubara et al, "Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase," Nucleic Acids Research, vol. 32, (17), pp. 5291-5302, 2004. DOI: 10.1093/nar/gkh859.

[20] J. Li et al, "Identification of a prototypical single-stranded uracil DNA glycosylase from Listeria innocua," DNA Repair, vol. 57, pp. 107-115, 2017. DOI: 10.1016/j.dnarep.2017.07.001.

[21] J. Hoseki et al, "Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8," Journal of Molecular Biology, vol. 333, (3), pp. 515-526, 2003. DOI: 10.1016/j.jmb.2003.08.030.

[22] X. Liu and J. Liu, "Characterization of family IV UDG from Aeropyrum pernix and its application in hot-start PCR by family B DNA polymerase," PLoS ONE, vol. 6, (11), pp. e27248, 2011. DOI: 10.1371/journal.pone.0027248.

[23] M. Sandigursky and W. A. Franklin, "Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes," Current Biology, vol. 9, (10), pp. 531-534, 1999. DOI: 10.1016/s0960-9822(99)80237-1.

[24] B. Xia et al, "Correlated mutation in the evolution of catalysis in uracil DNA glycosylase superfamily," Scientific Reports, vol. 7, (1), pp. 45978, 2017. DOI: 10.1038/srep45978.

[25] H. Kosaka et al, "Crystal structure of family 5 uracil-DNA glycosylase bound to DNA," Journal of Molecular Biology, vol. 373, (4), pp. 839-850, 2007. DOI: 10.1016/j.jmb.2007.08.022.

[26] A. A. Sartori et al, "A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site," The EMBO Journal, vol. 21, (12), pp. 3182-3191, 2002. DOI: 10.1093/emboj/cdf309.

[27] B. Xia et al, "Specificity and catalytic mechanism in family 5 uracil DNA glycosylase," The Journal of Biological Chemistry, vol. 289, (26), pp. 18413-18426, 2014. DOI: 10.1074/jbc.M114.567354.

[28] H. Lee, B. N. Dominy and W. Cao, "New family of deamination repair enzymes in uracil-DNA glycosylase superfamily," The Journal of Biological Chemistry, vol. 286, (36), pp. 31282-31287, 2011. DOI: 10.1074/jbc.M111.249524.

[29] P. B. Sang et al, "A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, vol. 43, (17), pp. 8452-8463, 2015. DOI: 10.1093/nar/gkv854.

[30] L. Jiang et al, "UdgX-mediated uracil sequencing at single-nucleotide resolution," J. Am. Chem. Soc., vol. 144, (3), pp. 1323-1331, 2022. DOI: 10.1021/jacs.1c11269.

[31] J. A. Stewart, G. Schauer and A. S. Bhagwat, "Visualization of uracils created by APOBEC3A using UdgX shows colocalization with RPA at stalled replication forks," Nucleic Acids Research, vol. 48, (20), pp. e118, 2020. DOI: 10.1093/nar/gkaa845.

[32] M. Datta et al, "Development of mCherry tagged UdgX as a highly sensitive molecular probe for specific detection of uracils in DNA," Biochemical and Biophysical Research Communications, vol. 518, (1), pp. 38-43, 2019. DOI: 10.1016/j.bbrc.2019.08.005.

[33] W. Ahn et al, "Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision," Nature Chemical Biology, vol. 15, (6), pp. 607-614, 2019. DOI: 10.1038/s41589-019-0289-3.

[34] Q. Jia et al, "Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability," DNA Repair, vol. 97, pp. 103008, 2021. DOI: 10.1016/j.dnarep.2020.103008.

[35] S. Kumar et al, "TimeTree: A resource for timelines, timetrees, and divergence times," Molecular Biology and Evolution, vol. 34, (7), pp. 1812-1819, 2017. DOI: 10.1093/MOLBEV/MSX116.

[36] I. Letunic and P. Bork, "Interactive Tree Of Life (iTOL) v4: recent updates and new developments," Nucleic Acids Research, vol. 47, (W1), pp. W256-W259, 2019. DOI: 10.1093/nar/gkz239.

[37] S. F. Altschul et al, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," Nucleic Acids Research, vol. 25, (17), pp. 3389-3402, 1997. DOI: 10.1093/nar/25.17.3389.

[38] C. L. Fisher and G. K. Pei, "Modification of a PCR-based site-directed mutagenesis method," Biotechniques, vol. 23, (4), pp. 570-574, 1997. DOI: 10.2144/97234bm01.

[39] O. Lanes et al, "Purification and characterization of a cold-adapted uracil-DNA glycosylase from Atlantic cod (Gadus morhua)," Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, vol. 127, (3), pp. 399-410, 2000. DOI: 10.1016/S0305-0491(00)00271-6.

[40] K. King, S. J. Benkovic and P. Modrich, "Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease," The Journal of Biological Chemistry, vol. 264, (20), pp. 11807-11815, 1989. DOI: 10.1016/S0021-9258(18)80137-5.

[41] C. L. M. Vermote and S. E. Halford, "EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition," Biochemistry, vol. 31, (26), pp. 6082-6089, 1992. DOI: 10.1021/bi00141a018.

[42] J. Tu et al, "Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation," Nature Chemical Biology, vol. 15, (6), pp. 615-622, 2019. DOI: 10.1038/s41589-019-0290-x.

[43] S. Aroli et al, "Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution," Nucleic Acids Research, vol. 51, (13), pp. 6554-6565, 2023. DOI: 10.1093/nar/gkad486.

[44] L. Aravind and E. V. Koonin, "The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates," Genome Biology, vol. 1, (4), pp. RESEARCH0007, 2000. DOI: 10.1186/gb-2000-1-4-research0007.

[45] J. T. Stivers and Y. L. Jiang, "A mechanistic perspective on the chemistry of DNA repair glycosylases," Chemical Reviews, vol. 103, (7), pp. 2729-2760, 2003. DOI: 10.1021/cr010219b.

[46] A. C. Drohat and A. Maiti, "Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA," Organic & Biomolecular Chemistry, vol. 12, (42), pp. 8367-8378, 2014. DOI: 10.1039/c4ob01063a.

CHAPTER THREE

BIOCHEMICAL PROPERTIES OF THE CROSSLINKING SITE OF UDGX

I. Abstract

UDGX is a newly identified member in the uracil-DNA glycosylase (UDG) superfamily with a unique DNA crosslinking activity. The crosslinking is conducted via a C-N covalent bond between the C1' carbon of deoxyribose and the ε^2 nitrogen in the imidazole ring of His109 of Msm UDGX. Previous studies have confirmed that the substitution for His109 causes the loss of DNA crosslinking activity but retains the uracil excision activity. However, there is an exception, the H109E mutant, which can also crosslink to the uracil-DNA substrates. To further study this special mutant, the H109E mutant as well as other substitution mutants for His109, H109G, H109N and H109Q, were constructed, and a series of biochemical and kinetics analyses were performed. The kinetics analyses showed that the reaction rates of all the mutants were reduced, compared with the UDGX wildtype. Although the H109E mutant could crosslink to the DNA substrates, the crosslinking product was hydrolysable by alkali. Similar to the wildtype, the H109E mutant could not crosslink directly to AP sites containing DNA substrates, either. The DNA crosslinking of the H109E mutant could be affected by the pH of reactions. Consistent with previous studies, His109 not only plays its role in crosslinking, but is also essential in the uracil excision step. The substitution mutant, H109E, converts the DNA-protein crosslinking from a C-N covalent bond to an ester

bond, and this crosslinking seems not as robust as the original crosslinking by the UDGX wildtype.

II. Introduction

Enzymes from uracil-DNA glycosylase (UDG) superfamily are a group of essential enzymes involved in the base excision repair (BER) of DNA base damage [1]. There have been at least six families of proteins identified in this superfamily. Although some of the UDG enzymes have a broad substrate specificity [2-4], the major function as DNA glycosylases is conserved, which is breaking the glycosidic bond between the damaged base and the C1' carbon of deoxyribose [5]. However, the discovery of a new family of UDG proteins, UDGX, refreshes the perception of the roles UDG enzymes can play in DNA repairs [6]. Unlike enzymes from the other UDG families, UDGX can form covalent crosslinking to the abasic site (AP site) via a histidine residue (H109 in Msm UDGX) after the uracil base excision [6-8].

As the crosslinking site, the substitution for His109 causes the loss of crosslinking activity of the proteins [6-8], but there is an exception that the mutant H109E still exhibits crosslinking to the uracil-DNA substrates [9]. With the crystal structures solved for the UDGX H109E mutant as well as the crosslinked complex of H109E and DNA, it has been found that the glutamate residue crosslinks the AP site via an ester bond (Figure 3.1), which is different from the C-N bond formed by C1' carbon of deoxyribose and ε^2 nitrogen of histidine [7-9]. Since there is a conversion of the crosslinking pattern from the wildtype UDGX to the H109E mutant, a series of quantitative and biochemical analyses

were performed to illustrate the characteristics of the H109E as well as other substitution mutants for H109, as part of our comprehensive analyses of the unique UDG, UDGX.

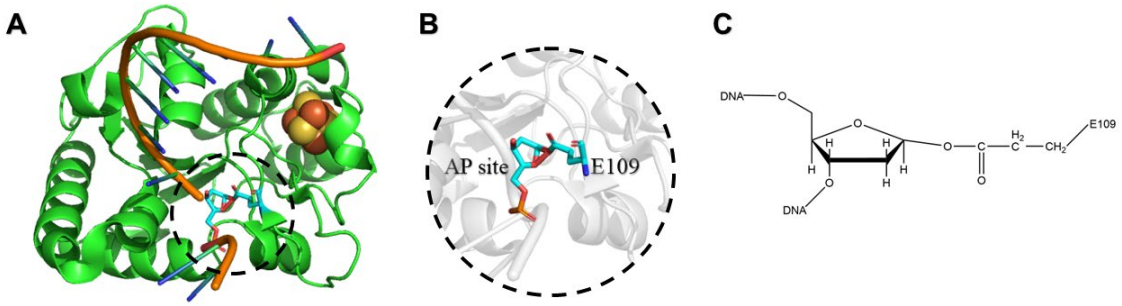

Figure 3.1 Crosslinking between UDGX H109E mutant and AP site. A. Structure of crosslinked complex of UDGX H109E mutant and DNA (PDB code: 6L6S). **B.** Close-up view of the crosslinking between E109 residue and AP site. **C.** Chemical structure of the ester bond formed by E109 and deoxyribose.

III. Materials and Methods

A. Reagents, media and strains

All routine reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), VWR (Suwanee, GA), ThermoFisher Scientific (Waltham, MA) or Gel Company (San Francisco, CA) 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. Plasmid miniprep kits and DNA gel extraction kits were purchased from New England Biolabs (Ipswich, MA) and ThermoFisher Scientific (Waltham, MA). Phusion DNA polymerase and dNTP were purchased from ThermoFisher Scientific (Waltham, MA). Cod UNG (heat-labile UDG from *Gadus morhua*) was purchased from ArcticZymes (Tromsø, Norway). HisTrap FF (1 mL), HiTrap Q FF (1 mL) and HiTrap SP FF (1 mL) 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. Oligonucleotide primers for PCR were synthesized from Eurofins Genomics (Huntsville, AL). Oligodeoxynucleotide substrates with carboxyfluorescein (FAM) fluorescence label were ordered from Integrated DNA Technologies Inc. (Coralville, IA). The LB medium was prepared according to standard recipes. The sonication buffer consisted of 50 mM Tris-HCl (pH 7.5), 300 mM NaCl and 40 mM imidazole. Buffer A for HisTrap FF columns 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, 500 mM imidazole and 10% glycerol. Buffer A for HiTrap Q/SP FF columns consisted of 50 mM Tris-HCl (pH 7.5) and buffer B consisted of 50 mM Tris-HCl (pH 7.5) and 1 M NaCl. Citrate-phosphate universal buffer (250 mM, pH 5.0, 5.5, 6.0, 6.5 and 7.0), Tris-HCl buffer (250 mM, pH 7.5, 8.0 and 8.5) and glycine-NaOH buffer (250 mM, pH 9.0) were prepared according to standard recipes. *E. coli* strain DH5α was purchased from ThermoFisher Scientific (Waltham, MA) and *E. coli* strain Rosetta was purchased from VWR (Suwanee, GA).

B. PCR-based site-directed mutagenesis, expression cloning and protein purification

PCR-based site-directed mutagenesis was performed using the pET28b-udgx recombinant plasmid as the template similarly as previously described [10]. The PCR reaction mixture (40 μL) consisted of 20 ng DNA template, 50 nM of each primer pair carrying the desired mutations (Table 3.1), 200 μM each dNTP, 1x Phusion DNA polymerase buffer and 1 Unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation step at 98°C for 2 min; 25 cycles of three-step amplification with each

cycle consisting of denaturation at 98°C for 10 s, annealing at 65°C for 20 s, and extension at 72°C for 7.5 min; and a final extension step at 72°C for 5 min. The PCR product was then treated with Dpn I and transformed into *E. coli* strain DH5α competent cells. Successfully mutated plasmid was confirmed by DNA sequencing.

Mutant	Primers
H109G	5'-CGACTGGGGGTCTTGCCGATGCGTCGTTTGCC-3' 5'-GGCAAACGACGCATCGGCAAGACCCCCAGTCG-3'
H109N	5'-GACTGGGGGTCTTGTTGATGCGTCGTTTGCC-3' 5'-GGCAAACGACGCATCAACAAGACCCCCAGTC-3'
H109Q	5'-CGACTGGGGGTCTTCTGGATGCGTCGTTT-3' 5'-AAACGACGCATCCAGAAGACCCCCAGTCG-3'
H109E	5'-CGACTGGGGGTCTTCTCGATGCGTCGTTTGC-3' 5'-GCAAACGACGCATCGAGAAGACCCCCAGTCG-3'

Table 3.1 Primers for mutagenesis.

Sequencing confirmed recombinant plasmids were finally transformed into *E. coli* strain Rosetta competent cells. Culture incubation, IPTG-induced protein expression and purification were performed as described below. A single *E. coli* colony transformed with recombinant plasmid was cultured in 500 mL LB medium supplemented with 50 μg/mL Kanamycin at 37°C with shaking at 250 rpm until the optical density at 600 nm was over 0.6. After adding IPTG to a final concentration of 0.8 mM, the culture was grown at 22°C overnight. The *E. coli* cells were harvested at 5000 rpm with a JLA8.1000 rotor (Beckman Coulter) at 4°C for 20 min. The cell pellet was resuspended in 7 mL sonication buffer and then sonicated at 5 s on/ 5 s off cycle for 10 min on ice using Qsonica model Q125. The sonicated solution was then centrifuged at 12000 rpm with a JLA16.250 rotor (Beckman Coulter) at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded

onto a 1 mL HisTrap FF column. After the sample loading, the column was washed with 15 mL of buffer A. The bound protein in the column was eluted with a linear gradient of 0-100% buffer B and collected in 1 mL fractions. The fractions identified by UV280 and SDS-PAGE were pooled and diluted 5-fold with 50 mM Tris-HCl (pH 7.5) buffer. The diluted solution was loaded onto a 1 mL HiTrap Q FF or HiTrap SP FF column. After the column was washed with 15 mL of buffer A, the bound protein in the column was eluted with a linear gradient of 10-100% buffer B and collected in 1 mL fractions. After identification, the fractions containing target proteins were pooled and concentrated. The protein concentration was determined by Nanodrop One (ThermoFisher Scientific). The protein solution was stored at -20°C in 50% glycerol.

C. Oligodeoxynucleotide substrates

Five types of uracil containing carboxyfluorescein (FAM) labelled oligodeoxynucleotide substrates were prepared as previously described [11]. AP site containing single-stranded oligodeoxynucleotide substrate was produced, as previously described, by incubating single-stranded U oligodeoxynucleotides with Cod UNG, followed by heating at 55°C to inactivate UNG [12].

D. Denaturing polyacrylamide gel electrophoresis

Assays were performed at 37°C for 30 min in 10 μL reaction mixtures containing 20 nM DNA substrate, excess enzymes and 25 mM reaction buffers. 10 μL of ice-cold water or 100 mM NaOH was added to the mixtures. After heating at 95°C for 5 min of NaOH-treated mixtures (while ice-cold water mixed samples were kept on ice), all the samples were supplemented with 5 μL of 50% glycerol loading buffer and electrophoresed at 130

V for 85 min on a 15% urea denaturing polyacrylamide gel in 0.5 x TB buffer (44.6 mM Tris base and 44.6 mM boric acid) supplemented with 2.5 mM EDTA. Fluorescence-labelled oligodeoxynucleotides in the gel were visualized by a Typhoon FLA 7000 imager (GE Healthcare).

E. Capillary electrophoresis

Assays were performed at 37°C for 30 min in 10 μL reaction mixtures containing 10 nM oligodeoxynucleotide substrate, excess enzymes, and 25 mM Tris-HCl buffer (pH 7.5). The resulting abasic sites were cleaved by heating at 95°C for 5 min after adding 1 μL of 1 M NaOH to stop the reaction. The mixtures (2 μL) were then mixed with 7.8 μL Hi-Di formamide and 0.2 μL GeneScan 500 LIZ dye size standard, and analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

F. Enzyme kinetics analysis

Excess enzymes (range from 100 nM to 2000 nM) of UDGX mutants were incubated with G/U or A/U DNA substrates at 37°C in 5 μL reaction mixtures supplemented with 25 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA and 1 mM DTT. Samples were collected at 5 min, 10 min, 15 min, 20 min, 30 min, 40 min, 55 min and 70 min for reactions with G/U base pairs; at 10 min, 20 min, 30 min, 40 min, 55 min, 70 min, 85 min and 100 min for reactions with A/U base pairs, by adding 5 μL of 100 mM NaOH to terminate the reactions. After incubation at 95 °C for 5 min, 2 μL of reaction mixtures were mixed with 7.8 μl Hi-Di formamide and 0.2 μl GeneScan 500 LIZ dye size standard. Samples were then analyzed

by ABI 3130xl with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software.

The apparent rate constants (k_{obs}) for each concentration of UDGX mutants were determined by nonlinear fitting using the integrated first-order rate equation (1):

$$P=P_{max}(1-e^{-k_{obs}t}) \quad (1)$$

where P is the product yield, P_{max} is the maximal yield, t is time and k_{obs} is the apparent rate constant.

The kinetic parameters k_2 and K_m were obtained from plots of k_{obs} against the total enzyme concentration ($[E_0]$) using a standard hyperbolic kinetic expression with the program GraphPad Prism 9 following the equation (2) [13-15]:

$$k_{obs} = \frac{k_2[E_0]}{K_m+[E_0]} \quad (2)$$

For the H109G mutant, because of a large K_m, in which $K_m \gg [E_0]$, the kinetic parameter k_2/K_m values were obtained from plots of k_{obs} against the total enzyme concentration ($[E_0]$) using a linear regression with the program GraphPad Prism 9 following the equation (3) [13-15]:

$$k_{obs} = \frac{k_2[E_0]}{K_m} \quad (3)$$

IV. Results

By PCR-based site-directed mutagenesis, four substitution mutants of Msm UDGX were constructed, in which the His109, the AP site crosslinking residue, was substituted with glycine, asparagine, glutamine or glutamic acid. Uracil excision and DNA crosslinking activities were tested first for the mutants by incubating them with G/U base pair containing DNA substrates. After alkali and heat treatment, truncated DNA products from AP sites, which were formed by uracil excision, were observed for all the four mutants, but no crosslinking product was detected (Figure 3.2A). However, when we ran the electrophoresis immediately after the incubation of reaction mixtures without alkali and heat treatment, a weak crosslinking activity was observed for the H109E mutant, whereas the other three mutants still could not crosslink any DNA substrates (Figure 3.2A).

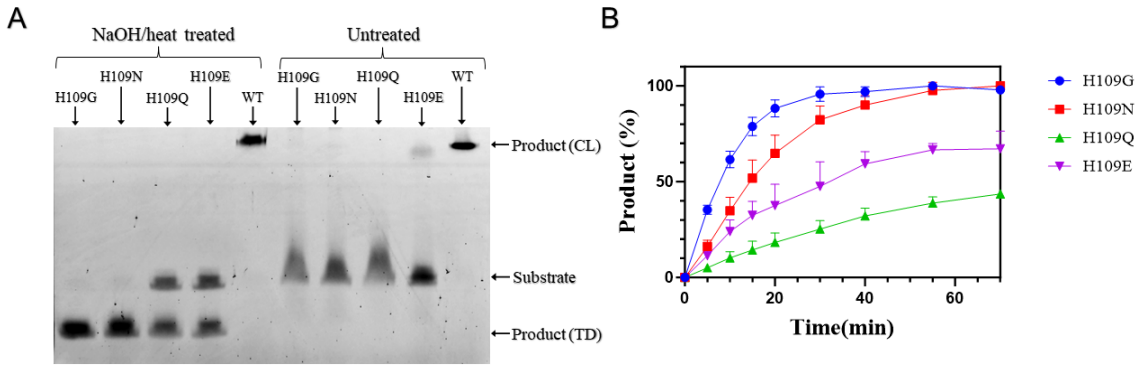

Figure 3.2 Activity determination of UDGX mutants. A. Uracil excision and DNA crosslinking activities of H109 mutants. Assays were performed as described in Materials and Methods with 10 nM of G/U substrates and 500 nM of UDGX H109G, H109N, H109Q and H109E mutants and wildtype. Products and free substrates were visualized by denaturing polyacrylamide gel electrophoresis and fluorescence scanning. Left five reaction mixtures were treated with NaOH and heat, right five were untreated. WT: Wildtype UDGX. Product (CL) indicates the DNA-UDGX crosslinking product. Product (TD) indicates the truncated DNA product after NaOH/heat treatment. **B.** Time-course analysis of uracil excision activity of UDGX H109G, H109N, H109Q and H109E mutants on G/U substrates. (●) H109G; (■) H109N; (▲) H109Q; (▼) H109E. The

assays were performed as described in Materials and Methods under Enzyme kinetics analysis with 500 nM of enzymes and 10 nM of G/U substrates. Data are shown as the average of three independent experiments and the standard deviation for each point is shown.

Since all four mutants exhibited their activities on uracil excision, a time-course analysis was performed with the mutants on G/U base pairs. As shown in Figure 3.2B, the H109G mutant had the fastest reaction rate, and converted all the substrates to products within 30 minutes. Though the reaction with the H109N mutant was a little slower than with the H109G mutant, the reaction still went to completion within 70 minutes. However, the H109E mutant could only convert around 60% of substrates to products within 70 minutes observed. The H109Q mutant showed the slowest reaction rate and could only produce about 40% of products within 70 minutes.

To quantitatively compare the activities of the mutants, kinetic analyses were performed (Table 3.2). Consistent with the time-course analysis, the H109G mutant had the largest k_2/K_m values with G/U base pair substrates among the four mutants. Surprisingly, although the k_2 value of H109E mutant with the A/U base pair was 10-fold slower than H109G mutant with the A/U base pair, the K_m of H109E was ~16-fold lower than H109G, resulting in a larger k_2/K_m value of H109E with the A/U base pair than H109G with the A/U base pair. Similarly, although the k_2 values of H109E mutant with both G/U and A/U base pair substrates were slower than H109N, because of the low K_m values, the catalytic efficiencies, reflected by k_2/K_m values, were higher for the H109E than the H109N mutant. The k_2/K_m values of H109G and H109N with G/U substrates were generally higher than with A/U substrates, 3-fold and 2-fold respectively, which is similar with the results of previous study on UDGX wildtype. However, the H109Q and

H109E mutants showed almost equal k_2/K_m values with the G/U and A/U base pairs. As the slowest enzyme among the four mutants, the H109Q mutant showed one-order of magnitude lower for the G/U base pair and 4-fold lower for the A/U base pair than the H109G mutant.

Enzyme	G/U			A/U		
	K_m (M)	k_2 (s^{-1})	k_2/K_m ($M^{-1}s^{-1}$)	K_m (M)	k_2 (s^{-1})	k_2/K_m ($M^{-1}s^{-1}$)
WT[b]	N.D.[c]	N.D.	$1.4\ (0.09) \times 10^{4}$	N.D.	N.D.	$7.4\ (1.8) \times 10^{3}$
H109G	N.D.	N.D.	$6.0\ (0.6) \times 10^{3}$	$1.3\ (0.2) \times 10^{-6}$	$3.0\ (0.9) \times 10^{-3}$	$2.3\ (0.3) \times 10^{3}$
H109N	$1.6\ (0.8) \times 10^{-6}$	$3.5\ (0.9) \times 10^{-3}$	$2.4\ (0.5) \times 10^{3}$	$1.2\ (0.4) \times 10^{-6}$	$1.4\ (0.4) \times 10^{-3}$	$1.2\ (0.03) \times 10^{3}$
H109Q	$4.2\ (1.7) \times 10^{-7}$	$2.2\ (0.2) \times 10^{-4}$	$5.9\ (2.4) \times 10^{2}$	$1.2\ (0.4) \times 10^{-7}$	$6.1\ (0.8) \times 10^{-5}$	$5.3\ (1.4) \times 10^{2}$
H109E	$1.2\ (0.3) \times 10^{-7}$	$5.3\ (0.8) \times 10^{-4}$	$4.3\ (0.4) \times 10^{3}$	$8.2\ (1.2) \times 10^{-8}$	$3.3\ (0.7) \times 10^{-4}$	$3.9\ (0.3) \times 10^{3}$

Table 3.2 Kinetic parameters of UDGX mutants with G/U and A/U base pairs[a]
[a]: Enzyme kinetics analysis was performed as described in Materials and Methods. Data are shown as the average of three independent experiments. S.D. values are shown in parentheses.
[b]: Data of WT (wildtype UDGX) are from Chapter two.
[c]: N.D. Not determined due to large K_m.

From the crystal structure of the crosslinked complex of UDGX H109E mutant and DNA, the crosslinking between the glutamate residue and the AP site is formed via an ester bond [9]. To further confirm this experimentally, two catalytic assays were performed with the H109E mutant and G/U base pair substrates. In the first assay, the reactions were stopped by fast cooling the mixtures on ice at 10 s, 30 s, 1 min, 3 min, 5 min, 10 min, 15 min, 25 min and 40 min, followed by being immediately electrophoresed (Figure 3.3A). In the second assay, reactions were stopped by adding sodium hydroxide at the same time points, and the mixtures were heated at 95°C for 5 minutes before the electrophoresis (Figure 3.3B). The results of the two assays confirmed that the crosslinking between the H109E mutant and the AP site was alkaline hydrolysable.

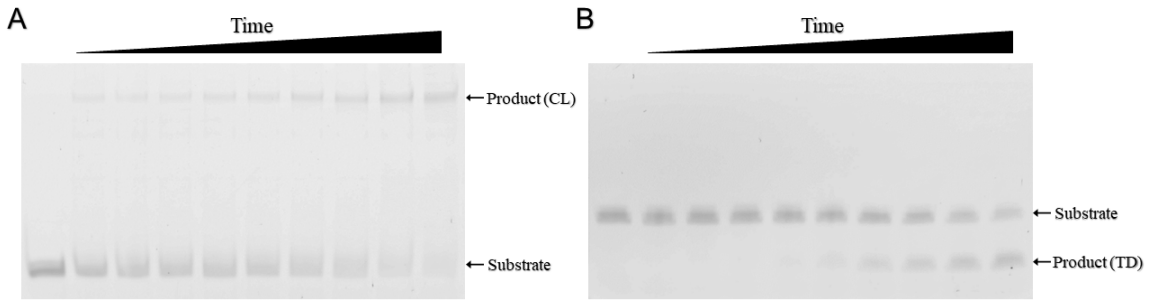

Figure 3.3 Alkaline hydrolysis of H109E-DNA crosslinking product. Catalytic reactions were performed with 1000 nM H109E and 20 nM G/U substrates. **A.** Untreated reactions. Reactions were stopped by fast cooling on ice. Product (CL) indicates the H109E-DNA crosslinking product. **B.** Alkaline treated reactions. Reaction was stopped by adding 1 µL 1 M NaOH, followed by heating at 95°C for 5 min. Product (TD) indicates the truncated DNA product after NaOH/heat treatment.

To investigate whether, like the UDGX wildtype, the uracil excision and AP site crosslinking are also coupled of the H109E mutant or not, the H109E proteins were subjected to either single-stranded U (ssU) substrates or single-stranded AP site (ssAP) substrates, which were produced by a heat-labile UNG from ssU oligonucleotides. As shown in Figure 3.4, H109E could crosslink ssU substrates to form complex, but no crosslinking was observed with ssAP substrates.

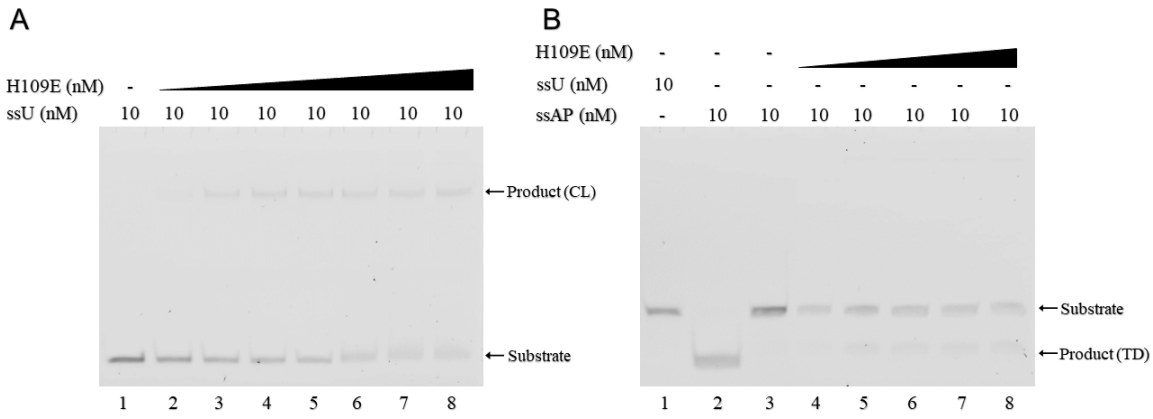

Figure 3.4 Uracil excision and DNA crosslinking coupling of UDGX H109E mutant. A. DNA crosslinking assay of H109E on single-stranded U DNA substrates. Assays were performed as described in Materials and Methods under denaturing polyacrylamide gel electrophoresis with

modification. Enzymes (ranging from 10 to 1000 nM) were incubated with 10 nM of ssU substrates at 37°C for 60 min. Product (CL) indicates the H109E-DNA crosslinking product. **B.** DNA crosslinking assay of H109E on single-stranded AP site substrates. Assays were performed as described in Materials and Methods under Denaturing polyacrylamide gel electrophoresis with modification. Enzymes (ranging from 100 to 1000 nM) were incubated with 10 nM of ssAP substrates at 37°C for 60 min. Lane 1 was ssU DNA only; lane 2 and 3 were ssAP DNA only, but sample in lane 2 was treated with NaOH/heat. Product (TD) indicates the truncated DNA product after NaOH/heat treatment.

Given that the formation of ester bonds can be influenced by pH, we performed the catalytic reactions with the H109E mutant on G/U base pair substrates at different pH. According to the results, the environmental pH had little effect on uracil excision of the H109E mutant (Figure 3.5A), but the crosslinking was significantly affected by pH (Figure 3.5B). A quantitative analysis of the pH effects on the catalytic reactions by H109E was consistent with the electrophoresis results (Figure 3.5C). Although the most truncated DNA products were detected at pH 8.5, the product yields at different pH tested, in general, were all around 40% without significant discrepancies. However, the most crosslinking products were detected at pH 7.0 and 8.5 with a yield around 30%, whereas, at other pH, the products were only around 20% or even less.

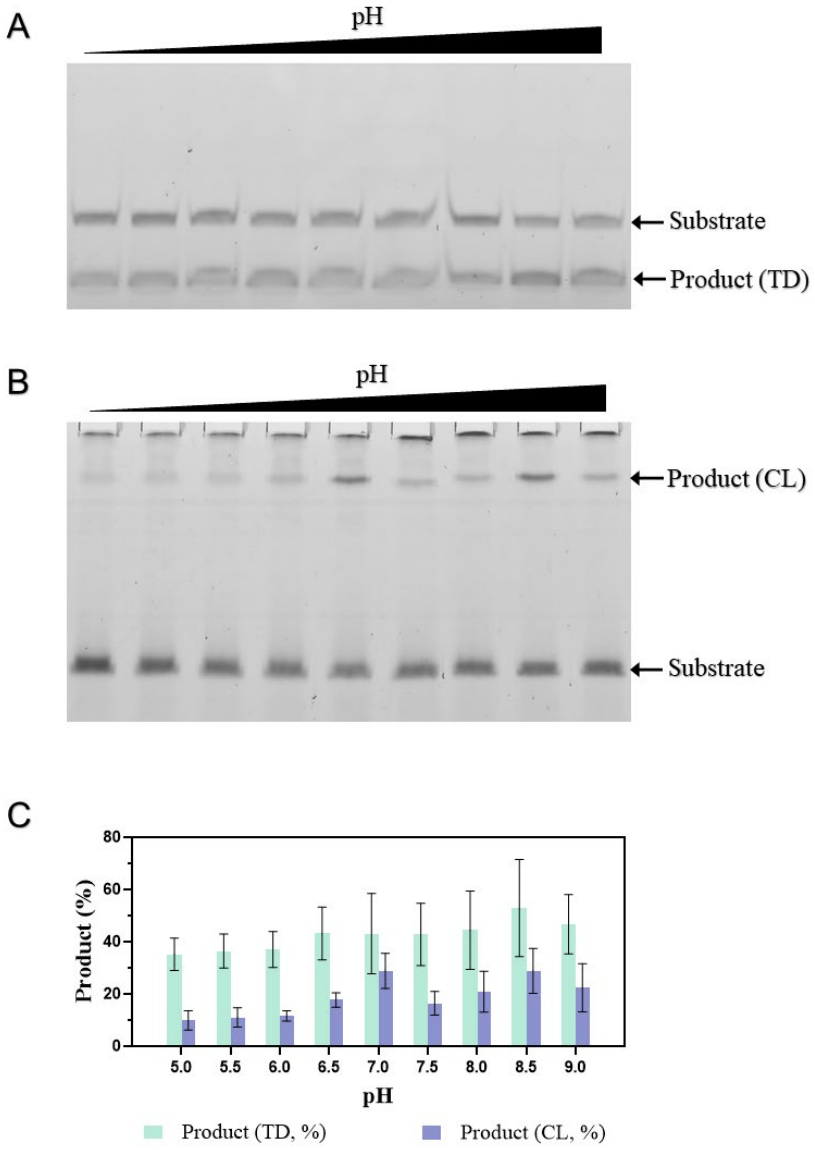

Figure 3.5 Effects of pH on uracil excision and DNA crosslinking of the H109E mutant. A. Effects of pH on uracil excision. Assays were performed as described in Materials and Methods under denaturing polyacrylamide gel electrophoresis with modification. 500 nM of H109E enzymes were incubated with 10 nM G/U substrates at pH 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0. Reactions were stopped by NaOH, followed by heating at 95°C for 5 min. **B.** Effects of pH on DNA crosslinking. Assays were performed under the same conditions as the assays in **Figure 3.5A**, but without NaOH/heat treatment. **C.** Quantitative analysis of pH effects on uracil excision and DNA crosslinking. Assays were performed as described in **Figure 3.5A** and **B**, intensities of fluorescence signals of products and free DNA species were quantified using a Typhoon 7000 FLA imager and ImageQuant TL software (GE Healthcare). Data are shown as the average of three independent experiments. Product (TD) indicates the truncated DNA product after NaOH/heat treatment. Product (CL) indicates the H109E-DNA crosslinking product.

V. Discussion

From the previous study, it has been known that the crosslinking and uracil excision by UDGX wildtype are coupled. During the breaking of the glycosidic bond between uracil and C1' carbon of deoxyribose, a new C-N covalent bond forms immediately between the C1' carbon and ε^2 nitrogen in the imidazole ring of His109. Because the His109 participates in both uracil excision and crosslinking steps, when the histidine is replaced by other amino acids, compared with the k_2/K_m values (1.4×10^4 $M^{-1}s^{-1}$ on G/U, 7.4×10^3 $M^{-1}s^{-1}$ on A/U) of the wildtype, the total reaction rates by the mutants were significantly reduced, especially the H109Q mutant, the reaction rate decreased by 24-fold with G/U base pair and 14-fold with A/U base pair (Table 3.1). As for the differences of catalytic efficiencies of the H109 substitution mutants, a crystal structure study reported that it's probably caused by their different topological changes from the wildtype in the active sites [16].

Although the H109E mutant can also crosslink to the DNA substrates, the mechanism seems different from the wildtype. According to the experimental results, the crosslinking between the H109E mutant and DNA was hydrolyzed by alkali, which supports the finding that the crosslinking is formed via an ester bond [9]. Albeit there are some examples of DNA-protein crosslinking by ester bonds, such as topoisomerase DNA-protein crosslinks (TOP-DPC) [17], which are formed via ester bonds between a tyrosine residue of the protein and the phosphate group of DNA, there has never been a case reported of DNA-protein crosslinking via an ester bond between the protein and the deoxyribose before. Based on the results of the assay with AP site containing DNA

substrates, the H109E mutant could not crosslink to the AP site without the uracil excision step (Figure 3.4B), which suggests the crosslinking of the H109E mutant is also coupled with uracil excision, just like the UDGX wildtype. Therefore, it can be hypothesized that the formation of such an unusual ester bond between the glutamate residue and the C1' carbon of deoxyribose requires the removal of the base from the C1' carbon simultaneously.

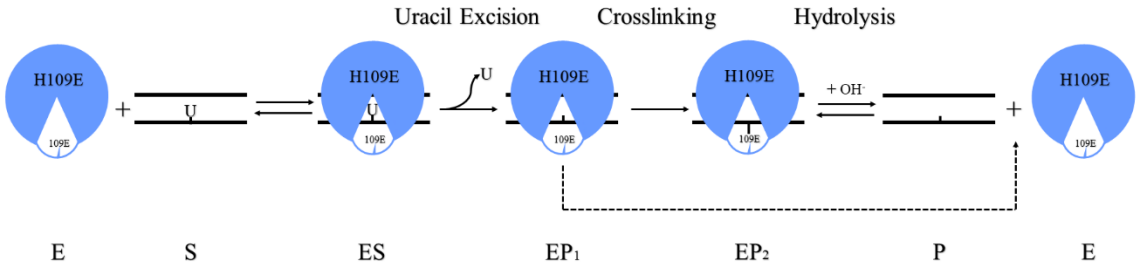

Figure 3.6 Catalytic scheme of UDGX H109E mutant. Catalytic scheme of UDGX H109E mutant on U-containing DNA substrate. E: Enzyme; S: Substrate; ES: Enzyme-Substrate complex; EP_1: Complex of enzyme and uracil excised DNA; EP_2: H109E-DNA crosslinking complex; P: AP site containing DNA product.

Interestingly, we found that the crosslinked product by H109E mutant was less than the truncated DNA product under the same conditions (Figure 3.2A and Figure 3.5C). Here, there are two possible explanations for the finding. The first is that the H109E mutant cannot crosslink all the AP sites produced by itself from uracil excision. The second is that some of the ester bonds formed were hydrolyzed since the formation and hydrolysis of ester are reversible. No matter which caused the reduction of the crosslinked product, to obtain more accurate kinetic parameters, which could reflect the actual rate of the catalytic reactions by the H109E mutant, truncated DNA product from NaOH/heat treatment was used to perform the kinetics analyses. The kinetic parameters

obtained in fact reflect the rate of the whole process shown in Figure 3.6. Since the concentration of NaOH added (50 mM) was much higher than the concentration of ester produced, which is in a nanomolar range, the effect of alkaline hydrolysis on the estimate of the reaction rate can be negligible [18].

In conclusion, the crosslinking site (residue H109 in Msm UDGX) plays important roles in both uracil excision and DNA crosslinking steps. Substitutions for it can cause a reduced reaction rate and loss of DNA crosslinking ability, except for the H109E mutant. Though the H109E mutant also shows a uracil excision coupled DNA crosslinking activity, unlike a C-N covalent bond formed by UDGX wildtype, it crosslinks the deoxyribose with its glutamate residue via an ester bond, which is not as robust as the C-N bond.

VI. Reference

[1] D. O. Zharkov, "Base excision DNA repair," Cell. Mol. Life Sci, vol. 65, (10), pp. 1544-1565, 2008. DOI: 10.1007/s00018-008-7543-2.

[2] M. Dizdaroglu et al, "Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage," Nucleic Acids Research, vol. 24, (3), pp. 418-422, 1996. DOI: 10.1093/nar/24.3.418.

[3] D. Cortázar et al, "The enigmatic thymine DNA glycosylase," DNA Repair, vol. 6, (4), pp. 489-504, 2007. DOI: 10.1016/j.dnarep.2006.10.013.

[4] H. Lee et al, "Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase," The Journal of Biological Chemistry, vol. 285, (53), pp. 41483-41490, 2010. DOI: 10.1074/jbc.M110.150003.

[5] H. E. Krokan, R. Standal and G. Slupphaug, "DNA glycosylases in the base excision repair of DNA," Biochemical Journal, vol. 325 (Pt 1), (1), pp. 1-16, 1997. DOI: 10.1042/bj3250001.

[6] P. B. Sang et al, "A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, vol. 43, (17), pp. 8452-8463, 2015. DOI: 10.1093/nar/gkv854.

[7] W. Ahn et al, "Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision," Nature Chemical Biology, vol. 15, (6), pp. 607-614, 2019. DOI: 10.1038/s41589-019-0289-3.

[8] J. Tu et al, "Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation," Nature Chemical Biology, vol. 15, (6), pp. 615-622, 2019. DOI: 10.1038/s41589-019-0290-x.

[9] Q. Jia et al, "Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability," DNA Repair, vol. 97, pp. 103008, 2021. DOI: 10.1016/j.dnarep.2020.103008.

[10] C. L. Fisher and G. K. Pei, "Modification of a PCR-based site-directed mutagenesis method," Biotechniques, vol. 23, (4), pp. 570-574, 1997. DOI: 10.2144/97234bm01.

[11] R. Mi et al, "Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG," Journal of Molecular Biology, vol. 385, (3), pp. 761-778, 2009. DOI: 10.1016/j.jmb.2008.09.038.

[12] O. Lanes et al, "Purification and characterization of a cold-adapted uracil-DNA glycosylase from Atlantic cod (Gadus morhua)," Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, vol. 127, (3), pp. 399-410, 2000. DOI: 10.1016/S0305-0491(00)00271-6.

[13] B. Xia et al, "Correlated mutation in the evolution of catalysis in uracil DNA glycosylase superfamily," Scientific Reports, vol. 7, (1), pp. 45978, 2017. DOI: 10.1038/srep45978.

[14] K. King, S. J. Benkovic and P. Modrich, "Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease," The Journal of Biological Chemistry, vol. 264, (20), pp. 11807-11815, 1989. DOI: 10.1016/S0021-9258(18)80137-5.

[15] C. L. M. Vermote and S. E. Halford, "EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition," Biochemistry, vol. 31, (26), pp. 6082-6089, 1992. DOI: 10.1021/bi00141a018.

[16] S. Aroli et al, "Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution," Nucleic Acids Research, vol. 51, (13), pp. 6554-6565, 2023. DOI: 10.1093/nar/gkad486.

[17] Y. Sun et al, "Debulking of topoisomerase DNA-protein crosslinks (TOP-DPC) by the proteasome, non-proteasomal and non-proteolytic pathways," DNA Repair, vol. 94, pp. 102926, 2020. DOI: 10.1016/j.dnarep.2020.102926.

[18] F. Jabeen, Q. I. Rahman and S. Zafar, "Determination of rate of reaction and rate constant of the hydrolysis of ester (ethyl acetate) with alkali(sodium hydroxide)," International Journal of Applied Chemistry, vol. 6, (3), pp. 18-23, 2019. DOI: 10.14445/23939133/IJAC-V6I3P104.

CHAPTER FOUR

URACIL-DNA CROSSLINKING ACTIVITY IN PUTATIVE UDGX HOMOLOGS

I. Abstract

Uracil-DNA glycosylases (UDG) are responsible for the removal of damaged bases from DNA in base excision repair (BER). There have been at least seven families of enzymes identified in the UDG superfamily. The last member of the superfamily is UDGX, which is a unique bifunctional enzyme. Not only excising uracil base from DNA, UDGX can also covalently crosslink to the apurinic/apyrimidinic site (AP site) formed by the base removal. Since the discovery of UDGX from *Mycobacterium smegmatis* (Msm UDGX), hundreds of proteins with similar sequences have been identified. In this work, the DNA crosslinking activity of six UDGX homologs were screened. The results showed that the proteins from *Saccharomonospora marina* (Sma UDGX), *Nonomuraea jiangxiensis* (Nji UDGX) and *Saccharomonospora viridis* (Svi UDGX) had DNA crosslinking activities, but proteins from *Burkholderia gladioli* (Bgl UDGX) and *Mycobacterium goodii* (Mgo UDGX1 and Mgo UDGX2) could not crosslink to the uracil-containing DNA. Mutational and biochemical analyses of Bgl UDGX, Mgo UDGX1 and Mgo UDGX2 suggest that some residues out of the catalytic motifs could also influence the activity of UDGX. A salt bridge, which is formed by R158 and D200 in Msm UDGX, was confirmed critical to the activity. In Mgo UDGX1 and Mgo UDGX2, the salt bridge is lost due to a substitution for the arginine by histidine, which

causes the loss of the activity. Such an activity loss might be a result of evolution to allow the fast proliferation of *M. goodii* bacteria.

II. Introduction

Uracil-DNA glycosylases (UDG) are a group of enzymes involved in the base excision repair (BER) of DNA base lesions, whose main function is to excise the damaged base off from the deoxyribose and produce an apurinic/apyrimidinic site (AP site). In the UDG superfamily, there have been seven families identified. Family 1 UNG (uracil-*N*-glycosylase) is a group of widely distributed DNA glycosylases with narrow specificity [1-3]. Family 2 MUG/TDG (mismatch-specific uracil-DNA glycosylase/ thymine-DNA glycosylase) is a series of enzymes, which are distributed from bacteria to eukaryotes, with broad substrate specificity [3-12]. Family 3 SMUG1/SMUG2 (single strand selective monofunctional UDG) contains multiple subfamilies of enzymes with a broad substrate spectra and wide distribution [13-19]. Family 4 UDGa and family 5 UDGb are both UDG enzymes found in thermophilic bacteria and archaea [20-22]. UDGa only has DNA glycosylase activity on double-stranded uracil-DNA [23, 24], whereas UDGb has a broad substrate specificity, such as hypoxanthine and xanthine [25, 26]. Family 6 HDG (hypoxanthine-DNA glycosylases) is a class of enzymes identified with a predominant hypoxanthine DNA glycosylase activity but no activity on uracil DNA [27].

The seventh family of UDG is a newly identified protein, UDGX. Unlike the other UDG families, besides the base excision activity, this unique enzyme can also covalently

crosslink to the AP site produced by itself [28-30]. The first reported UDGX was from *Mycobacterium smegmatis* [28]. Hundreds of proteins with similar sequences can be identified, but a previous phylogenetic study has confirmed that all the putative UDGX homologs are from bacteria. Despite of the similar sequences with Msm UDGX, whether these proteins also have the DNA crosslinking activity or not is still unrevealed. In this work, the crosslinking activity to uracil-DNA of several UDGX homologs were screened. Previously, our study has clarified the catalytic mechanism and the roles of some essential residues of Msm UDGX, we believe, by biochemical analyses on these UDGX homologs, it may shed light on the evolutionary process of UDGX among species, as well as provide a further understanding of the connection between structure and function of UDGX.

III. Materials and Methods

A. Reagents, media and strains

All routine reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), VWR (Suwanee, GA), ThermoFisher Scientific (Waltham, MA) or Gel Company (San Francisco, CA) 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. Plasmid miniprep kits and DNA gel extraction kits were purchased from New England Biolabs (Ipswich, MA) and ThermoFisher Scientific (Waltham, MA). Restriction enzymes, Phusion DNA polymerase, T4 DNA ligase and dNTP were purchased from ThermoFisher Scientific (Waltham, MA). HisTrap FF (1 mL),

HiTrap Q FF (1 mL) and HiTrap SP FF (1 mL) 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. Gene strands and oligonucleotide primers for PCR were synthesized from Eurofins Genomics (Huntsville, AL). Oligodeoxynucleotide substrates with carboxyfluorescein (FAM) fluorescence labels were ordered from Integrated DNA Technologies Inc. (Coralville, IA). The LB medium was prepared according to standard recipes. The sonication buffer consisted of 50 mM Tris-HCl (pH 7.5), 300 mM NaCl and 40 mM imidazole. Buffer A for HisTrap FF columns 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, 500 mM imidazole and 10% glycerol. Buffer A for HiTrap Q/SP FF columns consisted of 50 mM Tris-HCl (pH 7.5) and buffer B consisted of 50 mM Tris-HCl (pH 7.5) and 1 M NaCl. *E. coli* strain DH5α was purchased from ThermoFisher Scientific (Waltham, MA) and *E. coli* strain Rosetta was purchased from VWR (Suwanee, GA).

B. Cloning, expression and purification of UDGX homolog proteins

The optimized cDNA sequences of UDGX homologs from *Saccharomonospora marina* (GenBank accession number: WP_009153389.1), *Nonomuraea jiangxiensis* (GenBank accession number: WP_090946880.1), *Saccharomonospora viridis* (GenBank accession number: WP_015785655.1), *Burkholderia gladioli* (GenBank accession number: WP_036055803.1) and *Mycobacterium goodii* (GenBank accession number: MBU8812906.1 and MBU8821111.1) were synthesized and amplified by PCR using the primers listed in Table 4.1. The PCR reaction mixture (50 μL) consisted of 25 ng genomic

DNA template, 100 nM forward and reverse primers, 1x Phusion DNA polymerase buffer, 200 μM each dNTP, and 0.5 unit of Phusion DNA polymerase (ThermoFisher Scientific). The PCR procedure included a pre-denaturation step at 98°C for 1 min; 35 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 10 s, annealing at 60°C for 20 s, and extension at 72°C for 90 s; and a final extension step at 72°C for 7 min. The PCR product was purified and digested with restriction enzymes. After digestion, the PCR product was ligated with the same restriction enzyme digested pET21a vector with T4 DNA ligase. The ligation mixture was then transformed into *E. coli* strain DH5α competent cells. The recombinant plasmid was finally verified by DNA sequencing with T7 promoter and T7 terminator primers.

Primer Name	Species	Primer Sequence	Restriction Enzyme
Sma UDGX Forward	*Saccharomonospora marina*	5'-CGCGGATCCATGAATCAGAGATCG-3'	BamH I
Sma UDGX Reverse	*Saccharomonospora marina*	5'-CCGGAATTCGCAATAGCTTTAGCAACC-3'	EcoR I
Nji UDGX Forward	*Nonomuraea jiangxiensis*	5'-CGCGGATCCATGACCAAGGATGGC-3'	BamH I
Nji UDGX Reverse	*Nonomuraea jiangxiensis*	5'-CCGGAATTCGCGTGATTTCCTGAGTTTCC-3'	EcoR I
Svi UDGX Forward	*Saccharomonospora viridis*	5'-CGCGGATCCATGGTACGTTATGAAGGCG-3'	BamH I
Svi UDGX Reverse	*Saccharomonospora viridis*	5'-CCCAAGCTTCAACGCCGCCGCAAC-3'	Hind III
Bgl UDGX Forward	*Burkholderia gladioli*	5'-GGAATTCCATATGGAGTTAACCGATGCCATCCCA-3'	Nde I
Bgl UDGX Reverse	*Burkholderia gladioli*	5'-CCGCTCGAGCCGGGAATCTTCGTC-3'	Xho I
Mgo UDGX1 Forward	*Mycolicibacterium goodii*	5'-GCGGATCCATGGCCGGGGCTCAGGAC-3'	BamH I
Mgo UDGX1 Reverse	*Mycolicibacterium goodii*	5'-CGCGAATTCCGGGGCCGGATACCACTGCA-3'	EcoR I
Mgo UDGX2 Forward	*Mycolicibacterium goodii*	5'-GCGGATCCATGGCCGGCGCTCAAGAT-3'	BamH I
Mgo UDGX2 Reverse	*Mycolicibacterium goodii*	5'-CTCGAATTCCGAGGCCGGATTCCGCTACA-3'	EcoR I

Table 4.1 Primers to amplify cDNA sequences of UDGX homologs by PCR

PCR-based site-directed mutagenesis was performed using the pET21a-udgx recombinant plasmid as the template similarly as previously described [31]. The PCR reaction mixture (40 μL) consisted of 20 ng DNA template, 50 nM of each primer pair carrying the desired mutations (Table 4.2), 200 μM each dNTP, 1x Phusion DNA polymerase buffer and 1 Unit of Phusion DNA polymerase. The PCR procedure included a pre-denaturation step at 98°C for 2 min; 25 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 10 s, annealing at 65°C for 20 s, and extension at 72°C for 7.5 min; and a final extension step at 72°C for 5 min. The PCR product was then treated with Dpn I and transformed into *E. coli* strain DH5α competent cells. Successfully mutated plasmid was confirmed by DNA sequencing. The double mutant D57E-H96N of Bgl UDGX was generated by two rounds of PCR-based site-directed mutagenesis.

Mutant	Primers
Bgl H96N	5'-CTTGAAATGCTTTACGGCGTTCGTAAGATACACTTTATCTC-3' 5'-GAGATAAAGTGTATCTTACGAACGCCGTAAAGCATTTCAAG-3'
Bgl D57E	5'-ATGATCGCTTGGTTGCTCGCCTACCAGCATC-3' 5'-GATGCTGGTAGGCGAGCAACCAAGCGATCAT-3'
Mgo1 H158R	5'-AATGAAGAACCTCTCCGCGATGTTGTGTTACCCGG-3' 5'-CCGGGTAACACAACATCGCGGAGAGGTTCTTCATT-3'
Mgo2 H158R	5'-GTGTAAAACCTCACCACGGTGTTGCGTGACCCG-3' 5'-CGGGTCACGCAACACCGTGGTGAGGTTTTACAC-3'
Msm R158A	5'-ACCTCGCCGGCATGCTGCGTGACGCGAA-3' 5'-TTCGCGTCACGCAGCATGCCGGCGAGGT-3'
Msm D200A	5'-GCGACACGCAGAGCGTCGACCAGGC-3' 5'-GCCTGGTCGACGCTCTGCGTGTCGC-3'

Table 4.2 Primers for mutagenesis

To express the UDGX homolog proteins, the recombinant plasmids were transformed into *E. coli* strain Rosetta competent cells. Culture incubation, IPTG-induced protein expression and purification were performed as described below. A single *E. coli* colony transformed with recombinant plasmid was cultured in 500 mL LB medium supplemented with 100 μg/mL Ampicillin at 37°C with shaking at 250 rpm until the optical density at 600 nm was over 0.6. After adding IPTG to a final concentration of 0.8 mM, the culture was grown at 22°C overnight. The *E. coli* cells were harvested at 5000 rpm with a JLA8.1000 rotor (Beckman Coulter) at 4°C for 20 min. The cell pellet was resuspended in 7 mL sonication buffer and then sonicated at 5 s on/ 5 s off cycle for 10 min on ice using Qsonica model Q125. The sonicated solution was then centrifuged at 12000 rpm with a JLA16.250 rotor (Beckman Coulter) at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded onto a 1 mL HisTrap FF column. After the sample loading, the column was washed with 15 mL of buffer A. The bound protein in the column was eluted with a linear gradient of 0-100% buffer B and collected in 1 mL fractions. The fractions identified by UV280 and SDS-PAGE were pooled and diluted 5-fold with 50 mM Tris-HCl (pH 7.5) buffer. The diluted solution was loaded onto a 1 mL HiTrap Q FF or HiTrap SP FF column. After the column was washed with 15 mL of buffer A, the bound protein in the column was eluted with a linear gradient of 10-100% buffer B and collected in 1 mL fractions. After identification, the fractions containing UDGX homolog proteins were pooled and concentrated. The protein concentration was determined by Nanodrop One (ThermoFisher Scientific). The protein solution was stored at -20°C in 50% glycerol.

C. Oligodeoxynucleotide substrates

Oligodeoxynucleotide substrates containing uracil (Figure 4.1A) were prepared as previously described [18]. Carboxyfluorescein (FAM) fluorescence-labelled single-stranded oligodeoxynucleotides containing deoxyuridine (U) and 1.5-fold molar excess complementary single-stranded oligodeoxynucleotides were mixed and incubated at 85°C for 3 min, followed by annealing to form duplex DNA substrates at room temperature for more than 30 min.

D. DNA crosslinking assay

DNA crosslinking assays for UDGX homologs and their mutants were performed at 37°C for 30 min in 20 μL reaction mixtures containing 100 nM uracil-containing DNA substrate, 100 nM ([E]:[S]=1:1) or 10000 nM enzyme [E]:[S]=100:1), 25 mM Tris-HCl buffer (pH 7.5), 1 mM EDTA and 1 mM DTT. The mixtures were heated at 95°C for 5 min after adding 5 μL SDS-PAGE loading buffer, followed by loading into 15% SDS-PAGE gel. Electrophoresis was conducted at 300 V for 30 min. Fluorescence-labelled oligodeoxynucleotides in the gel were visualized by a Typhoon FLA 7000 imager (GE Healthcare).

IV. Results and Discussion

Among hundreds of putative UDGX homologs, four proteins were first selected to screen the DNA crosslinking activity based on the sequence alignment of the three catalytic motifs, which were from *Saccharomonospora marina* (Sma UDGX), *Nonomuraea jiangxiensis* (Nji UDGX), *Saccharomonospora viridis* (Svi UDGX) and

Burkholderia gladioli (Bgl UDGX). As shown in Figure 4.1B, Sma UDGX, Nji UDGX and Svi UDGX possess the same key residues in the three motifs with Msm UDGX, which have been confirmed essential to the activities in previous studies, except for a different residue in the position where it's Q53 in the Msm UDGX. The different residue is valine, glutamic acid or methionine in Sma UDGX, Nji UDGX and Svi UDGX, respectively. Among the known key residues in the motifs, Bgl UDGX has two variant residues, D57 and H96 in the positions of E52 and N91 of Msm UDGX according to the sequence alignment.

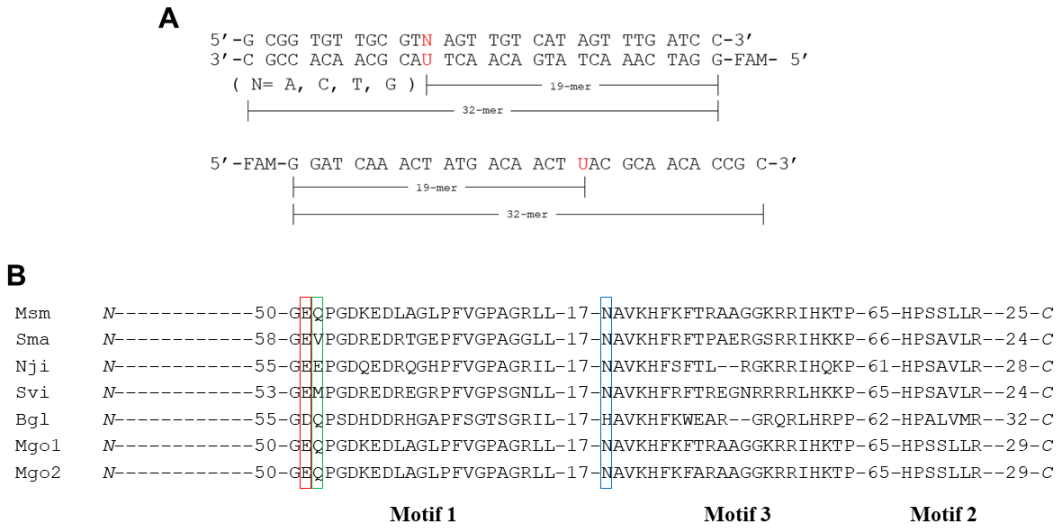

Figure 4.1 Substrates and sequence alignment. A. Sequences of uracil-containing DNA substrates (A/U, C/U, T/U and G/U base pair containing double-stranded DNA and single-stranded U DNA). **B.** Sequence alignment of UDGX homologs. Residues in the positions of E52, Q53 and N91 of Msm UDGX were highlighted. GenBank accession numbers are shown after the species names. Msm, *Mycobacterium smegmatis*, WP_011726794.1; Sma, *Saccharomonospora marina*, WP_009153389.1; Nji, *Nonomuraea jiangxiensis*, WP_090946880.1; Svi, *Saccharomonospora viridis*, WP_015785655.1; Bgl, *Burkholderia gladioli*, WP_036055803.1; Mgo1, *Mycobacterium goodii*, MBU8812906.1; Mgo2, *Mycobacterium goodii*, MBU8821111.1.

Results of DNA crosslinking assays with Sma UDGX, Nji UDGX and Svi UDGX showed that all three proteins could crosslink to the AP site produced after the uracil

excision in the uracil-DNA substrates, forming DNA-protein complex (Figure 4.2), in spite of the different crosslinking efficiencies. Sma UDGX and Nji UDGX exhibited their crosslinking abilities to all the five types of uracil-DNA substrates, including A/U, C/U, T/U, G/U base pairs and single-stranded U, especially the latter could convert all the C/U, T/U, G/U and single-stranded U substrates to crosslinking products when the E:S ratio was 100:1. The crosslinking efficiency of Nji UDGX on A/U base pair seemed lower than the other substrates probably due to the fact that A/U forms a normal Watson-Crick base pair whereas the other base pairs do not. Compared to Sma UDGX and Nji UDGX, the Svi UDGX showed robust crosslinking activity on T/U base pair and single-stranded U substrates, and weak activity on C/U and G/U base pairs, while no activity on A/U base pair was observed.

First of all, consistent with the previous analysis on the Q53A mutant of Msm UDGX, the substitution for the glutamine still retains the DNA crosslinking activity of the UDGX proteins, but the efficiency is significantly reduced compared to the wildtype Msm UDGX. This suggests that even though the glutamine does not take part in the uracil excision and AP site crosslinking steps directly, it likely plays a role in maintaining the activity of UDGX. Secondly, since the three UDGX homologs screened have different substitutions for the glutamine, the different crosslinking efficiencies seems reasonable. Nji UDGX showed the closest activity to Msm UDGX, which can be explained by the fact that glutamic acid (in Nji UDGX) has similar structural and chemical properties to glutamine (in Msm UDGX). Whereas, the residues valine (in Sma

UDGX) and methionine (in Svi UDGX) are obviously different from glutamine, which reduces the crosslinking activity.

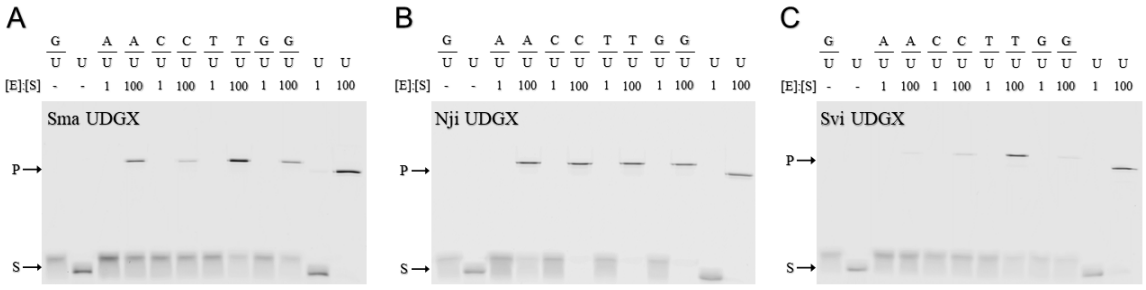

Figure 4.2 DNA crosslinking analysis of Sma UDGX, Nji UDGX and Svi UDGX. DNA crosslinking assays were performed as described in Materials and Methods. P indicates DNA-protein crosslinking products. S indicates uracil-DNA substrates. **A.** Crosslinking analysis of Sma UDGX. **B.** Crosslinking analysis of Nji UDGX. **C.** Crosslinking analysis of Svi UDGX.

From previous studies, we have known that the first residue in motif 3 of UDGX, N91 in the case of Msm UDGX, plays a critical role in the uracil excision via forming bidentate hydrogen bonds with the uracil base. In Bgl UDGX, this position is replaced by histidine (H96, Figure 4.1B), which might affect the activity of the protein. In addition, another key residue in motif 1, where it's E52 in Msm UDGX, is replaced by aspartic acid (D57, Figure 4.1B) in Bgl UDGX. It has been confirmed that the E52 in Msm UDGX is essential to the activity because of its interactions with the crosslinking site H109 and the uracil base. As expected, the DNA crosslinking analysis showed that Bgl UDGX had no crosslinking activity (Figure 4.3A). Then, a substitution mutant, H96N, was constructed. The results of DNA crosslinking analysis of the mutant did not exhibit any activity, either (Figure 4.3B). Finally, a double mutant of Bgl UDGX, D57E-H96N, was constructed and the crosslinking assay was performed with the double mutant, however, no crosslinking product was observed (Figure 4.3C). It has also been confirmed

that the two mutants as well as the wildtype Bgl UDGX have no uracil excision activity, either.

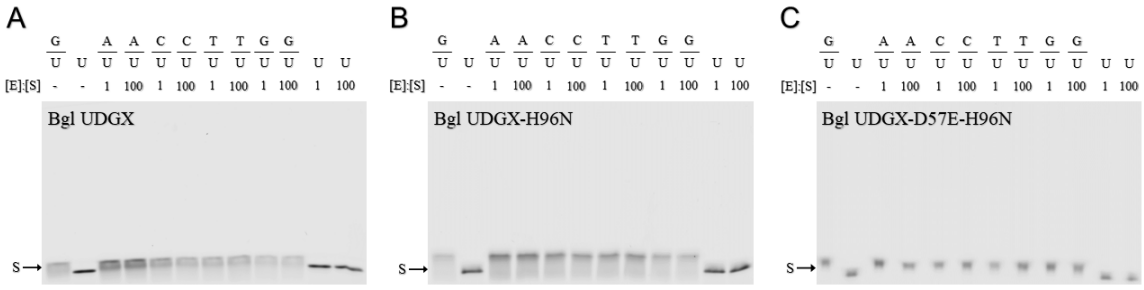

Figure 4.3 DNA crosslinking analysis of Bgl UDGX wildtype and mutants. DNA crosslinking assays were performed as described in Materials and Methods. S indicates uracil-DNA substrates. **A.** Crosslinking analysis of Bgl UDGX wildtype. **B.** Crosslinking analysis of Bgl UDGX-H96N mutant. **C.** Crosslinking analysis of Bgl UDGX-D57E-H96N mutant.

Since the key residues, which have been confirmed in previous studies, in the three motifs of Bgl UDGX D57E-H96N mutant are the same to the Msm UDGX, the loss of DNA crosslinking activity is probably caused by other residues out of the motifs. Thus, we aligned the whole sequence of Bgl UDGX with Msm UDGX; it has only 41.85% identity of the sequence, which increases the difficulty of searching for the target residues contributing to the activity. At last, two more UDGX homologs, Mgo UDGX1 and Mgo UDGX2 were selected by searching for similar sequences with the whole sequence of Msm UDGX on BLAST-P. As shown in Figure 4.1B, these two proteins share the same key residues in the three motifs with Msm UDGX. Furthermore, the whole sequences of Mgo UDGX1 and Mgo UDGX2 share 92.27% and 91.79% identity with the sequence of Msm UDGX, respectively (Figure 4.4).

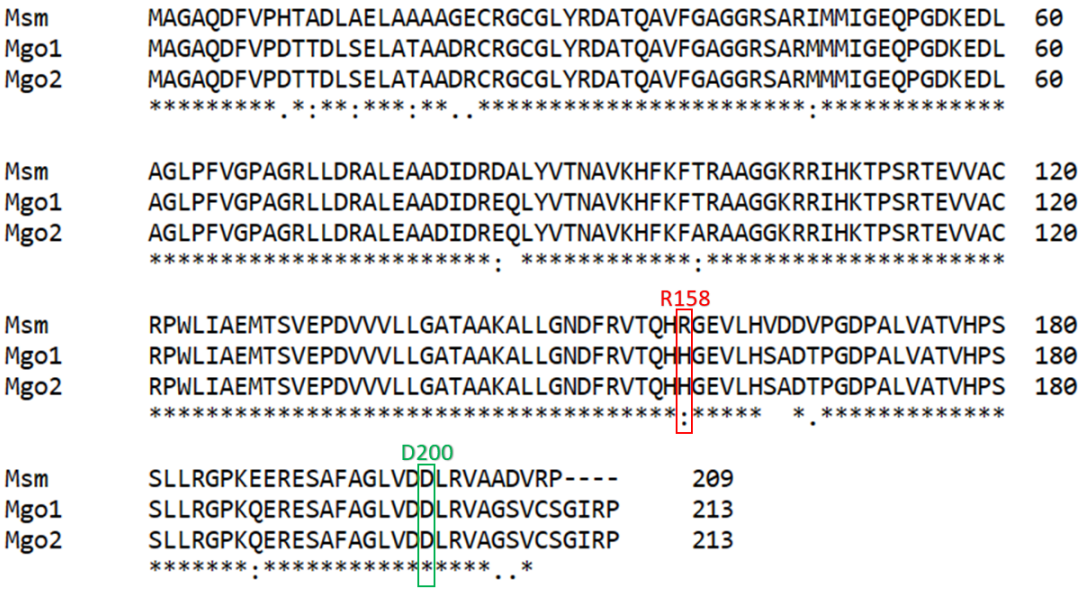

Figure 4.4 Whole sequence alignment of MsmUDGX, Mgo UDGX1 and Mgo UDGX2. The alignment was performed using the Clustal Omega tool from EMBL-EBI [32]. Msm, *Mycobacterium smegmatis*, WP_011726794.1; Mgo1, *Mycobacterium goodii*, MBU8812906.1; Mgo2, *Mycobacterium goodii*, MBU8821111.1. Residues R158 and D200 in Msm UDGX and their corresponding residues in Mgo UDGX1 and Mgo UDGX2 are highlighted.

The DNA crosslinking analyses of Mgo UDGX1 and Mgo UDGX2 showed surprising results that no crosslinking was observed with both proteins (Figure 4.5A and B). Besides, no uracil excision activities were detected by uracil excision assay for these two homologs. Combining the whole sequence alignment and the study of the structure of Msm UDGX, a potential residue was identified, which is the H158 in Mgo UDGX1 and Mgo UDGX2, where it's R158 in Msm UDGX. We constructed two mutants, Mgo UDGX1 H158R and Mgo UDGX2 H158R. The results of DNA crosslinking assays showed that there were crosslinking products with both the mutants on all the five types of uracil-DNA substrates (Figure 4.5C and D). Although the crosslinking efficiencies were not as robust as expected, the proteins indeed regained the DNA crosslinking ability, which suggests that the arginine is critical to the activity of UDGX.

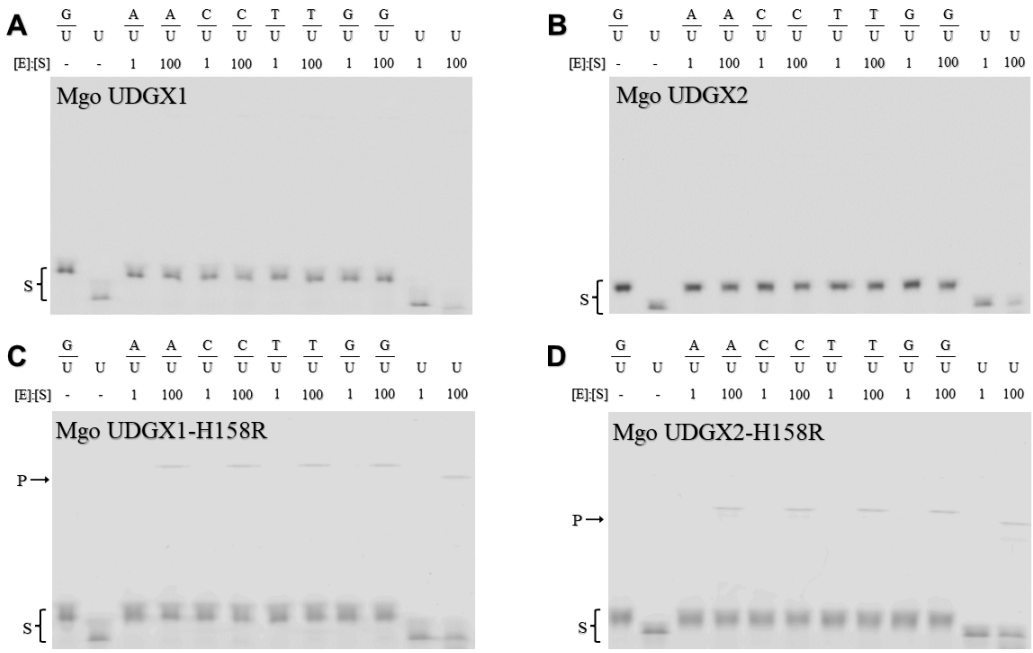

Figure 4.5 DNA crosslinking analysis of Mgo UDGX1 and Mgo UDGX2. DNA crosslinking assays were performed as described in Materials and Methods. P indicates DNA-protein crosslinking products. S indicates uracil-DNA substrates. **A.** Crosslinking analysis of Mgo UDGX1. **B.** Crosslinking analysis of Mgo UDGX2. **C.** Crosslinking analysis of Mgo UDGX1-H158R mutant. **D.** Crosslinking analysis of Mgo UDGX2-H158R mutant.

In the crystal structure of Msm UDGX, we find that the distance between the R158 and D200 is 2.7 Å, which suggests a salt bridge interaction could be formed by these two residues (Figure 4.6A). Interestingly, as we already know that UDGX is most closely related to family 4 UDGa in the UDG superfamily, a potentially similar interaction is also found in Tth UDGa, between the residues R139 and D177 (Figure 4.6B). To further verify the importance of this salt bridge to the activity, two mutants of Msm UDGX, R158A and D200A, were constructed. The DNA crosslinking analyses showed that both the mutants lost their crosslinking activities (Figure 4.6C and D). Though R158 and D200 are not located in the catalytic motifs, the salt bridge formed by these two residues seems to be essential to the activity, probably playing a critical role in maintaining the active

structure of UDGX. The residue R158 is located between motif 2 and motif 3, whereas D200 is in a helix at the C-terminal end of the protein. Therefore, the salt bridge is likely to support the position of motif 2 and motif 3 in the catalytic center and to stabilize the structure of the C-terminal helix in UDGX.

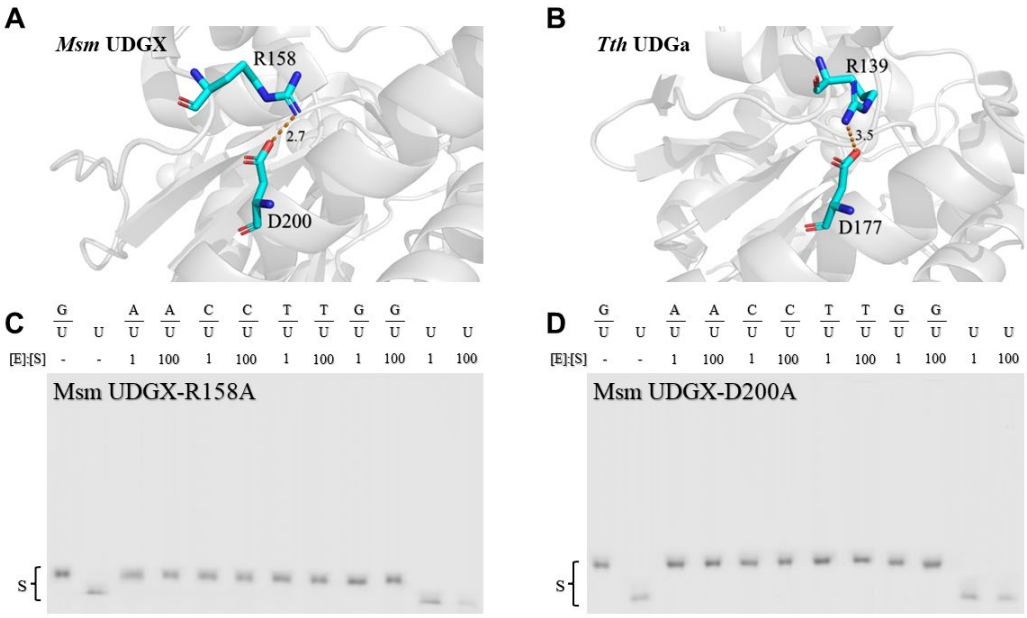

Figure 4.6 Role of R158 and D200 in Msm UDGX. DNA crosslinking assays were performed as described in Materials and Methods. S indicates uracil-DNA substrates. Distances in the crystal structure are shown in Å. **A.** Close-up view of the salt bridge formed by R158 and D200 in the crystal structure of Msm UDGX (PDB code 6IO9). **B.** Close-up view of the salt bridge formed by R139 and D177 in the crystal structure of Tth UDGa (PDB code 1UI1). **C.** Crosslinking analysis of Msm UDGX-R158A mutant. **D.** Crosslinking analysis of Msm UDGX-D200A mutant.

Since *Mycobacterium goodii* species was first isolated from *Mycobacterium smegmatis* species in 1999 [33], it has become a rapidly growing pathogen to mammals and humans, which could cause a series of infectious diseases [34]. Taking the close relationship between *M. goodii* and *M. smegmatis* into account, the loss of the DNA crosslinking activity of Mgo UDGX perhaps is a result of evolution from Msm UDGX.

As it's known, UDGX can crosslink to DNA strands, which could interfere the replication and transcription of DNA and even cause the cell death if the protein is not released from DNA. As a rapidly growing pathogen, *M. goodii* needs to complete the replication and transcription fast, thus, such a DNA-crosslinking protein could be harmful to the organism. As a consequence, the salt bridge is lost because of the mutation from arginine to histidine, so that the protein lost its DNA crosslinking activity to allow the fast bacterial proliferation.

In conclusion, the DNA crosslinking activities of six UDGX homologs were screened. Despite the different residues in motif 1, Sma UDGX, Nji UDGX and Svi UDGX still exhibited reduced DNA crosslinking activities. There is no crosslinking activity detected of Bgl UDGX, Mgo UDGX1 and Mgo UDGX2. In addition, even though all the key residues known in the motifs of Bgl UDGX were mutated back to the same with Msm UDGX, the mutants still could not crosslink to the DNA substrates. However, when H158 of Mgo UDGX1 and Mgo UDGX2 were substituted by arginine, the mutants regained part of the crosslinking activity. It seems that R158 in Msm UDGX forms a salt bridge with D200, which is necessary to the activity.

V. Reference

[1] T. Lindahl et al, "DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli," The Journal of Biological Chemistry, vol. 252, (10), pp. 3286-3294, 1977. DOI: 10.1016/S0021-9258(17)40386-3.

[2] L. Aravind and E. V. Koonin, "The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates," Genome Biology, vol. 1, (4), pp. RESEARCH0007, 2000. DOI: 10.1186/gb-2000-1-4-research0007.

[3] L. H. Pearl, "Structure and function in the uracil-DNA glycosylase superfamily," Mutation Research/DNA Repair, vol. 460, (3-4), pp. 165-181, 2000. DOI: 10.1016/s0921-8777(00)00025-2.

[4] P. Gallinari and J. Jiricny, "A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase," Nature, vol. 383, (6602), pp. 735-738, 1996. DOI: 10.1038/383735a0.

[5] U. Hardeland et al, "The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs," Nucleic Acids Research, vol. 31, (9), pp. 2261-2271, 2003. DOI: 10.1093/nar/gkg344.

[6] D. Lee et al, "A structural determinant in the uracil DNA glycosylase superfamily for the removal of uracil from adenine/uracil base pairs," Nucleic Acids Research, vol. 43, (2), pp. 1081-1089, 2015. DOI: 10.1093/nar/gku1332.

[7] H. Lee et al, "Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase," The Journal of Biological Chemistry, vol. 285, (53), pp. 41483-41490, 2010. DOI: 10.1074/jbc.M110.150003.

[8] L. Dong et al, "Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase," DNA Repair, vol. 7, (12), pp. 1962-1972, 2008. DOI: 10.1016/j.dnarep.2008.08.006.

[9] E. Lutsenko and A. S. Bhagwat, "The role of the Escherichia coli mug protein in the removal of uracil and 3,N(4)-ethenocytosine from DNA," The Journal of Biological Chemistry, vol. 274, (43), pp. 31034-31038, 1999. DOI: 10.1074/jbc.274.43.31034.

[10] M. Saparbaev and J. Laval, "3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase," Proceedings of the National Academy of Sciences - PNAS, vol. 95, (15), pp. 8508-8513, 1998. DOI: 10.1073/pnas.95.15.8508.

[11] M. Saparbaev, J. C. Mani and J. Laval, "Interactions of the human, rat, Saccharomyces cerevisiae and Escherichia coli 3-methyladenine-DNA glycosylases with DNA containing dIMP residues," Nucleic Acids Research, vol. 28, (6), pp. 1332-1339, 2000. DOI: 10.1093/nar/28.6.1332.

[12] E. Moe et al, "The crystal structure of mismatch-specific uracil-DNA glycosylase (MUG) from Deinococcus radiodurans reveals a novel catalytic residue and broad substrate specificity," The Journal of Biological Chemistry, vol. 281, (1), pp. 569-577, 2006. DOI: 10.1074/jbc.M508032200.

[13] K. A. Haushalter et al, "Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors," Current Biology, vol. 9, (4), pp. 174-185, 1999. DOI: 10.1016/s0960-9822(99)80087-6.

[14] P. Pang et al, "SMUG2 DNA glycosylase from Pedobacter heparinus as a new subfamily of the UDG superfamily," Biochemical Journal, vol. 474, (6), pp. 923-938, 2017. DOI: 10.1042/BCJ20160934.

[15] H. Nilsen et al, "Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase," The EMBO Journal, vol. 20, (15), pp. 4278-4286, 2001. DOI: 10.1093/emboj/20.15.4278.

[16] R. J. Boorstein et al, "Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1," The Journal of Biological Chemistry, vol. 276, (45), pp. 41991-41997, 2001. DOI: 10.1074/jbc.M106953200.

[17] B. Kavli et al, "hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup," The Journal of Biological Chemistry, vol. 277, (42), pp. 39926-39936, 2002. DOI: 10.1074/jbc.M207107200.

[18] R. Mi et al, "Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG," Journal of Molecular Biology, vol. 385, (3), pp. 761-778, 2009. DOI: 10.1016/j.jmb.2008.09.038.

[19] J. Li et al, "Identification of a prototypical single-stranded uracil DNA glycosylase from Listeria innocua," DNA Repair, vol. 57, pp. 107-115, 2017. DOI: 10.1016/j.dnarep.2017.07.001.

[20] A. Koulis et al, "Uracil-DNA glycosylase activities in hyperthermophilic micro-organisms," FEMS Microbiology Letters, vol. 143, (2-3), pp. 267-271, 1996. DOI: 10.1111/j.1574-6968.1996.tb08491.x.

[21] M. Sandigursky and W. A. Franklin, "Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes," Current Biology, vol. 9, (10), pp. 531-534, 1999. DOI: 10.1016/s0960-9822(99)80237-1.

[22] V. Starkuviene and H. Fritz, "A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus," Nucleic Acids Research, vol. 30, (10), pp. 2097-2102, 2002. DOI: 10.1093/nar/30.10.2097.

[23] B. Xia et al, "Correlated mutation in the evolution of catalysis in uracil DNA glycosylase superfamily," Scientific Reports, vol. 7, (1), pp. 45978, 2017. DOI: 10.1038/srep45978.

[24] Q. Gan et al, "Characterization of a Family IV uracil DNA glycosylase from the hyperthermophilic euryarchaeon Thermococcus barophilus Ch5," International Journal of Biological Macromolecules, vol. 146, pp. 475-481, 2020. DOI: 10.1016/j.ijbiomac.2019.12.202.

[25] A. A. Sartori et al, "A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site," The EMBO Journal, vol. 21, (12), pp. 3182-3191, 2002. DOI: 10.1093/emboj/cdf309.

[26] B. Xia et al, "Specificity and catalytic mechanism in family 5 uracil DNA glycosylase," The Journal of Biological Chemistry, vol. 289, (26), pp. 18413-18426, 2014. DOI: 10.1074/jbc.M114.567354.

[27] H. Lee, B. N. Dominy and W. Cao, "New family of deamination repair enzymes in uracil-DNA glycosylase superfamily," The Journal of Biological Chemistry, vol. 286, (36), pp. 31282-31287, 2011. DOI: 10.1074/jbc.M111.249524.

[28] P. B. Sang et al, "A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, vol. 43, (17), pp. 8452-8463, 2015. DOI: 10.1093/nar/gkv854.

[29] W. Ahn et al, "Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision," Nature Chemical Biology, vol. 15, (6), pp. 607-614, 2019. DOI: 10.1038/s41589-019-0289-3.

[30] J. Tu et al, "Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation," Nature Chemical Biology, vol. 15, (6), pp. 615-622, 2019. DOI: 10.1038/s41589-019-0290-x.

[31] C. L. Fisher and G. K. Pei, "Modification of a PCR-based site-directed mutagenesis method," Biotechniques, vol. 23, (4), pp. 570-574, 1997. DOI: 10.2144/97234bm01.

[32] F. Madeira et al, "Search and sequence analysis tools services from EMBL-EBI in 2022," Nucleic Acids Research, vol. 50, (W1), pp. W276-W279, 2022. DOI: 10.1093/nar/gkac240.

[33] B. A. Brown et al, "Mycobacterium wolinskyi sp. nov. and Mycobacterium goodii sp. nov., two new rapidly growing species related to Mycobacterium smegmatis and associated with human wound infections: a cooperative study from the International Working Group on Mycobacterial Taxonomy," International Journal of Systematic and Evolutionary Microbiology, vol. 49, (4), pp. 1493-1511, 1999. DOI: 10.1099/00207713-49-4-1493.

[34] N. M. Salas and N. Klein, "Mycobacterium goodii: An Emerging Nosocomial Pathogen: A Case Report and Review of the Literature," Infectious Diseases in Clinical

Practice (Baltimore, Md.), vol. 25, (2), pp. 62-65, 2017. DOI:

10.1097/IPC.0000000000000428.

CHAPTER FIVE

RESEARCH SIGNIFICANCE AND CONCLUDING REMARKS

DNA base damage is common due to various exogenous and endogenous factors, including base oxidation, methylation, deamination and so on [1-6]. Damaged bases can be repaired by multiple pathways, such as direct reversal of damaged DNA base, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and so on [1]. To initiate the BER, DNA glycosylases are required to excise the damaged base off from DNA strands first [7]. As one of the main DNA glycosylases, uracil-DNA glycosylase (UDG) plays a critical role in the removal of a series of DNA base lesions, such as uracil, xanthine, hypoxanthine and so on [8, 9]. Since the first UDG, uracil-*N*-glycosylase (UNG) was found in 1977 [10], there have been seven families of enzymes identified in the UDG superfamily. As a new member in the superfamily, UDGX exhibits a completely different characteristic from the other six families of UDG enzymes, which is the ability to covalently crosslink to the apurinic/apyrimidinic site (AP site) formed after the uracil base is excised from DNA via a histidine residue (H109 in Msm UDGX) [11-13]. In this work, a systematic analysis was performed on this unique DNA-crosslinking UDG, to investigate the structural and functional features, catalytic mechanism, and even distribution and activity discrepancies among species.

First, we found that, in the structure of UDGX, the motifs in the catalytic center, especially motif 1 and motif 3, are coupled by a series of interactions between residues. E52 in motif 1 could not only interact with the uracil base, but also interact with H109 to

fulfill its role in the uracil excision and DNA crosslinking. D56, D59 in motif 1 and R107 in motif 3 form tripartite salt bridges to anchor the H109-containing extended loop in motif 3 to the proximity of the active site to ready it for crosslinking reaction. In addition, K94 and N91, which are both located in motif 3, could form intra-motif interaction to support the uracil recognition and excision by N91. Based on structural coupling, the two functions of UDGX, uracil excision and AP site crosslinking are also tightly coupled, which is confirmed by mutational and kinetic analyses.

Then, a study focused on the crosslinking site of UDGX, H109, was performed. The functional coupling of uracil excision and DNA crosslinking is further confirmed by the reduced uracil excision efficiencies of substitution mutants for H109. In addition, though most of the substitution mutants for H109 lost their DNA crosslinking activity, a special mutant, H109E, was found to retain the crosslinking activity [14]. By biochemical and kinetic analyses on the H109E mutant, it's clarified that although the crosslinking between the glutamate residue and the AP site is via an ester bond, which is different from the C-N bond formed by UDGX wildtype, the crosslinking is still coupled with the prior uracil excision step.

Finally, we screened the DNA crosslinking activity of six UDGX homologs identified according to the sequence similarity. Consistent with the mutational analysis on Msm UDGX, a reduced crosslinking activity was observed of Sma UDGX, Nji UDGX and Svi UDGX, which have a substitution naturally for Q53 in motif 1 of Msm UDGX. By mutational analysis on Bgl UDGX, we found that not only the key residues we have identified in the catalytic motifs, some other residues outside motifs might affect the

activities of UDGX. By studying two homologs with highly similar sequences to the Msm UDGX, it's found that a salt bridge formed by R158 and D200 in Msm UDGX is essential to the activity.

In summary, this work identifies some key residues that maintain the structure and complete the uracil excision and DNA crosslinking functions of UDGX. Furthermore, a catalytic mechanism with structural and functional coupling features is proposed for this bifunctional enzyme. Phylogenetic analysis and biochemical study indicate that UDGX homologs only distribute in bacteria and their DNA crosslinking activity varies, which may shed light on the evolutionary process of this protein in bacterial species. Additionally, because of the specific ability to recognize and crosslink uracil-DNA, UDGX has been developed as a tool to detect and visualize uracil bases in DNA recently [15-17]. With our studies, a more comprehensive understanding of the connection between the structure and functions of UDGX is provided, which may help to develop more applications of this special DNA-crosslinking UDG. Last but not least, the discovery and research of UDGX have broadened the potential biological functions and significances of enzymes from the UDG superfamily.

Reference

[1] N. Chatterjee and G. C. Walker, "Mechanisms of DNA damage, repair, and mutagenesis," Environmental and Molecular Mutagenesis, vol. 58, (5), pp. 235-263, 2017. DOI: 10.1002/em.22087.

[2] L. J. Marnett, "Oxyradicals and DNA damage," Carcinogenesis, vol. 21, (3), pp. 361-370, 2000. DOI: 10.1093/carcin/21.3.361.

[3] H. Ikehata and T. Ono, "The mechanisms of UV mutagenesis," Journal of Radiation Research, vol. 52, (2), pp. 115-125, 2011. DOI: 10.1269/jrr.10175.

[4] G. T. Pauly and R. C. Moschel, "Mutagenesis by O(6)-methyl-, O(6)-ethyl-, and O(6)-benzylguanine and O(4)-methylthymine in human cells: effects of O(6)-alkylguanine-DNA alkyltransferase and mismatch repair," Chemical Research in Toxicology, vol. 14, (7), pp. 894-900, 2001. DOI: 10.1021/tx010032f.

[5] G. Fronza and B. Gold, "The biological effects of N3-methyladenine," Journal of Cellular Biochemistry, vol. 91, (2), pp. 250-257, 2004. DOI: 10.1002/jcb.10698.

[6] H. E. Krokan, F. Drabløs and G. Slupphaug, "Uracil in DNA -occurrence, consequences and repair," Oncogene, vol. 21, (58), pp. 8935-8948, 2002. DOI: 10.1038/sj.onc.1205996.

[7] D. O. Zharkov, "Base excision DNA repair," Cell. Mol. Life Sci, vol. 65, (10), pp. 1544-1565, 2008. DOI: 10.1007/s00018-008-7543-2.

[8] L. H. Pearl, "Structure and function in the uracil-DNA glycosylase superfamily," Mutation Research/DNA Repair, vol. 460, (3-4), pp. 165-181, 2000. DOI: 10.1016/s0921-8777(00)00025-2.

[9] D. Cortázar et al, "The enigmatic thymine DNA glycosylase," DNA Repair, vol. 6, (4), pp. 489-504, 2007. DOI: 10.1016/j.dnarep.2006.10.013.

[10] T. Lindahl et al, "DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli," The Journal of Biological Chemistry, vol. 252, (10), pp. 3286-3294, 1977. DOI: 10.1016/S0021-9258(17)40386-3.

[11] P. B. Sang et al, "A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, vol. 43, (17), pp. 8452-8463, 2015. DOI: 10.1093/nar/gkv854.

[12] W. Ahn et al, "Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision," Nature Chemical Biology, vol. 15, (6), pp. 607-614, 2019. DOI: 10.1038/s41589-019-0289-3.

[13] J. Tu et al, "Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation," Nature Chemical Biology, vol. 15, (6), pp. 615-622, 2019. DOI: 10.1038/s41589-019-0290-x.

[14] Q. Jia et al, "Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability," DNA Repair, vol. 97, pp. 103008, 2021. DOI: 10.1016/j.dnarep.2020.103008.

[15] L. Jiang et al, "UdgX-mediated uracil sequencing at single-nucleotide resolution," J. Am. Chem. Soc., vol. 144, (3), pp. 1323-1331, 2022. DOI: 10.1021/jacs.1c11269.

[16] J. A. Stewart, G. Schauer and A. S. Bhagwat, "Visualization of uracils created by APOBEC3A using UdgX shows colocalization with RPA at stalled replication forks," Nucleic Acids Research, vol. 48, (20), pp. e118, 2020. DOI: 10.1093/nar/gkaa845.

[17] M. Datta et al, "Development of mCherry tagged UdgX as a highly sensitive molecular probe for specific detection of uracils in DNA," Biochemical and Biophysical Research Communications, vol. 518, (1), pp. 38-43, 2019. DOI: 10.1016/j.bbrc.2019.08.005.