12-2014

Specificity and Catalytic Mechanism of DNA Glycosylases in UDG Superfamily

Bo Xia
Clemson University, bxia@g.clemson.edu

SPECIFICITY AND CATALYTIC MECHANISM OF DNA GLYCOSYLASES IN UDG SUPERFAMILY

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
Bo Xia
December 2014

Accepted by:
Dr. Weiguo Cao, Committee Chair
Dr. Abbott Bert
Dr. Brian Dominy
Dr. Liangjiang Wang

ABSTRACT

DNA can be damaged by several kinds of endogenous and exogenous reactive nitrogen species. Under nitosative stress, uracil (U), hypoxanthine (I), xanthine (X) and oxanine (O) are four major deaminated DNA bases derived from cytosine (C), adenine (A) and guanine (G) respectively. To repair this type of DNA damage, several different repair pathways are involved.

My dissertation work mainly focused on the uracil-DNA glycosylase (UDG) superfamily, which includes several groups of enzymes that recognize the damaged DNA bases and initiate the base excision repair (BER) pathway, one of the most important repair pathways to deal with deaminated DNA bases. Chapter 1 is a general introduction of different kinds of DNA damage and their corresponding repair pathways. Chapter 2 presents a detailed functional and structural analysis of family 5 UDGb from *Thermus thermophilus* HB8 in order to understand the specificity and catalytic mechanism of family 5 UDGb. Chapter 3 describes the biochemical properties and catalytic mechanism of family 4 UDGa from *Thermus thermophilus* HB8. A special double mutant has been identified with increased enzyme activity compared to single mutants. Chapter 4 is about a potential new group of enzymes within the UDG superfamily. Members from this new group of enzymes showed robust xanthine DNA glycosylase activities with unique catalytic mechanism and protein sequences. In summary, these functional and structural analyses provide new insights into substrate specificity and catalytic mechanism of UDG superfamily.

DEDICATION

This dissertation is dedicated to my wife, my parents and my family.

ACKNOWLEDGMENTS

First, I would like to deeply thank my advisor, Dr. Weiguo Cao who is supporting and advising me when I pursue my Ph.D. degree in Clemson University. He always tried to improve my scientific insight and his suggestion, encouragement and support was really helpful to my projects.

I would also like to sincerely thank my committee members (Dr. Abbott Albert, Dr. Brian Dominy, and Dr. Liangjiang Wang). Their support, assistance and discussion are very important and helpful for my research projects.

I would like to thank all Dr. Weiguo Cao current and previous lab members as well for their help and discussion (Dr. Guang-chen Fang, Dr. Wei Li, Dr. Ye Yang, Dr. Hyun-Wook Lee, Dr. Sung-Hyun Park, Dong-Hoon Lee, Jing Li and Celeste Jilich).

I would like to thank our collaborators, Dr. Brian N. Dominy and his lab members, Yinling Liu and Dr. Allyn R. Brice, and Dr. Liangjiang Wang and his lab members, Jose Guevara for their help on molecular modeling and bioinformatics analysis.

Last but not the least, I would like to thank my wife, my parents and my family for their love and support.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF TABLE

Table Page

LIST OF FIGURES

List of Figure (Continued)

Figure Page

List of Figure (Continued)

Figure Page

ABBREVIATIONS

AER, Alternative excision repair

APE1, Human AP endonuclease

AP site, Apurinic/apyrimidinic site

BER, Base excision repair

DSB, DNA double-strand breaks

DNA, Deoxyribonucleic acid

Endo V, Endonuclease V (Nfi)

HDG, Hypoxanthine DNA glycosylase

iNOS, Inducible nitric oxide synthase

MUG, Mismatch-specific uracil DNA glycosylase

NER, Nucleotide excision repair

NHEJ, Non-homologous end joining

OGG, 8-Oxoguanine DNA glycosylase

RNS, Reactive nitrogen species

ROS, Reactive oxygen species

SMUG1, Single strand-selective monofunctional uracil-DNA glycosylase

SSBs, Single strand breaks

TCR, Transcription-coupled repair

TDG, Thymine DNA glycosylase

UDG, Uracil DNA glycosylase

XDG, Xanthine DNA glycosylase

CHAPTER ONE

DEAMINATION OF DNA BASES AND URACIL-DNA GLYCOSYLASE

I. Introduction

Deoxyribonucleic acid (DNA) is a molecule that carries the majority of genetic information in all known living organisms. Coupled with proteins and RNA, DNA plays an important role in the central dogma. Although it is protected by many kinds of proteins and reducing environment in the cells, DNA is easily to be damaged. The changes in the genetic information might cause many problems and diseases in all known forms of life, such as cancer and Alzheimer's disease (13-16). To date, various sources of DNA mutagens and damage have been identified, deamination of DNA bases has been shown to be a very important type of DNA mutation, and uracil-DNA glycosylase has been suggested to be one of the major repair enzymes responsible for this kind of DNA damage (3,17). In this chapter, an overview of DNA damage and DNA repair pathways will be provided followed by specific introduction on the deamination of DNA bases, uracil-DNA glycosylase superfamily and base excision repair pathway.

II. DNA damages

A. Oxidative damage

The major sources of oxidative damage come from reactive oxygen species (ROS), a type of reactive molecule contains oxygen. In human, ROS is believed to be involved in the development of several diseases, such as cancer (18), Parkinson’s disease and

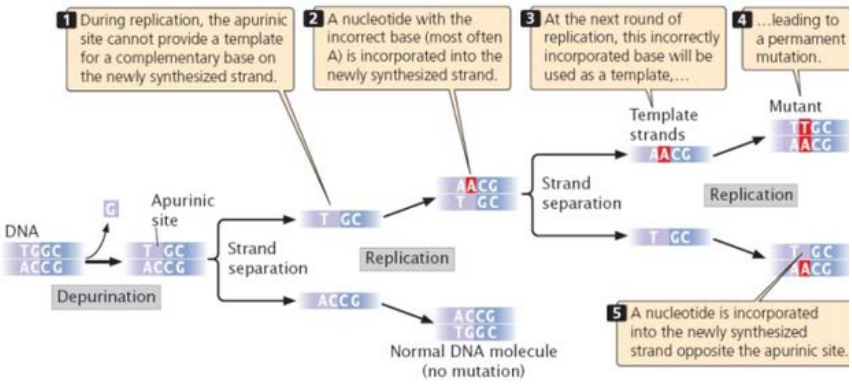


Figure 1.1 Mutation caused by depurination. After DNA bases undergo depurination, it leaves apurinic site. During replication, the apurinic site cannot provide a template for newly synthesized DNA strand. A nucleotide will be randomly incorporated into DNA and in most cases as adenine. This process will generate a mutation. Reproduced from (7) with permission.

Alzheimer's disease (19,20), although in some other cases, it can serve as a cell signaling factor and involve in the host defense towards pathogens. Hydrogen peroxide (H_2O_2), superoxide and hydroxyl radical are main contributors to DNA oxidative damage in cells caused by ROS (19,21). They can be produced during regular cell metabolism. For example, H_2O_2 is converted from superoxide that comes from the mitochondria (22). Since ROS is highly reactive agent, it can cause several kinds of DNA base damage in cells including 8-hydroxyguanine (8-oxoG), 8-hydroxydeoxyguanosine (8-oxodG) and 5-hydroxymethyluracil (23,24). The frequencies of mutations generated by oxidative stress could be as high as 10000 per day per cell in human (25). Besides the high rates of generations, both 8-oxoG and 8-oxodG are highly mutagenic (26), in which these non-canonical bases have the same tendency to pair with both adenine and cytosine (27,28). If these damaged bases are not being repaired, mutations will accumulate in the genome.

B. Depurination of DNA bases

Depurination of DNA is a process that the β-N-glycosidic bond of deoxyadenosine or deoxyguanosine is hydrolytically cleaved, generating an apurinic site (29). Since purine

is a good leaving group in terms of its N9 position, its hydrolytic rate is much higher compared to depyrimidination, although they share similar mechanism. According to previous studies, about 2000-10000 purines are lost in this way every day in a human cell (30). This estimation is confirmed later by other researchers (31). During replication, loss of DNA bases in one DNA strand will cause DNA polymerase randomly incorporate one nucleotide into that position, and in most cases, it is A instead of G, suggesting a transition mutation is generated (7) (Figure 1.1). It has been suggested that depurination might play an important role in the initial development of neoplasm (32).

C. UV light

Ultraviolet (UV) light has a wavelength shorter than that of visible light but longer than X-rays. A subtype of UV light, Ultraviolet B (UVB) is very important for human health. Exposure of UVB could induce the production of vitamin D, a well-known vitamin involved in several metabolism and immunity processes (33). However, too much UVB exposure might cause direct DNA damage (34). The resulting pyrimidine dimer (thymine or cytosine dimers) could be removed by nucleotide excision repair pathway, but unrepaired dimer is mutagenic (35).

D. Deamination of DNA bases and nitrosative stress

Nitrosative stress is usually coupled with oxidative stress, and the main component is reactive nitrogen species (RNS). RNS usually works together with ROS to fulfill its functions in terms of anti-bacteria. They are derived from nitric oxide and superoxide by inducible nitric oxide synthase 2 (NOS2 or iNOS) and NADPH oxidase in human cells (36). The gene of iNOS is primarily expressed in macrophages, and its expression and

production of nitric oxide is inducible by cytokines and microbial products, in most cases, by interferon-gamma and lipopolysaccharide (LPS) respectively (37). Nitric oxide is not only an important part of immune response towards pathogens, it can also serve as a cell to cell signal molecule (38). The nitric oxide produced by eNOS (endothelial nitric oxide synthase) in endothelial cells, can activate guanylate cyclase in smooth muscle cells near to the endothelial cells, and will cause vasodilation of the blood muscle (39). It has been suggested about 1 mM NO is produced per body per day in humans, indicating NO is an important secondary massager molecule (40). However, long-term exposure to nitric oxide and nitrous acid (HNO_2) (generated by nitric oxide when it react with oxygen and water) will cause severe problems to the cells, especially, in the long term inflammation in which the amount of nitric oxide will increase by about 10- to 100-fold (41). The stress coming from RNS/ROS might cause DNA mutations and alterations in microenvironment. Studies have indicated that an inflammatory environment and nitric oxide might promote proliferation, survival and metastasis of tumor cells (42,43). In addition, dinitrogen trioxide (N_2O_3), the anhydride derived from nitrous acid (HNO2), is highly reactive to DNA bases including cytosine, adenine and guanine (44), and the deamination of DNA bases is the primary DNA damage caused by RNS stress.

Uracil, hypoxanthine, xanthine and oxanine are corresponding deaminated products from cytosine, adenine and guanine (Figure1.2) (45). The amino group to keto group conversion will completely change the hydrogen bond properties of DNA bases, and these alterations in base pair preference will result in mutations in the genome. For example, the product from deamination of cytosine is uracil, which usually pair with adenine in RNA or DNA. Although different organisms use different ways to prevent uracil to incorporate into

DNA, the uracil generated from deamination of cytosine cannot be simply removed by depletion of dUTP from nucleotide pool (46). The resulting G:C to A:T transition might cause serious problem if it happened in a protein coding region (47,48). Deamination of adenine serves in a similar way as deamination of uracil. The resulting deamination product hypoxanthine has the potential to pair with all four regular bases, although it prefers to pair with cytosine (49,50), suggesting hypoxanthine is also highly mutagenic in DNA. Different from cytosine and adenine, there are two major products from deamination of guanine, which are xanthine and oxanine (40,51). Both of them are proved to be mutagenic if unrepaired. Studies have shown that large replacement of guanine to xanthine and oxanine in DNA will cause the helix instability in double-strand DNA (52). In addition, xanthine and oxanine prefer to pair with T and C during DNA replication, suggesting these two kinds of DNA bases might cause G:C to A:T transition (53,54). It is known that under exposure of reactive nitrous species, such as nitric oxide, large amount of xanthine can be detected in human cells (55). Defects in the nucleotide metabolic pathway will also result in the accumulation of them in the genome (56).

E. Other kinds of DNA damage (alkylation and DNA double-strand breaks)

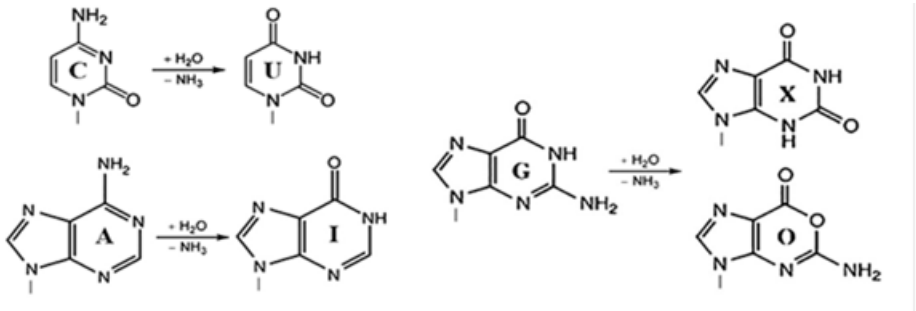


Figure 1.2 Schematic representation of DNA deamination of cytosine, adenine and guanine to uracil, hypoxanthine and xanthine/oxanine, respectively (9).

There are many other kinds of DNA damage, such as alkylation and DNA double-strand breaks which will also affect the integrity of the genome. Alkylation is a process including transferring an alkyl group from one molecular to another, in terms of DNA, this usually causes the formation of 7-methyguanine, 1-methyadenine and 6-O-Methylguanine. Since all these mutated bases are highly mutagenic, these chemicals are used to treat several different kinds of cancers. Although some of the alkylating agents are carcinogenic towards normal human cells, the drugs are more toxic to cancer cells during cell division (57).

Another popular type of DNA damage is double-strand breaks (DSBs) which might cause genome rearrangement, one of the most deleterious DNA damage (6). DSBs could be induced by several reasons (Figure 1.3). Firstly, exogenous insults from ionizing radiation (IR) or certain chemicals could cause severe DSBs (58). Coupled with other types of DNA damage introduced by ionizing radiation, low-level of radiation is enough to generate deadly DNA damage that will cause neoplasm and aging (59). On the other hand, DSBs can be also caused by free radicals that generated by regular metabolic pathways,

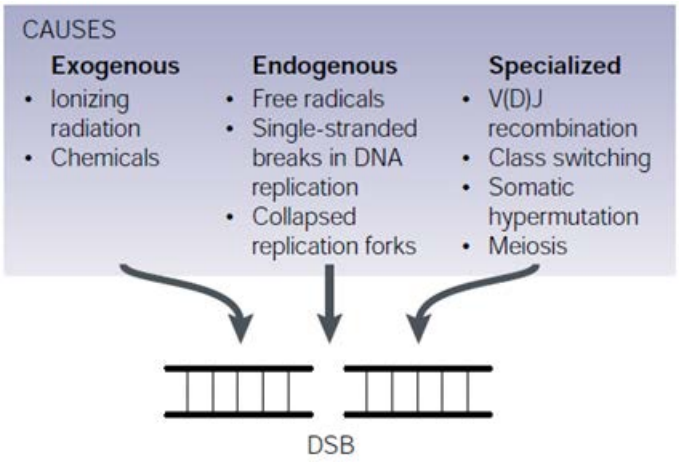


Figure 1.3 Origin of double-stranded breaks. Ionizing radiation and some certain chemicals are the major exogenous causes of double-stranded breaks. On the other hand, other endogenous sources such as free radicals and single-stranded breaks in DNA can also induce double-stranded breaks. Last but not least, double-stranded breaks is a natural process during meiosis. Picture is reproduced from (6) with permission.

although this type of DSBs has relative low complexity to the one caused by ionizing radiation. In addition, single-stranded break is likely to be converted to DSB during DNA replication (60,61). Last but not least, DSBs exist as normal intermediates during recombination in meiosis (62). Detailed repair pathways towards DSBs will be mentioned later in this chapter.

III. Enzymes and repair pathways for damaged DNA bases

A. DNA repair strategies

Damaged DNA will cause change and loss of information stored in the genome, and if the corruptions located in the essential part of the genome, cell may face serious problems such as apoptosis or transformation to cancer cells. Depending on the types of DNA damage, different repair strategies have been evolved (Figure 1.4) (11). Usually, cells would use unmodified complementary DNA strand or a DNA strand from a sister chromatid as a template to restore the original information. However, under some circumstances, the lack of a backup storage DNA strand will force cells to choose an error-prone replication mechanism known as translesion synthesis which can solve the problem temporarily but might cause more in the long term.

B. Direct reversal repair

One of the DNA repair strategies is called direct reversal. To date, three kinds of DNA base damage can be removed by direct chemical reversion. Most importantly, direct reversal repair does not need a template DNA: the enzymes will convert the modified DNA base directly back to the original base, so this mechanism does not involve the step of breakage of DNA backbone. The first type of DNA damage which can be repaired by direct

reversal pathway is O6-methylguanine, the most common O-alkyl lesion in DNA bases which prefers to pair with thymine in replication (63). The enzyme responsible to repair this lesion is called O6-methylguanine DNA methyltransferase (known as MGMT in humans). MGMT uses its Cys-145 as a receptor to the alkyl group to repair O6-methylguanine. Since the transfer is irreversible, after repair of the mutation base, the enzyme will be ubiquitinated and degraded by proteasome (64,65). N1-methyladenine and N3-methylcytosine are the other types of damaged DNA bases that can be directly reverted. AlkB, a well-known oxidative demethylase, discovered in 1983 in *E. coli*, is responsible for repairing the N-alkyl DNA bases (66). The enzyme removes the methyl group from DNA base and releases a formaldehyde by using Fe^{2+} and α-ketoglutarate as co-factors (67). Different from MGMT, AlkB is not a suicide enzyme, and its homologs in human cells such as hABH2 and hABH3 maintained similar enzymatic activities, suggesting this

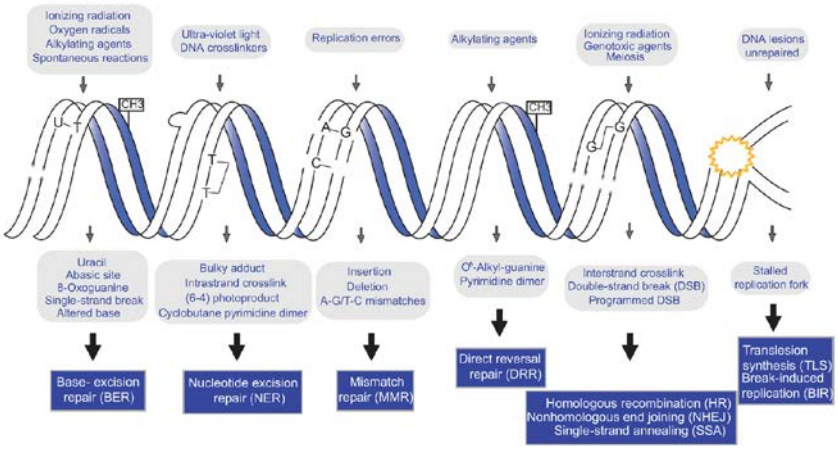


Figure 1.4 DNA damage and their corresponding repair pathways. Six major types of DNA damages are listed on the top of the picture. And their corresponding repair pathways are listed at the bottom of the figure. Ionizing radiation, oxygen radicals, alkylating agents and spontaneous mutations are likely to be repaired by BER pathway. The bulky adduct caused by Ultra-violet light and DNA crosslinkers will be fixed by NER pathway. If the damage is coming from replication, it would be repaired by MMR pathway. And double-stranded breaks result from ionizing radiation usually are mended by homologous recombination. However, in some circumstances, DNA damage is too severe to be repaired. In these cases, translesion synthesis of DNA will be initiated although it is an error-prone process. Picture is reproduced from (11) with permission.

group of enzymes might play an important role as a demethylase against alkylating agents (68-70). The third type of DNA damage can be directly reversed by repair enzyme is pyrimidine dimer caused by UV light (71). This dangerous DNA damage could jeopardize the genome integrity by blocking the replication and transcription. Photolyase is the enzyme that can separate the dimer and restore the genome information (71). The enzyme can recognize the lesion and repair it in a light-dependent manner (72). However, some species such as humans, do not have its homologs. Since the lack of photoreactivation repair enzyme, these species may only depend on the nucleotide excision repair pathway (NER) (73), which will be discussed in later paragraphs.

C. Single-strand damage repair

In most of DNA base damage, only one DNA strand will have mutated bases, and the other strand can serve as the template to restore the genome information. In order to remove the damaged bases, the damaged DNA strand will be cleaved by different excision repair enzymes, and the complementary DNA strand will be used as a template to guide DNA polymerase to put the original base back. There are several well-known repair pathways, such as base excision repair pathway (BER), nucleotide excision repair pathway (NER), and mismatch repair pathway (MMR) (Figure 1.4) (11).

As it sounds, base excision repair (BER) pathway is a repair pathway that will directly remove the mutated bases from the DNA and generate apurinic/apyrimidic (AP) sites (Figure 1.5). The modified bases will first be recognized by a group of enzymes-DNA glycosylases. Each class of DNA glycosylases is able to recognize a set of specific modified DNA bases. For example, enzymes in family 1 uracil DNA glycosylase are able

to specifically cleave uracil from DNA (17). There are two types of DNA glycosylase, monofunctional glycosylase and bifunctional glycosylase. Uracil-DNA glycosylases are usually monofunctional enzymes which means the enzyme will only function as a glycosylase to break N-glycosidic bond, while a bifunctional enzyme can work as both glycoylase and AP lyase. A classic example of bifunctional glycosylase is OGG1 (74), which is the primary enzyme to remove 8-oxoguanine, an important mutation base generated by reactive oxygen species (75,76). After the specific modified bases are removed by DNA glycosylase, the generated AP sites will be recognized by enzymes with AP endonuclease activity (Figure 1.5). To date, four classes of AP endonuclease have been found according to their sites of cleavage (77). Class I and class II AP endonucleases would leave a 3´-OH and a 5´-phosphate end, while class III and class IV enzymes would like to produce a 3´-phosphate and a 5´-OH termini. Human AP endonuclease (APE1) belongs to

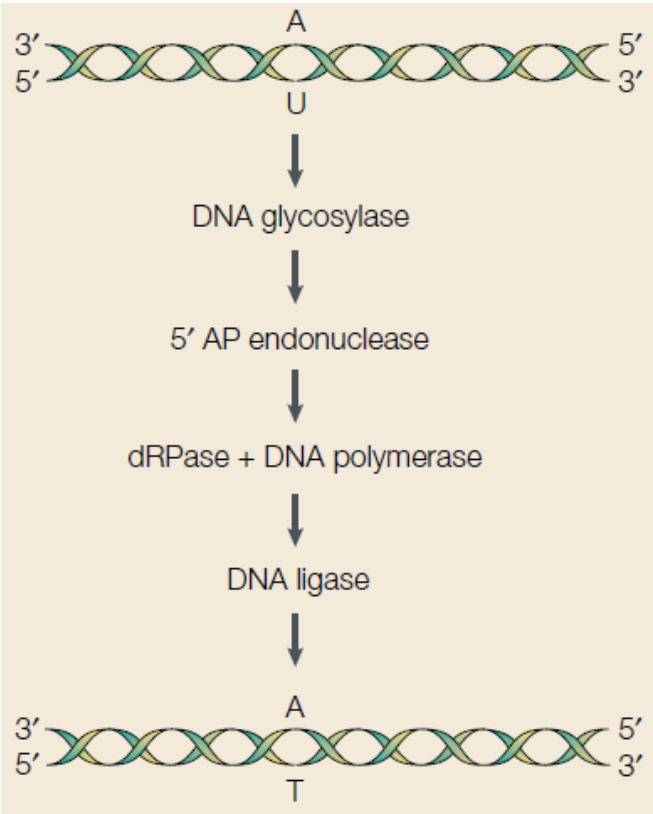


Figure 1.5 Base excision repair (BER) pathway. The specific modified bases are first recognized and removed by DNA glycosylase to generate an AP site. Afterwards, AP site will be processed by AP endonuclease. In the end, DNA polymerase and DNA ligase will incorporate the correct DNA base into the position and fill the gap (4,5). Figure is reproduced from (5) with permission.

class II AP endonuclease, and it needs an Mg^{2+} as a cofactor (78). Since the removal of AP sites is an essential step for BER pathway, and AP sites are highly mutagenic and toxic to cells, drugs have been designed to target APE1 to reduce the chance of chemotherapy resistance of cancer cells (79). The last step of BER pathway is accomplished by DNA polymerase and DNA ligase, a correct DNA base is incorporated into the DNA strand by these enzymes (Figure 1.5).

The second repair pathway involved in single-strand DNA break is called mismatch repair pathway (MMR). Different with BER pathway, by which damaged bases are removed as free DNA bases, mutation bases are released as mononucleotides by MMR pathway (80,81). The main feature of this repair pathway is that it is coupled with DNA replication and DNA damage. It could initiate a strand specific repair by recognizing the incorrect genomic information such as insertion, deletion or mismatch DNA base in the daughter strand during DNA synthesis (82,83). In *E. coli*, the strand specific signal is

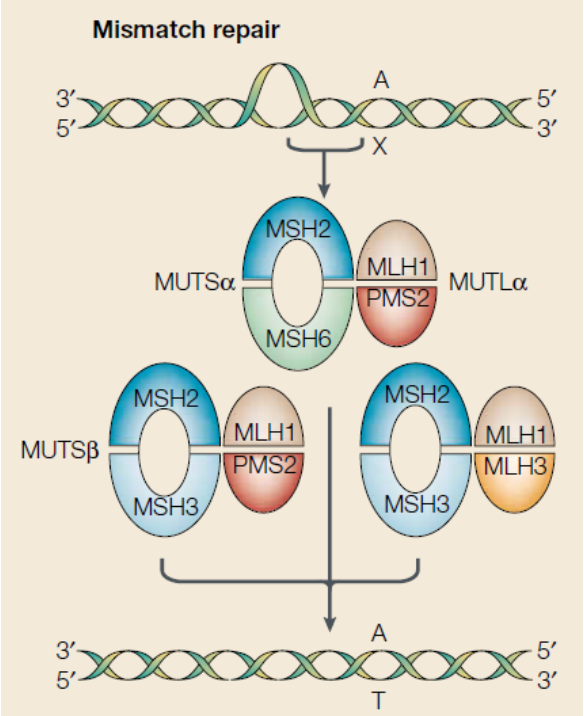


Figure 1.6 Mismatch repair pathway. Lesion is recognized by MUTSα and MUTSβ heterodimers, and then more components join this pathway to repair the DNA damage. Picture is reproduced from (5) with permission.

provided by the methylation status of DNA strand. Since the newly synthesized DNA strand is not labeled by methylation, it can be differed from template strand. However, the mechanism of specific strand recognition in eukaryotes is not clear yet (84). It has been suggested that the transient existence of nicks in the daughter strand could serve as a mark to be recognized by MMR pathway. In *E. coli*, a protein complex containing MutS and MutL homodimers are responsible for lesion recognition (85). In contrast, MutS homologs in eukaryotes are replaced by MSH2, MSH3 and MSH6. They form two heterodimers to initiate the MMR pathway (MSH2-MSH6 and MSH2-MSH3, named as MUTSα and MUTSβ respectively) (86). Both complexes possess a walker ATP-binding motif suggesting that the activities of these enzymes are ATP dependent (87). The mechanism subsequent to the lesion recognition in mammalian cells is not clear, however, it is known that several other proteins are involved in this repair pathway, such as MLH1, PMS2 and MLH3 (Figure 1.6) (5,88). Defects in this pathway, coupled with defects in nucleotide excision repair pathway (NER), are inclined to cancer, especially colon cancer (89), but also uterine, ovarian and gastric cancer (80).

Nucleotide excision repair (NER) pathway is another important DNA repair pathway to prevent mutations and diseases, by which damaged DNA bases are removed within an oligonucleotide fragment, not a free base as BER pathway or single nucleotide as MMR pathway (90-92). One important role of this repair pathway is to remove the damaged pyrimidine dimers caused by UV radiation (93,94). As mentioned before, since some species lack photolyase to directly reverse the pyrimidine dimers to their original forms, repair of this type of DNA damage exclusively depends on the NER pathway in these organisms (73). Other specific targets of NER pathway include cyclobutane

pyrimidine dimers (CPDs), and pyrimidine-(6,4)-pyrimidone products (6-4PPs), helix-distorting bulky adducts and intrastrand cross-links caused by chemicals or alkylating agents (95). Two different subpathways have been proposed in NER pathway due to their different ways in searching the DNA lesion. Global genome NER (GG-NER) focuses on scanning the DNA damage through whole genome. On the other hand, transcription-coupled NER (TC-NER) is specialized in recognizing DNA damage during RNA transcription (96) (Figure 1.7). To fulfill the DNA damage surveillance, GG-NER adopts a specific protein complex named as XPC-hRad23B with centrin2 (97-100). Although XPC

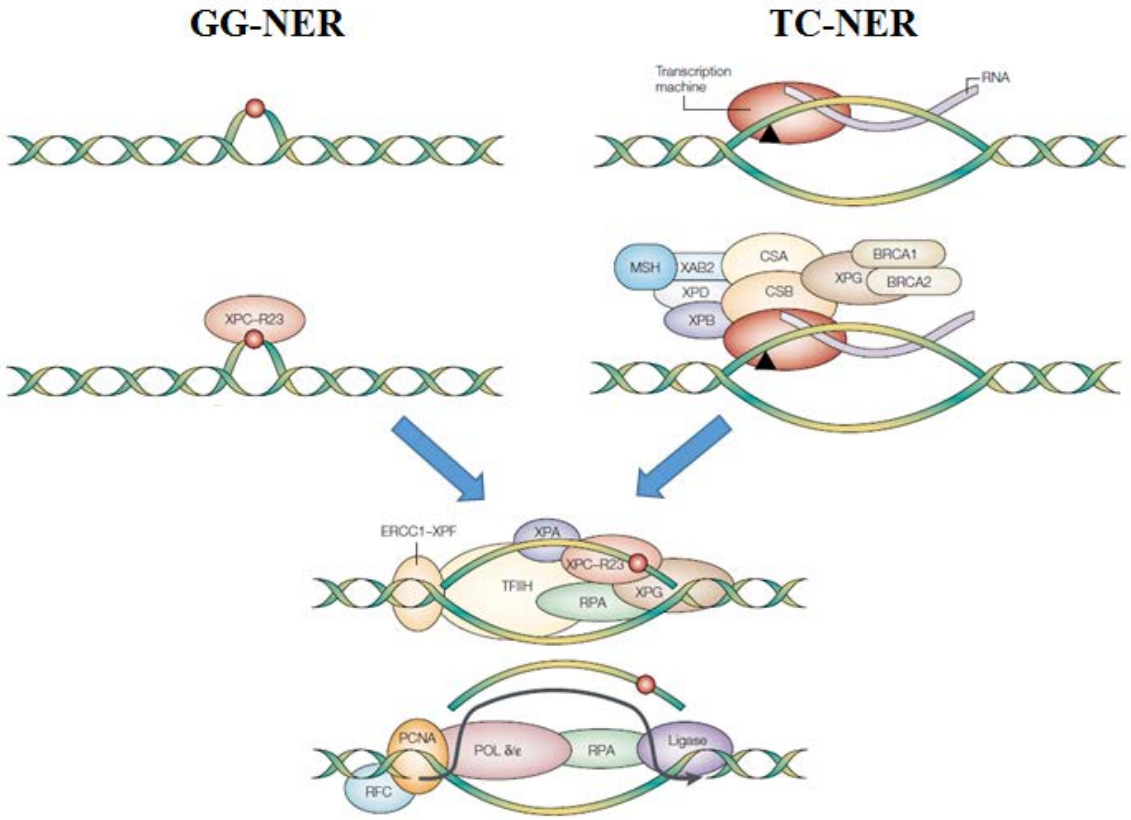


Figure 1.7 Global genome NER (GG-NER) and transcription-coupled NER (TC-NER). Two different subpathways have been proposed in NER pathway due to their different performance in searching the DNA lesion. GG-NER adopts a specific protein complex named as XPC-hRad23B. On the other hand, TC-NER pathway is believed to possess a large repair protein complex recruited by the arrested transcription complex (10). Picture is reproduced from (5) with permission.

is capable to recognize DNA damage independently *in vitro*, the efficiency is significantly improved by existence of hRad23 protein (101). Centrin2 has been proved to aid the stabilization of the interaction between XPC and hRad23 proteins, suggesting that the formation of this protein complex is very important for the GG-NER pathway (102). Afterwards, XPA and RPA are believed to bind to the DNA (5), and transcription factor II human (TFIIH) complex is suggested to unwind the DNA by two DNA helicases (XPB and XPD) within this super complex (103,104). The protein complex to cleave the DNA comprises of three proteins, including two endonucleases (80,89,103). The endonuclease XPG cuts at 3' end of damaged bases, whereas the endonuclease activity of the ERCC1–XPF heterodimers would cleave at 5' end of damaged bases. The released nucleotide fragment is about 30 nucleotides, however, the distance between the damaged bases to the incision site usually varies between different damaged bases (90,91,105-107). It has been proposed that the step of incision is followed by repair synthesis of DNA right away, since the formation of large single-strand gaps might become the target of nucleases which will cause the degradation of the genome. In the transcription-coupled NER (TC-NER) pathway, it is believed that the arrested transcription by RNA polymerase II results in the recruitment of a large repair protein complex (10). Although the exact components of this protein complex are not clear yet, two proteins, CSA and CSB are believed to be very important in lesion recognition, since inactivation of these two proteins would cause Cockayne syndrome (108). Several other proteins, such as MSH, BRCA1 and BRCA2, are also found to play important roles in TC-NER pathway. The subsequent biochemical events of TC-NER are similar with the GG-NER pathway, the only difference for these two subpathways is in the damage recognition step (Figure 1.7). Deficiency in NER pathway might cause

severe problems. Several human genetic disorders such as UV sensitivity syndrome (UVSS), xeroderma pigmentosum (XP), Cockayne's syndrome (CS), and trichothiodystrophy (TTD) have been linked to the deficiency of the NER pathway (64,109,110).

The other single-strand damage repair pathway is the endonuclease V initiated repair pathway (8) (Figure 1.8). Endonuclease V is the 5th endonuclease discovered in *E. coli* in 1977 (111). Endonuclease V from different species are able to target several deaminated DNA bases including uracil, hypoxanthine, xanthine and oxanine, suggesting the importance of these enzymes in anti-deamination of DNA bases (112-115). As the first enzyme in this repair pathway, endonuclease V is able to nick at the 3' end of the lesion. However, different from the other single-stranded repair pathway mentioned above, the

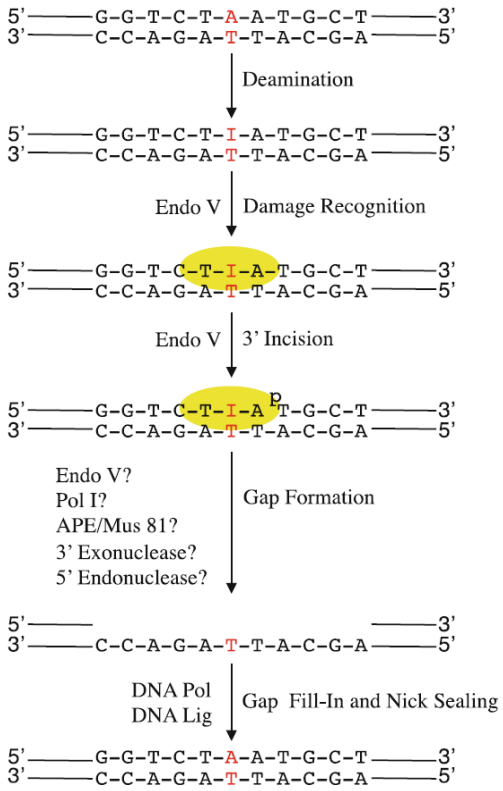


Figure 1.8 Proposed endonuclease V-initiated repair pathway. I is hypoxanthine, and endonuclease V is marked as yellow oval. Endonuclease V is the first enzyme initiating the pathway, although the downstream pathway is not clear yet. Picture is reproduced from (8) with permission.

damaged base is not removed (116,117). It has been proposed that a 3' exonuclease or an endonuclease is needed to remove the mutation. In one hypothesis, *E. coli* DNA polymerase I is proposed to remove hypoxanthine from the lesion due to its 3' exonuclease activity (118). In another hypothesis, endonuclease V itself possesses lesion specific 3'-exonuclease activity and non-specific 5'-exonuclease in the presence of Mn^{2+} (119). This finding suggests that the enzyme is able to remove the mutation base and enlarge the gap by its endonuclease activity, making it more easily to be recognized by DNA polymerase and DNA ligase. Surprisingly, human endonuclease V is identified as RNA endonuclease instead of DNA endonuclease recently (120,121). The homolog from human is able to remove hypoxanthine from RNA in a more efficient way compared to DNA. In additions, the cellular localization of human endonuclease is in cytoplasm rather than nuclei, and the enzyme is capable to recognize the deaminated RNA generated by ADAR2 enzyme, supporting the hypothesis that human endonuclease V might have evolved to play an important role in RNA metabolism in human cells.

D. Double strand break repair

To prevent genome rearrangement, double strand break (DSB) is a type of DNA damage need to be identified and repaired right away. Besides the rearrangement of the genome, single unrepaired DSB could also result in another type of deleterious DNA damage, aneuploidy (11,58). To date, two existing pathways have been indicated to repair this lethal damage: homologous recombination (HR) and non-homologous end joining (NHEJ) (6,122).

Homologous recombination (HR) fixes DSBs using a sister chromatid as a template, which is a high-fidelity DSB repair mechanism (122). The pathway is initiated by the

generation of a single-stranded DNA which is coated by single-strand DNA-binding protein and replication protein A (RPA). This protein-DNA filament will further recruit Rad52 which serves as the binding target of Rad 51 (123,124). Rad51 is responsible for mediating the essential step in HR, searching the homologous double-strand DNA and initiating the strand invasion (125) (Figure 1.9). The DNA structure formed during the invasion of template strand is named as holiday junction (126,127). The following biochemical events include DNA synthesis using sister chromatid strand as a template and resolution of the heteroduplex (Figure 1.9) (12).

In contrast to HR, non-homologous end joining (NHEJ) adopts little or no homologous sequence (6). Proteins involved in this pathway include a heterodimer KU80/70, a DNA-dependent protein kinase-DNA-PKcs, a complex of XRCC4 (X-ray-repair-cross-complementing defective repair in Chinese hamster mutant 4) and DNA ligase

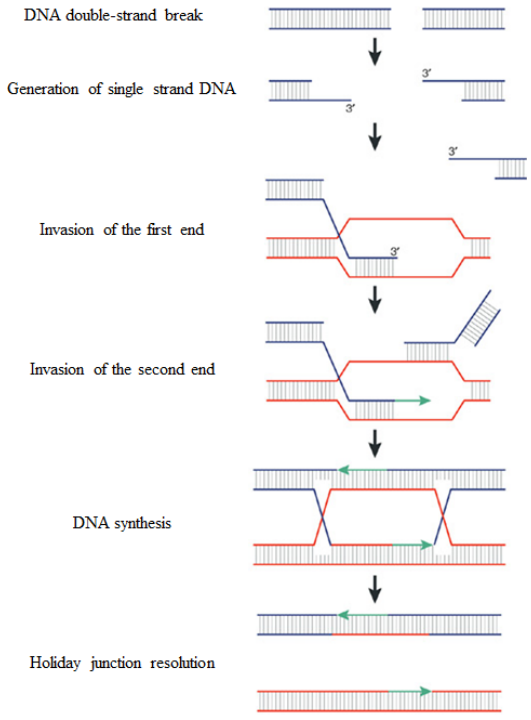


Figure 1.9 DSB repaired by homologous recombination. The end of DNA strand is digested to single-stranded DNA first. Afterwards, the ssDNA is coated and nucleoprotein filament is formed. Next, ssDNA pair with its homologous double stranded DNA in sister chromatid and strand exchange take place. The two end of ssDNA could serve as the primers for DNA synthesis. In the end, holiday junction are resolved and the two double strand DNA are separated. Picture is reproduced from (12) with permission.

IV and some other factors, such as Rad50, NBS1 and MRE11, although their functions are not clear yet. The DSB is first identified by KU80/70 heterodimer, which has very high affinity to free DNA ends (128). The formation of DNA-protein complex will recruit the DNA-PKcs to the damage site (129). The target of this protein kinase is not clear yet, but it might include p53 and itself. Afterwards, the XRCC4 and DNA ligase IV will accomplish the ligation step and seal the gap. Since NHEJ is not a template-based repair pathway, it might induce mutations to the genome.

Different DSB repair pathways play different roles in different cellular processes. For example, HR is believed to be more important during DNA replication, probably because the stalling replication fork is highly linked to DSBs which are the primary target for HR pathway during the cell cycle (130,131). Deficiency in HR pathway does not significantly affect the DNA damage response to ionizing radiation, but highly sensitive to DNA-crosslinking agent that blocks DNA replication (132). On the other hand, NHEJ is likely to deal with DSBs caused by ionizing radiation (58).

III. Uracil-DNA glycosylase superfamily

A. Introduction of uracil-DNA glycosylase (UDG)

Uracil-DNA glycosylase is the first enzyme in the BER pathway to recognize deaminated DNA bases (133). All the members in UDG superfamily are monofunctional enzymes, suggesting that they can only remove the DNA base and generate an AP site, but do not have endonuclease activity. To date, six families of enzymes have been found in UDG superfamily (134) (Figure 1.10). Most of them are able to remove uracil from DNA except family 6 HDG which is an exclusive hypoxanthine DNA glycosylase (HDG).

Although low similarities are found in their amino acid sequence, they share some highly conserved motif 1 and motif 2 (Figure 1.10) (48,134).

B. Family 1 UNG

Family 1 UDG, also called UNG, is the first UDG discovered in *E. coli* (17). The enzymes in family 1 UNG are specific for removing uracil. The enzymes can efficiently remove uracil from double-stranded and single-stranded DNA, and are the most efficient UDG towards uracil in the UDG superfamily (48). Besides keeping genome integrity by removing uracil from DNA caused by cytosine deamination, mammalian UNGs play an important role in maintaining the varieties of immunoglobulin molecules (Figure 1.11) (2). In this case, cytosine in DNA will be first deaminated to uracil by activation-induced (cytosine) deaminase (AID). And if uracil is left unrepaired, hypermutation (transitions only) will be induced. In contrast, family 1 UNG is able to remove uracil from DNA, generating an AP site which could induce both transition and transversion mutation by

```
Family 1 Eco N--61-GQDPYHGPGQAHGLAFSVRPGIAT-37-NTVLTVRAGQ-54-HPSPLSA--36-C
(UNG)
Family 2 Eco N--15-GINPGLSSAG-TGFPFAHPANRFW-29-KLVDRPTVQA-62-NPSGLSR--22-C
(MUG/TDG)
Family 3 Gme N--55-GMNPGPWGMAQTGVPFGEVAVVTE-56-NYCPLLFLTA-64-HPSPASP--21-C
(SMUG1)
Family 4 Tth N--39-GEGPGEEEDK-TGRPFVGKAGQLL-17-NIVKCRPPQN-65-HPAYLLR--44-C
(UDGa)
Family 5 Tth N--56-GLAPGAHGSNRTGRPFTGDASGAF-30-AAVRCAPPKN-69-HVSRQNT--23-C
(UDGb)
Family 6 Mba N--19-GSLPGDVSIR-KHQYYGHPGNDFW-31-DVFKAGKREG-52-SSSGANR--16-C
(HDG)                    Motif 1                          Motif 2
```

Figure 1.10 Sequence alignment of UDG superfamily. The alignment was based on BLAST and CLUSTALW analysis and constructed manually. Family 1 (UDG): *Eco, E. coli*, NP_289138. Family 2 (MUG/TDG): *Eco, E. coli*, P0A9H1. Family 3 (SMUG1): *Gme, G. metallireducens* GS-15,YP_383069. Family 4 (UDGa): *Tth, T. thermophilus* HB27, YP_004341.1. Family 5 (UDGb): *Tth, T. thermophilus* HB8, YP_144415.1. Family 6 (HDG): *Mba, Methanosarcina barkeri str. Fusaro,* YP_304295.1.

DNA replication. However, if family 1 UNG is inhibited by its specific inhibitor UGI, the varieties of immunoglobulin molecules will be decreased.

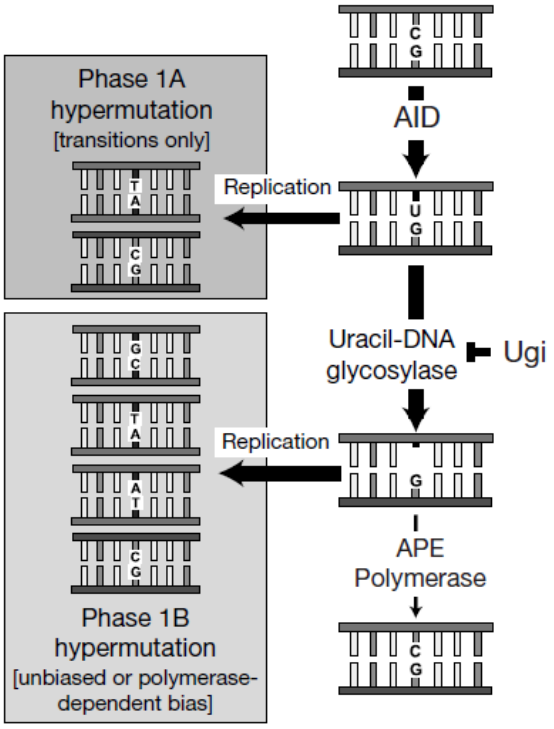


Figure 1.11 Proposed functions of UNG in hypermutation. When uracil-DNA glycosylase is not inhibited by Ugi, there is no bias on hypermutation induced by AID. However, if uracil-DNA glycosylase is inhibited by Ugi, only transitions will be induced. Picture is reproduced from (2) with permission.

UNG enzymes from human, *E. coli*, and herpes simplex virus type I are well studied by different methods such as mutagenesis, kinetic analysis, NMR analysis, transition state (TS) analysis and structural analysis (135-148). Multiple crystal structures with uracil or different uracil analogs have been solved (149). TS analysis demonstrated that *E coli* UNG might remove uracil through D_N*A_N mechanism, by which a uracil departure (D_N) followed by water nucleophilic attack (A_N). Several highly conserved amino acids within the active site have proved to be essential for catalysis (137-139). H187, a highly conserved amino acid in motif 2, is shown to be the most important amino acid during activation of leaving group. Its interaction with O2 of the imidate tautomer of uracil without transferring a proton, is able to lower the pKa of N1 of uracil from 9.8 to 6.4, significantly stabilizing uracil leaving as anion group. Phe158 in human UNG, another highly conserved amino

acid in motif 1 of family 1 UNG, is also shown to be necessary for this step. It can form a parallel stack with uracil substrate, and the notion here is the face-to-face staking provide an environment, which favors the hydrogen bond between O2 of the uracil and histidine in motif 2 (150). Besides of Phe in motif 1 and His in motif 2 of family 1 UNG, Asp64 in *E. coli* UNG is also conserved throughout family 1 UNG. It has been suggested that Asp64 promotes the catalysis by 8000-fold (137). One explanation for this is that Asp64 might be responsible for initiating the water attack, the other explanation is the backbone carbonyl of Asp64 is in contact with the uracil ring, together with Phe on the opposite side of uracil, providing the face-to-face stacking to promote the hydrogen bond between histidine and O2 of the uracil (151).

C. Family 2 TDG/MUG

Family 2 UDG is called thymine-DNA glycosylase (TDG) or mismatch-specific uracil DNA glycosylase, since the enzyme was first found as a glycosylase to repair G/T and G/U mismatch (152,153). Different from family 1 UNG, family 2 TDG/MUG lacks the key amino acids equivalent to aspartic acid in motif 1 and histidine in motif 2. The corresponding residues are Asn18 and Asn140 in *E. coli* MUG and Asn140 and Ser271 in human TDG respectively. Mutational analysis indicates that elimination of side chain of Asn140 in human TDG abolishes the activity of the enzyme but allowed the enzyme to bind to the G/U or G/T substrates, suggesting water attack might play a more important role compared to leaving group activation in family 2 UDGs catalysis (154). However, it was found later that uracil might not be the most favorable substrates for family 2 TDG/MUG. Xanthine was shown to be a better substrate for *E. coli* MUG (4), and mutagenesis and structural analysis indicated Ser23 in *E. coli* promotes xanthine leaving

group activation by interacting with N7 of xanthine. In addition, human TDG is found to play an essential role in the demethylation pathway by efficiently excising 5-caboxylcytosine (5-caC) from DNA, and the crystal structure of human TDG with 5-caC supports its robust activity on 5-caC (3,155). Similar with the findings in *E. coli* MUG, human TDG can form several hydrogen bonds to promote the leaving group activation of 5-caC by backbone of Asn140 and Ile139 and side chain of Asn157 (Figure 1.12). Consistent with the crystal structure result, mutation of Asn140 fails to abolish all the enzymatic activity of human TDG, supporting the notion that leaving group activation and water attack are both important in family 2 TDG/MUG (156,157).

D. Family 3 SMUG1

Family 3 UDG members are named as single-strand selective monofunctional UDG (SMUG1), which can efficiently remove uracil, xanthine, 5-hydroxymethyluracil and 5-hydroxyuracil (9,158). It is hypothesized that human SMUG1 might also involve in DNA

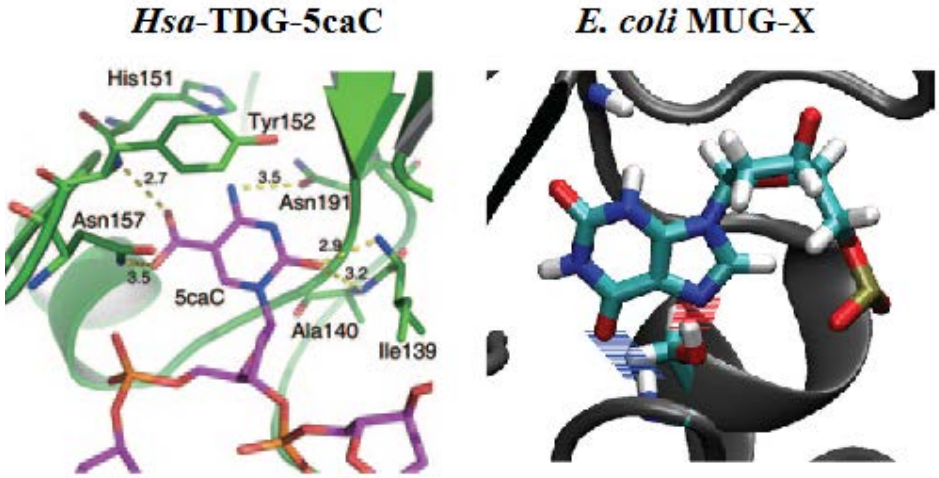


Figure 1.12 Active site of human TDG with 5-caC and E. coli MUG with xanthine. Amino acids interacted with 5-caC in human TDG is labeled and hydrogen bond is represented as yellow dash line (left). Interaction between N7 of xanthine and Ser23 of *E. coli* MUG is labeled by red dash line. Pictures are reproduced from (3,4) with permission.

demethylation pathway by removing 5-hydroxymethyluracil which is converted from 5-methylcytosine by an enzyme called AID (activation induced deaminase). However, no further evidence was provided and no activity was found towards modified methyl-cytosine by AID (159). Later on, human SMUG1 was linked to RNA quality control by specifically excising 5-hydroxyuracil from RNA. Knock-out of SMUG1 from human cells significantly decrease the amount of mature 47S rRNA (160). Different with family 2 TDG/MUG, family 3 SMUG1 possesses a similar activity towards uracil compared to family 1 UNG. SMUG1 from *xenopus laevis*, is able to process G/U at 9 s^{-1}, only 13-fold slower than *E. coli* UNG (158). The similar enzyme activities between family 1 enzymes and family 3 enzymes might be caused by the highly conserved amino acid sequence in motif 1 and motif 2 of these two families (Figure 1.10). Mutation of Phe98 and His239 in human SMUG1 caused reduction of k_{cat}/K_m about $2x10^3$ fold and $1x10^4$ fold respectively (161).

E. Family 4 UDGa

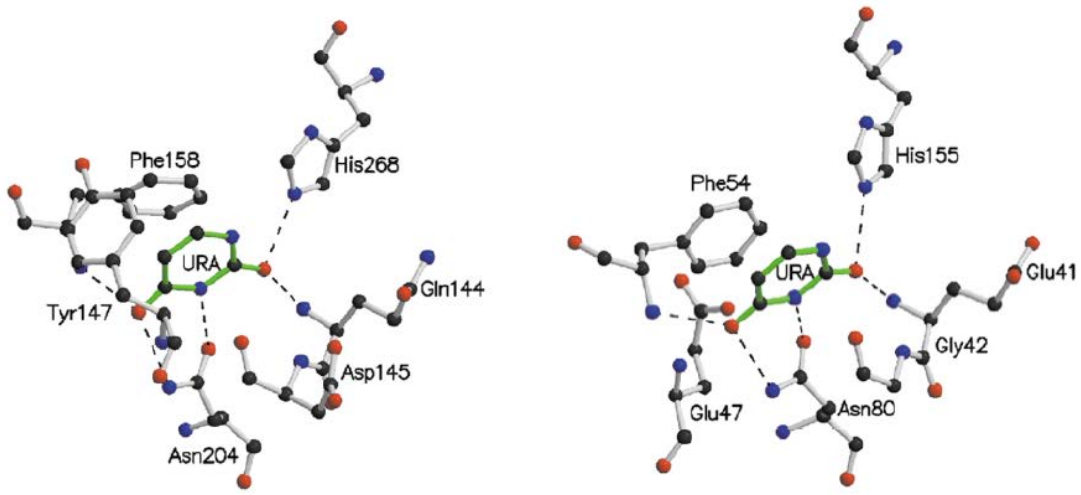


Figure 1.13 Active sites of UNG from *homo sapiens* and UDGa from *Thermus thermophilus* with uracil. The active site of *Has* UNG with uracil is shown on the left and active site of *Tth* UDGa with uracil is shown on the right. Picture is reproduced from (1) with permission.

Most enzymes belonging to family 4 UDGa are from thermophilic bacteria, such as *Thermos thermophiles (Tth)* (1), *Pyrobaculum aerophilum* (162,163), and *Thermotoga martima* (164,165). To date, family 4 UDGa was reported as a uracil-specific UDG, like family 1 UNG. The crystal structure of *Tth* UDGa reveals that family 4 UDGa has an active site very similar to family 1 UNG (Figure 1.13), although the corresponding site of Asp64 in *E. coli* UNG is replaced by a glycine (1). Family 4 UDGa is also characterized by a 4Fe-4S iron-sulfur cluster, although the function of the iron-sulfur cluster is unknown.

F. Family 5 UDGb

Similar with family 4 UDGa, family 5 UDGb exists in archaea and eubacteria, many of which are thermophilic bacteria. Biochemical studies of UDGb from *Pyrobaculum aerophilum* and *Mycobacterium tuberculosis* suggested that family 5 UDGb is a versatile enzyme which can remove several mutated purines and pyrimidines such as uracil, hydroxymethyluracil, fluorouracil, hypoxanthine and ethenocytosine (166,167). Consistently, genetic analysis reveals that family 5 UDGb from *Thermus thermophiles* (*Tth*) is a uracil DNA glycosylase that can reduce the mutation rate by three-fold (168,169). The crystal structure of UDGb from *T. thermophiles* confirms that UDGb adopts a structural fold similarly seen in other UDG enzymes (170). However, family 5 UDGb lacks Aspartic acid in the motif 1 as the catalytic residue to activate water molecule and mutagenesis of Ala68 in *P. aerophilium* UDGb back to Asp reduce the enzyme activity. This observation suggested that family 5 UDGb might adopt an alternative catalytic strategy to activate a water molecule which will be further discussed in chapter 2.

G. Family 6 HDG

The family 6 UDG is identified recently and named as hypoxanthine DNA glycosylase (HDG) since this family of UDG mainly act on hypoxanthine substrate (134). Sequence alignment analysis indicated that family 6 HDG shared similar amino acids in motif 1 and motif 2 with family 2 TDG/MUG, but not other UDG superfamily members. Mutational analysis and modelled structural analysis suggested family 6 HDG probably adopts different catalytic strategies and different catalytic center and Asn39 in *Methanosarcina barkeri* HDG is proposed to be potential catalytic residue (134). However, more information is needed to answer the question why family 6 HDG can only specifically act on hypoxanthine-containing substrates.

IV. References

1. Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S., and Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. *Journal of molecular biology* **333**, 515-526
2. Di Noia, J., and Neuberger, M. S. (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43-48
3. Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G. L., Luo, C., Jiang, H., and He, C. (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature chemical biology* **8**, 328-330
4. Lee, H. W., Brice, A. R., Wright, C. B., Dominy, B. N., and Cao, W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase. *The Journal of biological chemistry* **285**, 41483-41490
5. Friedberg, E. C. (2001) How nucleotide excision repair protects against cancer. *Nature reviews. Cancer* **1**, 22-33
6. van Gent, D. C., Hoeijmakers, J. H., and Kanaar, R. (2001) Chromosomal stability and the DNA double-stranded break connection. *Nature reviews. Genetics* **2**, 196-206
7. Pierce, B. (2005) Depurination produces an apurinic site. in *Genetics: A Conceptual Approach (Second Edition)*, 2nd Ed. pp 487
8. Cao, W. (2013) Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cellular and molecular life sciences : CMLS* **70**, 3145-3156

9. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N., and Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. *Journal of molecular biology* **385**, 761-778

10. Tsutakawa, S. E., and Cooper, P. K. (2000) Transcription-coupled repair of oxidative DNA damage in human cells: mechanisms and consequences. *Cold Spring Harbor symposia on quantitative biology* **65**, 201-215

11. Genois, M. M., Paquet, E. R., Laffitte, M. C., Maity, R., Rodrigue, A., Ouellette, M., and Masson, J. Y. (2014) DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. *Microbiology and molecular biology reviews : MMBR* **78**, 40-73

12. West, S. C. (2003) Molecular views of recombination proteins and their control. *Nature reviews. Molecular cell biology* **4**, 435-445

13. Liao, G. Y., An, J. J., Gharami, K., Waterhouse, E. G., Vanevski, F., Jones, K. R., and Xu, B. (2012) Dendritically targeted Bdnf mRNA is essential for energy balance and response to leptin. *Nature medicine* **18**, 564-571

14. Denissenko, M. F., Pao, A., Tang, M., and Pfeifer, G. P. (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* **274**, 430-432

15. Marchetti, F., Rowan-Carroll, A., Williams, A., Polyzos, A., Berndt-Weis, M. L., and Yauk, C. L. (2011) Sidestream tobacco smoke is a male germ cell mutagen. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12811-12814

16. Anand, P., Kunnumakkara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., Sung, B., and Aggarwal, B. B. (2008) Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research* **25**, 2097-2116

17. Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B., and Sperens, B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *The Journal of biological chemistry* **252**, 3286-3294

18. Halliwell, B. (2007) Oxidative stress and cancer: have we moved forward? *The Biochemical journal* **401**, 1-11

19. Valko, M., Leibfritz, D., Moncol, J., Cronin, M. T., Mazur, M., and Telser, J. (2007) Free radicals and antioxidants in normal physiological functions and human disease. *The international journal of biochemistry & cell biology* **39**, 44-84

20. Pohanka, M. (2013) Alzheimer s disease and oxidative stress: a review. *Current medicinal chemistry* **21**, 356-364

21. Cadet, J., Delatour, T., Douki, T., Gasparutto, D., Pouget, J. P., Ravanat, J. L., and Sauvaigo, S. (1999) Hydroxyl radicals and DNA base damage. *Mutation research* **424**, 9-21

22. Valko, M., Izakovic, M., Mazur, M., Rhodes, C. J., and Telser, J. (2004) Role of oxygen radicals in DNA damage and cancer incidence. *Molecular and cellular biochemistry* **266**, 37-56

23. Foksinski, M., Rozalski, R., Guz, J., Ruszkowska, B., Sztukowska, P., Piwowarski, M., Klungland, A., and Olinski, R. (2004) Urinary excretion of DNA repair products correlates with metabolic rates as well as with maximum life spans of different mammalian species. *Free radical biology & medicine* **37**, 1449-1454

24. Tudek, B., Winczura, A., Janik, J., Siomek, A., Foksinski, M., and Olinski, R. (2010) Involvement of oxidatively damaged DNA and repair in cancer development and aging. *American journal of translational research* **2**, 254-284

25. Helbock, H. J., Beckman, K. B., Shigenaga, M. K., Walter, P. B., Woodall, A. A., Yeo, H. C., and Ames, B. N. (1998) DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 288-293

26. Ames, B. N., and Gold, L. S. (1991) Endogenous mutagens and the causes of aging and cancer. *Mutation research* **250**, 3-16

27. Shibutani, S., Takeshita, M., and Grollman, A. P. (1991) Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349**, 431-434

28. Taddei, F., Hayakawa, H., Bouton, M., Cirinesi, A., Matic, I., Sekiguchi, M., and Radman, M. (1997) Counteraction by MutT protein of transcriptional errors caused by oxidative damage. *Science* **278**, 128-130

29. Lindahl, T., and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610-3618

30. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715

31. Nakamura, J., Walker, V. E., Upton, P. B., Chiang, S. Y., Kow, Y. W., and Swenberg, J. A. (1998) Highly sensitive apurinic/apyrimidinic site assay can detect

spontaneous and chemically induced depurination under physiological conditions. *Cancer research* **58**, 222-225

32. Cavalieri, E., Saeed, M., Zahid, M., Cassada, D., Snow, D., Miljkovic, M., and Rogan, E. (2012) Mechanism of DNA depurination by carcinogens in relation to cancer initiation. *IUBMB life* **64**, 169-179

33. Vieth, R. (1999) Vitamin D supplementation, 25-hydroxyvitamin D concentrations, and safety. *The American journal of clinical nutrition* **69**, 842-856

34. Goodsell, D. S. (2001) The molecular perspective: ultraviolet light and pyrimidine dimers. *The oncologist* **6**, 298-299

35. Huang, J. C., Svoboda, D. L., Reardon, J. T., and Sancar, A. (1992) Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 3664-3668

36. Fang, F. C. (2004) Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. *Nature reviews. Microbiology* **2**, 820-832

37. Iovine, N. M., Pursnani, S., Voldman, A., Wasserman, G., Blaser, M. J., and Weinrauch, Y. (2008) Reactive nitrogen species contribute to innate host defense against Campylobacter jejuni. *Infection and immunity* **76**, 986-993

38. Hou, Y. C., Janczuk, A., and Wang, P. G. (1999) Current trends in the development of nitric oxide donors. *Current pharmaceutical design* **5**, 417-441

39. Marletta, M. A. (1989) Nitric oxide: biosynthesis and biological significance. *Trends in biochemical sciences* **14**, 488-492

40. Terato, H., Masaoka, A., Asagoshi, K., Honsho, A., Ohyama, Y., Suzuki, T., Yamada, M., Makino, K., Yamamoto, K., and Ide, H. (2002) Novel repair activities of AlkA (3-methyladenine DNA glycosylase II) and endonuclease VIII for xanthine and oxanine, guanine lesions induced by nitric oxide and nitrous acid. *Nucleic acids research* **30**, 4975-4984

41. Ohshima, H., and Bartsch, H. (1994) Chronic infections and inflammatory processes as cancer risk factors: possible role of nitric oxide in carcinogenesis. *Mutation research* **305**, 253-264

42. Ungefroren, H., Sebens, S., Seidl, D., Lehnert, H., and Hass, R. (2011) Interaction of tumor cells with the microenvironment. *Cell communication and signaling : CCS* **9**, 18

43. Wink, D. A., Vodovotz, Y., Laval, J., Laval, F., Dewhirst, M. W., and Mitchell, J. B. (1998) The multifaceted roles of nitric oxide in cancer. *Carcinogenesis* **19**, 711-721

44. Wink, D. A., Kasprzak, K. S., Maragos, C. M., Elespuru, R. K., Misra, M., Dunams, T. M., Cebula, T. A., Koch, W. H., Andrews, A. W., Allen, J. S., and et al. (1991) DNA deaminating ability and genotoxicity of nitric oxide and its progenitors. *Science* **254**, 1001-1003

45. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N., and Cao, W. (2014) Specificity and Catalytic Mechanism in Family 5 Uracil DNA Glycosylase. *The Journal of biological chemistry* **289**, 18413-18426

46. Taylor, A. F., and Weiss, B. (1982) Role of exonuclease III in the base excision repair of uracil-containing DNA. *Journal of bacteriology* **151**, 351-357

47. Duncan, B. K., and Miller, J. H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560-561

48. Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutation research* **460**, 165-181

49. Hill-Perkins, M., Jones, M. D., and Karran, P. (1986) Site-specific mutagenesis in vivo by single methylated or deaminated purine bases. *Mutation research* **162**, 153-163

50. Martin, F. H., Castro, M. M., Aboul-ela, F., and Tinoco, I., Jr. (1985) Base pairing involving deoxyinosine: implications for probe design. *Nucleic acids research* **13**, 8927-8938

51. Suzuki, T., Yamaoka, R., Nishi, M., Ide, H., and Makino, K. (1996) Isolation and characterization of a novel product, 2'-deoxyoxanosine, from 2'-deoxyguanosine, oligodeoxynucleotide, and calf thymus DNA treated by nitrous acid and nitric oxide. *J Am Chem Soc* **118**, 2515-2516

52. Suzuki, T., Matsumura, Y., Ide, H., Kanaori, K., Tajima, K., and Makino, K. (1997) Deglycosylation susceptibility and base-pairing stability of 2'-deoxyoxanosine in oligodeoxynucleotide. *Biochemistry* **36**, 8013-8019

53. Eritja, R., Horowitz, D. M., Walker, P. A., Ziehler-Martin, J. P., Boosalis, M. S., Goodman, M. F., Itakura, K., and Kaplan, B. E. (1986) Synthesis and properties of oligonucleotides containing 2'-deoxynebularine and 2'-deoxyxanthosine. *Nucleic acids research* **14**, 8135-8153

54. Suzuki, T., Yoshida, M., Yamada, M., Ide, H., Kobayashi, M., Kanaori, K., Tajima, K., and Makino, K. (1998) Misincorporation of 2'-deoxyoxanosine 5'-triphosphate

by DNA polymerases and its implication for mutagenesis. *Biochemistry* **37**, 11592-11598

55. Caulfield, J. L., Wishnok, J. S., and Tannenbaum, S. R. (1998) Nitric oxide-induced deamination of cytosine and guanine in deoxynucleosides and oligonucleotides. *The Journal of biological chemistry* **273**, 12689-12695

56. Pang, B., McFaline, J. L., Burgis, N. E., Dong, M., Taghizadeh, K., Sullivan, M. R., Elmquist, C. E., Cunningham, R. P., and Dedon, P. C. (2012) Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2319-2324

57. Wiedemann, G. J., Robins, H. I., Gutsche, S., Mentzel, M., Deeken, M., Katschinski, D. M., Eleftheriadis, S., Crahe, R., Weiss, C., Storer, B., and Wagner, T. (1996) Ifosfamide, carboplatin and etoposide (ICE) combined with 41.8 degrees C whole body hyperthermia in patients with refractory sarcoma. *European journal of cancer* **32A**, 888-892

58. O'Driscoll, M., and Jeggo, P. A. (2006) The role of double-strand break repair - insights from human genetics. *Nature reviews. Genetics* **7**, 45-54

59. Ramzan, Z., Nassri, A. B., and Huerta, S. (2014) Genotypic characteristics of resistant tumors to pre-operative ionizing radiation in rectal cancer. *World journal of gastrointestinal oncology* **6**, 194-210

60. Kuzminov, A. (1995) Collapse and repair of replication forks in Escherichia coli. *Molecular microbiology* **16**, 373-384

61. Kogoma, T. (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and molecular biology reviews : MMBR* **61**, 212-238

62. Dresser, M. E. (2000) Meiotic chromosome behavior in Saccharomyces cerevisiae and (mostly) mammals. *Mutation research* **451**, 107-127

63. Fu, D., Calvo, J. A., and Samson, L. D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nature reviews. Cancer* **12**, 104-120

64. Christmann, M., Tomicic, M. T., Roos, W. P., and Kaina, B. (2003) Mechanisms of human DNA repair: an update. *Toxicology* **193**, 3-34

65. Rabik, C. A., Njoku, M. C., and Dolan, M. E. (2006) Inactivation of O6-alkylguanine DNA alkyltransferase as a means to enhance chemotherapy. *Cancer treatment reviews* **32**, 261-276

66. Falnes, P. O., and Rognes, T. (2003) DNA repair by bacterial AlkB proteins. *Research in microbiology* **154**, 531-538

67. Wyatt, M. D., and Pittman, D. L. (2006) Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chemical research in toxicology* **19**, 1580-1594

68. Sedgwick, B., Robins, P., and Lindahl, T. (2006) Direct removal of alkylation damage from DNA by AlkB and related DNA dioxygenases. *Methods in enzymology* **408**, 108-120

69. Drablos, F., Feyzi, E., Aas, P. A., Vaagbo, C. B., Kavli, B., Bratlie, M. S., Pena-Diaz, J., Otterlei, M., Slupphaug, G., and Krokan, H. E. (2004) Alkylation damage

in DNA and RNA--repair mechanisms and medical significance. *DNA repair* **3**, 1389-1407

70. Shankaracharya, Das, S., and Vidyarthi, A. S. (2011) Homology modeling and function prediction of hABH1, involving in repair of alkylation damaged DNA. *Interdisciplinary sciences, computational life sciences* **3**, 175-181

71. Sancar, A. (2003) Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chemical reviews* **103**, 2203-2237

72. Brettel, K., and Byrdin, M. (2010) Reaction mechanisms of DNA photolyase. *Current opinion in structural biology* **20**, 693-701

73. Eker, A. P., Quayle, C., Chaves, I., and van der Horst, G. T. (2009) DNA repair in mammalian cells: Direct DNA damage reversal: elegant solutions for nasty problems. *Cellular and molecular life sciences : CMLS* **66**, 968-980

74. Klungland, A., and Bjelland, S. (2007) Oxidative damage to purines in DNA: role of mammalian Ogg1. *DNA repair* **6**, 481-488

75. van der Kemp, P. A., Thomas, D., Barbey, R., de Oliveira, R., and Boiteux, S. (1996) Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5197-5202

76. Zharkov, D. O., Rosenquist, T. A., Gerchman, S. E., and Grollman, A. P. (2000) Substrate specificity and reaction mechanism of murine 8-oxoguanine-DNA glycosylase. *The Journal of biological chemistry* **275**, 28607-28617

77. Myles, G. M., and Sancar, A. (1989) DNA repair. *Chemical research in toxicology* **2**, 197-226

78. Marenstein, D. R., Wilson, D. M., 3rd, and Teebor, G. W. (2004) Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded DNA. *DNA repair* **3**, 527-533

79. Luo, M., Delaplane, S., Jiang, A., Reed, A., He, Y., Fishel, M., Nyland, R. L., 2nd, Borch, R. F., Qiao, X., Georgiadis, M. M., and Kelley, M. R. (2008) Role of the multifunctional DNA repair and redox signaling protein Ape1/Ref-1 in cancer and endothelial cells: small-molecule inhibition of the redox function of Ape1. *Antioxidants & redox signaling* **10**, 1853-1867

80. Buermeyer, A. B., Deschenes, S. M., Baker, S. M., and Liskay, R. M. (1999) Mammalian DNA mismatch repair. *Annual review of genetics* **33**, 533-564

81. Modrich, P. (1997) Strand-specific mismatch repair in mammalian cells. *The Journal of biological chemistry* **272**, 24727-24730

82. Iyer, R. R., Pluciennik, A., Burdett, V., and Modrich, P. L. (2006) DNA mismatch repair: functions and mechanisms. *Chemical reviews* **106**, 302-323

83. Larrea, A. A., Lujan, S. A., and Kunkel, T. A. (2010) SnapShot: DNA mismatch repair. *Cell* **141**, 730 e731

84. Heller, R. C., and Marians, K. J. (2006) Replisome assembly and the direct restart of stalled replication forks. *Nature reviews. Molecular cell biology* **7**, 932-943

85. Machado-Silva, A., Teixeira, S. M., Franco, G. R., Macedo, A. M., Pena, S. D., McCulloch, R., and Machado, C. R. (2008) Mismatch repair in Trypanosoma brucei:

heterologous expression of MSH2 from Trypanosoma cruzi provides new insights into the response to oxidative damage. *Gene* **411**, 19-26

86. McCulloch, S. D., Gu, L., and Li, G. M. (2003) Bi-directional processing of DNA loops by mismatch repair-dependent and -independent pathways in human cells. *The Journal of biological chemistry* **278**, 3891-3896

87. Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N., and Sixma, T. K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* **407**, 711-717

88. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nature reviews. Molecular cell biology* **7**, 335-346

89. Jiricny, J., and Nystrom-Lahti, M. (2000) Mismatch repair defects in cancer. *Current opinion in genetics & development* **10**, 157-161

90. Hoeijmakers, J. H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374

91. Friedberg, E. C. (2000) Biological responses to DNA damage: a perspective in the new millennium. *Cold Spring Harbor symposia on quantitative biology* **65**, 593-602

92. de Boer, J., and Hoeijmakers, J. H. (2000) Nucleotide excision repair and human syndromes. *Carcinogenesis* **21**, 453-460

93. Boyce, R. P., and Howard-Flanders, P. (1964) Release of Ultraviolet Light-Induced Thymine Dimers from DNA in E. Coli K-12. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 293-300

94. Setlow, R. B., and Carrier, W. L. (2003) The disappearance of thymine dimers from DNA: an error-correcting mechanism. 1963. *DNA repair* **2**, 1274-1279

95. Sugasawa, K. (2011) Multiple DNA damage recognition factors involved in mammalian nucleotide excision repair. *Biochemistry. Biokhimiia* **76**, 16-23

96. Melis, J. P., Luijten, M., Mullenders, L. H., and van Steeg, H. (2011) The role of XPC: implications in cancer and oxidative DNA damage. *Mutation research* **728**, 107-117

97. Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001) A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes & development* **15**, 507-521

98. Wood, R. D. (1999) DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie* **81**, 39-44

99. Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molecular cell* **2**, 223-232

100. Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J., Bootsma, D., and et al. (1994) Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *The EMBO journal* **13**, 1831-1843

101. Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., van der Spek, P. J., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1996) HHR23B, a human Rad23 homolog,

stimulates XPC protein in nucleotide excision repair in vitro. *Molecular and cellular biology* **16**, 4852-4861

102. Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., and Hanaoka, F. (2001) Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *The Journal of biological chemistry* **276**, 18665-18672

103. de Laat, W. L., Jaspers, N. G., and Hoeijmakers, J. H. (1999) Molecular mechanism of nucleotide excision repair. *Genes & development* **13**, 768-785

104. Petit, C., and Sancar, A. (1999) Nucleotide excision repair: from E. coli to man. *Biochimie* **81**, 15-25

105. Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair. *Science* **286**, 1897-1905

106. Lindahl, T., Karran, P., and Wood, R. D. (1997) DNA excision repair pathways. *Current opinion in genetics & development* **7**, 158-169

107. Sancar, A. (1996) DNA excision repair. *Annual review of biochemistry* **65**, 43-81

108. Friedberg, E. C. (1996) Cockayne syndrome--a primary defect in DNA repair, transcription, both or neither? *BioEssays : news and reviews in molecular, cellular and developmental biology* **18**, 731-738

109. Cleaver, J. E. (1968) Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218**, 652-656

110. Setlow, R. B., Regan, J. D., German, J., and Carrier, W. L. (1969) Evidence that xeroderma pigmentosum cells do not perform the first step in the repair of

ultraviolet damage to their DNA. *Proceedings of the National Academy of Sciences of the United States of America* **64**, 1035-1041

111. Gates, F. T., 3rd, and Linn, S. (1977) Endonuclease V of Escherichia coli. *The Journal of biological chemistry* **252**, 1647-1653

112. Hitchcock, T. M., Gao, H., and Cao, W. (2004) Cleavage of deoxyoxanosine-containing oligodeoxyribonucleotides by bacterial endonuclease V. *Nucleic acids research* **32**, 4071-4080

113. Feng, H., Dong, L., Klutz, A. M., Aghaebrahim, N., and Cao, W. (2005) Defining amino acid residues involved in DNA-protein interactions and revelation of 3'-exonuclease activity in endonuclease V. *Biochemistry* **44**, 11486-11495

114. Mi, R., Alford-Zappala, M., Kow, Y. W., Cunningham, R. P., and Cao, W. (2012) Human endonuclease V as a repair enzyme for DNA deamination. *Mutation research* **735**, 12-18

115. He, B., Qing, H., and Kow, Y. W. (2000) Deoxyxanthosine in DNA is repaired by Escherichia coli endonuclease V. *Mutation research* **459**, 109-114

116. Feng, H., Klutz, A. M., and Cao, W. (2005) Active site plasticity of endonuclease V from Salmonella typhimurium. *Biochemistry* **44**, 675-683

117. Feng, H., Dong, L., and Cao, W. (2006) Catalytic mechanism of endonuclease v: a catalytic and regulatory two-metal model. *Biochemistry* **45**, 10251-10259

118. Lee, C. C., Yang, Y. C., Goodman, S. D., Yu, Y. H., Lin, S. B., Kao, J. T., Tsai, K. S., and Fang, W. H. (2010) Endonuclease V-mediated deoxyinosine excision repair in vitro. *DNA repair* **9**, 1073-1079

119. Mi, R., Abole, A. K., and Cao, W. (2011) Dissecting endonuclease and exonuclease activities in endonuclease V from Thermotoga maritima. *Nucleic acids research* **39**, 536-544

120. Morita, Y., Shibutani, T., Nakanishi, N., Nishikura, K., Iwai, S., and Kuraoka, I. (2013) Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nature communications* **4**, 2273

121. Vik, E. S., Nawaz, M. S., Strom Andersen, P., Fladeby, C., Bjoras, M., Dalhus, B., and Alseth, I. (2013) Endonuclease V cleaves at inosines in RNA. *Nature communications* **4**, 2271

122. Sung, P., and Klein, H. (2006) Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature reviews. Molecular cell biology* **7**, 739-750

123. Takata, M., Sasaki, M. S., Sonoda, E., Fukushima, T., Morrison, C., Albala, J. S., Swagemakers, S. M., Kanaar, R., Thompson, L. H., and Takeda, S. (2000) The Rad51 paralog Rad51B promotes homologous recombinational repair. *Molecular and cellular biology* **20**, 6476-6482

124. Kanaar, R., Hoeijmakers, J. H., and van Gent, D. C. (1998) Molecular mechanisms of DNA double strand break repair. *Trends in cell biology* **8**, 483-489

125. Baumann, P., and West, S. C. (1998) Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends in biochemical sciences* **23**, 247-251

126. Sung, P., Trujillo, K. M., and Van Komen, S. (2000) Recombination factors of Saccharomyces cerevisiae. *Mutation research* **451**, 257-275

127. Thompson, L. H., and Schild, D. (1999) The contribution of homologous recombination in preserving genome integrity in mammalian cells. *Biochimie* **81**, 87-105

128. Ma, Y., Lu, H., Schwarz, K., and Lieber, M. R. (2005) Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. *Cell cycle* **4**, 1193-1200

129. Smith, G. C., and Jackson, S. P. (1999) The DNA-dependent protein kinase. *Genes & development* **13**, 916-934

130. Cha, R. S., and Kleckner, N. (2002) ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science* **297**, 602-606

131. Casper, A. M., Nghiem, P., Arlt, M. F., and Glover, T. W. (2002) ATR regulates fragile site stability. *Cell* **111**, 779-789

132. Thompson, L. H., and Schild, D. (2001) Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutation research* **477**, 131-153

133. Krokan, H. E., Drablos, F., and Slupphaug, G. (2002) Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**, 8935-8948

134. Lee, H. W., Dominy, B. N., and Cao, W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *The Journal of biological chemistry* **286**, 31282-31287

135. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-493

136. Sekino, Y., Bruner, S. D., and Verdine, G. L. (2000) Selective inhibition of herpes simplex virus type-1 uracil-DNA glycosylase by designed substrate analogs. *The Journal of biological chemistry* **275**, 36506-36508

137. Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L., and Stivers, J. T. (1999) Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **38**, 11876-11886

138. Dong, J., Drohat, A. C., Stivers, J. T., Pankiewicz, K. W., and Carey, P. R. (2000) Raman spectroscopy of uracil DNA glycosylase-DNA complexes: insights into DNA damage recognition and catalysis. *Biochemistry* **39**, 13241-13250

139. Drohat, A. C., and Stivers, J. T. (2000) Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry* **39**, 11865-11875

140. Jiang, Y. L., Drohat, A. C., Ichikawa, Y., and Stivers, J. T. (2002) Probing the limits of electrostatic catalysis by uracil DNA glycosylase using transition state mimicry and mutagenesis. *The Journal of biological chemistry* **277**, 15385-15392

141. Jiang, Y. L., Ichikawa, Y., and Stivers, J. T. (2002) Inhibition of uracil DNA glycosylase by an oxacarbenium ion mimic. *Biochemistry* **41**, 7116-7124

142. Jiang, Y. L., Cao, C., Stivers, J. T., Song, F., and Ichikawa, Y. (2004) The merits of bipartite transition-state mimics for inhibition of uracil DNA glycosylase. *Bioorganic chemistry* **32**, 244-262

143. Bianchet, M. A., Seiple, L. A., Jiang, Y. L., Ichikawa, Y., Amzel, L. M., and Stivers, J. T. (2003) Electrostatic guidance of glycosyl cation migration along the reaction coordinate of uracil DNA glycosylase. *Biochemistry* **42**, 12455-12460

144. Werner, R. M., Jiang, Y. L., Gordley, R. G., Jagadeesh, G. J., Ladner, J. E., Xiao, G., Tordova, M., Gilliland, G. L., and Stivers, J. T. (2000) Stressing-out DNA? The contribution of serine-phosphodiester interactions in catalysis by uracil DNA glycosylase. *Biochemistry* **39**, 12585-12594

145. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *The EMBO journal* **17**, 5214-5226

146. Shroyer, M. J., Bennett, S. E., Putnam, C. D., Tainer, J. A., and Mosbaugh, D. W. (1999) Mutation of an active site residue in Escherichia coli uracil-DNA glycosylase: effect on DNA binding, uracil inhibition and catalysis. *Biochemistry* **38**, 4834-4845

147. Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T., and Gilliland, G. L. (1999) Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited. *Proteins* **35**, 13-24

148. Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413**, 752-755

149. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. (2000) Uracil-DNA glycosylase-DNA substrate and

product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5083-5088

150. Shaw, R. W., Feller, J. A., and Bloom, L. B. (2004) Contribution of a conserved phenylalanine residue to the activity of Escherichia coli uracil DNA glycosylase. *DNA repair* **3**, 1273-1283

151. Berti, P. J., and McCann, J. A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chemical reviews* **106**, 506-555

152. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K., and Jiricny, J. (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *The Journal of biological chemistry* **271**, 12767-12774

153. Baba, D., Maita, N., Jee, J. G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**, 979-982

154. Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *The Journal of biological chemistry* **275**, 33449-33456

155. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307

156. Hashimoto, H., Zhang, X., and Cheng, X. (2013) Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *Journal of molecular biology* **425**, 971-976

157. Hashimoto, H., Zhang, X., and Cheng, X. (2013) Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH. *DNA repair* **12**, 535-540

158. Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L., and Pearl, L. H. (2003) Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Molecular cell* **11**, 1647-1659

159. Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., Zhang, Y., and Kohli, R. M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature chemical biology* **8**, 751-758

160. Jobert, L., Skjeldam, H. K., Dalhus, B., Galashevskaya, A., Vagbo, C. B., Bjoras, M., and Nilsen, H. (2013) The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control. *Molecular cell* **49**, 339-345

161. Matsubara, M., Tanaka, T., Terato, H., Ohmae, E., Izumi, S., Katayanagi, K., and Ide, H. (2004) Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase. *Nucleic acids research* **32**, 5291-5302

162. Hinks, J. A., Evans, M. C., De Miguel, Y., Sartori, A. A., Jiricny, J., and Pearl, L. H. (2002) An iron-sulfur cluster in the family 4 uracil-DNA glycosylases. *The Journal of biological chemistry* **277**, 16936-16940

163. Sartori, A. A., and Jiricny, J. (2003) Enzymology of base excision repair in the hyperthermophilic archaeon Pyrobaculum aerophilum. *The Journal of biological chemistry* **278**, 24563-24576

164. Sandigursky, M., and Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes. *Current biology : CB* **9**, 531-534

165. Sandigursky, M., Faje, A., and Franklin, W. A. (2001) Characterization of the full length uracil-DNA glycosylase in the extreme thermophile Thermotoga maritima. *Mutation research* **485**, 187-195

166. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H., and Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *The EMBO journal* **21**, 3182-3191

167. Wanner, R. M., Castor, D., Guthlein, C., Bottger, E. C., Springer, B., and Jiricny, J. (2009) The uracil DNA glycosylase UdgB of Mycobacterium smegmatis protects the organism from the mutagenic effects of cytosine and adenine deamination. *Journal of bacteriology* **191**, 6312-6319

168. Starkuviene, V., and Fritz, H. J. (2002) A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus. *Nucleic acids research* **30**, 2097-2102

169. Sakai, T., Tokishita, S., Mochizuki, K., Motomiya, A., Yamagata, H., and Ohta, T. (2008) Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus. *DNA repair* **7**, 663-669

170. Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S., and Masui, R. (2007) Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. *Journal of molecular biology* **373**, 839-850

CHAPTOR TWO

SPECIFICITY AND CATALYTIC MECHANISM IN FAMILY 5 URACIL DNA GLYCOSYLASE

I. Abstract

UDGb belongs to family 5 of the uracil DNA glycosylase (UDG) superfamily. Here, we report that family 5 UDGb from *Thermus thermophilus* HB8 is not only a uracil DNA glycosyase acting on G/U, T/U, C/U and A/U base pairs, it is also a hypoxanthine DNA glycosylase acting on G/I, T/I and A/I base pairs and a xanthine DNA glycosylase acting on all double-stranded and single-stranded xanthine-containing DNA. Analysis of potentials of mean force indicates that the tendency of hypoxanthine base flipping follows the order of G/I > T/I, A/I > C/I, matching the trend of hypoxanthine DNA glycosylase activity observed *in vitro*. Genetic analysis indicates that family 5 UDGb can also act as an enzyme to remove uracil incorporated into DNA through the existence of dUTP in the nucleotide pool. Mutational analysis coupled with molecular modeling and molecular dynamics analysis reveal that while hydrogen bonding to O2 of uracil underlies the UDG activity in a dissociative fashion, Tth UDGb relies on multiple catalytic residues to facilitate its excision of hypoxanthine and xanthine. This study underscores the structural and functional diversity in the UDG superfamily.

II. Introduction

DNA base deamination is a common mechanism of DNA damage caused by environmental and endogenous agents. Due to the reactivity of the exocyclic amino groups, DNA bases are subject to hydrolytic or oxidative deamination in which the amino group in

a DNA base is converted to keto group. Adenine (A), cytosine (C) and guanine (G) are deaminated to hypoxanthine (I), uracil (U), and xanthine (X) and oxanine (O), respectively (Fig. 2.1A). Due to the altered base pair preferences, base deamination may result in mutations.

The base excision repair (BER), initiated by DNA glycosylase, is a major pathway to repair damage caused by base deamination. The uracil DNA glycosylase (UDG) superfamily, consisting of six families, is involved in the repair of deaminated base damage. While family 1 UNGs show rather narrow specificity towards uracil and its derivatives (1,2), family 2 MUG/TDG enzymes have much broader specificity with some members showing activity toward all deaminated bases (3-6). Most interestingly, recent studies have indicated that human TDG is a DNA glycosylase that is involved in the removal of formyl-C (fC) and carboxyl-C (caC) during enzymatic demethylation (7,8), suggesting that some of the enzymes in the UDG superfamily have evolved functions beyond DNA repair. Family 3 SMUG1 enzymes, found in vertebrates and bacteria, act as both uracil DNA glycosylases and xanthine DNA glycosylases (XDG) (9,10). Family 4 UDGa was initially discovered in the hyperthermophilic bacterium *Thermotoga maritima* (11). Family 5 UDGb was first reported in the hyperthermophilic archaean *Pyrobaculum aerophilum* (12). Most recently, the family 6 enzymes was discovered as a class of enzymes with hypoxanthine DNA glycosylase (HDG) activity but not with uracil DNA glycosylase activity (13).

Family 5 UDGb exists in archaea and bacteria, many of which are hyperthermophiles or thermophiles. Biochemical characterization of *P. aerophilum* UDGb indicates that it can remove uracil, hydroxymethyluracil or fluorouracil opposite from

guanine base and hypoxanthine from T/I base pairs (12). The uracil DNA glycosylase and hypoxanthine DNA glycosylase activity prevents mutation resulting from cytosine and adenine deamination in cells (14). UDGb from *Mycobacterium tuberculosis* is also found to be active on ethenocytosine and the mutation rate in the absence of UDGb increases by two-fold (15,16). Similarly, biochemical and genetic analyses indicate that UDGb from thermophilic bacterium *Thermus thermophilus* (Tth) is a uracil DNA glycosylase that can reduce the mutation rate by three-fold (17,18). The crystal structure of UDGb from *T. thermophilus* confirms that UDGb adopts a structural fold similarly seen in other UDG enzymes (19).

Previous studies have provided valuable information on the structure-function relationship of Family 5 UDGb enzymes and their physiological roles. However, some fundamental questions remain to be answered. How broad is the specificity of UDGb towards other deaminated bases? How does the active site in UDGb catalyze the cleavage of the glycosidic bond in deaminated DNA? To answer these questions, we conducted a comprehensive biochemical, genetic, and molecular dynamics analysis using UDGb from *T. thermophilus* (Tth) as a model. Data presented here demonstrate that Tth UDGb can act as a uracil DNA glycosylase with enzymatic activity on double-stranded uracil-containing DNA, as a hypoxanthine DNA glycosylase with enzymatic activity on double-stranded hypoxanthine-containing DNA except for the C/I base pair, and as a xanthine DNA glycosylase with enzymatic activity on both double-stranded and single-stranded xanthine-containing DNA. This study also establishes the correlation between hypoxanthine DNA glycosylase activity and stability of hypoxanthine-containing base pairs; reveals the inverse correlation between the uracil DNA glycosylase activity against the A/U base pair and cell

survival in an *E. coli* cell deficient in *ung dut xth* genes; and identifies several catalytic residues that play an important role in the removal of deamianted bases in DNA. A model explaining the catalytic function of family 5 UDGb is proposed.

III. Experimental procedures

Cloning, Expression and Purification of Tth UDGb - The UDGb gene from *T. thermophilus* HB8 (GenBank accession number: YP_144415.1) was amplified by PCR using the forward primer Tth UDGbF (5'-TCAGGTGTGCATATGGACAG GGAAGCCTTCGTCCAAACC-3'; the NdeI site is underlined) and the reverse primer Tth UDGbR (5'-TGAATCAAGCTTAAGCCCGGCGAGGCG TTTAGCCTC-3'; the HindIII site is underlined). The PCR reaction mixture (20 μl) consisted 10 ng *T. thermophilus* HB8 genomic DNA, 500 nM forward and reverse primers, 1 x phusion DNA polymerase buffer, 200 μM each dNTP and 0.2 unit of phusion DNA polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 98°C for 30 s; 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, and extension at 72°C for 20 s; and a final extension step at 72°C for 10 min. The PCR product was purified by gel DNA recovery kit (Zymo Research). The purified PCR product and plasmid pET21a were digested by NdeI and HindIII, purified by gel DNA recovery kit, and ligated according to the manufacturer's instructional manual. The ligation mixture was transformed into *E. coli* strain HB101 competent cells by electroporation. The sequence of the Tth UDGb gene in the resulting plasmid (pET21a-Tth-UDGb) was confirmed by DNA sequencing. Site-directed mutagenesis was performed similarly as previously described (13).

The pET21a-Tth-UDGb was transformed into *E. coli* strain BH214 (*ung-*, *mug-*) by the standard protocol to express the C-terminal His-6-tagged Tth UDGb protein. Induction, sonication and purification were carried out as previously described with the following modifications (13). Prior to Hitrap chelating column chromatography, the sonicated solutions were incubated at 75°C for 15 min. Denatured proteins were removed by centrifugation at 12,000 rpm for 20 min. Fractions (300–400 mM imidazole, 60-80% chelating buffer B) containing the Tth UDGb protein as seen on 12.5% SDS-PAGE were pooled and concentrated by Amicon YM-10 (Millipore).

Oligodeoxynucleotide Substrates - Oligodeoxynucleotides containing deoxyuridine (U), deoxyinosine (I), deoxyxanthosine (X) or deoxyoxanosine (O) were obtained or constructed as previously described (Fig. 2.1B) (10).

DNA Glycosylase Activity Assay - DNA glycosylase cleavage assays for Tth UDGb were performed under optimized reaction conditions at 50°C for 60 min in a 10 µl reaction mixture containing 10 nM oligonucleotide substrate, an indicated amount of glycosylase, 20 mM Tris-HCl (pH 7.6), 100 mM KCl, 1 mM DTT, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95°C for 5 min after adding 1 µl of 1 N NaOH. The reaction mixtures (2 µl) were mixed with 7.8 µl Hi-D formamide and 0.2 µl GeneScan 500 LI Size Standard (Life Technologies) and analyzed by Applied Biosystems 3130xl sequencer with a fragment analysis module. Cleavage products and remaining substrates were quantified by GeneMapper software. For kinetics measurements, the reactions were carried out with 10 nM oligonucleotide substrate and 100 nM glycosylase. Increasing the enzyme concentration by three-fold did not change the kinetic profile, suggesting that the reactions reached saturation condition. Samples were withdrawn at 0, 2.5, 5, 10, 20, 40 and

60 min. For some substrates such as G/U and G/I with faster kinetics, samples were also withdrawn at 1.0 min time point to capture the early reaction. The apparent rate constants were determined by curve fitting using the integrated first-order rate equation:

$$P = P_{max}(1-e^{-kt})$$

Where P is the product yield, P_{max} is the maximal yield, t is time and *k* is apparent rate constant.

Base-Flipping Potential of Mean Force Calculations - Base flipping potentials of mean force (PMF) were constructed based on the dodecamer sequence d(GTCAGIGCATGG), where the hypoxanthine (I) was the base to be flipped out of the helix. Using the program 3DNA (20), the canonical B-form DNA structure of the sequence $d(GTCAGCGCATGG)_2$ was constructed. The base complementary to I was systematically modeled as guanine, adenine, cytosine and thymine. Starting from these four models of B-form DNA, umbrella sampling was performed to calculate the PMF associated with flipping deaminated DNA bases out of the double helix (21,22). A detailed description of the computational methods can be found in our previous work (21).

Survival Analysis of E. coli BW276 Strain Complement with Tth udgb - The wild type and mutant Tth *udgb* genes were subcloned into plasmid pBluescript II SK (+). The wt Tth *udgb* gene is described as an example. The wt Tth *udgb* gene was amplified by PCR using the plasmid pET21a-Tth-UDGb as a template and with the forward primer TthUDGbF-pBS (5'-CCGGAATTCCCATATGGACAGGGAAGCCTTCG-3'; the EcoRI site is underlined) and the reverse primer Tth UDGbR-pBS (5' ACGCGTCGACTCAGTGGTGGTGGTGGTGG-3'; the SalI site is underlined). The PCR

conditions were the same as the cloning of the Tth UDGb gene. The purified PCR products digested with a pair of EcoRI and SalI endonucleases were ligated to the cloning vector pBluescript II SK (+) treated with the same pair of restriction endonucleases. The recombinant plasmids were confirmed by DNA sequencing and transformed into *E. coli* BW276 (*dut*^{ts} *ung*$^{-}$ *xth*$^{-}$). *E. coli* BW276 (*thi-1, relA1, spoT1, dut-1, rfa-209*::Tn*10, ung-1* (*xth-pncA*)*90*) was a kind gift from Dr. Bernard Weiss (Emory University, Atlanta, GA) (23). Single colonies from each variant were inoculated into 1 ml LB medium (with 100 μg/ml ampicillin and 125 μg/ml thymidine) and grown at 22°C for 16 h. Overnight cultures (1 ml) were transferred to 4 ml fresh LB medium (with 100 μg/ml ampicillin and 125 μg/ml thymidine) and grown at 22°C until OD_{600} reached 0.6. After adding IPTG to a final concentration of 1 mM, the cultures were incubated at 22°C until OD_{600} reached 1.0. Before plating, 40 μl of 100 mM IPTG was spread on the LB plates containing ampicillin and thymidine. Afterwards, diluted cells (100 μl) were plated on LB plates. Cell numbers were scored after 24 h incubation at 42°C or after 72 h incubation at 22°C. The relative plating efficiencies (RPE) were taken as the ratios of the cell numbers between 42°C and 22°C.

Molecular Modeling and Molecular Dynamics Simulations - The crystal structure of family 5 Uracil DNA glycosylase (Tth UDGb) was acquired from the RCSB Protein Data Bank (accession code 2DEM), and used as a model for subsequent computational analysis (19). The initial structures of uridine, inosine, xanthosine and oxanosine nucleotides bound to the family 5 UDG enzyme were obtained by manually modifying the structure of 2-deoxy-5-phosphono-ribose extracted from the crystal structure of Tth UDGb using the Swiss-Pdb Viewer (SPDBV) program (24). Specifically, the 2-deoxy-5-phosphono-ribose was aligned to a cytidine and the cytosine was changed to uracil,

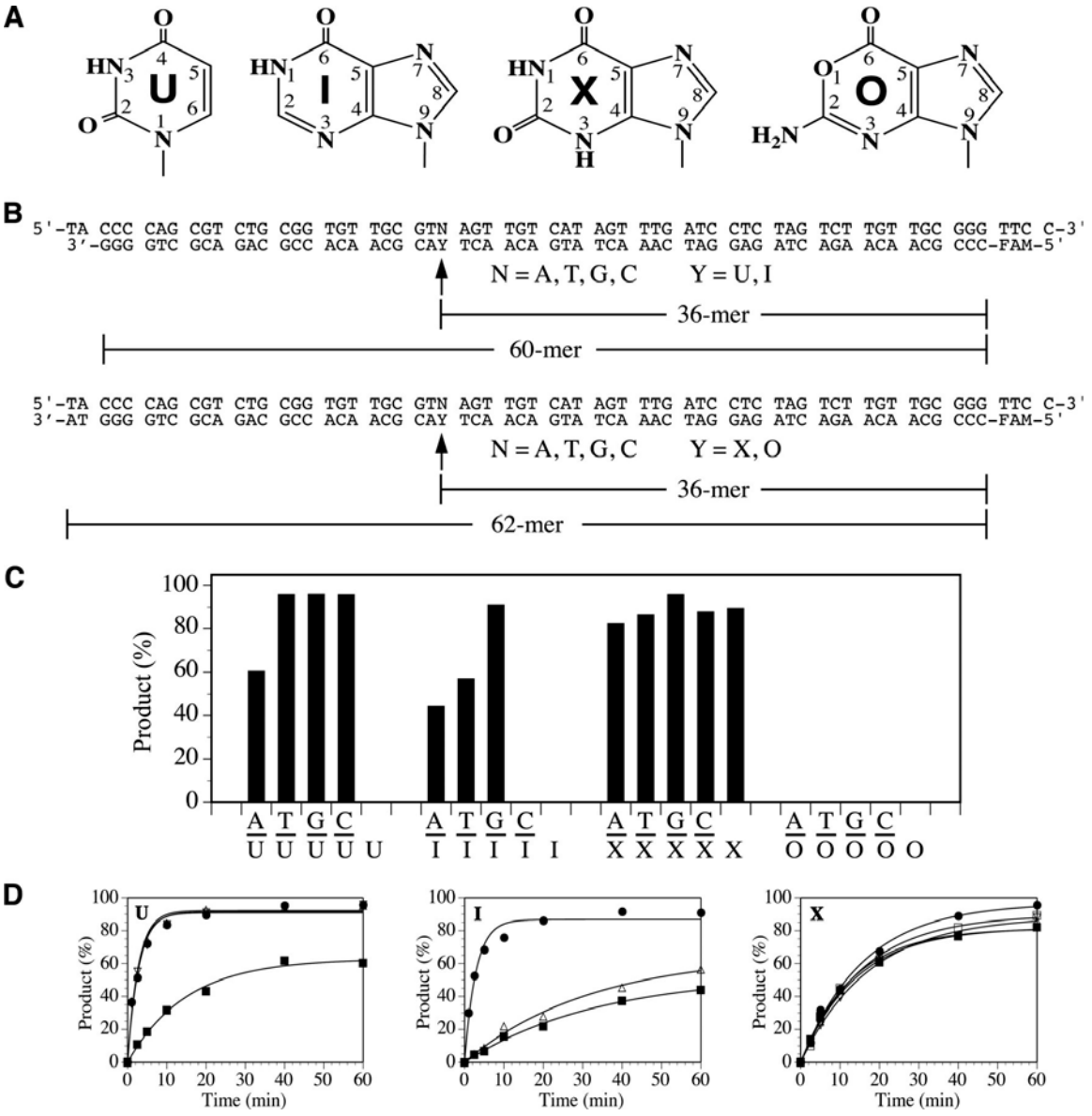


Figure 2.1 Deaminated DNA repair activity in Tth UDGb. *A*, chemical structures of deaminated DNA bases. *B*, sequences of xanthine (X)-, oxanine (O)-, hypoxanthine (I)-, and uracil (U)-containing oligodeoxyribonucleotide substrates. *FAM*, fluorophore. *C*, DNA glycosylase activity of Tth UDGb on U-, I-, O-, and X-containing substrates. Cleavage reactions were performed as described under "Experimental Procedures" with 100 nm WT Tth UDGb protein and 10 nm substrate. *D*, representative time course analysis of glycosylase activity of WT Tth UDGb on U-, I-, and X-containing DNA substrates. U-containing substrates: ■, A/U; ▵, T/U; ●, G/U; ▿, C/U. I-containing substrates: ■, A/I; ▵, T/I; ●, G/I. X-containing substrates: ■, A/X; ▵, T/X; ●, G/X; ▿, C/X; □, ss X.

hypoxanthine, xanthine and oxanine, respectively. After building the initial complex structures, an explicit solvent system using the TIP3P water model was constructed in the CHARMM c35b6 molecular mechanics package using a suitably sized box (25). The minimum distance between any of the atoms of the solvated UDG/DNA complex and the box boundary was maintained to at least 9 Å. Sodium chloride ions were added to the system to achieve an electrically neutral system. The CHARMM 27 all hydrogen force field for proteins and nucleic acids were used (26,27). Particle-mesh Ewald summation was applied in the periodic boundaries condition for the efficient calculation of long-range electrostatic interactions (28). Energy minimization was performed by using 4000 steepest descent steps followed by adopted basis Newton-Raphson (ABNR) method with the harmonic constraints from 10 to 1 kcal/(mol•$Å^2$) in decrements of 3 kcal/(mol•$Å^2$) every 1000 steps to remove any unfavorable van der Waals clashes. Using a Langevin barostat (29), an isothermal-isobaric ensemble (NPT) at 300 K was constructed using the NAMD program (30). An integration time step of 1 fs was used in order to avoid any significant structural deformation during equilibration and production run. Coordinates were saved every 2 ps. A total of 2 ns equilibration and 3 ns production simulation were performed for each structural analysis. VMD 1.9.1 was used for visualization (31).

IV. Results

Substrate Specificity of Tth UDGb - Previous studies have indicated that family 5 UDGb enzymes can act on some deaminated base damage. To gain a complete understanding of the specificity of the family 5 UDGb on deaminated bases, we measured the repair activity of Tth UDGb using all twenty possible uracil (U)-, hypoxanthine (I)-, oxanine (O)-, xanthine (X)-containing double-stranded and single-stranded DNA

substrates under the assay conditions in which the enzyme was in ten-fold excess over the substrate (Fig. 2.1C). Consistent with previous reports, Tth UDGb acted as a double-stranded uracil DNA glycosylase with a relatively low activity on the A/U base pair (Fig. 2.1C). No activity on single-stranded U-containing substrate was detected under the assay conditions. Tth UDGb could also act as a hypoxanthine DNA glycosylase with the strongest activity on the G/I base pair but no activity detected on the C/I base pair and the single-stranded I-containing substrate (Fig. 2.1C). The enzyme showed no activity on any of the oxanine-containing substrates, indicating that family 5 Tth UDGb is not an oxanine DNA glycosylase. On the other hand, Tth UDGb was able to incise xanthine in all xanthine-containing DNA, in which all five substrates were hydrolyzed to close to completion (Fig. 2.1C). Similar to some other enzymes in the UDG superfamily, no glycosylase activities were detected with the deaminated substrates under the assay conditions in which the enzyme: substrate ratio was 1:1 (data not shown). Consistent with previous observations, Tth UDGb was not inhibited by Ugi peptide and acted as a monofunctional glycosylase ((12) and data not shown).

To determine the catalytic efficiencies on the deaminated DNA, we measured the apparent rate constants. A representative time course analysis is shown in Figure 2.1D and complete data are summarized in Table 2.1. For the U-containing substrates, Tth UDGb

Table 2.1 Apparent rate constants of Tth UDGb wild type and mutants

Apparent rate constants of DNA glycosylase activity in wild type and mutant Tth UDGb enzymes

The reactions were performed as described under "Experimental Procedures" with 100 nM Tth UDGb protein and 10 nM substrate. The data are averages of three independent experiments. NA, no activity was detected under the assay conditions. ND, not determined because of a low level of activity.

		Rate constant				
	Bottom strand	Top strand A	Top strand T	Top strand G	Top strand C	
				min^{-1}		
WT	U	0.067 ± 0.0048	0.37 ± 0.031	0.35 ± 0.032	0.36 ± 0.032	NA
D75A		NA	0.028 ± 0.0040	0.031 ± 0.0063	ND	NA
N120A		NA	0.0065 ± 0.00053	0.095 ± 0.0052	0.029 ± 0.0037	NA
WT	I	0.030 ± 0.0042	0.031 ± 0.00059	0.35 ± 0.014	NA	NA
D75A		ND	ND	0.0061 ± 0.00043	NA	NA
N120A		NA	NA	0.0074 ± 0.00029	NA	NA
H190A		NA	NA	0.0022 ± 0.00063	NA	NA
WT	X	0.075 ± 0.0027	0.071 ± 0.0078	0.063 ± 0.0074	0.057 ± 0.0041	0.064 ± 0.0074

was most active on three mismatched base pairs (T/U, C/U, G/U) with apparent rate constants around 0.36 per min (Table 2.1). The apparent rate constant for the A/U pair was five-fold slower than that for G/U pair. For the I-containing substrates, the apparent rate constant of 0.35 per min for the G/I base pair was comparable to the mismatched U-containing base pairs (Table 2.1). However, the apparent rate constants for the T/I and A/I base pairs were more than ten-fold lower. For the X-containing substrates, the apparent rates constants ranged from 0.057 per min for the C/X base pair to 0.075 per min for the A/X base pair (Table 2.1).

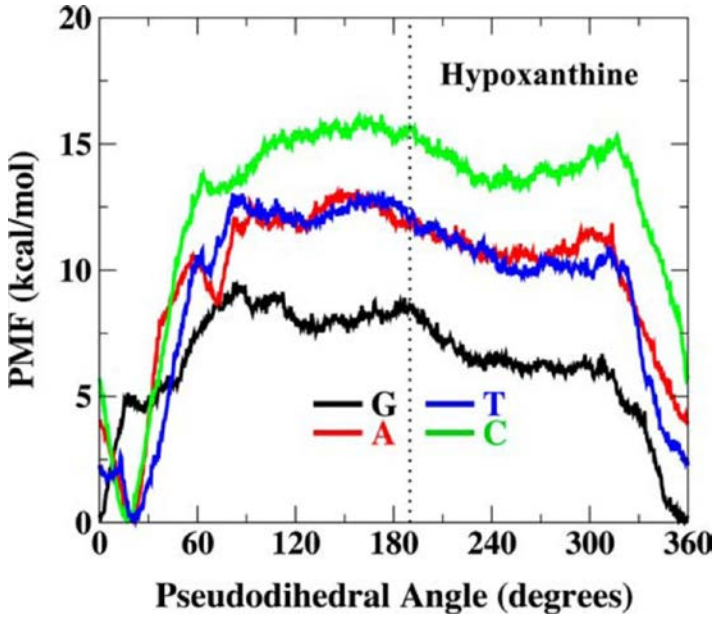
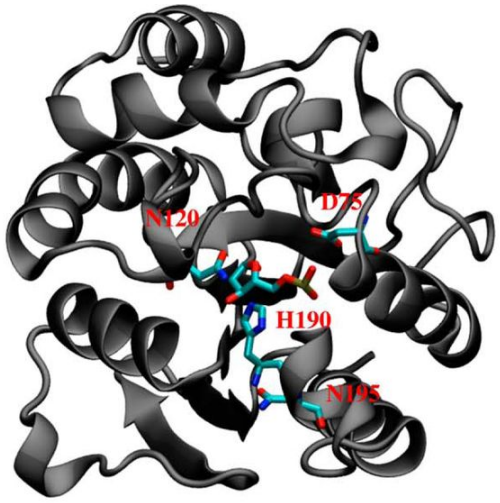


Figure 2.2 PMFs of hypoxanthine-containing base pairs along the pseudo-dihedral angle coordinate. Watson-Crick base pairing is ~10–30° pseudodihedral angle, and the flipped out state is ~190°. PMF profiles were generated in TIP3P explicit water solvent.

HDG Activity and Stability of I-containing Base Pairs – The base flipping tendencies for uracil- and xanthine-containing base pairs have been studied previously (21). The UDG and XDG activity of Tth UDGb followed the general trend of base pair stability. The kinetic analysis indicated that the hypoxanthine DNA glycosylase in Tth UDGb was

A



B

		D75	N120 H190 N195
Family 5 (UDGb)	Tth	*N*--56-GLAPGAHGSNRTGRPFTGDASGAF-30-AAVRCAPPKN-69-HVSRQNT--23-*C*	
	Pae	*N*--65-GLAPAAHGGNRTGRMFTGDASAQN-31-SAVKCAPPKN-65-HPSPLNV--24-*C*	
	Sso	*N*--46-GLAPAGNGGNRTGRMFTGDESSNN-31-SAVKCAPPQN-76-HPSPRNM--25-*C*	
	Tvo	*N*--61-GLAPAATGGNRTGRVFTGDKSSDF-31-AAVKCVPPDN-71-HPSPRNV--23-*C*	
	Sco	*N*--60-GLAPAAHGGNRTGRMFTGDRSGDV-31-SPVHCAPPAN-79-HVSQRNT--26-*C*	
	Mtu	*N*-101-GLAPAAHGANRTGRMFTGDRSGDQ-31-APVRCAPPGN-71-HPSQQNM--24-*C*	
Family 1 (UNG)	Eco	*N*--61-GQDPYHGPGQAHGLAFSVRPGIAT-37-NTVLTVRAGQ-54-HPSPLSA--36-*C*	
Family 2 (MUG/TDG)	Eco	*N*--15-GINPGLSSAG-TGFPFAHPANRFW-29-KLVDRPTVQA-62-NPSGLSR--22-*C*	
Family 3 (SMUG1)	Gme	*N*--55-GMNPGPWGMAQTGVPFGEVAVVTE-56-NYCPLLFLTA-64-HPSPASP--21-*C*	
Family 4 (UDGa)	Tth	*N*--39-GEGPGEEEDK-TGRPFVGKAGQLL-17-NIVKCRPPQN-65-HPAYLLR--44-*C*	
Family 6 (HDG)	Mba	*N*--19-GSLPGDVSIR-KHQYYGHPGNDFW-31-DVFKAGKREG-52-SSSGANR--16-*C*	
		Motif 1	**Motif 2**

Figure 2.3 Tth UDGb structure and sequence alignment.*A*, Tth UDGb-AP site co-crystal structure (Protein Data Bank code 2DEM). The four residues (Asp[75], Asn[120], His[190], and Asn[195]) that were subjected to mutational analysis are shown in *red. B*, sequence alignment in family 5 UDGb and comparison with other UDG families. The alignment was based on BLAST and CLUSTALW analysis and constructed manually. Four conserved residues that are matched in *A* are shown in *red*. Family 5 (UDGb): *Tth*, *T. thermophilus* HB8, YP_144415.1; *Pae*, *P. aerophilum* str. IM2, NP_559226; *Sso*, *Sulfolobus solfataricus* P2, NP_344053.1; *Tvo*, *Thermoplasma volcanium* GSS1, NP_111346.1; *Sco*, *Streptomyces coelicolor* A3(2), NP_626251.1; *Mtu*, *M. tuberculosis* H37Rv, P64785 (Rv1259). Family 1 (UDG): *Eco*, *E. coli*, NP_289138. Family 2 (MUG/TDG): *Eco*, *E. coli*, P0A9H1. Family 3 (SMUG1): *Gme*, *G. metallireducens* GS-15,YP_383069. Family 3 (SMUG1): *Gme*, *G. metallireducens* GS-15,YP_383069; Family 4 (UDGa): *Tth*, *T. thermophilus* HB27, YP_004341.1.

strongest on the G/I base pair followed by the T/I and A/I base pairs. The enzyme was least active with the C/I base pair. To assess how the order of HDG activity is correlated with the stability of the I-containing base pairs, we analyzed potentials of mean force (PMF) for the double-stranded I-containing base pairs. The PMF essentially represents the change in free energy associated with transitioning the DNA from the base paired state to base flipped-out state, and is therefore capable of indicating the tendency of a deaminated base to flip out of an isolated B-form DNA double helix. Among the four base pairs, hypoxanthine demonstrates the greatest tendency (lowest PMF barrier) to flip when paired with guanine (Fig. 2.2). The thermodynamic stability of T/I and A/I base pairs were found to be similar to each other and more stable (higher PMF barrier) than the G/I pair. The most stable base pair is the C/I base pair, which can adopt a natural Watson-Crick base pair conformation. Interestingly, the PMF profile of the I-containing base pairs is quite consistent with activity pattern (Fig. 2.1 and Table 2.1), suggesting that the tendency of base flipping contributes to the recognition and consequently the catalysis.

Catalytic Residues in UDGb - To identify the amino acid residues in the family 5 Tth UDGb enzyme that may play an important role in its catalytic function, we took advantage of the Tth UDGb crystal structure complexed with an abasic site (19). The presence of the AP site in the structure helped us to locate potential catalytic residues that are in the vicinity of the scissile bond. Aspartate and asparagine are known catalytic residues in family 1 UNG and family 2 MUG, respectively. A close examination of the structure revealed several potential catalytic residues: D75, N120, H190 and N195 (Fig. 2.3A). All these residues are highly conserved in family 5 UDGb homologs (Fig. 2.3B). To test the role of these residues in catalysis, D75 was substituted with A, N, E and Q;

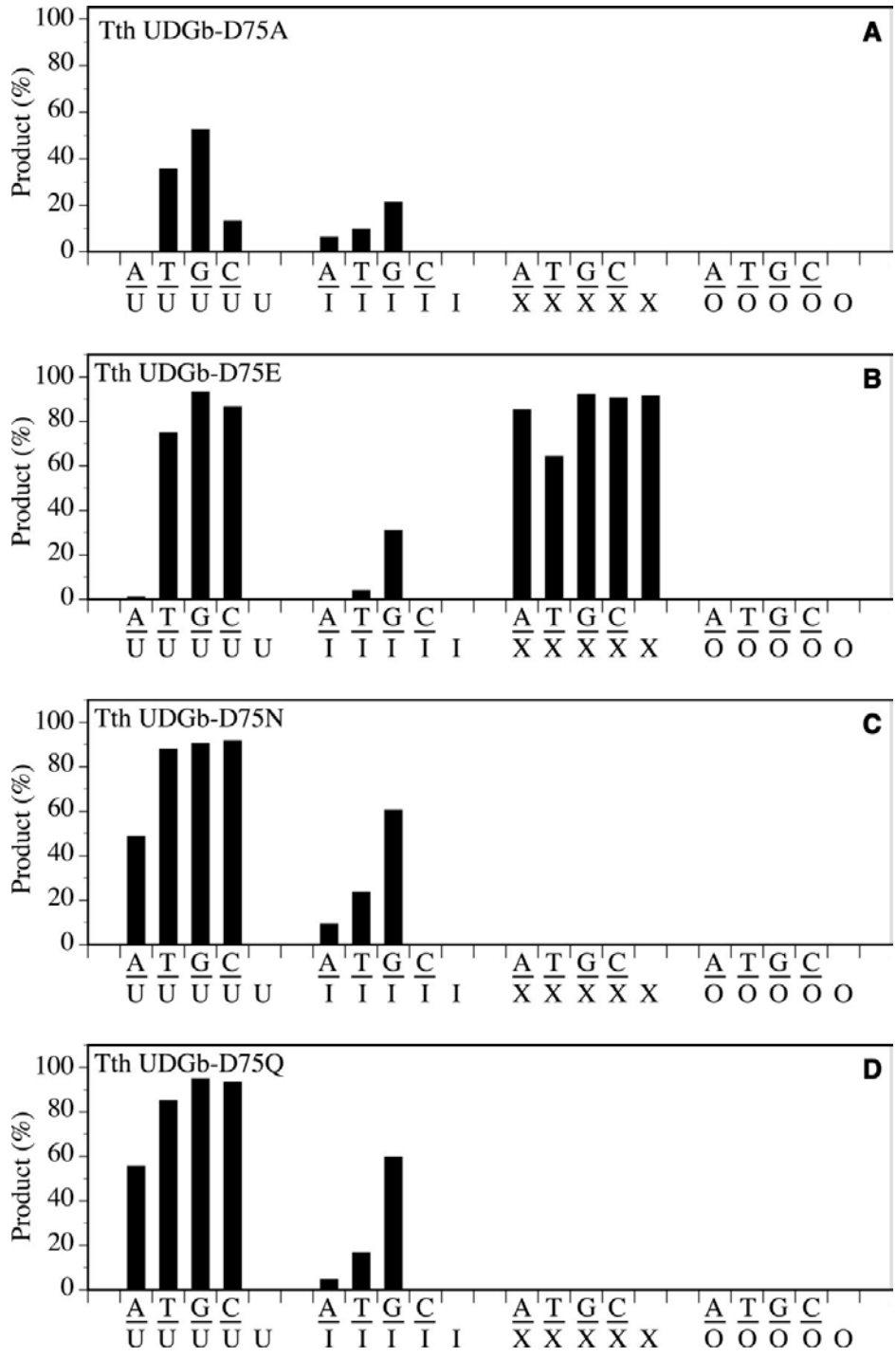


Figure 2.4 Glycosylase activity of Tth UDGb D75 mutants on U-, I-, X-, and O-containing DNA substrates. Cleavage reactions were performed as described under "Experimental Procedures" at 50 °C for 60 min with 100 nm protein and 10 nm substrate. *A*, Tth UDGb-D75A. *B*, Tth UDGb-D75E. *C*, Tth UDGb-D75N. *D*, Tth UDGb-D75Q.

N120 and N195 with A, D, E and Q; and H190 with A, N and S.

For the D75 position, substitution with an Ala residue exhibited the most profound effect on X-containing substrates since D75A completely lost all of its XDG activity (Fig. 2.4A). The UDG activity was also reduced, in particular for the A/U and C/U base pairs in which the activity was not detectable under the assay conditions or rather weak (Fig. 2.4A). For the T/U and G/U base pairs, the apparent rate constants were reduced by more than ten-fold (Table 2.1). For the I-containing substrates, the HDG activity on A/I and T/I was rather weak and that on G/I was reduced by fifty-seven-fold (Fig. 2.4A and Table 2.1). The effect caused by substitution of D75 with Glu was in general much less severe than D75A as the D75E mutant still retained much of the activity (Fig. 2.4B). The substitutions of the carboxyl side chain with an amide side chain resulted in similar effects. Both D75N and D75Q mutants lost XDG activity while retaining much of their UDG and HDG activity (Fig. 2.4C-D).

The effect caused by an Ala substitution at the N120 position was in general similar to D75A in which the XDG activity was lost and the UDG and HDG activity was reduced (Fig. 2.5A). The main difference is that N120A retained a higher C/U and G/U activity but completely lost HDG activity on A/I and T/I (Fig. 2.5A and Table 2.1). The reduction of G/I activity, as judged by the apparent rate constants, was forty-seven-fold (Table 2.1). Extension of the amide side chain by one methylene group, as shown in the N120Q mutant, exhibited the same effect as removal of the amide side chain as shown in N120A (Fig. 2.5B). Conversion of the amide group to a carboxyl group as shown in N120D mutant had a minimal effect on UDG and XDG activity except for a small reduction on the A/U base pair (Fig. 2.5C). The main effect of the Asp substitution was the reduction of HDG activity

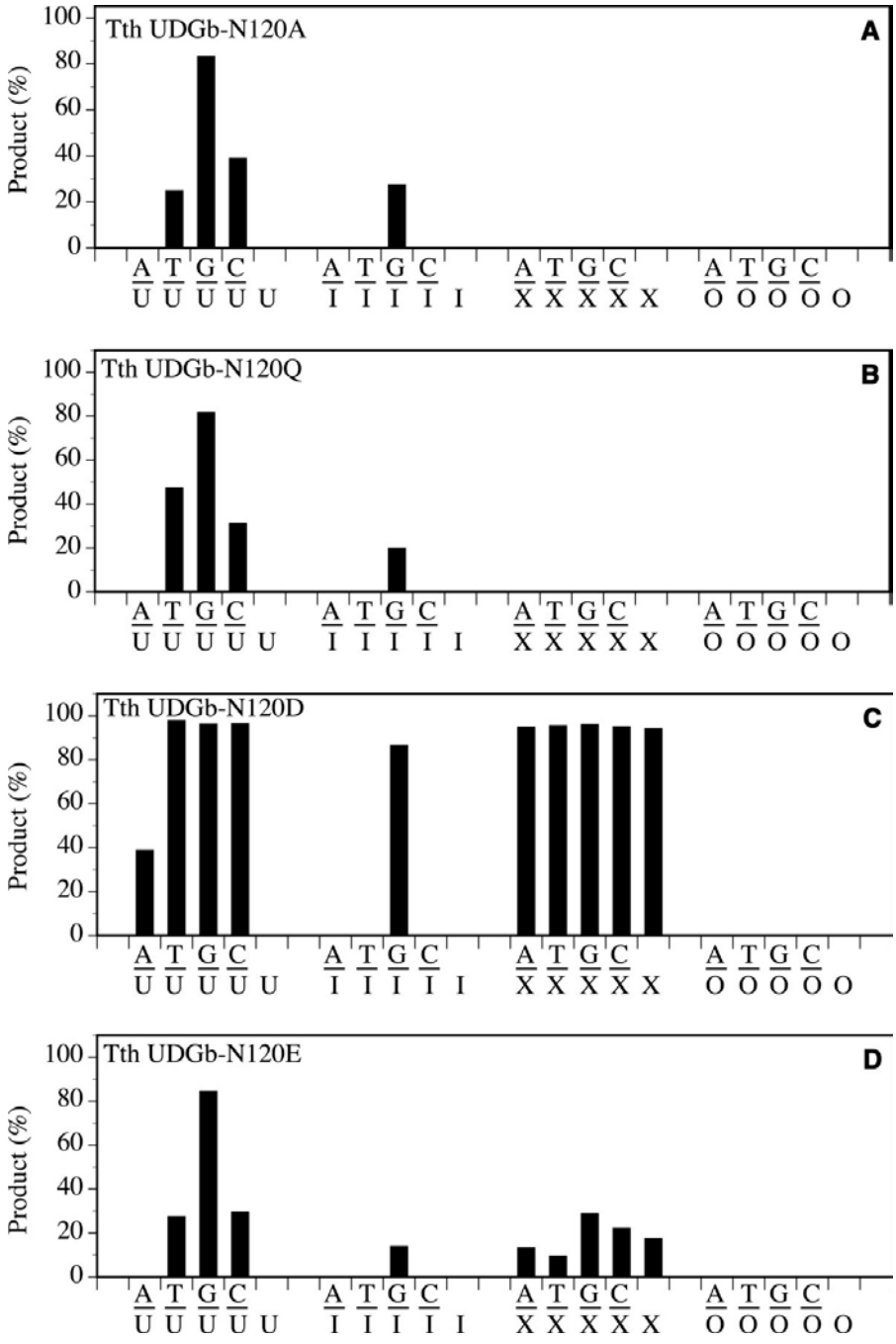


Figure 2.5 Glycosylase activity of Tth UDGb N120 mutants on U-, I-, X-, and O-containing DNA substrates. Cleavage reactions were performed as described under "Experimental Procedures" at 50 °C for 60 min with 100 nm protein and 10 nm substrate. *A*, Tth UDGb-N120A. *B*, Tth UDGb-N120Q. *C*, Tth UDGb-N120D. *D*, Tth UDGb-N120E.

on A/I and T/I base pairs to below the level of detection under the assay conditions (Fig. 2.5B). N120E mutant amplified the effect of carboxyl substitution as UDG, HDG and XDG activities were all reduced substantially (Fig. 2.5D). The effects observed in N120E and N120Q underscore the sensitivity of UDGb catalytic activity toward small length changes in the N120 side chain.

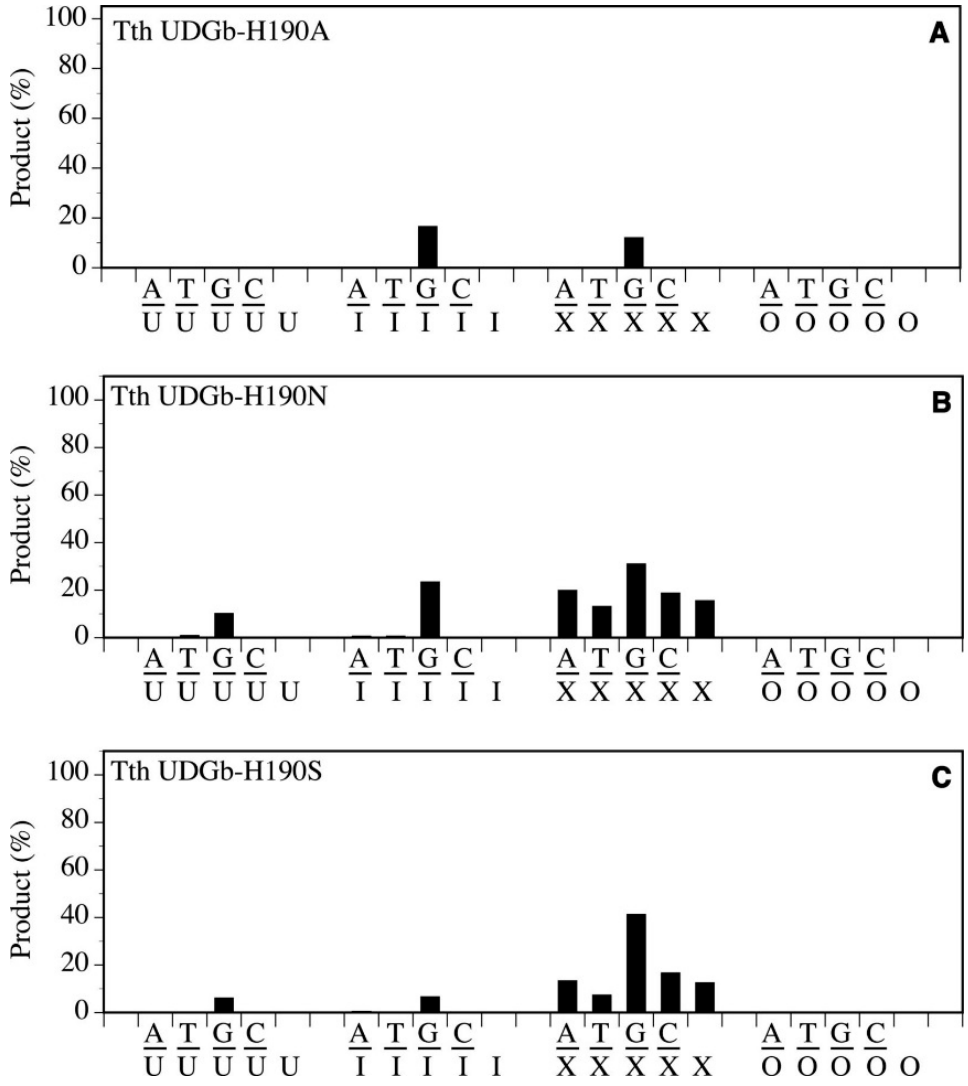


Figure 2.6 Glycosylase activity of Tth UDGb H190 mutants on U-, I-, X-, and O-containing DNA substrates. Cleavage reactions were performed as described under "Experimental Procedures" at 50 °C for 60 min with 100 nm protein and 10 nm substrate. *A*, Tth UDGb-H190A. *B*, Tth UDGb-H190N. *C*, Tth UDGb-H190S.

The mutational effect was most profound with the substitution of H190 with Ala. H190A mutant lost all UDG, HDG and XDG activity with the exception of retaining a low level of activity on G/I base pairs and a rather minor activity observed on G/X base pairs (Fig. 2.6A). The functional role of H190 could be accommodated by Asn or Ser to a very

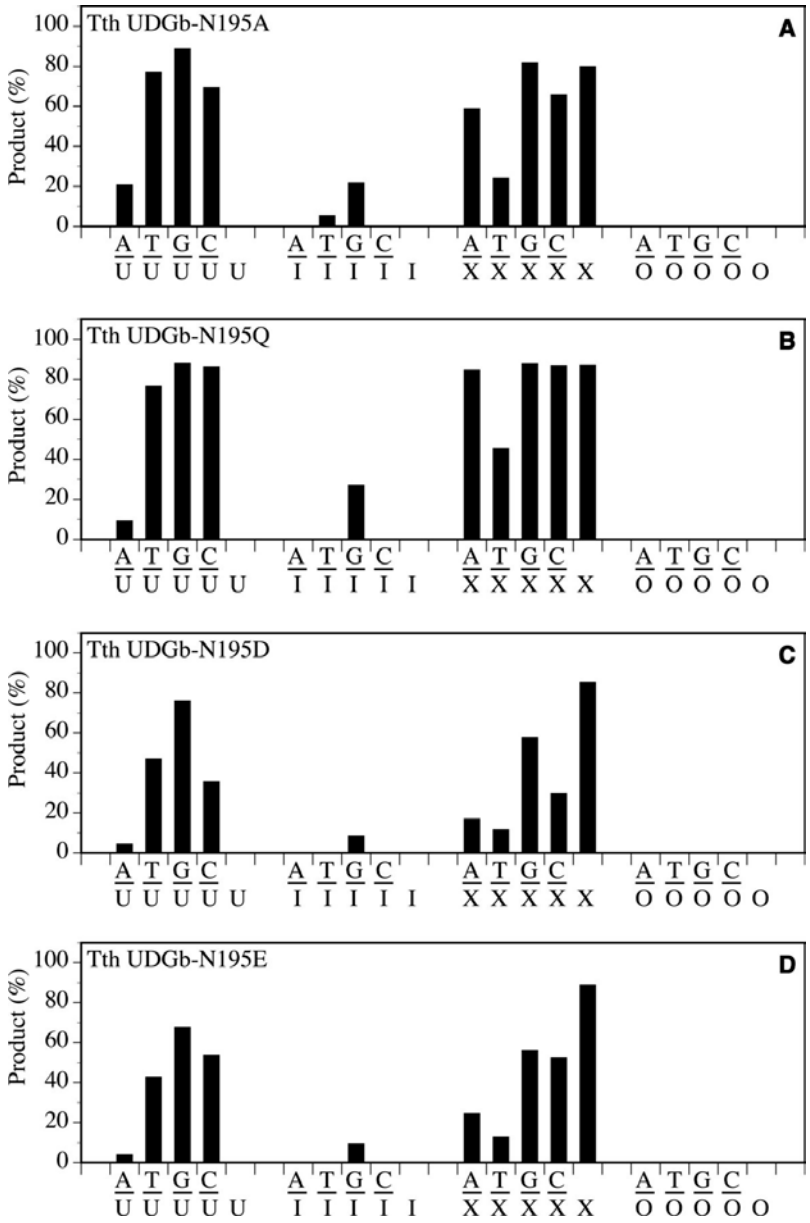


Figure 2.7 Glycosylase activity of Tth UDGb N195 mutants on U-, I-, O-, and X-containing DNA substrates. Cleavage reactions were performed as described under "Experimental Procedures" at 50 °C for 60 min with 100 nm protein and 10 nm substrate. *A*, N195A. *B*, N195Q. *C*, N195D. *D*, N195E.

limited degree as indicated by the retention of weak UDG, HDG and XDG activity (Fig. 2.6B-C).

The effects of amino acid substitutions at the N195 position were less severe than those at the D75, N120 and H190 positions. In general, reduced UDG, HDG and XDG activity was still observable in N195A, N195Q, N195D and N195E mutants (Fig. 2.7). The most noticeable reduction of activity was observed using A/I and T/I substrates in which the HDG activity was nearly or completely lost. Given that motif 2 serves as a wedge to insert into the space vacated by a flipped out base, mutations at the conserved N195 position may affect the wedging mediated by motif 2. The UDG activity on A/U base pairs was also reduced to varying degrees in N195A, N195Q, N195D, and N195E mutants. These mutants were used in the genetic analysis of A/U repair *in vivo* as described below.

Survival of UDGb in dut Temperature Sensitive E. coli Strain - Uracil may occur in DNA either through cytosine deamination or through incorporation of dUTP during DNA replication. In normal cells, incorporation of dUTP into DNA to form A/U base pairs is minimized due to the hydrolysis catalyzed by dUTPase (encoded by *dut*) (Fig. 2.8A). The dUMP thus produced is a precursor for the synthesis of dTMP by thymidylate synthase (encoded by *tms*), which is then phosphorylated to form dTTP as one of the regular nucleoside triphosphates for DNA synthesis. The *E. coli* strain BW276 is deficient in *ung xth* and contains a temperature sensitive (ts) *dut-1* gene (23). *E. coli* cells deficient in *xth* (encodes an AP endonuclease) is lethal but the lethality is rescued by deletion of *ung* (23). At permissible temperature (22°C), the active dUTPase prevents dUTP from incorporating into DNA during replication (Fig. 2.8A, left). At high temperature (42°C), uracil is incorporated into DNA to form A/U base pairs due to the loss of much of the dUTPase

activity. In the absence of the endogenous UNG, which removes uracil from A/U base pair, cells remain viable since a limited amount of uracil incorporated into the genome is tolerated (32). However, cells lose viability in the presence of active UNG, which constantly removes uracil from A/U base pairs and creates toxic AP sites (Fig. 2.8A, right).

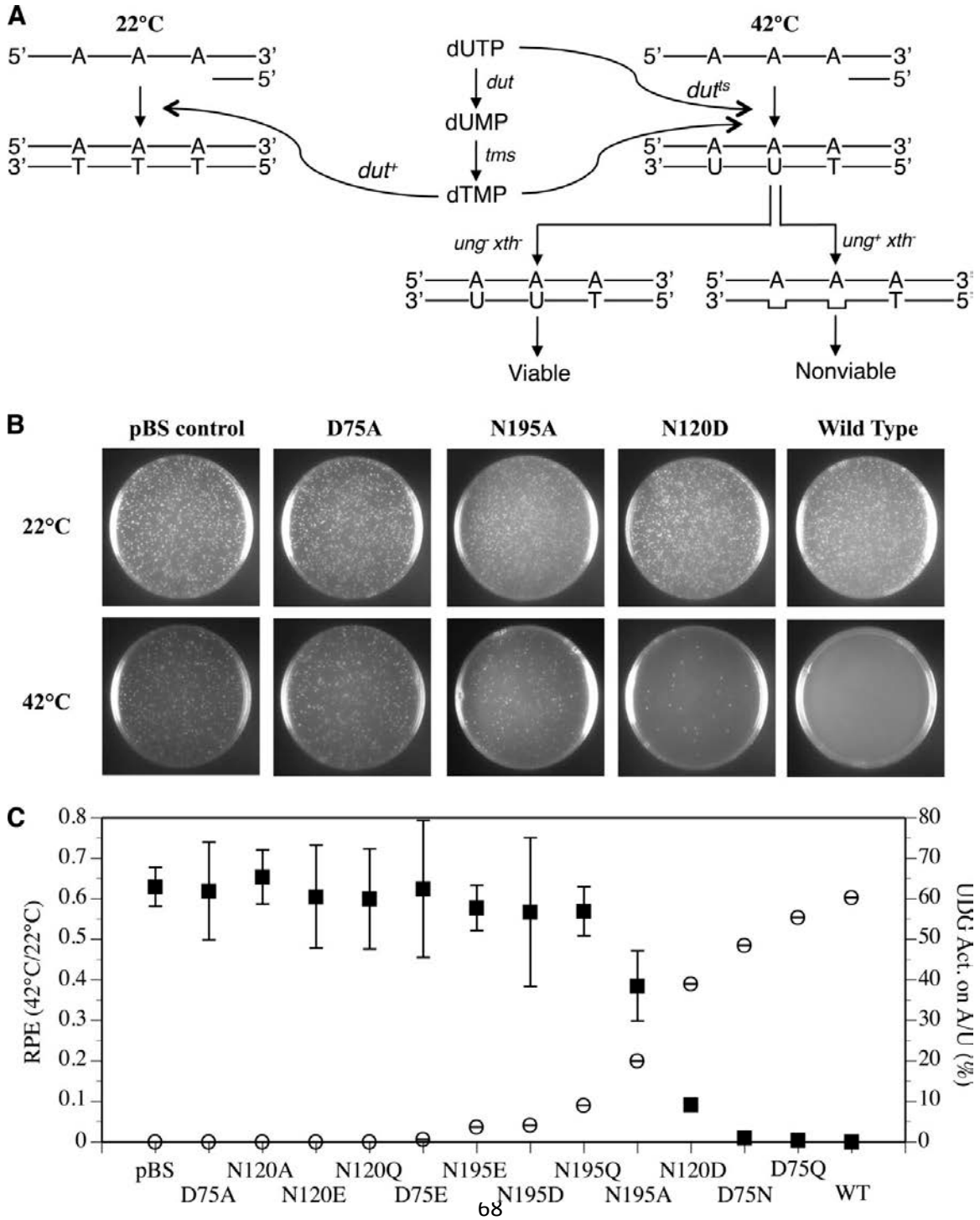


Figure 2.8 The survival of E. coli BW276 containing Tth udgb genes at 42 and 22 °C and their correlation with enzymatic activity of Tth UDGb variants on the A/U base pair. *A*, principle of the *in vivo* assay. *dut*, gene for dUTPase; *tms*, gene for thymidylate synthase; *xth*, AP endonuclease; *ts*, temperature-sensitive. *B*, the survival of *E. coli* BW276 containing pBS-Tth-UDGb variants vectors at 42 and 22 °C. Cells with A_{600} value of 1 were diluted 1×10^4 times, and 150 μl of diluted cells were spread on the LB plates containing ampicillin and thymidine and incubated for 24 h at 42 °C or 72 h at 22 °C. *pBS*, pBluescript. *C*, correlation between A/U enzymatic activities of Tth UDGb variants and the survival of *E. coli* BW276 containing pBS-Tth-UDGb variants vectors at 42 and 22 °C. The cleavage reactions of Tth UDGb variants were performed as described under "Experimental Procedures" at 50 °C for 60 min with 100 nm protein and 10 nm substrate. The data of cleavage activities were plotting according to the right *y* axis. For the convenience of counting colonies grown at 22 °C, *E. coli* cells were diluted 1×10^6 times prior to plating. For counting colonies grown at 42 °C, *E. coli* cells were diluted as follows prior to plating: pBluescript, D75A, N120A, N120E, N120Q, D75E, N195E, N195Q, N195D, and N195A, 1×10^6 times; N120D, 1×10^5 times; D75N, 2×10^3 times; D75Q, 1×10^3 times; wild type, 1×10^2 times. Cell numbers were counted on LB plates containing ampicillin and thymidine after incubation at 42 °C for 24 h or incubation at 22 °C for 72 h. The relative plating efficiencies were calculated by the ratios of the cell numbers between 42 and 22 °C. The data of relative plating efficiencies (*RPE*) are plotted as the left *y* axis. The data are the averages of at least three independent experiments. ■, relative plating efficiencies; ○, UDG activity on the A/U base pair.

We were interested in understanding how the ability of a UDG to remove uracil from an A/U base pair *in vitro* correlates to its ability to do so *in vivo*. In examining the UDG activity of Tth UDGb on the A/U base pair, we found a descending trend of the activity in the order of WT > D75Q > D75N > N120D > N195A > N195Q > N195D, N195E > D75E > N120Q, N120E, N120A, D75A in the Tth UDGb mutants (Figs. 2.4-2.7). We thought that this spectrum of UDG activity on the A/U base pair could provide an excellent opportunity to test how the *in vitro* activity correlates to *in vivo* uracil removal. As illustrated in Fig. 2.8A, the UDG activity on A/U base pairs in the wild type Tth UDGb was expected to render the *E. coli* BW 276 strain much less viable. On the other hand, the lack of activity on A/U base pairs in the D75A mutant should allow the cells become much more viable even at 42°C. Mutants with intermediate levels of *in vitro* activity toward A/U base pairs should also exhibit intermediate cell viability. Indeed, we found an inverse correlation between the *in vitro* activity on A/U base pairs and the viability as measured by

the relative plating efficiencies (RPE) at 42°C/22°C (Fig. 2.8B-C). These results indicate that a higher level of the activity on the A/U base pairs will lead to a better removal of uracil from A/U base pairs in the genome and thus limits the survival of the *dut*^{ts} cells.

V. Discussion

Comparison of Substrate Specificity in UDG Superfamily - UDG superfamily encompasses six diverse families. This study reveals that family 5 UDGb is a deamination repair enzyme with rather broad specificity. Among the UDG enzymes we have investigated, only *Schizosaccaromyces pombe* TDG shows a broader specificity as it acts on all deaminated bases (6). Family 5 UDGb can act as a uracil, hypoxanthine and xanthine DNA glycosylase (Fig. 2.1). In comparison, family 1 UNG only works on uracil, while family 2 *E. coli* MUG and family 3 SMUG1 act on both uracil and xanthine. Therefore, family 5 UDGb is a versatile repair enzyme that can deal with multiple types of deaminated base lesions. Interestingly, family 5 UDGb is implicated in playing a role in the repair of hypoxanthine *in vivo* in addition to its role in the repair of uracil (14). Recently, we reported on a new family of enzymes in the UDG superfamily that act as hypoxanthine DNA glycosylases rather than uracil DNA glycosylases (13). Taken together, these data indicate that enzymes in the UDG superfamily can play a variety of roles in the repair of deaminated base lesions. Structurally, the active sites of the enzymes in the superfamily appear to be adaptable to accommodate both pyrimidine and purine deaminated products. Evolutionarily, different families have evolved a rather impressive spectrum of specificity, from family 1 UNG acting exclusively on one type of lesion (uracil) to family 6 HDG acting primarily on one type of lesion (hypoxanthine); from family 2 *E. coli* MUG and family 3 SMUG1 acting on two types of lesions (uracil and xanthine) to family 5 UDGb

acting on three types of lesions (uracil, hypoxanthine and xanthine) and family 2 *S. pombe* TDG acting on four types of lesions (uracil, hypoxanthine, xanthine and oxanine). Additionally, the family 2 human TDG has evolved to act as a demethylase to remove fC and caC to initiate an enzymatic demethylation process in epigenetics.

HDG Activity and Base Flipping - The hypoxanthine DNA glycosylase activity from Tth UDGb showed a trend of G/I > A/I and T/I (Fig. 2.1 and Table 2.1). This trend, in general, is consistent with the observations we previously made in *Methanosarcina bakeri* HDG and human endonuclease V (13,33). The PMF analysis indicates that the G/I base pair is the least stable among the four base pairs (Fig. 2.2). The T/I and A/I base pairs have a similar level of higher stability than the G/I base pair. The C/I base pair has the highest level of stability, in keeping with its structural similarity to a C/G base pair (34). The outcome of the computational analysis on the stability of the hypoxanthine-containing base pairs is consistent with experimental analyses previously reported (35-37). The stability of the base pairs as estimated by the PMF calculation suggests that the G/I base pair is more prone to base flipping. Indeed, the wild type Tth UDGb showed ten-fold higher activity with the G/I base pair than the A/I or the T/I base pair (Table 2.1). In addition, a large number of mutants still maintained some level of activity on the G/I base pair while the activity on the A/I or T/I base pair was either negligible or not detected (Figs. 2.4-2.7). This is in accord with the notion that base flipping renders the glycosidic bond more reactive (38). The consistency between the hypoxanthine DNA glycosylase activity on different base pairs and the tendency of base flipping suggests that some of the repair enzymes rely on the instability of the base pairs to capture the deaminated base, hypoxanthine. Therefore, these data underscore the importance of base pair stability in

promoting the catalysis of deaminated bases. This is an example in which the observed catalytic activity of an enzyme is perturbed not by chemically altering the substrate functional groups directly involved in enzyme recognition but by altering the conformational equilibrium of a substrate.

Inverse Correlation of Viability and the UDG Activity on A/U Base Pairs - The UDG enzymes may play two roles *in vivo*. First, when cytosine is deaminated, UDG can remove the deaminated cytosine--uracil from the G/U base pair. Second, uracil may be incorporated into a DNA through dUTP to form A/U base pairs. The A/U base pair can be removed by a UDG with activity on A/U base pairs. The family 5 UDG enzymes possess repair activity on both G/U and A/U base pairs. The physiological role of the UDG in removing uracil from G/U base pairs has been studied (14,18). However, whether the UDG activity of UDGb on A/U base pairs can remove uracil *in vivo* remained unclear. Based on the genetic properties of *dut*^{ts} *ung xth* genes in *E. coli*, we adopted strain BW 276 for our investigation for the purpose of understanding the physiological role of the UDG activity on A/U base pairs.

Since the transition from the RNA world to the DNA world, nature has chosen thymine over uracil as the fourth base in DNA. Because DNA polymerase can incorporate dUTP to DNA, the level of dUTP in a cell has to be kept minimal to prevent appearance of uracil in DNA genomes. The highly active dUTPase encoded by the *dut* gene efficiently hydrolyzes dUTP to dUMP to serve two important functions. First, it minimizes the dUTP in the nucleotide pool to prevent its incorporation into DNA. Second, it produces the dUMP, which is converted to dTMP by thymidylate synthase for DNA synthesis. Once a dUTP is accidentally incorporated into DNA to form an A/U base pair, the UDG with activity on

A/U base pairs can remove it through the base excision repair pathway. Thus, dUTPase and the BER pathway serve as preventative and corrective mechanisms to ensure a uracil-free DNA genome in the DNA world at both the nucleotide and DNA levels. The inverse correlation between the UDG activity on A/U base pairs and the viability of the BW 276 cells indicates that Tth UDGb can remove uracil from A/U base pairs (Fig. 2.8). The increase of UDG activity on A/U base pairs accompanied by the corresponding decrease in viability suggests that the genetic assay correlates well with the *in vitro* biochemical assay. Therefore, the genetic system described here can be a useful tool to study other uracil DNA glycosylases with activity on A/U base pairs.

Unique Catalytic Mechanism - The catalytic mechanisms are most extensively studied in family 1 UNGs and to a lesser degree in family 2 enzymes. A common theme in the hydrolysis of N-glycosidic bonds is that an oxacarbenium ion intermediate is formed in the transition state (38,39). The catalytic power to accelerate the hydrolysis reaction comes from a combination of activation of the leaving group, stabilization of the oxacarbenium ion, and activation of water as a nucleophile. For family 1 UNG enzymes, structural, biochemical, mutational, and kinetic investigations have identified a His residue (H187 in *E. coli* UNG) in motif 2 as a critical residue that forms a short hydrogen bond to O2 of uracil to promote the departure of the uracil anion leaving group (40-44). In addition, a negatively charged Asp residue (D64 in *E. coli* UNG) is proposed to act as a general base to activate a water molecule (40,43,45). For family 2 MUG/TDG enzymes, it is proposed that an Asn residue (N18 in *E. coli* MUG), which is located in an equivalent position as D64 in *E. coli* UNG, may activate a water molecule to initiate the nucleophilic attack at the glycosidic bond (46,47). Family 5 UDGb enzymes lack a catalytic Asp/Asn residue in

motif 1 as seen in family 1 and family 2 enzymes. However, a water molecule was observed near the conserved N120 in the Tth UDGb structures complexed with AP/G- or AP/A-containing DNA (19). The conserved H190 in motif 2 is also implicated in the removal of uracil by the Pae UDGb (12). In light of the broad deaminated base excision specificity demonstrated in this study (Fig. 2.1B), we set out to identify residues that are important for the excision of both pyrimidine base damage and purine base damage.

In the Tth UDGb crystal structure, D75 is involved in coordination of a water molecule through its side chain (19). The D75E mutant, which retained the negative charge on the side chain, still maintained UDG, HDG and XDG activity (Fig. 2.4B). Other substitutions result in substantial reduction of XDG, HDG and UDG activity (Fig. 2.4). To understand the role of D75 in the excision of both pyrimidine and purine base damage, we modeled U, I and X into the crystal structure and carried out molecular dynamics simulation analysis. In the modeled Tth UDGb-U structure, the side chain of D75 is in close proximity to the O4 of uracil. A distance averaging 4.9 Å is consistent with a bridging water molecule mediating an interaction between D75 and the O4 of uracil (Fig. 2.9A). A closer look at the modeled Tth UDGb structures suggest that D75 is in close proximity to the 5' phosphate of the uridine, which can raise the pKa and facilitate protonation of the D75 sidechain. This will allow the bridging water molecule to form a hydrogen bond with the O4, which facilitates the removal of uracil (Fig. 2.10A and 2.10D). Mutations at the D75 position exhibited more profound effects on HDG and XDG activity (Fig. 2.4). In the modeled UDGb-I and UDGb-X structures, the protonated sidechain carboxylate of D75 can interact with the N7 in the hypoxanthine and xanthine through a bridging water molecule in a similar fashion (Fig. 2.10B-E, 2.10C-F). MD analysis indicates that the

carboxyl side chain is within a distance capable of forming a water-mediated hydrogen bond with the N7 in the purine bases (Fig. 2.9B-C). On the other hand, the lack of interaction to the N7 moiety in oxanine may in part be responsible for the lack of ODG activity although the interaction with the N3 in oxanine appears feasible (Fig. 2.9D-E). Activation of a purine base through hydrogen bonding or protonation has been proposed as a catalytic mechanism for the cleavage of glycosidic bonds in purine nucleotides. In acid-catalyzed hydrolysis of purine nucleosides, N7 and N3 protonations promote the departure of the purine base (48). In enzyme-catalyzed reactions, MutY, a DNA repair glycosylase involved in removal of adenine from G/A base pairs, catalyzes the excision of adenine by protonating the N7 position (49,50). Biochemical and structural studies identify a Glu residue (E37 in *E. coli* MutY and E43 in *Bacillus stearothermophilus* MutY) in the active site that can serve as a general acid in promoting N7 protonation (51-55). The close proximity of a water molecule between the E43 in B. *stearothermophilus* MutY and N7 indicates that the proton for N7 protonation can come from water coordinated by E43 (56). Within the UDG superfamily, we identified an S23-N7 interaction in the *E. coli* family 2 MUG enzyme and an M64 main chain-N7 interaction in the family 3 *Geobacter metallireducens* SMUG1 enzyme that play an important role in the excision of xanthine bases (10,21). Apparently, N7 interaction or protonation is a common catalytic mechanism for leaving group activation in the hydrolysis of purine deaminated bases.

According to the AP site cocrystal structure, N120 is involved in coordinating a water molecule that is located on the opposite side of the deoxyribose (19). Elimination of the amide group or lengthening of the amide group by one methylene carbon leads to the loss of XDG activity and a substantial reduction of HDG and UDG activity (Fig. 2.5).

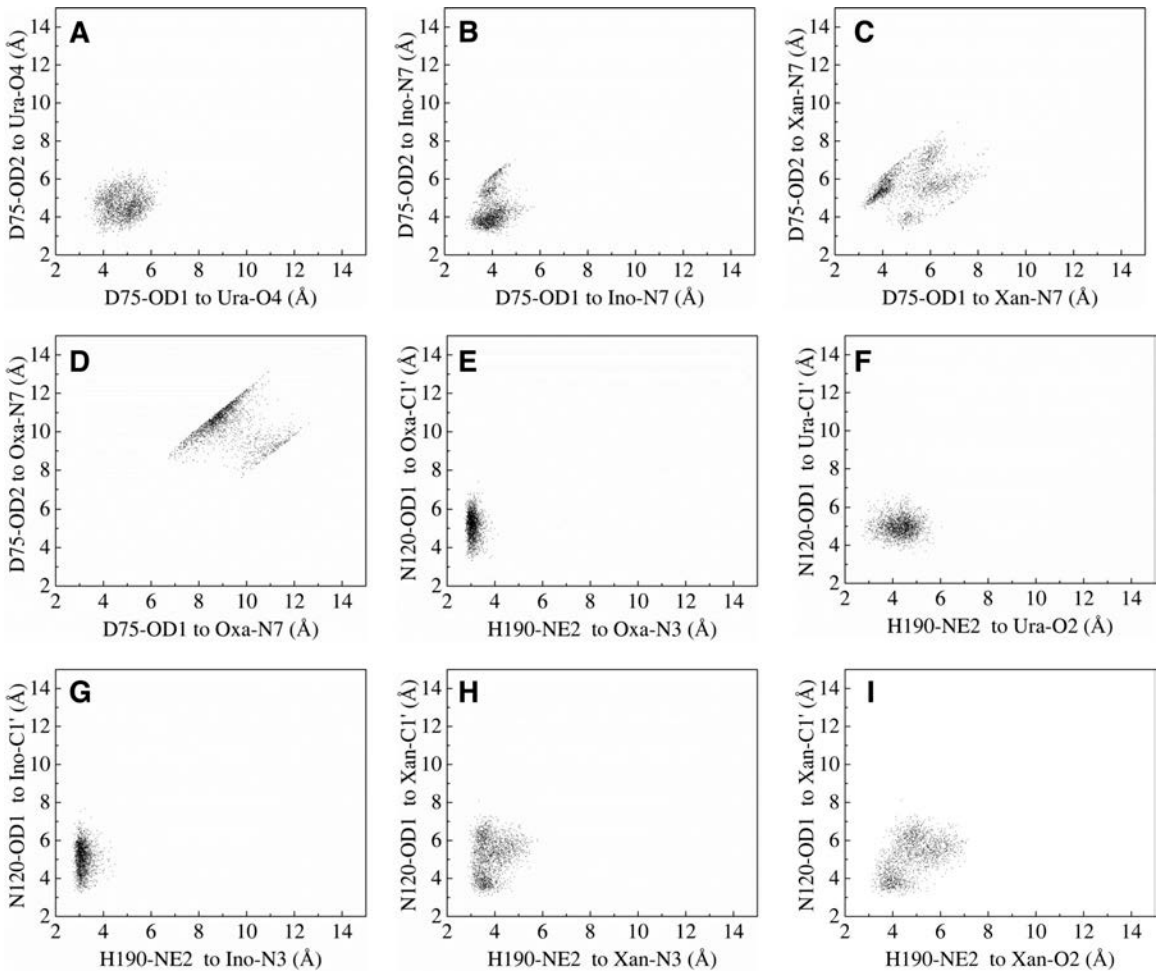


Figure 2.9 Two-dimensional scatter plots of heavy atom distances in the active site of enzyme-DNA complexes obtained from MD trajectories. *A*, distances for interactions between Asp^{75} and uridine. *B*, distances for interactions between Asp^{75} and inosine. *C*, distances for interactions between Asp^{75} and xanthosine. *D*, distances for interactions between Asp^{75} and oxanosine. *E*, distances for interactions between Asn^{120}, His^{190}, and oxanosine. *F*, distances for interactions between Asn^{120}, His^{190}, and uridine. *G*, distances for interactions between Asn^{120}, His^{190}, and inosine. *H* and *I*, distances for interactions between Asn^{120}, His^{190}, and xanthosine.

Based on the structural information and biochemical analysis, we speculate that N120 may perform a functional role similar to N18 in *E. coli* family 2 MUG, in which the Asn helps activate/position a water molecule for initiating a nucleophilic attack on the glycosidic bond. According to MD analysis, an average distance between OD1 of N120 and C1' carbon of deoxyribose is approximately 5.0 Å (Fig. 2.9F-H). A structural comparison

between N18 in MUG and N120 in Tth UDGb is shown in Fig. 2.11. While differences in the sequence lengths between the Tth UDGb and the *E. coli* MUG enzyme prevent a perfect alignment, strong structural similarity is noted in the core secondary structural elements including five β-sheets and four α-helices (Fig. 2.11A). The strong structural similarity is also noted in the active sites (Fig. 2.11B). The N18 in MUG and N120 in Tth UDGb hydrogen bond with water through their sidechains, positioning the water molecule proximal to the anomeric carbon of the bound nucleotide. Thus, different families in the UDG superfamily have adopted the same amino acid residue in different structural locations to perform a similar function.

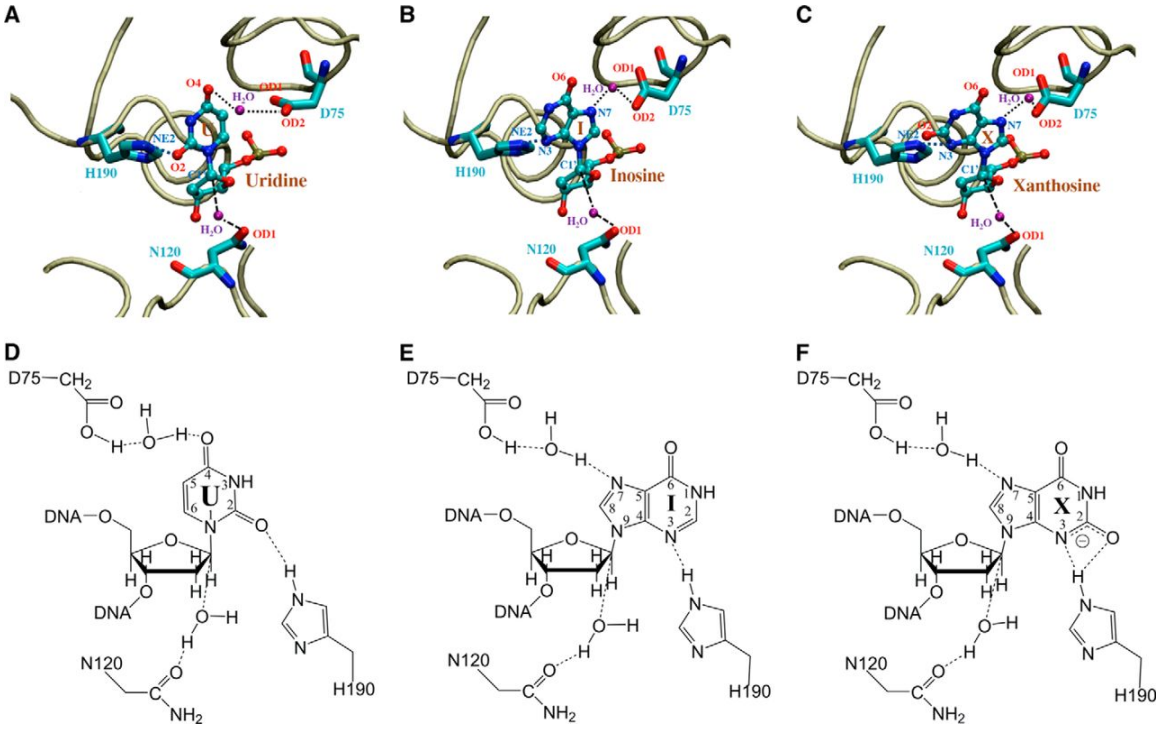


Figure 2.10 Close-up views of the Tth UDGb-DNA active site interactions in the energy minimized structures. Tube trace of the protein is colored in *tan*. Uridine (U), inosine (I), and xanthosine (X), as well as amino acids Asp^{75}, Asn^{120}, and His^{190}, are colored by atom type. Water molecules are labeled in *purple. Dashed lines* indicate inferred hydrogen bonds or water association. *A*, modeled Tth UDGb-U interactions. *B*, modeled Tth UDGb-I interactions. *C*, modeled Tth UDGb-X interactions. *D*, chemical illustration of Tth UDGb-U interactions. *E*, chemical illustration of Tth UDGb-I interactions. *F*, chemical illustration of Tth UDGb-X interactions.

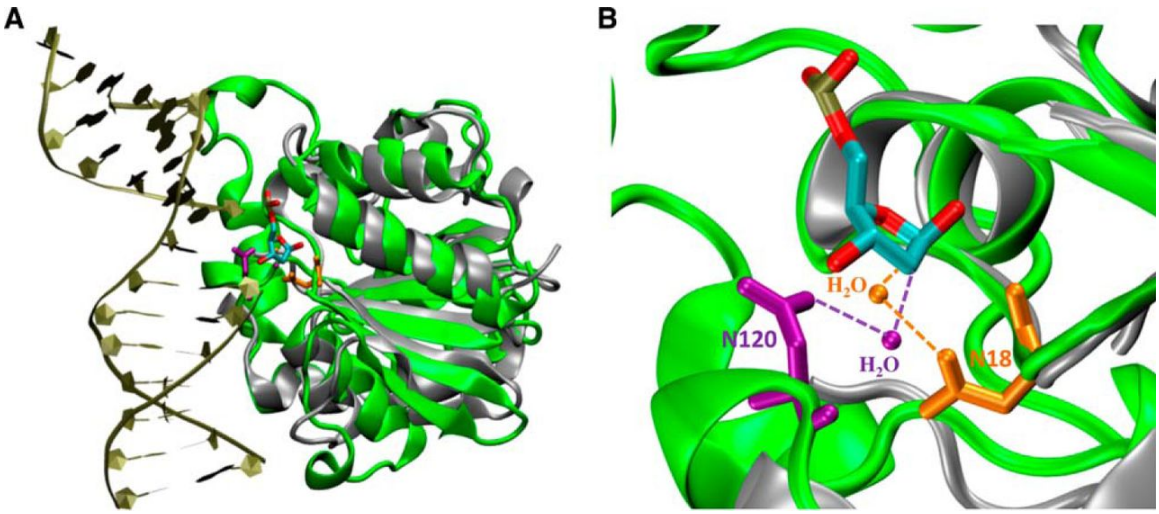


Figure 2.11 Comparison of interactions between Tth UDGb-N120 with water and E. coli MUG-N18 with water.*A*, superimposition of UDGb-AP structure (Protein Data Bank code 2DEM; *green*) with *E. coli* MUG structure (Protein Data Bank code 1MUG; *silver*). The AP site is colored by atom type. The two structures were superimposed using the program VMD. *B*, close-up view of Tth UDGb-N120-water and *E. coli* MUG-N18-water interactions using the same coloring scheme described in *A*. Asn^{120} and the interacting water in Tth UDGb structure are shown in *purple*. Asn^{18} and the interacting water in the MUG structure are shown in *orange*.

Among the six families within the UDG superfamily, families 1, 3, 4 and 5 contain a His residue at the beginning of motif 2 (Fig. 2.3B). Mutational studies in family 1 enzymes confirm that the His residue plays an important role in catalysis and suggests that H187 in *E. coli* UNG can form a hydrogen bond with O2 of uracil (41-43). Spectroscopic analyses indicate that H187 in *E. coli* UNG is neutral and forms a short hydrogen bond with the O2 group (44,57). Mutational and biochemical investigations in family 3 SMUG1 enzymes also underscore the importance of the His residue in motif 2 in catalysis (10,58). Data from this work indicate that H190 in family 5 Tth UDGb is critical for all deaminated base glycosylase activity (Fig. 2.6). The elimination of the imidazole side chain renders the enzyme essentially inactive except for a minor HDG activity on G/I base pairs and a barely detectable XDG activity on G/X base pairs (Fig. 2.6). H190N and H190S can rescue the

activity to a very limited extent (Fig. 2.6). Evidently, UDG activity relies on H190 in motif 2 since H190A is the only mutant that results in a loss of UDG activity. D75A and N120A mutants reduce but do not completely eliminate the UDG activity. Similar to family 1 UNG enzymes, MD analysis suggests that H190 can form a hydrogen bond with O2 of uracil (Fig. 2.9F). The predominant role H190 plays in UDG activity suggests that the catalysis of glycosidic bond hydrolysis during uracil excision is likely to follow a stepwise D_N*A_N mechanism in which the uracil base departs as a uracil anion. Similar to the catalytic mechanism in family 1 UNG enzymes, the hydrogen bonding provided by H190 can promote the departure of the leaving group by stabilizing the uracil anion. This appears to be the main catalytic power endowed in family 5 Tth UDGb for uracil excision.

H190 is also important for hypoxanthine and xanthine excision (Fig. 2.6). Similar to nonenzymatic purine nucleoside hydrolysis (48), the hydrogen bonding provided by H190 to N3 of hypoxanthine could promote the departure of hypoxanthne. Biochemical studies using adenine analogs also indicate that *E. coli* MutY utilizes the N3 interaction to enhance adenine excision (59). In the modeled structure, the hydrogen in NE2 of H190 of Tth UDGb is within 2.1 Å of the N3 of hypoxanthine (Fig. 2.10B and 2.10E), consistent with a moderately strong hydrogen bond. MD analysis also supports the hydrogen bonding model between H190 and N3 of hypoxanthine (Fig. 2.9G). For the xanthine-containing base pairs, although normally the N3 position is shown as protonated, xanthosine exists as a monoanion under physiological conditions (60,61). Thus, as supported by the modeled structure and MD analysis (Fig. 2.9H, Fig. 2.10C and 2.10F), H190 can promote the departure of xanthine through hydrogen bonding with the N3 moiety of xanthosine. H190 is 3.5 Å and 4.2 Å from N3 and O2 of xanthosine based on the energy minimized structure,

raising the possibility of forming a bidentate hydrogen bond (Fig. 2.10C and 2.10F). MD analysis also supports the possibility of a bidentate hydrogen bond (Fig. 2.9H-I). The caveat here is that for the excision of deaminated purine bases, the potential water activation by N120 and the water mediated N7 contact by D75 are also important. Therefore, family 5UDGb enzymes may rely on the concerted action of multiple catalytic residues to excise hypoxanthine and xanthine and are likely to catalyze the N-glycosidic bond hydrolysis through a dissociative A_ND_N mechanism.

In summary, this study for the first time comprehensively investigated deaminated base repair by the family 5 Tth UDGb enzyme. The data presented here reveals that family UDGb enzymes are uracil, hypoxanthine and xanthine DNA glycosylases. The inverse correlation between cellular viability and the UDG activity on A/U base pairs offers a tool to study the role of different UDG enzymes in keeping uracil out of DNA in the DNA world. In addition to taking advantage of greater tendency for base flipping that occurs in damaged DNA base pairs, family 5 UDGb enzymes have adapted multiple mechanisms to achieve deaminated base excision. Whereas the UDG activity heavily relies on the leaving group activation mediated by O2 of uracil and H190 to excise a uracil base, the family UDGb enzymes combine an existing catalytic element (H190 in Tth UDGb) with additional nucleophile activation and purine specific leaving group activation mechanisms to expand their DNA repair capacity. As such, family 5 UDGb offers an example of how an enzyme may strategically acquire catalytic elements to broaden its specificity.

VI. Reference

1. Parikh, S. S., Putnam, C. D., and Tainer, J. A. (2000) Lessons learned from structural results on uracil-DNA glycosylase. *Mutation research* **460**, 183-199
2. Krokan, H. E., Drablos, F., and Slupphaug, G. (2002) Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**, 8935-8948
3. Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007) The enigmatic thymine DNA glycosylase. *DNA repair* **6**, 489-504
4. Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J., and Schar, P. (2001) Thymine DNA glycosylase. *Progress in nucleic acid research and molecular biology* **68**, 235-253
5. Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutation research* **460**, 165-181
6. Dong, L., Mi, R., Glass, R. A., Barry, J. N., and Cao, W. (2008) Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase. *DNA repair* **7**, 1962-1972
7. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307
8. Maiti, A., and Drohat, A. C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* **286**, 35334-35338

9. Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W., and Verdine, G. L. (1999) Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Current biology : CB* **9**, 174-185

10. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N., and Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. *Journal of molecular biology* **385**, 761-778

11. Sandigursky, M., and Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes. *Current biology : CB* **9**, 531-534

12. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H., and Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J* **21**, 3182-3191

13. Lee, H. W., Dominy, B. N., and Cao, W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *The Journal of biological chemistry* **286**, 31282-31287

14. Wanner, R. M., Castor, D., Guthlein, C., Bottger, E. C., Springer, B., and Jiricny, J. (2009) The uracil DNA glycosylase UdgB of Mycobacterium smegmatis protects the organism from the mutagenic effects of cytosine and adenine deamination. *Journal of bacteriology* **191**, 6312-6319

15. Srinath, T., Bharti, S. K., and Varshney, U. (2007) Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis. *DNA repair* **6**, 1517-1528

16. Malshetty, V. S., Jain, R., Srinath, T., Kurthkoti, K., and Varshney, U. (2010) Synergistic effects of UdgB and Ung in mutation prevention and protection against commonly encountered DNA damaging agents in Mycobacterium smegmatis. *Microbiology* **156**, 940-949

17. Starkuviene, V., and Fritz, H. J. (2002) A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus. *Nucleic acids research* **30**, 2097-2102

18. Sakai, T., Tokishita, S., Mochizuki, K., Motomiya, A., Yamagata, H., and Ohta, T. (2008) Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus. *DNA repair* **7**, 663-669

19. Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S., and Masui, R. (2007) Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. *Journal of molecular biology* **373**, 839-850

20. Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research* **31**, 5108-5121

21. Lee, H. W., Brice, A. R., Wright, C. B., Dominy, B. N., and Cao, W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase. *The Journal of biological chemistry* **285**, 41483-41490

22. Priyakumar, U. D., and MacKerell, A. D., Jr. (2006) Computational approaches for investigating base flipping in oligonucleotides. *Chemical reviews* **106**, 489-505

23. Taylor, A. F., and Weiss, B. (1982) Role of exonuclease III in the base excision repair of uracil-containing DNA. *Journal of bacteriology* **151**, 351-357

24. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723

25. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545-1614

26. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586-3616

27. MacKerell, A. D., and Banavali, N. K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *Journal of Computational Chemistry* **21**, 105-120

28. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089-10092

29. Adelman, S. A., and Doll, J. D. (1976) Generalized Langevin equation approach for atom/solid - surface scattering: General formulation for classical scattering off harmonic solids. *Journal of Chemical Physics* **64**, 2375-2388

30. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802

31. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38

32. Warner, H. R., Duncan, B. K., Garrett, C., and Neuhard, J. (1981) Synthesis and metabolism of uracil-containing deoxyribonucleic acid in Escherichia coli. *Journal of bacteriology* **145**, 687-695

33. Mi, R., Alford-Zappala, M., Kow, Y. W., Cunningham, R. P., and Cao, W. (2012) Human endonuclease V as a repair enzyme for DNA deamination. *Mutat Res*

34. Kumar, V. D., Harrison, R. W., Andrews, L. C., and Weber, I. T. (1992) Crystal structure at 1.5-A resolution of d(CGCICICG), an octanucleotide containing inosine, and its comparison with d(CGCG) and d(CGCGCG) structures. *Biochemistry* **31**, 1541-1550

35. Case-Green, S. C., and Southern, E. M. (1994) Studies on the base pairing properties of deoxyinosine by solid phase hybridisation to oligonucleotides. *Nucleic acids research* **22**, 131-136

36. Martin, F. H., Castro, M. M., Aboul-ela, F., and Tinoco, I., Jr. (1985) Base pairing involving deoxyinosine: implications for probe design. *Nucleic acids research* **13**, 8927-8938

37. Watkins, N. E., Jr., and SantaLucia, J., Jr. (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic acids research* **33**, 6258-6267

38. Berti, P. J., and McCann, J. A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chemical reviews* **106**, 506-555

39. Stivers, J. T., and Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical reviews* **103**, 2729-2759

40. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-493

41. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80**, 869-878

42. Shroyer, M. J., Bennett, S. E., Putnam, C. D., Tainer, J. A., and Mosbaugh, D. W. (1999) Mutation of an active site residue in Escherichia coli uracil-DNA glycosylase: effect on DNA binding, uracil inhibition and catalysis. *Biochemistry* **38**, 4834-4845

43. Drohat, A. C., Jagadeesh, J., Ferguson, E., and Stivers, J. T. (1999) Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry* **38**, 11866-11875

44. Drohat, A. C., and Stivers, J. T. (2000) Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry* **39**, 11865-11875

45. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87-92

46. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92**, 117-129

47. Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *The EMBO journal* **18**, 6599-6609

48. Zoltewicz, J. A., Clark, D. F., Sharpless, T. W., and Grahe, G. (1970) Kinetics and mechanism of the acid-catalyzed hydrolysis of some purine nucleosides. *Journal of the American Chemical Society* **92**, 1741-1749

49. McCann, J. A., and Berti, P. J. (2008) Transition-state analysis of the DNA repair enzyme MutY. *Journal of the American Chemical Society* **130**, 5789-5797

50. Michelson, A. Z., Rozenberg, A., Tian, Y., Sun, X., Davis, J., Francis, A. W., O'Shea, V. L., Halasyam, M., Manlove, A. H., David, S. S., and Lee, J. K. (2012) Gas-phase studies of substrates for the DNA mismatch repair enzyme MutY. *Journal of the American Chemical Society* **134**, 19839-19850

51. Wright, P. M., Yu, J., Cillo, J., and Lu, A. L. (1999) The active site of the Escherichia coli MutY DNA adenine glycosylase. *J Biol Chem* **274**, 29011-29018

52. Brinkmeyer, M. K., Pope, M. A., and David, S. S. (2012) Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Chemistry & biology* **19**, 276-286

53. Fromme, J. C., Banerjee, A., Huang, S. J., and Verdine, G. L. (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652-656

54. Lee, S., and Verdine, G. L. (2009) Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18497-18502

55. Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, S., and Tainer, J. A. (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nature structural biology* **5**, 1058-1064

56. Brunk, E., Arey, J. S., and Rothlisberger, U. (2012) Role of environment for catalysis of the DNA repair enzyme MutY. *Journal of the American Chemical Society* **134**, 8608-8616

57. Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L., and Stivers, J. T. (1999) Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA

glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **38**, 11876-11886

58. Matsubara, M., Tanaka, T., Terato, H., Ohmae, E., Izumi, S., Katayanagi, K., and Ide, H. (2004) Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase. *Nucleic acids research* **32**, 5291-5302

59. Francis, A. W., Helquist, S. A., Kool, E. T., and David, S. S. (2003) Probing the requirements for recognition and catalysis in Fpg and MutY with nonpolar adenine isosteres. *Journal of the American Chemical Society* **125**, 16235-16242

60. Kulikowska, E., Kierdaszuk, B., and Shugar, D. (2004) Xanthine, xanthosine and its nucleotides: solution structures of neutral and ionic forms, and relevance to substrate properties in various enzyme systems and metabolic pathways. *Acta biochimica Polonica* **51**, 493-531

61. Poznanski, J., Kierdaszuk, B., and Shugar, D. (2003) Structural properties of the neutral and monoanionic forms of xanthosine, highly relevant to their substrate properties with various enzyme systems. *Nucleosides, nucleotides & nucleic acids* **22**, 249-263

CHAPTER THREE

CATALYTIC MECHANISM OF FAMILY 4 UDGA AND ENHANCING CATALYTIC EFFICIENCY BY CORRELATED MUTATION

I. Abstract

Uracil DNA glycosylase (UDG) is an essential enzyme for the removal of uracil from DNA introduced by cytosine deamination or misincorporation of dUMP by DNA polymerase during replication. Family 4 UDGa is one of six families in the UDG superfamily found in nature. Here, we report that family 4 UDGa from *Thermus thermophiles* is a robust uracil DNA glycosylase that only acts on double-stranded and single-stranded uracil-containing DNA. A series of contacts orchestrated by E41, E47, F54, N80 and H155 in Tth UDGa defines a tight uracil binding pocket. Based on mutational, kinetic and modeling analyses, a catalytic mechanism involving leaving group stabilization by H155 and water coordination by N89 is proposed. The correlation of E41 with G42 positions provides insight on the need of coevolution for enhancing catalytic efficiency during protein divergence.

II. Introduction

Cytosine (C) bases in DNA is prone to deamination to become uracil (U) bases (1). Because U pairs with adenine during DNA replication, G/C base pairs can be mutated to A/T base pairs due to deamination. C to T transition mutation is a prominent genetic change (2). Uracil in DNA is in general removed by uracil DNA glycosylase (UDG) through base excision repair (BER) pathway (3). The uracil DNA glycosylase (UDG) superfamily consists of six families with distinct enzymatic and repair properties. With the exception of family 6 hypoxanthine DNA glycosylases, families from 1 to 5 all contain uracil DNA

glycosylase activity. Family 1 UNG stands out as an extraordinarily robust UDG that removes uracil from both double-stranded and single-stranded uracil-containing DNA (4). The UDG activity in families 2, 3, 5 is orders of magnitude lower than family 1 UNG but can act on a variety of deaminated bases from hypoxanthine, a deamination product of adenine; to xanthine or oxanine, deamination products of guanine (5-8).

Family 4 UDGa was initially discovered in the hyperthermophilic bacterium *Thermotoga maritima* (9), then later in archaea (10-12). UDGa from thermophilic bacterium *Thermus thermophilus* (Tth) can remove uracil in vitro and reduce mutation rate in vivo (13,14). The crystal structure of Tth UDGa with uracil base is solved, which indicates that family 4 enzymes adopt a similar structural fold as seen in other families within the UDG superfamily (15). While previous studies have provided valuable information on family 4 UDGa enzymes, some fundamental questions remain to be answered. How broad is the specificity of UDGa towards other deaminated bases? How does the active site in UDGa catalyze the cleavage of the glycosidic bond in deaminated DNA? How efficient are the family 4 UDGa enzymes? To answer these questions, we conducted a comprehensive biochemical, mutational, kinetics and molecular dynamics analysis using UDGa from *T. thermophilus* as a model. Resembling family 1 UNG, Tth UDGa acts exclusively as an efficient UDG on double-stranded and single-stranded uracil-containing DNA, with no detectable activity on hypoxanthine-, xanthine-, or oxanine-containing DNA. Extensive mutational, enzyme kinetics studies coupled with molecular modeling and molecular dynamics analyses have led to a model that relies on a histidine residue in motif 2 to stabilize a departing negatively charged uracilate. Strikingly, a double substitution of E41-G42 by E41Q-G42D in motif 1 is able to rescue the detrimental effects

of single substitution by one to two orders of magnitude. Molecular modeling and molecular dynamics analysis reveal that the correlated mutation of E41Q-G42D brings the catalytic histidine in a closer position to stabilize the leaving group. This study underscores the significance of correlated mutation in achieving enzyme catalytic efficiency.

III. Materials and methods

Reagents, media, and strains

All routine chemical reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), or VWR (Suwanee, GA). Restriction enzymes, Phusion DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). Bovine serum albumin and dNTPs were purchased from Promega (Madison, WI). Gel DNA recovery Kit was purchased from Zymo Research (Irvine, CA). Oligodeoxyribonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA) and Eurofins Genomics (Huntsville, AL). The LB medium was prepared according to standard recipes. Hi-Di Formamide and GeneScan 500 LIZ dye Size Standard for ABI3130xl were purchased from Applied Biosystems. The *Tth* UDGa sonication buffer consisted of 20 mM Tris-HCl (pH7.5), 1 mM ethylenediaminetetraacetic acid (EDTA) (pH8.0), 2.5 mM DTT, 0.15 mM PMSF, and 50 mM NaCl. The GeneScan stop buffer consisted of 80% formamide (Amresco, Solon,OH), 50 mM EDTA (pH8.0), and 1% blue dextran (Sigma Chemicals). The TE buffer consisted of 10 mM Tris-HCl (pH8.0) and 1 mM EDTA.

Cloning, Expression and Purification of Tth UDGa

The uracil DNA glycosylase gene from *T. thermophilus* HB8 (TtUDGA) (GenBank accession number: AB109239.1) was amplified by PCR using the forward primer Tth

UDGaF (5’ TCG TATGTCCATATGACCCTGGAACTGCTTCAGGC -3’ (*Nde*I)) and the reverse primer Tth UDGaR (5’ ATCGTACTCGAGGAAGAGGGGCTCCTGGC TCACC -3’ (*Xho*I)). The PCR reaction mixture (20 μl) consisted 10 ng *T. thermophilus* HB8 genomic DNA, 500 nM forward and reverse primers, 1x phusion polymerase buffer, 200 μM each dNTP and 0.2 unit of phusion polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 98°C for 30 s; 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, and extension at 72°C for 20 s; and a final extension step at 72°C for 10 min. The PCR product was purified and cloned into pET21a. The recombinant plasmid was confirmed by DNA sequencing.

Site-directed mutagenesis was performed by using an overlapping extension PCR procedure similarly as previously described (6). Taking the mutant E41Q as an example: The first round of PCR was carried out using plasmid pET21a-Tth-UDGa as template DNA with two pairs of primers, Tth-UDGaF and E41QR (5’- CTCCTCCCCGGGG CCCTGCCCCACGATCATGAGCT-3’) pair; E41QF (5’-CTCATGATCGTGGG GCAGGGCCCCGGGGAGGAGGA-3’) and Tth-UDGaR pair. The PCR products were electrophoresed on 1% agarose gel and the expected PCR fragments were purified from gel slices by Gel DNA clean Kit. The second run of the PCR reaction mixture (20 μl), which contained 1 μl of each of the first run PCR fragments, 200 μM dNTPs, 1×Phusion DNA polymerase buffer, and 0.2 units of Phusion DNA polymerase (New England Biolabs), was initially carried out with a predenaturation step at 95 °C for 30 s; 5 cycles with each cycle of denaturation at 98 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72 °C for 30 s; and a final extension at 72 °C for 5 min. Afterward, 500 nM of outside

primers (Tth-UDGaF and Tth-UDGaR) was added to the above PCR reaction mixture. The subsequent overlapping PCR amplification included a predenaturation step at 98 °C for 15 s; 30 cycles with each cycle of denaturation at 98 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72°C for 30 s; and a final extension at 72°C for 10 min. Subsequent molecular cloning procedures were performed as previously described. The purified PCR products digested with a pair of *Bam*HI and *Xho*I endonucleases were ligated to the cloning vector pET21a treated with the same pair of restriction endonucleases. The recombinant plasmids containing the desired mutations were confirmed by DNA sequencing and transformed into *E. coli* strain BL21 (DE3).

The pET21a-Tth-UDGa was transformed into *E. coli* strain BL21 (DE3) by the standard protocol to express the C-terminal His-6-tagged *Tth* UDGa protein. Briefly, the protein was induced by 0.5 mM IPTG at 16 °C for 12h. After sonication and purification, fractions (300–400 mM imidazole, 60-80% chelating buffer B) containing the Tth UDGa protein as seen on 12.5% SDS-PAGE were pooled and concentrated by Amicon YM-10 (Millipore). The concentration of *Tth* UDGa protein was determined by SDS-PAGE analysis using bovine serum albumin as a standard and confirmed by measuring absorption at A_{280}. The protein was stored in aliquots at -80 °C. Prior to use, the protein was diluted with 2 x storage buffer (20 mM Tris-HCl pH8.0, 2 mM DTT, 2 mM EDTA, 400 μg/ml BSA, 100% Glycerol).

Oligodeoxynucleotide substrates

Oligodeoxynucleotides containing deoxyuridine (U), deoxyinosine (I), deoxyxanthosine (X) or deoxyoxanosine (O) were obtained or constructed as previously described (6).

DNA glycosylase activity assays

DNA glycosylase cleavage assays for *Tth* UDGa were performed under optimized reaction conditions at 60°C in a 10 μl reaction mixture containing 10 nM oligonucleotide substrate, 100 nM glycosylase, 20 mM Tris-HCl (pH 7.6), 100 mM KCl, 1 mM DTT, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95 °C for 5 min after adding 1 μl of 1 M NaOH. Samples for ABI 377 sequencer (Applied Biosystem) were prepared by mixing equal volume of GeneScan stop buffer and reaction mixture. After incubation at 95 °C for 5 min, 3.5 μl samples were loaded into 10% denaturing polyacrylamide gel. Electrophoresis was conducted at 1500 V for 1.5 h using the ABI 377 sequencer. Cleavage products and remaining substrates were quantified using the GeneScan analysis software. Samples for ABI 3130xl sequencer (Applied Biosystems) were prepared by mixing 2 μl of reaction mixture with 7.8 μl Hi-Di Formamide and 0.2 μl GeneScan 500 LIZ Size Standard. A total of 10 μl sample was loaded into ABI 3130xl and run with a fragment analysis module. Cleavage products and remaining substrates were analyzed by Gene Mapper.

Enzyme Kinetic Analysis.

Uracil DNA glycosylase assay was performed at 60°C with 20 nM G/U substrates with enzyme in excess ranging from 100 nM to 3200 nM. Samples were collected at time points from 2 s, 5 s, 10 s, 30 s, 1 min, 2.5 min, 5 min, 10 min, 15 min, 25 min, 30 min, 40 min and 60 min. The apparent rate constants for each concentration were determined by curve fitting using the integrated first-order rate equation [1]:

$$P = P_{max} (1\text{-}e^{-k_{obs} t}) \qquad [1]$$

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

The kinetic parameters k_2 and *K*m were obtained from plots of k_{obs} against total enzyme concentration ($[E_0]$) using a standard hyperbolic kinetic expression with program GraphPad 4.1 following the equation [2] (16)

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

For some mutants with a large *K*m in which Km >> $[E_0]$, the kinetic parameter k2/Km values were obtained from plots of k_{obs} against total enzyme concentration ($[E_0]$) using a linear regression with program GraphPad 4.1 following the equation [3] (17).

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

Molecular modeling and mutant making

The crystal structure of TthUDGa and product complex was acquired from the RCSB Protein Data Bank (accession code 1UI0), and used as a model for subsequent computational analysis. Structure of substrates flipped-out DNA was extracted from the crystal structure of human UDG-DNA complex (PDB accession code 1EMH) (18) using the Swiss-Pdb Viewer (SPDBV) program (19). Mutants E41Q, G42D and E41Q/G42D of TthUDGa were also made using the mutation tool in Swiss-Pdb Viewer program and the "best rotamer" was chosen with the lowest clash score.

Molecular dynamics simulations

After building the initial complex structures, an explicit solvent system using the TIP3P water model was constructed in the CHARMM c35b6 molecular mechanics package (20) using a suitably sized box. The minimum distance between any of the atoms of the solvated TthUDGa/DNA complex and the box boundary was maintained to at least 9 Å. Sodium chloride ions were added to the system to achieve an electrical neutral system. The CHARMM 27 all hydrogen force field for proteins (21) and nucleic acids (22) were used. Particle-mesh Ewald summation (23) was applied in the periodic boundaries condition for the efficient calculation of long-range electrostatic interaction. Energy minimization was performed by using 4000 steepest descent steps followed by adopted basis Newton-Raphson (ABNR) method with the harmonic constraints from 10 to 1 kcal/(mol•$Å^2$) in decrements of 3 kcal/(mol•$Å^2$) every 1000 steps to remove any unfavorable van der Waals clashes while minimally perturbing the original model x-ray structure. Using a Langevin barostat (24), an isothermal-isobaric ensemble (NPT) was constructed in NAMD program (25) and system has been heated gradually from 100K to 300K over a period of 400ps. An integration time step of 1fs was used in order to avoid any significant structural deformation during equilibration and production run. Coordinates were saved every 2ps. A total of 2 ns equilibration and 3 ns production simulation were performed for each structural analysis. VMD 1.9.1(26) had been used for visualization purposes.

MM-PBSA interaction energy calculation of ground state and transition state

Molecular mechanics Poisson-Boltzmann solvent accessible surface area (MM-PBSA) approach, $G = < G_{MM} > + < G_{PB} > + < G_{SA} >$, was used to calculate binding free energy difference of each constructed TthUDGa/DNA complexes. The molecular

mechanics energy, G_{MM}, used an infinite cutoff to evaluate the non-bonded interactions. Poisson-Boltzmann polar solvation interaction energies, G_{PB}, were solved with CHARMM using the PBEQ module, which allows the setting up and a successive over relaxation method to solve the Poisson-Boltzmann equation on a discretized grid for the complexes. The bulk solvent dielectric constant and protein interior dielectric constant were set to 80.0 and 4.0, respectively. The non-polar solvation energy, G_{SA}, was approximated with $G_{SA} = \gamma(SASA) + \beta$, where $SASA$ was solvent-accessible surface area, $\gamma = 0.00542\ kcal/mol\text{Å}^2$ and $\beta = 0.92\ kcal/mol$. Solvent accessible surface with a 1.4 Å solvent probe radius was constructed for the solvent-solute dielectric boundary. Non-entropic contribution of MM-PBSA interaction energies change between each amino acid of enzyme and flipping out nucleotide after binding, $\Delta G = G_{complex} - G_{enzyme} - G_{sub}$, were also evaluated. Mulliken charges from reference (27) was scaled and applied to the Charmm partial atomic charge for the purpose of simulating transition state. Two binding free energies of TthUDGa/DNA complexes (ΔG_{TS} and ΔG_G) based on different partial atomic charges was compared for MMPBSA analysis. Differences between them was then estimated according to $\Delta\Delta G = \Delta G_{TS} - \Delta G_G$, which can be correlated to enzyme efficiency.

IV. Results

Substrate specificity of Tth UDGa

Family 4 UDGa is a distinct family in UDG superfamily with limited sequence homology with other families (Fig. 3.1A). Previous reports show that UDGa is a uracil DNA glyocosylase that can act on both double-stranded and single-stranded uracil-containing DNA (10,11,13,28). However, it is not known that whether family 4 UDGa can excise other deaminated bases. Using deoxyoligonucleotide-containing uracil, hypoxanthine, xanthine, and oxanine, we tested the deaminated repair glycosylase activity of *Thermus thermophilus* UDGa on all double-stranded and single-stranded substrates. Tth UDGa exhibited robust glycosylase activity on all uracil substrates but did not show any detectable activity on other deaminated bases (Fig. 3.1B and data not shown). The robust UDG activity was further confirmed by a time course analysis (Fig. 3.1C). In the initial measurement, the reactions were largely completed within a minute. This was confirmed by a 60 sec time course analysis. The excision of all uracil-containing substrates except for the A/U base pair was essentially completed within 30 sec (Fig. 3.1C). Since enzymes in families 2, 3, 5 and 6 can excise other deaminated bases. Apparently, family 4 UDGa has similar narrow substrate specificity as family 1 UNG.

Uracil binding pocket and catalytic mechanism

The availability of crystal structures of both family 4 Tth UDGa and family 1 UNG allows a structural comparison of the uracil binding pockets (15). In Tth UDGa, the uracil binding pocket is defined by E41, G42, E47, F54, N80 and H155, whereas in family 1 UNG, uracil is surrounded by Q63, D64, Y66, F77, N123 and H187 (Fig. 3.1D and 3.1E). To understand the importance of these residues in binding and catalysis, we made a series

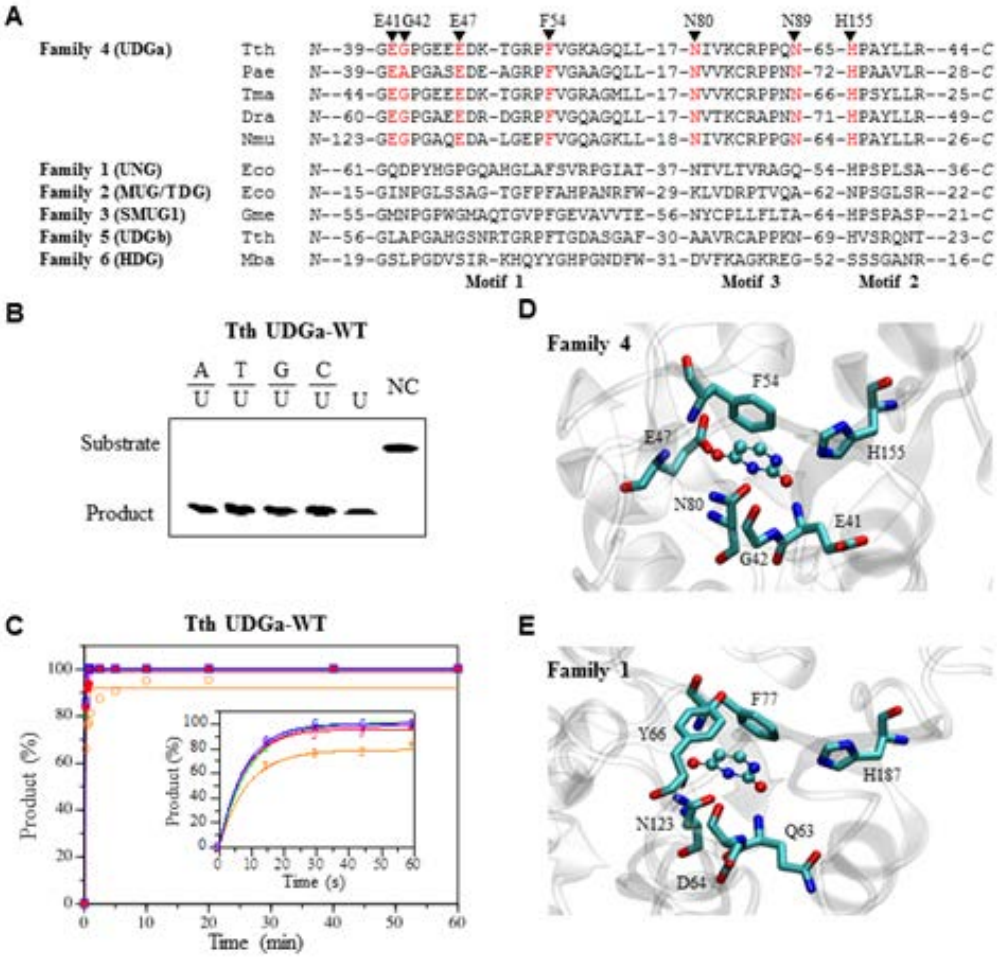


Figure 3.1 Sequence alignment, UDG activity and uracil binding pocket of Family 4 Tth UDGa. A. Sequence alignment in family 4 UDGa and comparison with other UDG families. The alignment was based on BLAST and CLUSTALW analysis and constructed manually. Family 4 (UDGa): *Tth*, *T. thermophilus* HB27, YP_004341.1; *Pae*, *P. aerophilum str*. IM2, NP_558739.1; Dra, *D. radiodurans* R1, NP_295474 (DR 1751); Tma, *Thermotoga maritima* MSB8, NP_228321.1; Nmu, *Nitrosospira multiformis*, YP_412806. Family 1 (UDG): *Eco*, *E. coli*, NP_289138. Family 2 (MUG/TDG): *Eco*, *E. coli*, P0A9H1. Family 3 (SMUG1): *Gme*, *G. metallireducens* GS-15,YP_383069. Family 5 (UDGb): *Tth*, *T. thermophilus* HB8, YP_144415.1. Family 6 (HDG): *Mba*, *Methanosarcina barkeri* str. Fusaro, YP_304295.1. **B**. DNA glycosylase activity of Tth UDGa on U-containing DNA substrates. Cleavage reactions were performed as described under "Materials and Methods" with 100 nM WT Tth UDGa protein and 10 nM substrate. NC, negative control without addition of enzyme. **C**. Time course analysis of DNA glycosylase activity of WT Tth UDGa on U-containing DNA substrates. (△) C/U; (●) G/U; (○) A/U; (▽) T/U; (□) single-stranded U. **D**. Uracil binding pocket in the active site of Tth UDGa crystal structure (PDB code 1UI0). Uracil is colored by atom type. Amino acid residues interacting with the uracil are highlighted in licorice in program VMD. **E**. Uracil binding pocket in the active site of E. coli UNG crystal structure (PDB code 2EUG). Uracil is colored by atom type. Amino acid residues interacting with the uracil are highlighted in licorice in program VMD.

of amino acid substitutions. N89 was also investigated because it is located in an identical position that could play a catalytic role as previously indicated in the study of family 5 Tth UDGb (7). Initially, we screened the UDG activity of all 29 mutants using all five uracil-containing substrates (Table 3.1). The impairment on UDG activity varied depending on the positions and substitutions. The most severe reduction was at H155 position while the least one was at N89 position. Other mutants also showed substantial effects on UDG activity. To more accurately quantify the mutational effects on binding and catalysis, we measure the kinetic constants for the wild type and mutant UDGa. Because the loss of catalytic activity was too great to allow the use of conventional steady state kinetics, we adopted a kinetic method that was previously used for the study of noncognate site in EcoRI and EcoRV restriction enzymes (16,17). In the case that the k_{obs} was plateaued with increasing enzyme concentrations, K_m and k_2 would be obtained (Fig. 3.2A-

Table 3.1 Glycosylase activity of Tth UDGa on uracil substrates[a]

Enzymes	A	T	G	C	
	U	U	U	U	U
WT	100	100	100	100	100
E41A	1	15	10	30	10
E41D	0	4	0	9	0
E41N	1	34	18	62	15
E41Q	0	19	10	29	5
G42D	2	33	14	64	6
G42E	0	12	8	34	0
G42H	0	0	0	0	0
G42L	0	0	0	0	0
G42N	0	1	0	2	0
E47A	0	32	10	57	1
E47D	0	1	0	4	0
E47N	0	0	0	0	0
E47Q	0	0	0	0	0
E47S	0	0	0	0	0
F54A	6	44	13	61	16
F54H	0	40	10	57	7
F54Y	77	98	98	99	96
N80A	11	52	21	85	19
N80Q	1	6	2	14	0
N89A	12	99	82	100	44
N89D	5	96	70	100	36
N89E	0	45	16	65	5
N89Q	19	90	83	95	32
H155F	0	0	0	0	0
H155M	0	0	0	0	0
H155N	0	0	0	0	0
H155S	0	1	0	0	7
H155Y	0	0	0	0	0
E41Q-G42D	7	59	41	89	36

[a]: The reactions were performed as described in Materials and Methods. Numbers are percentage of product signal over total signal. Data are an average of at least two independent experiments.

B). In the case that the K_m was increased to a degree that the plot of k_{obs} versus the total enzyme concentration was linear, only the k_2/K_m would be determined (Fig. 3.2C).

Based on the Tth UDGa structure complexed with a uracil base (15), the mainchain NH of E41 interacts with the O2 of uracil. Substitution of E41 with Ala, Asp, Asn, and Gln all caused a substantial reduction in UDG activity, in particular for the A/U base pair and the single-stranded uracil-containing substrate. Kinetic measurements for the E41Q mutant showed that the k_2/K_m was reduced by three orders of magnitude (Table 3.2). Similar effect was observed for substitution in the adjacent G42 position. Interestingly, two substitutions with a carboxyl sidechain (G42D and G42E) were relatively more active than the other substitutions. In family 1 UNG, the equivalent position is occupied by an Asp residue (Fig. 3.1A). The G42D mutant lowered the k_2/K_m by two orders of magnitude (Table 3.2). E47A

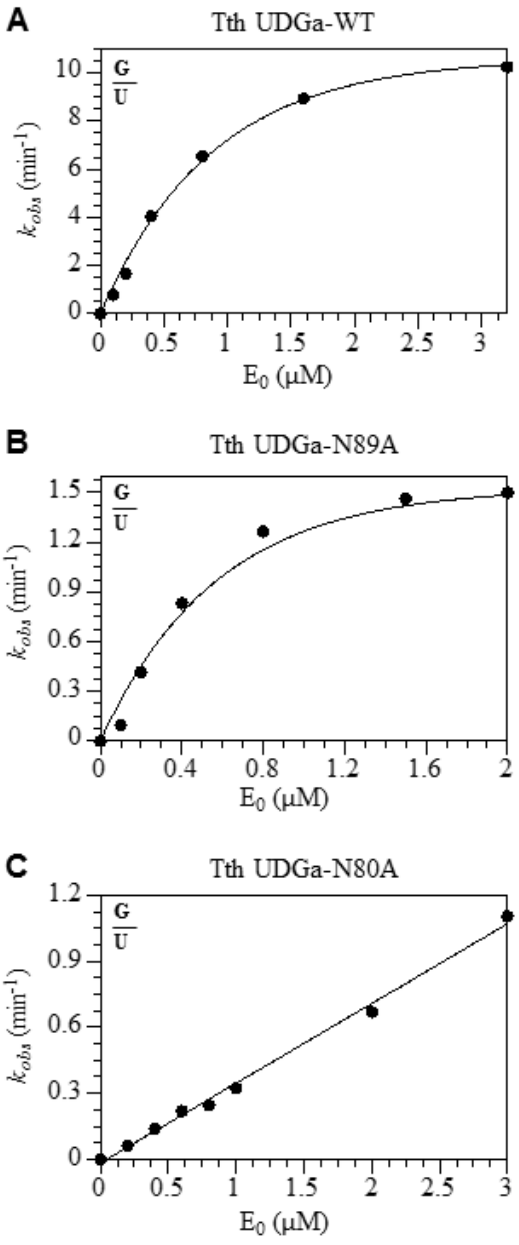


Figure 3.2 Representative kinetics analysis of the wild type and mutant Tth UDGa glycosylase. See Materials and Methods for details. **A**. Tth UDGa-WT. **B**. Tth UDGa-N89A. **C**. Tth UDGa-N80A.

substitution also caused a similar two orders of magnitude reduction in UDG activity on the G/U base pair (Table 3.2). The mutational effects on F54 depended on the nature of substitution. Whereas F54A and F54H had a significant effect on the UDG activity, the conserved change by replacement of F54 with the aromatic Tyr largely retained the UDG activity (Table 3.1). The loss of the aromatic sidechain caused a 17-fold reduction in k_2/K_m value (Table 3.2). N80 makes bidentate hydrogen bonds to the N3 and O4 of uracil (Fig. 3.1D). Substitutions at N80 lowered the k_2/K_m by over 40-fold (Table 3.2). N89A mutant reduced the UDG activity to a lesser degree and was one of the mutants that both K_m and

Table 3.2 Kinetic constants of Tth UDGa on G/U substrate[a]

Enzymes	Substrate	K_m (nM)	k_2 (min^{-1})	k_2/K_m (min^{-1} nM^{-1})
Wild type	G/U	995	13.9	1.4×10^{-2}
E41Q		N.D.[b]	N.D.	2.4×10^{-5}
G42D		N.D.	N.D.	1.5×10^{-4}
E47A		N.D.	N.D.	1.2×10^{-4}
F54A		N.D.	N.D.	8.3×10^{-4}
N80A		N.D.	N.D.	3.3×10^{-4}
N89A		634	2.1	3.3×10^{-3}
H155S		735	4.3×10^{-3}	5.9×10^{-6}
E41Q-G42D		262	6.7×10^{-1}	2.6×10^{-3}

[a]: The reactions were performed as described in Materials and Methods. Data are an average of at least two independent experiments.

[b]: Not determined. Individual K_m and k_2 values were not determined due to a relative large K_m.

k_2 could be obtained (Fig. 3.2B, Tables 3.1 and 3.2). Whereas the K_m for N89A was slightly reduced as compared with the wild type enzyme, the k_2 was reduced by almost 7-fold (Table 3.2). These results indicate a role of N89 in catalysis and will be discussed later. H155S exhibited its effects mostly on k_2 while the K_m was only slightly reduced. The k_2 effect was much more profound than the N89A mutant, resulting in over three orders of magnitude difference as compared with the wild type Tth UDGa (Table 3.2).

Correlated mutations in motif 1

The robust and exclusive glycosylase activity on uracil-containing DNA prompted us to compare the sequences of family 4 UDGa and family 1 UNG closely. Whereas several important structural elements for the UDG function are highly conserved, a notable difference is that in motif 1 the E41-G42 doublet is replaced by Q63-D64 (Fig. 3.1A). The single mutations described above have already shown that substitutions in E41 and G42 are detrimental to the catalytic function of Tth UDGa. In light of the highly conserved nature of the QD doublet in family 1 UNG enzymes, we thought that these two positions may be correlated during evolution for some family enzymes in the UDG superfamily. To test this possibility, we replaced the EG doublet in family 4 Tth UDGa with the QD doublet in family 1 UNG. Indeed, the E41Q-G42D mutant was more robust than any of the single mutants (Table 3.1). To quantitatively compare the catalytic efficiencies, we measured the kinetic constants. The k_2/K_m of the Tth UDGa E41Q-G42D was only 5-fold lower than the wild type enzyme (Table 3.2 and 3.3). Remarkably, the double mutant enhanced the catalytic efficiencies of E41Q and G42D by 108-fold and 17-fold, respectively. These results underscore the important structural and functional correlation of QD doublet in both family 1 UNG and family 4 UDGa.

V. Discussion

Family 4 UDGa enzymes are found in prokaryotes, while family 1 UNG are common in eukaryotes and bacteria. This study comprehensively investigated the deaminated repair activity of family 4 UDGa. To understand the mutational effects and the catalytic mechanism, we also modeled DNA into the Tth UDGa crystal structure (Fig. 3.3A). Data presented here indicate that family 4 UDGa is a glycosylase with rather narrow substrate specificity. Despite its low sequence homology, the uracil binding pocket of family 4 UDGa shares some similar features as seen in family 1 UNG (Fig. 3.1 and 3.3). The mainchain of E41 in motif 1 (equivalent to Q63 in Eco UNG) and the sidechain of H155 (equivalent to H187 in Eco UNG) in motif 2 interact with O2 of uracil, whereas the sidechain of N80 (equivalent to N123 in Eco UNG) form bidentate hydrogen bonds with N3 and O4 of uracil (Figs. 3.1 and 3.3). The aromatic sidechain of F54 (equivalent to F77 in Eco UNG) stacks on top of the uracil ring while its mainchain contacts the O4 of uracil (Figs. 3.1 and 3.3). A distinctly different arrangement is E47 in Tth UDGa, which blocks the entry of thymine (Figs. 3.1 and 3.3). In Eco UNG, Y66 plays a similar role in distinguishing uracil from thymine. The crystal structures complexed with uracil show that E47 in Tth UDGa and Y66 in Eco UNG are located in different structure contexts (Fig.

Table 3.3 Enhancement of Tth UDGa E41Q-G42D double substitution on UDG

Table 3.3. Enhancement of Tth UDGa E41Q-G42D double substitution on UDG activity and free energy[a]

Enzyme	Substrate	k_2/K_m (min^{-1} nM^{-1})	Activity Change (fold)[b]	Fold Enhancement over Single Substitution[c]	ΔΔG (kj mol^{-1}) [d]
Wild type	G/U	1.4×10^{-2}	1		-0.91
E41Q		2.4×10^{-5}	583	108	0.39
G42D		1.5×10^{-4}	93	17	0.02
E41Q-G42D		2.6×10^{-3}	5.4		-0.87

[a]: The reactions were performed as described in Materials and Methods. Data are an average of at least two independent experiments.

[b]: Activity change was calculated by the ratio of k_2/K_m of the wild type to k_2/K_m of a mutant.

[c]: Fold enhancement over single substitution was calculated by the ratio of of k_2/K_m of E41Q-G42D to k_2/K_m of single mutant.

[d]: ΔΔG was calculated by the ΔG between the transition state and the ground state.

3.4). In Tth UDGa, the sidechain of E47 is brought to close proximity to C5 of uracil by an α-helix, while the sidechain of Y66 in Eco UNG located in the loop faces the C5 of uracil (Fig. 3.4A-3.4B). The helix form does not seem possible with Eco UNG because the position equivalent to E47 is occupied by a highly conserved proline residue (P69). The alignment of E47 and Y66 with their corresponding uracil is further underscored by superimposition of Tth UDGa with Eco UNG (Fig. 3.4B). Mutations in the uracil binding pocket appears to affect the binding affinity to uracil by substantially increasing the K_m (Table 3.2). Among the six families in the UDG superfamily, glycine (G44 in Tth UDGa) appears in all five families with the exception of family 1, which is occupied by a tyrosine (Y66 in Eco UNG) (Fig. 3.1A). While the Tyr residue blocks thymine which even differs

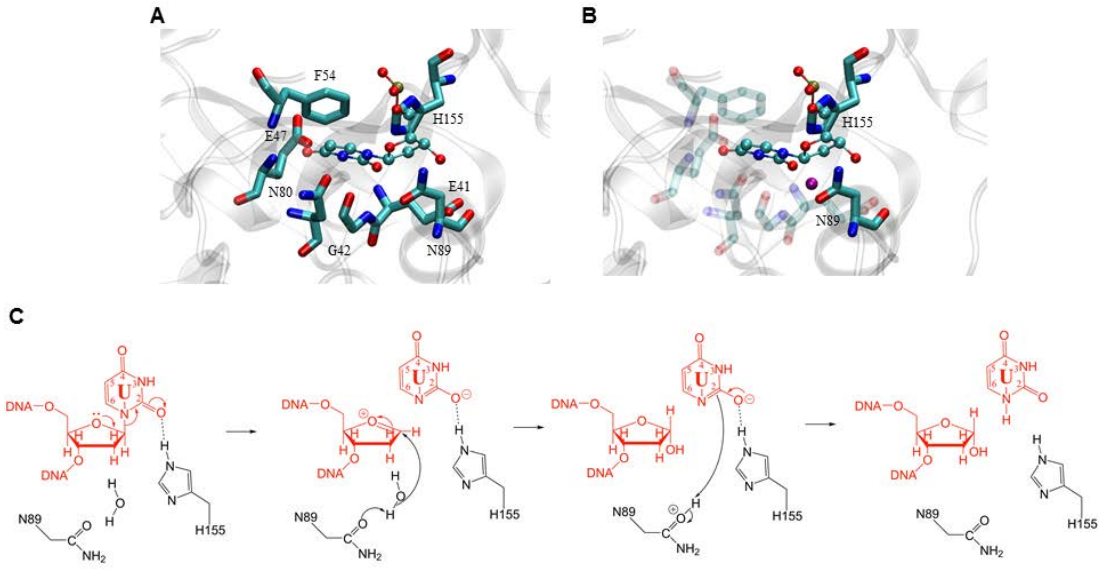


Figure 3.3 Modeled structure and proposed catalytic mechanism of Tth UDGa. A. Modeled structure of Tth UDGa complexed with uracil-containing DNA in the energy minimized structure. The protein structure is shown in the background in light gray. dUMP is colored by atom type. Amino acid residues in the active site of Tth UDGa are shown in licorice in program VMD. **B**. Interactions of N89 and H155 with dUMP in the modeled structure. The water molecule found in the modeled structure between N89 and the C1' carbon is shown in purple. Water molecule is shown as a sphere in purple. **C**. Proposed catalytic mechanism of family 4 Tth UDGa glycosylase. See text for details.

from uracil only by a methyl group from entering the binding pocket, the adaption of a Gly residue in families 2, 3, 5 and 6 may play a role in enabling these families to act on a more bulky purine deaminated bases such as hypoxanthine or xanthine or methylcytosine derivatives such as carboxylcytosine or formylcytosine (6-8,29,30). Indeed, structural

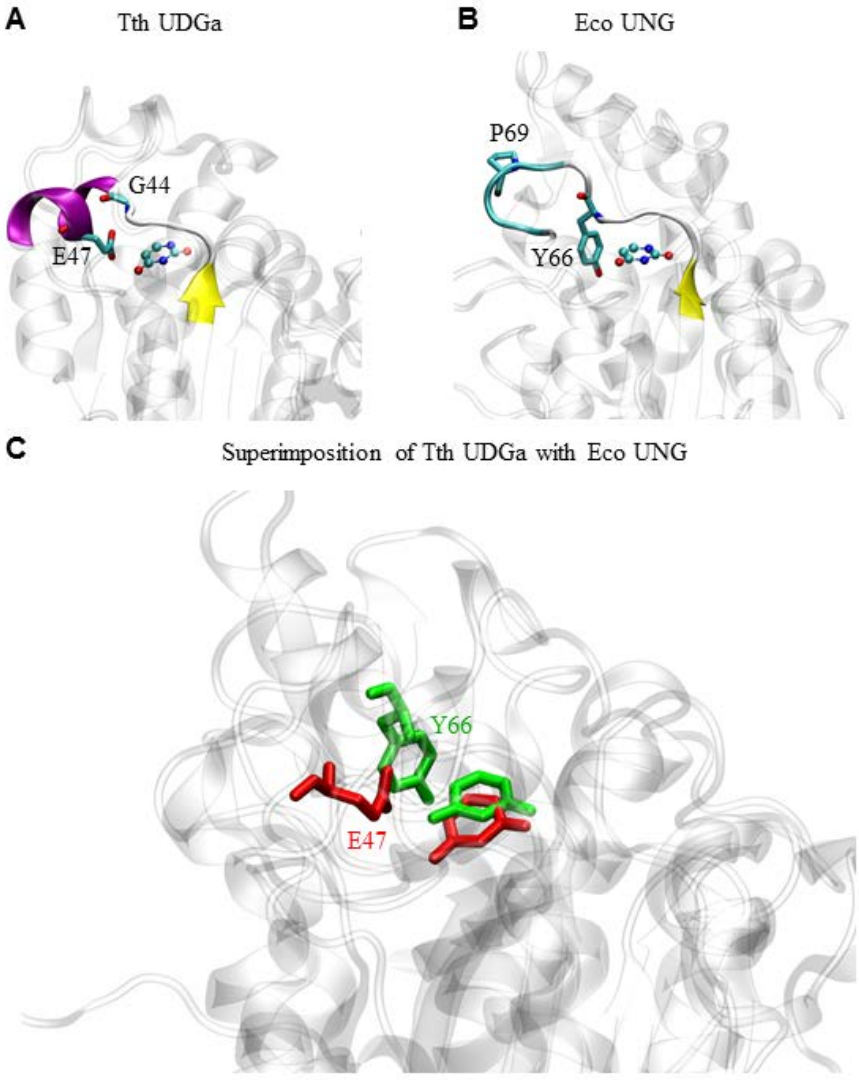


Figure 3.4 Comparison of E47 of Tth UDGa with Y66 of Eco UNG. A. Amino acid residues 40-50 of Tth UDGa and uracil in the crystal structure (PDB code 1UI0). Uracil is colored by atom type. Amino acid residues are shown in licorice in program VMD. **B**. Amino acid residues 62-72 of Eco UNG and uracil in the crystal structure (PDB code 2EUG). **C**. Superimposition of Tth UDGa with Eco UNG. E47 of Tth UDGa and the uracil in the crystal structure are shown in red. Y66 of Eco UNG and the uracil in the crystal structure are shown in green. The two structures were superimposed using the program Topmatch.

comparisons show that the Gly residue is located in an extensive loop region, which could allow the flipping of a bulkier base into the binding pocket (Fig. 3.5). On the other hand, replacement of the Gly with a Tyr can block the flipped out base from entering the binding pocket, as seen in family 3 SMUG1 (6).

The cleavage of the N-glycosidic bond between the uracil and deoxyribose is achieved through the formation of an oxacarbenium ion intermediate and attacking of the anomeric carbon by a water molecule (31,32). Activation of the leaving group, stabilization

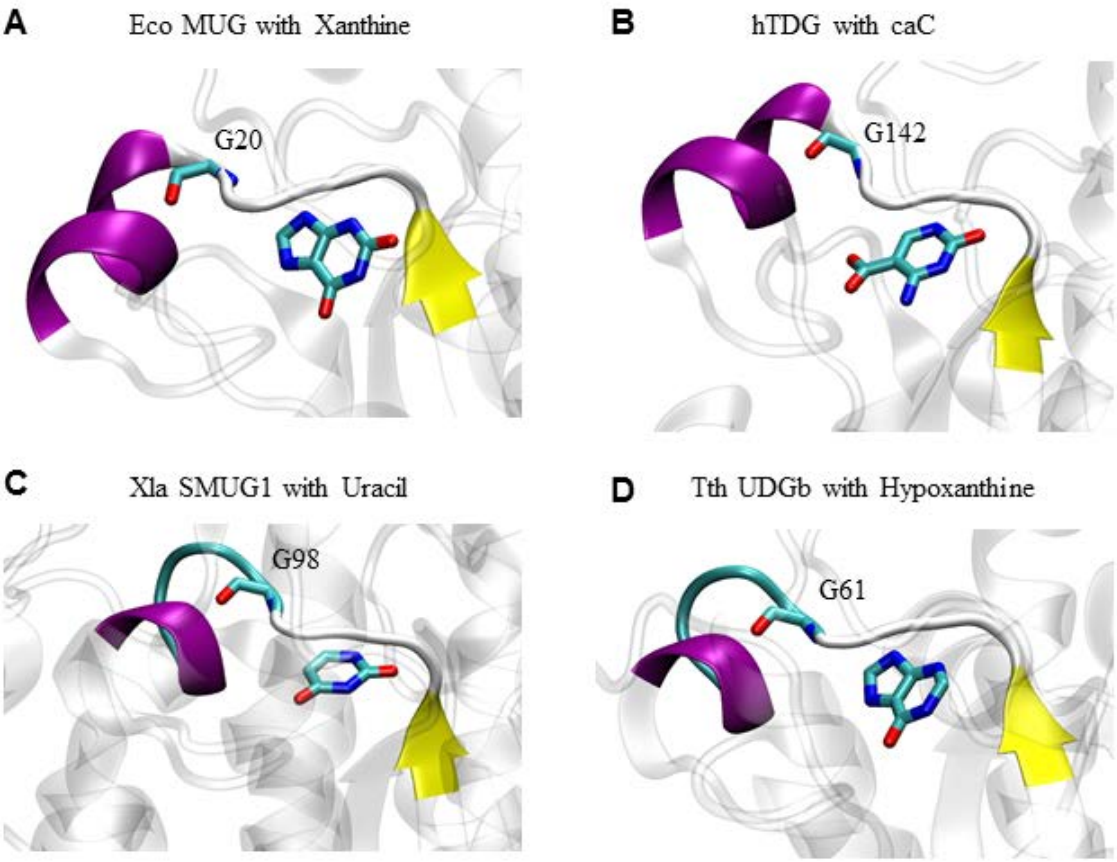


Figure 3.5 Comparison of UDG families with family 1 Eco UNG in the secondary structural segment around Y66. A. Modeled family 2 Eco MUG structure (PDB code 1MUG) complexed with xanthine. **B**. Family 2 hTDG structure (PDB code 3UOB) complexed with carboxylcytosine. **C**. Family 3 Xenopus laevis structure (PDB code 3UOB) complexed with carboxylcytosine. **D**. Modeled family 5 Tth UDGa structure (PDB code 2DEM) complexed with hypoxanthine.

of the oxacarbenium ion and activation/positioning of water as a nucleophile may contribute to the catalysis. The catalytic mechanism underlying the hydrolysis of the N-glycosidic bond in family 4 UDGa is not understood. In family 1 UNG, a His residue (H187 in Eco UNG) in motif 2 can act as a general acid to stabilize the uracil leaving group and an Asp residue (D64 in Eco UNG) in motif 1 is proposed to activate a water molecule as a general base (33-35). Part of the challenge in suggesting a catalytic mechanism for family 4 UDGa lies at the fact that the water-activating Asp residue in motif 1 of family 1 UNG is a Gly or Ala residue in motif 1 of family 4 UDGa (Fig. 3.1A). This work implicates two residues as playing an important role in catalysis. Mutational effects at N89 and H155 positions are mainly at the catalytic step (Table 3.2). The four orders of magnitude change in k_2 and k_2/K_m by H155S substitution indicates that H155 in motif 2 is critical for catalysis. The contact made between the H155-NE2 and O2 of uracil can stabilize the uracil leaving group, thus promoting the cleavage of the N-glycosidic bond (Fig. 3.3B). Similarly, H187 in Eco UNG makes a large contribution to transition state stabilization by forming a hydrogen bond (34,36). In the modeled structure, N89 in a sequence segment we named motif 3 is located on the opposite site of the uracil relative to the deoxyribose (Figs. 3.1A and 3.3B). In the sequence alignment shown in Fig. 1A, N89 corresponds to N120 in family 5 Tth UDGb. The kinetics analysis shows that N89 in Tth UDGa plays a significant catalytic role (Table 3.2). Previously, we proposed that N120 in family 5 Tth UDGb can contribute to catalysis by activating/positioning a water molecule observed in the crystal structure (7). Analogously, we suggest that N89 in family 4 Tth UDGa can activate/position a water molecule for attacking the anomeric carbon (Fig. 3.3B). Overall, we propose an S_N1-like catalytic mechanism for the family 4 Tth UDGa, in which H155 stabilizes the

uracil leaving group and N89 activates/positions a water molecule for attacking the anomeric carbon (Fig. 3.3C).

To understand the structural and functional correlation between E41 and G42 positions in family 4 UDGa, we conducted molecular dynamics (MD) analysis. In the wild type enzyme, the average hydrogen bond distances between the mainchain of E41Q and O2 of uracil and between the sidechain of H155 and O2 of uracil are 3.26 Å and 2.86 Å, respectively (Fig. 3.6A and 3.6E). The short distance between H155-NE2 to the O2 of uracil is suggestive of a strong hydrogen bond. E41Q mutation increased the distances between the O2 of uracil to the mainchain of E41Q and the sidechain of H155 to 3.38 Å and 3.39 Å, respectively (Fig. 3.6B and 3.6F). This outcome would substantially weaken the hydrogen bonds to O2, resulting in a large loss of UDG activity. The structural effect caused by G42D mutation is more profound for the hydrogen bond distance between the uracil and the E41 than that between the uracil and H155. The increases in average distance are 4.12 Å and 3.03 Å, respectively (Fig. 3.6C and 3.6G). The concurrent change of E41Q and G42D, however, shortens the hydrogen bond distances between O2 of uracil and the mainchain of E41Q and between O2 of uracil and the sidechain of H155 to 3.27 Å and 2.91 Å, respectively (Fig. 3.6D and 3.6H). The molecular mechanics Poisson-Boltzmann solvent accessible surface area (MM-PBSA) binding free energies were also performed during MD analysis. The calculated energy differences $\Delta\Delta G$ between ground state and transition state for wile type, E41Q, G42D, and E41Q/G42D are -0.91, 0.39, 0.02 and -0.87kj/mol, respectively. The binding free energy changes demonstrated that concurrent change of E41Q and G42D could effectively enhance the TthUDGa/DNA complexes interaction when compared to the single mutant E41Q or G42D, which agreed well with

the k_2/K_m determined in enzyme kinetic analysis. (Table 3.3). The structural alignment of the two important hydrogen bonds brought about by E41Q-G42D doublet is in line with the large recovery of the lost UDG activity in single mutants (Table 3.3). These analyses suggest that these two positions are intrinsically correlated and the EG doublet or the QD doublet works in concert to exert its structural and functional impact on family UDGa.

In summary, this study reveals that family 4 UDGa is a narrow specificity but robust uracil DNA glycosylase with a binding pocket evolved for accommodating a uracil base. While both families 1 and 4 glycosylases use histidine-mediated transition state stabilization for the cleavage of the N-glycosidic bond, they differ by how to

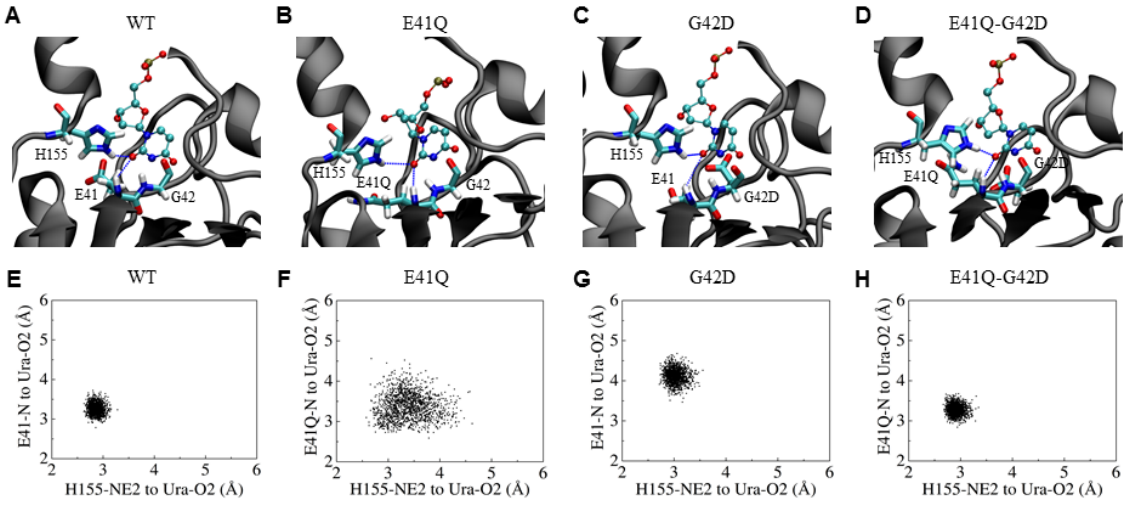


Figure 3.6 Interactions and two-dimensional scatter plots of the wild type and mutant Tth UDGa proteins with O2 of uracil in the active site. Modeled interactions with O2 of uracil in the active site of Tth UDGa-WT (**A**), Tth UDGa-E41Q (**B**), Tth UDGa-G42D (**C**) and Tth UDGa-E41Q-G42D (**D**). dUMP is colored by atom type. Amino acid residues in the active site of Tth UDGa are shown in licorice in program VMD. Two-dimensional scatter plots of heavy atom distances between E41, H155 and uridine in Tth UDGa-WT (**E**), Tth UDGa-E41Q (**F**), Tth UDGa-G42D (**G**) and Tth UDGa-E41Q-G42D (**H**). The distances were obtained from MD trajectories in the modeled enzyme-DNA complexes.

activate/position a water molecule for attacking the anomeric carbon. The correlation between the E41 and the G42 positions underscores the importance of coevolution in the divergence of UDG superfamily.

VI. References

1. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715
2. Duncan, B. K., and Miller, J. H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560-561
3. Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., and Ellenberger, T. (2006) *DNA repair and mutagenesis*, Second ed., ASM Press, Washington, DC
4. Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B., and Sperens, B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *The Journal of biological chemistry* **252**, 3286-3294
5. Dong, L., Mi, R., Glass, R. A., Barry, J. N., and Cao, W. (2008) Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase. *DNA repair* **7**, 1962-1972
6. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N., and Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. *Journal of molecular biology* **385**, 761-778
7. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N., and Cao, W. (2014) Specificity and Catalytic Mechanism in Family 5 Uracil DNA Glycosylase. *The Journal of biological chemistry* **289**, 18413-18426

8. Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2003) The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic acids research* **31**, 2261-2271

9. Haas, B. J., Sandigursky, M., Tainer, J. A., Franklin, W. A., and Cunningham, R. P. (1999) Purification and characterization of Thermotoga maritima endonuclease IV, a thermostable apurinic/apyrimidinic endonuclease and 3'-repair diesterase. *Journal of bacteriology* **181**, 2834-2839

10. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H., and Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J* **21**, 3182-3191

11. Sandigursky, M., and Franklin, W. A. (2000) Uracil-DNA glycosylase in the extreme thermophile Archaeoglobus fulgidus. *The Journal of biological chemistry* **275**, 19146-19149

12. Liu, X. P., and Liu, J. H. (2011) Characterization of family IV UDG from Aeropyrum pernix and its application in hot-start PCR by family B DNA polymerase. *PloS one* **6**, e27248

13. Starkuviene, V., and Fritz, H. J. (2002) A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus. *Nucleic acids research* **30**, 2097-2102

14. Sakai, T., Tokishita, S., Mochizuki, K., Motomiya, A., Yamagata, H., and Ohta, T. (2008) Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus. *DNA repair* **7**, 663-669

15. Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S., and Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. *Journal of molecular biology* **333**, 515-526

16. King, K., Benkovic, S. J., and Modrich, P. (1989) Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease. *The Journal of biological chemistry* **264**, 11807-11815

17. Vermote, C. L., and Halford, S. E. (1992) EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition. *Biochemistry* **31**, 6082-6089

18. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5083-5088

19. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *ELECTROPHORESIS* **18**, 2714-2723

20. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545-1614

21. MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *The Journal of Physical Chemistry B* **102**, 3586-3616

22. MacKerell, A. D., and Banavali, N. K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *Journal of Computational Chemistry* **21**, 105-120

23. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *Journal of Chemical Physics* **98**, 10089-10092

24. Adelman, S. A., and Doll, J. D. (1976) Generalized Langevin equation approach for atom/solid‐surface scattering: General formulation for classical scattering off harmonic solids. *Journal of Chemical Physics* **64**, 2375-2388

25. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802

26. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38

27. Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature (London, U. K.)* **413**, 752-755

28. Sandigursky, M., and Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes. *Current biology : CB* **9**, 531-534

29. Lee, H. W., Brice, A. R., Wright, C. B., Dominy, B. N., and Cao, W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase. *The Journal of biological chemistry* **285**, 41483-41490

30. Lee, H. W., Dominy, B. N., and Cao, W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *The Journal of biological chemistry* **286**, 31282-31287

31. Berti, P. J., and McCann, J. A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chemical reviews* **106**, 506-555

32. Stivers, J. T., and Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical reviews* **103**, 2729-2759

33. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-493

34. Drohat, A. C., Jagadeesh, J., Ferguson, E., and Stivers, J. T. (1999) Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry* **38**, 11866-11875

35. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87-92

36. Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L., and Stivers, J. T. (1999) Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **38**, 11876-11886

CHAPTER FOUR

A NEW CLASS OF DEAMINATION REPAIR ENZYME IN URACIL DNA GLYCOSYLASE SUPERFAMILY

I. Abstract

Uracil DNA glycosylases plays an important role in repairing deaminated DNA bases. Previous studies have identified six different families within UDG superfamily. Here, we report a putative new UDG family from *Streptococcus mutans* and *Methylobacterium radiotolerans*, with robust activities on xanthine-containing DNA. Mutational analysis combined with molecular modeling and molecular dynamics analysis revealed that the new group of UDG utilized multiple residues to specifically recognize xanthine and catalyze the hydrolysis of N-glycosidic bond. Additional phylogenetic analysis indicated the new group of enzymes formed a separate group within UDG superfamily. This study provides new insights into the evolution and the catalytic mechanisms of UDG superfamily.

II. Introduction

DNA is vulnerable to chemical damage under normal or stressed conditions (1). One of the common types of DNA base damage is deamination, including cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine, usually generated under nitrous stress condition. Uracil DNA glycosylases (UDG) are a group of the enzymes to repair the deaminated DNA bases (2). To date, six UDG families have been studied in the UDG superfamily (3). Functionally, they are able to remove at least one of four kinds of deaminated DNA bases, and most of them exhibit uracil DNA glycosylase activities except family 6 hypoxanthine DNA glycosylase (HDG), which has no detectable uracil DNA glycosylase activity. Structurally, they share secondary structure elements, including five β-sheets and four α-helices. Family 1 UNG are first discovered in *Escherichia coli* and characterized as a very efficient enzyme to repair uracil in both double-stranded DNA and single-stranded DNA (4,5). Family 2 MUG/TDG is first found as a uracil DNA glycosylase due to its activity on mismatched T/U, G/U and C/U base pairs, but identified as a xanthine DNA glycosylase (XDG) in *E. coli* later (6). Family 3 SMUG1 is able to repair uracil from double- and single-stranded DNA, however, detailed study of *Geobacter metallireducens* SMUG1 and human SMUG1 revealed that the enzymes also possess xanthine DNA glycosylase activity (7). Family 4 UDGa is found as uracil DNA glycosylases in several hyperthermophilic bacterium (8,9). The preceding chapter provides experimental evidence that family 4 UDGa is an exclusive uracil DNA glycosylase. Similar with family 4 UDGa, family 5 UDGb is found in several hyperthermophilic bacterium, however, a detailed study on UDGb from *T. thermophilus* demonstrated the enzyme have activities on hypoxanthine- and xanthine-containing DNA as well (10). Different from other UDG families, the newly

identified family 6 enzymes are hypoxanthine DNA glycosylase with minor xanthine DNA glycosylase activity (3).

Streptococcus mutans genome contains an open reading frame with its protein structure (PDB:3IKB) remotely related to known enzymes in UDG superfamily. Recent phylogenetic analysis also indicates the existence of a new UDG family for UDG superfamily, with a unique amino acid sequence "GQAPG" in motif 1 (11). However, the biochemical properties of this group of enzymes have not been characterized. Here, we report that the enzyme from *S. mutans* belongs to a new group of DNA glycosylases in the UDG superfamily. Unlike previously known UDG families, enzymes from this new family exhibit robust activity on all xanthine-containing DNA compared to its uracil and hypoxanthine repair activity. Detailed mutagenesis analysis and molecular dynamics analysis demonstrated that it possesses a unique active site to distinguish xanthine from uracil and hypoxanthine. Based on its robust xanthine repair activity, special active site architecture and phylogenetic analysis, we propose this class of UDG as family 7 xanthine DNA glycosylase (XDG). The discovery of this new UDG family provides valuable information about the mechanism of xanthine repair and evolution of UDG superfamily.

III. Experimental Procedures

Reagents, media, and strains

All routine chemical reagents were purchased from Sigma Chemicals (St. Louis, MO), Fisher Scientific (Suwanee, GA), or VWR (Suwanee, GA). Restriction enzymes, Phusion DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). Bovine serum albumin and dNTPs were purchased from Promega (Madison, WI). Gel DNA recovery Kit was purchased from Zymo Research (Irvine, CA). Oligodeoxyribonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA) and Eurofins Genomics (Huntsville, AL). Genomic DNA from *Streptococcus mutans* UA159 and *Methylobacterium radiotolerans* were purchased from ATCC. The LB medium was prepared according to standard recipes. Hi-Di Formamide and GeneScan 500 LIZ dye Size Standard for ABI3130xl were purchased from Applied Biosystems. The protein sonication buffer consisted of 20 mM Tris-HCl (pH7.5), 1 mM ethylenediaminetetraacetic acid (EDTA) (pH8.0), 2.5 mM DTT, 0.15 mM PMSF, and 50 mM NaCl. The TE buffer consisted of 10 mM Tris-HCl (pH8.0) and 1 mM EDTA.

Cloning, Expression and Purification of Smu UDG and Mra UDG

The uracil DNA glycosylase gene from *Streptococcus mutans* (*Smu* UDG) (NCBI Reference Sequence: WP_002263204.1) and *Methylobacterium radiotolerans* (*Mra* UDG) (NCBI Reference Sequence: YP_001754430.1) was amplified by PCR using the forward primer Smu-UDGF (5'-CCGGAATTCCGGATCCATGACAAGTCTTGAAGAAATT ACC -3' (*Bam*HI)) and the reverse primer Smu-UDGR (5'-CCGCTCGAGTGATGA TTGAATAATTTGCTG-3' (*Xho*I)) for Smu DNA glycosylase and Mra-UDGF (5'-AAAGGGGAACATATGCCCGCCTTCGACGACACCGCC -3' (*Nde*I)) and Mra UDGR

(5’-TGTGACGGATCCCCCCGCCATCACCCGCGCCACCTC-3’ (*Bam*HI)) for Mra DNA glycosylase. The PCR reaction mixture (20 μl) consisted 10 ng genomic DNA, 500 nM forward and reverse primers, 1x phusion polymerase buffer, 200 μM each dNTP and 0.2 unit of phusion polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 98°C for 30 s; 30 cycles of three-step amplification with each cycle consisting of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, and extension at 72°C for 20 s; and a final extension step at 72°C for 10 min. The PCR product was purified and cloned into pET21a vector. The recombinant plasmid was confirmed by DNA sequencing.

Site-directed mutagenesis was performed by using an overlapping extension PCR procedure similarly as previously described (10). Taking the mutant Q42A as an example: The first round of PCR was carried out using plasmid pET21a-Smu-UDG as template DNA with two pairs of primers, Smu-UDGF and Q42AR (5’-GGCCTTAATT CCCGGTGCCGCACCAACAATATTAATACG-3’) pair; Q42AF (5’-CGTATTAA TATTGTTGGTGCGGCACCGGGAATTAAGGCC-3’) and Smu-UDGR pair. The PCR products were electrophoresed on 1% agarose gel and the expected PCR fragments were purified from gel slices by Gel DNA clean Kit. The second run of the PCR reaction mixture (20 μl), which contained 1 μl of each of the first run PCR fragments, 200 μM dNTPs, 1×Phusion DNA polymerase buffer, and 0.2 units of Phusion DNA polymerase (New England Biolabs), was initially carried out with a predenaturation step at 95 °C for 30 s; 5 cycles with each cycle of denaturation at 98 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72 °C for 30 s; and a final extension at 72 °C for 5 min. Afterward, 500 nM of outside primers (Smu-UDGF and Smu-UDGR) was added to the above PCR reaction

mixture. The subsequent overlapping PCR amplification included a predenaturation step at 98 °C for 15 s; 30 cycles with each cycle of denaturation at 98 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72°C for 30 s; and a final extension at 72°C for 10 min. Subsequent molecular cloning procedures were performed as previously described. The purified PCR products digested with a pair of *Bam*HI and *Xho*I endonucleases were ligated to the cloning vector pET21a treated with the same pair of restriction endonucleases. The recombinant plasmids containing the desired mutations were confirmed by DNA sequencing and transformed into *E. coli* host strain BH214 (*ung-, mug-, nfi-*).

Protein expression and purification were performed as previously described (7). The pET21a-Smu-UDG or pET21a-Mra-UDG was transformed into *E. coli* strain BH214 [thr-1, ara-14, leuB6, tonA31, lacY1, tsx-78, galK2, galE2, dcm-6, hisG4, rpsL, xyl-5, mtl-1, thi-1, ung-1, tyrA::Tn10, mug::Tn10, supE44, (DE3)] by the standard protocol to express the C-terminal His-6-tagged protein. Briefly, the protein was induced by 0.5 mM IPTG at 16 °C for 12h. After sonication and purification, fractions (300–400 mM imidazole, 60-80% chelating buffer B) containing the UDG protein as seen on 12.5% SDS-PAGE were pooled and concentrated by Amicon YM-10 (Millipore). The concentration of *Smu* UDG *and Mra* UDG protein was determined by SDS-PAGE analysis using bovine serum albumin as a standard and confirmed by measuring absorption at A_{280}. The protein was stored in aliquots at -80 °C. Prior to use, the protein was diluted with 2 x storage buffer (20 mM Tris-HCl pH8.0, 2 mM DTT, 2 mM EDTA, 400 μg/ml BSA, 100% Glycerol).

Oligodeoxynucleotide substrates

Oligodeoxynucleotides containing deoxyuridine (U), deoxyinosine (I), or deoxyxanthosine (X) were obtained or constructed as previously described. (7)

DNA glycosylase activity assays

DNA glycosylase cleavage assays for Smu DNA glycosylase and Mra DNA glycosylase were performed under optimized reaction conditions at 37°C in a 10 μl reaction mixture containing indicated amount of oligonucleotide substrate, an indicated amount of glycosylase, 20 mM Tris-HCl (pH7.6), 100 mM KCl, 1mM DTT, and 1 mM EDTA. The resulting abasic sites were cleaved by incubation at 95 °C for 5 min after adding 0.5 μl of 1 M NaOH. Samples for ABI 3130xl sequencer (Applied Biosystems) were prepared by mixing 0.5 μl of reaction mixture with 5 μl Hi-Di Formamide and 0.15 μl GeneScan 500 LIZ Size Standard, and loaded into ABI 3130xl and run with a fragment analysis module. Cleavage products and remaining substrates were analyzed by Gene Mapper software.

Enzyme Kinetic Analysis

Uracil DNA glycosylase assay was performed at 37°C with 20 nM deaminated substrates when enzyme is in excess. Enzyme concentration ranged from 100 nM to 3200 nM. Samples were collected at time point from 2s, 5s, 10s, 30s, 1 min, 2.5 min, 5 min, 10 min, 15 min, 25 min, 30 min, 40 min, and 60 min. The apparent rate constants for each concentration were determined by curve fitting using the integrated first-order rate equation [1]:

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

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

The kinetic parameters k_2 and *K*m were obtained from plots of k_{obs} against enzyme concentration ($[E_0]$) using a standard hyperbolic kinetic expression with program GraphPad 4.1 following the equation [2].(12)

$$k_{obs} = \frac{k_2[\mathrm{E_0}]}{K_\mathrm{m} + [\mathrm{E_0}]} \quad [2]$$

For some mutants with a large *K*m in which $K_m >> [E_0]$, the kinetic parameter k_2/K_m were obtained from plots of k_{obs} against enzyme concentration ($[E_0]$) using a linear regression with program GraphPad 4.1 following the equation [3].(13)

$$k_{obs} = \frac{k_2[\mathrm{E_0}]}{K_\mathrm{m}} \quad [3]$$

Molecular modeling and molecular dynamics simulation

The crystal structure of family 7 UDG was acquired from the RCSB Protein Data Bank (accession code 3IKB), and used as an enzyme model for subsequent computational analysis. Structure of substrates DNA d(ATGTTGCBTTAGTCC), where the B was the base to be flipped out of the helix, was extracted from the crystal structure of family 5 TthUDGb-DNA complex (PDB accession code 2DEM) (14) using the Swiss-Pdb Viewer (SPDBV) program (15). The base complementary to Uridine, Inosine and Xanthosine were systematically modified to guanine, thymine and cytosine, respectively. Mutants P84N of family 7 UDG was also made using the mutation tool in Swiss-Pdb Viewer program. After building the initial complex structures, an explicit solvent system using the TIP3P water model was constructed in the CHARMM c35b6 molecular mechanics package (16). A total of 2 ns equilibration and 3 ns production simulation were performed for each structural analysis in NAMD program (17). A detailed description of the methods of molecular dynamics simulation can be found in our previous work (18). Similarly, the modeled structures of family 2 E. coli MUG and family 3 Gme SMUG1 was built as previously

described (6,7). The snapshots were generated with the Visual Molecular Dynamics (VMD) 1.9.1 program (19).

Phylogenetic analysis

Proteins sequence alignment was generated by ClusteralW program with following parameters for pairwise alignment: gap opening penalty, 10; gap extension penalty, 0.2. The parameters for multiple alignment were as follows: gap opening penalty, 10; gap extension penalty, 0.1. Other parameters were: protein weight matrix, Blosum; residue-specific penalties, on; hydrophilic penalties, on; gap separation distance, 4; end gap separation, off. The resulting alignment was curated manually by comparing alignment result from a structural based protein sequence alignment from PROMALS3D. The phylogenetic tree was generated with a neighbor-joining algorithm within MEGA6 software package.

IV. Results

Repair activity of Smu DNA glycosylase

The structure of the putative DNA glycosylase from *S. murans* (PDB code 3IKB) is similar to known UDG structures (Fig. 4.1), which is consistent with previous phylogenetic analysis for UDG superfamily indicating a new putative UDG family (11). Taking advantage of the available crystal structure, we modeled deaminated nucleotides into the protein structure. The structural elements that may interact with deoxyxanthosine are shown in Fig. 4.2A. BLAST search identified a group of proteins that are homologous with the putative Smu DNA glycosylase (Fig. 4.2B). Within the three motifs that are critical to the glycosylase function, some distinct differences were noted (Fig. 4.2B). As shown in previous phylogenetic analysis, the first few amino acids in motif 1 in this putative new class are "GQAP" instead of "GQDP" as seen in family 1 UNG. D64 in motif 1 in family 1 UNG is known for its role in water activation. Motif 3 is also unique in which it starts with a Pro residue and ends with an Asp residue (Fig. 4.2B).

To study the potential DNA repair activity of this putative new class of UDG enzymes, we first produced recombinant protein from a homologous gene in *Streptococcus*

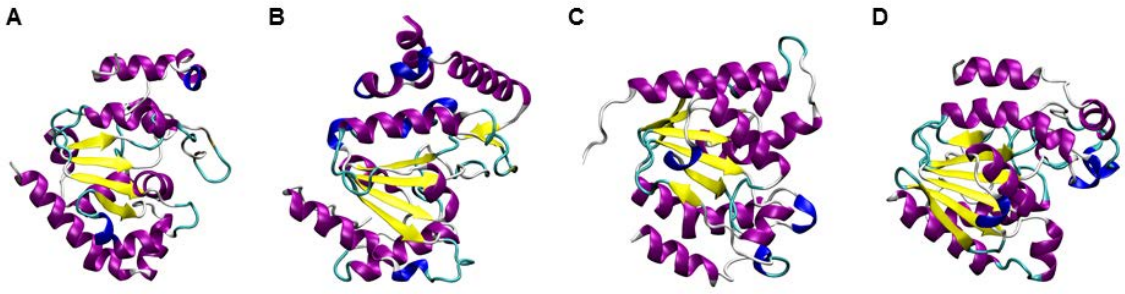


Figure 4.1 Comparison of total protein structure of Smu DNA glycosylase and other UDG enzymes. A. Smu DNA glycosylase (PDB: 3IKB). **B.** Family 1 UNG from Human herpesvirus 1 (PDB: 1LAU). **C.** Family 4 UDGa from *Thermus thermophiles* (PDB: 1UI0). **D.** Family 5 UDGb from *Thermus thermophiles* (PDB: 2D3Y).

mutans (Smu) and measured its enzymatic activities towards all twenty deaminated DNA substrates under conditions in which enzyme was in ten-fold excess over substrate (Fig. 4.3). The putative Smu DNA glycosylase exhibited broad repair activity, as it excised deaminated bases from uracil-, hypoxanthine- and the xanthine-containing DNA (Fig. 4.3C). Its uracil DNA glycosylase activity was higher with the C/U and the T/U base pairs and lower with the G/U base pair (Fig. 4.3C). Low-level UDG activity was observed with the single-stranded uracil substrate and the A/U base pair. Hypoxanthine DNA glycosylase activity was also observed with all hypoxanthine-containing substrates. Xanthine DNA

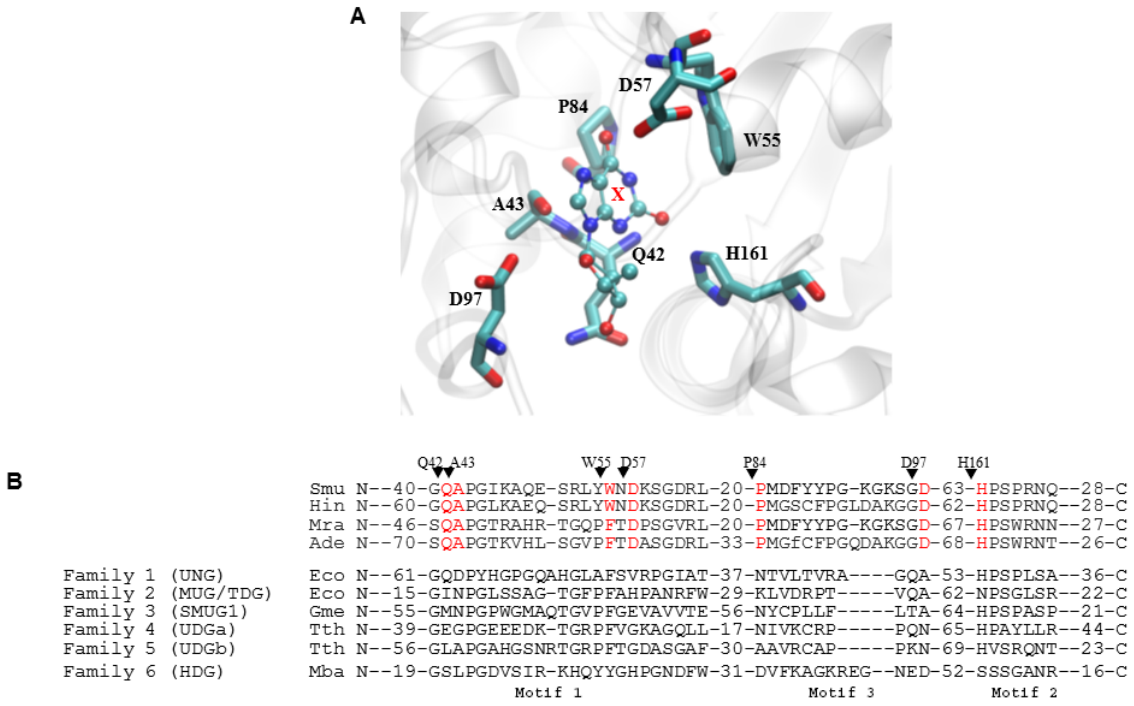


Figure 4.2 Active site of Smu DNA glycosylase and sequence alignment of UDG superfamily with homologous genes of Smu DNA glycosyalse. A. Active site of Smu DNA glycosylase with xanthosine. Residues were subjected to mutational analysis are labelled in *dark.* **B.** Sequence alignment of putative family UDG with UDG superfamily. Homologous genes of Smu DNA glycosyalse (XDG): Ade, *Aquamicrobium defluvii*, EXL01735.1; Hin, *Haemophilus influenza*, WP_005694076.1; Mra, *Methylobacterium radiotolerans* JCM 2831, YP_001754430.1; Smu, *Streptococcus mutans* UA159, NP_721617.1. Family 1 (UDG): Eco, *E. coli*, NP_289138. Family 2 (MUG/TDG): Eco, *E. coli*, P0A9H1. Family 3 (SMUG1): Gme, *G. metallireducens* GS-15,YP_383069. Family 4 (UDGa): Tth, *T. thermophilus* HB27. Family 5 (UDGb): Tth, *T. thermophilus* HB8, YP_144415.1. Family 6 (HDG): Mba, *Methanosarcina barkeri str. Fusaro*, YP_304295.1.

glycosylase activity was the highest but no oxanthine DNA glycosylase activity was detectable under the assay conditions (Fig. 4.3C and data not shown). To verify the glycosylase activity pattern of the new class of DNA glycosylase, another homologous gene from *Methylobacterium radiotolerans* (Mra) was investigated (Fig. 4.3D). A similar activity pattern was observed in the Mra DNA glycosylase with XDG activity being the highest (Fig. 4.3D).

To more quantitatively define the catalytic efficiency towards different deaminated DNA bases, kinetic parameters of the Smu DNA glycosylase were determined using the G/U, T/I and C/X base pairs, which are three biologically relevant substrates. Since UDG and HDG enzymatic activities are not high enough to allow us to use conventional steady state kinetics method, we adopted a kinetic method which allow us to measure the kinetic parameters under the condition that enzyme is in excess (12,13). As shown in Table 4.2, the Smu DNA glycosyase was not very active towards G/U and T/I substrates. For the G/U base pair, the k_2 and K_m values were 0.2 min^{-1} and 832 nM, respectively. For the T/I base pair, the k_2 and K_m values were 0.64 min^{-1} and 629 nM, respectively. In contrast, the Smu DNA glycosyase showed a more robust activity towards the C/X base pair. Time course analysis indicated that k_{obs} of Smu DNA glycosyalse towards the C/X base pair stayed constant under different enzyme concentrations suggesting k_2 towards C/X substrates is 3.2 min^{-1} (Fig. 4.4A and Table 4.2). Since the lowest enzyme concentration we used is 100 nM, this result indicated the exact value of *K*m of Smu DNA glycosyalse towards C/X containing substrates is smaller than 100 nM (if *K*m is around 100 nM, the curve should be similar as Fig. 4.4B; If *K*m is much large than 1.6 μM, the curve should be similar as Fig. 4.4C). Compared to the G/U and the T/I base pairs, the K_m value for the C/X base pair was

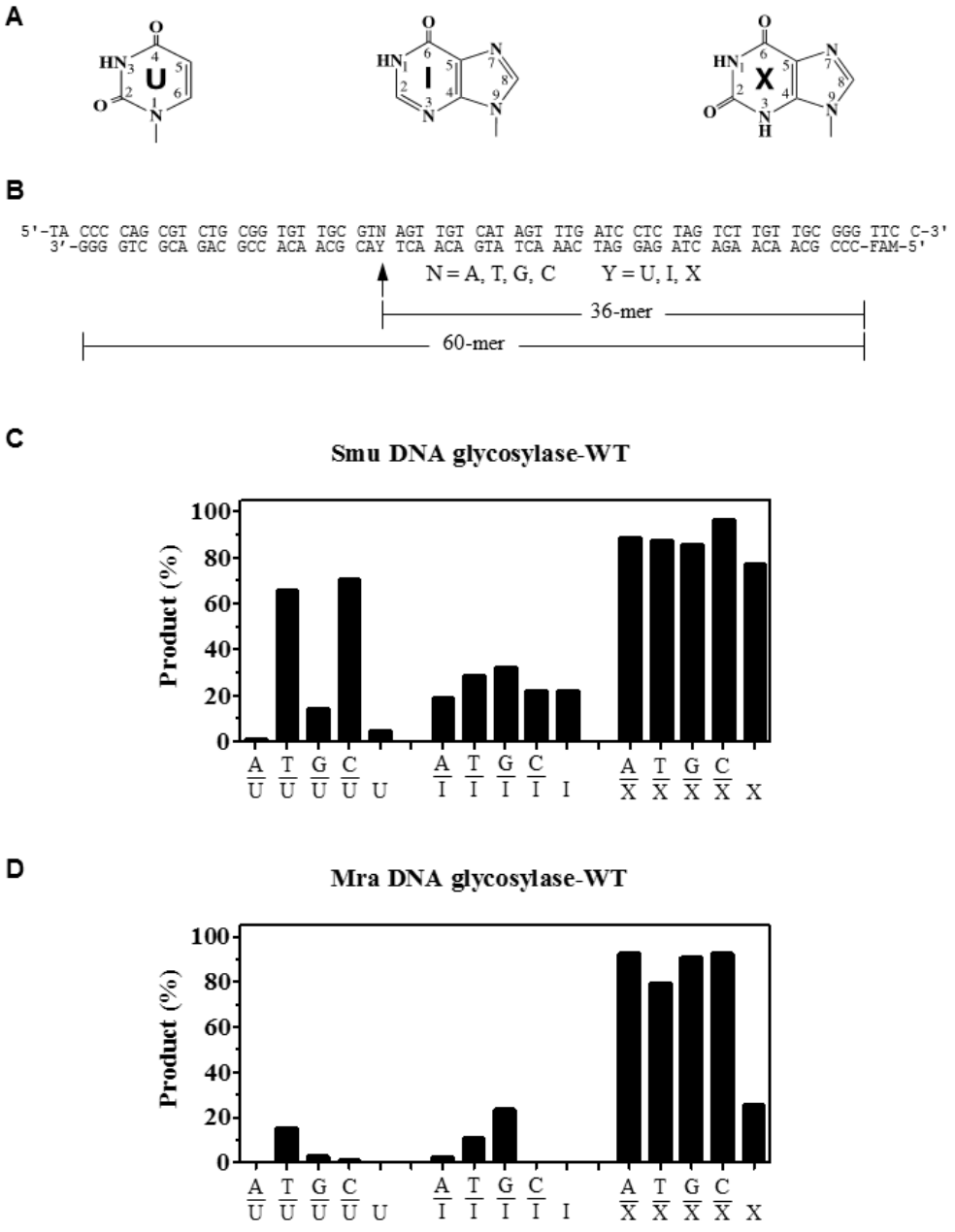


Figure 4.3 Deaminated DNA repair activity in Smu DNA glycosyalse and Mra DNA glycosyalse. A. chemical structures of deaminated DNA bases. **B.** sequences of uracil (U)-, hypoxanthine (I)- and xanthine (X)-containing oligodeoxyribonucleotide substrates. *FAM*, fluorophore. **C.** DNA glycosylase activity of Smu DNA glycosyalse on U-, I- and X-containing substrates. **D.** DNA glycosylase activity of Mra DNA glycosyalse on U-, I- and X-containing substrates. Cleavage reactions were performed as described under "Experimental Procedures" with 100 nM enzyme and 10 nM substrate.

at least several-fold smaller and the k_2 value was sixteen-fold and five-fold greater, resulting in a k_2/K_m value that was two orders and one order of magnitude greater than the G/U base pair and the T/I base pair, respectively (Table 4.2).

Table 4.1 Glycosylase activitiy of Smu DNA glycosylase on deaminated substrates[a]

Smu UDG	A/U	T/U	G/U	C/U	/U	A/I	T/I	G/I	C/I	/I	A/X	T/X	G/X	C/X	/X
Wild type	1	66	14	71	5	19	29	32	22	22	89	88	86	96	77
Q42A	0	0	0	0	0	0	0	0	0	0	8	9	10	32	3
Q42E	0	0	0	0	0	0	0	0	0	0	16	24	18	59	0
Q42I	0	0	0	0	0	0	0	0	0	0	74	73	76	84	18
A43D	0	0	0	0	0	0	0	0	0	0	9	8	9	16	5
A43N	0	7	0	15	0	0	0	0	0	0	19	24	27	53	7
W55A	0	0	0	0	0	0	0	0	0	0	7	8	7	15	5
W55F	16	74	50	88	0	9	34	33	14	0	97	97	97	98	97
D57A	0	16	5	26	0	0	0	2	0	0	16	12	17	3	3
P84A	15	37	26	49	25	12	20	28	18	22	89	78	87	85	62
P84N	11	69	32	84	47	2	5	6	1	3	64	72	60	79	13
D97A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H161A	0	0	0	0	0	0	0	0	0	0	8	9	3	49	0

[a] The reactions were performed as described in Materials and Methods. Numbers are percentage of product signal over total signal. Data are an average of at least two independent experiments. Enzyme concentration is 100 nM, the substrate concentration is 10 nM.

Mutagenesis analysis of Smu UDG

To identify the amino acid residues that are important for enzymatic activity in the Smu DNA glycosylase, we took the advantage of the available crystal structure (PDB code 3IKB). Although there is no substrate or product within the protein structure, the highly conserved structural fold in the UDG superfamily helped us locate the active site of the

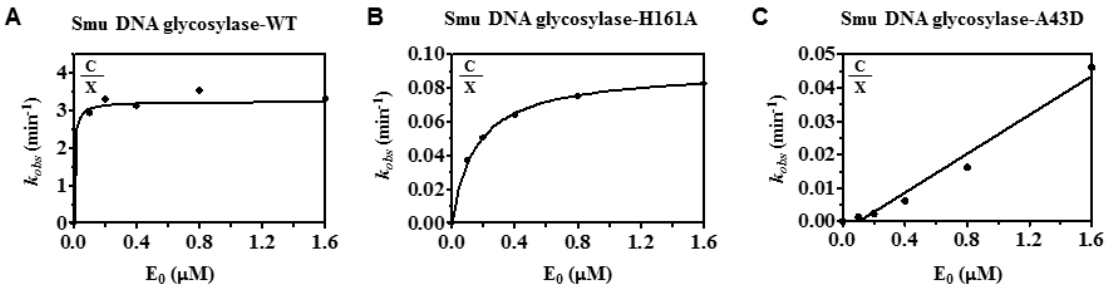


Figure 4.4 Representative kinetics analysis of the wild type and mutant Smu DNA glycosylase. See Materials and Methods for details. **A.** Smu DNA glycosylase-WT. **B.** Smu DNA glycosylase-H161A. C. Smu DNA glycosylase-A43D.

Smu DNA glycosylase. A modeled structure of the enzyme with xanthine-containing DNA was constructed. This modeled structure defined a putative damaged base binding pocket and hinted at potential critical catalytic residues (Fig. 4.2A). To test the roles of these residues for the enzymatic activity, Q42 was substituted with Ala, Glu and Ile; A43 with Asp and Asn; W55 with Ala and Phe; D57 with Ala; P84 with Ala and Asn; D97 with Ala, and H161 with Ala.

The DNA glycosylase activity of these mutants was first examined under the condition that enzyme was in ten-fold excess over substrate (Table 4.1). Afterwards, a more detailed kinetics analysis was performed on Q42A, A43D, W55A, D57A, P84N, D97A and H161A (Table 4.2). For the Q42 position, substitution of Gln with Ala, Glu and Ile significantly reduced the glycosylase activity on all uracil, hypoxanthine and xanthine substrates (Table 4.1). Kinetic analysis showed that substitution of Q42 with Ala caused an increase of K_m to 255 nM, a five-fold decrease of k_2 to 0.61 per min, resulting in a more than thirteen-fold reduction in k_2/K_m value towards the C/X base pair, as compared to wild-type enzyme (Table 4.2). For the A43 position, substitution of Ala with Asp and Asn also caused a dramatic reduction of glycosylase activity (Tables 4.1 and 4.2). A43D completely lost uracil and hypoxanthine activities and only kept very low-level activity on xanthine substrates. Kinetic analysis indicated that the k_2/K_m value for A43D mutant on the C/X base pair was three orders of magnitude lower than that for the wild type enzyme. For the W55 position, substitution of Trp with Ala had a profound effect on the glycosylase activity for all the substrates while W55F still maintained similar level of UDG, HDG and XDG activity compared to wild type (Table 4.1). These results indicate that the aromatic ring that stacks on the deaminated uracil, hypoxanthine and xanthine base is important for the

Table 4.2 Kinetic constants of Smu DNA glycosylase on deaminated substrates[a]

Smu UDG	Substrate	K_m (nM)	k_2 (min^{-1})	k_2/K_m (min^{-1} nM^{-1})
Wild type	G/U	832	2.0×10^{-1}	2.5×10^{-4}
P84N		255	2.8	1.1×10^{-2}
Wild type	T/I	629	6.4×10^{-1}	1.0×10^{-3}
Wild type	C/X	<100	3.2	$>3.2 \times 10^{-2}$
Q42A		255	6.1×10^{-1}	2.4×10^{-3}
A43D		N.D.[b]	N.D.	2.9×10^{-5}
W55A		890	1.0×10^{-1}	1.1×10^{-4}
D57A		118	9.6×10^{-3}	8.1×10^{-5}
P84N		<100	1.7×10^{-1}	$>1.7 \times 10^{-3}$
D97A		N.A.[c]	N.A.	N.A.
H161A		154	9.0×10^{-2}	5.8×10^{-4}

[a] The reactions were performed as described in Materials and Methods. Data are an average of at least two independent experiments.
[b] N.D.: K_m and k_2 is not determined due to a relative large K_m.
[c] N.A.: No activities are detected under all available conditions.

glycosylase activity. Correspondingly, the kinetic analysis showed a large increase in K_m and thirty-two-fold decrease in k_2, resulting in a two orders of magnitude lower k_2/K_m value as compared with the wild type enzyme (Table 4.2). D57A mutant still retained some activities on uracil substrates but the HDG and XDG activity was significantly reduced (Table 4.1). Kinetic analysis revealed that D57A mutation had a profound effect on k_2, causing a four-hundred-fold reduction on k_2/K_m (Table 4.2). For the P84 position, P84A mutant still retained significant glycosylase activity. However, P84N mutant showed an interesting differentiating effect on pyrimidine and purine deaminated bases. Whereas the HDG and XDG activity was reduced, the UDG activity was increased (Table 4.1). The increase in the UDG activity on the G/U base pair can be attributed to a three-fold reduction in K_m and fourteen-fold increase in k_2, resulting in a forty-four-fold increase in k_2/K_m (Table 4.2). The most dramatic change occurred at D97 position as the D97A substitution

completely eliminated all DNA glycosylase activity (Table 4.1). No glycosylase activity was detected even when the enzyme concentration was increased by ten-fold (data not shown). At the H161 position, H161A mutation reduced all glycosylase activity with the UDG, and HDG activity was not detectable under the assay conditions (Table 4.1). The k_2 value and k_2/K_m value for the C/X base pair were reduced to thirty-fold and forty-five-fold, respectively (Table 4.2).

V. Discussion

Deaminated base binding pocket

Many DNA repair enzymes including enzymes in the UDG superfamily use a base flipping mechanism to recognize damaged DNA bases. The broad deaminated base repair activity in Smu DNA glycosylase indicates that base binding pocket of the enzyme can accommodate uracil, hypoxanthine and xanthine. Molecular modeling coupled with mutational and enzyme kinetic analyses allow us to define how different bases are accommodated in the binding pocket. Similar to other family enzymes in the superfamily, Smu DNA glycosylase adopts several key structural elements in motifs 1, 2 and 3 to bind to deaminated bases (Fig 4.5A, 4.5C and 4.5D).

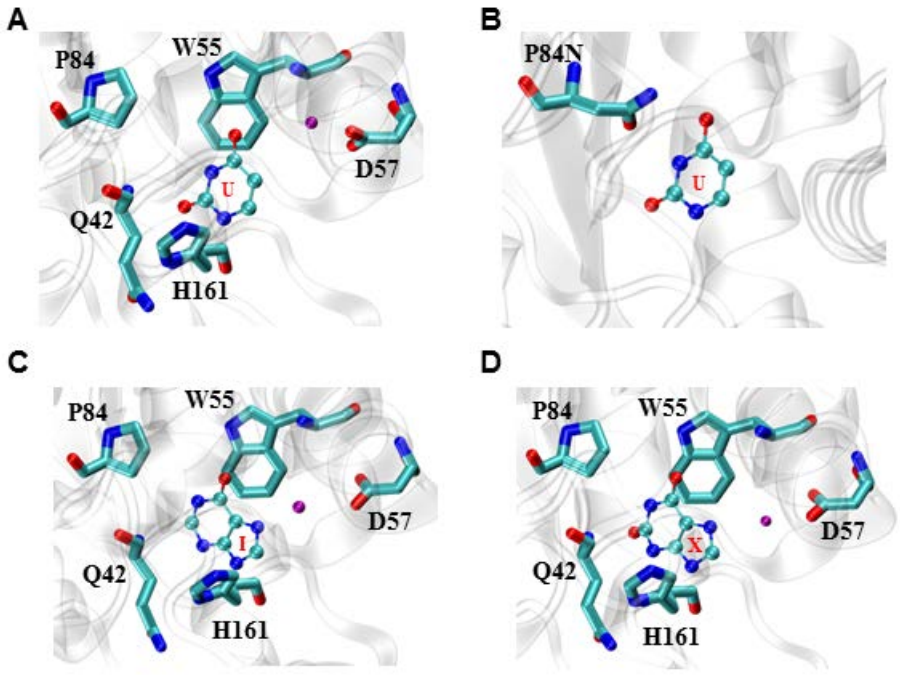


Figure 4.5 Modeled structure of Smu DNA glycosylase wild type and P84N with U, I, and X. The protein structure is shown in the background in white. U, I and X is colored by atom type. Amino acid residues in the active site of Smu DNA glycosylase are shown in licorice in program VMD. **A.** Smu DNA glycosylase wild type with U. **B.** Smu DNA glycosylase P84N with U. **C.** Smu DNA glycosylase wild type with I. **D.** Smu DNA glycosylase with X.

For a uracil base, interactions to O2 of uracil are provided by Q42 in motif 1 and H161 in motif 2 (Fig. 4.5A). However, the typical interactions with N3 and O4 of uracil provided by Asn residue in motif 3 is occupied by P84 in Smu DNA glycosylase (Fig. 4.2B and 4.5A). In family 1 UNG, the sidechain of this Asn residue in motif 3 forms bidentate hydrogen bonds to N3 and O4 of uracil, locking the uracil in the binding pocket (20-22). The lack of the interactions may underlie the relatively higher K_m value for the G/U base pair as compared with the T/I and the C/X base pairs (Table 4.2). On the other hand, when the P84 is substituted by an Asn residue, we indeed observed significant gain of UDG activity while the HDG and XDG activity declined (Tables 4.1 and 4.2). Molecular modeling analysis suggests that the P84N mutant enabled the Smu DNA glycosylase to form the bidentate hydrogen bonds with N3 and O4 of uracil (Fig. 4.5B). As a result, the K_m value for the G/U base pair was reduced by three-fold and the k_2 value was increased by fourteen-fold. In a recent study, we showed that the K68N substitution in family 2 E. coli MUG substantially enhances the UDG activity by reducing the K_m and greatly increasing the single turnover rate constant (manuscript under review). Similar phenomenon is also seen in family 5 Tth UDGb when the A111 in the equivalent position is replaced with an Asn. These results suggest that enzymes in different families in the UDG superfamily use this common mechanism to enhance the UDG activity.

For the hypoxanthine base, the N3 interaction is provided by the mainchain of Q42 and the sidechain of H161 (Fig. 4.5C). In contrast, the mainchain of Q42 and the sidechain of H161 might form hydrogen bonds with O2 of xanthine instead of N3 of xanthine (Fig. 4.5D). The N7 moiety in the purines is an important contact for recognition and catalysis (6,7,10,23). In family 2 E. coli MUG, the contact to N7 of xanthine is provided by S23 in

motif 2 (6). In family 3 SMUG1 from *Geobacter metallireducens* (Gme), the same contact is provided by the mainchain of M64 (7). In family 5 Tth UDGb, the N7 interaction appears to be mediated by a bridging water molecule that is coordinated by a protonated sidechain of D75 and other residues (10). In the modeled Smu DNA glycosylase structure complexed with hypoxanthine and xanthine, D57 is within the distance to interact with N7 of hypoxanthine and xanthine by a bridging water molecule (5.77 Å and 5.88 Å, respectively) (Fig. 4.5C and 4.5D). Interestingly, these amino acid residues that interact with N7 of a purine base are located in different secondary structures in motif 1 (Fig. 4.6). S23 in family 2 E. coli MUG and M64 in family 3 Gme SMUG1 are located in the first helix in the motif 1 (Fig. 4.6A-4.6B). On the other hand, D75 in family 5 Tth UDGb and D57 in Smu DNA glycosylase are located in the second helix, presumably interacting with the N7 via a bridging water (Fig. 4.6C-4.6D). These arrangements suggest that there are multiple ways to interact with the N7 within the secondary structural framework in motif 1.

Regardless of the deaminated bases, loss of base stacking by mutations in W55 would significantly compromise the glycosylase activity by affecting the K_m and k_2 (Tables 4.1 and 4.2). In several family 1 UNG crystal structures solved to date, the highly conserved Phe residue can stack with the uracil base parallel (24-27). This face to face stacking bents the N-glycoside bond nearly 90° in order to bring the O2 of uracil to hydrogen bond distance with the highly conserved His residue in motif 2 (H161 in Smu DNA glycosylase) for catalysis. The highly conserved nature of this aromatic residue across different families suggests that this is a common approach to facilitate base recognition and catalysis in the UDG superfamily.

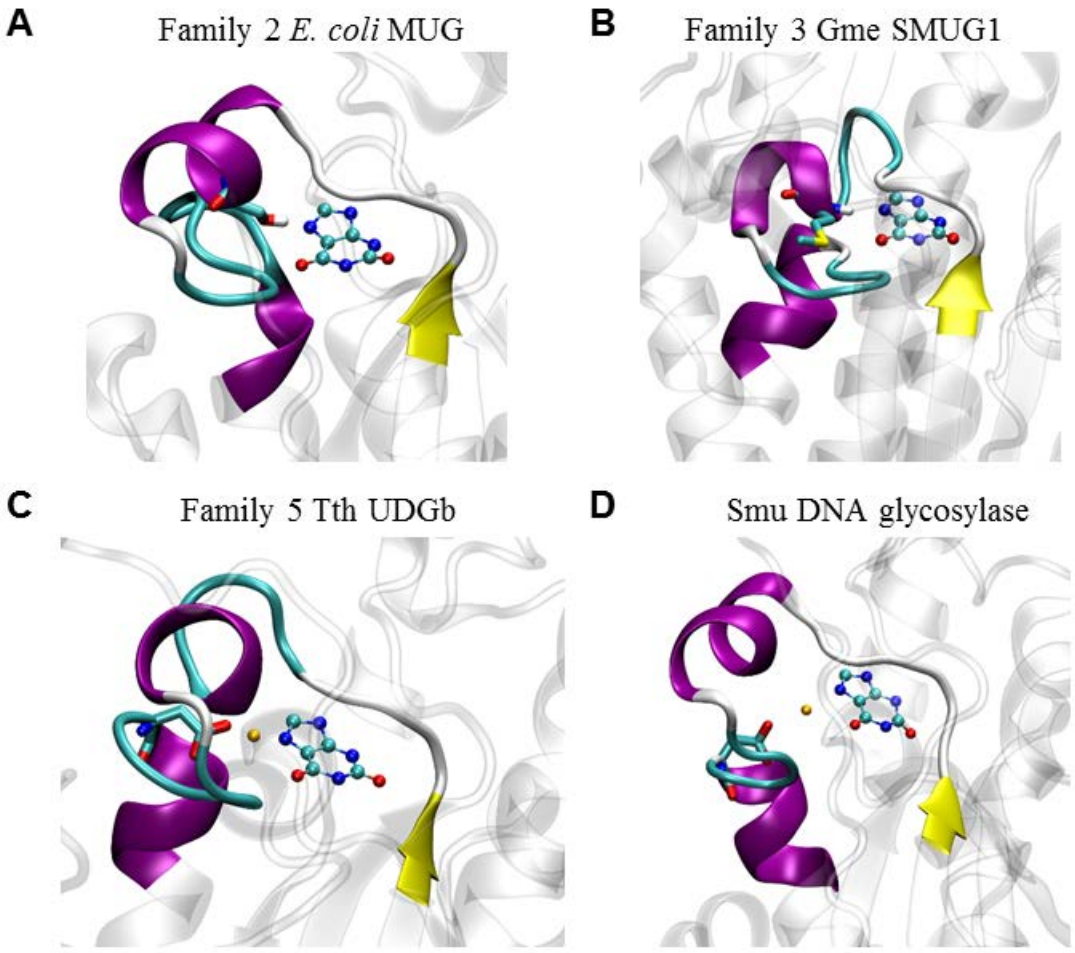


Figure 4.6 Comparison of interactions of UDG families with N7 of xanthine. **A**. Modeled family 2 Eco MUG structure (PDB: 1MUG) complexed with xanthine. **B**. Modeled family 3 Gme SMUG1 structure complexed with xanthine. **C**. Modeled family 5 Tth UDGb structure complexed with xanthine (PDB: 2DEM). **D**. Modeled Smu DNA glycosylase structure (PDB: 3IKB) complexed with xanthine.

Catalytic mechanism

A common theme in the hydrolysis of N-glycosidic bonds is that an oxacarbenium ion intermediate is formed in the transition state (28,29). The catalytic power to accelerate the hydrolysis reaction comes from a combination of activation of the leaving group, stabilization of the oxacarbenium ion, and activation of water as a nucleophile. To remove uracil from DNA, UDG relies more on O2 leaving group activation (30-34). To remove

purine from DNA, enzymes relies on N3, O2 and N7 leaving group activation and activation of water attack (10,35-39). Smu DNA glycosylase appears to use multiple structural elements to catalyze the hydrolysis of the glycosidic bond. As described above, Q42 and H161 contribute to catalysis by interacting with the O2 of uracil, N3 of hypoxanthine, and O2 of xanthine (Fig. 4.7). These interactions promote the formation of oxacarbenium ion intermediate by stabilizing the deaminated base leaving group (Fig. 4.5 and Fig. 4.7). Molecular dynamics analysis was carried out with uracil-, hypoxanthine- and xanthine-containing DNA (Fig. 4.8). The interactions between the O2 of uracil and Q42 and H161 of Smu DNA glycosylase appeared weak, with an average distance of 4.94 Å and 3.57 Å, respectively (Fig. 4.8A). These weak interactions may account for the low UDG activity (Table 4.2). Likewise, the weak hydrogen bonds to the N3 of hypoxanthine explains the relatively low HDG activity (Fig. 4.8B and Table 4.2). On the other hand, the MD analysis suggests that the mainchain of Q42 and the sidechain of H161 of Smu DNA glycosylase is capable of forming strong hydrogen bonds with the O2 of xanthine with an average distance of 2.79 Å and 2.95 Å, respectively (Fig. 4.8C). These short distance

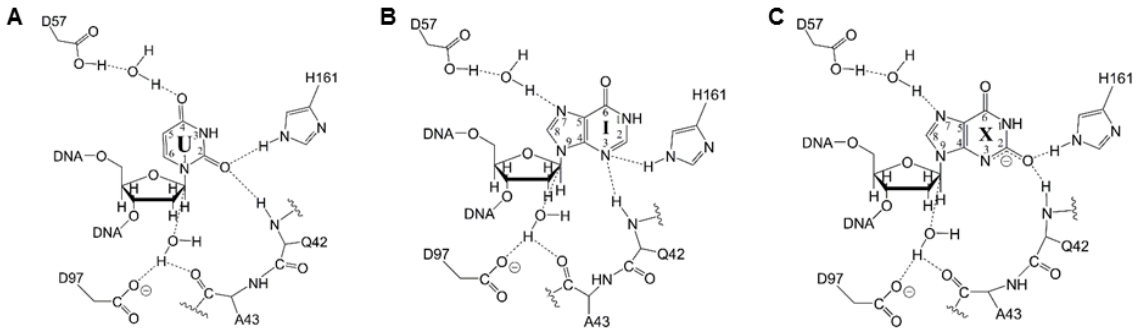


Figure 4.7 Chemical illustration of interaction between Smu DNA glycosylase and deaminated bases. A. chemical illustration of Smu DNA glycosylase -U interactions. **B.** chemical illustration of Smu DNA glycosylase -I interactions. **C.** chemical illustration of Smu DNA glycosylase -X interactions.

hydrogen bonds could promote the departure of the xanthine leaving group, contributing to a robust XDG activity (Table 4.2).

Protonation of N7 of purine has been suggested as a general catalytic mechanism to remove damaged purine base from DNA (10,35-39). As mentioned above, D57 is proposed to play this important role (Fig. 4.7). As shown in the modeled structures and the MD analysis, D57 can interact with O4 of uracil, N7 of hypoxanthine and N7 of xanthine through a bridging water molecule (Fig. 4.6A, 4.6C-D and Fig. 4.8D-F). Kinetics analysis indicates that D75A substitution has a large effect on k_2 towards C/X substrates (Table 4.2). These results underscore the role of N7 interaction in catalysis.

Activation of water molecule is another important step to remove the deaminated DNA base by UDG (21,30,33). It is suggested that different UDG families adopts different

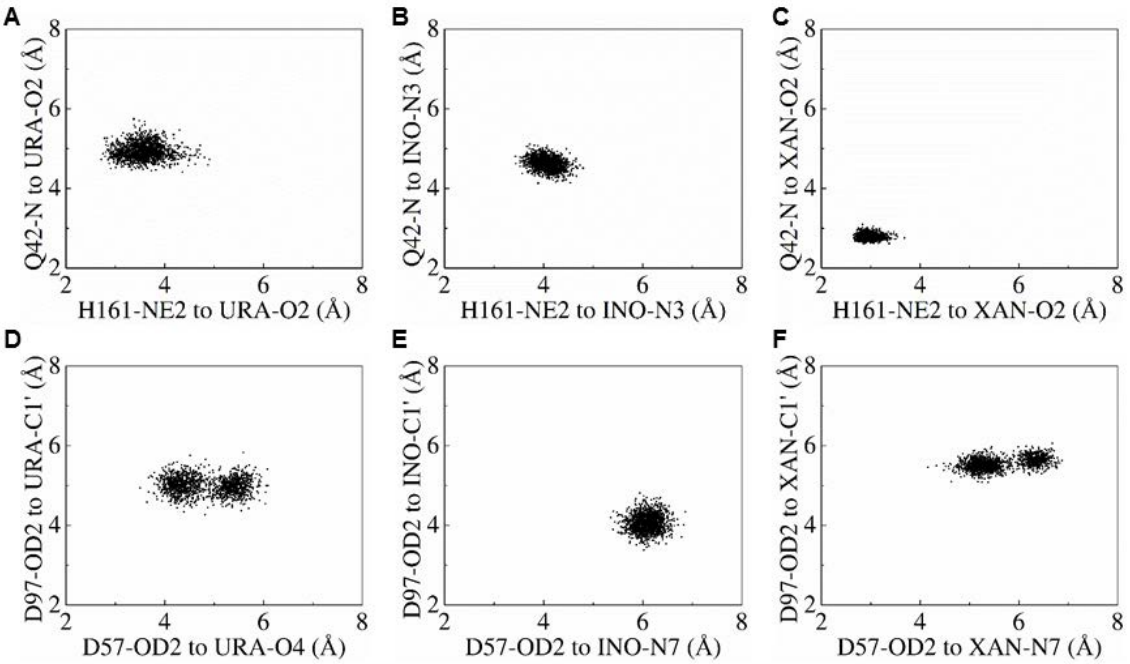


Figure 4.8 Two-dimensional scatter plots of heavy atom distances in the active site of enzyme-DNA complexes obtained from MD trajectories. **A.** distances for interactions between Q42, H161 and uridine. **B.** distances for interactions between Q42, H161 and inosine. **C**. distances for interactions between Q42, H161 and xanthosine. **D**. distances for interactions between D97, D57 and uridine. **E**. distances for interactions between D97, D57 and inosine. **F**. distances for interactions between D97, D57 and xanthosine.

ways to activate water molecule (10). D64 in family 1 E. coli UNG, N18 in family 2 E. coli MUG and N58 in family 3 Gme SMUG1 are important for catalysis by activating a water molecule for the nucleophilic attack on the C1' anomeric carbon (7,24,27,30,40,41). On the other hand, the role of water activation/positioning is played by the last Asn residue in motif 3 in families 4 and 5 enzymes (10,42) (also see chapter 3). A particularly interesting

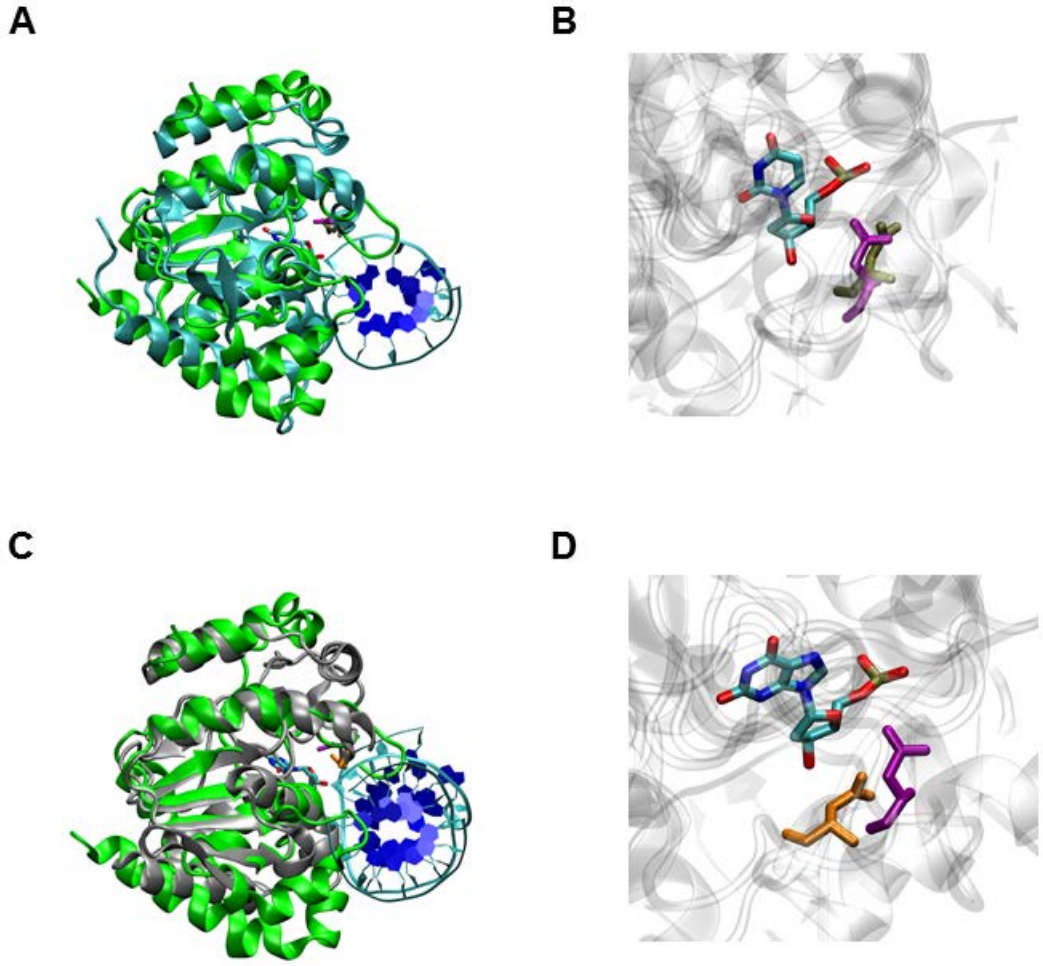


Figure 4.9 Comparison of Smu DNA glycosylase-D97 with Family 4 Tth UDGa-N89 and Family 5 Tth UDGb-N120. **A.** superimposition of modeled Smu DNA glycosylase-U structure (PDB: 3IKB, green, D97 is shown in purple) with Tth UDGa structure (PDB: 1UI0, N89 is shown in tan). **B.** close-up view of Smu DNA glycosylase-D97 and Tth UDGa-N89 using the same coloring scheme described in A. **C.** superimposition of modeled Smu DNA glycosylase-X structure (PDB: 3IKB, green, D97 is shown in purple) with Tth UDGb structure (PDB: 2DEM, N120 is shown in orange). **D.** close-up view of Smu DNA glycosylase-D97 and Tth UDGb-N120 using the same coloring scheme described in C.

structural element that makes a profound contribution to the hydrolysis of the N-glycosidic bond in Smu DNA glycosylase is D97 in motif 3. This residue is located on the opposite site of the deaminated base below the deoxyribose (Fig. 4.5). Superimposition with family 4 Tth UDGa and family 5 Tth UDGb indicates that D97 can play a similar role in water activation/positioning (Fig. 4.9). A distinct difference is that the mutational effect by substituting D97 in Smu DNA glycosylase is more severe as D97A becomes completely inactive in UDG, HDG and XDG activity (Tables 4.1-4.2). Besides, mainchain oxygen of A43 also showed a close localization towards the N-glycosidic bond and side chain of D97 (data not shown). Considering the dramatic effect caused by mutating A43 (Table 4.1 and table 4.2), this residue might be able to aid D97 to coordinate water molecule as similar as the A59 in family 5 Tth UDGb (the equivalent position of A43 in Smu DNA glycosylase) (42). These results suggest that water activation/positioning by D97 and A43 plays a critical role in catalysis. Thus, combined with leaving group activation provided by Q42, H161 and D57, Smu DNA glycosylase appears to rely on a S_N2-type mechanism for the hydrolysis of the N-glycosidic bond especially for a deaminated purine base (Fig. 4.7)

A new family

Smu DNA glycosylase represent a new class of enzymes that is structurally homologous to families in the UDG superfamily but presents its distinct DNA repair and catalytic features. To understand its evolutionary relationship with other families, a phylogenetic analysis was performed. As shown in Fig. 4.10, this new class of enzymes establishes a distinct group within the UDG superfamily. We designate this new class of enzymes as family 7 XDG. As such, the UDG superfamily now has seven families with

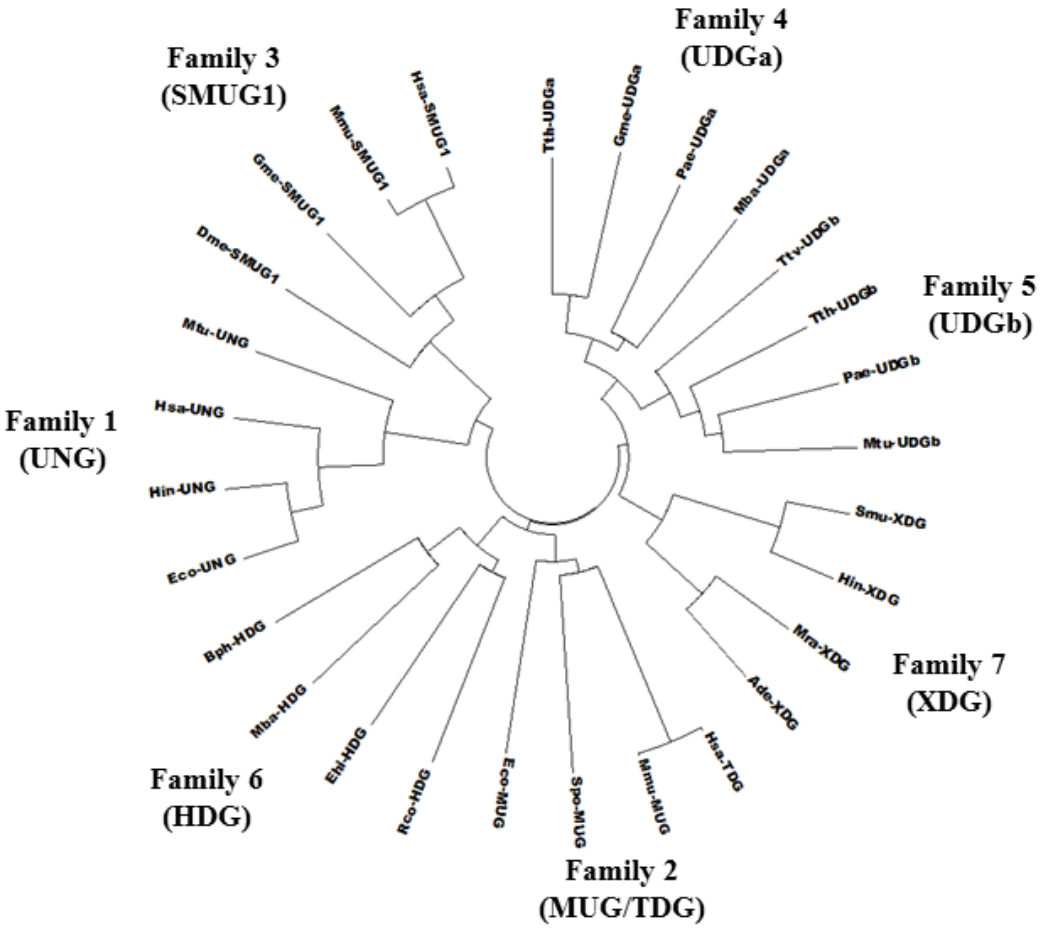


Figure 4.10 Phylogenetic analysis of UDG superfamily. Family 7 (XDG): *Ade, Aquamicrobium defluvii*, EXL01735.1; *Hin, Haemophilus influenza*, WP_005694076.1; *Mra, Methylobacterium radiotolerans* JCM 2831, YP_001754430.1; *Smu, Streptococcus mutans* UA159, NP_721617.1. Family 1 (UDG): *Eco, E. coli*, NP_289138; *Hin*, *Haemophilus influenzae* KR494, YP_008544610.1; *Mtu, Mycobacterium tuberculosis*, WP_003908950.1; *Hsa*, *H. sapiens*, NP_003353. Family 2 (MUG/TDG): *Eco*, *E. coli*, P0A9H1; *Spo, S. pombe*, O59825; *Hsa*, *H. sapiens*, NP_003202; *Mmu, Mus musculus*, XP_003945901.1. Family 3 (SMUG1): *Gme, G. metallireducens* GS-15,YP_383069; *Hsa, Homo sapiens*, NP_055126; *Mmu, Mus musculus*, NP_082161; *Dme, Drosophila melanogaster*, NP_650609.1. Family 4 (UDGa): *Tth*, *T. thermophilus* HB27, YP_004341.1. ; *Pae*, *P. aerophilum str*.IM2, NP_558739.1; *Gme, Geobacter metallireducens* GS-15, YP_006721625.1; *Mba, Methanosarcina barkeri str. Fusaro*, YP_305330.1. Family 5 (UDGb): *Tth*, *T. thermophilus* HB8, YP_144415.1; *Pae*, *P. aerophilum* str. IM2, NP_559226; *Tvo*, *Thermoplasma volcanium* GSS1, NP_111346.1; *Mtu*, *M. tuberculosis* H37Rv, P64785 (Rv1259). Family 6 (HDG): *Bph, Burkholderia phymatum* STM815, YP_001858334.1; *Mba, Methanosarcina barkeri str. Fusaro*, YP_304295.1; *Rco, Ricinus communis*, XP_002536323.1; *Ehi, Entamoeba histolytica* HM-1:IMSS, XP_655177.1.

diverse DNA repair activities. This study illustrates the evolutionary potential of the glycosylases in the UDG superfamily to offer functional and catalytic diversity.

VI. Reference

1. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715
2. Lindahl, T., and Wood, R. D. (1999) Quality control by DNA repair. *Science* **286**, 1897-1905
3. Lee, H. W., Dominy, B. N., and Cao, W. (2011) New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *The Journal of biological chemistry* **286**, 31282-31287
4. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 3649-3653
5. Huffman, J. L., Sundheim, O., and Tainer, J. A. (2005) DNA base damage recognition and removal: new twists and grooves. *Mutation research* **577**, 55-76
6. Lee, H. W., Brice, A. R., Wright, C. B., Dominy, B. N., and Cao, W. (2010) Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase. *The Journal of biological chemistry* **285**, 41483-41490
7. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N., and Cao, W. (2009) Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: switching SMUG1 to UDG. *Journal of molecular biology* **385**, 761-778

8. Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S., and Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. *Journal of molecular biology* **333**, 515-526

9. Sandigursky, M., and Franklin, W. A. (1999) Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes. *Current biology : CB* **9**, 531-534

10. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N., and Cao, W. (2014) Specificity and Catalytic Mechanism in Family 5 Uracil DNA Glycosylase. *The Journal of biological chemistry* **289**, 18413-18426

11. Lucas-Lledo, J. I., Maddamsetti, R., and Lynch, M. (2011) Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Molecular biology and evolution* **28**, 1307-1317

12. King, K., Benkovic, S. J., and Modrich, P. (1989) Glu-111 is required for activation of the DNA cleavage center of EcoRI endonuclease. *The Journal of biological chemistry* **264**, 11807-11815

13. Vermote, C. L., and Halford, S. E. (1992) EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition. *Biochemistry* **31**, 6082-6089

14. Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S., and Masui, R. (2007) Crystal Structure of Family 5 Uracil-DNA Glycosylase Bound to DNA. *Journal of molecular biology* **373**, 839-850

15. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *ELECTROPHORESIS* **18**, 2714-2723

16. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545-1614

17. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802

18. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N., and Cao, W. (2014) Specificity and Catalytic Mechanism in Family 5 Uracil DNA Glycosylase. *Journal of Biological Chemistry* **289**, 18413-18426

19. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38

20. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5083-5088

21. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87-92

22. Wuenschell, G. E., O'Connor, T. R., and Termini, J. (2003) Stability, miscoding potential, and repair of 2'-deoxyxanthosine in DNA: implications for nitric oxide-induced mutagenesis. *Biochemistry* **42**, 3608-3616

23. Dong, L., Mi, R., Glass, R. A., Barry, J. N., and Cao, W. (2008) Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase. *DNA repair* **7**, 1962-1972

24. Parikh, S. S., Putnam, C. D., and Tainer, J. A. (2000) Lessons learned from structural results on uracil-DNA glycosylase. *Mutation research* **460**, 183-199

25. Bianchet, M. A., Seiple, L. A., Jiang, Y. L., Ichikawa, Y., Amzel, L. M., and Stivers, J. T. (2003) Electrostatic guidance of glycosyl cation migration along the reaction coordinate of uracil DNA glycosylase. *Biochemistry* **42**, 12455-12460

26. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *The EMBO journal* **17**, 5214-5226

27. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92**, 117-129

28. Berti, P. J., and McCann, J. A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chemical reviews* **106**, 506-555

29. Stivers, J. T., and Jiang, Y. L. (2003) A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chemical reviews* **103**, 2729-2759

30. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-493

31. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80**, 869-878

32. Shroyer, M. J., Bennett, S. E., Putnam, C. D., Tainer, J. A., and Mosbaugh, D. W. (1999) Mutation of an active site residue in Escherichia coli uracil-DNA glycosylase: effect on DNA binding, uracil inhibition and catalysis. *Biochemistry* **38**, 4834-4845

33. Drohat, A. C., Jagadeesh, J., Ferguson, E., and Stivers, J. T. (1999) Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry* **38**, 11866-11875

34. Drohat, A. C., and Stivers, J. T. (2000) Escherichia coli uracil DNA glycosylase: NMR characterization of the short hydrogen bond from His187 to uracil O2. *Biochemistry* **39**, 11865-11875

35. Zoltewicz, J. A., Clark, D. F., Sharpless, T. W., and Grahe, G. (1970) Kinetics and mechanism of the acid-catalyzed hydrolysis of some purine nucleosides. *Journal of the American Chemical Society* **92**, 1741-1749

36. McCann, J. A., and Berti, P. J. (2008) Transition-state analysis of the DNA repair enzyme MutY. *Journal of the American Chemical Society* **130**, 5789-5797

37. Michelson, A. Z., Rozenberg, A., Tian, Y., Sun, X., Davis, J., Francis, A. W., O'Shea, V. L., Halasyam, M., Manlove, A. H., David, S. S., and Lee, J. K. (2012) Gas-phase studies of substrates for the DNA mismatch repair enzyme MutY. *Journal of the American Chemical Society* **134**, 19839-19850

38. Brunk, E., Arey, J. S., and Rothlisberger, U. (2012) Role of environment for catalysis of the DNA repair enzyme MutY. *Journal of the American Chemical Society* **134**, 8608-8616

39. Francis, A. W., Helquist, S. A., Kool, E. T., and David, S. S. (2003) Probing the requirements for recognition and catalysis in Fpg and MutY with nonpolar adenine isosteres. *Journal of the American Chemical Society* **125**, 16235-16242

40. Drohat, A. C., Xiao, G., Tordova, M., Jagadeesh, J., Pankiewicz, K. W., Watanabe, K. A., Gilliland, G. L., and Stivers, J. T. (1999) Heteronuclear NMR and crystallographic studies of wild-type and H187Q Escherichia coli uracil DNA glycosylase: electrophilic catalysis of uracil expulsion by a neutral histidine 187. *Biochemistry* **38**, 11876-11886

41. Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *The EMBO journal* **18**, 6599-6609

42. Kosaka, H., Hoseki, J., Nakagawa, N., Kuramitsu, S., and Masui, R. (2007) Crystal structure of family 5 uracil-DNA glycosylase bound to DNA. *Journal of molecular biology* **373**, 839-850

CHAPTER FIVE

RESEARCH SIGNIFICANCE AND CONCLUDING REMARKS

Cells are constantly attacked by endogenous and exogenous factors (1-3).To maintain the genome integrity, BER repair pathway is essential to remove the deaminated bases from DNA (2-5). As the enzyme initiating the BER repair pathway, members in UDG superfamily play a critical role in deaminated DNA base recognition and removal (5-9). In this study, the specificities of several enzymes from different UDG families are investigated and their catalytic mechanisms are illustrated.

Family 5 UDGb was first identified as a UDG and HDG (10,11). It can play an anti-mutation role in several organisms (12). In this work, we show that family 5 UDGb from *Thermus thermophilus* is also a xanthine DNA glycosyalse. In addition, several amino acids are identified to be important for its enzymatic activities. Although its UDG activity mainly depends on the leaving group activation mediated by O2 of uracil and H190, family 5 Tth UDGb combines leaving group activation by H190 and D75 with a nucleophile activation by N120 to expand its DNA repair capacity to deaminated purine bases. As such, this work on family 5 UDGb illustrate a specific catalytic strategy adopted by UDG to broaden its specificity (13).

In this study, family 4 UDGa has been identified to be an exclusive UDG. Consistent with previous studies, family 4 UDGa from *Thermus thermophilus* possesses robust activities on all uracil containing substrates (8). Further enzymatic analysis and structural modeling analysis identified a specific binding pocket for uracil. A catalytic mechanism including leaving group activation by interaction between O2 of uracil and

H155 and nucleophile activation by N89 is proposed. Most importantly, the concurrent change of E41Q and G42D is found to be able to restore the enzymatic activities compared to single mutants, suggesting that these two positions are structurally and functionally correlated. This study provide insights into how family 4 UDGa specifically recognizes and removes uracil from DNA.

Based on sequence and structural similarity, we found a putative new UDG family from *Streptococcus mutans* and *Methylobacterium radiotolerans* with robust activities on xanthine-containing DNA substrates. Detailed kinetic and structural analysis revealed that enzymes from *Streptococcus mutans* can specifically recognize xanthine by interacting O2 of xanthine with H161 and Q42. The structural comparison of family 5 UDGb with this putative family UDG suggests that a similar nucleophile activation approach is adopted by these two different group of enzymes. Further structural comparison within UDG superfamily suggests that several UDG families utilize a similar strategy to interact with N7 of deaminated purine. This study shed on light how different UDG families broaden their specificity to purine by using a similar catalytic mechanism.

In summary, different families in UDG superfamily adopt different residues to initiate leaving group activation and water activation. In the future, one interesting question is whether this difference is the cause of different enzyme activities towards different deaminated DNA bases. For example, enzymes from family 2 TDG/MUG utilize a highly conserved asparagine to activate water molecule within the motif 1 of UDG superfamily. In contrast, enzymes from family 6 HDG use a highly conserved aspartic acid within the motif 3 of UDG superfamily. Considering the dramatic different enzyme activities between family 2 enzymes and family 6 enzymes, it would be interesting to substitute the catalytic

residue of family 6 enzymes with the catalytic residue of family 2 enzymes, to see whether this alternation would expand substrate specificity of family 6 enzymes. Another interesting topic about UDG superfamily is whether there are more residues correlated with each other. In chapter 3, we found E41 and G42 are highly correlated within family 4 Tth UDGa. In the future, we hope to use multiple bioinformatics tools combined with our current biochemical system to identify more correlated residues within the UDG superfamily.

Reference

1. Duncan, B. K., and Miller, J. H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560-561
2. Myles, G. M., and Sancar, A. (1989) DNA repair. *Chemical research in toxicology* **2**, 197-226
3. Ames, B. N., and Gold, L. S. (1991) Endogenous mutagens and the causes of aging and cancer. *Mutation research* **250**, 3-16
4. Sancar, A. (1996) DNA excision repair. *Annual review of biochemistry* **65**, 43-81
5. Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B., and Sperens, B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *The Journal of biological chemistry* **252**, 3286-3294
6. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-493
7. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92**, 117-129
8. Hoseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S., and Kuramitsu, S. (2003) Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. *Journal of molecular biology* **333**, 515-526
9. Berti, P. J., and McCann, J. A. (2006) Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chemical reviews* **106**, 506-555

10. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H., and Jiricny, J. (2002) A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *The EMBO journal* **21**, 3182-3191

11. Starkuviene, V., and Fritz, H. J. (2002) A novel type of uracil-DNA glycosylase mediating repair of hydrolytic DNA damage in the extremely thermophilic eubacterium Thermus thermophilus. *Nucleic acids research* **30**, 2097-2102

12. Sakai, T., Tokishita, S., Mochizuki, K., Motomiya, A., Yamagata, H., and Ohta, T. (2008) Mutagenesis of uracil-DNA glycosylase deficient mutants of the extremely thermophilic eubacterium Thermus thermophilus. *DNA repair* **7**, 663-669

13. Xia, B., Liu, Y., Li, W., Brice, A. R., Dominy, B. N., and Cao, W. (2014) Specificity and Catalytic Mechanism in Family 5 Uracil DNA Glycosylase. *The Journal of biological chemistry* **289**, 18413-18426