12-2008

DNA DEAMINATION REPAIR ENZYMES IN BACTERIAL AND HUMAN SYSTEMS

Rongjuan Mi
Clemson University, rmi@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Part of the Biochemistry Commons

DNA DEAMINATION REPAIR ENZYMES IN BACTERIAL AND HUMAN SYSTEMS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biochemistry

---

by
Rongjuan Mi
December 2008

Accepted by:
Dr. Weiguo Cao, Committee Chair
Dr. Chin-Fu Chen
Dr. James C. Morris
Dr. Gary Powell

ABSTRACT

DNA repair enzymes and pathways are diverse and critical for living cells to maintain correct genetic information. Single-strand-selective monofunctional uracil DNA glycosylase (SMUG1) belongs to Family 3 of the uracil DNA glycosylase superfamily. We report that a bacterial SMUG1 ortholog in *Geobacter metallireducens* (Gme) and the human SMUG1 enzyme are not only uracil DNA glycosylases (UDG), but also xanthine DNA glycosylases (XDG). Mutations at M57 (M57L) and H210 (H210G, H210M, H210N) can cause substantial reductions in XDG and UDG activities. Increased selectivity is achieved in the A214R mutant of Gme SMUG1 and G60Y completely abolishes XDG and UDG activity. Most interestingly, a proline substitution at the G63 position switches the Gme SMUG1 enzyme to an exclusive uracil DNA glycosylase. Mutational analysis and molecular dynamics (MD) simulations of Gme SMUG1 identify important structural determinants in conserved motifs 1 and 2. Our study offers insights on the important role that modulation of conformational flexibility may play in defining specificity and catalytic efficiency.

Endonuclease V is an enzyme that initiates a conserved DNA repair pathway by making an endonucleolytic incision at the 3' side one nucleotide from a deaminated base lesion. This study defines the endonuclease and exonuclease activity in endonuclease V from *Thermotoga maritima* (Tma) in an assay condition with Mn^{2+} as a metal cofactor. Tma endonuclease V exhibits inosine-dependent 3'-exonuclease activity. Detailed kinetic analysis using 3'-labeled DNA indicates that Tma endonuclease V also possesses

nonspecific 5'-exonuclease activity. The multiplicity of the endonuclease and exonuclease activity is discussed with respect to deaminated base repair.

Biochemical properties of human endonuclease V with respect to repair of deaminated base lesions were reported. We determined repair activities of human endonuclease V on inosine (I)-, xanthosine (X)-, oxanosine (O)- and uridine (U)-containing DNA. Human endonuclease V is most active with inosine-containing DNA; however, with minor activity on xanthosine-containing DNA. Mg^{2+} and to a much less extent, Mn^{2+}, Ni^{2+}, Co^{2+} can support the endonuclease activity. Introduction of human endonuclease V into *Escherichia coli* cells caused two-fold reduction in mutation frequency. This is the first report of deaminated base repair activity from human endonuclease V.

DEDICATION

I dedicate my dissertation to my parents. They helped me face all these difficulties and frustrations, and shared my happiness. Their wisdom and unselfish love always encourage me to pick myself up and go further.

Thank you!

ACKNOWLEDGMENTS

In all these years for my Ph.D. study, so many people have helped me. Here I would like to express my gratitude to them.

Firstly, I would like to acknowledge and extend my heartfelt gratitude to my advisor, Dr. Weiguo Cao who offered me this opportunity to study at Clemson University. He supported me and gave me advice through my Ph.D. study. This dissertation could not have been written without Dr. Weiguo Cao. He has special intuition for the projects and moves them in correct directions.

I would like to thank my graduate committee members (Dr. Chin-Fu Chen, Dr. James C. Morris, and Dr. Gary Powell) for their consistent help and advice on my research and career during my Ph.D. study.

I also would like to thank all the people from Cao lab (Dr. Hong Feng, Dr. Thomas Hitchcock, Dr. Honghai Gao, Dr. Liang Dong, Hyun-Wook Lee, and Sung-Hyun Park) who helped me and exchanged ideas with me all the time.

I want to express thanks to our collaborator, Dr. Brian N. Dominy, for his help on molecular modeling data and preparation of the manuscript.

Finally, I would like to thank my parents again for their unselfish support and encouragement in my life.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

ABBREVIATIONS

Alkyladenine DNA glycosylase gene, AAG

Activation-Induced cytidine Deaminase, AID

BER, Base Excision Repair

BSA, bovine serum albumin

DNA, deoxyribonucleic acid

dATP, deoxyadenosine triphosphate

dCTP, deoxycytidine triphosphate

dGTP, deoxyguanosine triphosphate

dITP, deoinosine triphosphate

dOTP, deoxyoxanosine triphosphate

dTTP, deoxythymidine triphosphate

dUTP, deoxyuridine triphosphate

dXTP, deoxyxanthosine triphosphate

DTT, dithiothreitol

Endonuclease V, Endo V

Formamidopyrimidine-DNA glycosylase, Fpg

3-methyladenine-DNA glycosylase II, AlkA

mCpG binding domain protein 4, MBD4

Mismatch-specific Uracil-DNA glycosylase, MUG

Oxanine, Oxa

Reactive nitrogen species, RNS

Abbreviations (Continued)

Reactive oxygen species, ROS

Single strand-selective Monofunctional Uracil-DNA Glycosylase, SMUG1

Somatic hypermutation, SHM

Thymine DNA N-glycosylase, TDG

Uracil DNA Glycosylase, UDG

Xanthine, Xan

CHAPTER ONE

DNA DAMAGE AND REPAIR

1. Introduction

It has been long known that DNA is the carrier of the genetic information which needs to be passed accurately from one generation to the next. Since each cell inherits traits from its parents to maintain the basic characteristics of the family, DNA should be stable enough to conserve and pass this genetic information. DNA is composed of a double helix structure. Its stability is maintained through hydrogen bonds and other noncovalent interactions. Research shows that the hydrodynamic diameter of DNA molecules ranges from 22 to 26 Å in dilute aqueous solutions (*1*).

In nature, many chemicals can damage the DNA bases and subsequently cause mutations. The damage frequency is low *in vivo*; however, a genome is quite large. For example, the human genome is ~ 3.2 Gb (*2*). As a result, mutations generated by damage are considerable, with more than one million damage incidences occurring in DNA per cell every day (*3*). DNA repair enzymes which can recognize and repair this damage play an important role in reducing mutations by removing damaged DNA. To date, various kinds of responses after DNA damage and many kinds of DNA repair enzymes and pathways recognizing and repairing the damage have been studied.

This chapter first introduces the sources of DNA damage, second describes the types of DNA damage, third discusses the mechanisms responsible for DNA repair, and finally presents the consequences of DNA damage

2. Sources of DNA damage

Generally, the sources of DNA damage can be divided into two groups, endogenous and exogenous damage. Previously, exogenous chemicals were thought to be the major causes of DNA damage in cells; however, recent research suggests that the damage caused by endogenous reagents is more prevalent (*4*). Endogenous damage can arise from metabolic byproducts, with numerous endogenous reagents having been discovered (*5, 6*). For example, 4-hydroxynonenal, a reactive oxygen species (ROS) toxic to DNA, is a lipid peroxide product (*7*). Oestrogen, a natural hormone, can induce direct or indirect DNA damage including single-strand breaks and 8-hydroxylation of guanine bases (*8*).

Several types of exogenous damage caused by external chemicals have been studied. It is known that nucleic acid can absorb UV from the sun, damaging DNA forming UV-specific dipyrimidine photoproducts. However, this damage also provides benefits as a driving source of mutation evolution (*9*). It can be recognized by deoxyribodipyrimidine photolyase (*10*) or several other enzyme, such as UV damage endonuclease (*11*). Ionizing radiation (IR), generating hydroxyl (OH) radicals (*12*), is another frequently seen source of DNA damage. Research shows the cells exhibit proliferation delay after exposure to ionizing radiation (*13-15*). A possible reason for this delay is that it takes the cells some time to deal with the damaged DNA after exposure to ionizing radiation (*16, 17*). Several chemicals, inhibiting DNA repair enzymes and increasing instability of the genome, are regarded as mutagens (*18*). DNA oxidation and deamination, frequently observed damage, can be caused by these exogenous reagents. In

2

the mean time, it has been found that many human-made chemicals are harmful and can damage DNA. The disadvantages of using these chemicals become the focus of much research (*19-23*).

3. Types of DNA damage

Many forms of DNA damage have been extensively studied (*24-26*), including base modification, base loss, replication error, DNA cross-linkage, DNA-protein cross-linkage and double-strand breaks. DNA damage caused by deamination and its repair are the focus of my research.

A. Base modification

Four kinds of bases compose genomic DNA, adenine, thymine, guanine, and cytosine. Base modifications, such as cytosine methylation, is widely used in cells to control gene expression (*27*). However, most of these modifications occur can cause mutations and diseases. Commonly observed base modifications, such as alkylation, oxidation, depuination and depyrimidination, and deamination, have been studied intensively.

B. Base deamination

Among the four normal bases in genomic DNA, cytosine and adenine, can be deaminated to uracil and hypoxanthine, respectively, and guanine can be deaminated to

either xanthine or oxanine (*28-30*) (Fig. 1.1). Because thymine does not have an amino group (-NH$_2$), it does not experience deamination as other bases do.
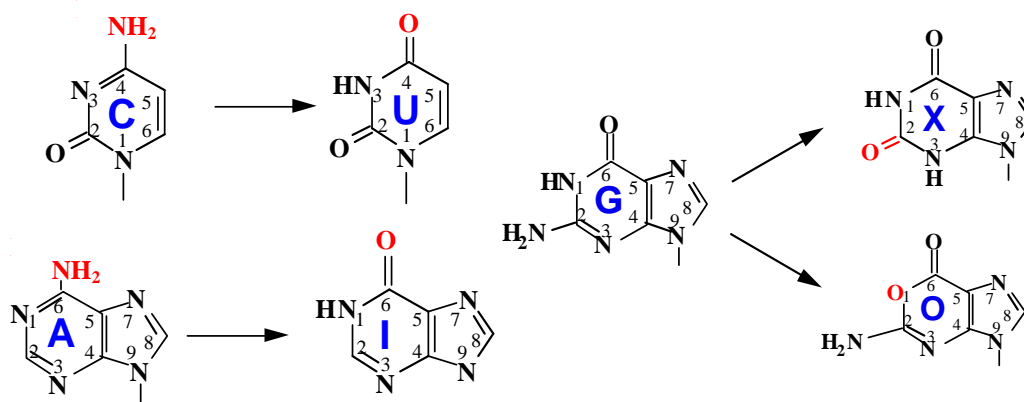
Figure 1.1 Schematic representation of the deamination products of cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine and oxanine. Taken from (*28*)

Uracil can form a stable base pair with adenine. After another cycle of DNA replication, the original C:G base become the A:T base pair. As such, it can introduce the C:G to A:T transition mutation into genomic DNA if uracil is not recognized and repaired in time (*31, 32*). Hypoxanthine, which prefers to be paired with cytosine, can cause the A:T to C:G base pair mutation. Research shows hypoxanthine can also be integrated into the chromosome DNA under certain circumstances, such as in the *rdgB* deletion mutant (*33*), and ITP, in conjunction with T7 polymerase, can be used in RNA extension with (*34*).

Xanthine and oxanine can be paired with thymine and it can cause the G:C to T:A mutation. Xanthine is relatively stable at pH 7.0 at 37°C, its half life in double-stranded DNA being 2.4 years (*35*). Drosophila DNA polymerase alpha has been used to determine which base(s) can be paired with xanthine *in vitro*. The results demonstrated

that the relative nucleotide incorporation rates of xanthine follow this order: $T > G > A = C$ (*36*). However, since *in vivo*, the nucleotide which can be paired with xanthine is determined by replisome, and they may not follow this order.

Reactive nitrogen species (RNS) are known to introduce deamination in DNA bases. Reactive oxygen species (ROS) can also generate deaminations in DNA (*37*). Toyokuni and Spencer originally assumed that oxidative stress primarily produced oxidative damage with a small amount of deamination products. However, recent research has demonstrated that reactive oxygen species can generate large amounts of deaminated bases under certain conditions. This increased ratio of deaminated DNA bases corresponds to the increased ratio of ROS to cells (*37, 38*).

When male Wistar rats were treated with ferric nitrilotriacetate (Fe-NTA), many modified bases including xanthine, 5-hydroxy-5-methylhydantoin, 5-(hydroxymethyl) uracil, 5-hydroxy-cytosine, thymine glycol, 5,6-dihydroxyuraril, 4,6-diamino-5-form-amidopyrimidine, 8-hydroxyadenine, 2-hydroxyadenine and 8-hydroxyguanine were detected in the renal chromatin (*37*). When both calf thymus DNA and human skin epidermal keratinocytes were treated with peroxynitrite, only a small dose of oxidative damage was detected, while large amounts of nitrated and deaminated purine bases, hypoxanthine, xanthine, 8-nitro-guanine were found (*37*). These results indicated that peroxynitrite can generate more deamination damage than oxidation damage in DNA. In addition to OH•, a known radical generating oxidative damage to bases, is not a major source from peroxynitrite stress according to a previous study (*38*).

The profiles of the base deamination products and the intermediates of the process caused by diazonium ions has been determined (*39*) and the deamination rate is very low. For example, the half life of cytosine in single-stranded DNA is approximately 200 years (*40*). However, the events of cytosine deamination are numerous due to the large size of the genome. It has been estimated that approximately 60-500 cytosines can be deaminated in 24 hours in the human body based on the assumption that cytosine forms approximately 20% of the human genome (*41*). Direct evidence was provided that the deamination of cytosine is a significant source for spontaneous mutations (*32*).

Base deamination can also occur when the DNA is treated with UV, ionizing radiation (IR) (*42*) or nitric oxide (*43, 44*). Nitrous acid is widely found in environment and in *in vivo* metabolic processes. It causes deamination by interacting with polynucleotides (*45*).

There are several common reagents which can cause deamination, such as nitrous acid (HNO_2), hydroxylamine ($HONH_2$) and bisulfite (HSO_3^-). Nitrous acid can penetrate into the cell, causing damage, and HSO_3^- reacts with single-stranded DNA. Various bacteria can denitrify nitrate to NO• directly, causing damage stress; however, in *E. coli*, nitrite can be reduced to ammonia (*46*) and nitrosating agents are produced during metabolic processes when oxygen is scarce because the anaerobic pathway allows nitrate and nitrite to be used as electron acceptors to replace oxygen (*46-48*). The products of nitrate and nitrite are known to be harmful to DNA bases. Nitrosating reagents can cause both deamination and alkylation (*49*). Cells contain several enzymes, for example, Alka, alkylguanine DNA alkyltransferases (*50*), to repair this alkylation damage.
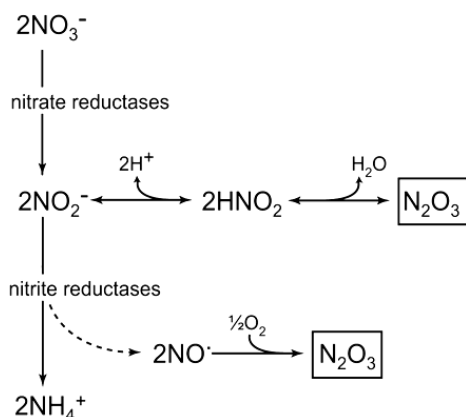
Figure 1.2 Production of the mutagenic nitrosating agent N_2O_3 during nitrate/nitrite metabolism in *E. coli*. NO·here is a minor by-product of nitrite reductase activity. Taken from (*51*)

Dinitrogen trioxide (N_2O_3), the anhydride of nitrous acid (Fig. 1.2), causes damage to DNA bases through N-nitrosation. N_2O_3 is a result from HNO_2 or NO• autooxidation. NO• is generated in *E. coli* cells during nitrate respiration (*52*). Nitrite reductases bind to nitrite, maintaining at low level (*47*). In acid condition, HNO_2 can cause mutations to *E. coli* DNA at stationary phase (*53*). The damage to DNA caused by nitrate or nitrite is at similar levels, indicating that these molecules use the same pathway. Research demonstrates that NO•, not HNO_2, is the major source producing N_2O_3 during nitrate and nitrite metabolism processes. It was also verified that N_2O_3 produced from HNO_2 condensation is the predominant product in *E. coli* when the cells are treated with HNO_2 (*51*).

The immunity to retroviruses, such as to HIV, involves DNA deamination of the retroviral first strand cDNA at certain stage (*54*). The protein (Vpr) encoded by HIV type 1 has been found to have the ability to degrade UNG or SMUG1 of the host cell, preventing AP sites from occurring (*55*). The mispair U:G is an important intermediate

for Ig genes during the process of somatic hypermutation (SHM) and class switch recombination (CSR) (*56, 57*).

Cytosine can be methylated to 5-methyl-cytosine, a process commonly observed in areas known for high mutation frequency. This type of research is a popular because of its close relation to cancer. 5-methyl-cytosine can be deaminated by nitrosative stress generating thymine, a normal base in DNA, to form a G:T mismatch pair, which is removed by mismatch repair enzymes (*58*).

Guanosine monophosphate reductase (*guaC*), one of the enzymes involved in the purine *de novo* synthesis pathway, can catalyze the deamination from GMP to IMP (*59-61*). IMP can be converted to ITP which can be mistakenly incorporated into chromosomal DNA generating mutations.

It has been found that the abundance of uracil increased in mammalian cells after the treatment with 5-fluorouracil. This can disturb the normal process of pyrimidine metabolism, increasing the possibility of the incorporation of uracil in the chromosome (*62*).

4. Repair mechanisms and pathways

DNA damage is diverse, and the consequences are serious. Therefore, cells have complex repair mechanisms to deal with this issue. Many DNA repair pathways involving multiple enzymes have been discovered. Since many types of DNA damage exist, one pathway can repair several kinds of DNA damage and several DNA repair enzymes can even recognize wide broad substrates. However, several kinds of repair

pathways are still required. The repaired DNA lengths in each repair pathways are quite different, from one nucleotide to approximately 1000 nucleotides. For example, in the short patch base excision repair pathway, only one nucleotide is removed while in mismatch repair pathway in either the *E. coli* or mammalian system, up to 1000 nucleotides are removed and regenerated.

A. Base excision repair pathway (BER)

Base excision repair pathway is a relatively simple; however, very important repair pathway and it recognizes many types of damaged DNA. Research demonstrates that BER is involved in the repair of the damage caused by ionizing radiation, alkylation, deamination, methylation chemicals. It is speculated that BER is a tumor suppressor mechanism (*63*).

I. Base excision repair pathway outline

BER includes several steps and usually repairs non-bulky (a single nucleotide or 2-10 nucleotides) DNA damage (*64*). BER pathway is initiated by a glycosylase which recognizes the damaged base. The specific glycosylase determines the patch size of the repair pathway (*65*). Two patch size pathways have been found in BER. Short patch size pathway (*66-69*), approximate 80-90% of all BER, is more frequently observed than long patch size pathway which is also called back up pathway (*70*). In certain rare cases, one enzyme may also be involved in both patch size pathways, such as human 3-methyladenine DNA glycosylase (*71*).
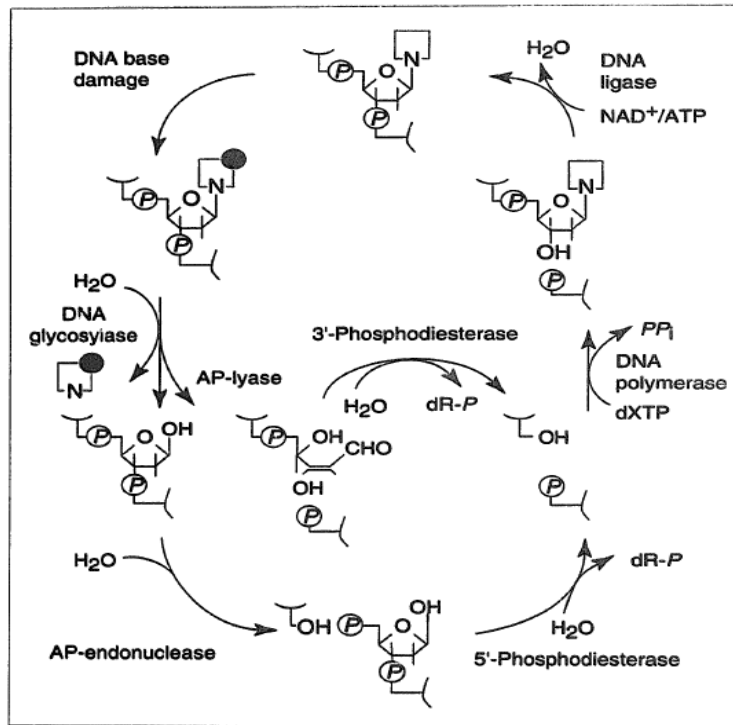
Figure 1.3 The base excision repair (BER) pathway. The nucleotide structure drawn is assumed to be part of a longer double-stranded DNA. Repair by DNA glycosylases without associated AP-lyase activity generally follows the pathway using AP-endonuclease and 5'-deoxyribophosphodiesterase activity. However, repair with N-glycosylases having associated AP-lyase activity uses the 3'-phosphodiesterase activity of the AP-endonuclease to make a single nucleotide gap that is filled by DNA polymerase and ligase. The β elimination reaction of the AP-lyase converts the deoxyribose residue to the aldehyde form. Taken from (*72*).

The glycosylic bond between the base and the sugar phosphate backbone is cleaved by one glycosylase and an apurinic/apyrimidinic (AP) sites is generated. If the glycosylase only removes the base, it is called monofunctional glycosylase. Then an AP-lyase (3' to the AP site is cleaved) or AP-endonuclease (usually APE1) (5' to the AP site is cleaved) is recruited depending on different glycosylases. If the glycosylase removes the base and possess AP-lyase activity, it is called bifunctional glycosylase. After AP lyase cleavage, the 3'-phosphate is removed by 3'-phosphodiesterase enzymes. While

10

after AP endonuclease (APE1) cleavage, the 5'-phosphate is removed by 5'-phosphodiesterase enzymes. Finally DNA polymerase and ligase fill the gap. The whole pathway is shown in Fig. 1.3 (*72*).

The abasic sites (AP site) generated after the cleavage of glycosylases is very mutagenic and cytotoxic (*73*). During DNA replication, the normal DNA polymerase does not use an AP site as a template and pass it. Because the AP site blocks the replication, other enzymes, such as DNA polymerases involved in translesion synthesis (TLS) usually insert a base following this order: G>A>T>C (*74*) and generate a site mutation, which is harmful to the cells (*75*). AP site also increases the possibility of double-strand breaks (*76*).

II. Uracil DNA glycosylase (UDG) superfamily

Uracil DNA glycosylase (UDG) superfamily is one of the families responsible for BER pathway. Until now five groups of uracil DNA glycosylases, which recognizes and repairs uracil have been found. Family 1 UDG (UDG or UNG) exclusively recognizes uracil containing substrates and repair most of the uracil on chromosome, while certain UDG superfamily members recognize more substrates. For example, Family 2 and 3 UDGs recognize xanthine. The details will be discussed in following sections.

The amino acid sequences are quite diverse among the UDG superfamily. However, they all share a similar crystal structural fold. According to amino acid sequence alignment and structure comparison, two highly conserved motifs have been found in the UDG superfamily. Motif 1 and the short helix after the motif 1 are used to

11

recognize the substrate. Motif 2 is used to insert into the DNA a "wedge" to replace the gap after the damaged base has been flipped out by glycosylase.

a. Uracil DNA glycosylase (UDG) Family 1

UDG Family 1 (UDG or UNG) is a group of highly conserved enzymes. Family 1 UDG glycosylase recognizes uracil exclusively on both single-stranded and double-stranded DNA. In *E. coli*, it is the primary enzyme responsible for removal of uracil which may be incorporated on chromosome or caused by deaminiation of cytosine. *E. coli* UNG mutant strains exhibit a 3 to 5 times increase in spontaneous mutation frequencies compared with wild type strains (*32, 77*). In *E. coli* UNG and dUTPase double mutants, approximately 20% of the thymine in genomic DNA is replaced by uracil although this does not harm the cells because uracil still prefers to be paired with adenine (*78*).

Family 1 UDG is wildly spread among prokaryotic, eukaryotic and viruses, while in archaea and *Drosophila melanogaster*, it is absent. In mammalian systems, there are two forms of UNG. Two different RNA splicing methods generate two UNGs. UNG1 and UNG2 come from one gene with different promoters and mRNA slicing (*79*). UNG1 was found in mitochondria (*79*) and UNG2 was located in nuclei which interacted with replication factor A (RPA) and proliferating cell nuclear antigen (PCNA). UNG2 was found in S phase during cell amplification (*80*) and it was regulated by cell cycles and repair the uracil in post-replication stage. UNG2 plays an important role in both short patch and long patch BER (*80*). In nucleoplasm and in pre-replicative BER which is in

front of replication fork, UNG2 repairs uracil and its analogs with short patch BER. While in post replicative BER which may be caused by misincorperation of uracil, UNG2 is utilized in long patch BER to remove uracil (*81*).

UNG2 has been found important in acquired immune response. It is up regulated in activated B-cells (*56*). The protein RPA, which binds to UNG2, also binds to activation-induced cytosine deaminase (AID) (*80*). AID is important in immunoglobulin's diversification. It catalyzes cytidine to uracil on single-stranded DNA, which is recognized by UNG2. Although uracil in single-stranded DNA is the best substrate for UNG2, only a small part of uracil from this deamination in IgV is repaired and unrepaired uracil is paired with adenine and this process generates many kinds of immunoglobulin. Research demonstrated that human cells with deficiency in UNG2 have impaired immunoglobulin class-switch recombination (*56*).

The crystal structures of Family 1 UDG from *E. coli*, human and Herpes have been obtained (*82-84*). The enzymes form a pocket with highly conserved residues among the family to recognize and bind the substrates. In *E. coli* UNG, the Gly 86-Gln 87 interacts with the O2 carbonyl of uracil to fix the base orientation. The Asn 123 can interact with the O4 of carbonyl and identify the bases, which helps UNG exclude cytosine and thymine while accept uracil as substrate (*82*).

b. Uracil DNA glycosylase (UDG) Family 2

The UDG Family 2 (MUG/TDG) was found initially as a G:T mismatch N-glycosylase. Later it was found that it possesses uracil DNA glycosylase activity (*85*).

Two types of enzymes with different properties have been found until now, thymine DNA glycosylase (TDG) and mismatch UDG (MUG). Family 2 UDG recognizes G:U base pair as substrate. It is insensitive to ugi, an inhibitor structurally specific to UDG Family 1, because UDG Family 2 does not have the conserved residues corresponding to UDG Family 1's catalytic sites and key substrate recognition residues.

The crystal structure of *E. coli* MUG has been resolved in 1998 (*86*). Although UNG and MUG in *E. coli* only share approximately 10% homology at amino acid sequence level, they share a similar 3-D structural fold. TDG/MUG was called as UDG Family 2 based on the structural similarity.

c. Uracil DNA glycosylase (UDG) Family 3

SMUG1 was firstly found in eukaryotic systems, *Xenopus*, as a protein binding to DNA repair enzyme inhibitors. The crystal structure of *Xenopus laevis* SMUG1 has been obtained (*87*). Human SMUG1 localizes at the human cell nucleus (*88*). It was named as a new family of uracil DNA glycosylase because it has low amino acid sequence similarity to UDG Family 1 (UDG or UNG) and Family 2 (MUG/TDG) enzymes; however, with structural similarity. Sequences analysis demonstrated that SMUG1 has hybrid active sites of UNG and MUG (*89*).

UNG is not the only primary repair enzyme in mammals responsible for uracil repair because a mouse UNG knock out mutant did not show high mutation frequency (*90*). This result also leads to the identification of SMUG1. SMUG1 plays a very important role in mice and maintains the similar level in both proliferating and non

proliferating cells, which is different with UNG2 (*91*). SMUG1 is not related to replication repair directly and SMUG1 exists with high abundance in heart and kidney. PCNA binds to UNG2 *in vivo*; however, not to SMUG1 by recognizing the motif QxxL/I/MxxF/HF/Y (*92*) which does not exist in SMUG1. It is estimated that SMUG1 evolves from UNG 550 million years ago (*91*).

The relative amount of UNG2 and SMUG1 are similar *in vivo*. Nilsen *et al.* once thought that SMUG1 was the backup enzyme of UNG2. However, with different assay conditions, SMUG1 was found to be as important as UNG2 (*91*). When damaged bases in mice cells are in excess, UNG2 is found to play a major role to repair the bases because of the enzyme's high turnover number. SMUG1 is reported to be a primary enzyme when substrates are not in excess because of the lower Km value compared with UNG2. Because the products of SMUG1, abasic sites, inhibits SMUG1's activity, adding APE1 enhances SMUG1's activity *in vitro* assays and increases the multi-turnover ability by eliminating these abasic sites (*90*).

Previously, SMUG1 were thought to exist only in eukaryotic systems. It was found exclusively in mammals, insects, parasites and ocean animals. While in our research, we investigated SMUG1's uracil and xanthine DNA glycosylase activities from a bacterium and it will be addressed in details in chapter 2.

SMUG1 was also found to be involved in removing uracil from immunoglobulin genes when it was overexpressed; however, the major function of SMUG1 is still in base excision repair pathway (*93*).

d. Uracil DNA glycosylase (UDG) Family 4

Thermatoga maritima UDGa was found to contain uracil DNA glycosylase activity; however, have different amino acid sequences from 1-3 UDG families. It was named as the UDG Family 4 (*94*). Later, its homologues were also found in certain archaea and eubacteria. Tma UDGa shows part of sequence homology to MUG; however, according to amino acid sequence BLAST, the total 192 Tma UDGa demonstrates approximately 44% identities and 60% positives with C terminal of *Stigmatella aurantiaca* phage SPO1 DNA polymerase domain protein.

An iron sulfur cluster was found first in UDG Family 4 in *Pyrobaculum aerophilum* (Pae-UDGa) firstly, which is conserved among UDG Family 4; however, is rare in other UDG families (*95*). In certain DNA repair enzymes, such as Nth/MutY family, iron sulfur also exists; however, it was not involved in catalytic activity (*96*). The crystal structure of *Thermus thermophilus* HB8 UDGa demonstrated that this iron sulfur is essential to stabilize a nearby loop. Besides iron sulfur, the salt bridge and ion pairs on the enzyme's surface and prolines on the enzyme's turns and loops were found to help Tth UDGa maintain activity at high temperature. *Thermus thermophilus* UDGa (Tth UDGa) recognizes both double-stranded and single-stranded uracil containing substrates and this property is similar as UDG Family 1 (*97*).

e. Uracil DNA glycosylase (UDG) Family 5

Most of UDG Family 4 and 5 were found in thermophilic prokaryotic organisms. Different from 1-4 UDG Families, a polar residue, used to hydrolyze the N-glycosidic bond's hydrolysis with a water molecule is missed in Family 5 UDG.

Family 5 UDGb is a broad substrate enzyme and has a classified sequence motif in the active site. *Thermus thermophilus* HB8 UDGb (Tth UDGb) recognizes uracil from double-stranded DNA while not be active on single-stranded DNA (*98*). It repairs the G:T mismatch which cannot be repaired by Tth UDGa. Tth UDGb also excises uracil analogues from DNA, such as 5'-hydroxymethyluracil (hmU) and 5'-fluorouracil (fU).

Similar to UDG Family 4 UDGa, UDG Family 5 UDGb also possesses four conserved cysteines to form an iron sulfur cluster. The crystal structure of Tth UDGb was obtained (*98*). Tth UDGb demonstrates 55% similarity and 32% identity to Tth UDGa on amino acid sequence level while with very low similarity to the other three UDG families. Both Family 4 and 5 UDGs share similar orders of secondary structures with other UDG family members.

B. Mismatch repair pathways (MMR)

Mismatch repair pathway is a relatively comprehensive repair pathway involved in correcting the errors generated during DNA replication. It requires several proteins' cooperation. The purified enzymes have been used to reconstitute the repair system in both *E. coli* (*99, 100*) and human cells (*101, 102*).
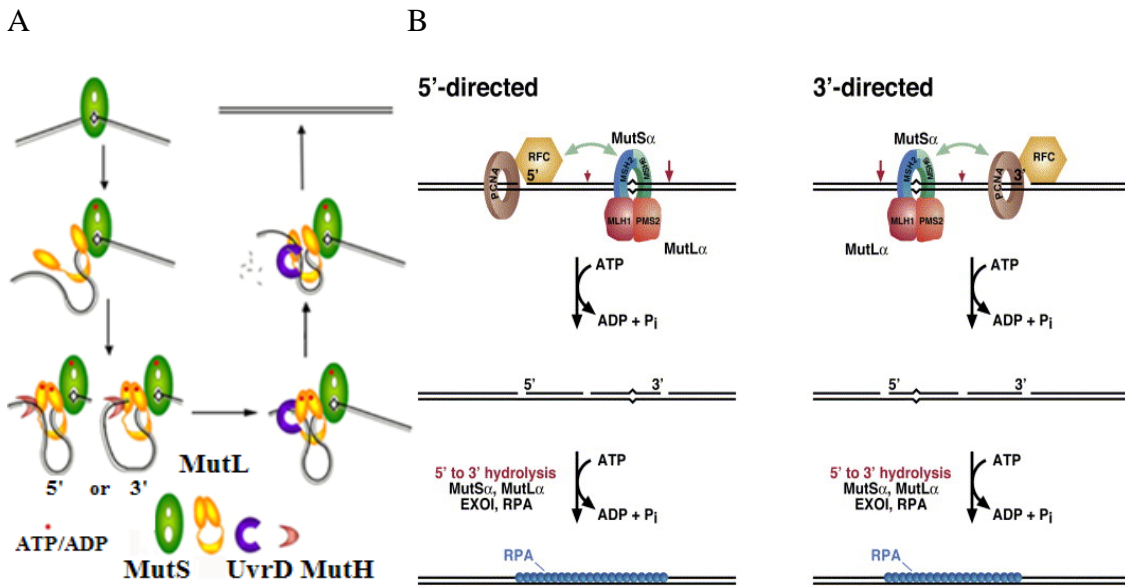
Figure 1.4 Mismatch repair pathway (MMR) in *E. coli* and human. (A) A working hypothesis of MutL-mediated mismatch repair in *E. coli*. MutS, MutL, MutH and UvrD are shape- and color-coded, and methylated (template) and unmethylated (daughter) DNA strands are shown in black and gray, respectively. Adapt from (*103*). (B) Incision of the Discontinuous Heteroduplex Strand in Human Mismatch Repair. MutSα, PCNA, and RFC activate a latent MutLα endonuclease, which incises the discontinuous strand of 5′ or 3′ heteroduplex DNAs in an ATP-dependent reaction. Incision displays a bias for occurrence on the distal side of the mismatch relative to the location of the original strand break (large arrows); however, can also occur proximal to the mispair (small arrows). For a 3′ heteroduplex, this yields a new 5′ terminus on the distal side of the mismatch that serves as an entry site for MutSα-activated ExoI, which removes the mismatch in a 5′-to-3′ hydrolytic reaction controlled by RPA (*102*). The strong bias for incision of the discontinuous strand implies signaling along the helix contour, which may involve ATP-promoted movement of MutSα or the MutSα-MutLα complex along the helix. Adapted from (*104*)

E. coli employs d(GATC) methylation mechanism to identify whether a strand is methylated or not and the methylated strand is not cleaved (Fig. 1.4A). First, MutS (Fig. 1.4A left) binds to the mismatch position and recruits MutL which demonstrates a MutS and ATP (shown as a red dot) dependent property. MutL binds DNA at both the mismatch and the nicked GATC sites, which releases adjacent DNA sequences. MutH is

18

a GATC sequence-specific endonuclease enzyme and it recognizes and is activated by the MutS-MutL-DNA complex (Fig. 1.4A middle left). MutH cleaves the unmethylated strand at GATC position, which is be regarded as a signal to synthesize the new DNA strand. MutH nicks both 3' and 5' of the mispaired DNA. After that, MutS-MutL-UvrD-DNA complex is formed with the presence of ATP. After ATP is hydrolyzed by MutL, the flexibility of MutS-MutL-UvrD-DNA complex becomes loose and reduces the size of DNA loop bound before. Then UvrD unwinds the DNA (Fig. 1.4A right). DNA helicase II and DNA polymerase III holoenzyme have been proved to be involved in the following repair pathway (*99, 100, 103*).

In mammalian systems, the process is quite similar to that in *E. coli*. The repair pathway is initiated by a discontinuous strand, a strand with either a nick or a gap (Fig. 1.4B). The *E. coli* MutS homolog MutSα includes MSH2-MSH6 and MutSβ includes MSH2-MSH3. MutSα is the major enzyme responsible for mismatch repair pathway (MMR) in mammals. PMS2 domain in MutLα (MLH1-PMS2) is conserved with many MutL homologs while it is absent in *E. coli*. It has been found that PMS2 has a $DQHA(X)_2E(X)_4E$ domain which is responsible for the endonuclease activity. Human MutLα demonstrates endonuclease activity in the presence of MutSα, PCNA, RFC and nicked DNA substrate (*104*). MutH homologue, an endonuclease in *E. coli*, is missing in human system. MutLα increases the mismatch dependent cleavage by decreasing excision of exonuclease I (ExoI) to regular DNA bases (*102*). With Mn^{2+}, MutLα also demonstrates endonuclease activity *in vitro* and it is enhanced by the presence of ATP. The nick produced by MutLα facilitates the entry of MutSα/β (*104*) which recruits ExoI

and enhances ExoI's ability to cleave the DNA strand. Without RPA (Fig. 1.4 B left), ExoI cleaves a few DNA bases from the strand following 5' to 3' direction, and with RPA, the DNA strand is bound with single-stranded binding proteins (RPA). As long as the mismatch bases exist, MutSα and MutLα maintain both ExoI and MutSα to bind with DNA. ExoI and MutSα will be released from DNA strands after removal approximately 250 nucleotides (*102*). Besides ExoI, MutSα, MutLα and RPA, PCNA (replication clamp), RFC (clamp loader), DNA polymerase δ and HMGB1 (DNA binding protein) are found to take part in human mismatch repair events (*99, 100*).

C. Nucleotide excision repair pathway (NER)

Nucleotide excision repair pathway is a complicated and versatile repair pathway. It recognizes many kinds of DNA damage, such as (6-4) photoproducts (6-4 PPs), and cyclobutane pyrimidine dimmers (CPDs) caused by short wave length UV from sun. It also repairs large bulky damage caused by chemical reactions.

NER in *E. coli* is mediated by UvrABC. In DNA repair enzymes, besides certain active site domains, there are several domains which are commonly used in the enzymes. One of them is adaptor domain and it mainly exists in eukaryotic systems. However, in prokaryotic cells, an adaptor domain from UvrB and UvrC has been found and it is used to connect these two proteins to form a complex (*105*).

NER core factors in mammals have been identified and NER pathway has been reconstituted with purified enzymes. Two types of NER pathways are found to date, including global genome NER (GG-NER) and transcription coupled NER (TC-NER).

20
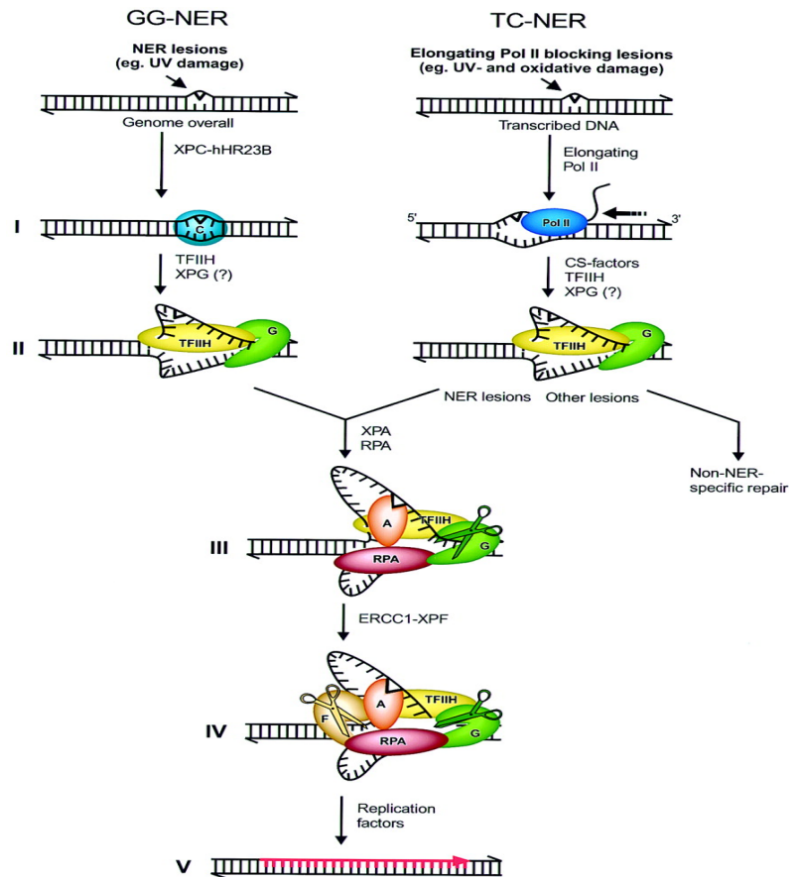
Figure 1.5 Molecular models for the incision stage of NER. Taken from (*106*)

Briefly, (Fig. 1.5) the pathway starts from the lesion recognition. (I) XPC (xeroderma pigmentosum C) -hHR23B recognizes the damage from DNA substrates by detecting the structure distortion in GG-NER. The lesions in TC-NER are recognized by elongating RNA Pol II because the enzyme is blocked by the lesions. (II) In GG-NER, the protein XPC-hHR23B binds with DNA and recruits TFIIH and possibly XPG. TFIIH is composed of XPB and XPD helicases, which can unwind DNA with approximately 10 to 20 nucleotides to form a so called "opened DNA complex". The helicases requires ATP to perform this function. Then XPC-hHR23B is disassociated from DNA or released

at later steps. In TC-NER, CS factors and TFIIH (XPG possibly involved) replace elongating RNA Pol II to interact with DNA. (III) XPA and RPA are recruited and participate in the complex which stabilizes this 10 to 20 nucleotide opened DNA complex. XPA binds to the damaged strand and RPA binds to the undamaged DNA strand. XPG here is also involved in stabilizing the whole complex. (IV) In the following step, XPG makes a 3' incision from the lesion with the help of TFIIH and RPA to locate the position. ERCC1-XPF makes the second incision 5' of the former lesion which is located by RPA and XPA. (V) After double cleavage, the gap will be filled by DNA synthesis and following ligation (*106*).

The NER defects cause several kinds of diseases, such as lung cancer (107, 108). The proteins within the NER pathway may have extra functions *in vivo*, because it has been found recently that the protein deficiency causes more kinds of disease beyond lung cancer.

D. Other repair pathways

Besides the repair pathways mentioned above, other repair pathways are also observed, homologous recombinational repair (HR) and nonhomologous end joining (NHEJ). Both of them deal with double-strand breaks damage. Homologous recombinational repair, involved in S and G2 phases, is accurate because it employs the homologous template. To the contrary, nonhomologous end joining, found in the G1 phase, introduces several errors and loses several nucleotides because it only joints the break ends. Double-strand breaks are harmful to cells because they cause chromosomal

instability if it is not repaired. The deficiency repair of double-strand breaks or the deficiency of its regulation is related to certain diseases, such as Ataxia telangiectasia (AT) and related disorders (*109*).

E. Endonuclease V mediated repair pathway

Endonuclease V mediated repair pathway is regarded as an alternative repair pathway. *E. coli* endonuclease V (endo V) was first purified in 1977 by biochemical methods as a 25 kDa DNA repair enzyme which could recognize single-stranded circular phage fd DNA. It was estimated that there are approximately 80-100 molecules in one cell, which is similar to endonuclease III and IV (*34*). Endo V is an authentic endonuclease and it generates 3'hydroxy and 5'phosphate terminus.

I. Endonuclease V in bacteria

a. Biochemical properties of *Escherichia coli* endo V

E. coli endo V is a broad substrate enzyme and it recognizes the DNA treated with OsO_4, UV or acid (pH5.0). It cleaves many sites on fd DNA; however, it cannot recognize either RNA or DNA/RNA hybrids. Endo V digests deaminated bases, AP sites, uracil, FLAP, pseudo Y structure, and small insertions or deletions (*34, 110-112*). Because endo V's activity is enhanced with uracil containing substrates, it was thought it provided an alternative way to repair uracil (*113, 114*). Until now, the knowledge about endo V is limited to the biochemical and genetic studies.

Endo V requires divalent metals to facilitate its activity, such as Mg^{2+}, while Co^{2+} and Mn^{2+} recover its activity to certain extent (*34, 115*). Endo V demonstrates different cleavage abilities with different metals and it also demonstrates activity in a broad pH range, from pH 6-8 (*34*). It digests substrates processively because it moves to next substrate after it digests multiply damaged sites on one substrate completely (*114*).

b. *Escherichia coli* endo V mediated repair pathway

All deaminated bases, including uracil, hypoxanthine, xanthine and oxanine, are substrates of *E. coli* endonuclease V (*34, 116-119*). In *E. coli*, 3-methyladenine DNA glycosylase II (AlkA) (*120*) has been found to recognize hypoxanthine. *E. coli* MutSLH is a mismatch repair enzyme complex which recognizes I:T pairs in newly synthesized DNA strands (*121*). *E. coli* AlkA recognizes hypoxanthine; however, it has been estimated that 93-99% hypoxanthine is repaired by endo V, which provide a solid foundation to use genetic methods to detect the patch length of DNA involved in the repair pathway (*122*). The spontaneous mutation frequency or HNO_2 induced frequency in AlkA mutant strain is not increased dramatically compared with wild type strain, which is quite different with the results from endo V mutant. This result proves that AlkA is not the major enzyme to repair hypoxanthine in *E. coli*.

Different from uracil DNA glycosylase superfamily members, *E. coli* endo V recognizes the inosine on DNA strand with no preference to the opposite bases and cleave 3' of the 2nd phosphodiester bond (*34, 116, 118, 119, 123*), and endo V binds both

substrate and product tightly. The damaged base is not cleaved and still remains on the DNA strand which indicates that endo V may just initiate the repair pathway.

The patch length of DNA involved in the repair pathway initiated by endo V has been reported recently (*122*). Then full-length deoxyhypoxanthine (dI)-containing oligodeoxyribonucleotides and their corresponding cleavage products by endo V were transformed to *E. coli* strains contacting different sets of *lacZ* mutations. The strain with full-length dI-containing oligo grows on selective plates while the strain with the cleavage product of the dI-containing oligo cannot grow. The results indicate that the dI-containing oligo is incorporated onto the chromosome before cleavage by endo V. This is very important to the genetic study to be sure that the oligodeoxyribonucleotides will not suffer digestion before incorporation into chromosome. G:T base pairs are used to replace the correct base pairs as an indicator because the G:T base pair is a relative stable mismatch pair with a shorter half life and a longer open-state lifetime than normal correct base pairs and it does not introduce too much distortion for the DNA double helix structure. dI-containing oligodeoxyribonucleotides with dG, paired with dT on the other strand, at different distance positions were chosen.

After the oligodeoxyribonucleotides were incorporated onto the chromosome, the endo V repair pathway recognized the inosine and repaired the adjacent nucleosides. According to Dr. Weiss' studies, the oligodeoxyribonucleotides are incorporated into the lagging strand of the chromosome. G:T base pairs in *E. coli* have approximately   twenty minutes to be repaired. It is important for the cell to have enough time to recruit endoV to cleave the dI. After repair, the correct nucleosides recovers the wild type *lacZ* and allow

E. coli to grow on selection medium. It is the first demonstration of the repair patch for the pathway mediated by endo V. It was found that the patch DNA size in endo V initiated repair pathway is approximately 5 nucleotides, with 2 nucleotides on the 5' and 3 nucleotides on the 3' of the cleavage cut position (*122*). The repair patch length of this repair pathway is close to the very short patch mismatch repair pathway which removes 2-10 nucleotides according to genetic experiments (*124*). The proteins used to remove the nucleotides from the patch position and the interaction behaviors are still unknown.

c. *In vivo* function study of *Escherichia coli* endo V

There are several uracil DNA glycosylases in *E. coli*, such as UNG and MUG. They are all responsible for uracil repair, it was thought that endo V's activity to uracil may not be as important as UDG family enzymes (*123*) although endo V demonstrates robust uracil cleavage activity. And *in vivo* genetic study demonstrates that neither overexpression nor knockout of endo V gene influences the growth of single-stranded DNA phages M13 or uracil containing bacteriophage λ (*125*). The spontaneous mutation frequencies of endo V mutant in either *ung* overexpression or knockout strains are similar (*126*). *In vitro* binding study also shows that endo V may not play a critical role to repair uracil because endo V fails to bind the uracil containing substrates tightly (*123*).

E. coli nfi also demonstrates the anti-mutator property. The strain grows with sodium nitrate or sodium nitrite at neutral pH condition, which indicates that *nfi* is an anti-mutator in nitrate and nitrite metabolism process. *E. coli* endo V mutant strain

26

demonstrated two times higher for spontaneous mutation frequency than that of in the wild type strain (*125, 127*) and 12 times higher than wild type strain under nitrous acid stress (*125*).

d. Biochemical properties of *Thermotoga maritima* endo V

Thermotoga maritima (Tma) endo V shares 44% amino acid identities with *E. coli* Endo V and it has been extensively studied. Tma endo V possesses similar properties with *E. coli* endo V. For example, it recognizes inosine, abasic site (AP site), uracil and mismatched substrates. It cleaves inosine-containing strand and also nicks its complementary strand at two different positions and these two nick events are independent (*123*).

Tma endo V cleaves 2-3 nucleotides at the 5' side of complementary strand in I:C and I:T substrates under certain conditions. Tma endo V cleaves both strands and demonstrates a higher preference for purine mismatch bases than for pyrimidine mismatch bases. For inosine-containing substrates, with Mg^{2+}, Tma endo V shows higher cleavage activity than that with Mn^{2+}. For mismatched substrates, with Mn^{2+}, Tma endo V possesses higher activity than that with Mg^{2+} (*123*). Tma endo V demonstrates different cleavage preferences from *E. coli* endo V. *E. coli* endo V does not cleave C:A, C:C and C:T mismatch bases, while it cleaves all inosine-containing mismatch base pairs without any preference to the opposite bases; however, if the base pairs have G:C pairs in adjacent area, inosine cannot be cleaved by *E. coli* endo V.
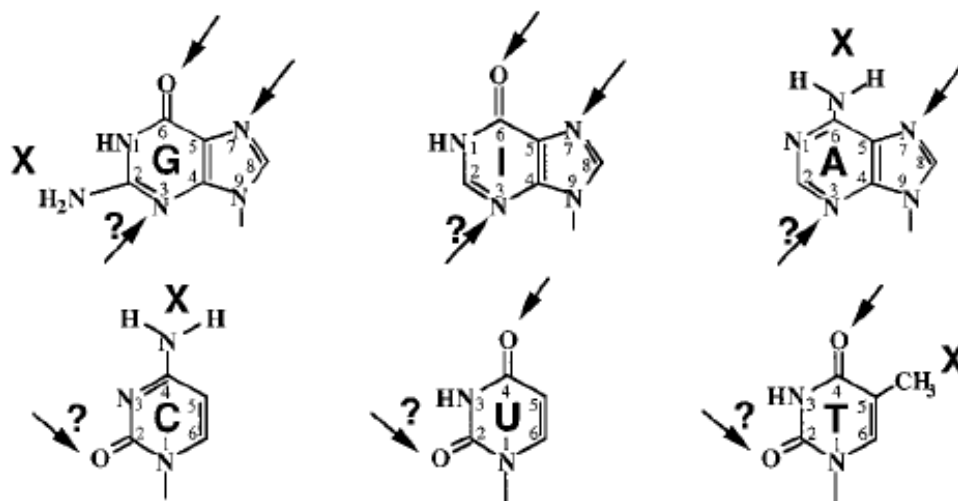
Figure 1.6 Potential base contacts with Tma Endonuclease V. The possible favorable interactions are marked by an arrow. The unfavorable interactions are marked by an X. Taken from (*123*)

It has been demonstrated that Tma endo V may interact with 6-keto of purine (*110*) and 4-keto of uracil and N7 of purine may also be involved in the interaction. The N3 in purines is possible to interact with enzymes (Fig. 1.6), which may help enzyme distinguish purine from pyrimidine (*123*).

Tma Endo V is known as an endonuclease; however, interestingly, exonuclease activity has been detected in Tma endo V (*118*). Both endonuclease and exonuclease activities are studied in our group, and exonuclease activities will be discussed in details in chapter 3. WT Tma endo V demonstrated robust exonuclease activities in the reaction condition with $MnCl_2$. Although it is known that Mn^{2+}'s concentration is not very high *in vivo*. It has been demonstrated that the metal Mn^{2+} may help enzyme to show its latent property and it may require several enzymes or cofactors to reveal the activity *in vivo*. For example, human MutLα demonstrates endonuclease activity with metal Mn^{2+} and

low KCl concentration *in vitro* and no endonuclease activity is detected with metal Mg^{2+}, while this endonuclease activity requires three proteins' (MutSα, RFC and PCNA's) cooperation *in vivo* (*104*). Tma endo V's exonuclease activity starts after endonuclease cleavage with hypoxanthine containing substrates. WT Tma endo V possesses both 3' hypoxanthine independent and 5' exonuclease activity and it also demonstrates non-specific endonuclease activity on regular DNA substrates (*123*).

Although endo V's biochemical properties have been studied extensively, and endo V's function *in vivo* has been studied intensively in *E. coli* under the nitrosive condition in Dr. Weiss group with genetic methods, the repair pathway remains an enigma.

II. Endonuclease V in mammals

Endo V not only exists in prokaryotic microorganism; however, also in archaea, eukaryotic organisms, such as yeast, mustard, rice, pig, mouse and human and it is highly conserved among these species which were found by ETS database searching. Based on amino acid sequence alignment, seven motifs were identified.

Mouse endo V, with the accession number of XP_203558, was identified in 2003 with 338 amino acids and 37.2 kDa molecular weight. It has 32% amino acid sequence identity with *E. coli* endo V. It was amplified from cDNA and cloned into *E. coli* vector. After expression and purification, mouse endo V was tested with biochemical assays and demonstrated highest activity on single-stranded inosine-containing substrates. However, the activity was very low. Even under the condition that mouse endo V's concentration is

100 times higher than that of *E. coli* endo V, the cleavage activity (80%) is still lower than that of *E. coli* endo V (100%). Because genetic study is more sensitive than biochemical study, mouse endo V was found to suppress spontaneous mutation frequency which means mouse endo V is also an anti-mutator (*128*).

In mouse, it has been found that AAG is currently the only hypoxanthine DNA glycosylase, which recognizes and repair hypoxanthine containing DNA. The AAG knockout mice are still viable and do not show too many phenotypes difference from wild type mice (*129, 130*). Hypoxanthine DNA glycosylase from calf thymus recognizes hypoxanthine from both I:T and I:C base pairs and showed 15 to 20 times faster on I:C than I:T base pair (*131*).

Mouse endo V expresses widely in tissues. According to sequence data search, mouse endo V was found express widely in eyeballs, lung, and brain. Experimental data show that it is abundant in liver, heart, kidney and testis (*128*). The wide distribution indicates that endo V is very important to life events. Mouse endo V doesn't recognize broad substrates, which indicates that bacteria endo V, such as *E. coli* endo V and Tma endo V, may have gained the ability to repair broad substrates during evolution (*132*).

Human endo V DNA sequences were first obtained from human genome project. It is located on chromosome 17q25.3 with 22,915 bp genomic DNA. It has 10 exons and with 7 corresponding homologous motifs among endo V family. Human endo V has 282 amino acids with 31.9 kDa molecular weight and the pI is 6.46. Now in our lab, we have cloned human endo V gene and expressed in *alka⁻*, *ung⁻*, and *mug⁻* BL21(DE3) mutant strain. Soluble and active protein was purified. Biochemical and genetic methods were

employed to investigate human endo V's properties, which will be addressed in details in Chapter four. Briefly, human endo V shows highest activity in Tris-HCl pH 8.0 buffer with low concentration of NaCl. Similar to bacteria endo V, human endo V requires divalent metal ion and the best one is Mg^{2+}. Human endo V exhibits lower activity with metal Mn^{2+}. Its substrates preference follows this order single-stranded I> G/I> T/I>> A/I and C/I. It is not a broad substrate enzyme and only shows activity for inosine-containing substrates and very low activity to xanthine-containing substrates, which may prove the notion that bacteria endo V may gain the broad substrate recognition ability during evolution (*132*).

Human endo V is also an anti-mutator enzyme. Mispaired, oxidized, or hydrolytically deaminated damage occurs without external reagent. Investigations in our lab demonstrated that expressing human endo V in *E. coli* complements *E. coli* endo V to decrease the spontaneous mutation frequency for *E. coli* endo V mutant strain. It has been reported that *E. coli* endo V mutant have a huge survival rate (12 to 1,000-fold) on streptomycin (Str) containing medium than wild type strain after treated with nitrous acid (*125*). Streptomycin binds to 16S rRNA of the bacterial ribosome and inhibits protein synthesis; however, does not cause cell death (*133*). It has been verified that mutations in gene *rpsL* (ribosomal protein S12) helps strains resistant to streptomycin (*134*). Mutations to *rpsL* gene are mainly occur at one of two lysine codons, AAA (*135*). An *E. coli* endo V mutant that does not have the ability to repair inosine causing mutations to *rpsL* gene and allow *E. coli* to grow on Str containing medium. The better ability to

31

survive on Str containing plates demonstrated that endo V plays a major role in repairing inosine damage *in vivo* (*125*).

III. Endonuclease V's homologues

With the help of software Metabasic, combining both amino acid sequences and protein secondary structure, it has been found that endo V is a homolog of UvrC, a subunit of UvrABC complex (*136*). Both enzymes cleave the nucleotide next to the damaged nucleotide. UvrABC complex recognizes pyrimidine dimmers and nucleotide containing bulky adducts, and serves as an endonuclease which is involved in NER pathway. UvrABC complex cleaves both sides of the damaged DNA bases. It removes 15 nucleotides from 3' and 8 nucleotides from 5' of the damaged DNA bases. *E. coli* UvrC mutants at position D399 and D466 diminish 5' cleavage while maintain 3' cleavage (*137, 138*). Based on sequence homology analysis, endo V only possesses one domain which is responsible for 5' cleavage (*139*).

5. DNA damage and consequences

DNA damage exists all the time. If cells fail to repair the damage which may be caused by inactive DNA repair enzyme or errors in repair pathway, there will be many consequences. One of the frequently observed consequences is disease.

Superoxide ($O_2\bullet$) and nitric oxide ($NO\bullet$), generated from active phagocytes, can cause inflammation, reperfusion and other pathological conditions. It has been reported that chronic infection and inflammation may cause carcinogenesis diseases (*140, 141*).

XPD is a helicase involved in transcription and nucleotide excision repair pathways. XP mutant proteins failed to repair the DNA damage treated after oxygen free radicals (*142*). XPD mutation causes hypersensitivity to UV and increase the risk for skin cancer (*143*). The study of XPD polymorphisms showed a surprising result that individuals with Lys751, on the common allele, maintain low DNA repair proficiency (*144*) . Cockayne syndrome (CS) is the second human nucleotide excision repair related disease. The patients are sensitive to UV light and unable to synthesis RNA after UV radiation. Also the patients have several of these following problems: neurological abnormalities, dwarfism, premature aging of certain tissues, facial and limb abnormalities, and early death due to neurodegeneration (*145, 146*). Another disease found to be related to nucleotide excision repair pathway is trichothiodystrophy (TTD). The patients are photosensitivity and the main symptoms are brittle hair and nails (*147*). These results are all obtained from clinical studies and the specific enzymes found responsible for these diseases are enzymes in XP group.

8-hydroxy-2-deoxyguanine is found to be related to UVB-induced skin cancer. The level of 8-oxo-dG is increased when the mice were treated with chronic UVB exposure (*148*), while the addition of exogenous OGG1 helps decrease the tumor size on mice (*149*). Approximately 20 different variants of OGG1 have been identified to date and several of them are associated with cancer. For example, ser326cys variant was found in lung cancer patient (*63, 150*).

In hereditary nonpolyposis colorectal cancer (HNPCC) patients, mutations in several mismatch repair enzymes have been found. They are hMLH1, hMSH2, hPMS1 and hPMS2 (*151*).

MBD4 gene (methyl-CpG binding domain protein 4 gene) has one polyA(10) and three polyA(6). The mutations in these repeats are found in several tumor cells, such as colon carcinomas cells (*150*).

Another commonly observed consequence for DNA damage is related to evolution. The studies about DNA repair enzymes' properties, such as substrates recognition and enzymatic functions' divergence may help explain the evolution. Because the DNA repair enzymes are evolved under both intracellular and extracellular pressure, such as oxidative stress, the relationships among those enzymes and the repair pathways reflect the evolution to certain extent (*152*).

Novel properties emerge during evolution corresponding to environmentally selective pressure according to Darwinian's theory. The process is slow and has to accumulate enough mutations to certain degree to generate new characters. The theory of sudden origins (*153, 154*) suggests that unrepaired DNA damage increases the number of the mutation events and change the gene expression. Although DNA repair enzymes repair damage and provide stability of chromosome, under certain conditions, the damaged DNA cannot be repaired in time and correctly. According to this theory, a large amount of mutations occur in a short time to generate novel properties (*154*).

6. References

1.	Mandelkern, M., Elias, J. G., Eden, D., and Crothers, D. M. (1981) The dimensions of DNA in solution. *Journal of Molecular Biology 152*, 153-161.

2.	McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., Whittaker, A., Holden, J. L., McCann, O. T., Dunham, A., Soderlund, C., Scott, C. E., Bentley, D. R., Schuler, G., Chen, H. C., Jang, W., Green, E. D., Idol, J. R., Maduro, V. V., Montgomery, K. T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J. H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P. J., Catanese, J. J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V. G., Kirsch, I. R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J. F., Hawkins, T., Myers, R. M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N. E., Cox, D. R., Haussler, D., Kent, W. J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X. N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H. S., Sakaki, Y., Shimizu, N., Asakawa, S., et al. (2001) A physical map of the human genome. *Nature 409*, 934-941.

3.	Lodish H, B. A., Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J. (2004) *Molecular Biology of the Cell*, 5th ed., WH Freeman, New York, NY.

4.	Marnett, L. J., and Plastaras, J. P. (2001) Endogenous DNA damage and mutation. *Trends in Genetics 17*, 214-221.

5.	De Bont, R., and van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis 19*, 169-185.

6.	Burcham, P. C. (1999) Internal hazards: baseline DNA damage by endogenous products of normal metabolism. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis 443*, 11-36.

7.	Esterbauer, H., Eckl, P., and Ortner, A. (1990) Possible mutagens derived from lipids and lipid precursors. *Mutation Research/Reviews in Genetic Toxicology 238*, 223-233.

8.	Liehr, J. G. (2000) Is Estradiol a Genotoxic Mutagenic Carcinogen? *Endocr Rev 21*, 40-54.

9.   Friedberg, E. C., McDaniel, L. D., and Schultz, R. A. (2004) The role of endogenous and exogenous DNA damage and mutagenesis. *Current Opinion in Genetics & Development 14*, 5-10.

10.   Eker, A. P., and Fichtinger-Schepman, A. M. (1975) Studies on a DNA photoreactivating enzyme from Streptomyces griseus. II. Purification of the enzyme. *Biochim Biophys Acta 378*, 54-63.

11.   Kanno, S., Iwai, S., Takao, M., and Yasui, A. (1999) Repair of apurinic/apyrimidinic sites by UV damage endonuclease; a repair protein for UV and oxidative damage. *Nucl. Acids Res. 27*, 3096-3103.

12.   Peak, J. G., Ito, T., Robb, F. T., and Peak, M. J. (1995) DNA damage produced by exposure of supercoiled plasmid DNA to high- and low-LET ionizing radiation: effects of hydroxyl radical quenchers. *Int J Radiat Biol 67*, 1-6.

13.   Maity, A., McKenna, W. G., and Muschel, R. J. (1994) The molecular basis for cell cycle delays following ionizing radiation: a review. *Radiother Oncol 31*, 1-13.

14.   Bernhard, E. J., Maity, A., Muschel, R. J., and McKenna, W. G. (1995) Effects of ionizing radiation on cell cycle progression. A review. *Radiat Environ Biophys 34*, 79-83.

15.   Iliakis, G. (1997) Cell cycle regulation in irradiated and nonirradiated cells. *Semin Oncol 24*, 602-615.

16.   Walters, R. A., Gurley, L. R., and Tobey, R. A. (1974) Effects of caffeine on radiation-induced phenomena associated with cell-cycle traverse of mammalian cells. *Biophys J 14*, 99-118.

17.   Tobey, R. A. (1975) Different drugs arrest cells at a number of distinct stages in G2. *Nature 254*, 245-247.

18.   Jin, Y. H., Clark, A. B., Slebos, R. J., Al-Refai, H., Taylor, J. A., Kunkel, T. A., Resnick, M. A., and Gordenin, D. A. (2003) Cadmium is a mutagen that acts by inhibiting mismatch repair. *Nat Genet 34*, 326-329.

19.   Ford, G. P., and Herman, P. S. (1992) Relative stabilities of nitrenium ions derived from polycyclic aromatic amines. Relationship to mutagenicity. *Chem Biol Interact 81*, 1-18.

20.   Grant, D. M., Josephy, P. D., Lord, H. L., and Morrison, L. D. (1992) Salmonella typhimurium strains expressing human arylamine N-acetyltransferases: metabolism and mutagenic activation of aromatic amines. *Cancer Res 52*, 3961-3964.

21.   Sarkar, F. H., Radcliff, G., and Callewaert, D. M. (1992) Purified prostaglandin synthase activates aromatic amines to derivatives that are mutagenic to Salmonella typhimurium. *Mutat Res 282*, 273-281.

22.   Sinsheimer, J. E., Hooberman, B. H., Das, S. K., Brezzell, M. D., and You, Z. (1992) The in vivo and in vitro genotoxicity of aromatic amines in relationship to the genotoxicity of benzidine. *Mutat Res 268*, 255-264.

23.   Yamazaki, H., Oda, Y., and Shimada, T. (1992) Use of a newly developed tester strain Salmonella typhimurium NM2009 for the study of metabolic activation of carcinogenic aromatic amines by rat liver microsomal cytochrome P-450 enzymes. *Mutat Res 272*, 183-192.

24.   Rouse, J., and Jackson, S. P. (2002) Interfaces Between the Detection, Signaling, and Repair of DNA Damage. *Science 297*, 547-551.

25.   Demple, B., and Harrison, L. (1994) Repair of Oxidative Damage to DNA: Enzymology and Biology. *Annual Review of Biochemistry 63*, 915-948.

26.   Sinha, R. P., and Hader, D. P. (2002) UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci 1*, 225-236.

27.   Cosma, M. P., Tanaka, T., and Nasmyth, K. (1999) Ordered Recruitment of Transcription and Chromatin Remodeling Factors to a Cell Cycle- and Developmentally Regulated Promoter. *Cell 97*, 299-311.

28.   Hitchcock, T. M. (2005) in *Genetics* pp 7, Clemson University, Clemson.

29.   Suzuki, T., Ide, H., Yamada, M., Endo, N., Kanaori, K., Tajima, K., Morii, T., and Makino, K. (2000) Formation of 2'-deoxyoxanosine from 2'-deoxyguanosine and nitrous acid: mechanism and intermediates. *Nucleic Acids Res 28*, 544-551.

30.   Suzuki, T., Kanaori, K., Tajima, K., and Makino, K. (1997) Mechanism and intermediate for formation of 2'-deoxyoxanosine. *Nucleic Acids Symp Ser*, 313-314.

31.   Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature 274*, 775-780.

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

33. Bradshaw, J. S., and Kuzminov, A. (2003) RdgB acts to avoid chromosome fragmentation in Escherichia coli. *Mol Microbiol 48*, 1711-1725.

34. Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli. *J Biol Chem 269*, 16260-16268.

35. Vongchampa, V., Dong, M., Gingipalli, L., and Dedon, P. (2003) Stability of 2'-deoxyxanthosine in DNA. *Nucleic Acids Res 31*, 1045-1051.

36. 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 Res 14*, 8135-8153.

37. Toyokuni, S., Mori, T., and Dizdaroglu, M. (1994) DNA base modifications in renal chromatin of Wistar rats treated with a renal carcinogen, ferric nitrilotriacetate. *Int J Cancer 57*, 123-128.

38. Spencer, J. P., Wong, J., Jenner, A., Aruoma, O. I., Cross, C. E., and Halliwell, B. (1996) Base modification and strand breakage in isolated calf thymus DNA and in DNA from human skin epidermal keratinocytes exposed to peroxynitrite or 3-morpholinosydnonimine. *Chem Res Toxicol 9*, 1152-1158.

39. Glaser, R., Rayat, S., Lewis, M., Son, M. S., and Meyer, S. (1999) Theoretical Studies of DNA Base Deamination. 2. Ab Initio Study of DNA Base Diazonium Ions and of Their Linear, Unimolecular Dediazoniation Paths. *J. Am. Chem. Soc. 121*, 6108-6119.

40. Miller, R. V., and Day, M. J. (2004) *Microbial Evolution Gene Establishment, Survival, and Exchange*, Wiley.

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

42. Ponnamperuma, C., Lemmon, R. M., Bennett, E. L., and Calvin, M. (1961) Deamination of Adenine by Ionizing Radiation. *Science 134*, 113.

43. Schuster, H. (1959) The infectivity of viral nucleic acids. *Dtsch Med Wochenschr 84*, 1868-1870.

44.   Nguyen, T., Brunson, D., Crespi, C. L., Penman, B. W., Wishnok, J. S., and Tannenbaum, S. R. (1992) DNA damage and mutation in human cells exposed to nitric oxide in vitro. *Proc Natl Acad Sci U S A 89*, 3030-3034.

45.   Shapiro, R., and Pohl, S. H. (1968) The reaction of ribonucleosides with nitrous acid. Side products and kinetics. *Biochemistry 7*, 448-455.

46.   Berks, B. C., Ferguson, S. J., Moir, J. W., and Richardson, D. J. (1995) Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochim Biophys Acta 1232*, 97-173.

47.   Rupp, W. D. (1996) *Escherichia coli and Salmonella : cellular and molecular biology*, 2nd ed., ASM Press, Washington, D.C. .

48.   Cole, J. (1996) Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? *FEMS Microbiol Lett 136*, 1-11.

49.   Taverna, P., and Sedgwick, B. (1996) Generation of an endogenous DNA-methylating agent by nitrosation in Escherichia coli. *J Bacteriol 178*, 5105-5111.

50.   Sedgwick, B. (1997) Nitrosated peptides and polyamines as endogenous mutagens in O6-alkylguanine-DNA alkyltransferase deficient cells. *Carcinogenesis 18*, 1561-1567.

51.   Weiss, B. (2006) Evidence for mutagenesis by nitric oxide during nitrate metabolism in Escherichia coli. *J Bacteriol 188*, 829-833.

52.   Corker, H., and Poole, R. K. (2003) Nitric oxide formation by Escherichia coli. Dependence on nitrite reductase, the NO-sensing regulator Fnr, and flavohemoglobin Hmp. *J Biol Chem 278*, 31584-31592.

53.   Zimmermann, F. (1977) Genetic effects of nitrous acid. *Mutat Res 39*, 127-148.

54.   Harris, R. S., Sheehy, A. M., Craig, H. M., Malim, M. H., and Neuberger, M. S. (2003) DNA deamination: not just a trigger for antibody diversification but also a mechanism for defense against retroviruses. *Nat Immunol 4*, 641-643.

55.   Schrofelbauer, B., Yu, Q., Zeitlin, S. G., and Landau, N. R. (2005) Human immunodeficiency virus type 1 Vpr induces the degradation of the UNG and SMUG uracil-DNA glycosylases. *J Virol 79*, 10978-10987.

56.   Imai, K., Slupphaug, G., Lee, W. I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H. E., Ochs, H. D., Fischer, A., and Durandy, A. (2003) Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol 4*, 1023-1028.

57.   Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T., and Neuberger, M. S. (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol 12*, 1748-1755.

58.   Jones, M., Wagner, R., and Radman, M. (1987) Mismatch repair of deaminated 5-methyl-cytosine. *J Mol Biol 194*, 155-159.

59.   Li, J., Wei, Z., Zheng, M., Gu, X., Deng, Y., Qiu, R., Chen, F., Ji, C., Gong, W., Xie, Y., and Mao, Y. (2006) Crystal Structure of Human Guanosine Monophosphate Reductase 2 (GMPR2) in Complex with GMP. *Journal of Molecular Biology 355*, 980-988.

60.   Andrews, S. C., and Guest, J. R. (1988) Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12. *Biochem J 255*, 35-43.

61.   Saxild, H. H., and Nygaard, P. (1991) Regulation of levels of purine biosynthetic enzymes in Bacillus subtilis: effects of changing purine nucleotide pools. *J Gen Microbiol 137*, 2387-2394.

62.   Longley, D. B., Harkin, D. P., and Johnston, P. G. (2003) 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer 3*, 330-338.

63.   Sweasy, J. B., Lang, T., and DiMaio, D. (2006) Is base excision repair a tumor suppressor mechanism? *Cell Cycle 5*, 250-259.

64.   Dianov, G., Price, A., and Lindahl, T. (1992) Generation of single-nucleotide repair patches following excision of uracil residues from DNA. *Mol. Cell. Biol. 12*, 1605-1612.

65.   Fortini, P., Parlanti, E., Sidorkina, O. M., Laval, J., and Dogliotti, E. (1999) The Type of DNA Glycosylase Determines the Base Excision Repair Pathway in Mammalian Cells. *J. Biol. Chem. 274*, 15230-15236.

66.   Matsumoto, Y., and Kim, K. (1995) Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science 269*, 699-702.

67.   Sobol, R. W., Horton, J. K., Kuhn, R., Gu, H., Singhal, R. K., Prasad, R., Rajewsky, K., and Wilson, S. H. (1996) Requirement of mammalian DNA polymerase-[beta] in base-excision repair. *Nature 379*, 183-186.

68. Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E., and Lindahl, T. (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *Embo J 15*, 6662-6670.

69. Wang, Z., Wu, X., and Friedberg, E. C. (1997) Molecular mechanism of base excision repair of uracil-containing DNA in yeast cell-free extracts. *J Biol Chem 272*, 24064-24071.

70. Matsumoto, Y., Kim, K., and Bogenhagen, D. F. (1994) Proliferating cell nuclear antigen-dependent abasic site repair in Xenopus laevis oocytes: an alternative pathway of base excision DNA repair. *Mol. Cell. Biol. 14*, 6187-6197.

71. Parker, A., Gu, Y., Mahoney, W., Lee, S.-H., Singh, K. K., and Lu, A. L. (2001) Human Homolog of the MutY Repair Protein (hMYH) Physically Interacts with Proteins Involved in Long Patch DNA Base Excision Repair. *J. Biol. Chem. 276*, 5547-5555.

72. Seeberg, E., Eide, L., and Bjoras, M. (1995) The base excision repair pathway. *Trends in Biochemical Sciences 20*, 391-397.

73. Auerbach, P., Bennett, R. A., Bailey, E. A., Krokan, H. E., and Demple, B. (2005) Mutagenic specificity of endogenously generated abasic sites in Saccharomyces cerevisiae chromosomal DNA. *Proc Natl Acad Sci U S A 102*, 17711-17716.

74. Neto, J. B., Gentil, A., Cabral, R. E., and Sarasin, A. (1992) Mutation spectrum of heat-induced abasic sites on a single-stranded shuttle vector replicated in mammalian cells. *J Biol Chem 267*, 19718-19723.

75. Loeb, L. A., and Preston, B. D. (1986) Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet 20*, 201-230.

76. Kingma, P. S., Corbett, A. H., Burcham, P. C., Marnett, L. J., and Osheroff, N. (1995) Abasic sites stimulate double-stranded DNA cleavage mediated by topoisomerase II. DNA lesions as endogenous topoisomerase II poisons. *J Biol Chem 270*, 21441-21444.

77. Fix, D. F., and Glickman, B. W. (1987) Asymmetric cytosine deamination revealed by spontaneous mutational specificity in an Ung- strain of Escherichia coli. *Mol Gen Genet 209*, 78-82.

78. Tye, B. K., Chien, J., Lehman, I. R., Duncan, B. K., and Warner, H. R. (1978) Uracil incorporation: a source of pulse-labeled DNA fragments in the replication of the Escherichia coli chromosome. *Proc Natl Acad Sci U S A 75*, 233-237.

79.   Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T. A., Skorpen, F., and Krokan, H. E. (1997) Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res 25*, 750-755.

80.   Nagelhus, T. A., Haug, T., Singh, K. K., Keshav, K. F., Skorpen, F., Otterlei, M., Bharati, S., Lindmo, T., Benichou, S., Benarous, R., and Krokan, H. E. (1997) A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to XPA interacts with the C-terminal part of the 34-kDa subunit of replication protein A. *J Biol Chem 272*, 6561-6566.

81.   Otterlei, M., Warbrick, E., Nagelhus, T. A., Haug, T., Slupphaug, G., Akbari, M., Aas, P. A., Steinsbekk, K., Bakke, O., and Krokan, H. E. (1999) Post-replicative base excision repair in replication foci. *Embo J 18*, 3834-3844.

82.   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.

83.   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. *Embo J 17*, 5214-5226.

84.   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.

85.   Neddermann, P., and Jiricny, J. (1994) Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci U S A 91*, 1642-1646.

86.   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.

87.   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. *Mol Cell 11*, 1647-1659.

88.   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. *Curr Biol 9*, 174-185.

89. 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 Res 32*, 5291-5302.

90. Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H. E., Lindahl, T., and Barnes, D. E. (2000) Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol Cell 5*, 1059-1065.

91. Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., and Lindahl, T. (2001) Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *Embo J 20*, 4278-4286.

92. Warbrick, E. (1998) PCNA binding through a conserved motif. *Bioessays 20*, 195-199.

93. Di Noia, J. M., Rada, C., and Neuberger, M. S. (2006) SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *Embo J 25*, 585-595.

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

95. 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. *J Biol Chem 277*, 16936-16940.

96. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) Endonuclease III is an iron-sulfur protein. *Biochemistry 28*, 4450-4455.

97. 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. *J Mol Biol 333*, 515-526.

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

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

100. Kunkel, T. A., and Erie, D. A. (2005) DNA mismatch repair. *Annu Rev Biochem 74*, 681-710.

101. Zhang, Y., Yuan, F., Presnell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L., and Li, G. M. (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell 122*, 693-705.

102. Genschel, J., and Modrich, P. (2003) Mechanism of 5'-directed excision in human mismatch repair. *Mol Cell 12*, 1077-1086.

103. Guarne, A., Ramon-Maiques, S., Wolff, E. M., Ghirlando, R., Hu, X., Miller, J. H., and Yang, W. (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *Embo J 23*, 4134-4145.

104. Kadyrov, F. A., Dzantiev, L., Constantin, N., and Modrich, P. (2006) Endonucleolytic Function of MutL[alpha] in Human Mismatch Repair. *Cell 126*, 297-308.

105. Moolenaar, G. F., Franken, K. L., Dijkstra, D. M., Thomas-Oates, J. E., Visse, R., van de Putte, P., and Goosen, N. (1995) The C-terminal region of the UvrB protein of Escherichia coli contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. *J Biol Chem 270*, 30508-30515.

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

107. Kiyohara, C., and Yoshimasu, K. (2007) Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: a meta-analysis. *Int J Med Sci 4*, 59-71.

108. De Ruyck, K., Szaumkessel, M., De Rudder, I., Dehoorne, A., Vral, A., Claes, K., Velghe, A., Van Meerbeeck, J., and Thierens, H. (2007) Polymorphisms in base-excision repair and nucleotide-excision repair genes in relation to lung cancer risk. *Mutat Res 631*, 101-110.

109. Ismail, I. H., Nystrom, S., Nygren, J., and Hammarsten, O. (2005) Activation of ataxia telangiectasia mutated by DNA strand break-inducing agents correlates closely with the number of DNA double strand breaks. *J Biol Chem 280*, 4649-4655.

110. Yao, M., and Kow, Y. W. (1994) Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3'-endonuclease from Escherichia coli. *J Biol Chem 269*, 31390-31396.

111.   Yao, M., and Kow, Y. W. (1996) Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from Escherichia coli. *J Biol Chem 271*, 30672-30676.

112.   Yao, M., and Kow, Y. W. (1997) Further characterization of Escherichia coli endonuclease V. Mechanism of recognition for deoxyinosine, deoxyuridine, and base mismatches in DNA. *J Biol Chem 272*, 30774-30779.

113.   Gates, F. T., 3rd, and Linn, S. (1977) Endonuclease V of Escherichia coli. *J Biol Chem 252*, 1647-1653.

114.   Demple, B., and Linn, S. (1982) On the recognition and cleavage mechanism of Escherichia coli endodeoxyribonuclease V, a possible DNA repair enzyme. *J Biol Chem 257*, 2848-2855.

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

116.   Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Deoxyinosine 3' endonuclease, a novel deoxyinosine-specific endonuclease from Escherichia coli. *Ann N Y Acad Sci 726*, 315-316.

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

118.   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.

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

120.   Saparbaev, M., and Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proc Natl Acad Sci U S A 91*, 5873-5877.

121.   Modrich, P., and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem 65*, 101-133.

122.   Weiss, B. (2008) Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation. *DNA Repair (Amst) 7*, 205-212.

123.   Huang, J., Lu, J., Barany, F., and Cao, W. (2001) Multiple cleavage activities of endonuclease V from Thermotoga maritima: recognition and strand nicking mechanism. *Biochemistry 40*, 8738-8748.

124.   Lieb, M., Allen, E., and Read, D. (1986) Very short patch mismatch repair in phage lambda: repair sites and length of repair tracts. *Genetics 114*, 1041-1060.

125.   Guo, G., and Weiss, B. (1998) Endonuclease V (nfi) mutant of Escherichia coli K-12. *J Bacteriol 180*, 46-51.

126.   Schouten, K. A., and Weiss, B. (1999) Endonuclease V protects Escherichia coli against specific mutations caused by nitrous acid. *Mutat Res 435*, 245-254.

127.   Weiss, B. (2001) Endonuclease V of Escherichia coli prevents mutations from nitrosative deamination during nitrate/nitrite respiration. *Mutat Res 461*, 301-309.

128.   Moe, A., Ringvoll, J., Nordstrand, L. M., Eide, L., Bjoras, M., Seeberg, E., Rognes, T., and Klungland, A. (2003) Incision at hypoxanthine residues in DNA by a mammalian homologue of the Escherichia coli antimutator enzyme endonuclease V. *Nucleic Acids Res 31*, 3893-3900.

129.   Engelward, B. P., Weeda, G., Wyatt, M. D., Broekhof, J. L., de Wit, J., Donker, I., Allan, J. M., Gold, B., Hoeijmakers, J. H., and Samson, L. D. (1997) Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proc Natl Acad Sci U S A 94*, 13087-13092.

130.   Hang, B., Singer, B., Margison, G. P., and Elder, R. H. (1997) Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1,N6-ethenoadenine and hypoxanthine but not of 3,N4-ethenocytosine or 8-oxoguanine. *Proc Natl Acad Sci U S A 94*, 12869-12874.

131.   Dianov, G., and Lindahl, T. (1991) Preferential recognition of I.T base-pairs in the initiation of excision-repair by hypoxanthine-DNA glycosylase. *Nucleic Acids Res 19*, 3829-3833.

132.   Liu, X., and Roy, R. (2002) Truncation of amino-terminal tail stimulates activity of human endonuclease III (hNTH1). *J Mol Biol 321*, 265-276.

133.   Sreevatsan, S., Pan, X., Stockbauer, K. E., Williams, D. L., Kreiswirth, B. N., and Musser, J. M. (1996) Characterization of rpsL and rrs mutations in streptomycin-resistant Mycobacterium tuberculosis isolates from diverse geographic localities. *Antimicrob Agents Chemother 40*, 1024-1026.

134. Morris, S., Bai, G. H., Suffys, P., Portillo-Gomez, L., Fairchok, M., and Rouse, D. (1995) Molecular mechanisms of multiple drug resistance in clinical isolates of Mycobacterium tuberculosis. *J Infect Dis 171*, 954-960.

135. Timms, A. R., Steingrimsdottir, H., Lehmann, A. R., and Bridges, B. A. (1992) Mutant sequences in the rpsL gene of Escherichia coli B/r: mechanistic implications for spontaneous and ultraviolet light mutagenesis. *Mol Gen Genet 232*, 89-96.

136. Rand, T. A., Ginalski, K., Grishin, N. V., and Wang, X. (2004) Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. *Proc Natl Acad Sci U S A 101*, 14385-14389.

137. Lin, J. J., and Sancar, A. (1992) Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. *J Biol Chem 267*, 17688-17692.

138. Lin, J. J., Phillips, A. M., Hearst, J. E., and Sancar, A. (1992) Active site of (A)BC excinuclease. II. Binding, bending, and catalysis mutants of UvrB reveal a direct role in 3' and an indirect role in 5' incision. *J Biol Chem 267*, 17693-17700.

139. Aravind, L., Walker, D. R., and Koonin, E. V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res 27*, 1223-1242.

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

141. Ames, B. N., Shigenaga, M. K., and Hagen, T. M. (1993) Oxidants, antioxidants, and the degenerative diseases of aging. *Proc Natl Acad Sci U S A 90*, 7915-7922.

142. Satoh, M. S., Jones, C. J., Wood, R. D., and Lindahl, T. (1993) DNA excision-repair defect of xeroderma pigmentosum prevents removal of a class of oxygen free radical-induced base lesions. *Proc Natl Acad Sci U S A 90*, 6335-6339.

143. Eveno, E., Bourre, F., Quilliet, X., Chevallier-Lagente, O., Roza, L., Eker, A. P., Kleijer, W. J., Nikaido, O., Stefanini, M., Hoeijmakers, J. H., and et al. (1995) Different removal of ultraviolet photoproducts in genetically related xeroderma pigmentosum and trichothiodystrophy diseases. *Cancer Res 55*, 4325-4332.

144. Lunn, R. M., Helzlsouer, K. J., Parshad, R., Umbach, D. M., Harris, E. L., Sanford, K. K., and Bell, D. A. (2000) XPD polymorphisms: effects on DNA repair proficiency. *Carcinogenesis 21*, 551-555.

145. Vermeulen, W., Jaeken, J., Jaspers, N. G., Bootsma, D., and Hoeijmakers, J. H. (1993) Xeroderma pigmentosum complementation group G associated with Cockayne syndrome. *Am J Hum Genet 53*, 185-192.

146. Sarasin, A., and Stary, A. (1997) Human cancer and DNA repair-deficient diseases. *Cancer Detect Prev 21*, 406-411.

147. Stefanini, M., Vermeulen, W., Weeda, G., Giliani, S., Nardo, T., Mezzina, M., Sarasin, A., Harper, J. I., Arlett, C. F., Hoeijmakers, J. H., and et al. (1993) A new nucleotide-excision-repair gene associated with the disorder trichothiodystrophy. *Am J Hum Genet 53*, 817-821.

148. Hattori, Y., Nishigori, C., Tanaka, T., Uchida, K., Nikaido, O., Osawa, T., Hiai, H., Imamura, S., and Toyokuni, S. (1996) 8-hydroxy-2'-deoxyguanosine is increased in epidermal cells of hairless mice after chronic ultraviolet B exposure. *J Invest Dermatol 107*, 733-737.

149. Wulff, B. C., Schick, J. S., Thomas-Ahner, J. M., Kusewitt, D. F., Yarosh, D. B., and Oberyszyn, T. M. (2008) Topical treatment with OGG1 enzyme affects UVB-induced skin carcinogenesis. *Photochem Photobiol 84*, 317-321.

150. Weiss, J. M., Goode, E. L., Ladiges, W. C., and Ulrich, C. M. (2005) Polymorphic variation in hOGG1 and risk of cancer: a review of the functional and epidemiologic literature. *Mol Carcinog 42*, 127-141.

151. Planck, M., Koul, A., Fernebro, E., Borg, A., Kristoffersson, U., Olsson, H., Wenngren, E., Mangell, P., and Nilbert, M. (1999) hMLH1, hMSH2 and hMSH6 mutations in hereditary non-polyposis colorectal cancer families from southern Sweden. *Int J Cancer 83*, 197-202.

152. O'Brien, P. J. (2006) Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev 106*, 720-752.

153. Kultz, D. (2003) Evolution of the cellular stress proteome: from monophyletic origin to ubiquitous function. *J Exp Biol 206*, 3119-3124.

154. Maresca, B., and Schwartz, J. H. (2006) Sudden origins: a general mechanism of evolution based on stress protein concentration and rapid environmental change. *Anat Rec B New Anat 289*, 38-46.

CHAPTER TWO

INSIGHTS FROM XANTHINE AND URACIL DNA GLYCOSYLASE ACTIVITIES

OF BACTERIAL AND HUMAN SMUG1: SWITCHING SMUG1 TO UNG

1. Summary

Single-strand-selective monofunctional uracil DNA glycosylase (SMUG1) belongs to Family 3 of the uracil DNA glycosylase superfamily. Here, we report that a bacterial SMUG1 ortholog in *Geobacter metallireducens* (Gme) and the human SMUG1 enzyme are not only uracil DNA glycosylases (UDG) but also xanthine DNA glycosylases (XDG). In addition, mutational analysis and molecular dynamics (MD) simulations of Gme SMUG1 identify important structural determinants in conserved motifs 1 and 2 for XDG and UDG activities. Mutations at M57 (M57L) and H210 (H210G, H210M, H210N), both of which are involved in interactions with C2 carbonyl oxygen in uracil or xanthine, cause substantial reductions in XDG and UDG activities. Increased selectivity is achieved in the A214R mutant of Gme SMUG1, which corresponds to a position involved in base flipping. This mutation results in an activity profile resembling a human SMUG1-like enzyme as exemplified by the retention of UDG activity on mismatched base pairs and weak XDG activity. MD simulations indicate that M57L increases the flexibility of the motif 2 loop region and specifically A214, which may account for the reduced catalytic activity. G60Y completely abolishes XDG and UDG activity, which is consistent with a modeled structure in which G60Y blocks the entry of either xanthine or uracil to the base binding pocket. Most interestingly, a

proline substitution at the G63 position switches the Gme SMUG1 enzyme to an exclusive uracil DNA glycosylase as demonstrated by the uniform excision of uracil in both double-stranded and single-stranded DNA and the complete loss of XDG activity. MD simulations indicate that a combination of a reduced free volume as well as altered flexibility in the active site loops may underlie the dramatic effects of the G63P mutation on the activity profile of SMUG1.   This study offers insights on the important role that modulation of conformational flexibility may play in defining specificity and catalytic efficiency.

2. Introduction

DNA bases adenine (A), cytosine (C), and guanine (G) are subject to deamination caused by endogenous and environmental agents (*1-7*). Hypoxanthine (I) and uracil (U) are generated by deamination of adenine and cytosine, respectively (*2, 3*). Treatment of deoxyguanosine or DNA with nitrous acid, nitric oxide, or 1-nitrosoindole-3-acetonitrile yields xanthine (X) and oxanine (O) (*8, 9*). The deaminated base damage may cause mutations if they are not removed by DNA repair systems (*10-17*).

The frequently generated deamination product uracil is removed by DNA glycosylases (*18*). The uracil DNA glycosylase superfamily is classified into five families based on conserved motifs and structural similarity (*19, 20*). A common structural feature of the UDG superfamily is a four-stranded β-sheet surrounded by α-helices. Family 1 UDGs (UDG or UNG), as represented by *Escherichia coli* (*E. coli*), human, and herpes simplex virus 1 UDGs, are highly conserved enzymes with exquisite specificity toward uracil in both double-stranded (ds) and single-stranded (ss) DNA. Family 2 is comprised of human thymine DNA glycosylase (TDG), *E. coli* mismatch-specific uracil-DNA glycosylase (MUG), and a broad substrate specificity fission yeast *Schizosaccromyces pombe* TDG (*21, 22*). Family 3 enzymes were previously thought to be eukaryotic-specific UDG's as represented by African clawed frog *Xenopus laevis* SMUG1 (single-strand-selective monofuntional uracil-DNA glycosylase) and human SMUG1. Family 4 UDG enzymes are a group of prokaryotic iron-sulfur-containing enzymes that act on both single-stranded and double-stranded uracil-containing DNA (*23-26*). Family 5 UDG enzymes are found in a small number of prokaryotic species,

51

one of which, from the hyperthermophilic crenarchaeon *Pyrobacculum aerophilum,* demonstrates glycosylase activity toward G/U and to a lesser degree toward T/I substrates (*27*).

Xanthine is a stable lesion in DNA under physiological conditions, indicating the need for repair (*15, 28*). The content of xanthine in DNA increases when exposed to reactive nitrogen species such as nitric oxide (*2, 29-31*). *E. coli* AlkA was reported to possess xanthine DNA glycosylase activity, while the possible XDG activity of *E. coli* endo VIII remains controversial (*32, 33*). More importantly, endonuclease V initiates a repair pathway to remove xanthine from DNA in bacteria. Biochemical analyses have detected deoxyxanthosine endonuclease activities in several bacterial endonuclease V homologs (*32-35*). In an earlier study, repair of xanthine lesions was observed in human lymphoblast cells (*36*). Later, it was found that human alkyladenine DNA glycosylase (hAAG) exhibited strong xanthine DNA glycosylase activity ((*15, 33*), L. Dong & W. Cao, unpublished data). hAAG is active toward xanthine in both single-stranded and double-stranded DNA regardless of the opposite base (L. Dong & W. Cao, unpublished data).

SMUG1 was initially discovered as a uracil DNA glycosylase by screening proteins that bound to transition state analogs of monofunctional DNA glycosylases (*37*). In mammalian cells, SMUG1 is responsible for removal of uracil generated from cytosine deamination in premutagenic G/U base pairs (*38-40*). In addition to excision of uracil in double-stranded and single-stranded DNA, hSMUG1 also removes 5-formyluracil (fU), 5-hydroxyuracil (hoU), 5-hydroxymethyluracil (hmU), and 3,N^4-ethenocytosine (*41, 42*).

A crystal structure of a *Xenopus laevis* SMUG1-DNA complex reveals how hmU is accommodated while thymine is excluded from the active site   (*43*).

SMUG1 was considered a eukaryote-only uracil DNA glycosylase since homologs were found previously in vertebrates and insects. A search of genome databases showed the existence of SMUG1 orthologs in several bacteria including *Geobacter metallireducens* (Gme), *Azoarcus species* (Asp), *Rhodopirellula baltica* (Rba) and *Opitutaceae bacterium* (Oba) (Fig. 2.1).   This study investigated deaminated repair activity in Gme SMUG1 and XDG activity in hSMUG1.   As expected, bacterial Gme SMUG1 is a uracil DNA glycosylase.   Surprisingly, Gme SMUG1 is also a xanthine DNA glycosylase (XDG) that removes xanthine in both double-stranded and single-stranded DNA.   Interestingly, human SMUG1 is also found to be active toward xanthine, indicating that XDG activity is universal in both prokaryotic and eukaryotic SMUG1.   This work, along with a previous report on oxanine DNA glycosylase activity from SMUG1 (*44*), extends the substrate specificity of SMUG1 from pyrimidine deamination/oxidation products to purine deamination products.   How the active site of Gme SMUG1 accommodates both uracil and xanthine was investigated by site-directed mutagenesis and molecular modeling. Through a single amino acid change, G63P switches Gme SMUG1 to a Family 1 UDG- or UNG-like exclusive uracil DNA glycosylase.

```
                          Motif 1                    Motif 2
                           M64G
                           G63P
                           W62F                       A214R
                          G60Y                       H210N
                         N58D                        H210M
                        M57L                         H210G
                        ▼▼ ▼ ▼▼▼                      ▼     ▼
Family 3     Gme   N--55-GMNPGPWGMAQTGVP-F-64-N-73-HPSPASP--21-C
(SMUG1)      Asp   N--57-GMNPGPFGMTQTGVP-F-64-N-72-HPSPASP--23-C
             Rba   N--68-GMNPGPWGMAQSGVP-F-64-N-76-HPSPASP--20-C
             Oba   N--58-GMNPGPFGMAQTGVP-F-64-N-76-HPSPASP--22-C
             Spu   N-115-GMNPGPFGMAQNGVP-F-64-N-76-HPSPINP--25-C
             Hsa   N--82-GMNPGPFGMAQTGVP-F-64-N-75-HPSPRNP--25-C
             Mmu   N--84-GMNPGPFGMAQTGVP-F-64-N-75-HPSPRSA--32-C
             Xla   N--93-GMNPGPFGMAQTGVP-F-64-N-75-HPSPRNP--25-C
             Dme   N--91-GMNPGPNGMAQTGIP-F-64-N-75-HPSPRST--26-C
             Ame   N-105-GMNPGPWGMSQTGVP-F-64-N-75-HPSPRAV--24-C
             Tca   N--77-GMNPGPFGMCQTGVP-F-64-N-73-HPSPRSK--31-C


Family 1     Eco   N--61-GQDPYHGPGQAHGLA-F-45-N-63-HPSPLSA--36-C
(UDG)        Dra   N—-80-GQDPYHGPNQAHGLS-F-45-N-63-HPSPLSE--35-C
             Mtu   N--65-GQDPYPTPGHAVGLS-F-45-N-63-HPSPLSA--30-C
             Hsa   N-142-GQDPYHGPNQAHGLC-F-45-N-63-HPSPLSV--30-C
             Mmu   N-133-GQDPYHGPNQAHGLC-F-45-N-63-HPSPLSV--30-C
             Xla   N-142-GQDPYHGPNQAHGLC-F-45-N-63-HPSPLSV--30-C
             HSV1  N--85-GQDPYHHPGQAHGLA-F-45-N-62-HPSPLSK--28-C


Family 2     Eco   N--15-GINPGLSSAGTGFP--F-37-K-71-NPSGLSR--22-C
(MUG/TDG)    Bce   N--18-GINPGLIAAATGHH--F-37-A-71-NPSGRNL--27-C
             Dra   N--31-GTAPSGISARARAY--Y-38-D-74-STSPLGH--34-C
             Swi   N--38-GSLPGEASLRAARY--Y-30-R-73-SSSPAFT--26-C
             Csp   N--20-GSLPGEASLAVQQY--Y-29-A-71-SSSPAHA--23-C
             Dge   N--27-GTAPSRISARAKAY--Y-37-D-71-STSPLGH--37-C
             Acl   N-175-GVNPGILTGTTGFA--Y-38-N-87-TTSGLAA--42-C
             Spo   N-155-GLNPGITSSLKGHA--F-39-N-77-GISSSGR--31-C
             Hsa   N-137-GINPGLMAAYKGHH--Y-38-N-77-MPSSSAR-135-C
             Dme   N-792-GINPGLFAAYKGHH--Y-37-N-71-MPSSSAR-815-C


Family 4     Pae   N--39-GEAPGASEDEAGRP--F-25-N-81-HPAAVLR--28-C
(UDGa)       Dra   N--60-GEGPGAEEDRDGRP--F-25-N-80-HPAYLLR--49-C (DR 1751)
             Dra   N--42-LEAPGPQAAQSRGGSGF-30-N-68-HPSGQAL--34-C (DR 0022)
             Tma   N--44-GEGPGEEEDKTGRP--F-25-N-75-HPSYLLR--25-C
             Nmu   N-123-GEGPGAQEDALGEP--F-26-N-73-HPAYLLR--26-C
             Tth   N--39-GEGPGEEEDKTGRP--F-25-N-74-HPAYLLR--44-C


Family 5     Pae   N--65-GLAPAAHGGNRTGRM-F-39-S-74-HPSPLNV--23-C
(UDGb)       Sso   N--46-GLAPAGNGGNRTGRM-F-39-S-85-HPSPRNM--25-C
             Tvo   N--61-GLAPAATGGNRTGRV-F-39-A-80-HPSPRNV--23-C
             Sco   N--60-GLAPAAHGGNRTGRM-F-39-S-88-HVSQRNT--26-C
             Mtu   N-101-GLAPAAHGANRTGRM-F-39-S-80-HPSQQNM--24-C
             Tth   N--56-GLAPGAHGSNRTGRP-F-38-A-78-HVSRQNT--23-C
```

Figure 2.1 Sequence alignment of bacterial and eukaryotic SMUG1.   Genbank accession numbers are shown after the species names.   Family 3 (SMUG1): Gme, *Geobacter metallireducens* GS-15,YP_383069; Asp, *Azoarcus sp.* BH72, YP_935478; Rba, *Rhodopirellula baltica* SH 1, NP_869403; Oba, *Opitutaceae bacterium* TAV2, ZP_02013615.1; Spu, *Strongylocentrotus purpuratus*, XP_782746.1; Hsa, *Homo sapiens*,

NP_055126; Mmu, *Mus musculus*, NP_082161; Xla, *Xenopus laevis*, AAD17300; Dme, *Drosophila melanogastser*, NP_650609.1; Ame, *Apis mellifera*, XP_396883.2; Tca, *Tribolium castaneum*, XP_971699.1. Family 1 (UDG or UNG): Eco, *Escherichia coli*, NP_289138; Dra, *Deinococcus radiodurans* R1, NP_294412; Mtu, *Mycobacterium tuberculosis* H37Rv, CAB05436.1; Hsa, *Homo sapiens*, NP_003353; Mmu, *Mus musculus*, NP_035807; Xla, *Xenopus laevis*, NP_001085412; HSV1, *Herpes Simplex Virus-1*, 1UDI. Family 2 (MUG/TDG): Eco, *Escherichia coli*, P0A9H1; Bce, *Burkholderia cenocepacia* HI2424, YP_836419; Dra, *Deinococcus radiodurans* R1, NP_294438; Swi; *Sphingomonas wittichii* RW1, ZP_01607068; Csp, *Caulobacter sp.* K31, ZP_01418424.1; Dge, *Deinococcus geothermalis* DSM 11300, YP_605182.1; Acl, *Aspergillus clavatus* NRRL 1, XP_001268386.1; Spo, *Schizosaccharomyces pombe*, O59825; Hsa, *Homo sapiens*, NP_003202; Dme, *Drosophila melanogaster*, CAB93525; Xla, *Xenopus laevis*, AAH77465. Family 4 (UDGa): Pae, *Pyrobaculum aerophilum* str. IM2, NP_558739.1; Dra (DR 1751), *Deinococcus radiodurans* R1, NP_295474; Dra (DR 0022), *Deinococcus radiodurans* R1, AAF09614; Tma, *Thermotoga maritima* MSB8, NP_228321.1; Nmu, *Nitrosospira multiformis*, YP_412806; Tth, *Thermus thermophilus* HB27, YP_004341.1. Family 5 (UDGb): Pae, *Pyrobaculum 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, *Mycobacterium tuberculosis* H37Rv, P64785 (Rv1259); Tth, *Thermus thermophilus* HB27, YP_004757.1

3. 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, *Taq* DNA polymerase and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). BSA and dNTPs were purchased from Promega (Madison, WI). HiTrap chelating and Q columns were purchased from Amersham-Pharmacia Biotech (Piscataway, NJ). Oligodeoxyribonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA). The LB medium was prepared according to standard recipes. The Gme SMUG1 sonication buffer consisted of 50 mM HEPES-KOH (pH 7.4), 1 mM EDTA (pH 8.0), 2.5 mM DTT, 0.15 mM PMSF, 10% glycerol and 50 mM NaCl.

The GeneScan stop buffer consisted of 80% formamide (Amresco, Solon, OH), 50 mM EDTA (pH 8.0), and 1% blue dextran (Sigma Chemicals). The TB buffer (1 x) consisted of 89 mM Tris base and 89 mM boric acid. The TE buffer consisted of 10 mM Tris-HCl (pH 8.0), and 1 mM EDTA. The *E. coli* host 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*::Tn*10, mug*::Tn*10, supE44*, (DE3)] was a kind gift of Dr. Ashok Bhagwat (Wayne state university, Detroit, MI) and JM109 [e14⁻(McrA⁻) *endA1*, *recA1*, *gyrA96*, *thi-1*, *hsdR17*(r_k^-, m_k^+), *supE44*, *relA1* Δ(*lac-proAB*), [F', *traD36*, *proAB*, *lacI*^{q}ZΔM15]] was from our laboratory collection. The genomic DNA from *Geobacter Metallireducens* GS-15 was a kind gift of Dr. Derek R. Lovley and Muktak Aklujkar (University of Massachusetts, Amherst, MA). The Human SMUG 1 protein was a kind gift of Drs. Geir Slupphaug and Bodil Kavli (Norwegian University of Science and Technology, Trondheim, Norway).

Plasmid Construction, Cloning and Expression of Gme SMUG1.

The putative SMUG1 gene from *G. Metallireducens* GS-15 was amplified by PCR using the forward primer Gme.smug1.01F (5' TAA GGT ACC CC ATG GCC GGT CTC GCG GCT ATT TCC 3'; the *Nco*I site is underlined) and the reverse primer Gme.smug1.02R (5' TGA ACG CGT GGA TCC TCA GAA ATC GAC ACC AAG CTC CG 3'; the *BamH*I site is underlined). The PCR reaction mixture (50 µL) consisted of 10 ng of *G. Metallireducens* GS-15 genomic DNA, 200 nM forward primer Gme.smug1.01F and reverse primer Gme.smug1.02R, 1 x *Taq* PCR buffer (New England

Biolabs), 200 μM each dNTP, and 2 units of Phusion DNA polymerase (New England Biolabs). The PCR procedure included a predenaturation step at 94°C for 4 min, 30 cycles of three-step amplification with each cycle consisting of denaturation at 94°C for 15 s, annealing at 47°C for 30 s and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR product was purified with Gene Clean 2 Kit (Qbiogene). The purified PCR product and plasmid pET32a were digested with *Nco*I and *BamH*I, purified with Gene Clean 2 Kit and ligated according to the manufacturer's instructional manual. The ligation mixture was transformed into *E. coli* strain JM109 competent cells prepared by a CaCl$_2$ method. The sequence of the Gme smug1 gene in the resulting plasmid (pET32a-Gsmug1) was confirmed by DNA sequencing.

To express the N-terminal His-6-tagged Gme SMUG1 protein, pET32a-Gsmug1 was transformed into *E.coli* strain BH214 by standard protocol. An overnight *E. coli* culture containing pET32a-Gsmug1/BH214 was diluted 100-fold into LB medium (1 Liter) supplemented with 50 μg/mL ampicillin. The *E. coli* cells were grown at 37°C while being shaken at 250 rpm until the optical density at 600 nm reached approximately 0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was added to a final concentration of 0.5 mM. The culture was grown at room temperature for an additional 16 h. The cells were collected by centrifugation at 4,000 rpm with JS-4.2 rotor in J6-MC centrifuge (Beckman Coulter) at 4°C and washed once with precooled sonication buffer.

To purify the Gme SMUG1 protein, the cell paste from the 1 L culture was suspended in 10 mL of sonication buffer and sonicated at output 5 for 3 x 1 min with 5 min rest on ice between intervals. The sonicated solution was clarified by centrifugation

at 12,000 rpm with JA-17 rotor in Avanti J-25 centrifuge (Beckman Coulter) at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded into a 1 mL HiTrap chelating column. The bound protein in the column was eluted with a linear gradient of 10 column volumes of 0-1 M imidazole in chelating buffer B [20 mM Tris-HCl (pH 7.6), 10% glycerol and 50 mM NaCl] using a Bio-Rad BioLogic chromatographic system.

Fractions (200-400 mM imidazole) containing the Gme SMUG1 protein as seen on 15% SDS-PAGE were pooled and dialyzed against HiTrap Q buffer A [20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 10% glycerol and 0.2 mM DTT] overnight at 4°C. The dialysis sample was then loaded onto a 1 mL HiTrap Q column and eluted with a linear gradient of 10 column volumes of 0-1 M of NaCl in HiTrap Q buffer B (HiTrap Q buffer A containing 1 M NaCl). The putative Gme SMUG1 protein was eluted at 200 - 400 mM NaCl. The homogeneity of the protein was examined by 12.5% SDS-PAGE analysis and found to be greater than 90% pure. The Gme SMUG1 protein concentration was determined by SDS-PAGE using BSA as a standard.

Site-Directed Mutagenesis.

An overlapping extension PCR procedure was used for site-directed mutagenesis. The construction of the M57L mutant is described as an example. The first round of PCR was carried out using pET32a-Gsmug1 as the template DNA with two pairs of primers, Gme.smug1.01F and GSM57L.R (5' TCC TTT TCG TCG GC<u>C</u> <u>TGA</u> ACC CCG GCC CCT GGG GGA T 3', the M57L site is underlined) pair & Gme.smug1.02R

and GSM57L.F (5' CAG GGG CCG GGG TTC AGG CCG ACG AAA AGG ACT TCC

T 3', the M57L site is underlined) pair.   The PCR mixtures (50 µL) contained 1 ng of

pET32a-Gsmug1 DNA as template, 200 nM of each primer pair, 50 µM of each dNTP, 1

× Phusion DNA polymerase buffer, and 1 unit of *Taq* DNA polymerase (New England

Biolabs).   The PCR procedure involved a predenaturation step at 95°C for 2 min, 30

cycles of three-step amplification with each cycle consisting of denaturation at 94°C for

15 s, annealing at 49°C for 30 s and extension at 72°C for 1 min, and a final extension

step at 72°C for 10 min.   The PCR products were electrophoresed on 1% agarose gel

and the expected PCR fragments were purified from gel slices by spin-squeeze method.

The second run of the PCR reaction mixture (100 µL), which contained 3 µL of each of

the first run PCR fragments, 50 µM of each dNTP, 1 × *Taq* DNA polymerase buffer, and

2 units of Phusion DNA polymerase (New England Biolabs), was initially carried out

with a predenaturation step at 95°C for 2 min, five cycles with each cycle of denaturation

at 94°C for 15 s and annealing at 55°C for 30 s and extension at 72°C for 2 min, and a

final extension at 72°C for 5 min.   Afterward, 100 nM of outside primers

(Gme.smug1.01F and Gme.smug1.02R) were added to the above PCR reaction mixture.

The subsequent overlapping PCR amplification included a predenaturation step at 95°C

for 2 min, 25 cycles with each cycle of denaturation at 94°C for 15 s and annealing at

49°C for 1.5 min and extension at 72°C for 2 min, and a final extension at 72°C for 5

min.   The purified PCR products digested with a pair of *Nco*I and *Bam*HI endonucleases

were ligated to the cloning vector pET32a 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 for expression and protein purification as described above.

Oligodeoxynucleotide Substrates.

Oligonucleotides containing xanthine (X) or oxanine (O) were prepared as previously described (*44*). The sequences of the oligonucleotides are shown in Fig. 2.2A. Oligonucleotides containing hypoxanthine or uridine were ordered from IDT, purified by PAGE, and dissolved in TE buffer at a final concentration of 10 μM. The two complementary strands with the unlabeled strand in 1.2-fold molar excess were mixed, incubated at 85°C for 3 min, and allowed to form duplex DNA substrates at room temperature for more than 30 min.

DNA Glycosylase Activity Assays.

DNA glycosylase cleavage assays for Gme SMUG1 were performed under optimized reaction conditions at 33°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), 30 mM NaCl, 1 mM Dithiothreitol, 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. Reactions were quenched by addition of an equal volume of GeneScan stop buffer. After incubation at 95°C for 3 min, samples (3.2 µL) were loaded onto a 7 M urea-10% denaturing polyacrylamide gel. Electrophoresis was conducted at 1500 V for 1.5 h using an ABI 377 sequencer (Applied Biosystems). Cleavage products and remaining

substrates were quantified using the GeneScan analysis software. DNA glycosylase cleavage assays for human SMUG1 were performed under the same reaction conditions except that the reaction mixtures were incubated at 37ºC.

Homology Modeling.

A PSI-BLAST alignment of the amino acid sequence from chain A of 1oe5.pdb (*Xenopus* SMUG1) and the UDG superfamily sequence from *Geobacter metalloreducins* GS-15 (Gene ID: 3739421 Gmet_0095) resulted in a 64% similarity and 50% identity between the two sequences. In addition, a PSI-BLAST alignment was performed between the *Xenopus* sequence and that for the human SMUG1 enzyme (Gene ID: 23583 SMUG1), resulting in 78% similarity and 66% identity. Based on these sequence alignments and the 1oe5 pdb structure, homology models were constructed for the Gme and human SMUG1 enzymes using the NEST program (*45*). Coordinates for the uracil base complexed with the *Xenopus* SMUG1 enzyme were transferred directly to the coordinate files containing the homology modeled Gme and human SMUG1 enzymes. From these initially modeled complexes, models containing the deaminated guanine base xanthine were constructed. Parameters for xanthine were constructed within the molecular modeling package charmm (c32b1) by homology (*46, 47*). Within the charmm modeling package, the coordinates for the Gme SMUG1 complex with uracil were fixed. Nonbonded energy terms (Coulomb and van der Waals components) were eliminated using the "skipe" command in charmm. Distance restraints were placed between the nitrogen and oxygen atoms that are associated with the six-membered rings and

analogous between xanthine and uracil (uracil/xanthine restraints: N1/N3, N3/N1, O2/O2, O4/O6). The system was then minimized and resulted in an overlap between the six-membered ring of xanthine and the uracil molecule in the complex. The uracil coordinates were then deleted, leaving a model for xanthine bound to the bacterial Gme SMUG1 based on the initial model of SMUG1 bound to uracil. The identical procedure was carried out on the human SMUG1-uracil complex.

All model complexes were then minimized gently by first fixing the coordinates of all amino acids that do not contain atoms within 10Å of the bound substrate (uracil or xanthine). Harmonic restraints, using a force constant of 10 kcal/mol/$Å^2$, were applied to amino acids within 10Å of the bound substrate. A short adopted-basis Newton Raphson minimization of 200 steps was carried out to remove van der Waals clashes close to the substrate. Harmonic restraints were removed, a generalized Born implicit solvent was applied (GBSW algorithm in charmm) and the complex was minimized a further 1000 steps.

In order to generate mutants of the Gme SMUG1, the mutate.pl script from the MMTSB toolset (*48*) was applied to the initial Gme SMUG1 complexes (prior to minimization) bound to uracil and xanthine. The mutated complexes were then gently minimized using the protocol described above.

Molecular Dynamics Simulations.

Canonical ensemble (NVT) Molecular dynamics on the wild-type and mutant Gme SMUG1, were carried out using the charmm molecular mechanics package.

Simulations were carried out using an implicit generalized Born solvent model (the GBSW algorithm in charmm) with a 1 fs timestep. The salt concentration (modeled through a Debye-Huckel term in the GB solvent) was maintained at 0.15 M, while the non-polar surface tension coefficient (γ) and the smoothing length (sw) were assigned as 0.03 kcal/mol Å^2 and 0.3 Å, respectively. Prior to the molecular dynamics, the initial unbound Gme SMUG1 model was minimized within a generalized Born solvent under successively reduced harmonic restraints applied to all atoms (50 kcal/mol Å^2, 10, 1, 0). This was done to eliminate high-energy clashes that could impact the simulation. During the molecular dynamics run, the system was gradually heated over 20 ps from an initial temperature of 100 K to a final temperature of 300 K. The system was equilibrated for 1.11 ns and the production run was carried out for 2 ns.

Based on the production portion of each molecular dynamics calculation, isotropic root mean square fluctuations (rmsf) were calculated for each of the Gme SMUG1 mutant systems. Configurational snapshots taken every 1 ps during the production run were oriented with respect to the first production run conformation. The isotropic root mean squared fluctuations for each atom (with respect to the average of the oriented structures) were determined using the "coor dyna" command in charm (*49*). The same trajectories were used to calculate the free volume of the active site using the "coor volume" command in charmm. First, amino acids selected to represent the active site were chosen based on the initial models of Gme SMUG1 bound to xanthine (the larger of the two ligands). Any amino acid that contained an atom within 8Å of the xanthine ligand was identified as part of the binding site. Active site free volumes were

calculated for each production run snapshot based on the difference between the free volume calculated with and without the "holes" keyword.

4. Results

Bacterial SMUG1 as xanthine DNA glycosylase (XDG)

SMUG1 was previously thought as a eukaryote-only uracil DNA glycosylase, yet a search of rapidly accumulating genome databases discovered SMUG1 in four bacterial genomes (Fig. 2.1). Human SMUG1 shows a 54% amino acid sequence identity to *Geobacter metallireducens* SMUG1, 46% to *Azoarcus sp.* SMUG1, 43% to *Rhodopirellula baltica* SMUG1, and 46% to *Opitutaceae bacterium* SMUG1. In comparison, the sequence identity between hSMUG1 and fruit fly SMUG1 is 48%. Similar to their eukaryotic counterparts, bacterial SMUG1 contains the highly conserved sequences GMNPGP in motif 1 and HPSP in motif 2 (Fig. 2.1). To understand the potential role of SMUG1 in DNA repair in bacteria, we cloned and expressed SMUG1 from *Geobacter metallireducens* GS-15. Using fluorescently labeled oligonucleotide substrates (Fig. 2.2A), we assayed for the glycosylase activity of Gme SMUG1 toward all four deaminated bases, hypoxanthine (I), uracil (U), xanthine (X) and oxanine (O) (Fig. 2.2B). As expected, Gme SMUG1 exhibited robust activity in excising all U-containing substrates including the single-stranded U-containing substrate, suggesting that bacterial SMUG1 is a bona fide uracil DNA glycosylase (Fig. 2.2C). Both the adenine deamination product hypoxanthine and second guanine deamination product oxanine were not substrates for Gme SMUG1 even in our assay conditions in which the enzyme

was in excess (Fig. 2.2C).   In contrast, Gme SMUG1 demonstrated significant excision of xanthine as indicated by ~70% cleavage of all double- and single-stranded xanthine-containing substrates (Fig. 2.2C).   These results suggested that Gme SMUG1 is also a xanthine DNA glycosylase.

Table 2.1: Apparent rate constants for cleavage of uracil (U) and xanthine (X) substrates by Gme SMUG1 (min^{-1})[a]

Bottom Strand		Top Strand				
		A	C	T	G	ss[b]
WT		0.18	0.19	0.21	0.22	0.17
WT [c] (E:S=1:10)		0.048	0.051	0.063	0.18	0.015
M57L		0.13	0.16	0.17	0.18	0.12
G63P	U	0.24	0.24	0.24	0.25	0.24
H210G		0.0032	0.014	0.019	0.016	n.a.[d]
H210M		0.0027	0.013	0.015	0.010	n.a.
H210N		0.0085	0.012	0.015	0.012	n.a.
A214R		0.062	0.23	0.21	0.20	0.013
WT		0.056	0.046	0.044	0.056	0.051
M57L	X	0.011	0.013	0.011	0.017	0.013
A214R		0.012	0.016	0.012	0.013	0.014

[a]: The reactions were performed as described in Materials and Methods with 100 nM SMUG1 and 10 nM substrate, unless otherwise stated.   The apparent rate constants were determined by fitting the time course data into a first-order rate equation using Deltagraph (SPSS Inc.).   Data are an average of two independent experiments.
[b]: Single-stranded uracil- or xanthine-containing substrate.
[c]: The enzyme concentration was 1 nM and the substrate concentration was 10 nM.
[d]: n.a.: no activity was detected under assay conditions.

To dissect the kinetic differences between the UDG and XDG activities, we performed a time course analysis on five U-containing substrates and five X-containing substrates. Under the assay conditions in which the enzyme was in excess (E:S ratio = 10:1), Gme SMUG1 excised uracil from all five U-containing substrates to near completion, with an apparent rate constant of approximately 0.20 per min (Fig. 2.3A, E:S = 10:1; Table 2.1). The XDG activity from Gme SMUG1 was not as robust as UDG, as indicated by a slower cleavage of X-containing substrates and a less than 70% completion (Fig. 2.3B, E:S = 10:1). The apparent rate constant for XDG was around 0.05 per min, representing a four-fold difference (Table 2.1). Under the assay conditions in which the substrate was in excess (E:S ratio = 1:10), Gme SMUG1 excised uracil from G/U base pairs rapidly in a manner similar to that observed when the enzyme was in excess (Fig. 3A, G/U). The cleavage of the remaining U-containing substrates followed the order of T/U > C/U, A/U > single-stranded U (Fig. 2.3A, Table 2.1). These results indicated that G/U was a better substrate for the Gme SMUG1. Cleavage of the X-containing substrates was low under the assay conditions, with less than 5% of the substrate being converted to product (Fig. 2.3B, E:S = 1:10).
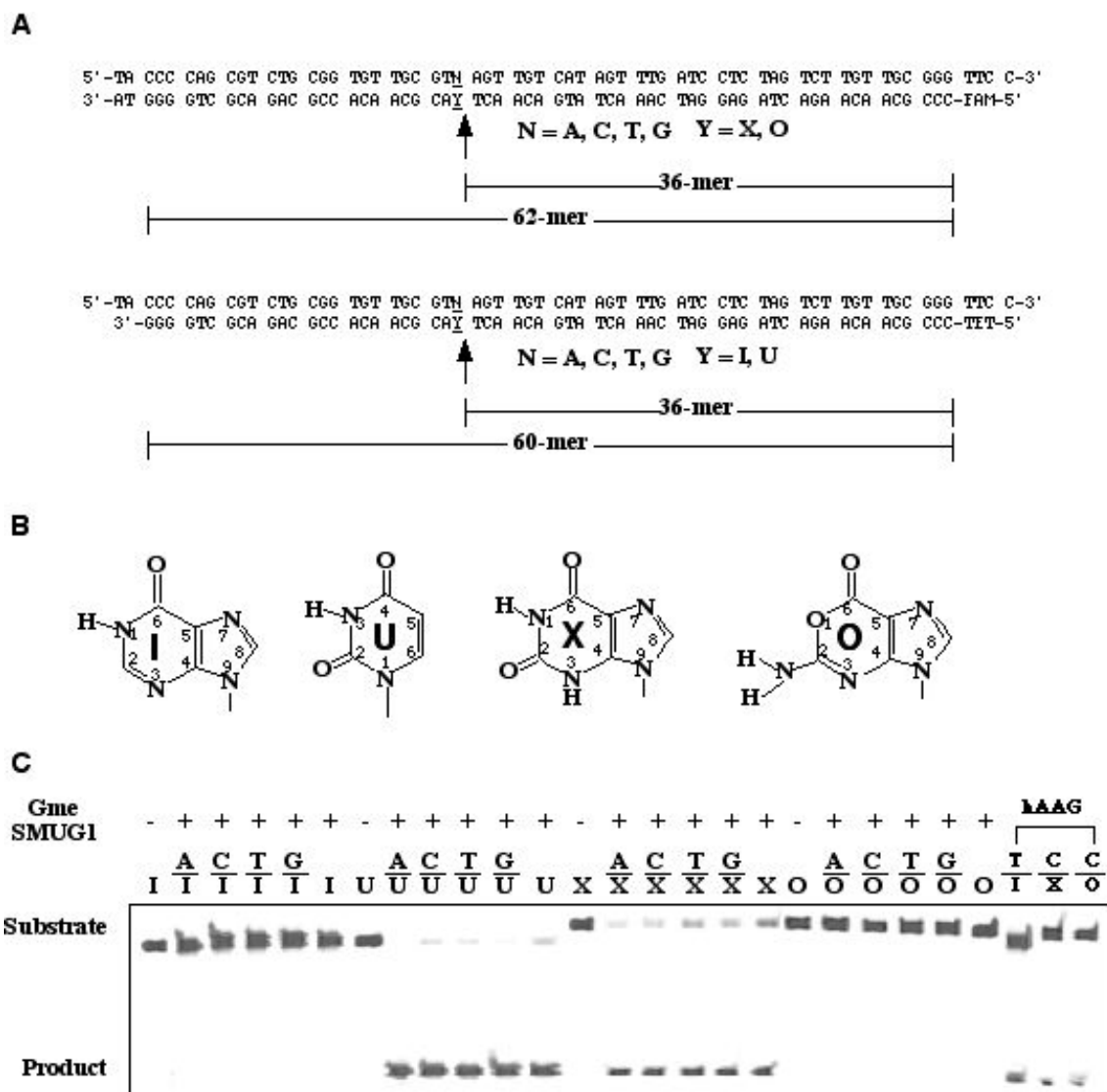
Figure 2.2 Cleavage of uracil- and xanthine-containing DNA substrates by wild-type Gme SMUG1.   A. Sequences of xanthine (X)- and oxanine (O)-, and hypoxanthine (I)- and uracil (U)-containing oligodeoxyribonucleotide substrates.   B. Chemical structures of deaminated DNA bases.   C. DNA glycosylase activity of wt Gme SMUG1 on I-, U-, X- and O-containing substrates. Cleavage reactions were performed as described in Materials and Methods with 100 nM wt Gme SMUG1 protein and 10 nM substrate. hAAG was assayed as a control with 10 nM hAAG protein and 10 nM substrate as described previously (*16*).

67
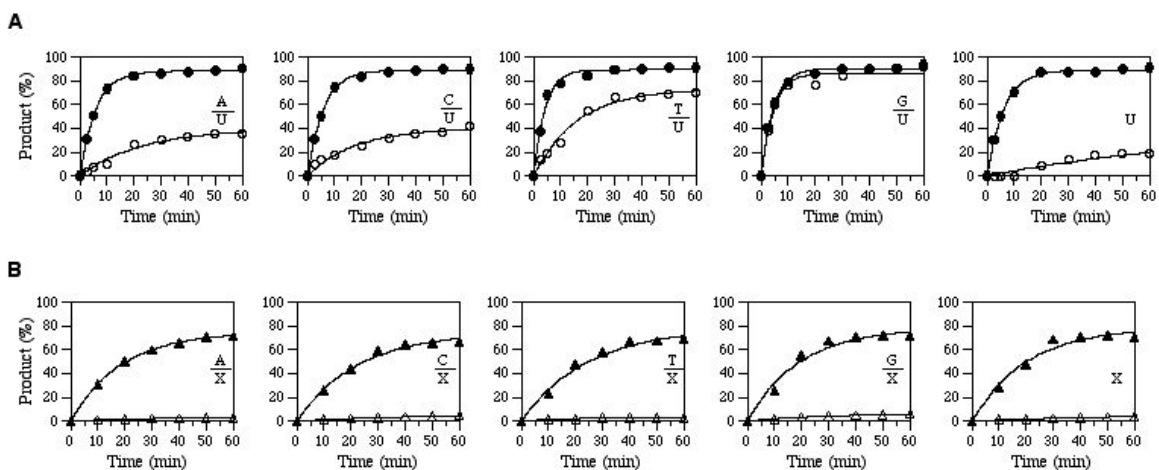
Figure 2.3 Kinetic analysis of glycosylase activity of wt Gme SMUG1 on U- and X-containing substrates. A. Time course analysis of cleavage activity on U-containing substrates. S = 10 nM. (●) E:S = 10:1; (○) E:S = 1:10. B. Time course analysis of cleavage activity on X-containing substrates. S = 10 nM. (▲) E:S = 10:1; (△) E:S = 1:10.

Human SMUG1 as xanthine DNA glycosylase (XDG)

The surprising discovery of XDG activity from Gme SMUG1 prompted us to investigate XDG activity in human SMUG1. Under the assay conditions in which the enzyme was in excess, we indeed observed cleavage of X-containing substrates by hSMUG1 (Fig. 2.4A). Similar to the bacterial enzyme, hSMUG1 cleaves both double-stranded and single-stranded X-containing DNA without regard to the opposite base (Fig. 2.4B). The XDG activity from hSMUG1 was not as robust as the one from Gme SMUG1 as judged by the percent cleavage yield. Since hSMUG1 was shown to maintain good activity on the positive control substrate (G/U), these results indicate that hSMUG1 is a less efficient enzyme for xanthine substrates (Figs. 2.3 and 2.4).
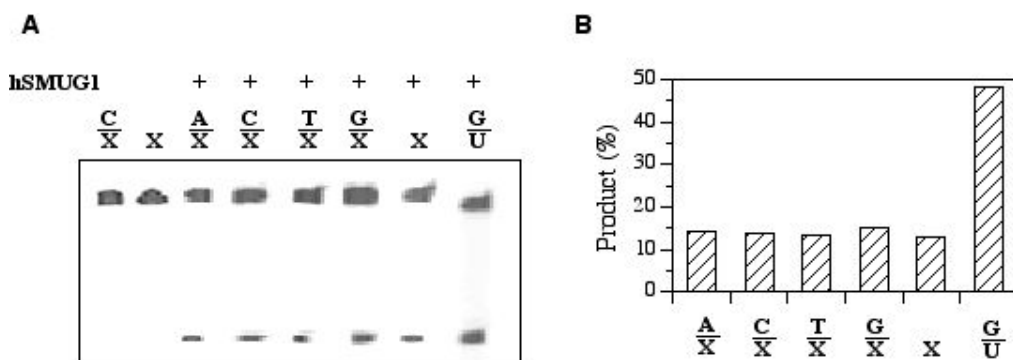
Figure 2.4 Cleavage activity of hSMUG1 on X-containing substrates. Cleavage reactions were performed as described in Materials and Methods at 37°C for 60 min with 100 nM hSMUG1 protein and 10 nM substrate. A. GeneScan gel picture of cleavage activity of hSMUG1 on X-containing substrates. B. Quantification of XDG activity from hSMUG1. Cleavage products and remaining substrates were quantified using GeneScan analysis software.

Site-directed mutagenesis of Gme SMUG1

The hallmark of the UDG superfamily is that all the glycosylases contain two conserved motifs that are involved in the recognition of deaminated bases (Fig. 2.1). Apparently, some of the amino acid residues within these motifs may play an important role in determining substrate specificity. Using the SMUG1 from *Geobacter* as a model system, we made ten site-directed mutants at eight sites in both motifs (Fig. 2.1). The rationale for choosing each site targeted in this study is described below along with a description of the activity changes associated with each corresponding mutation. To characterize these mutants, we assayed DNA cleavage activity using double-stranded and single-stranded uracil-, xanthine-, hypoxanthine-, and oxanine-containing substrates. When appropriate, detailed kinetic analyses were performed to compare the mutant Gme SMUG1 and the wild-type (wt) enzyme. Overall, neither hypoxanthine DNA glycosylase activity nor oxanine DNA glycosylase activity was detected, indicating that

none of mutants acquired additional deamination repair activity. However, significant changes in both xanthine DNA glycosylase activity and in uracil DNA glycosylase activity were observed and are described in the following sections.

M57, N58D, G60Y, W62F and M64G in motif 1

M57 is an invariant residue among SMUG1 family proteins; however, is not conserved in other UDG families (Fig. 2.1). The M57L substitution renders the Gme SMUG1 identical in this position to Family 5 UDGb proteins. Under the assay conditions in which the enzyme was in excess (E:S ratio = 10:1), M57L excised uracil from all five U-containing substrates to near completion (Fig. 2.5A-2.5B), although the efficiencies were around 15-30% lower than the wt enzyme (Table 2.1). The cleavage of X-containing substrates was substantially lower, with 20% cleavage for A/X, 36% for C/X, 30% for T/X, 42% for G/X, and 33% for single-stranded X observed when the enzyme was in excess (Fig. 2.5C). Under the assay conditions in which the substrate was in excess (E:S ratio = 1:10), Gme SMUG1 excised uracil from G/U base pair with the highest efficiency, followed by T/U > C/U > A/U (Fig. 2.5B). Cleavage of single-stranded U was less than 3% (Fig. 2.5B). No cleavage was detectable in any X-containing substrate when the substrate was in excess (data not shown).

N58 is an invariant residue in SMUG1 that was changed to aspartate in this study. N58D showed no DNA glycosylase activity toward any deaminated substrate (data not shown). It was proposed and later proved by a structural study that motif 1 in SMUG1 is more Family 2 MUG/TDG-like than Family 1 UDG- or UNG-like (*37, 43*). The
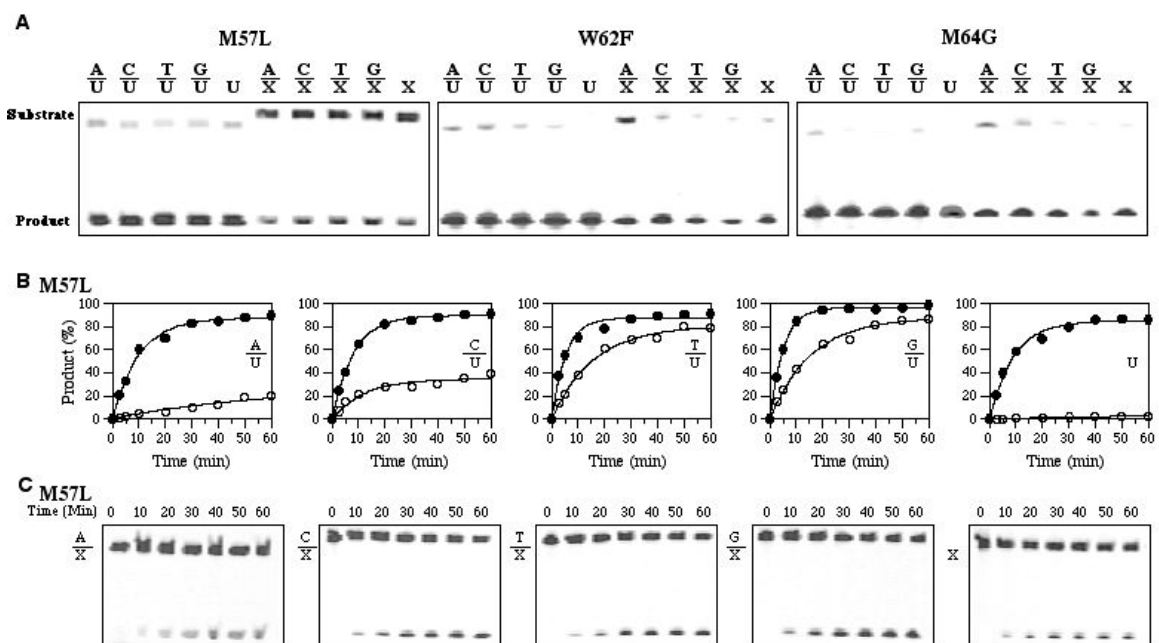
Figure 2.5 Kinetic analysis of glycosylase activity of M57L, W62F and M64G mutants of Gme SMUG1 on U- and X-containing substrates.   A. GeneScan gel pictures of cleavage activity of M57L, W62F and M64G mutants on U- and X-containing substrates.   B. Time course analysis of cleavage activity of M57L on U-containing substrates.   S = 10 nM.   (●) E:S = 10:1; (○) E:S = 1:10.   C. GeneScan gel pictures of time course analysis of cleavage activity of M57L on X-containing substrates with E:S = 10:1. S = 10 nM.

outcome of the N58D substitution is consistent with the biochemical analysis of *E. coli* MUG and human TDG, which reveals that indeed this asparagine residue plays a critical role in catalysis (*50*).   Within the UDG superfamily, the G60 position in Gme SMUG1 is typically occupied by either glycine or alanine except for UDG Family 1, which contains an invariant tyrosine residue (Fig. 2.1).   The G60Y substitution rendered the mutant inactive toward all deaminated substrates (data not shown), suggesting that a small sidechain is required at this position for the SMUG1 family and perhaps for Families 2, 4 and 5 as well.   These results are consistent with a previous biochemical

analysis of hSMUG1 G87Y mutant, which showed no detectable activity against the A/U or G/U substrate (*40*).   A conservative change was made at W62 in an attempt to convert the activity of Gme SMUG1 to that observed in hSMUG1.   W62F essentially maintained wt level activity toward xanthine and uracil substrates (Fig. 2.5A). Likewise, although M64 is invariant in SMUG1, M64G did not seem to alter its catalytic activity (Fig. 2.5A). These results, consistent with the mutational analysis of hSMUG1 (*40, 51*), suggest that the W62 and M64 sidechains in motif 1 are not involved in determining DNA base specificity in SMUG1.
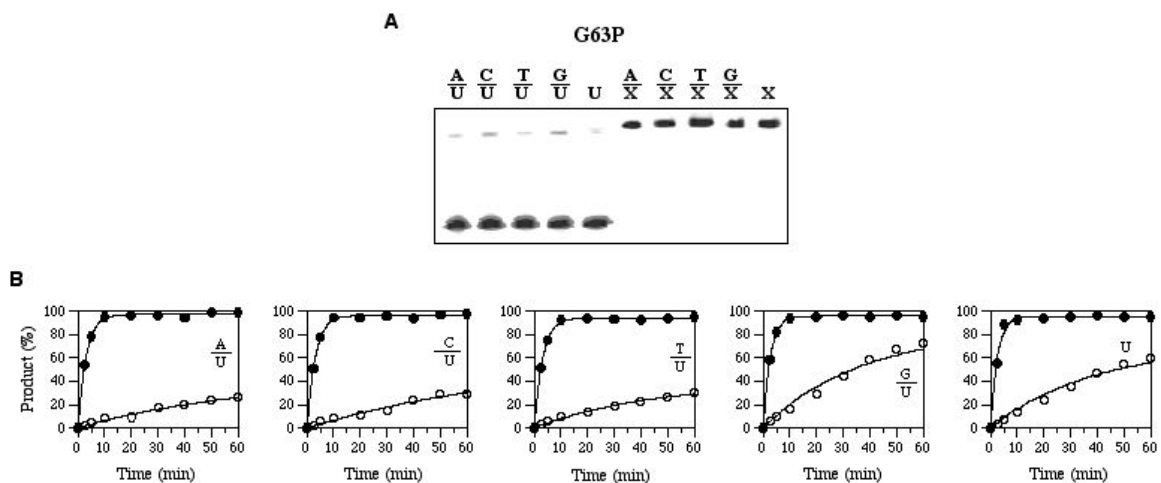
Figure 2.6 Cleavage activity of G63P mutant of Gme SMUG1 on U- and X-containing substrates and kinetic analysis of cleavage activity on U-containing substrates.   A. GeneScan gel picture of cleavage activity of G63P mutant on U- and X-containing substrates.   B. Time course analysis of cleavage activity of G63P mutant on U-containing substrates.   S = 10 nM.   E:S = 10:1.   (●) E:S = 10:1; (○) E:S = 1:10.

G63P

The G63P substitution in Gme SMUG1 mimicked the UNG enzyme (Family 1) at this position (Fig. 2.1).   Similar to wt Gme SMUG1 (Fig. 2.2C), the G63P mutant retained a high level of uracil DNA glycosylase activity, as demonstrated by close to

complete cleavage under the same assay conditions (Fig. 2.6A). In stark contrast to the wt enzyme and other mutants, G63P completely lost xanthine DNA glycosylase activity toward any xanthine-containing DNA substrate, indicating that this mutant became an exclusive uracil DNA glycosylase (compare Fig. 2.6A with Fig. 2.2C). Furthermore, the subtle difference in UDG activity between the wt enzyme and G63P was revealed by kinetic analysis. Although both of them showed similar levels of activity when the enzyme was in excess (E:S = 10:1) (Fig. 2.2C and Fig. 2.6A), G63P differed from the wt enzyme in cleavage of G/U and single-stranded U substrates. Under the assay conditions in which the substrate was in excess (E:S = 1:10), while the wt enzyme was most active with G/U and least active with ss U (Fig. 2.3A), G63P showed a somewhat lower activity with G/U; however, elevated activity with ss U (Fig. 2.6B). As a consequence, the UDG activity from G63P is more uniform with respect to all uracil substrates relative to the wt SMUG1 with the apparent rate constants in the range of 0.24-0.25 per min (Table 2.1), demonstrating a kinetic character that is consistent with the Family 1 UDG or UNG (*19*).

H210 and A214R in motif 2

H210 in motif 2, a conserved residue in all UDG families except for Family 2 MUG/TDG, was substituted by glycine, methionine and asparagine (Fig. 2.1). Overall, the mutants were significantly less active toward uracil substrates than the wt enzyme (Fig. 2.7A-B). Excision of uracil in C/U, T/U and G/U experienced a more than ten-fold reduction in activity as compared with the wt enzyme (Table 2.1). Cleavage of an

A/U substrate suffered 56-fold, 67-fold and 21-fold decreases for H210G, H210M, and H210N, respectively (Fig. 2.7A and Table 2.1). In addition, these mutants demonstrated a significant loss of single-stranded UDG activity (Fig. 2.7A, U), indicating a catalytic feature that is more in-line with Family 5 UDGb enzymes (25). Additionally, the XDG activity of H210G, H210M and H210N was reduced to such a level that the cleavage product was barely detectable even when the enzyme was in excess (Fig. 2.7A).
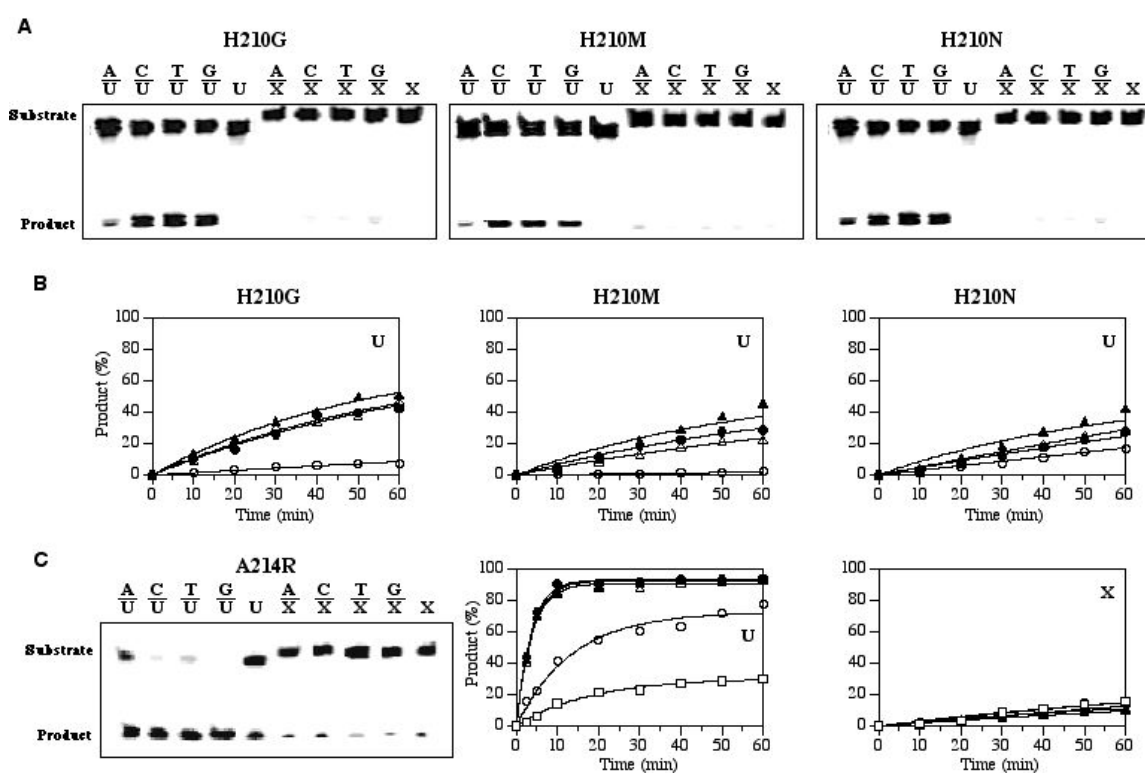
Figure 2.7 Cleavage activity of H210 mutants of Gme SMUG1 on U- and X-containing substrates and kinetic analysis of cleavage activity on U-containing substrates. A. GeneScan gel pictures of cleavage activity of H210 mutants on U- and X-containing substrates. U or X: single-stranded uracil- or xanthine-containing substrate. B. Time course analysis of cleavage activity on U-containing substrates. S = 10 nM. E:S = 10:1. (●) C/U; (○) A/U; (▲) T/U; (△) G/U. C. GeneScan gel picture and time course analysis of cleavage activity of the A214R mutant on U- and X-containing substrates. S = 10 nM. E:S = 10:1. (●) C/U or C/X; (○) A/U or A/X; (▲) T/U or T/X; (△) G/U or G/X; (□) ss U or ss X.

74

All bacterial homologs contain an alanine at the position homologous to position 214 in Gme SMUG1, while eukaryotic SMUG1 proteins typically contain an arginine residue in this position (Fig. 2.1).   The A214R substitution was made in Gme SMUG1 to test its effects on the glycosylase activity.   Under the assay conditions in which the enzyme was in excess (E:S = 10:1), the A214R mutant maintained a similar level of UDG activity as the wt enzyme on C/U, T/U and G/U substrates as indicated by very similar apparent rate constants (Fig. 2.7C and Table 2.1).   On the other hand, the efficiencies of the A214R mutant on A/U and ss U substrates were reduced by 2.9- and 13-fold, respectively (Table 2.1).   Cleavage of all five X-containing substrates was reduced, ranging from 3-fold for C/X to approximately   4-fold for T/X, G/X and ss X to five-fold for A/X (Fig. 2.7C and Table 2.1).   Under the assay conditions in which the substrate was in excess (E:S = 1:10), only the cleavage of U-containing substrates was observed with the activity following the descending order of C/U > T/U > G/U >> A/U, ss U (data not shown).

5. Discussion

Bacterial SMUG1

The rapid sequencing of bacterial genomes has changed our view on SMUG1 as a eukaryote-only DNA repair enzyme (this work and (*40*)).   To the best of our knowledge, this work is the first report to provide experimental evidence to prove SMUG1 homologs from bacteria act as DNA glycosylases.   The distribution of SMUG1 in nature appears to be limited.   So far, it has been found only in four bacterial species

(Fig. 2.1). *Geobacter metallireducens*, isolated from the Potomac River near Washington D. C., is capable of oxidizing organic compounds using iron oxides as the electron acceptor (*52*). In addition to SMUG1, the Gme genome contains a Family 4 UDGa homolog. Given that it lacks a Family 1 UDG or UNG commonly found in prokaryotic and eukaryotic organisms, SMUG1 may play an important role in the repair of cytosine and guanine deamination in DNA in *Geobacter metallireducens*.

SMUG1 as xanthine DNA glycosylase

One of the major findings from this study is that SMUG1 enzymes, regardless of their origins, contain xanthine DNA glycosylase activity. This is surprising since eukaryotic SMUG1 enzymes are only known to be active toward uracil, uracil derivatives, or ethenocytosine (*41, 42*). However, we recently reported oxanine DNA glycosylase activity from mammalian cell extracts and purified human SMUG1, indicating several deaminated purine base lesions are also accommodated by the active site in SMUG1 (*44*). This work offers a comprehensive analysis of xanthine DNA glycosylase activity between bacterial SMUG1 and human SMUG1. The XDG activity from bacterial SMUG1 appears far more robust than that from hSMUG1 (Figs. 2.3 and 2.4). At the maximal incubation time (60 min) and with the enzyme in excess, the Gme SMUG1 was able to remove 70% of xanthine from either double-stranded or single-stranded DNA (Fig. 2.3B). On the other hand, human SMUG1 can only remove less than 15% xanthine from DNA (Fig. 2.4B). It should be noted that the XDG activity from hSMUG1 (< 15%) is more robust than its ODG activity (~3%) (*44*). Despite the

difference in the level of XDG activity, both the bacterial and human SMUG1 can remove xanthine regardless of strandness or opposite base. It should be pointed out that the XDG activity of Gme SMUG1 is distinct from its UDG activity in two respects. First, the overall XDG activity of Gme SMUG1 is less robust than the UDG activity (Fig. 2.3). Second, the UDG activity is sensitive toward strandness and somewhat toward the opposite base (Fig. 2.3A).

Table 2.2: Interaction energies (kcal/mol) between substrates and SMUG1[a]

	Uracil	Xanthine	Uracil (min)	Xanthine (min)
Gme SMUG1	-34.4	-36.8	-39.0	-68.1
Human SMUG1	-36.4	-26.2	-41.0	-53.4

[a]: Interaction energies, the sum of the intermolecular nonbonded van der Waals and Coulomb energies between the enzyme and ligand, were calculated under vacuum conditions without cutoffs and determined for both the initial model complexes as well as the minimized (min) model complexes.

In order to gain further insight into the experimental results, molecular modeling calculations were performed to characterize the structural and energetic properties of the Gme and human SMUG1 enzymes bound to uracil and xanthine. We investigated the significant differences in xanthine DNA glycosylase activity between Gme and human SMUG1 enzymes. This was accomplished by first homology modeling the Gme and human SMUG1 enzymes complexed with xanthine and uracil and then followed up with a qualitative estimate of the relative interaction energies associated with these complexes (Table 2.2). Within the charmm molecular modeling package, simple interaction energies were calculated between the SMUG1 enzymes and their substrates uracil and xanthine. Although the comparison of interaction energies (intermolecular Coulomb

and van der Waals energies) is not an accurate way of determining binding free energy values, it provides a rapid, qualitative means of accessing key interactions between the ligand and the protein. We examined both the initial models of the enzymes bound to their substrates as well as the models following a gentle minimization protocol described in the Materials and Methods section. These basic calculations, when applied to both sets of models, indicate that uracil is accommodated in both the Gme and human SMUG1 enzymes similarly (Table 2.2). However, xanthine is accommodated much more readily by the Gme SMUG1 than the human homolog as indicated by the more negative (favorable) interaction energies determined in the Gme SMUG1-xanthine complex. As shown in Table 2.2, these interaction energy trends are consistent when examining both the initial models of the complexes as well as the optimized (minimized) models.

A   B
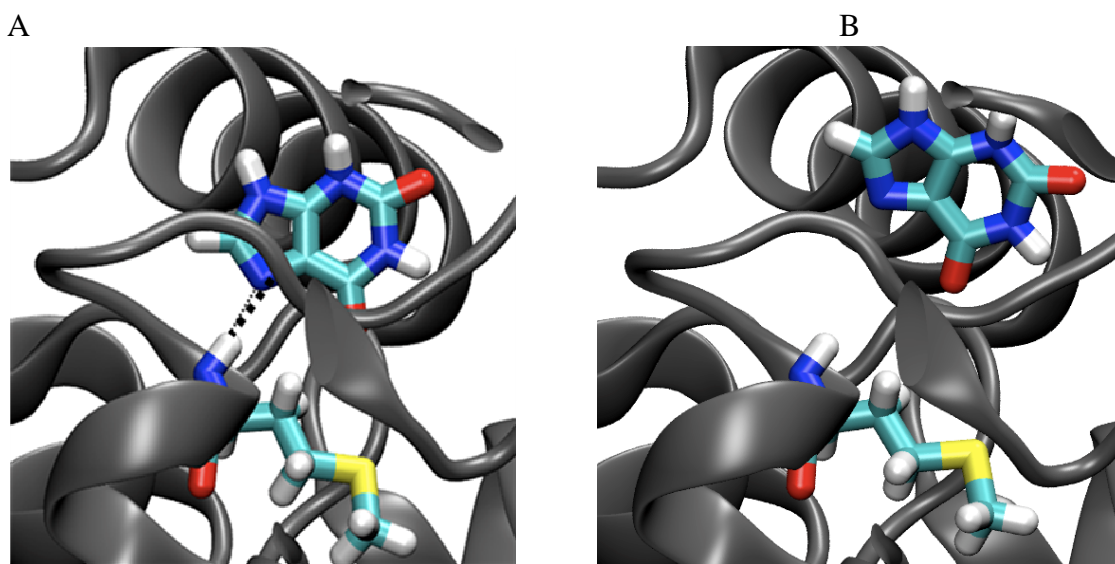
Figure 2.8 Minimized structures of Gme SMUG1 and human SMUG1. A. Gme SMUG1-xanthine complex structure. M64 and xanthine are shown in color. B. Human SMUG1-xanthine complex structure. M91 and xanthine are shown in color. Hydrogen bond formed between the backbone amide of M64 and N7 of xanthine in Gme SMUG1 (A) is missing in the human SMUG1 (B).

By closely examining the contributions from each amino acid in the Gme and human SMUG1 enzyme to their vdW and Coulombic interactions with xanthine, we find that a key interaction with M64, which forms a backbone hydrogen bond with N7 of xanthine in the Gme SMUG1-xanthine complex, is significantly weakened in the human SMUG1-xanthine complex (Fig. 2.8). This weakened interaction may partly contribute to the lower activity of human SMUG1 for xanthine (Figs. 2.2 and 2.4).

G60Y

All UDG superfamily enzymes consist of two motifs involved in the activation of a water molecule for attacking the glycosidic bond and recognition of deaminated bases (*19, 53*). The mutational analysis described here reveals contributions from motifs 1 and 2 in the recognition of both the pyrimidine deaminated base uracil and the purine deaminated base xanthine. As an experimental control, we constructed the W62F mutant and modeled its structure. Amino acid substitutions at the non-conserved W62 did not seem to significant affect its catalytic activity in Gme SMUG1 and hSMUG1 (Fig. 2.5 and (*51*)), suggesting that W62 is not part of the recognition pocket. Indeed, the sidechain of W62F is directed away from the binding pocket according to the molecular modeling (Supplemental Fig. 2.1, W62 and W62F).

The mainchains of G98 (equivalent to G60 in Gme SMUG1 and G87 in hSMUG1) and M102 (equivalent to M64 in Gme SMUG1 and M91 in hSMUG1) form hydrogen bonds to a well-ordered water molecule in the x-ray structure of *Xenopus* SMUG1 (*43*). Mutations at the corresponding positions in Gme SMUG1 had very

79

different effects.   Neither M64G in Gme SMUG1 described in this study nor M91A in

hSMUG1 significantly alter the catalytic activity (Fig. 2.5 and (*51*)).   On the other hand,

G60Y in Gme SMUG1 and G87Y in hSMUG1 rendered the enzymes completely inactive

(this work and (*40*)).   Note all UDG superfamily enzymes take a glycine or alanine in

this position except for UDG Family 1.   These results, combined with biochemical

analysis of G87A, G87S, G87V, and G87F in hSMUG1 (*40, 51*), suggest that a bulky

sidechain may prevent recognition of both uracil and xanthine in the active site. This is

substantiated by examining molecular models for the G60Y and M64G mutants.   In the

case of M64G, the sidechain is directed away from the base binding pocket and thus may

cause minimal changes to the activity of the enzyme (Supplemental. 2.1, M64 and M64G;

located at the end of this chapter).   On the other hand, models of the G60Y mutation in

Gme SMUG1 indicate that the bulky tyrosine is in a position that may inhibit a flipped

out base from entering the binding pocket (Fig. 2.9A).   Models indicate that the base

binding pocket in G60Y still has enough room to accommodate a free DNA base;

however, the proximity of the tyrosine at position 60 to the N1 of uracil suggests the

possibility that this amino acid is blocking the channel to the binding site.   As a result,

both UDG and XDG activities are eliminated.   Interestingly, in the human UNG

structure, the equivalent Y147 blocks binding of pyrimidines other than uracil   (*54*).

Mutational analysis shows that Y147A, Y147C and Y147S of human UNG become

active on cytosine and thymine bases (*55*).   To understand the distinctly different

specificity exhibited by the SMUG1 and UNG enzymes, we compared the human UNG

structure with the modeled Gme SMUG1 G60Y mutant structure (Fig. 2.9A).   The

molecular models strongly indicate that the tyrosine at position 60 in the G60Y structure is sterically blocking the binding channel (Fig. 2.9A, compare Gme-WT with Gme-G60Y). However, an examination of the binding channel of the human UNG structure clearly shows that the corresponding tyrosine is in a conformation that does not interfere with the binding channel (Fig. 2.9A, hUNG-WT and Gme-G60Y) (*40, 56*).
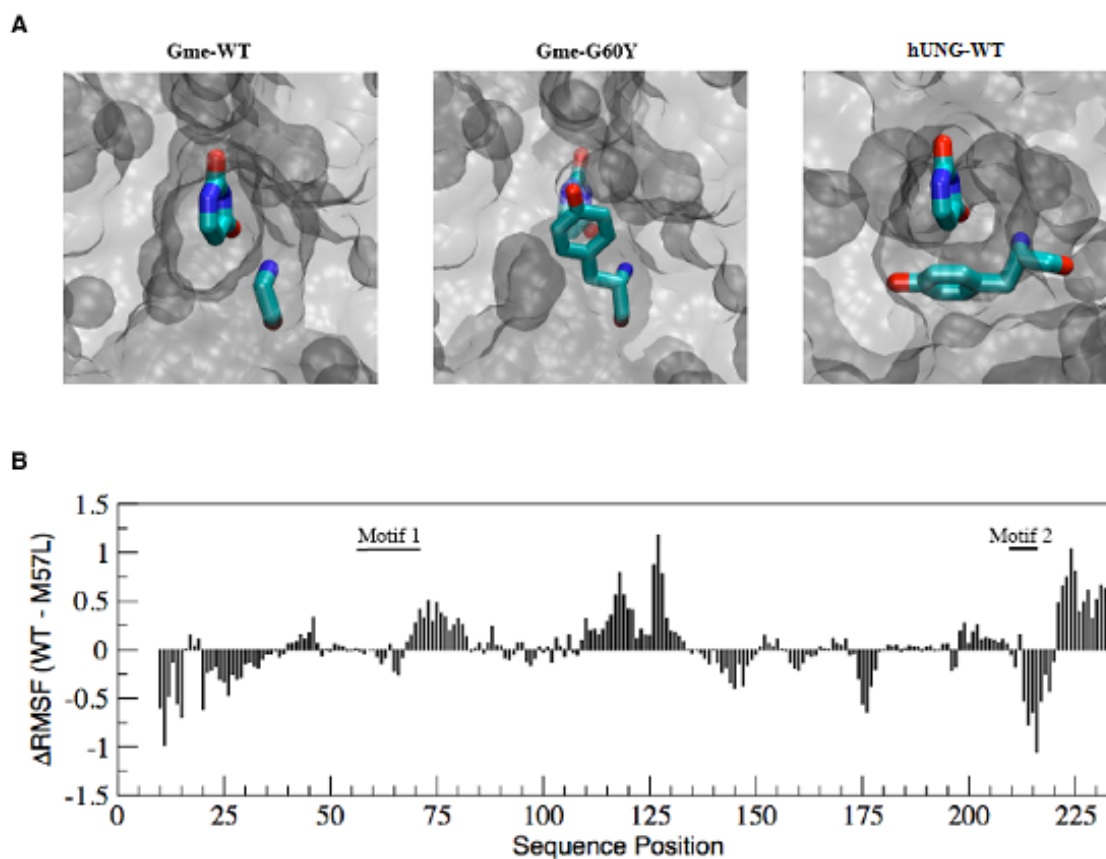
Figure 2.9 Molecular modeling and molecular dynamics simulation of Gme SMUG1 mutants. A. Models depicting the effect of Gme SMUG1 G60Y mutation on the entry of a uracil base to the binding pocket. Uracil, G60 and G60Y in Gme SMUG1, and Y147 in human UNG (pdb code: 1ssp) are shown in color. B. Difference in isotropic root mean squared fluctuations between the wt Gme SMUG1 and the M57L mutant. Positive values indicate that C-α's are more rigid in the mutant.

M57L and A214R

The M57L substitution reduced; however, did not abolish both UDG and XDG activities. In the *Xenopus* SMUG1 crystal structures and the modeled Gme SMUG1 structures, the mainchain amide of M57 interacts with the C2-keto group of uracil or xanthine. Mutating the sidechain would not directly perturb this interaction. Further, the sidechain of methionine and the sidechain of leucine in the models both appear to be directed away from the binding site into the core of the protein (Supplemental Fig. 2.1, M57 and M57L). However, given the fact that M57 directly neighbors the catalytic N58 residue, subtle changes in structure or dynamics in this region could significantly impact the specificity and catalytic efficiency of the Gme SMUG1 enzyme. In order to explore the impact of the M57L mutation on the Gme SMUG1 enzyme, we ran molecular dynamics simulations examining any relevant changes to the flexibility of the enzyme. The results indicate the flexibility of the M57L mutant is perturbed, most notably in the motif 2 region, which demonstrates greater flexibility (Fig. 2.9B). Since part of motif 2 (equivalent to [211]PSPAS[215] in Gme SMUG1) is a "wedge" that occupies the space vacated by the flipped-out base (*43*), we surmise that this change may result in a less effective "wedge" that will not sufficiently stabilize the double-stranded DNA in the flipped out conformation. As a result, the M57L mutant shows a reduction in the xanthine DNA glycosylase activity (Fig. 2.5 and Table 2.1).

To test this hypothesis, we substituted alanine with arginine at the 214 position. The A214R substitution in Gme SMUG1, which mimics the human SMUG1, reduced its XDG activity (Fig. 2.7C). This is consistent with the low XDG activity observed in

hSMUG1 (Fig. 2.4). Furthermore, the reduced activity on A/U and ss U resembles *Xenopus* SMUG1 in which G/U is a more efficient substrate than either A/U or ss U (*43*). These results may suggest that the A214R mutant is less capable of flipping a deaminated base. Similarly to the M57L mutant, the A214R mutant displays increased selectivity for uracil. Therefore, either a direct mutation at the "wedge" or a remote change such as M57L that alters the flexibility of the "wedge" may achieve a similar effect on SMUG1's glycosylase activity.

H210G, H210M and H210N

In the crystal structure of the *Xenopus* SMUG1 enzyme, H250 in motif 2 (equivalent to H210 in Gme SMUG1) appears to contribute a hydrogen bond interaction with the C2-keto group in uracil (*43*). All three Gme SMUG1 mutants (H210N, H210M, H210G) exhibited similar effects, i.e., close to complete loss of XDG activity and substantial loss of UDG activity (Fig. 2.7 and Table 2.1). In the modeled structures, H210G and H210M lose the hydrogen bond interaction, which may contribute to the substantial loss of activity by these mutants. The loss of catalytic activity on the A/U substrate is less severe with the H210N mutant (Fig. 2.7A and Table 2.1). This may be due to the fact that H210N potentially can still form a hydrogen bond via its sidechain amide, which may help maintain the interaction with the uracil base in an A/U base pair. However, even H210N lost substantial UDG activity as compared with the wt enzyme, suggesting that H210 may play other subtle roles beyond a simple hydrogen bond

interaction. One possibility is that the H210 mutants may impact the structure or flexibility of the neighboring "wedge" region of the SMUG1 enzyme.

G63P

The G63P substitution in motif 1 showed the most striking effect as it abolished Gme SMUG1's xanthine DNA glycosylase activity (Fig. 2.6). Consistent with a proline at this position in UDG Family 1 (Fig. 2.1), G63P is kinetically quite similar to UDG or UNG (Table 2.1). We surmised that a mutation from glycine to a more conformationally restrictive proline would alter movement within the activation loop of the Gme SMUG1 enzyme. A change of the dynamics within this critical region of the enzyme may contribute to the observed loss of xanthine DNA glycosylase activity. To understand the effect of G63P on the catalytic activity, we modeled G63P into the Gme SMUG1 structure (Fig. 2.10A). Several interesting insights emerged from molecular dynamics simulations performed on the wild-type Gme SMUG1 and the G63P mutant. First, the equilibrated trajectories were analyzed to determine whether there were any significant changes in the C-α isotropic root mean squared fluctuations (RMSF) of each protein. The results indicate that indeed the flexibility of the loop regions in motifs 1 and 2 are altered by the G63P substitution (Fig. 2.10B). Consistent with the M57L mutation (Fig. 2.9B), the increased flexibility of the "wedge" region in motif 2 may contribute to the increased selectivity of this mutant for uracil over xanthine. Second, an examination of the base binding pocket indicates that the G63P mutation causes a reduction in the average free volume, which may also contribute to this mutant being able to more
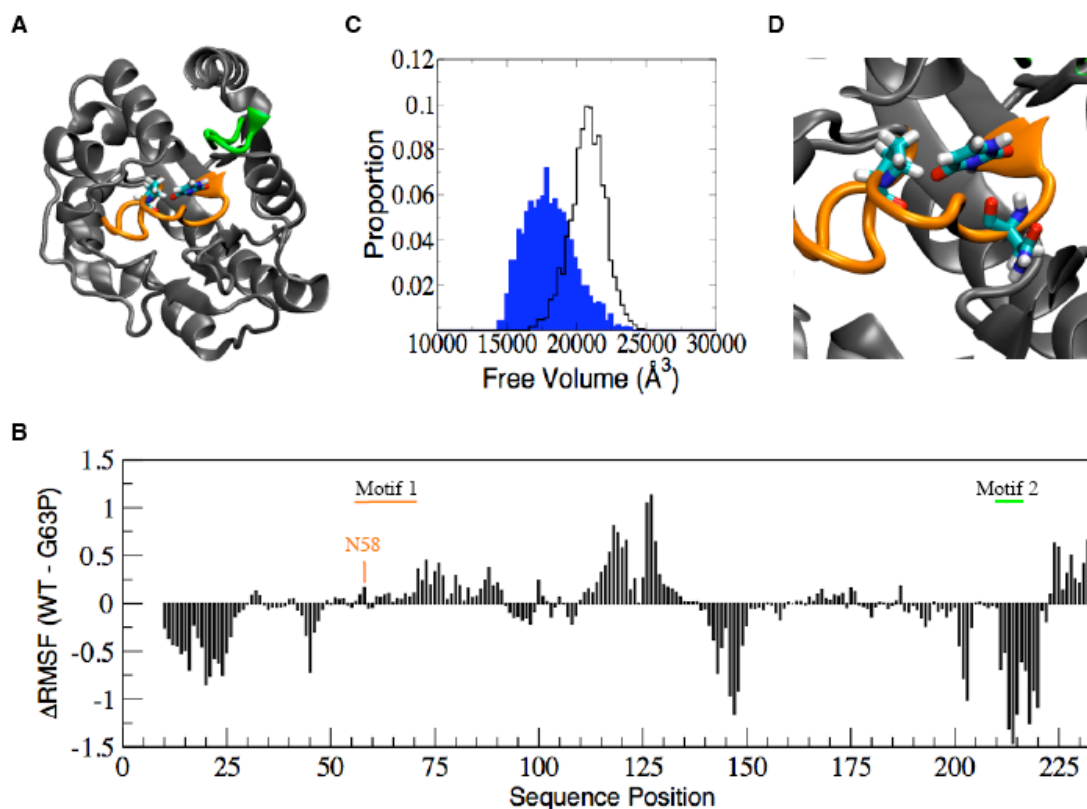
Figure 2.10 Molecular modeling and molecular dynamics simulation of Gme SMUG1 G63P mutant.   A. Model of G63P mutant bound to uracil.   Motif 1 is shown in orange and motif 2 is shown in green.   G63P and uracil are shown in color.   B. Difference in isotropic root mean squared fluctuations between the wt Gme SMUG1 and the G63P mutant.   Positive values indicate that C-α's are more rigid in the mutant.   C. Free volume distributions of WT Gme SMUG1 and G63P mutant.   D. A closeup view of the catalytic residue N58 and its vicinity.   N58, G63P and uracil are shown in color.

effectively eliminate binding to the larger xanthine base while maintaining strong interactions with the smaller uracil base (Fig. 2.10C).   In contrast, there is no significant change in free volume observed between M57L and the wt enzyme (data not shown). Third, G63P maintains robust and uniform uracil DNA glycosylase activity toward all five U-containing substrates (Fig. 2.6B and Table 2.1).   N58 is the catalytic residue that activates a water molecule to attack the N-glycosidic bond.   The RMSF chart shows that

N58 appears to be situated in a more rigid loop in the G63P mutant than that in the wt enzyme (Fig. 2.10B and D). This change in the dynamic character may help to more efficiently align the catalytic center to the scissile glycosidic bond, thus contributing to the robust glycosylase activity with respect to all five uracil-containing substrates (Table 2.1). Overall, it may be that the structural changes in the active site result in a conformation more accommodating to uracil in addition to excluding xanthine. This effect might explain the uniformity of activity against all uracil substrates. As such, Gme SMUG1 is switched to a Family 1 UDG- or UNG-like enzyme by the G63P mutation. Protein dynamics have been shown to be closely associated with enzyme catalysis (*57-61*), where interconversions between sub-states have been observed (*62*). By modulating the flexibility of specific protein regions, G63P may confine the motion of Gme SMUG1 toward more productive conformations.

In summary, this study reveals that both bacterial and human SMUG1 enzymes are not only uracil DNA glycosylases but also xanthine DNA glycosylases. The difference in XDG activity between the bacterial SMUG1 and human SMUG1 may be attributable to the ability of the bacterial enzyme to effectively flip out the xanthine base and accommodate it in the active site by specific interactions. The striking switching of SMUG1 to an exclusive Family 1 UDG- or UNG-like enzyme by a single amino acid substitution underscores the evolutionary conservation between Family 3 SMUG1 and Family 1 UDG or UNG. Molecular modeling and molecular dynamics analyses uncover the power a single amino acid substitution may exert on substrate specificity as well as catalytic efficiency through the tuning of protein conformational flexibility.

Given that loop regions are commonly involved in the active site of many enzymes, this mechanism for evolving distinct and specialized enzyme functions (such as the truly selective UNG function generated through a glycine to proline mutation) may be a general mechanism applicable to other DNA repair enzymes and perhaps other enzyme classes.

Acknowledgements

6. References

1. Shapiro, R. (1981) Damage to DNA caused by hydrolysis, in *Chromosome Damage and Repair* (Seeberg, E., and Kleppe, K., Eds.) pp 3-18, Plenum Press, New York.

2. Nguyen, T., Brunson, D., Crespi, C. L., Penman, B. W., Wishnok, J. S., and Tannenbaum, S. R. (1992) DNA damage and mutation in human cells exposed to nitric oxide in vitro. *Proc Natl Acad Sci U S A 89*, 3030-3034.

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

4. Wink, D. A., and Mitchell, J. B. (1998) Chemical biology of nitric oxide: Insights into regulatory, cytotoxic, and cytoprotective mechanisms of nitric oxide. *Free Radic Biol Med 25*, 434-456.

5. Burney, S., Caulfield, J. L., Niles, J. C., Wishnok, J. S., and Tannenbaum, S. R. (1999) The chemistry of DNA damage from nitric oxide and peroxynitrite. *Mutat Res 424*, 37-49.

6. Spencer, J. P., Whiteman, M., Jenner, A., and Halliwell, B. (2000) Nitrite-induced deamination and hypochlorite-induced oxidation of DNA in intact human respiratory tract epithelial cells. *Free Radic Biol Med 28*, 1039-1050.

7. Dedon, P. C., and Tannenbaum, S. R. (2004) Reactive nitrogen species in the chemical biology of inflammation. *Arch Biochem Biophys 423*, 12-22.

8. Lucas, L. T., Gatehouse, D., and Shuker, D. E. (1999) Efficient nitroso group transfer from N-nitrosoindoles to nucleotides and 2'-deoxyguanosine at physiological pH. A new pathway for N- nitrosocompounds to exert genotoxicity. *J Biol Chem 274*, 18319-18326.

9. 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.

10. Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature 274*, 775-780.

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

12.   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.

13.   Kamiya, H., Sakaguchi, T., Murata, N., Fujimuro, M., Miura, H., Ishikawa, H., Shimizu, M., Inoue, H., Nishimura, S., Matsukage, A., Masutani, C., Hanaoka, F., and Ohtsuka, E. (1992) In vitro replication study of modified bases in ras sequences. *Chem Pharm Bull (Tokyo) 40*, 2792-2795.

14.   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.

15.   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.

16.   Hitchcock, T. M., Dong, L., Connor, E. E., Meira, L. B., Samson, L. D., Wyatt, M. D., and Cao, W. (2004) Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase. *J Biol Chem 279*, 38177-38183.

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

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

19.   Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat Res 460*, 165-181.

20.   Huffman, J. L., Sundheim, O., and Tainer, J. A. (2005) DNA base damage recognition and removal: new twists and grooves. *Mutat Res 577*, 55-76.

21.   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 Res 31*, 2261-2271.

22.   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 (Amst) 7*, 1962-1972.

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

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

25. 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 Res 30*, 2097-2102.

26. 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. *J Mol Biol 333*, 515-526.

27. 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.

28. Vongchampa, V., Dong, M., Gingipalli, L., and Dedon, P. (2003) Stability of 2'-deoxyxanthosine in DNA. *Nucleic Acids Res 31*, 1045-1051.

29. Caulfield, J. L., Wishnok, J. S., and Tannenbaum, S. R. (1998) Nitric oxide-induced deamination of cytosine and guanine in deoxynucleosides and oligonucleotides. *J Biol Chem 273*, 12689-12695.

30. Dong, M., Wang, C., Deen, W. M., and Dedon, P. C. (2003) Absence of 2'-deoxyoxanosine and presence of abasic sites in DNA exposed to nitric oxide at controlled physiological concentrations. *Chem Res Toxicol 16*, 1044-1055.

31. Dong, M., and Dedon, P. C. (2006) Relatively small increases in the steady-state levels of nucleobase deamination products in DNA from human TK6 cells exposed to toxic levels of nitric oxide. *Chem Res Toxicol 19*, 50-57.

32. 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 Res 30*, 4975-4984.

33. Dong, M., Vongchampa, V., Gingipalli, L., Cloutier, J. F., Kow, Y. W., O'Connor, T., and Dedon, P. C. (2006) Development of enzymatic probes of oxidative and nitrosative DNA damage caused by reactive nitrogen species. *Mutat Res 594*, 120-134.

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

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

36.   Jaruga, P., and Dizdaroglu, M. (1996) Repair of products of oxidative DNA base
       damage in human cells. *Nucleic Acids Res 24*, 1389-1394.

37.   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. *Curr Biol 9*, 174-185.

38.   Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., Slupphaug, G.,
       Daly, G., Krokan, H. E., Lindahl, T., and Barnes, D. E. (2000) Uracil-DNA
       glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during
       DNA replication. *Mol Cell 5*, 1059-1065.

39.   Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., and
       Lindahl, T. (2001) Excision of deaminated cytosine from the vertebrate genome:
       role of the SMUG1 uracil-DNA glycosylase. *Embo J 20*, 4278-4286.

40.   Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E., and
       Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the
       initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res 35*,
       3879-3892.

41.   Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P.
       A., Hagen, L., Krokan, H. E., and Slupphaug, G. (2002) hUNG2 is the major
       repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U
       in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J Biol
       Chem 277*, 39926-39936.

42.   Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H.,
       Ohyama, Y., Karino, N., Matsuda, A., and Ide, H. (2003) Mammalian 5-
       formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in
       repair of 5-formyluracil and other oxidized and deaminated base lesions.
       Biochemistry 42, 5003-5012.

43.   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. *Mol Cell 11*, 1647-1659.

44.   Dong, L., Meira, L. B., Hazra, T. K., Samson, L. D., and Cao, W. (2008) Oxanine DNA glycosylase activities in mammalian systems. *DNA Repair (Amst) 7*, 128-134.

45.   Petrey, D., Xiang, Z., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T., Soto, C. S., Goldsmith-Fischman, S., Kernytsky, A., Schlessinger, A., Koh, I. Y., Alexov, E., and Honig, B. (2003) Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins 53 Suppl 6*, 430-435.

46.   Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization and dynamic calculations. *J. Comp. Chem. 4*, 187-217.

47.   MacKerell Jr., A. D., Bashford, D., Bellot, M., Dunbrack Jr., 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 III, 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. *J. Phys. Chem. 102*, 3586-3616.

48.   Feig, M., Karanicolas, J., and Brooks, C. L., 3rd. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model 22*, 377-395.

49.   Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999) Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Accts. 101*, 426-434.

50.   Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem 275*, 33449-33456.

51.   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 Res 32*, 5291-5302.

52.   Lovley, D. R. (1991) Dissimilatory Fe(III) and Mn(IV) reduction. *Microbiol Rev 55*, 259-287.

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

54.   Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W., and Tainer, J. A. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell 82*, 701-708.

55.   Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A., and Krokan, H. E. (1996) Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *Embo J 15*, 3442-3447.

56.   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. *Embo J 17*, 5214-5226.

57.   Cannon, W. R., Singleton, S. F., and Benkovic, S. J. (1996) A perspective on biological catalysis. *Nat Struct Biol 3*, 821-833.

58.   Daniel, R. M., Dunn, R. V., Finney, J. L., and Smith, J. C. (2003) The role of dynamics in enzyme activity. *Annu Rev Biophys Biomol Struct 32*, 69-92.

59.   Agarwal, P. K. (2005) Role of protein dynamics in reaction rate enhancement by enzymes. *J Am Chem Soc 127*, 15248-15256.

60.   Huang, Y. J., and Montelione, G. T. (2005) Structural biology: proteins flex to function. *Nature 438*, 36-37.

61.   Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006) An NMR perspective on enzyme dynamics. *Chem Rev 106*, 3055-3079.

62.   Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature 438*, 117-121.
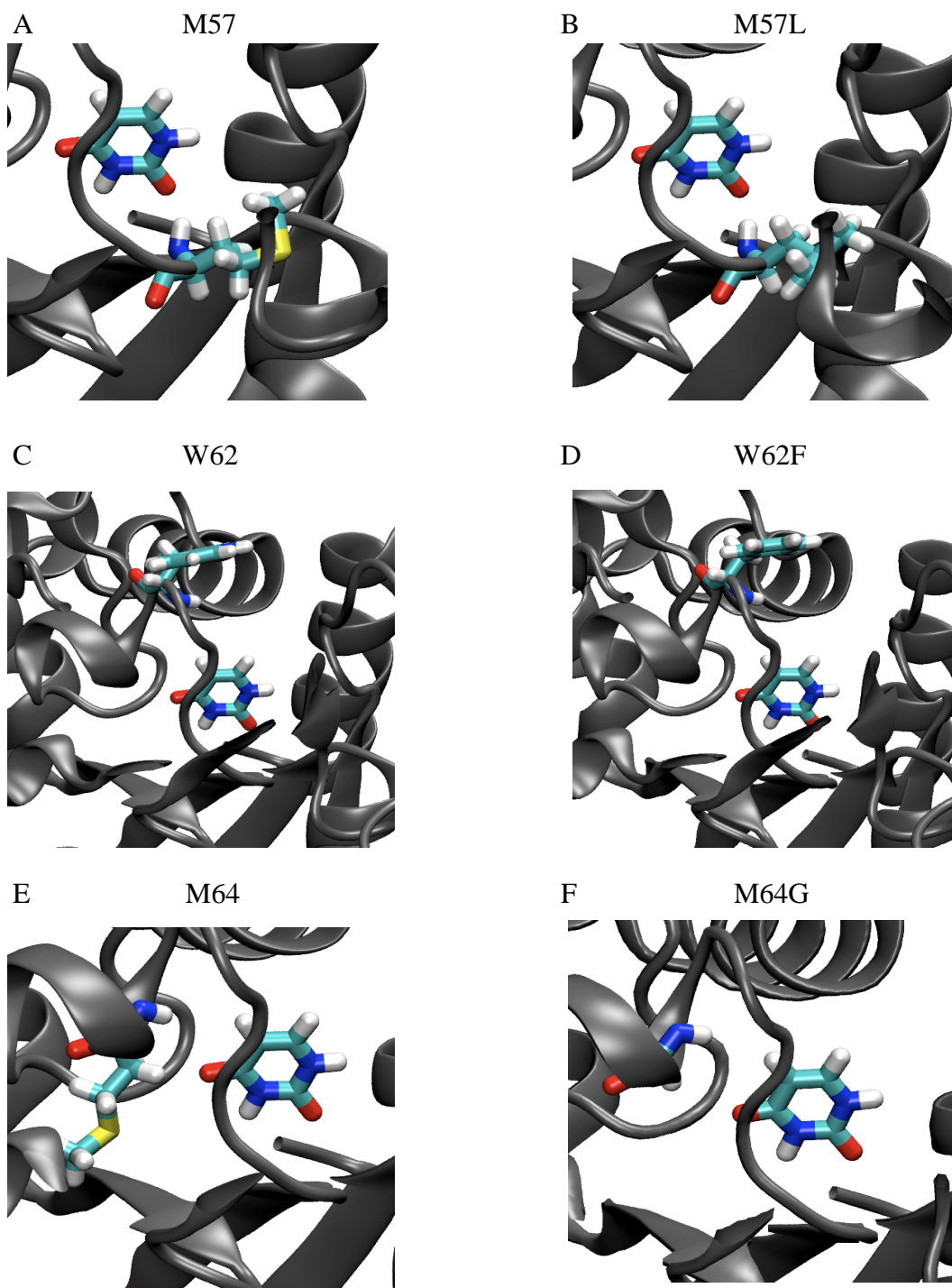
Supplemental Figure 2.1 Molecular modeling of Gme SMUG1.   Uracil, M57 (A), M57L (B), W62 (C), W62F (D), M64 (E), and M64G (F) are shown in color.

CHAPTER THREE

DISSECTING ENDONUCLEASE AND EXONUCLEASE ACTIVITIES IN

ENDONUCLEASE V FROM *THERMOTOGA MARITIMA*

1. Abstract

Endonuclease V is an enzyme that initiates a conserved DNA repair pathway by making an endonucleolytic incision at the 3' side one nucleotide from a deaminated base lesion. DNA cleavage analysis using mutants defective in DNA binding and Mn^{2+} as a metal cofactor reveals a novel 3'-exonuclease activity in endonuclease V. This study defines the endonuclease and exonuclease activity in endonuclease V from *Thermotoga maritima* (Tma). In addition to its well-known inosine-dependent endonuclease, Tma endonuclease V also exhibits inosine-dependent 3'-exonuclease activity. The dependence on inosine site and the exonuclease nature of the 3'-exonuclease activity was demonstrated using 5'-labeled and internally-labeled inosine-containing DNA and H214D mutant that is defective in nonspecific nuclease activity. Detailed kinetic analysis using 3'-labeled DNA indicates that Tma endonuclease V also possesses nonspecific 5'-exonuclease activity. The multiplicity of the endonuclease and exonuclease activity is discussed with respect to deaminated base repair.

2. Introduction

Endonuclease V (endo V) is a DNA repair enzyme which hydrolyzes the second phosphodiester bond 3' from a deaminated base lesion (*1-4*). Inosine derived from adenosine deamination, xanthosine and oxanosine from guanosine deamination, and uridine from cytidine deamination are all substrates *in vitro* for endo V (*1-3, 5, 6*). The endonuclease activity towards inosine and its role in repair of deaminated inosine damage *in vivo* have been studied intensely. Biochemical and kinetic analysis reveals that endo V remains bound to inosine-containing DNA after strand cleavage (*1, 3, 7, 8*). It was proposed that this unique property may act as a sensor to recruit other proteins for downstream repair process (*3*). Genetic analysis using endo V (encoded by *nfi* gene) deletion mutants indicates that endo V is a primary repair enzyme for the repair of inosine lesions produced under nitrosative stress condition (*9-11*). After the 3' cut by endo V downstream of the inosine site, a repair patch needs to be created to remove the lesion. Taking advantage of an oligonucleotide-mediated transformation system in *Escherichia coli* (*E. coli*), a genetic study demonstrates that repair of inosine in chromosome creates a five-nucleotide (nt) gap (*12*). After the cleavage at the inosine site, the repair patch is defined by removal of two nt from the 5' side and three nt from the 3' side (*12*).

The structure and function relationship of this repair enzyme has been investigated intensely using thermostable endonuclease V from the thermophilic bacterium *Thermotoga maritima* (Tma) as a model system. A systematic site-directed mutagenesis analysis on all seven conserved motifs defines the role of a series of residues

in motifs I, III, IV, V, Vi, VII in interactions with inosine in DNA (*13*). D43, E89, D110, and H214 are part of the active site that coordinates metal binding and catalytic function (*14*). Several conserved amino acid residues in motifs III and IV play an important role in protein-DNA interactions and recognition of deamianted DNA bases.

In addition to the endonuclease activity nicking the inosine site, Tma endo V also contains nonspecific endonuclease activity (*3*). An important finding from previous systematic site-directed mutagenesis analysis is the revelation of 3'-exonuclease activity from Tma endo V (*13*). Previous biochemical characterization of the nuclease activity from Tma endo V primarily focused on endonuclease activity on deaminated lesions (*2, 3, 7, 13-15*). In this study, we investigated endonuclease and exonuclease activities from Tma endo V using 5'-labeled, 3'-labeled, 5'-internally labeled inosine-containing DNA. Using 5'-labeled T/I and T/A substrates and 5'-internally labeled T/I substrate, we detected inosine-dependent 3'-exonuclease activity. Using H214D mutant that primarily retained inosine-dependent endonuclease and 3'-exonuclease activity, we defined the kinetic property of the inosine-dependent 3'-exonuclease activity. Using 3'-labeled T/I and T/A substrates, we detected and defined nonspecific 5'-exonuclease activity. Using 5'-labeled dual inosine-containing substrate, we defined the directionality of endo V's movement on DNA.

3. 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). Deoxyoligonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA). LB medium was prepared according to standard recipes. GeneScan Stop Buffer consisted of 80% formamide (Amresco, Solon, OH), 50 mM EDTA (pH 8.0), and 1% blue dextran (Sigma Chemicals). TB Buffer (1 x) consisted of 89 mM Tris base and 89 mM boric acid. TE buffer consisted of 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA. *Escherichia coli* host strain AK53 (*mrr*B⁻, MM294) was from our laboratory collection. Tma endo V mutant proteins were prepared as previously described (13).

Oligonucleotide substrates.

The sequences of the oligonucleotides are shown in Fig. 2.2A. Oligonucleotides containing inosine or uridine were ordered from IDT and purified by PAGE. The oligonucleotides were dissolved in TE buffer at a final concentration of 10 µM. The two complementary strands with the unlabeled strand in 1.2-fold molar excess was mixed and incubated at 85°C for 3 min and allowed to form duplex DNA substrates at room temperature for more than 30 min. Oligonucleotides containing xanthosine (X) or oxanosine (O) were prepared as previously described (*1, 17*).

Tma endo V cleavage assays.

The cleavage reaction mixtures (10 µL) containing 10 mM HEPES-KOH (pH 7.4), 1 mM DTT, 2% glycerol, 5 mM $MnCl_2$ (unless otherwise specified), 10 nM oligonucleotide DNA substrate, and 100 nM Tma endo V protein (unless otherwise specified) were incubated at 65℃ for 30 min. The reactions were terminated by addition of an equal volume of GeneScan Stop Buffer. The reaction mixtures were then heated at 94 ℃ for 3 min and cooled on ice. Samples (3.6 µL) were loaded onto a 16% denaturing polyacrylamide gel containing 7 M urea. Electrophoresis was conducted at 1500 V for 3.2 h using an ABI 377 sequencer (Applied Biosystems). Cleavage products and remaining substrates were quantified using GeneScan analysis software version 3.0.

4. Results

Cleavage of 5'-labeled inosine- and non-inosine-containing substrates

Endonuclease V is an authentic nuclease involved in repair of DNA after deaminat base damage such as inosine. For example, the thermostable ortholog from *Thermotoga maritima* has been a valuable model system for biochemical investigation given its thermostability and ease to generate site-directed mutants. Previously, we have conducted a systematic site-directed mutagenesis of Tma endo V and generated over sixty site-directed mutants (*13*). Using the mutants with relatively low binding affinity to inosine-containing DNA, we observed 3'-exonuclease activity of Tma endo V in the presence of Mn^{2+} metal cofactor. However, the biochemical properties of the exonuclease activity were not defined.

A

B

5'-Labeled Full-Length T/I Substrate
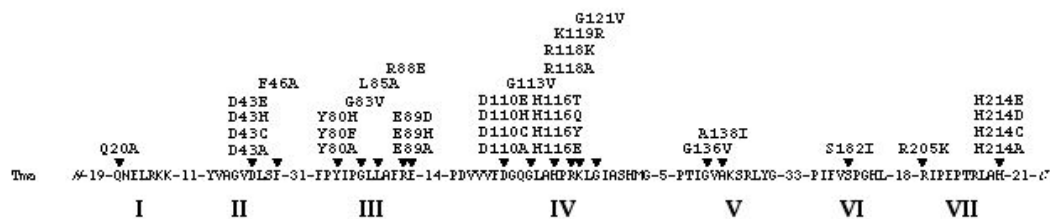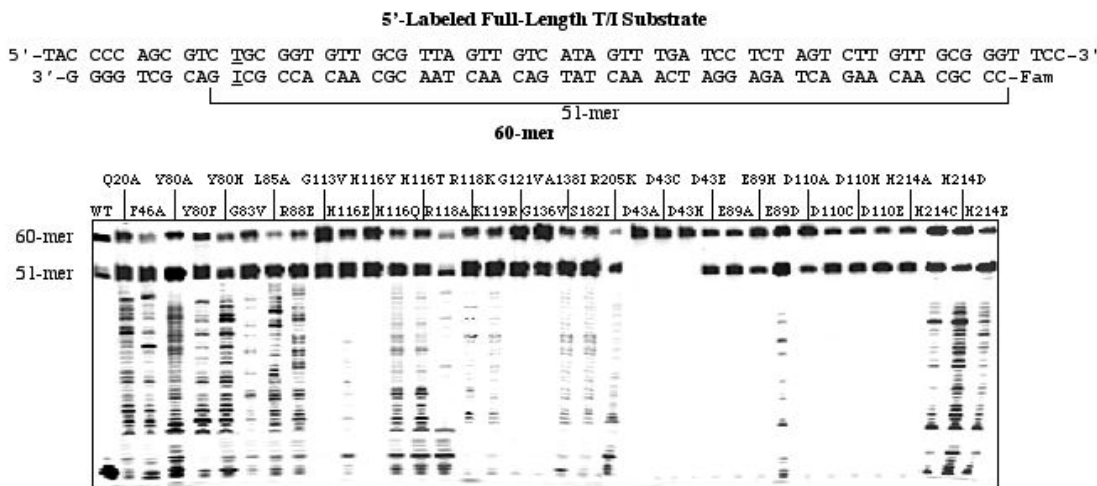
Figure 3.1 Cleavage activity of Tma endo V on 5'-labeled full-length T/I substrate. A. Conserved motifs of Tma endo V. Site-directed mutants used in this study were listed above the arrows. B. Cleavage activity of wt and mutant Tma endo V on 5'-labeled full-length T/I substrate. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzyme and 10 nM substrate. The 5' end of the bottom strand was labeled with the Fam fluorophore. The T/I base pair was underlined.
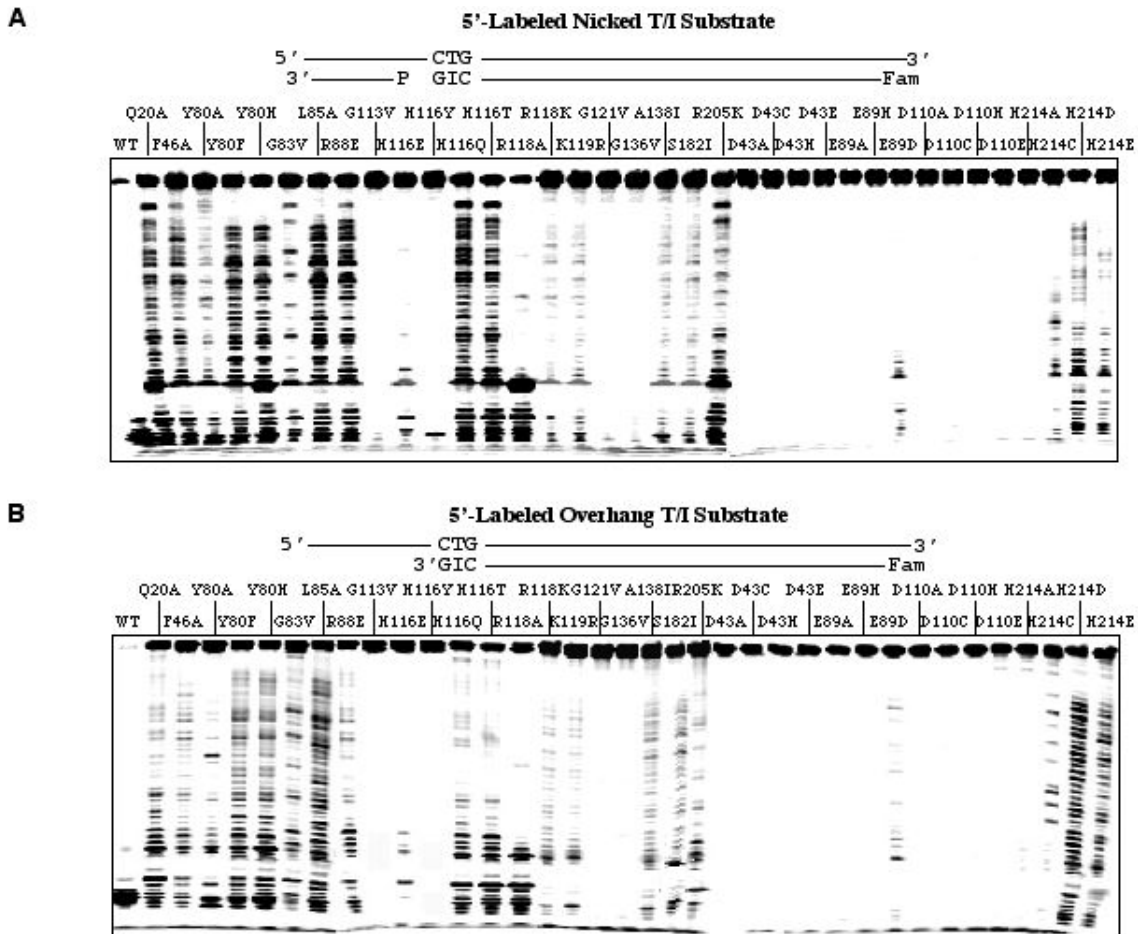
Figure 3.2 Cleavage activity of Tma endo V on 5'-labeled nicked and overhang T/I substrates.   Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzyme and 10 nM substrate.   A. Cleavage activity of wt and mutant Tma endo V on 5'-labeled nicked T/I substrate.   B. Cleavage activity of wt and mutant Tma endo V on 5'-labeled overhang T/I substrate.

This study attempts to understand the exonuclease activity in greater detail using thirty six mutants and the wt Tma endo V (Fig. 3.1A).   D43, E89, D110 and H214 are part of the active site that coordinates metal binding (*14*).   We initially tested the nuclease activity using a 5' fluorescently labeled inosine-containing substrate (Fig. 3.1B). The inosine was placed at the 11th position from the 3' side so that endo V cleavage

would generate a 51-mer since it cuts at the 3'-side one nucleotide from inosine. If a mutant exhibited 3'-exonuclease activity, a long ladder below the 51-mer was anticipated. Under the assay condition in which Mn^{2+} was used as a metal cofactor, all the mutants except D43A, D43C and D43H showed a strong 51-mer band, resulting from the specific inosine-dependent endonuclease activity (Fig. 3.1B). Due to nonspecific nuclease activities, the wt enzyme cleaved the specific 51-mer to small low molecular weight products (Fig. 3.1B). Similar to the previous observation (*13*), several mutants produced a ladder that was indicative of 3'exonuclease activity (Fig. 3.1B).

The active site mutants in general did not show the 3'-exonuclease activity except E89D, H214C, H214D and H214E (Fig. 3.1B). A few low activity mutants such as G113V, G121V and G136V also showed minimal ladder formation. Consistent with the substantial reduction of binding affinity to inosine-containing DNA, H116E and in particular H116Y had little exonuclease activity (Fig. 3.1B). Overall, the uniform ladders generated from several of the mutants confirmed that Tma endo V possesses 3'-exonuclease activity.

To further confirm the 3'-exonuclease activity, we assayed DNA cleavage using a synthesized nicked inosine-containing DNA with (Fig. 3.2A, T/I nicked substrate) or without the 3' downstream complementary DNA (Fig. 3.2B, T/I overhang substrate). Regardless of which substrate tested, they all showed a very similar laddering pattern, suggesting that the 3'-exonuclease activity did not depend on the 3' downstream complementary sequence. We then tested whether the same pattern holds for single-stranded inosine-containing substrates. Indeed, the wt enzyme and the mutants all

showed a similar laddering pattern with the full-length and the short single-stranded inosine-containing substrates (Fig. 3.S1, located at the end of this chapter).

Since all the substrates tested so far contained a single inosine in the sequences, we set out to determine how the wt enzyme and mutants cleave non-inosine-containing DNA. Using a double-stranded substrate in which the T/I base pair was replaced with a T/A base pair, we measured the nonspecific nuclease activities. As expected, the wt Tma endo V degraded the substrate to low molecular weight fragments due to nonspecific activities (Fig. 3.3A and (2)). The majority of the binding mutants showed varying degree of nonspecific activities as indicated by the low molecular weight fragments in the bottom of the gel (Fig. 3.3A, Q20A-R205K). The active site mutants exhibited little low molecular weight fragments (Fig. 3.3A, D43A-H214E), suggesting that the nonspecific activities were minimized. A very similar pattern was observed in the single-stranded non-inosine-containing substrate (Fig. 3.3B). A significant difference of the 3'-exonuclease pattern between the inosine-containing and non-inosine-containing substrates was noted in certain active site mutants (Figs. 3.1 and 3.3). In particular, H214D mutant only generated the well-formed ladder only with the inosine-containing DNA; however, not with non-inosine-containing DNA. This result indicates that the 3'-exonuclease activity requires the presence of an inosine site in the DNA.

Figure 3.3 Cleavage activity of Tma endo V on 5'-labeled non-inosine substrates. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzyme and 10 nM substrate.   A. Cleavage activity of wt and mutant Tma endo V on 5'-labeled full-length T/A substrate.   B. Cleavage activity of wt and mutant Tma endo V on 5'-labeled full-length single-stranded substrate.
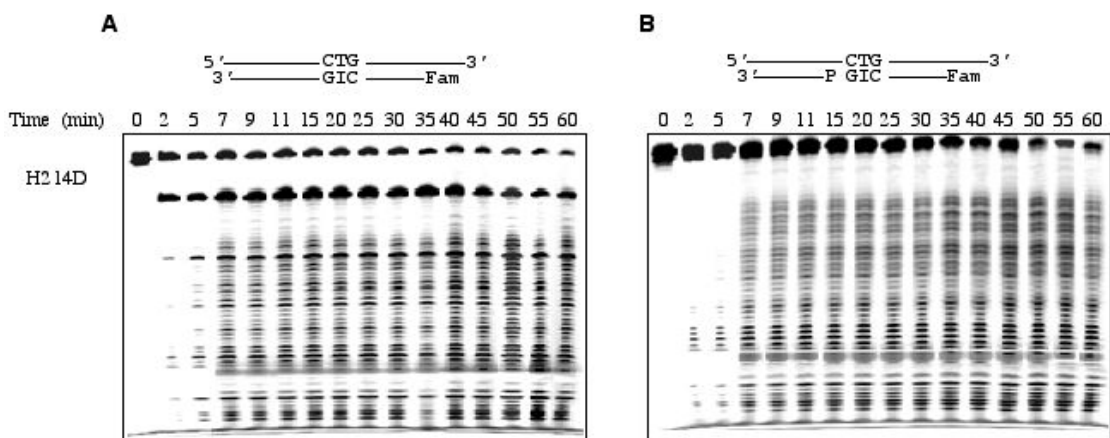
Figure 3.4 Time course analysis of cleavage activity by H214D on 5'-lableled full-length and nicked T/I substrates. Cleavage reactions were performed as described in Experimental Procedures with 100 nM H214D and 10 nM substrate. Reactions were stopped on ice at the indicated time points, followed by addition of an equal volume of GeneScan Stop Buffer. A. Time course analysis of cleavage activity by H214D on 5'-labeled full-length T/I substrate. B. Time course analysis of cleavage activity by H214D on 5'-labeled nicked T/I substrate.

The minimal nonspecific activity exhibited by the H214D mutant allowed us to better characterize the 3'-exonuclease activity without the interference of nonspecific activity as seen in the wt enzyme. Taking advantage of this unique nuclease property, we studied the kinetics of the 3'-exonuclease activity using the H214D mutant. At the initial time points, we observed significant cleavage at the inosine site to generate the 51-mer fragment and small amount of lower molecular weight fragments on a T/I substrate (Fig. 3.4A, 2-5 min). After that, a long ladder was produced by the 3'-exonuclease cleavage (Fig. 3.4A). Interestingly, the ladder did not seem to start right below the 51-mer band, instead, the vast majority of the ladder started around 43-mer position, leaving a 8-nucleotide gap in between (Fig. 3.4A). To verify this kinetic property, we then performed a time course analysis using the nicked inosine-containing substrate. The

106

cleavage of the nicked inosine-containing substrate essentially showed the same

laddering pattern, with a gap between the 51-mer and the start of the ladder (Fig. 3.4B).
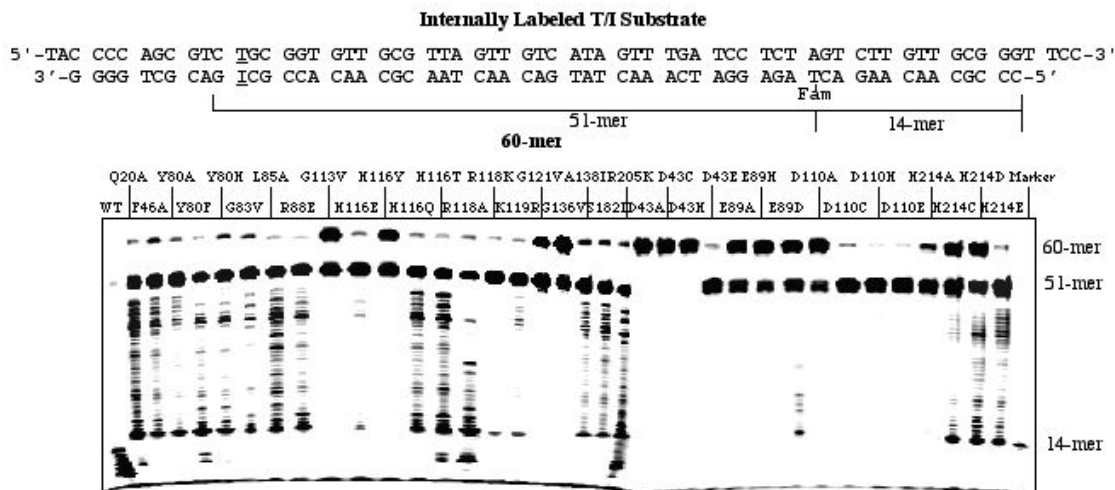
Figure 3.5 Cleavage activity of Tma endo V on internally labeled T/I substrate. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzyme and 10 nM substrate. The thymidine at position 14 from 5' end of the bottom strand was labeled with the Fam fluorophore.

Cleavage of internally labeled inosine-containing substrate

To better understand the 3'-exonuclease activity, we designed an internally labeled substrate in which the fluorescent group was moved from the 5' end to the 14th position thymine (Fig. 3.5). The overall cleavage pattern was similar to the 5'-end labeled T/I substrates. All the mutants except D43A, D43C and D43H generated a 51-mer specific inosine-dependent cleavage product at the 51-mer position (Fig. 3.5). The same set of binding and active site mutants showed a 3'-exonuclease ladder. A significant deviation from the cleavage of the internally labeled T/I substrate is that for many mutants in particular H214A, H214C and H214D the ladder halted at around 14-

mer position (Fig. 3.5).   The implication of these observations will be discussed in detail later.
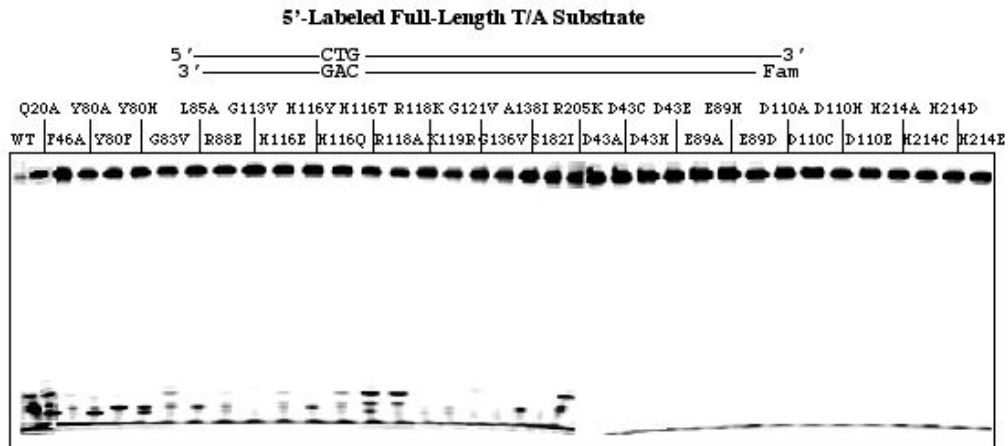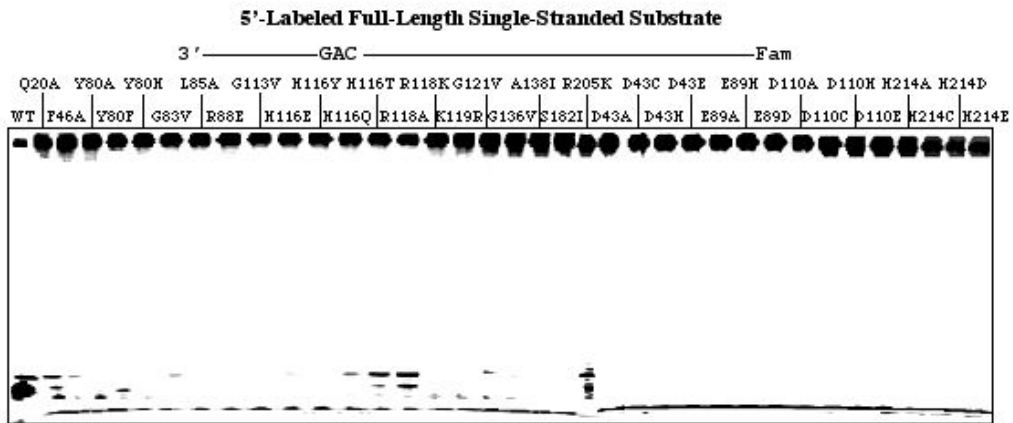

Figure 3.6   Cleavage activity of Tma endo V on 3'-labeled full-length T/I and T/A substrates. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzyme and 10 nM substrate.   The 3' end of the bottom strand was labeled with the Fam fluorophore.   A. Cleavage activity of wt and mutant Tma endo V on 3'-labeled full-length T/I substrate. B. Cleavage activity of wt and mutant Tma endo V on the 3'-labeled full-length T/A substrate.

Cleavage of 3'-labeled inosine- and non-inosine-containing substrates

To characterize any potential nuclease activities that initiates from the 5' side, we

switched the fluorescent label to the 3' side of the oligonucleotide substrate with the

108

inosine placed at the 33th position from the 3'-label (Fig. 3.6A). As expected, the endonuclease cleavage at the inosine site generated the 31-mer product (Fig. 3.6A). Again, the wt enzyme degraded the product to low molecular weight fragments as seen in the bottom of the gel. Many binding mutants in particular Q20A, H116T, R118A and R205K still showed nonspecific activity; however, only R118A and R205K showed complete degradation of the 31-mer. G113V, H116Y, R118K, K119R, G121V, G136V and S182I generated little low molecular weight fragments, indicating attenuated nonspecific activity (Fig. 3.6A).

We then tested the nuclease activities on the 3'-labeled non-inosine substrate. The wt enzyme exhibited nonspecific activities to degrade the DNA to low molecular weight fragments (Fig. 3.6B). Consistent with the results from the inosine-containing substrate, Q20A, H116T, R118A and R205K showed stronger nonspecific activity than other binding mutants (Fig. 3.6B, Q20A-R205K). The active site mutants exhibited minimal nonspecific activity (Fig. 3.6B, D43A-H214E).

To further understand the kinetic property of the nonspecific activity, we performed a time course analysis on the 3'-labeled non-inosine substrate. At the intial 5 sec time point, we observed small amount of high molecular weight fragments generated from degradation of the non-inosine substrate (Fig. 3.7). With increased incubation time, the intensities of the fragments also increased. After one min incubation, the high molecular weight fragments started to disappear while the bands of low molecular weight started to intensify. These results indicated that Tma endo V possesses endonuclease or exonuclease activity that initiates nonspecific degradation from the 5' end.

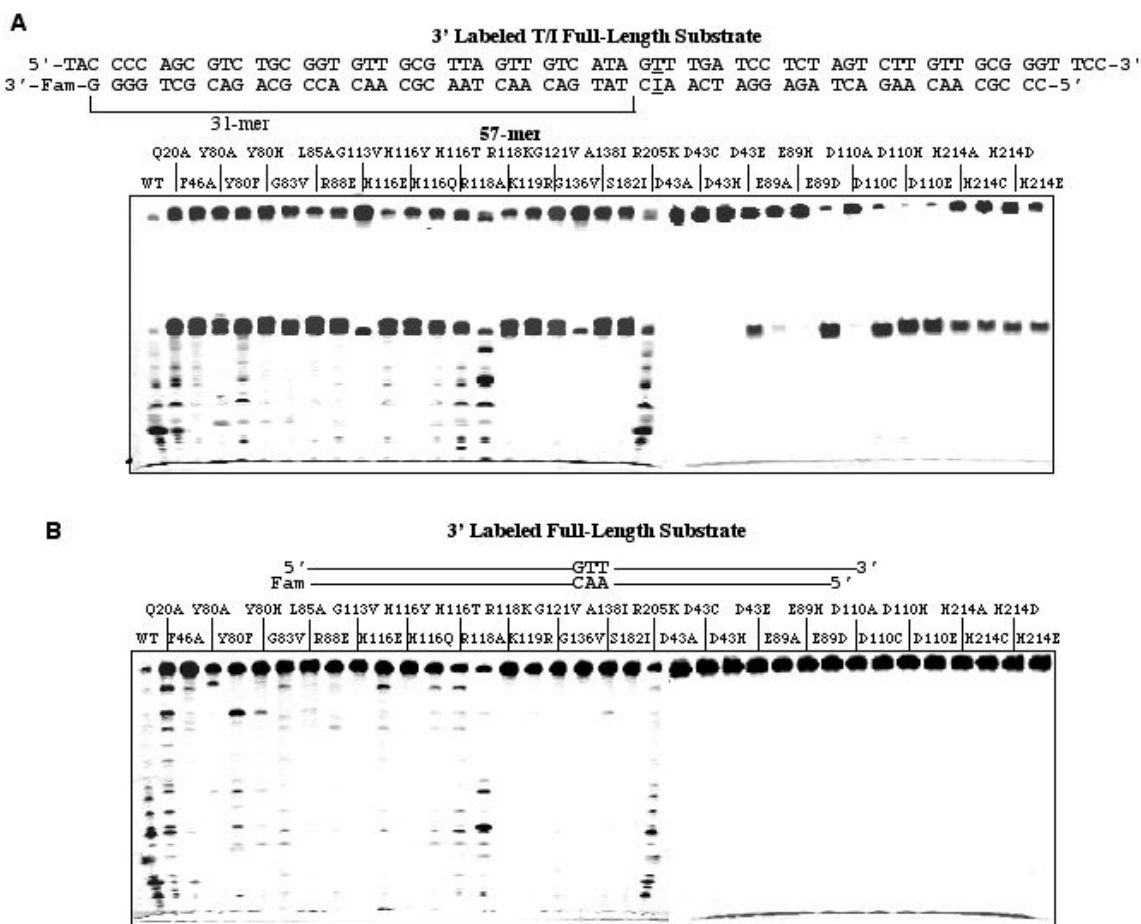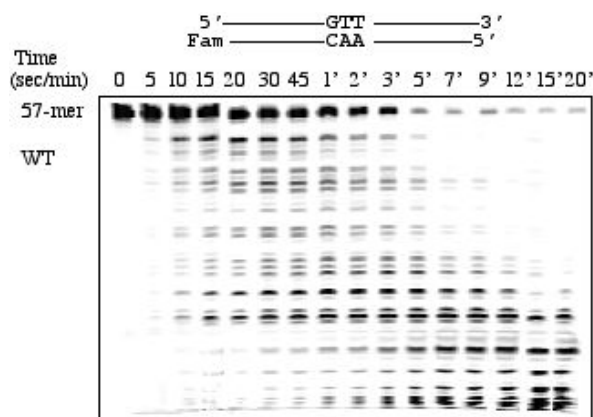Figure 3.7 Time course analysis of cleavage activity by wt Tma endo 3'-labeled full-length T/A substrate. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzyme and 10 nM substrate. Reactions were stopped on ice at the indicated time points, followed by addition of an equal volume of GeneScan Stop Buffer.

5. Discussion

Endonuclease V is a versatile DNA repair enzyme with multiple nuclease activities on different DNA substrates, including deaminated bases and mismatched base pairs. Previous studies have indicated that endo V possesses 5'-exonuclease, nonspecific endonuclease, 3'-exonuclease activities, in addition to the inosine-specific endonuclease activity that nicks the inosine-containing strand at the one nucleotide downstream of the lesion from the 3' side (*3, 4, 18-20*). However, the nature of the multiple nuclease activities is not defined. This study takes advantage of the site-directed mutant reportire we accumulated previously and uses different end labeled and internally labeled oligonucleotide substrates to investigate the nuclease activities. The biochemical data presented here allows us to dissect the nature of the multiple nuclease activities.

Inosine-dependent and nonspecific endonuclease activity

Endo V is known as a deaminated base repair enzyme. The inosine-dependent endonuclease activity is well documented and is a relevant enzymatic activity *in vivo* (*1, 3-5, 8, 9, 11, 21, 22*). The inosine-dependent endonuclease activity is the primary enzymatic activity that is well maintained within the protein structure. This is evidenced by the retention of the activity by most of the Tma endo V mutants tested except three active site mutants (D43A, D43C, D43H) (Figs. 3.1 and 3.S1). Tma endo V is also active on a nonspecific circular plasmid in particular when Mn^{2+} is used as a metal cofactor (*3*). Thus, Tma endo V is an inosine-dependent endonuclease with nonspecific endonuclease activity.

Inosine-dependent 3'-exonuclease activity

Previously we detected 3'-exonuclease activity from Tma endo V using binding mutants (*13*). Here we establish that the 3'-exonuclease activity is inosine-dependent. Support for this notion primarily comes from biochemical analysis of a few active site mutants. H214C, H214D and H214E failed to show any ladder on the 5'-labeled non-inosine substrate; however, formed a distinct ladder on all 5'-labeled inosine-containing substrate (Figs. 3.1-3.3 and 3.S1), suggesting that the 3-exonuclease activity relies on the initial primary cut at the inosine site to initiate the secondary enzymatic action. It is well recognized that Tma endo V remains bound to the nicked product after cleavage at the inosine site (*1, 3, 7, 13*). The affinity to the inosine site appears to be essential for initiating the 3'-exonuclease action.

The more interesting kinetic property of the inosine-dependent 3'-exonuclease activity is revealed from the time course analysis using the H214D mutant that has minimal nonspecific nuclease activity (Fig. 3.4). First, the vast majority of the cleavage bands appeared around 8-nt below the 51-mer. These results indicate that the first hydrolysis reaction may not occur at the first nucleotide; instead, the enzyme makes a first cut at a few nucleotides at the 5'-side from the inosine specific cleavage, leaving a 8-nt gap in between. After the initial cut to remove the short inosine-containing oligonucleotide fragment, the enzyme then proceeds at the 3' to 5' direction in a complete 3'-exonuclease manner. By cleaving one nucleotide at a time on the 5'-labeled DNA, the fall-off of the enzyme at different sites allows formation of a laddering pattern. The laddering pattern shows little change over time, i.e., the pattern at 60 min is quite similar to that at 7 min (Fig. 3.4B). This indicates that the 3'-exonuclease action only occurs once on an inosine-containing DNA. Once the enyzme removes the short inosine-containing fragment and falls off the DNA, it has little chance to re-bind to the 3'-side to reinitiate the 3'-exonulcease action. This kinetic pattern reinforces the notion that the 3'-exonuclease activity is inosine-dependent.

An alternative explanation for the even laddering pattern throughout the time course is that it is caused by singularly random endonuclease action that allows the enzyme to jump randomly in the 3' to 5' direction at a downstream site after the first cut to remove the short inosine-containing fragment. We found this cut and jump model not plausible in comparison with the above cut and chew model due to the distinct even one-nt laddering pattern (Fig. 3.4). More importantly, if the enzyme adopts the cut and jump

112

model, we could have seen the ladder extends beyond the 14-mer mark in the internally labeled substrate (Fig. 3.5).   Instead, the ladder stopped at the 14-mer position for most of the mutants (Fig. 3.5).

Does the wt enzyme possess nonspecific 3-exonuclease activity?   Although the data presented here may not give an absolute answer.   We tend to believe the answer is no based on the following reasonings. First, no obvious ladders were observed with 5'-labeled nonspecific substrates (Fig. 3.3).   Second, the time course data from the 3'-labeled substrate does not show appearance of low molecular weight fragments at the bottom of the gel (Fig. 3.7), which would indicate nonspecific 3'-exonulcease activity. Last, the experimental data presented below already demonstrates that the enzyme tracks DNA in a 5' to 3' direction.

5'-exonuclease activity

It has been proposed that Tma or *E. coli* endo V enzymes contain 5'-exonuclease activity (*18-20*).   We measured the 5'-exonuclease activity using 3'-labeled substrates. A ladder formed by several mutants such as Q20A and R205K is a good indication of 5'-exonuclease activity (Fig. 3.6).   The 5'-exonuclease activity is nonspecific since it does not dependent on the existence of an inosine site (Fig. 3.6).   A more definitive proof of the 5'-exonuclease activity comes from the time course analysis of the 3'-labeled subtrate (Fig. 3.7).   At the initial time points, the 5'-exonuclease action produces a ladder consisting of higher molecular weight fragments (Fig. 3.7, 5-20 sec).   At the later time points, the continuous 5'-exonuclease action results in the disappearance of the higher

molecular weight fragments with concurrent accumulation of lower molecular weight fragments (Fig. 3.7, 5-20 min).   The detail kinetic analysis also indicates that the mode of action of the 5'-exonuclease is somewhat different from the inosine-dependent 3'-exonuclease discussed above.   Whereas the 3'-exonuclease action seems to remove one nucleotide at a time, the nonspecific 5'-exonuclease generates a combination of one nucleotide and oligonucleotide products.

Nuclease activities and inosine repair

Since endo V cuts at the 3' side of the lesion, the inosine damage still remains in the DNA.   Additional enzymatic actions are required for the removal of the damage. Based on the result obtained from an oligonucleotide-based study in *E. coli*, after nicking by the inosine-dependent endonuclease activity of *E. coli* endo V, two-nt from 5' side and three nt from 3' side are removed to create a gap (*12*).   The enzymes responsible for the removal of inosine and subsequent creation of the 5-nt gap in the repair process are not identified yet.   However, the inosine-dependent 3'-exonuclease and nonspecific 5'-exonuclease raise the possibility that endo V, is facilitated by another protein factor, and if so could potentially be involved in creation of the short gap during the repair process. If this is the case, then endo V can latch on to a DNA and search for an inosine site. Once the enzyme nicks at the 3' side of an inosine site, it may recruit a protein partner to allow it use its inosine-dependent 3'-exonuclease and nonspecific 5'-exonuclease to remove the inosine in the DNA.   The gap created by the multiple enzymatic actions can then be filled by a DNA polymerase and the resulting nick sealed by a DNA ligase.

114

Interestingly, in the case of mismatch repair, human MutLa is found to be a Mn^{2+}-dependent endonuclease that can be activated in the presence of a mismatch, ATP and multiple protein factors (*23*). Of course, it is also possible that the *in vitro* exonuclease activities are not relevant to the *in vivo* process. In this scenario, independent nuclease(s) instead of endo V is involved in the post-inosine cleavage process. Regardless of whether the multiple nuclease activities are physiologically relevant, this work reveals a series of concerted enzymatic functions that may define the inner working of endonuclease V.

Acknowledgments

6. References

1. Feng, H., Klutz, A. M., and Cao, W. (2005) Active Site Plasticity of Endonuclease V from Salmonella typhimurium, *Biochemistry 44*, 675-683.

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

3. Huang, J., Lu, J., Barany, F., and Cao, W. (2001) Multiple Cleavage Activities of Endonuclease V from Thermotoga maritima: Recognition and Strand Nicking Mechanism, *Biochemistry 40*, 8738-8748.

4. Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli, *J Biol Chem 269*, 16260-16268.

5. Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Deoxyinosine 3' endonuclease, a novel deoxyinosine-specific endonuclease from Escherichia coli, *Ann N Y Acad Sci 726*, 315-316.

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

7. Huang, J., Lu, J., Barany, F., and Cao, W. (2002) Mutational analysis of endonuclease V from Thermotoga maritima, *Biochemistry 41*, 8342-8350.

8. Yao, M., and Kow, Y. W. (1995) Interaction of deoxyinosine 3'-endonuclease from Escherichia coli with DNA containing deoxyinosine, *J Biol Chem 270*, 28609-28616.

9. Guo, G., and Weiss, B. (1998) Endonuclease V (nfi) mutant of Escherichia coli K-12., *J Bacteriol 180*, 46-51.

10. Weiss, B. (2001) Endonuclease V of Escherichia coli prevents mutations from nitrosative deamination during nitrate/nitrite respiration, *Mutat Res 461*, 301-309.

11. Schouten, K. A., and Weiss, B. (1999) Endonuclease V protects Escherichia coli against specific mutations caused by nitrous acid, *Mutat Res 435*, 245-254.

12. Weiss, B. (2008) Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation, *DNA Repair (Amst) 7*, 205-212.

13.   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.

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

15.   Lin, J., Gao, H., Schallhorn, K. A., Harris, R. M., Cao, W., and Ke, P. C. (2007) Lesion recognition and cleavage by endonuclease V: a single-molecule study, *Biochemistry 46*, 7132-7137.

16.   Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymease chain reaction, *Gene 77*, 51-59.

17.   Hitchcock, T. M., Dong, L., Connor, E. E., Meira, L. B., Samson, L. D., Wyatt, M. D., and Cao, W. (2004) Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase, *J Biol Chem 279*, 38177-38183.

18.   Gao, H., Huang, J., Barany, F., and Cao, W. (2007) Switching base preferences of mismatch cleavage in endonuclease V: an improved method for scanning point mutations, *Nucleic Acids Res 35*, e2.

19.   Pincas, H., Pingle, M. R., Huang, J., Lao, K., Paty, P. B., Friedman, A. M., and Barany, F. (2004) High sensitivity EndoV mutation scanning through real-time ligase proofreading, *Nucleic Acids Res 32*, e148.

20.   Yao, M., and Kow, Y. W. (1996) Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from Escherichia coli, *J Biol Chem 271*, 30672-30676.

21.   Yao, M., and Kow, Y. W. (1997) Further characterization of Escherichia coli endonuclease V, *J Biol Chem 272*, 30774-30779.

22.   Moe, A., Ringvoll, J., Nordstrand, L. M., Eide, L., Bjoras, M., Seeberg, E., Rognes, T., and Klungland, A. (2003) Incision at hypoxanthine residues in DNA by a mammalian homologue of the Escherichia coli antimutator enzyme endonuclease V, *Nucleic Acids Res 31*, 3893-3900.

23.   Kadyrov, F. A., Dzantiev, L., Constantin, N., and Modrich, P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair, *Cell 126*, 297-308.

A

Full-Length Single-Stranded I-Containing Substrate

3'————————GIC————————————————————————Fam

| | Q20A | Y80A | Y80H | L85A | G113V | H116Y | H116T | R118K | G121V | A138I | R205K | D43C | D43E | | E89H | D110A | D110H | H214A | H214D |
| WT | F46A | Y80F | G83V | R88E | H116E | H116Q | R118A | K119R | G136V | S182I | D43A | D43H | | E89A | E89D | D110C | D110E | H214C | H214E |

B

Short Single-Stranded I-Containing Substrate

3'GIC————————————————————————————Fam

| | Q20A | Y80A | Y80H | L85A | G113V | H116Y | H116T | R118K | G121V | A138I | R205K | D43C | D43E | | E89H | D110A | D110H | H214A | H214D |
| WT | F46A | Y80F | G83V | R88E | H116E | H116Q | R118A | K119R | G136V | S182I | D43A | D43H | | E89A | E89D | D110C | D110E | H214C | H214E |

Figure 3.S.1 Cleavage activity of Tma endo V on 5'-labeled single-stranded and short I-containing substrates. A. Exonuclease activity of Tma endo V mutants on 5' labeled single-stranded I-containing substrate. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzymes and 10 nM substrate. B. Exonuclease activity of Tma endo V mutants on 5' labeled short I-containing substrate. Cleavage reactions were performed as described in Experimental Procedures with 100 nM enzymes and 10 nM substrate.

CHAPTER FOUR

HUMAN ENDONUCLEASE V AS A DEAMINATION REPAIR ENZYME

1. Abstract

Human endonuclease V located in chromosome 17q25.3 is a 282 amino acid protein that shares approximately 30% sequence identity with bacterial endonuclease V. This study reports biochemical properties of human endonuclease V with respect to repair of deaminated base lesions. Using soluble proteins from recombinant sources, we determined repair activities of human endonuclease V on inosine (I)-, xanthosine (X)-, oxanosine (O)- and uridine (U)-containing DNA. Human endonuclease V is most active with inosine-containing DNA; however, with minor activity on xanthosine-containing DNA. Endonuclease activities on oxanosine and uridine were not detected. The endonuclease activity on inosine-containing DNA follows the order of single-stranded I > G/I > T/I > A/I > C/I. The preference of the catalytic activity correlates with the binding affinity of these inosine-containing DNA. Mg^{2+} and to a much less extent, Mn^{2+}, Ni^{2+}, Co^{2+} can support the endonuclease activity. Introduction of human endonuclease V into *Escherichia coli* cells caused two-fold reduction in mutation frequency. This is the first report of deaminated base repair activity from human endonuclease V. The relationship between the endonuclease activity and deaminated adenine (inosine) repair is discussed.

120

2. Introduction

Endonuclease V (endo V) was initially discovered in *Escherichia coli* (*E. coli*) as a nuclease that acted on a variety of DNA damage (*1,2*). In the course of identifying hypoxanthine DNA repair activity, *E. coli* endo V was rediscovered as a deoxyinosine 3' endonuclease (*3,4*). Endo V in general hydrolyzes the second phosphodiester bond 3' to the aberrant site (*3,5-11*). Inosine (known as hypoxanthine as a base), xanthosine and oxanosine, and uridine are deamination products derived from adenosine, guanosine, and cytidine (*12-15*). In addition of endonuclease activity on DNA containing inosine (*5,7,9,16*), endo V was also found active on xanthosine (*5,17-19*), uridine (*2,4,5,7,20*), and oxanosine (*5,6,17*). Genetic analysis indicates that endo V is involved in repair of inosine, xanthosine, and N^6-hydroxylaminopurine (*21-23*). Homologs from endo V family proteins are diverse in substrate specificity. While bacterial endo V enzymes exhibit broad endonuclease activity towards different deaminated bases, endo V proteins from archaea species *Archaeoglobus fulgidus* and mouse seem to only active on inosine-containing DNA (*8,9*). Endo V from *Salmonella typhimurium* appears to possess high affinity to deaminated bases as it is the only endo V enzyme tested that shows detectable binding to oxanosine-containing DNA (*5,6*). Interestingly, bacterial endo V enzymes also demonstrate endonuclease activity on mismatches (*5,7,10,24*), small insertions/deletions (indel) (*11*), flap and pseudo-Y structures (*11*). The mismatch and indel cleavage activity has been exploited for development of mutation identification or scanning methods (*25-28*).

The structure-functional relationship has been extensively studied using endo V from the thermophilic bacterium *Thermotoga maritima* (Tma) as a model system. Sequence alignment has identified seven conserved motifs in endo V family proteins (Fig. 3.1A). D43 in motif II, E89 in motif III, D110 in motif IV, and H214 in motif VII are identified as catalytic residues involved in coordination of metal ion (*17,29*). Based on peculiar cleavage pattern occurred when Mn^{2+} was used as a metal cofactor and two-metal competition assays, it was proposed that the active site of Tma endo V may contain two metal binding sites. This model posits that the high affinity site (M1) holds a catalytic metal ion and the low affinity site (M2) retains another metal ion to modulate the nuclease activity (*6,30*). Through a large-scale site-directed mutagenesis analysis, Y80, G83 and L85 in motif III, G113, H116, R118 and G121 in motif IV, G136 and A138 in motif V, and S182 in motif VI were identified as residues that affect protein-DNA interactions (*17*). The importance of Y80 in base recognition has been demonstrated by switching of base preference in mismatch cleavage by Y80A mutant (*24*).

Endo V family proteins are ubiquitous in bacteria, archaea, and eykaryotes (Fig. 3.1A). The human genome contains an endo V homolog located in chromosome 17q25.3. The biochemical and enzymatic properties of human endo V are not known. In this study, we report that human endo V is a deoxyinosine and deoxyxanthosine endonuclease that cleaves the second phosphodiester bond at the 3' side of the lesion. The single-stranded inosine endonuclease activity is 14-fold stronger than the single-stranded xanthosine endonuclease activity. Human endo V is active with Mg^{2+}, Mn^{2+}

and to a much less extent Ni^{2+} and Co^{2+} as a metal cofactor. The biochemical analysis indicates that human endo V possesses robust inosine endonuclease activity that may play an important role for repair of deaminated purine damage *in vivo*.

3. 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, *Taq* DNA polymerase, Phusion hifidelity polymerase and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). BSA and dNTPs were purchased from Promega (Madison, WI). Anti-His (N-term) antibody and anti-rabbit IgG, HRP-linked antibody were purchased from Cell Signaling Technology (Danvers, MA). The horseradish peroxidase substrate Opti-4CN for western blot and PVDF membrane were purchased from Bio-Rad (Hercules, CA). HiTrap chelating and Q columns were purchased from GE Healthcare (Piscataway, NJ). Oligodeoxyribonucleotides were ordered from Integrated DNA Technologies Inc. (Coralville, IA). LB medium was prepared according to standard recipes. Human endo V sonication buffer consisted of 50 mM Tris HCl (pH 7.4), 1 mM EDTA (pH 8.0), 2.5 mM DTT, 0.15 mM PMSF, 10% glycerol and 50 mM NaCl. GeneScan stop buffer consisted of 80% formamide (Amresco, Solon, OH), 50 mM EDTA (pH 8.0), and 1% blue dextran (Sigma Chemicals). TB buffer (1 x) consisted of 89 mM Tris base and 89 mM boric acid. TE buffer consisted of 10 mM Tris-HCl (pH 8.0), and 1 mM EDTA. *E. coli* host strain

BL21(DE3)Δ3 (F⁻, *ompT*, *hsdS_B*, (r_B⁻ m_B⁻), *gal*, *dcm*, *sly,* (DE3), *nfi*, *ung*, *mug*) and JM109 (e14⁻(McrA⁻) *endA1*, *recA1*, *gyrA96*, *thi-1*, *hsdR17* (r_k⁻, m_k⁺), *supE44*, *relA1* Δ(*lac-proAB*), (F⁻, *traD36*, *proAB*, *lacI*qZΔM15)) are from our laboratory collection. Plasmid pET28a-hnfi was constructed by PCR amplification of human cDNA. *E. coli* wild type K-12 strain was obtained as a generous gift from *E. coli* Genetic Stock Center in Yale University (New Haven, CT)

Confirmation of the genotype of BL21(DE3)Δ3.

The *E. coli* BL21(DE3)Δ3 strain was confirmed by PCR with wild type *E. coli* K-12 stain as positive control using the following primers. Ec.NFI.F 5'- TAA AGT ACC <u>CCA TGG</u> GTG ATT ATG GAT CTC GCG TC-3'; the *Nco*I site is underlined. Ec.NFI.R 5'- TAA AGG GT<u>G GAT CC</u>T AGG GCT GAT TTG CTG T-3'; the *Bam*HI site is underlined. Ec.MUG.F 5'-TGG GGT ACC <u>CCA TGG</u> GTT GAG GAT ATT TTG GCT CCA GGG-3'; the *Nco*I site is underlined. Ec.MUG.R 5'-CCC <u>GGA TCC</u> TTA TCG CCC ACG CAC TAC CAG CGC CTG GTC-3'; the *Bam*HI site is underlined. Ec.UNG.F 5'-GGG AAT TC<u>C ATA TG</u>G CTA ACG AAT TAA CCT GGC ATG AC-3'; the *Nde*I site is underlined. Ec.UNG.R 5'-CCC <u>AAG CTT</u> CTC ACT CTC TGC CGG TAA TAC TGG-3'; the *Hind*III site is underlined.The PCR mixtures (50 µL) contained 40 ng of genomic DNA as template, 200 nM each primer pair, 50 µM each dNTP, 1 × *Taq* DNA polymerase buffer, and 1 unit of *Taq* DNA polymerase. The PCR procedure was composed of a predenaturation step at 95°C for 2 min, 30 cycles with each cycle consisting of denaturation at 94°C for 15 s, annealing at 56°C for 30 s, and extension at

72°C for 1 min, and a final extension step at 72°C for 10 min. The PCR products were electrophoresed on 1% agarose gel.

Plasmid construction, cloning, and expression of human endo V.

The human endo V gene in pET28a-hnfi and plasmid pET32a were digested with *Nco*I and *Eco*RI. The fragments containing human endo V (*nfi*) gene and digested plasmid pET32a, recovered from agarose gel, were purified with Gene Clean 2 Kit (MP Biomedicals) and ligated according to the manufacturer's instruction manual. The ligation mixture was transformed into *E. coli* strain JM109 competent cells prepared by a CaCl$_2$ method (*31*).

To express the N-terminal His-6-tagged human endo V gene, pET32a-hnfi was transformed into *E.coli* strain BH21(DE3)Δ3 by standard protocol (*31*). A single colony of BH21(DE3)Δ3 containing pET32a-hnfi was selected to inoculate into LB medium and incubate at 37°C. This overnight *E. coli* culture was diluted 100-fold into LB medium (1 Liter) supplemented with 50 μg/mL ampicillin. The *E. coli* cells were grown at 37°C while being shaken at 250 rpm until the optical density at 600 nm reached about 0.6. IPTG was added to a final concentration of 0.5 mM. After growing at room temperature for an additional 16 h, the cells were collected by centrifugation at 4,000 rpm with JS-4.2 rotor in J6-MC centrifuge (Beckman Coulter) at 4°C and washed once with precooled sonication buffer.

To purify the human endo V protein, the cell paste from a 1 L culture was suspended in 10 mL of sonication buffer and sonicated at output 5 for 3 x 1 min with 5

125

min rest on ice between intervals. The sonicated solution was clarified by centrifugation at 12,000 rpm with JA-17 rotor in Avanti J-25 centrifuge (Beckman Coulter) at 4°C for 20 min. The supernatant was transferred into a fresh tube and loaded into a 1 mL HiTrap chelating column. The bound protein in the column was eluted with a linear gradient of 15 column volumes of 0-1 M imidazole in chelating buffer A (20 mM Tris-HCl (pH 7.6), 10% glycerol and 50 mM NaCl) using a Bio-Rad BioLogic chromatographic system.

Fractions (200-400 mM imidazole) containing the human endo V protein as seen on 15% SDS-PAGE were pooled and dialyzed against HiTrap Q column buffer A (20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 10% glycerol and 0.2 mM DTT) overnight at 4°C. The dialysis sample was then loaded onto a 1 mL HiTrap Q column and eluted with a linear gradient of 15 column volumes of 0-1 M of NaCl in HiTrap Q buffer A. The putative human endo V protein was eluted at 200-400 mM NaCl. The homogeneity of the protein was examined by 15% SDS-PAGE analysis. Fractions (200-400 mM NaCl) containing the human endo V protein as seen on 15% SDS-PAGE were pooled and dialyzed against HiTrap chelating column buffer A (20 mM Tris-HCl (pH 7.6), 10% glycerol and 50 mM NaCl) overnight at 4°C. The dialysis sample was then loaded onto a 1 mL HiTrap chelating column and eluted with a linear gradient of 15 column volumes of 0-1 M imidazole in chelating buffer A. The putative human Endo V protein was eluted at 200-400 mM imidazole. The homogeneity of the protein was examined by 15% SDS-PAGE analysis. The human endo V protein concentration was determined on SDS-PAGE using BSA as a standard.

Site-directed mutagenesis.

An overlapping extension PCR procedure was used for construction of D52A mutant (32). The first round of PCR was carried out using pET32a-hnfi as template DNA with two pairs of primers, Hnfi.01F (5' TCA GGT ACC <u>CCA TGG</u> CC CTG GAG GCG GCG GG 3', *Nco*I site underlined) and HVD52A.02R (5' GGG TCG GGG GCG TT<u>G CC</u>G TGT CCT TCG TGA AAG GGG A3', the D52A site is underlined) pair & Hnfi.04R (5' TCA AGC TTT <u>GAA TTC</u> ATT ACA AAG TGC TGA GGA CTC TC 3', *Eco*RI site underlined) and HVD52A.03F (5' CAG GGG CCG GGG TTC <u>AG</u>G CCG ACG AAA AGG ACT TCC T 3', the D52A site is underlined) pair.   The PCR mixtures (50 µL) contained 1 ng of pET32a-hnfi DNA as template, 200 nM each primer pair, 50 µM each dNTP, 1 × *Taq* DNA polymerase buffer, and 1 unit of *Taq* DNA polymerase. The PCR procedure was composed of a predenaturation step at 95°C for 2 min, 30 cycles with each cycle consisting of denaturation at 94°C for 15 s, annealing at 53°C for 30 s, and extension at 72°C for 1 min, and a final extension step at 72°C for 10 min.   The PCR products were electrophoresed on 1% agarose gel and the expected PCR fragments were purified from gel slices by spin-squeeze method.   This second run of PCR reaction mixture (100 µL), which contained 3 µL of each of the first run PCR fragment, 50 µM each dNTP, 1 × Phusion hifidelity polymerase buffer, and 2 units of Phusion hifidelity polymerase, was initially carried out with a predenaturation step at 98°C for 2 min, five cycles with each cycle of denaturation at 94°C for 15 s and annealing and extension at 60°C for 4 min, and a final extension at 72°C for 5 min.   Afterward, 100 nM outside primers (Hnfi.01F and Hnfi.04R) were added to the above PCR reaction mixture to

continue the overlapping PCR reaction with the same condition for 25 cycles. The purified PCR products, digested with a pair of *Nco*I *and Eco*RI endonucleases, were ligated to cloning vector pET32a treated with the same pair of restriction endonucleases. The recombinant plasmids containing the desired mutations were confirmed by DNA sequencing and transformed into *Escherichia coli* host strain BH21(DE3)Δ3 for expression and protein purification.

Western blot analysis.

Western Blot analysis was carried out using an antibody raised against the N-terminal His-tag to confirm that the protein was overexpressed in *E. coli* cells. The protein samples were first separated on 15% SDS-PAGE, and then transferred onto a PVDF membrane by electro-blotting at 100 V for 1 h using a Bio-Rad Mini Trans-Blot apparatus. The membrane was blocked with 1% low-fat milk. Anti-His (N-term) antibody (1 µl) diluted in 5 ml of 1% BSA solution was added onto the membrane sealed in a plastic bag at room temperature for 1 h while shaking gently. After washing, anti-rabbit IgG, HRP-linked antibody (1 µl) diluted in 5 ml of 1% BSA solution was added onto the membrane sealed in a plastic bag at room temperature for 1 h with shaking gently. The color reaction was developed using Opti-4CN as a substrate.

Oligodeoxynucleotide substrates.

The fluorescently labeled inosine- and uridine-containing substrates were prepared as described (*7*). The sequences of the oligonucleotides are shown in Figure

128

2A.   The oligodeoxyribonucleotides were dissolved in TE buffer at a final concentration

of 10 μM.   The two complementary strands with the unlabeled strand in 1.2-fold molar

excess were mixed, incubated at 85°C for 3 min, and allowed to form duplex DNA

substrates at room temperature for more than 30 min.   The fluorescently labeled

xanthosine- and oxanosine-containing substrates were constructed as previously

described (*5,33*).

Human endonuclease V activity assay.

DNA cleavage assays for human endo V were performed at 37°C for 60 min in a

10 µL reaction mixture containing 10 nM oligonucleotide substrate, an indicated amount

of human endo V, 10 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 1 mM EDTA, 5 mM

$MgCl_2$, and 2% glycerol.   Reactions were quenched by addition of an equal volume of

GeneScan stop buffer.   Samples were treated at 94°C for 3 min and 3.6 µL of samples

were loaded onto a 7 M urea-10% denaturing polyacrylamide gel.   Electrophoresis was

conducted at 1500 V for 1.5 h using an ABI 377 sequencer (Applied Biosystems).

Cleavage products and remaining substrates were quantified using GeneScan analysis

software.

Gel mobility shift assay.

The binding reactions were performed at 37°C for 60 min in a 10-µl volume

containing 50 nM DNA substrate, 10 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol, 5 mM

$MgCl_2$, and 2% glycerol, and 500 nM of human endo V protein.   Samples were

supplemented with 5 µl of 50% glycerol and electrophoresed at 200 V on a 6% native polyacrylamide gel in 1 x TB buffer (89 mM Tris base and 89 mM boric acid) supplemented with 5 mM EDTA. The bound and free DNA species were analyzed using a Typhoon 9400 Imager (GE Healthcare) with the following settings: photomultiplier tube at 600 V, excitation at 495 nm, and emission at 535 nm.

Spontaneous mutation frequency assay.

A single colony was selected, inoculated into 4 mL liquid LB supplemented with 50 μg/mL ampicillin and grown at $30^{°}$ C overnight (*34*). IPTG was added to a final concentration of 0.5 mM. The culture continued to grow at $37^{°}$ C for an additional 5 h. The cell culture was diluted and plated on LB plates with 50 μg/mL ampicillin. One mL cell culture was mixed with 3 mL 0.7% soft agar and plated on LB plates with both 50 μg/mL ampicillin and 100 μg/mL rifampicin. The plates were incubated at $37^{°}$ C for 24 h and cell numbers on amp^R and rif^R plates were counted respectively. The mutation frequency was calculated as the results of rif^R colony number per 10^8 amp^R colony number.

A

```
                        D52A
                          ▼
Hsa  /-20-QARLKAH-19-RVGGVDVSF-33-APYVSGFLAFRE-19-PQVLLVDGHGVLHHRGFGVACHLG-5-PCVGVAKKLLQV-45-PLYISVGHR-16-RIPEPVRQAD-42-/
Mmu  /-20-QARLKAR-19-KVGGVDVSF-33-APYVSGFLAFRE-19-PQVVLVDGHGVLHQRGFGVACHLG-5-PCIGVAKKLLQV-45-PLYVSVGHR-16-RIPEPIRQAD-98-/
Cel  /-435-QIRERLI-17-LVAGIDTSA-33-LPYIPQYLAVRE-19-PDVILCDGFGFHSRGCGMACHVG-5-ASIGVAKNLTLH-51-GVFVSAGYG-18-TTIEPIRAAD-18-/
Ath  /-23-QDELKKK-25-YVGGVDMSF-33-VPYVPGFLAFRE-19-PQVLMVDGHGILHPRGFGLACHLG-5-PTIGVAKNLHHV-45-PIYVSVGHR-16-RVPEPIRQAD-28-/
Spo  /-16-QIELKGM-14-YVVGLDISF-33-EDYVPGFLSFRE-12-IDIILVDGHGVLHPVGFGLACHLG-5-PVVGVAKNYLHC-51-PVYVSIGNQ-18-RVPEPIRQAD-16-/

Tma  /-19-QNELRKK-11-YVAGVDLSF-31-FPYIPGLLAFRE-14-PDVVVFDGQGLAHPRKLGIASHMG-5-PTIGVAKSRLYG-33-PIFVSPGHL-18-RIPEPTRLAH-21-/
Eco  /-11-QIELASS-13-LIAGADVGF-32-MPYIPGFLSFRE-14-PDLVFVDGHGISHPRRLGVASHFG-5-PTIGVAKKRLCG-33-PLFIATGHR-17-RLPEPTRWAD-17-/
Sty  /-09-QIELASS-13-FIGGADVGF-32-MPYIPGFLSFRE-14-PDLLFVDGHGISHPRRLGVASHFG-5-PTIGVAKKRLCG-33-PLFIATGHR-17-RLPEPTRWAD-17-/
Ype  /-22-QRQRASE-13-FIAGADVGF-32-LPYIPGLLSFRE-14-PDLILVDGQGIAHPRRLGVASHFG-5-PTIGVAKSRLCG-33-PLFISPGHR-17-RLPEPTRWAD-27-/
Sco  /-22-QDELRAR-13-RVTGVDVRY-32-FPYVPGLLAFRE-14-PGLVVCDGYGLAHPRRFGLASHLG-5-PTIGVAKNPFTT-33-PVFVSVGHR-17-RLPETTRRAD-14-/
Bsu  /-20-QFHLKNR-13-TGAGVDLAY-34-VPYVSGFLAFRE-14-PDVFLFDGHGYLHYNHMGVATHAA-5-PTIGIAKTYLKI-33-PIFLSCGNY-18-RLPIPVRLAD-15-/

Afu  /-12-QEEMSRS-13-YVVGVDQAF-34-FPYIPTFLMFRE-14-RAAIMVDGSGIAHPRRCGLATYIA-5-PTVGITKKRLFG-30-PIFISPGSY-17-KLPEPIRIAD-17-/
Tac  /-127-QEYIARK-10-TVRAFDVSY-27-FPYIPGYLAFRE-10-ADLLLIDGHGILHPRFAGLATHAG-5-ASIGVAKHLTRS-22---MIVSPGNM-16-RYPWPLKLAH-17-/
Fac  /-117-QEKIASM-10-HYRAIDVSY-27-FPYIPGYLGYRE-11-DALLLIDANGLLHPRKCGLATFAG-5-ATIGVAKSLLMG-23---IVSPGNM-16-KYPAILKMVH-13-/
Sso  /-13-QFLISKM-10-KICGVDIAY-29-FPYIPGFLFMRE-12-CHLLLVDGHGIAHPRKSGIAAVIG-5-PTIGVAKSRLTG-24--YFYSPGNK-11-GYPKVLKIAD-09-/
Pfu  /-10-QRRLSRK-12-RVGAVDVSY-30-FPYIPTFFFLRE-12-FDVLLVEGHGKAHPRRYGLASHIG-5-PTIGVAKRLLRG-12--AFVSVGHL-17-GYPKPLNIAD-12-/
```

I   II   III   IV   V   VI   VII

B

WT   +   +   +
ufi nmg ung   +   +   +
ufi Primers   +   +
nmg Primers   +   +
ung Primers   +   +

M

C

175 kD
83 kD
62 kD
48 kD
33 kD
25 kD
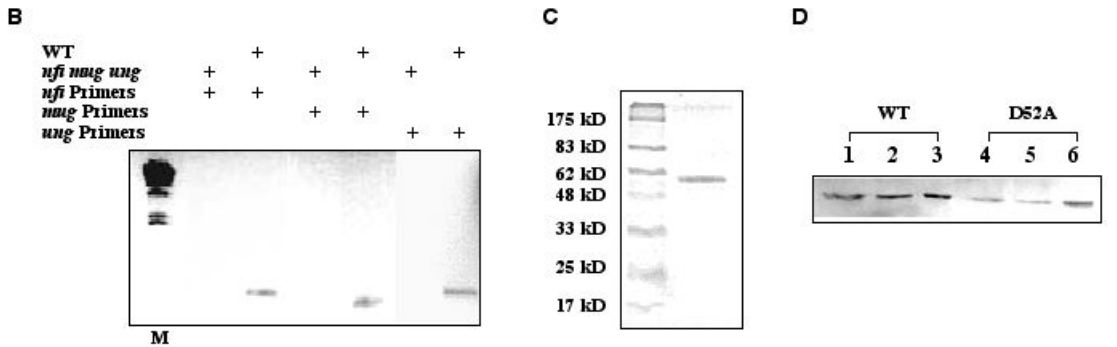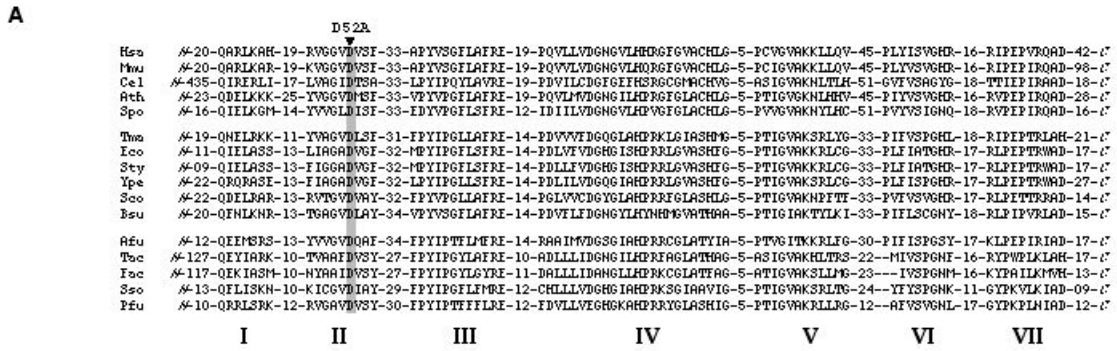17 kD

D

WT   D52A
1   2   3   4   5   6

Figure 4.1 Expression of human endonuclease V. A. Sequence alignment of endonuclease V. Amino acid residues selected for site-directed mutagenesis are highlighted and the resulting mutants are indicated above the arrows. Genbank accession numbers are shown after the species names. Hsa: *Homo sapiens*, BAC04765; Mmu: *Mus musculus*, XP_203558; Cel: *Caenorhabditis elegans*, 1731299; Ath: *Arabidopsis thaliana*, T10669; Spo: *Schizosaccharomyces pombe*, 1723511; Tma: *Thermotoga maritima*, NP_229661; Eco: *Escherichia coli*, NP_418426; Sty *Salmonella typhimurium*, NP_463037; Ype: *Yersinia pestis*, NP_667835; Sco: *Streptomyces coelicolor*, CAB40676; Bsu: *Bacillus subtilis*, BSUB0019; Afu: *Archaeoglobus fulgidus*, NP_068968; Tac: *Thermoplasma acidophilum*, CAC11602; Fac: *Ferroplasma acidarmanus*, ZP_00001774; Sso: *Sulfolobus solfataricus*, NP_343804; Pfu: *Pyrococcus furiosus*, NP_578716. B. Confirmation of the genotype of BL21(DE3)Δ3 by PCR. WT: *E. coli* K-12 strain. C. SDS-PAGE analysis of the wt h Endo V. Purified protein (~1 μg) was electrophoresed on a 15% polyacrylamide gel containing 0.1% SDS. Protein bands were visualized by Coomassie staining. D. Western blot analysis of the wt h endoV and mutant D52A proteins. Lanes 1 and 4: 10 μl cell extract. Lanes 2 and 5: Supernatant collected after sonication (equivalent to 30 μl cell culture). Lanes 3 and 6: purified protein (~0.5 μg).

4. Results and Discussion

Expression and purification of human endonuclease V

Despite biochemical characterization of homologs from a variety of species, nothing is known on human endonuclease V. This is due to the difficulty of obtaining soluble and active human endo V protein from recombinant sources. Previously, we attempted to detect endo V activity from mammalian tissues without success, probably due to low activity in fractionated protein extracts and interference from nonspecific nucleases (L. Dong & W. Cao, unpublished data). In order to obtain soluble and active human endo V, we tried different constructs and induction conditions to express the nfi gene. Co-expression of human nfi gene with heat shock genes only yielded insoluble protein (H. Gao & W. Cao, unpublished data). Refolding of endo V from inclusion bodies did not generate active protein (H. Gao & W. Cao, unpublished data). Expression of S. pombe and Arabidopsis thaliana nfi genes met with similar difficulty (H. Feng & W. Cao, unpublished data). We then cloned human nfi into pET32a vector, in which the human nfi was fused to the downstream of thioredoxin domain. The resulting plasmid pET32a-hnfi was then transformed into a special E. coli expression strain BL21(DE3)Δ3, in which the endogenous E. coli nfi, mug and ung were deleted. The deletion of these three genes prevents contamination of endo V, MUG, and UNG proteins from interfering with assays of deaminated base repair activities. The genotype of triple mutant strain was confirmed by PCR, which showed positive for the wt strain; however, negative for the mutant strain (Fig. 4.1B). Human nfi gene was overexpressed in the mutant strain and purified by metal chaleting and Q column chromatography (Fig. 4.1C).

To confirm the soluble nature of the expressed protein, Western blot analysis was performed on cell extract, and supernatant after sonication and purified protein (Fig. 4.1D). Both the wt and the active site mutant D52A protein turned out positive, indicating that the expressed human endo V was soluble.

Deaminated base repair activity.

Using the soluble human endo V protein purified from the mutant strain, we measured the repair activity towards all four deaminated bases in DNA, inosine (I), uridine (U), xanthosine (X) and oxanosine (O). Under the assay condition in which the enzyme was in excess ((E):(S) = 10:1), human endo V showed the strongest endonuclease activity towards inosine-containing DNA (Fig. 4.2B). The same activity was not detectable in the active site mutant D52A, indicating that the observed activity was intrinsic to human endo V (Fig. 4.2C). Human endo V was most active on single-stranded inosine-containing DNA (56%) followed by G/I (35%), T/I (24%), A/I (8%) and C/I (4%) (Fig. 4.2D). The enzyme also showed low level endonuclease activity on xanthosine-containing DNA (Fig. 4.2B), ranging from 4% for single-stranded X to 3% for G/X and T/X, 2% for A/X, and 1% for C/X (Fig. 4.2E). The activity on xanthosine was also intrinsic to human endo V since D52A did not show detectable activity (data not shown). On the other hand, human endo V did not show any endonuclease activity on either uridine- or oxanosine-containing DNA even under the assay condition in which the enzyme was in excess (Fig. 4.2B). Endo V enzymes from bacteria are also active on mismatch base pairs and contain exonuclease activities. Human endo V did not seem to
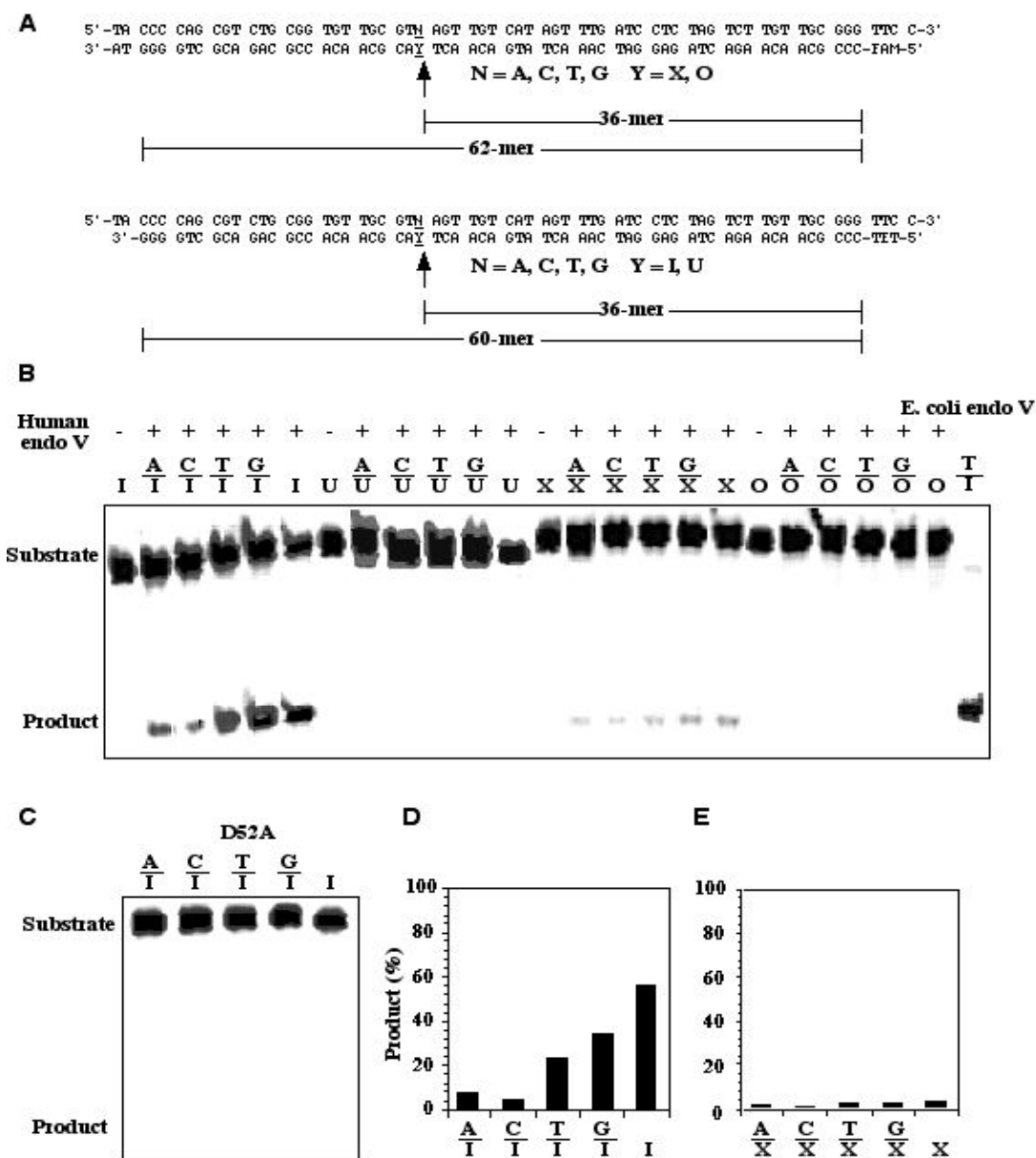
133

A

5'-TA CCC CAG CGT CTG CGG TGT TGC GT<u>N</u> AGT TGT CAT AGT TTG ATC CTC TAG TCT TGT TGC GGG TTC C-3'
3'-AT GGG GTC GCA GAC GCC ACA ACG CA<u>Y</u> TCA ACA GTA TCA AAC TAG GAG ATC AGA ACA ACG CCC-FAM-5'

N = A, C, T, G   Y = X, O

———— 36-mer ————
———————— 62-mer ————————

5'-TA CCC CAG CGT CTG CGG TGT TGC GT<u>N</u> AGT TGT CAT AGT TTG ATC CTC TAG TCT TGT TGC GGG TTC C-3'
3'-GGG GTC GCA GAC GCC ACA ACG CA<u>Y</u> TCA ACA GTA TCA AAC TAG GAG ATC AGA ACA ACG CCC-TET-5'

N = A, C, T, G   Y = I, U

———— 36-mer ————
———————— 60-mer ————————

B

Human endo V   - + + + + + - + + + + + - + + + + + - + + + + +   E. coli endo V

A/I C/I T/I G/I I A/U C/U T/U G/U U A/X C/X T/X G/X X A/O C/O T/O G/O O   T/I

Substrate

Product

C

D52A

A/I C/I T/I G/I I

Substrate

Product

D

Product (%) — 100, 80, 60, 40, 20, 0

A/I C/I T/I G/I I

E

100, 80, 60, 40, 20, 0

A/X C/X T/X G/X X

Figure 4.2 Cleavage of deamianted DNA by wt human endonuclease V and D52A mutant. A. Sequences of xanthine (X)- and oxanine (O)-, and hypoxanthine (I)- and uracil (U)-containing oligodeoxyribonucleotide substrates. FAM or TET: fluorophore. B. Cleavage activity of wt human endo V on I-, U-, X-, and O-containing substrates. Cleavage reactions were performed as described in Materials and Methods with 100 nM wt h endo V protein and 10 nM substrate. C. Cleavage activity of D52A mutant on I-containing substrates. Cleavage reactions were performed as described in Materials and Methods with 100 nM mutant D52A protein and 10 nM substrate. D. Quantification of cleavage by wt human endo V on I-containing substrates. E. Quantification of cleavage by wt human endo V on X-containing substrates.
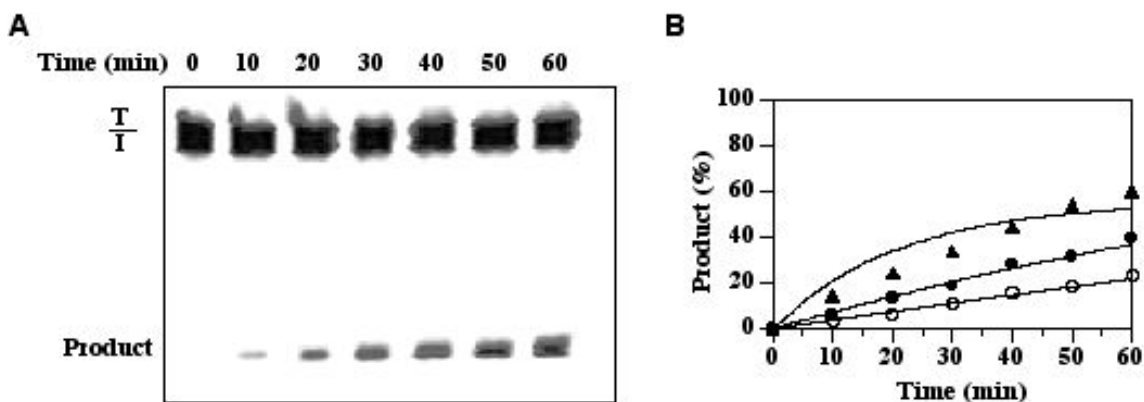
134

Figure 4.3 Kinetic analysis of cleavage activity of wt human endonuclease V on T/I substrates. A. GeneScan gel picture of cleavage activity of wt human endo V on T/I substrate. B. Time course analysis of cleavage activity. (S) = 10 nM. (E):(S) = 10:1 (▲) ss I; (●) G/I; (○) T/I.
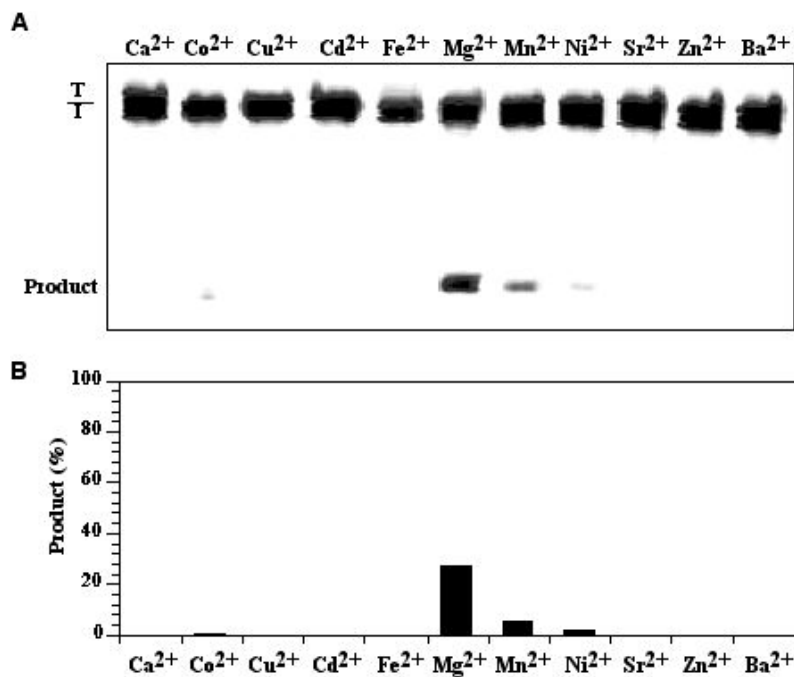

Figure 4.4 Metal effects on T/I cleavage by the wt human endonuclease V. A. Cleavage activity of wt h endo V with different metals. Cleavage reactions were performed as described in Materials and Methods with 100 nM protein and 10 nM substrate with different metals replacing Mg^{2+}. B. Quantification of T/I cleavage with different metal ions.
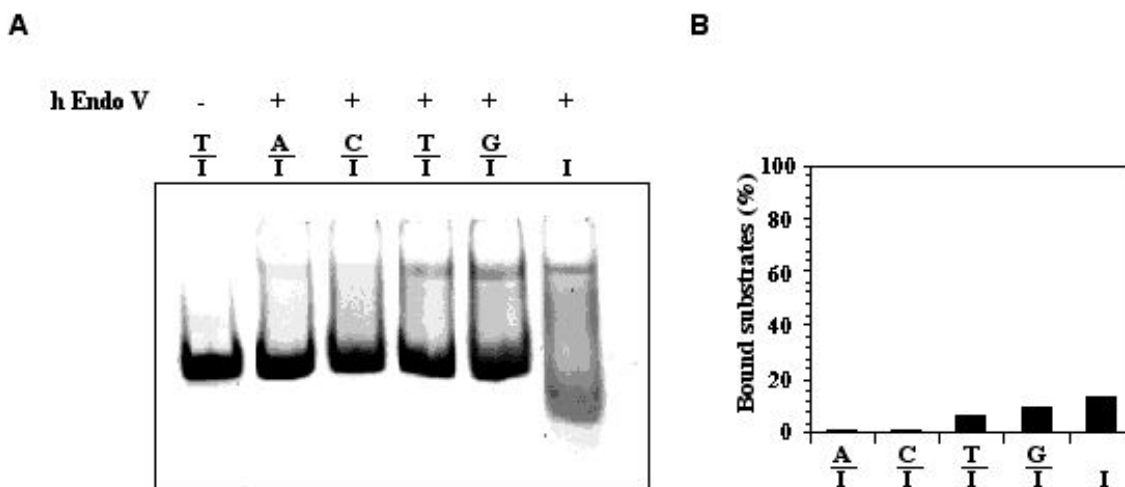
A

B

Figure 4.5 Binding analysis of wt h endo V on I-containing DNA substrates. A. Gel mobility shift analysis of binding of h endo V to I-containing substrate with 5 mM $MgCl_2$. Gel mobility shift assays were performed as described in Experimental Procedures with 500 nM endo V protein, 500 nM substrates and 5 mM $MgCl_2$. The first lane is the T/I substrate only used as a control. B. Quantification of wt h endo V's binding activity on I-containing substrates.
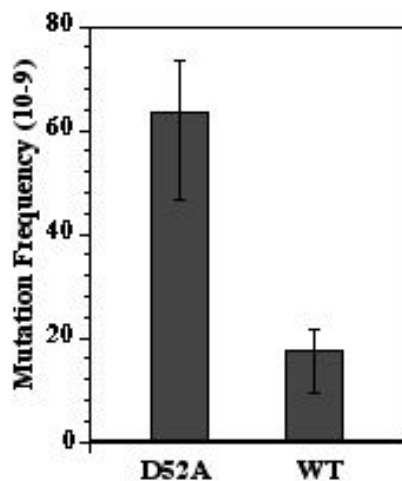

Figure 4.6 Antimutator effect of h endo V in *E.coli* BL21(DE3)Δ3 (*nfi, mug* and *ung*). Mutant *E.coli* BL21(DE3)Δ3 (*nfi, mug* and *ung*) cells transformed with pET-32a-hnfi or pET32a-D52A were plated both in the presence and absence of rifampicin. Results are from three independent experiments. Error bars are indicated.

136

contain any of these activities (data not shown). These data suggest that human endo V is primarily an inosine endonuclease with minor activity on xanthosine.

To better understand the kinetic differences of endonuclease activity on different inosine-containing DNA, a time course analysis was performed (Fig. 4.3). The apparent rate constant was the highest with single-stranded I substrate (0.025 min^{-1}), followed by G/I (0.017 min^{-1}) and T/I (0.014 min^{-1}). The rate constants for A/I and C/I were not measured; however, should be significantly lower than that of the other three substrates. Since endo V requires divalent metal ion for its endonuclease activity, we examined the cleavage of T/I substrate using eleven metal ions (Fig. 4.4A). Human endo V was most active with Mg^{2+} as the metal factor. The activity with Mn^{2+}, Ni^{2+} and Co^{2+} was detectable; however, significantly less as compared with Mg^{2+} (Fig. 4.4B). To determine the binding affinity of human endo V to the inosine-containing DNA, we performed gel mobility shift analysis. As shown in Figure 4.5, human endo V was able to generate a shifted band in all five substrates. The intensities of the shifted bands roughly followed the order of its endonuclease activity with C/I being the weakest. These results suggest that the binding affinity of human endo V to inosine-containing base pairs correlates with its catalytic activity.

Human endonuclease V as an antimutator

To assess whether of human endo V acts as an antimutator, we measured mutation frequency on rifampicin plates (*34*). The active site mutant D52A was used as a control.

As shown in Figure 6, the mutation frequency was reduced two-fold in the presence of the wt human endo V. This result suggests that human endo V is an antimutator *in vivo*.

Biochemical properties of human endonuclease V

Human endonuclease V is an elusive repair enzyme that has escaped biochemical characterization for years. The successful production of soluble human endo V makes it possible to understand well the biochemical properties of this important repair enzyme, which initiates so called Alternative Excision Repair pathway to remove deaminated base lesions (*35*). The primary target of the endonuclease activity from human endo V appears to be inosine lesions with minor activity on xanthosine lesions. On the other hand, bacterial endo V enzymes have broader endonuclease activities on deaminated bases, including inosine, xanthosine, oxanosine and uridine (*3-5,16-18*). However, genetic and biochemical studies indicate that bacterial endo V primarily is involved in repair of inosine and xanthosine lesions (*7,22,36,37*). It is interesting that few amino acid substitutions influence the endonuclease activity on inosine and xanthosine of Tma endo V in an extensive site-directed mutagenesis analysis, indicating that endo V has evolved to maintain its activity on inosine and xanthosine lesions.

The inosine endonuclease activity appears to be distinct in human endo V. Whereas bacterial endo V enzymes do not appear to be sensitive to the regular DNA base opposite to inosine (*5,7*), human endo V does. Besides the single-stranded inosine-containing DNA, the enzyme is most active on G/I followed by T/I > A/I > C/I (Figs. 4.2-4.3). Such a trend in activity in general is consistent with the stability of the inosine-

138

containing base pairs in which the stability follows the order of C/I > A/I > T/I, G/I (*38-40*). These results indicate that human endo V may significantly rely on the instability of the inosine-containing base pairs for recognition of the inosine base and subsequent endonucleolitic cleavage. The bacterial counterparts, on the other hand, may have evolved intrinsic exquisite ability to recognize any inosine in any base pairs.

We now know human genome contains two repair activities for inosine, with human endo V being a primarily inosine-specific endonuclease and Alkyl Adenine Glycosylase (AAG) being a hypoxanthine DNA glycosylase. These two enzymes share certain similarity in repair of deaminated adenine base, in which they both are quite active towards T/I base pairs as well as G/I and A/I base pairs. The activity on C/I appears to be low for both enzymes (Fig. 4.2 and (*33,41-44*)). The most significant contrast is that while human endo V is quite active on single-stranded inosine-containing DNA, human AAG has little activity on ss I (*33,45*). The physiological significance of this difference is unknown; perhaps it allows human endo V, in coupling with other repair machinery, to repair single-stranded inosine lesions during DNA replication and transcription more effectively.

Acknowledgements

5. References

1.   Demple, B., and Linn, S. (1982) On the recognition and cleavage mechanism of Escherichia coli endodeoxyribonuclease V, a possible DNA repair enzyme, *J Biol Chem 257*, 2848-2855.

2.   Gates, F. T., 3rd, and Linn, S. (1977) Endonuclease V of Escherichia coli, *J Biol Chem 252*, 1647-1653.

3.   Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli, *J Biol Chem 269*, 16260-16268.

4.   Yao, M., and Kow, Y. W. (1997) Further characterization of Escherichia coli endonuclease V, *J Biol Chem 272*, 30774-30779.

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

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

7.   Huang, J., Lu, J., Barany, F., and Cao, W. (2001) Multiple Cleavage Activities of Endonuclease V from Thermotoga maritima: Recognition and Strand Nicking Mechanism, *Biochemistry 40*, 8738-8748.

8.   Liu, J., He, B., Qing, H., and Kow, Y. W. (2000) A deoxyinosine specific endonuclease from hyperthermophile, Archaeoglobus fulgidus: a homolog of Escherichia coli endonuclease V, *Mutat Res 461*, 169-177.

9.   Moe, A., Ringvoll, J., Nordstrand, L. M., Eide, L., Bjoras, M., Seeberg, E., Rognes, T., and Klungland, A. (2003) Incision at hypoxanthine residues in DNA by a mammalian homologue of the Escherichia coli antimutator enzyme endonuclease V, *Nucleic Acids Res 31*, 3893-3900.

10.   Yao, M., and Kow, Y. W. (1994) Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3'-endonuclease from Escherichia coli, *J Biol Chem 269*, 31390-31396.

11.   Yao, M., and Kow, Y. W. (1996) Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from Escherichia coli, *J Biol Chem 271*, 30672-30676.

12. Shapiro, R. (1981) Damage to DNA caused by hydrolysis, in *Chromosome Damage and Repair* (Seeberg, E., and Kleppe, K., Eds.), pp 3-18, Plenum Press, New York.

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

14. Lucas, L. T., Gatehouse, D., and Shuker, D. E. (1999) Efficient nitroso group transfer from N-nitrosoindoles to nucleotides and 2'-deoxyguanosine at physiological pH. A new pathway for N- nitrosocompounds to exert genotoxicity, *J Biol Chem 274*, 18319-18326.

15. 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.

16. Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Deoxyinosine 3' endonuclease, a novel deoxyinosine-specific endonuclease from Escherichia coli, *Ann N Y Acad Sci 726*, 315-316.

17. 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.

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

19. Dong, M., Vongchampa, V., Gingipalli, L., Cloutier, J. F., Kow, Y. W., O'Connor, T., and Dedon, P. C. (2006) Development of enzymatic probes of oxidative and nitrosative DNA damage caused by reactive nitrogen species, *Mutat Res 594*, 120-134.

20. Guo, G., Ding, Y., and Weiss, B. (1997) nfi, the gene for endonuclease V in Escherichia coli K-12, *J Bacteriol 179*, 310-316.

21. Burgis, N. E., Brucker, J. J., and Cunningham, R. P. (2003) Repair system for noncanonical purines in Escherichia coli, *J Bacteriol 185*, 3101-3110.

22. Guo, G., and Weiss, B. (1998) Endonuclease V (nfi) mutant of Escherichia coli K-12., *J Bacteriol 180*, 46-51.

23. Weiss, B. (2001) Endonuclease V of Escherichia coli prevents mutations from nitrosative deamination during nitrate/nitrite respiration, *Mutat Res 461*, 301-309.

24.   Gao, H., Huang, J., Barany, F., and Cao, W. (2007) Switching base preferences of mismatch cleavage in endonuclease V: an improved method for scanning point mutations, *Nucleic Acids Res 35*, e2.

25.   Bazar, L., Collier, G., Vanek, P., Siles, B., Kow, Y., Doetsch, P., Cunningham, R., and Chirikjian, J. (1999) Mutation identification DNA analysis system (MIDAS) for detection of known mutations., *Electrophoresis 20*, p1141-1148.

26.   Favis, R., Huang, J., Gerry, N. P., Culliford, A., Paty, P., Soussi, T., and Barany, F. (2004) Harmonized microarray/mutation scanning analysis of TP53 mutations in undissected colorectal tumors, *Hum Mutat 24*, 63-75.

27.   Huang, J., Kirk, B., Favis, R., Soussi, T., Paty, P., Cao, W., and Barany, F. (2002) An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. *Oncogene 21*, 1909-21.

28.   Pincas, H., Pingle, M. R., Huang, J., Lao, K., Paty, P. B., Friedman, A. M., and Barany, F. (2004) High sensitivity EndoV mutation scanning through real-time ligase proofreading, *Nucleic Acids Res 32*, e148.

29.   Huang, J., Lu, J., Barany, F., and Cao, W. (2002) Mutational analysis of endonuclease V from Thermotoga maritima, *Biochemistry 41*, 8342-8350.

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

31.   Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning-A Laboratory Manual*, 3rd ed., Cold Spring Harbor laboratory Press, Cold Spring Harbor, New York.

32.   Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymease chain reaction, *Gene 77*, 51-59.

33.   Hitchcock, T. M., Dong, L., Connor, E. E., Meira, L. B., Samson, L. D., Wyatt, M. D., and Cao, W. (2004) Oxanine DNA glycosylase activity from Mammalian alkyladenine glycosylase, *J Biol Chem 279*, 38177-38183.

34.   Otterlei, M., Kavli, B., Standal, R., Skjelbred, C., Bharati, S., and Krokan, H. E. (2000) Repair of chromosomal abasic sites in vivo involves at least three different repair pathways, *EMBO J 19*, 5542-5551.

35.   Kow, Y. W. (2002) Repair of deaminated bases in DNA, *Free Radic Biol Med 33*, 886-893.

36.   Schouten, K. A., and Weiss, B. (1999) Endonuclease V protects Escherichia coli against specific mutations caused by nitrous acid, *Mutat Res 435*, 245-254.

37.   Yao, M., and Kow, Y. W. (1995) Interaction of deoxyinosine 3'-endonuclease from Escherichia coli with DNA containing deoxyinosine, *J Biol Chem 270*, 28609-28616.

38.   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 Res 22*, 131-136.

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

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

41.   Abner, C. W., Lau, A. Y., Ellenberger, T., and Bloom, L. B. (2001) Base excision and DNA binding activities of human alkyladenine DNA glycosylase are sensitive to the base paired with a lesion, *J Biol Chem 276*, 13379-13387.

42.   Dianov, G., and Lindahl, T. (1991) Preferential recognition of I.T base-pairs in the initiation of excision-repair by hypoxanthine-DNA glycosylase, *Nucleic Acids Res 19*, 3829-3833.

43.   Saparbaev, M., Mani, J. C., and Laval, J. (2000) Interactions of the human, rat, Saccharomyces cerevisiae and Escherichia coli 3-methyladenine-DNA glycosylases with DNA containing dIMP residues, *Nucleic Acids Res 28*, 1332-1339.

44.   Wyatt, M. D., and Samson, L. D. (2000) Influence of DNA structure on hypoxanthine and 1,N(6)-ethenoadenine removal by murine 3-methyladenine DNA glycosylase, *Carcinogenesis 21*, 901-908.

45.   Saparbaev, M., and Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases, *Proc Natl Acad Sci U S A 91*, 5873-5877.

CHAPTER FIVE

RESEARCH SIGNIFICANCE AND CONCLUDING REMARKS

Repair enzymes are essential factors for maintaining DNA stability against attack from endogenous and exogenous sources. A defect in these repair enzymes or their pathways poses a significant problem for living cells because it can cause mutation, disease or even death.

SMUG1, believed to include hybrid active sites from UDG Family 1 and Family 2, evolutes at a relatively late stage (*1*). This enzyme, which belongs to Family 3 of the UDG superfamily, was previously thought to be an eukaryotic exclusive enzyme, primarily recognizing uracil and its derivatives (*2-6*). Based on genome sequence analysis, the research reported here identified four SMUG1 orthologs in bacteria strains, meaning that these strains place SMUG1's initial appearance back to a much earlier time period. This research is the first report studying SMUG1's activity from bacterial species. To characterize the bacteria SMUG1, the recombinant protein, Gme SMUG1, was cloned, expressed and purified from the *E. coli* system. This protein exhibits robust uracil DNA glycosylase activity and xanthine DNA glycosylase activity on both double- and single-stranded substrates. Ten Gme SMUG1 mutants were constructed to study the key active site residues involved in the interactions between the DNA substrates and this enzyme, and in the ability to recognize specific damaged bases. Molecular modeling and molecular dynamic simulation methods were employed to analyze the experimental results. These results indicate that although several key residues are typically involved in

the enzyme-DNA substrate complex, a single amino acid change in one of the mutants switches the UDG Family 3 SMUG1 to Family 1 UDG or UNG. Thus, this mutant eliminates the XDG activity while maintaining robust UDG activity and demonstrating an equal preference for all uracil containing substrates unlike the Family 3 UDG (*7*).

Bacterial endonuclease V is a small protein which recognizes many types of substrates (*8-11*). Thus far, only two enzymes, bacterial endo V and Spo TDG (*12*), were found to possess the ability to recognize hypoxanthine, xanthine, uracil and oxanine with endo V being the primary enzyme repairing inosine in bacteria *in vivo* (*13*); however, the precise repair pathway remains unknown. According to an *E. coli* endo V repair patch study, 3 nucleotides from the 3' end and 2 nucleotides from the 5' end of the endonuclease cleavage are removed *in vivo* (*13*). In addition, our group investigated Tma endo V, a thermal stable enzyme, which not only exhibits endonuclease activity, but also exonuclease activity under specific assay conditions. This exonuclease activity may potentially play a vital role in investigating the endo V-initiated repair pathway. The research reported here expanded the understanding of endonuclease activity and exonuclease activity in Tma endo V. In addition to inosine specific and nonspecific endonuclease activity, Tma endo V was found to have 3' inosine dependent exonuclease activity and 5' nonspecific exonuclease activity. These results may help in determining the endo V-initiated repair pathway.

The genetic and biochemical studies of bacterial endo V are extensive, while the studies of eukaryotic endo V are limited because of the difficulty of obtaining this soluble recombinant protein. For this research; however, soluble and active human endo V was

obtained, facilitating the characterization of this enzyme *in vitro*. Human endo V is also an authentic endonuclease having different properties from bacterial endo V. It recognizes the inosine-containing DNA substrates having the highest activity on single-stranded DNA and exihibts limited cleavage activity on xanthosine containing substrates under the assay conditions reported here. These findings are consistent with the results indicating that *E. coli* endo V *in vivo* is primaly responsible for inosine repair (*13*). These results can help future research of the endo V-initiated repair pathway in mammalian systems.

References

1. 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. *Nucl. Acids Res. 32*, 5291-5302.

2. Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., and Slupphaug, G. (2002) hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J Biol Chem 277*, 39926-39936.

3. Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A., and Ide, H. (2003) Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. *Biochemistry 42*, 5003-5012.

4. Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., and Lindahl, T. (2001) Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *Embo J 20*, 4278-4286.

5. Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H. E., Lindahl, T., and Barnes, D. E. (2000) Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol Cell 5*, 1059-1065.

6. Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E., and Kavli, B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res 35*, 3879-3892.

7. Mi, R., Dong, L., Kaulgud, T., Hackett, K. W., Dominy, B. N., and Cao, W. Insights from xanthine and uracil DNA glycosylase activities of bacterial and human SMUG1: Switching SMUG1 to UNG. *Journal of Molecular Biology In Press*.

8. Yao, M., Hatahet, Z., Melamede, R. J., and Kow, Y. W. (1994) Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli. *J Biol Chem 269*, 16260-16268.

9. Yao, M., and Kow, Y. W. (1994) Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3'-endonuclease from Escherichia coli. *J Biol Chem 269*, 31390-31396.

10. Yao, M., and Kow, Y. W. (1996) Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from Escherichia coli. *J Biol Chem 271*, 30672-30676.

11. Yao, M., and Kow, Y. W. (1997) Further characterization of Escherichia coli endonuclease V. Mechanism of recognition for deoxyinosine, deoxyuridine, and base mismatches in DNA. *J Biol Chem 272*, 30774-30779.

12. Dong, L., Mi, R., Glass, R. A., Barry, J. N., and Cao, W. Repair of deaminated base damage by Schizosaccharomyces pombe thymine DNA glycosylase. *DNA Repair In Press*.

13. Weiss, B. (2008) Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation. *DNA Repair (Amst) 7*, 205-212.