The contribution of different uracil-DNA glycosylases to removal of uracil from DNA in different mouse organs - the significance of sequence context and proliferative status

Acknowledgements

This thesis was carried out at the Department of Cancer Research and Molecular Medicine (IKM) at the Norwegian University of Science and Technology (NTNU) from 2013 to 2014, leading to the degree “Master of Science in Molecular Medicine”.

I would like to express my heartfelt gratitude to my main supervisor Prof. Hans E. Krokan for providing me a chance to work in his renowned research group, for sharing his great knowledge, supporting me and spending his quality time to answer all my queries. I am also grateful to my co- supervisor Mr. Antonio Sarno. His scientific supervision, knowledge, enthusiasm, support and thoroughly prepared feedback along the way have been invaluable for the progress of this thesis. My vote of thanks to Dr. Berit Doseth for teaching me some important research experiments and her constant support and feedback for the results has been a much help to complete my work.

I would like to thank all my friends and colleagues at the DNA repair group for useful discussions, helpfulness, advices and inspiration.

Lastly, I would like to thank my husband Tuhin, my parents and in-laws, and my grandmother for their constant love and support that helped me to successfully complete this work.

Trondheim, July 2014

Romi Roy Choudhury-Bhakta

Abstract

Uracil is a non-canonical base in DNA that can arise through misincorporation of dUMP instead of dTMP during replication or from cytosine deamination. Uracil in DNA can be removed by the four different DNA glycosylases UNG, SMUG1, TDG or MBD4 as the first step in base excision repair. These glycosylases have different catalytic efficiencies, different substrate preferences and different expression patterns. We wanted to elucidate the contribution of the different DNA glycosylases in removal of uracil, using protein extracts from mouse brain, small intestine, kidney, liver, lung and muscle. Importantly, $UNG^{+/+}$ and $UNG^{-/-}$ mice were available for these studies. Furthermore, we used the UNG-specific inhibitor Ugi to inhibit UNG-activity and an anti-SMUG1 antibody PSM1 to inhibit SMUG1 activity. In addition, we used a combination of Ugi and PSM1 to inhibit both UNG and SMUG1 activities. To investigate the contribution of the DNA glycosylases in different sequence contexts, we used uracil-containing DNA substrates with uracil in a U:G mismatch, a U:A pair, or in single stranded DNA. We found that UNG was by far the most important DNA glycosylase removing uracil from U:A pairs and from single-stranded DNA, even in extracts from non-proliferative tissues such as brain and muscle. However, we found that UNG and SMUG1 are almost equally active in removal of uracil from a U:G context in extracts from non-proliferative tissues. In contrast, SMUG1 contributed very little to the removal of uracil from small intestines, which presumably contain a higher fraction of rapidly dividing cells. Furthermore, we detected essentially no residual uracil-DNA glycosylase activity when both UNG and SMUG1 were specifically inhibited, indicating that TDG and MBD4 do not contribute measurable activity in any of the organs investigated under the conditions used here. Possibly these DNA glycosylases can only contribute in sequence patterns that have not been investigated here. We also carried out a preliminary study on the possible correlation between genomic uracil contents (measured by mass spectrometry) and uracil excision activities, but failed to observe a significant positive correlation or inverse correlation. In conclusion, UNG is the dominant uracil-DNA glycosylase in proliferating tissues in removal of uracil from a U:G context, whereas in non-proliferating tissues the contribution of UNG and SMUG1 is essentially similar. Furthermore, UNG is the only glycosylase removing uracil from U:A and single stranded DNA contexts.

Contents

List of figures

List of tables

Abbreviations

5-FU	5-fluorouracil
5-hmU	5-hydroxymethyluracil
5-hU	5-hydroxyuracil
6-4 PP	6-4 photoproduct
6-FAM	6-fluoroscein amidite
Ade	Adenine
AID	Activation induced deaminase
AP	Apurinic/ apyrimidinic or abasic
APE	AP endo-/exonuclease
APOBEC	Apolipoprotein B mRNA-editing catalytic polypeptide
BER	Base excision repair
Cyt	Cytosine
CPD	Cyclobutane pyrimidine dimer
CSR	Class switch recombination
C-terminal	Carboxyl terminal
dCyd	Deoxycytidine
dTMP	2'-deoxythymidine-5'-monophosphate
dTTP	2'-deoxythymidine-5'-triphosphate
dUMP	2'-deoxyuridine-5'-monophosphate
dUDP	2'-deoxyuridine-5'-diphosphate
dUTP	2'-deoxyuridine-5'-triphosphate
DHFR	Dihydrofolate reductase
DNA	Deoxyribonucleic acid
dRP	Deoxyribose phosphate
ds	double stranded
DSB	Double stranded break
dUrd	Deoxyuridine
E. coli	*Escherichia coli*
EXO	Exonuclease
FEN1	Flap endonuclease
Gua	Guanine
HIGM	Hyper- immunoglobulin M
HPLC	High performance liquid chromatography
hSMUG1	Human SMUG1
Ig	Immunoglobulin
IR	Ionizing radiation
KO	Knock out
LC-MS/MS	Liquid chromatography coupled tandem mass spectrometry
LIG	DNA ligase
LP	Long patch
MBD4	Methyl-binding domain protein 4
MLH	Human homolog of *E. coli* MutL
MMR	Mismatch repair protein
MSH	Human homolog of *E. coli* MutS
MTHF	N5, N10-methylene-tetrahydrofolate
NDPK	Nucleoside diphosphate phosphatase/kinase
NEIL	Nei endonuclease VIII-like (*E. coli*)

NEMO	Nuclear factor kappa B essential modulator
NER	Nucleotide incision repair
NHEJ	Non homologous end joining
N-terminal	Amino terminal
PCNA	Proliferating cell nuclear antigen
POL	Polymerase
PSM1	Anti-SMUG1 antibody
RPA	Replication protein A
SAM	S-adenosylmethionine
SHM	Somatic hypermutation
SMUG1	Single-strand-selective monofunctional Uracil-DNA glycosylase 1
SP	Short-patch
ss	Single strand
SSB	Single strand break
SUMO	Small ubiquitin-like modifier
Thd	Thymine
TDG	G/T mismatch specific Thymine DNA glycosylase
TLS	Translesion synthesis
TS	Thymidylate synthase
Ura	Uracil
U:A	Uracil base paired with adenine
U:G	Uracil base paired with guanine
UDG	Uracil DNA glycosylase
Ugi	Uracil DNA glycosylase inhibitor protein
UNG1/2	Uracil-N glycosylase (mitochondrial/nuclear forms)
UV	Ultraviolet
WT	Wild type
XRCC1	X-ray repair cross-complementing protein 1
εC	$3,N^4$-ethenocytosine

1. Introduction

Eukaryotic cells undergo strict DNA replication and repair processes to ensure proper transmission of genetic material from a mother to a daughter cell. DNA lesions may arise endogenously from e.g. oxidation, hydrolysis, and deamination, and exogenous threats emerge from sources like ionizing radiation and various chemical agents. If not repaired, these lesions may lead to mutagenesis, apoptosis, cancer, and neurological disorders (1). DNA repair pathways have evolved in all organisms that repair most DNA lesions.

Nucleobases are sub-categorized into purines and pyrimidines (2). The purine bases are adenine (Ade) and guanine (Gua) and the pyrimidines are thymine (Thy) and cytosine (Cyt). In the Watson-crick model, pyrimidines base-pair with purines by hydrogen-bonding, with Ade opposite to Thy (A:T) and Gua opposite to Cyt (G:C). Uracil (Ura) is a canonical RNA base which, instead of Thy, base pairs with Ade under normal conditions. Ura may also be found in very small amounts in DNA as a result of 2'-deoxyuridine-5'-triphosphate (dUTP) misincorporation during replication and Cyt deamination. Misincorporation of 2'-deoxyuridine-5'-triphosphate (dUTP) instead of 2'-deoxythymidine-5'-triphosphate (dTTP) leads to U:A pairing while Cyt deamination results in a more mutagenic U:G pair (3). Genomic Ura, regardless of the source, is recognized by uracil-DNA-glycosylases (UDGs), which recognize and excise the lesion to begin repair by the base excision repair (BER) pathway (4); however, studies reveal that genomic uracilation, in selected regions, is also crucial for antibody diversification.

1.1. Sources of uracil in DNA

1.1.1. dUTP misincorporation during DNA synthesis

In a mammalian cell, dUTP is a natural intermediate in the pyrimidine metabolism pathway (5). dUTP emerges from nucleoside diphosphate kinase (NDPK)-dependent phosphorylation of 2'-deoxyuridine-5'-diphosphate (dUDP) following ribonucleoside diphosphate reductase (rNDPase) activity and by random phosphorylation of 2'-deoxyuridine-5'-monophosphate (dUMP), a crucial precursor of 2'-deoxythymidine-5'monophosphate (dTMP). In turn, dUMP stems either from thymidine kinase dependent phosphorylation of deoxyuridine (dU) or by 2'-deoxycytidine-5'monophosphate (dCMP) deamination (3).

However, dUTP levels are regulated by 2'-deoxyuridine-5'-triphosphate nucleotide hydrolase (dUTPase) by hydrolyzing dUTP to dUMP (6).

Due to the structural similarity of dUTP and 2'-deoxythymidine-5'triphosphate (dTTP), most replicative polymerases fail to differentiate between the two and therefore aberrantly incorporate dUTP with a similar efficiency to dTTP during DNA synthesis (1). The frequency of misincorporation in a growing chain in human DNA has estimated to be about one dUTP residue per 10^4 dTTPs per cell per day (7). The degree of dUMP mis-incorporation into the growing DNA chain depends on the relative quantities of dTTP and dUTP, the respective concentrations of which vary between 30-37 µM to 0.2 µM (8). Therefore, the intracellular concentration of dUTP is as low as 1% of that of dTTP (9). Studies support that most of the genomic uracil is incorporated during replication: about 3600 uracil per genome in proliferating cells (embryonic fibroblasts) and about 900-1400 Ura residues per genome in non-dividing cells (hepatocytes) (10).

dUTP misincorporation leads to a U:A base-pair which is not inherently mutagenic because U:A replicates to T:A (1). Genomic uracil-induced cytotoxicity might occur when misincorporation occurs in promoter regions altering the binding of transcription factors and in turn compromises transcription (11). It may also occur by abasic site formation resulting from glycosylase activity leading to topoisomerase halt and frequent single stranded breaks all over the genome (12).

The dUTP misincorporation rate is increased with decreased dUTPase activity, cytostatic drug treatment, folate deficiency, and reduced thymidylate synthase (TS) activity (13). Folate is an important cofactor for dTMP generation with co-substrate substrate N5, N10-methylene-tetrahydrofolate (MTHF) and aids the conversion of dUMP to dTMP and dihydrofolate (DHF) by the catalyzing action of TS. During thymidylate biosynthesis, TS uses MTHF as a methyl donor to dUMP and dihydrofolate reductase (DHFR) enhances MTHF regeneration in every step of continued dTMP synthesis (3). TS activity may be inhibited by the folate antagonist methotrexate, and/or folic acid deficiency and leads to an increase in the dUMP level in cells (14). Thus, this pathway has been the target of several anticancer drugs like 5-fluorouracil (5-FU), 5-fluorodeoxyuridine (5-FUrd), and pemetrexed.

1.1.2. Cytosine deamination

Cytosine deamination is a major source of uracil in DNA and may arise in several ways as discussed below.

1.1.2.1. Spontaneous deamination

Under normal physiological conditions, cytosine deamination occurs in two different pathways (15). As shown in Figure 1-1, the direct hydrolytic deamination pathway follows an intramolecular process of nucleophilic attack of a hydroxyl ion (OH^-) on an unstable exocyclic amino group at the C_4 position of the base. This reaction releases ammonia to give rise to an intermediate product- uridine hydrate (6-hydroxy-5,6-dihydro-uridine) that subsequently dehydrates to uridine (I->III->IV). The addition-elimination pathway is comprised of a hydroxylic attack at the C5,C6 double bond of the protonated base to form cytidine hydrate that eventually gives rise to uracil in the double helix (II->V->IV).

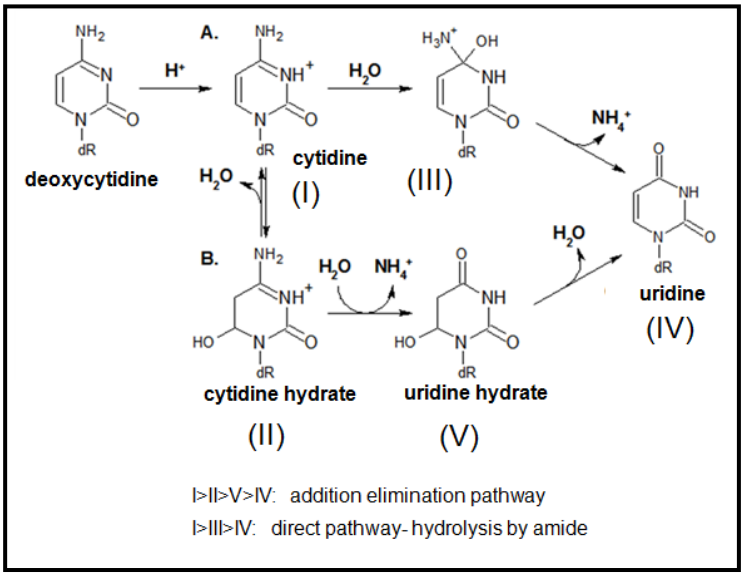

Figure 1-1: Spontaneous chemical deamination of dC. (A.) Direct hydrolytic deamination via alkali-catalyzed hydrolysis or (B.) water attack on the protonated base (modified from Shapiro (15))

These reactions depend on various parameters like temperature, pH, and the state of DNA: single or double stranded (16). Acidic pH (within the range of 1-3) accelerates the deamination rate by converting the substrate to a more protonated form, whereas, alkalic pH converts the substrate into a less anionic form by substituting an effective nucleophilic agent- OH^-. In an acidic condition of pH 3 at 25^oC, one out of every 10,000 protonated

deoxycytidine residues exists as hydrates. In order to extrapolate the cytosine deamination rate, Lindahl *et al.* conducted an experiment by increasing the temperature range. At 100^{o}C and at neutral pH, they measured one uracil per 5000 cytosine residues (16). Under normal biological conditions, double-stranded DNA is less prone (<1%) to undergo deamination than single-stranded DNA. Based on the results of Frederico *et al*., an estimated 70-200 uracil residues could be formed per cell per day provided 0.1% of the genome becomes single stranded at any time point (17). These mutagenic deaminations lead to U:G mispairs and result in transition mutations from G:C to A:T, if tolerated until the next round of replication (18).

1.1.2.2. Non-enzymatic deamination

Several chemicals have been shown to increase the cytosine deamination rate. Nitrous anhydride (N_2O_3) induces deamination upon oxidation of nitric oxide ($NO^{\cdot}$) that originates from arginine (Arg) oxidation (19). This intermediate works on both ss- and ds-DNA with similar efficiency and is widely produced *in vivo* in response to various types of chronic inflammations (20). Sodium bisulphite also causes to cytosine deamination and has been found in many food items (3).

Apart from chemicals, cytosine deamination may arise from different types of radiation. The major source is UV irradiation, which leads to the formation of cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4-PP). CPDs are the one of the important mutagenic lesions in mammalian cells (21) and they are generated at a much higher rate, about 5-10 fold higher frequency, than the 6-4 photoproducts (22,23). Cytosine involved in a CPD undergoes higher rate of deamination (24-26). Uracil lesions produced in this way are difficult to alter as most mammals lack photolyase to reverse the CPD crosslink and rely only on the nucleotide excision pathway (NER) for repair (1).

1.1.2.3. Enzymatic deamination of cytosine

There are two main types of enzymes responsible for deamination: most importantly deoxycytidine deaminases (27) and the less potent cytosine-5-methyltransferase (28). The latter is mainly responsible for DNA modifications during epigenetic gene silencing by converting CpG cytosine to 5-methylcytosine (5-mC). Methyltransferases transfer a methyl

group to cytosine utilizing S-adenosylmethionine (SAM) as a methyl donor. Methyltransferases have been shown to deaminate Cyt when SAM concentration is low (29).

In higher vertebrates, the apolipoprotein B mRNA editing polypeptide (APOBEC) family of proteins has been shown to deaminate cytosine, among which AID plays an important role in adaptive immunity (30). APOBEC2, APOBEC4, and AID are present in both humans and mice (reviewed in (3)).

AID deaminates cytosine in targeted immunoglobin gene loci to activate antibody diversification through class switch recombination (CSR) and somatic hypermutation (SHM) (31). This enzyme is usually active in B-cell germinal centers (BGCs) (32). Apart from adaptive immunity, AID may has a role in DNA demethylation during reprogramming of pluripotent stem cells (33) and in primordial germ cells (PGCs) (34). Cytosine deamination by AID occurs in 2-7kbp long hot-spots of DNA using the following motif: WRCY (W=A/T, R=A/G, C/Y=T/A). It has been found that AID deaminates cytosine in about 25% of expressed genes in mouse BGC thus correlating its importance as a source or uracil in DNA (32).

1.2. Consequences of Uracil in DNA

Eukaryotic organisms respond to DNA damage in a variety of ways. They eradicate or tolerate the damage or respond by inducing apoptosis.

1.2.1. DNA mutagenicity

U:A base-pairing in DNA originates from dUTP misincorporation, which is not directly mutagenic but can give rise to mutagenic abasic sites (10). Cytosine deamination is more efficient on deoxynucleosides (dNs), less in single stranded DNA (ss-DNA), and the least efficient in double stranded DNA (35). Uracil occurring from in this process results in U:G mispair that produces C>T transition mutations after one cycle of replication resulting in a change in the original sequence (Figure 1-2). Uracil N'-glycosylase (UNG) has a high range of selectivity for ss-DNA substrate over double stranded, and U:G over U:A (36). Likewise, AP site formation in DNA is cytotoxic because it irreversibly traps topoisomerase I, an enzyme responsible for unwinding and over winding of DNA during replication which halts the replication and transcription processes (2,37).

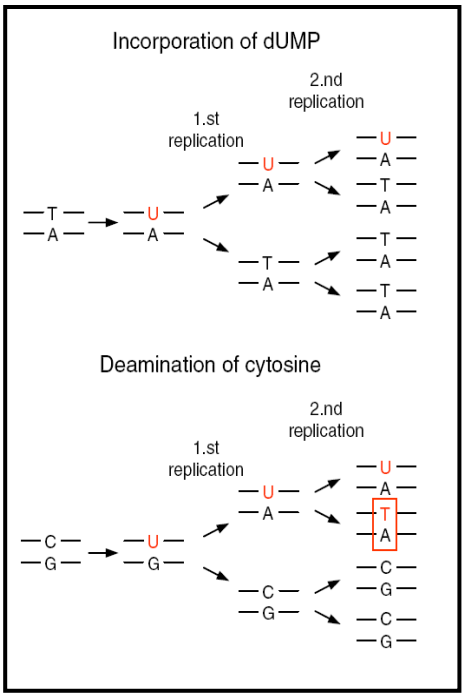

Figure 1-2: Consequences of misincorporation of dUMP and cytosine deamination that are not repaired. U:A base pair in DNA results from incorporation of dUTP instead of dUMP by polymerases. U:G mispair occur after deamination of Cyt to Ura, resulting in C:G to T:A mutation transition after second round of replication (modified from Sousa 2007)

An estimated 200 Cyt residues are spontaneously deaminated per mammalian diploid genome per day (18), but recent studies show the global genomic uracil content in the genome of UNG-active cell to be about 400-600 residues with a 10 fold increase in UNG-deficient cells (38).

1.2.2. Repair of genomic uracil

Genomic uracil is repaired by the base excision repair (BER) pathway (39). BER is typically responsible for repair of a large number of small base lesions that do not distort the helical structure of DNA (2,40,41). These types of base lesions are the result of deamination, misincorporation, base methylation, oxidative base damage, and DNA replication faults that occur spontaneously or by environmentally induced factors. BER occurs both in the nucleus and mitochondria that frequently use the isoform of glycosylases originating from splicing of the same gene. It is one of the important pathways in relation to cancer, neurodegeneration, and aging (42).

The basic steps of BER are damage recognition, base removal, strand incision, end trimming/ processing, repair synthesis by nucleotide insertion, and ligation (43). These

"common steps" may differ *in vivo* depending on the recruitment of different glycosylases as well as the physiological state of the cell. The common mechanism followed by all DNA glycosylases is to flip the faulty base out of the helix into the substrate recognition site of the enzyme so as to create an apurinic (AP)- site by cleaving N'-glycosyl bond between C1-carbon of 2'-deoxyribose and the inappropriate/ damaged base. The second step is carried forward by AP endonuclease 1 (APE1), which recognizes the abasic site, and cut the phosphodiester bond on the 5'end of AP-site, forming a free 3'-OH and 5'-deoxyribose-phosphate (5'-dRP) moiety. This 3'-OH serves as the substrate for new strand synthesis in case of monofunctional glycosylases. Bifunctional glycosylases carry an additional AP-lyase activity site that cleaves the phosphodiester bond on the 3' end of the AP site producing a 3'-dRP (unsaturated hydroxyaldehyde) and a 5'-P end. In this case 5'-end processing is required as it is considered a polymerase blocking lesion and also DNA ligases need free 5'-phosphate and 3'-OH ends on either sides to seal the nick. The end-cleaning action is completed by APE1 or polynucleotide kinase 3'-phosphatase (PNPK) prior to base insertion. The downstream step recruits DNA polymerase β (POL-β) which adds one nucleotide and removes the dRP from the DNA chain. Polymerization takes place following recruitment different enzymes or complexes of enzymes: short patch BER (SP-BER) that replaces a single nucleotide and long patch BER (LP-BER) replaces 3-12 nucleotides. As shown in Figure 1-3, in SP-BER, which is also known as "single-nucleotide" BER, gap-filling is carried out by Pol-β which inserts a single base and removes the dRP alongside. The nick is sealed by DNA ligase I (LIG1) in complex with scaffold protein X-ray repair cross-complementing protein 1 (XRCC1). LP-BER replaces several nucleotides by the aid of replication enzymes. Pol ∂/ε or Pol-β synthesizes the new strand in complex with proliferating cell nuclear antigen (PCNA) and replication factor C (RCF). The strand produced in this way is digested by flap endonuclease (FEN1); dRP is removed as a part of the flap in LP-BER. To complete the repair process, the nick is ligated by LIG1 (44,45) (reviewed in (41)).

The factors that decide whether SP-BER or LP-BER takes place are still unelucidated. One explanation for the switch to LP-BER is the presence of 5'-dRP which restricts the activity of Pol-β lyase (46). Nevertheless, bifunctional glycosylases tend to undergo SP-BER while monofunctional glycosylases exhibit no preference (47). However, SP-BER is a dominant pathway in both proliferating and non-proliferating cells, while LP-BER has mainly been found in proliferating cells. Other factors include proteins involved in postreplicative

BER, such as UNG2 and NEIL1 (48) which are found in highest level in S-phase of a proliferating cell (49).

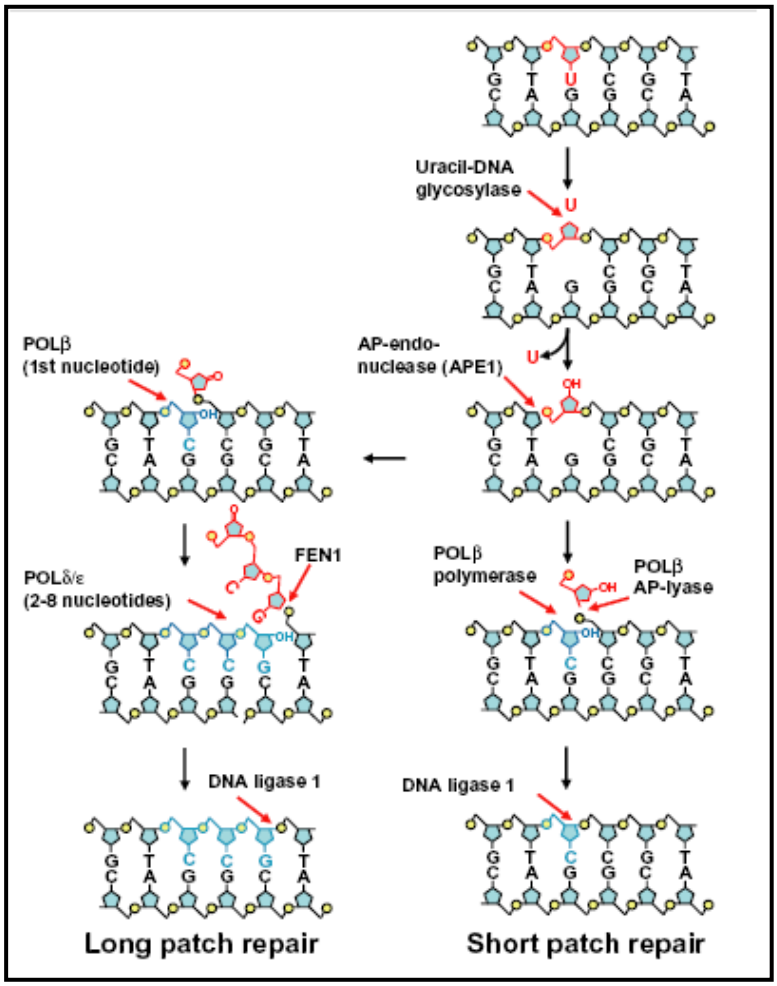

Figure 1-3: Schematic overview of SP- and LP-BER pathways. UNG recognizes and excises Ura and APE1 cleaves the DNA at the resulting AP site. In short patch BER, Pol-β replaces excised nucleotide and LIG1 ligates the strand. In long patch BER, Pol ∂/€ insert the rest of the nucleotides, the flap is removed by FEN1, and DNA is ligated by LIG1 (modified from Sousa, 2007).

1.2.2.1. Different types of Uracil DNA glycosylases

The human genome encodes for four different uracil DNA glycosylases (UDGs) namely mitochondrial and nuclear uracil N'-glycosylase (UNG1 and UNG2 respectively), single strand selective monofunctional uracil DNA-glycosylase (SMUG1), thymine DNA glycosylase (TDG), and methyl-binding domain protein 4 (MBD4) (3,4,50). These glycosylases vary from each other with respect to preferred sequence context, type of DNA

(single or double stranded), cell cycle expression, and whether Ura is an U:A or U:G. These enzymes are discussed in detail below Table 1-1 (41).

Table 1-1: Mammalian DNA glycosylases (adapted from (41))

Enzyme	Subcellular localization	Functionality	Substrates	Knockout mice outcome	Human disease
UNG2	Nuclei	Monofunctional (M)	U, 5-FU in ss and ds-DNA; U:A & U:G context (alloxan, 5-hU, isodialuric acid)	Partial defect in CSR, skewed SHM, B-cell lymphomas	Complete defect in CSR, HIGM syndrome, infections, lymphoid hyperplasia
UNG1	Mitochondria	M	Same as UNG2	Same as UNG2	Unknown
SMUG1	Nucleus	M	5-hmU, U:G>U:A>ss-U, 5-FU, εC in ss and ds-DNA	Viable and fertile, SMUG1/UNG/MSH triple KO reduced longevity (51)	Unknown
TDG	Nucleus	M	U:G>T:G (5-hmU dsDNA, 5-FU)	Embryonic lethal, epigenetic role in development	Unknown
MBD4	Nucleus	M	U:G and T:G, 5-hmU in CpG context (epsilon-C, 5-FU in ds-DNA)	Viable and fertile, C to T transitions, intestinal neoplasia	Mutated in carcinomas with microsatellite instability

1.2.2.1.1. Uracil DNA N'-glycosylases

UNG1 and UNG2 are generated by alternative promoter usage (P_A and P_B) and subsequent splicing of the transcript of the same gene for mitochondria and nucleus, respectively (52,53). The active site consists of highly conserved regions that catalyze glycosidic bond cleavage. Both isoforms share an identical catalytic domain but have different N'-terminal domains, which result in different cellular localizations. UNG2 operates in post

replicative base excision repair of misincorporated uracil. A study conducted by Kavli *et al.* revealed an exceptional high rate of catalytic turnover of human UNG2 *in vitro* (54).

UNG2 shows a high preference for ss-DNA over ds-DNA and U:G over U:A. Though at a very low efficiency, UNG2 can also excise uracil analogues, like the synthetic chemotherapeutic drug- 5-FU, oxidized pyrimidines like 5-hydroxyuracil (5-hU), alloxan, and isodialuric acid. These analogues contain minor modifications at the 5'-position of their pyrimidine ring (54-56).

UNG1 is expressed in all tissue types while UNG2 is more abundant in proliferating tissues, like small intestine, colon, thymus, and testis (49). The protein and mRNA levels of UNG2 are cell cycle dependent and are upregulated several fold in S phase, in which DNA replication takes place and the tendency of dUTP misincorporation is greater (57,58).

Human UNG2 is found to take part in the mutagenic processing of uracil in antigen stimulated B-cells as induced by AID to initiate CSR and SHM (54). The mechanism is discussed briefly in the next major section. Nilsen *et al.* showed a high genomic uracil accumulation rate in the isolated nuclei of mouse embryonal fibroblast (MEF) cells obtained from *UNG* knockout mice with respect to MEF cells from wild-type mouse (35).

1.2.2.1.2. Single-strand-selective monofunctional uracil DNA glycosylase1

Single-strand-selective monofunctional uracil DNA-glycosylase (SMUG1) was discovered in *Xenopus laevis* and received its name for high binding affinity to Ura in single stranded DNA (59). It is active for both U:A and U:G in double strand and single stranded DNA (60,61). SMUG1 has a low turnover number compared to UNG, but can be involved in repairing premutagenic U:G lesions in non-replicating chromatin due it its high binding affinity to U:G substrate (62). SMUG1 acts as an immediate backup for UNG2 and is constitutively expressed during the cell cycle. In a study conducted by Doseth *et al.*, SMUG1 was shown to be essentially specific for ds-DNA in the presence of Mg^{2+} and other divalent salts, despite its name. The same study reveals that this enzyme, in particular, is 200-fold less efficient than UNG2 in removing uracil from ss-DNA in AID hotspot sequences in the immunoglobin (*Ig*) loci, thus making it inefficient in antibody diversification. SMUG1 is 8-fold more efficient in mice than in human (63).

A unique ability of SMUG1 is to excise oxidized pyrimidines with bulky substitutions at the C5 position, such as 5-hydroxymethyluracil (5-hmU) and 5-formyluracil (5-foU). It is also able to excise the other uracil analogues (54,64,65).

1.2.2.1.3. G/T mismatch specific Thymine DNA glycosylase

TDG is an important mismatch specific glycosylase and is involved in epigenetic regulation during embryonal development. It removes uracil lesions from CpG islands in DNA (66). TDG shows high preference for U:G over T:G lesions and 5-methyl cytosine (5-mC), and a low affinity towards U:A substrate (67). It has a much lower catalytic turnover than UNG2 and binds with high affinity to AP- sites until released by APE1 in the following step of the BER pathway (68-70). This makes TDG slower and less efficient than UNG2. *In vitro* studies conducted by Hardeland U. *et al.* show that SUMOylated (small ubiquitin like modifier) TDG loses efficiency to excise T:G substrate and promotes its detachment from the AP sites (reviewed in (3)).

TDG is highly expressed in G1 and G2/M and found in low amounts in S-phase (70). TDG acts as the main backup to excise most of the modified pyrimidine rings that are actively removed by UNG and SMUG1 (reviewed in (3)).

1.2.2.1.4. Methyl-CpG-binding domain protein 4 (Methyl-CpG-binding endonuclease 4)

Methyl binding domain protein 4 is probably involved in epigenentic regulation in which it targets methylated CpG DNA (71). It purpotedly helps preventing mutagenesis from the spontaneous deamination of cytosine and 5-mC by removing U:G and T:G mismatches. In addition, MBD4 removes 5-FU, 5-mC, and ε-C opposite to G. Functional loss of MBD4 is associated with microsatellite instability and observed in various types of human cancer (reviewed in (3)).

1.2.3. Mismatch repair pathway (MMR)

Mismatch repair (MMR) proteins are recruited to repair single base mismatches and erroneous DNA replication loops (72). Defects in the MMR pathway increase mutation frequency and thus lead to oncogenesis. Human mismatch repair proteins vary with each other

in respect to substrate specificity and structure. MutS homolog 2/ MutS homolog6 (MSH2/MSH6) heterodimer recognize both single base mismatches and defective insertion/deletion loops while MutS homolog 2/ MutS homolog3 (MSH2/MSH3) heterodimer mostly binds to loops. Human MLH1 and post-meiotic segregation protein 1 (PMS1) or PMS2 act downstream of MSH2/MSH6 by forming a heteroduplex and migrate towards the end of a free the 3'-OH. In bidirectional MMR, if a nick is positioned at the 5'-end, then 5'->3' exonuclease I (EXO1) degrades the strand from the damage site. If a nick is in 3'-position, MLH1/PSM2 nicks a 150-nucleotide long stretch of DNA and EXO1 subsequently digests the strand. RPA and PCNA recruit Polδ/ε to finish the polymerization process and DNA is ligated by LIG1 (Figure 1-4).

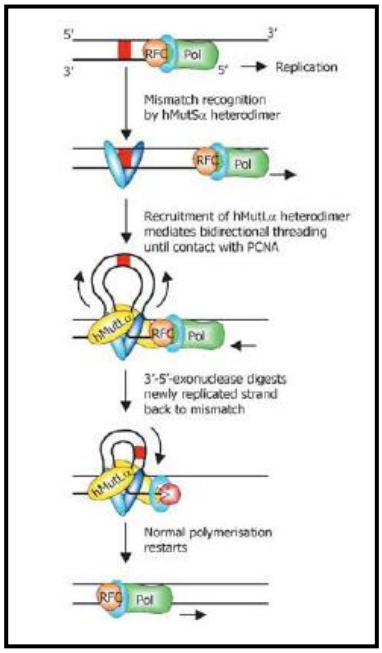

Figure 1-4: Visual summary of mismatch repair pathway. MSH2/MSH6 recognizes mismatches and MSH2/MS3H binds to loops. PMS1 or PMS2 act downstream of MSH2/MSH6 by forming a heteroduplex and migrate towards the end of a free 3'-OH. 5'->3' EXO1 degrades the strand from the damage site and DNA is ligated by LIG1 (Krokan-Kavli, 2004)

Defects in MMR proteins are found in several cancer types including hereditary non-polyposis colon cancer (HNPCC). A hallmark of HNPCC is microsatellite instability that is less frequently found in sporadic colon cancers. Mutations in MLH1 and MSH2 are found in almost 60% and 35% of genetically characterized HNPCC cases.

The MSH2/MSH6 heterodimer plays an important role in somatic hypermutation in which it recognizes U:G mismatch and converts it to A:T mutations by error prone translesion

synthesis (TLS) polymerase, Pol eta (Pol η). Recent findings reveal the importance of MMR proteins in a mouse cohort study. Triple knockout mice (*UNG$^{-/-}$SMUG1$^{-/-}$MSH2$^{-/-}$)* have relatively much shorter life span than single (*UNG$^{-/-}$* or *SMUG1$^{-/-}$* or *MSH2$^{-/-}$)* and double knockout mice (*UNG$^{-/-}$SMUG1$^{-/-}$)* (51). Thus, it could be said that MMR proteins can be considered as the last line of defense in uracil processing.

1.2.4. Mutagenic uracil processing in adaptive immunity

U:G mismatches are necessary for proficient adaptive immunity, but deleterious if left unrepaired in a normal stretch of genome other than the *Ig* genes. In response to antigen stimulation, the B-cell specific AID triggers class switch recombination (CSR) and somatic hypermutation (SHM) by targeted cytosine deamination within the recombined *Ig* gene locus (73). The derived uracil is then carried forward via replication or processed by error-prone repair giving rise to mechanisms that generate very specific antibodies (SHM) with different effector functions (CSR).

To initiate SHM, AID works on ss-DNA by introducing U:G mismatches in the switch region (S) and variable region (V) of *Ig* loci. The uracils are recognized and excised by UNG2 creating abasic/ AP sites in the respective *Ig* regions. These AP sites are subjected to three different fates: <a> error-free BER restoring original sequence by insertion of the correct base by a DNA polymerase in a multi-step mechanism <b> generation of any possible mutation by using AP-sites as template during replication, which enables the insertion of any base and any possible transition and transversion at the C:G pair; <c> nicking of the AP-site by APE1 or APE2 and successive gap filling by low-fidelity polymerases (74) (reviewed in (3)).

CSR depends on generation of single stranded breaks in DNA at the S- region of *Ig* loci most likely in the G1-phase of cell cycle (75). UNG2 initially removes the uracil generating AP-sites that are next nicked by APE1 and APE2 to yield single stranded breaks (76). These breaks arise in two separate S regions and are joined by non-homologous end joining (NHEJ) to give rise to switched isotypes. The next step follows by converting the 5' and 3' overhangs into blunt end by EXO1 (Figure 1-5).

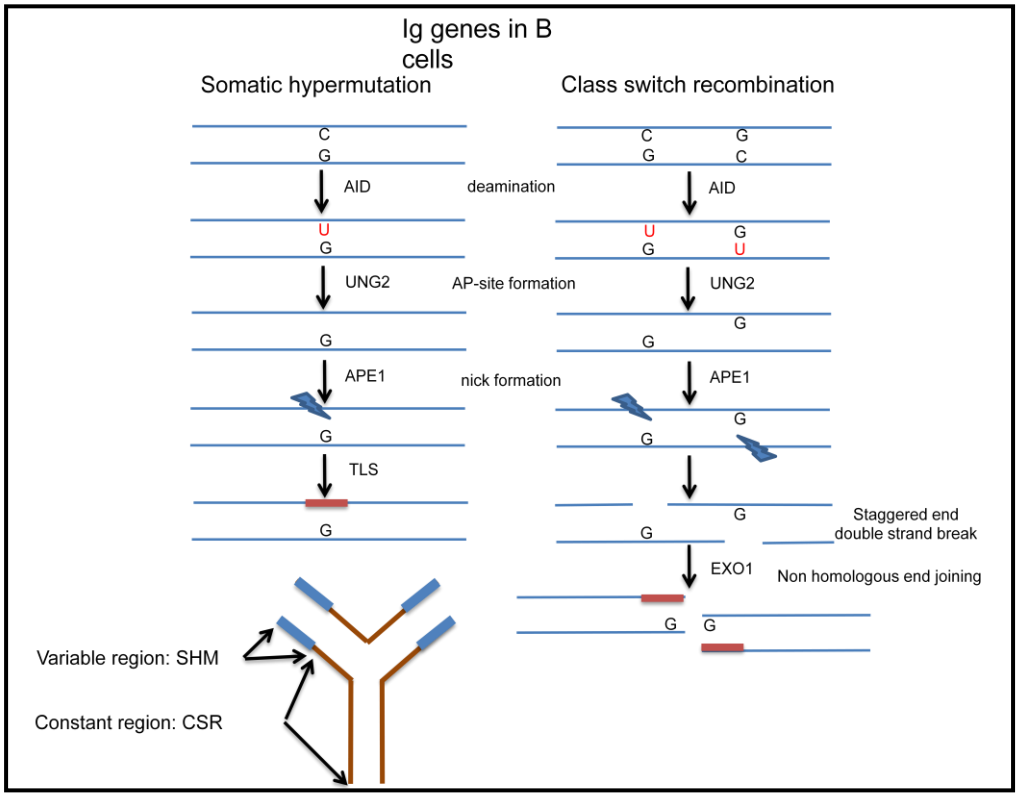

Figure 1-5: Model of somatic hypermutation and class switch recombination

1.2.5. Pathophysiology- Role of AID and glycosylases in tumorigenesis

AID is considered the "master molecule" for uracil generation in the *Ig* genes in adaptive immunity (77). AID, which is not only expressed in germinal centre B-cells but also in cells of other tissues, has several pathophysiological conditions related to mutation of the *AICDA* gene (78). Figure 1-6 shows the different physiological roles of AID. Details are explained in the subsequent sections.

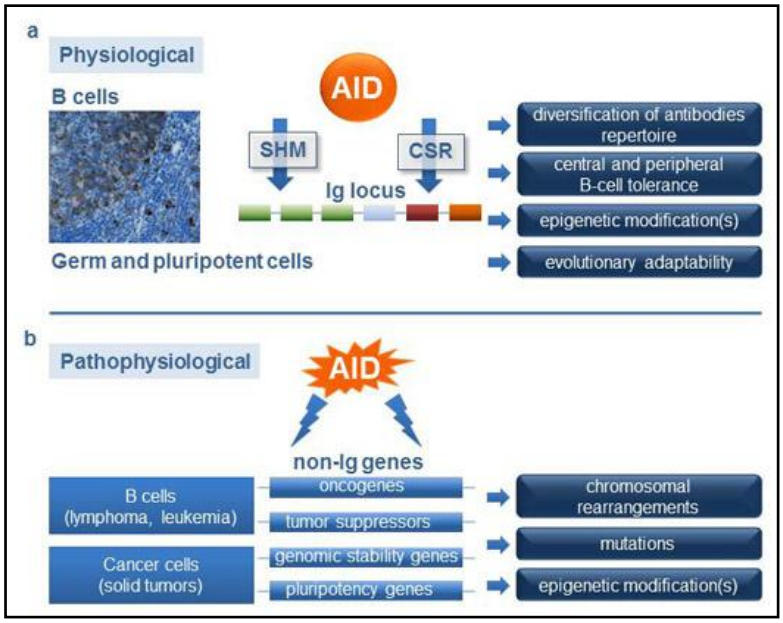

Figure 1-6: Physiological and pathophysiological function of AID (79)

1.2.5.1. Types of hematological malignancies and hyperplasia

B-cell malignancies are marked by the presence of chromosomal translocation between an *Ig* locus and a proto-oncogene, and the consequences are dysregulated CSR (80). Studies conducted on c-myc overexpressing and BCL-6 overexpressing transgenic mice indicate that AID induced deamination in the B-cell germinal centre during antibody diversification lead to lymphomagenesis (81). Experiments conducted by Dorsett Y. *et al* show that AID is an indispensible molecule in the translocation between c-myc and IgH variable region subsequently leading to hyperplasia. Removal of the enzyme from malignant B-cell lines lessens the occurrence of chromosomal breaks and rearrangement in the transcription start sites and decreases elevated levels of transcription-coupled deamination in ss-DNA (82). However, mice models with *AID* gene knockout exhibit a reduced mutation rate in connection to B-cell lymphomas (83) (reviewed in (43)).

Studies have revealed 180,000 different types of rearrangements from about 400 million B-cells. To name some general types are: Mantle zone lymphoma is associated with *bcl-1/Ig* translocations, follicular lymphoma with *bcl-2/Ig* translocations and *c-myc/Ig* translocation is found in Burkitt's lymphoma (80) (reviewed in (43). Mouse model studies suggest that AID is overexpressed in cells affected by infections, which may consequently give rise to B-cell lymphoma (by Epstein Barr virus), gastric carcinoma (*H. pylori*) (84), and

liver cancer (HCV). Apart from this, ubiquitously expressed AID leads to only T cell malignancies in some mouse models (reviewed in (43)).

1.2.5.2. Somatic hyperplasia/ *Kataegis*

Mutations at C:G base pairs are prevalent in most common types of human cancer. Cytosine deamination leading to U:G mismatches or translesion bypass of an abasic site give rise to C to T transitions, C to G transversions, and less frequently C to A transversions, particularly on TCA, TCC, and TCT contexts (85).

The majority of the APOBEC family members may potentially cause cytosine deamination in DNA leading to cancers in breast, bladder, thyroid, cervix, lung, acute lymphoblastic lymphomas, multiple myeloma, chronic lymphocytic leukemia, and B-cell lymphomas. Certain types of malignancies do not have APOBEC mutational signatures, but develop from heavy carcinogenic exposures e.g. melanoma, small cell lung cancer and others (85) (reviewed in (43)).

Kataegis is a Greek term meaning rainfall that refers to a pattern of localized hypermutations characterized in some cancer genome (85). High deaminase expression in yeast cells demonstrated the direct link between APOBEC proteins and the kataegis clusters (86). Several malignancy studies have revealed a unique observation of regional localized hypermutations in the same strand of a cancer related gene and its association to genomic rearrangement. These hypermutations most commonly are linked to C to T transitions and C to G transversions (85) (reviewed in (43).

1.2.5.3. Physiological deficiencies

AID deficiency is associated with a hyper IgM phenotype. Mutations in the *UNG* gene cause impaired CSR with partial SHM deficiency in some types of the hyper IgM syndrome. Both SHM and CSR are impaired in AID-deficient patients, but individuals with UNG deficiency generate an altered SHM with no or reduced CSR. UNG deficiency is likely related to malignancy, as in the absence of UNG, Ura is accumulated as an end product which is supposed to initiate mutation (as shown in Figure 1-3). Studies show that UNG deficient mice likely exhibit lymphatic hyperplasia and develop B-cell malignancies at very high rates (87).

1.2.5.3.1. HIGM syndrome

HIGM syndrome is a rare hetergenous disorder (1/500,000 births) that is directly linked with normal or abundant level of low affinity IgM along with reduced or null serum concentrations of IgG, IgA, and IgE. These patients suffer from impaired CSR and variable consequences of SHM leading to recurrent respiratory tract infections and digestive disorders (88,89).

These patients are grouped into several types: HIGM2 with AID deficiency following defective CSR and deficient SHM, HIGM5 with UNG2 deficiency exhibiting reduced SHM, and the third variant with mutation in the C-terminal *AICDA* gene following preserved SHM. Other types include CD-40 deficiency, CD-40 ligand deficiency, and NEMO deficiency syndrome. Treatments for these patients include intravenous *Ig* doses (60).

1.2.6. Previous attempts to measure genomic uracil

Uracil in the genome is far less abundant than the four canonical bases A, T, G, and C and therefore difficult to directly quantify. Several indirect and direct attempts have been made to quantify genomic uracil (32).

1.2.6.1. Indirect approaches

Andersen *et al.* carried out an indirect size dependent alkaline elusion assay to quantify dUMP incorporation in $UNG^{+/+}$ and $UNG^{-/-}$ mouse embryonic fibroblast (MEF) DNA following the assay protocol from Kohn *et al.* (90). Alkaline elution works by forcing DNA in a cell lysate through a filter under alkaline pH that separates DNA fragments of different sizes as a function of time. DNA from MEF cells was pretreated with UNGΔ84 (a recombinant human UNG enzyme that lacks 84 amino acids in the N'-terminal) to excise uracil (10). Subsequent cleavage of the AP-sites by T4 endonuclease was used to introduce single stranded breaks (SSB) in DNA. Next, alkaline elution was performed to detect the size of DNA fragments. Thus the fragments eluted were compared to an untreated DNA sample. This method has a drawback as AP-sites can be formed spontaneously which hinders the actual measurement of uracil content in DNA.

Horvath *et al.* used a single-step real time PCR based method to quantify genomic uracil on synthetic DNA and genomic DNA from *E.coli* and mouse cell lines with perturbed thymidylate synthesis. They used wild type *P. furiosus* polymerase that halts immediately when encountered with Ura. They had a control polymerase from the same source that lacks the uracil specific catalytic site in it. Results from synthetic DNA were verified by direct isotopic measurements. By using different sets of primers, they analyzed the heterogeneity of Ura distribution throughout the genome (91). This assay has low sensitivity compared to others methods used to measure Ura, which marks one of the major disadvantages of this method. Another drawback is that it produces relative measurement of uracil.

A luminescence technique had been undertaken by Atamna *et al.* to measure uracil in DNA. They incubated the DNA from cell lines with UNG to create AP-sites which were then treated biotin-containing aldehyde-reacting prove (ARP) that binds irreversibly to AP sites. These cells were lysed and the isolated DNA was treated with horse-radish peroxidase (HRP) that forms a HRP-DNA complex with the biotin. HRP-DNA adduct was separated by treating them with free avidin-HRP complex and precipitating DNA with DNA precipitating dye. A chromogenic assay was then performed to measure the number of AP sites by the luminescence activity of HRP (92).

1.2.6.2. Direct approaches

Direct methods of uracil quantification, excised *in vitro* by UNG, or deoxyuridine (dUrd) employ mass spectrometric detection of these small molecules with different initial chromatographic separation procedures. Measurement of dUrd has certain advantages over Ura as the former contains N'-glycosylic bond which on fragmentation produces Ura. A higher energy is needed to break Ura which increases the background during quantification.

UNG excision has certain disadvantages in the quantification method: it has been previously reported that UNG has a substrate preference of single-strand over double strand DNA, U:G pair over U:A pair, and an explicit sequence preference. Moreover, buffer compositions can have an adverse effect on the enzyme activity. Mashiyama *et al.* performed a linearity testing by using increasing concentrations of the same genomic DNA extracted from lymphocytes that were pre-treated with restriction digestion enzymes that makes the DNA more fragmented thereby decreasing the chance of supercoiling and making its otherwise obscured uracil moieties easily accessible for UNG (93). In the experiments

conducted by Ren *et al.,* a 198-bp long DNA had been PCR amplified using dUTP instead of dTTP that counts uracil in one-fourth fraction with all other bases (94). Therefore, the uracil content in this oligomer is drastically higher than that in genomic DNA. At the same time, it is unclear whether UNG acts with the same efficiency on an oligomer than on a natural genomic DNA. Burns *et al.* measured genomic Ura by HPLC-ESI+-MS/MS method from the UDG treated DNA obtained from breast cancer cell lines and spiked with (+6) heavy labeled Ura (^{13}C and ^{15}N) (95).

Hydrolysis of DNA has been used by Chango *et al.* and Dong *et al.* to measure deoxyuridine by liquid chromatography coupled single mass spectrometry and liquid chromatography coupled tandem mass spectrometry, respectively (96,97).

The analysis of dUrd has been a challenge because of the relative abundance of naturally occurring (^{13}C)-2'-deoxycytidine (^{13}C-dCyd) which is isobaric to and MS indistinguishable from dUrd. The relative natural abundance of (^{13}C) is approximately 1.1% of all carbons (98). As dCyd elutes first from the column and there is a tendency to collect a considerable amount of dCyd with dUrd, therefore, only reverse-phase chromatographic separation is insufficient as dUrd signal overlaps with the (^{13}C)-dCyd peak tail. This difficulty was successfully overcome by Galashevskaya *et al.*, in 2013. They introduced a precursory HPLC step with a reverse-phase chromatography column (brand name Primesep 200) prior to reverse phase C18 column chromatography coupled tandem mass spectrometer. It has been reported using their assay that dUrd count in *UNG* $^{-/-}$ mouse embryonic fibroblast (MEF) DNA is 5-fold higher than that of *UNG* $^{+/+}$ MEF DNA and 11-fold higher in UNG2 mutated human lymphoblastoid cell lines than in normal UNG2 functional lymphoblastoid cells. They report lower basal genomic uracil level than previously measured by other groups (38).

1.3. Aim of the study

The focus of this thesis was to calculate the amount of uracil excised by the activity of different glycosylases. We measured the activity using different substrates to elucidate the substrate specificity of different glycosylases present in different organs of *UNG*$^{-/-}$ and *UNG*$^{+/+}$ mice. Genomic uracil was also measured on the same organs by LC-MS/MS and the number of uracil moieties present per 10^6 deoxynucleosides has been quantified. Finally, we correlated genomic uracil and UDG activity to ascertain to what extent glycosylase activity is related to the number of uracil present in the genome.

2. Materials and methods

2.1. Genotyping of transgenic mice

2.1.1. Method of genotyping

Background- Genotyping confirms whether the transgenic mice are wild-type, heterozygous or knockout.

Procedure- Ear clips obtained from Balb/C $UNG^{-/-}$ and $UNG^{+/+}$ transgenic mice (35) were subjected to DNA isolation using Qiagen Blood and Tissue kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instruction. The concentration of DNA was measured by Nanodrop ND-1000 spectrophotometer (Thermo Fisher scientific, Waltham MA, USA). A final volume of 20 µl contained 2 µl of 2-50 ng/µl of DNA, 1 U Platinum Taq polymerase (Invitrogen, Eugene, OR, USA), IX Platinum Taq buffer, 1.5 mM of $MgCl_2$, 0.2 mM dNTP mix (Finnzymes, Espoo, Finnland) and 0.2 µM of forward and reverse primers (forward primer- 5'-GGCCACCCTGACAAATCCC-3'and reverse primer- 3'-CACGGACCTAATCAAGCTCACG-5') (obtained from Sigma-Aldrich, St. Louis, MO, USA). The thermal cycler parameters are as follows in Table 2-1:

Table 2-1: PCR parameters

Initial denaturation	98°C	30 seconds	
Denaturation	98°C	10 seconds	30 cycles
Annealing	65°C	30 seconds	
Elongation	72°C	2 minutes	
Final elongation	72°C		

Agarose gel was prepared by using 2% w/v of agarose, 1X Tris- acetic acid-EDTA (TAE) buffer and 5000X gel red. 5 µl of the PCR products were mixed with 1 µl of Biotium 6X loading dye (0.25% bromo-phenol-blue and 30% glycerol) and electrophoresis was performed with 1X TAE buffer at 100V for 25 minutes and analyzed on a Gel Logic 200 imaging system by Kodak Molecular Imaging software (Kodak, Rochester, NY, USA).

2.2. Mice selection and organs harvest

After strain determination by PCR, three *UNG* $^{+/+}$ and *UNG* $^{-/-}$ mice were selected. These mice were euthanized by cervical dislocation and their organs were harvested at the quarantine animal laboratory. Organs include kidney, liver, brain, lung, muscle, and intestine.

2.3. Preparation of protein extract

2.3.1. Background

Protein extracts prepared from tissues by homogenization was used for measuring different glycosylase activities. The tissue extracts contain all type of proteins including uracil DNA glycosylases, which when incubated with uracil-containing DNA/oligomer *in vitro* at physiological conditions excises the uracil that is measured using two different methods as described in the following sections. Proteins may become denatured, unfolded, or damaged during extraction and purification. To maintain a physiological pH, Tris-HCl and KCl were used as components in our buffer solution. Osmolytes like glycerol stabilizes the conformation of protein while in store thus rendering equal *in vitro* enzymatic activity when in reaction under physiological conditions. Non-ionic detergent NP-40 is less harsh on the protein than other ionic detergents during extraction and minimizes loss of enzymatic activity during analysis. At lower concentrations, dithiothreitol (DTT) reduces the disulphide bonds and maintains the reduced monothiol state of the protein. Thus *in vitro*, DTT stabilizes the free sulfhydryl bonds of the enzyme and restores its activity which could be lost by oxidation. Protease inhibitors inhibit the activity of proteases that could otherwise digest the proteins during purification. Tissue homogenization was achieved by passing the buffered tissues through needles of two different diameters to ensure absolute lysis. The concentration of protein was measured at 595 nm by the Bradford assay. This is a colorimetric protein assay that measures the absorbance shift of the dye coomassie brilliant blue G-250 which gradually turns blue from its original red color, in acidic conditions. The positively charged amine groups of the protein binds to the negatively charged dye molecules by ionic interaction that produces the intensity of the blue color of the dye. Therefore, the intensity varies with the number of amino groups present in the protein.

2.3.2. Procedure

We used two standard lysis buffers for protein isolation of which the first buffer consisted of 10mM of Tris-HCl and 200mM of potassium chloride (KCl) and the second buffer contained 10 mM of Tris-HCL, 200 mM of potassium chloride, 40% glycerol and 0.5% NP-40. To 1 ml of each these buffers 0.001 µg of DTT, 1 µl of 1X complete, protease inhibitors- protease inhibitor cocktail 2 (PIC2) and protease inhibitor cocktail 3 (PIC3) were added. Each buffer complete were added in equal volumes, the amount being 0.6 µl of each buffer per 1 µg of tissue. With each addition of buffer, the mixture was minced by scalpel, thoroughly vortexed and subjected to homogenization through 21G and 25G needles. Reaction tubes were maintained on ice throughout the process. Homogenized extracts were incubated for two hours at 4^{o}C and then centrifuged at 13.200 rpm for 15 minutes at 4^{o}C to precipitate the cell debris. Supernatant containing protein was collected and concentration was determined by Bradford assay (Bio-Rad) using bovine serum albumin (BSA) as standard. Tissue extracts were aliquoted, snap frozen in liquid nitrogen and stored at -80^{o}C. The purpose of aliquoting was to retain the enzyme activity intact in the extract and to use each aliquot tube only once during the assay.

2.4. UDG activity assay

2.4.1. ^{3}H Ura release assay to measure glycosylase activity

2.4.1.1. Principle

Kavli *et al.* described the UDG activity assay which is employed in our experiment to measure UDG activity in our tissue extracts (54). The principle of this assay is to incubate a specific amount of tissue/cell extract with nick-translated calf thymus DNA containing tritiated dUMP (^{3}H-dUMP), specific to U:A. and then to precipitate the DNA and protein reaction mixture in order to measure the liberated ^{3}H-Ura excised by the present glycosylases in the extract. The amount of excised ^{3}H-Ura measured by radioactive tritium decay in a scintillation counter relates to the amount of glycosylase present in the extract and their efficiency to initiate BER by excising uracil. The concept behind scintillation is to incorporate the measuring substance into a scintillation fluid- consisting of solvent and scintillators. Radioactive decay of the radio-substances present in sample constantly emits β-particles that

excite the scintillator which can then be measured by a very narrow range of electromagnetic radiation (light) in a scintillator counter. In our case, the tritium present in uracil moiety is the source of radioactivity. We measured glycosylase activity from the obtained disintegration per minute (DPM) values.

2.4.1.2. Procedure

Briefly, the protocol uses a total volume of 20 µl of assay mixture containing 5 µg of tissue extract and 1.8 µM of nick translated ^{3}H-dUMP-labeled calf thymus DNA (specific activity 0.5 mCi/µmol) in 1X UDG buffer. The 1X buffer mixture contains 20 mM Tris-HCl, pH 7.5, 60 mM NaCl, 1 mM DTT, and 0.5 mg/ml of BSA. The reaction mixture was incubated for 10 minutes at 30^{o}C and terminated by placing on ice. Next, 50 µl of 1 mg/ml salmon sperm DNA carrier (in 10 mM Tris-HCl, pH= 7.5 and 1 mM EDTA) and 500 µl of fresh ice-cold 5% trichloroacetic acid (TCA) were added subsequently and incubated on ice for 10 minutes to precipitate DNA. The mixture was centrifuged at 13,200 rpm for 10 minutes at 4^{o}C to separate precipitated DNA from (^{3}H)-U containing supernatant. Ultimately, 500 µl of supernatant were added to 5 ml of Ready Protein+ Liquid Scintillation Cocktail (Beckmann Coulter). Acid-soluble tritiated uracil was quantitated in a scintillation counter by Tri-Carb 2900TR liquid scintillation analyzer using Quantasmart v.2.02 software (Perkin Elmer, Waltman, MA, USA).

2.4.2. Oligonucleotide UDG cleavage assay

2.4.2.1. Principle

In the oligonucleotide cleavage assay, the 5'-end is either labeled with radioactive phosphate (^{32}P or ^{33}P) or a fluorescent compound, whereas the damaged base (in our case Ura) is in the middle of the same oligonucleotide. The labeled oligonucleotide can be used as a substrate either in the single-stranded form or after annealing to a complementary strand. The single or double stranded substrate is then incubated with cell/tissue extract or purified glycosylase that will release the damaged base, leaving an apurinic/apyrimidinic site that can be cleaved with pipiridine. Next, the uncleaved substrate and the cleaved product strands are separated by 12% polyacrylamide gel electrophoresis (PAGE) and quantified, in our case by ImageQuant TL software (GE Healthcare). Figure 2-1 shows the resulting product and substrate band formed after PAGE.

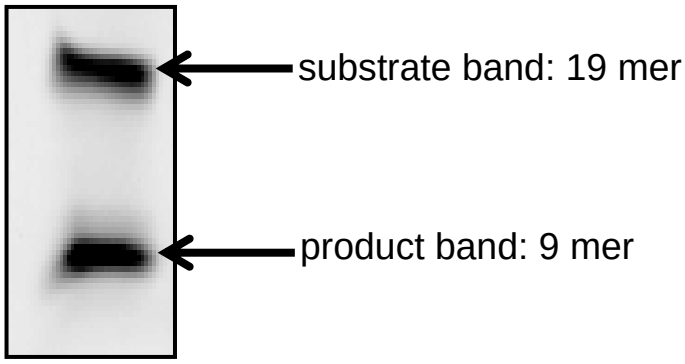

Figure 2-1: Gel image showing subsequent substrate band and product band

2.4.2.2. Procedure

Double stranded U:G and U:A substrates were prepared by annealing a 6-flourescein-amidite (6-FAM)-labeled 19 oligonucleotide long strand containing uracil to a complementary strand containing either A (match) or G (mismatch) opposite to Ura. The sequences of the mentioned substrates are listed in Table 2-3. 6-FAM is attached to the 5'-end of the oligo. A total volume of 10 µl assay mixture contains 20 nM oligonucleotide substrate with 15 µg (for double stranded substrate) and 5 µg (for single strand substrate) of tissue extract dissolved in UDG assay buffer. This reaction buffer contains 20 mM of Tris-HCl, 60 mM sodium chloride, 1 mM EDTA, 0.5 mg/ml BSA, and 1 mM DTT. For each substrate, four sets of reaction tubes were prepared to measure the activity of different glycosylases. 1 µg of UNG inhibitor (Ugi) obtained from *Bacillus subtilis* phage PBS2 (99) and 4 µg of anti-hSMUG1-antibody (polyclonal PSM1 which was prepared in our lab) were added to inhibit endogenous glycosylases so as to measure SMUG1 and UNG activity in the respective tubes. In the third set of tubes, both types of inhibitors were added to observe the residual activity representing TDG and MBD4. The fourth set contained no inhibitors and measures the total UDG activity. Table 2-2 provides a summary of each of the tubes:

Table 2-2: Compositions in the four sets of tubes used during experiment:

Tube A	+ PSM1- inhibits SMUG1, measures TDG and MBD4 activity but mostly UNG activity
Tube B	+ Ugi- inhibits UNG, measures TDG and MBD4 activity but mostly SMUG1 activity
Tube C	+ PSM1 +Ugi- inhibits both SMUG1 and UNG and represents activity from TDG and MBD4
Tube D	Without any inhibitors to measure the total glycosylase activity

The reaction mixes were incubated at 37^{o}C for 35 minutes for U:G substrate, 70 minutes for U:A substrate and 30 minutes for ss-U substrate. The reactions were stopped by placing on ice and immediately adding 50 µl 10% piperidine following heating at 90^{o}C for 20 minutes to denature DNA and cleave the AP-sites. Piperidine was vacuum evaporated for 60 minutes at 60^{o}C and redissolved with 30 µl of 60% loading buffer. Each 25 ml of 12% polyacrylamide gel contained 420 mg/ml of urea, 10X TBE (Tris-Boric acid-EDTA) buffer, 12% acrylamide/bis (19:1); 0.1% of ammonium persulphate (APS) and 25 µl of tetramethylethylenediamine (TEMED) were added freshly during gel casting to ensure proper polymerization. Wells of the newly polymerized gel were washed with 0.5X TBE buffer and allowed to prerun for more than 20 minutes at 200V. Wells were washed again and samples were loaded to separate product and substrate by 12% polyacrylamide gel electrophoresis (PAGE). Subsequently, gel slabs were laser scanned on Typhoon Trio imager and quantified by ImageQuant TL software (GE Healthcare).

Table 2-3: Sequences of the different Ura substrates used in assay: (6'-FAM is attached to the 5'-end of each strand that together count 19 nucleotides.)

Substrate	Sequence
U:G	5'-CATAAAG**GUA**AAGCCTGG-3' 3'-GTATTTCC**G**TTTCGGACC-5'
U:A	5'-CATAAAG**GUA**AAGCCTGG-3' 3'-GTATTTCC**A**TTTCGGACC-5'
ss-U	5'-CATAAAG**GUA**AAGCCTGG-3'
T:G	5'-CATAAAG**GTA**AAGCCTGG-3' 3'-GTATTTCC**G**TTTCGGACC-5'

2.5. DNA isolation and genomic uracil quantification

2.5.1. Background

DNA was isolated from mice organs using Qiagen Blood and Tissue kit according to the manufacturer's instruction. For complete lysis, tissues were passed through a homogenizer and a series of needles varying in diameter. DNA was eluted with milliQ water as it is a polar molecule and more soluble in water. Alkaline phosphatase treatment of the purified DNA was followed to dephosphorylate any cellular dUMP that might have been co-isolated during the procedure, which is considered as a source of background interference in absolute dUrd measurement. UNGΔ84, a recombinant form the complete catalytic domain of human UNG2 (100) was used to excise uracil from DNA prior to hydrolysis. Along with UNGΔ84, HindIII restriction enzyme was added to allow fragmentation that makes the otherwise obscured uracil in DNA easily accessible to UNG. Later, the hydrolyzing enzymes DNase I, nuclease P1 (DNA cleaved to short oligomers and single dNMPs) and alkaline phosphatase (dephosphorylating dNMPS to dNs) are used to convert DNA into deoxynucleosides.

2.5.2. Procedure

2.5.2.1. Isolation of DNA and removal of intracellular dUMP

25 mg tissue (10 mg spleen) were suspended in 200 µl of buffer ATL and 5 µl of 10mg/ml RNase A from Qiagen Blood and tissue kit (Qiagen, Germantown, MD, USA) and

with 0.7µl of 100mM tetrahydrouridine (THU) which is a deaminase inhibitor. Samples were then passed through Dounce homogenizers and though 19G, 21G, 23G and 25G needles, incubated at 37^{o}C for 5 minutes. 20 µl of proteinase K was added next, incubated at 37^{o}C for 55 minutes and DNA was extracted by DNeasy kit over milliQ water. The concentration was determined by Nanodrop ND-1000 spectrophotometer (Thermo Fisher scientific, Waltham MA, USA). To remove of the intracellular 2'-deoxyribonucleotides, DNA was incubated in 0.2 U/µl of alkaline phosphatase with 100 mM ammonium bicarbonate (pH 7.6) for 30 minutes, precipitated in isopropanol and washed with 70% ethanol.

2.5.2.2. Uracil excision, DNA hydrolysis and sample preparation

In a total reaction volume of 30 µl, 15 µg of DNA was buffered with 20 mM Tris-HCl (pH 7.5), 60 mM NaCl, 1 mM of DTT, 1 mg/ml of BSA and treated with HindIII and 0.075 U of UNGΔ84 at 37^{o}C for 1 hour to excise the uracil contained in DNA. This deuracilated DNA assists in calculating the rate of Cyt deamination during sample workup, though a probable disadvantage is it may increase the background if all the Ura present in DNA is not excised. In another reaction tube, the same parameters were maintained, except for the addition of UNGΔ84 so as to keep the uracils in DNA intact. Then DNA was precipitated with isopropanol as described in the previous paragraph and resuspended with 30 µl of 100 mM ammonium acetate (pH 6.0), 1 mM $MgCl_2$, 1 mM $CaCl_2$, 2 U DNase I and 0.2 U nuclease P1, incubated for 30 minutes at 37^{o}C. This step marks hydrolysis of the DNA. (2-^{13}C;1,3-$^{15}N_2$)-2'-deoxyuridine (^{13}C-$^{15}N_2$-dU) was used as an internal standard (IS). Afterwards, the reaction was buffered in ammonium bicarbonate (pH 7.6) to achieve a final concentration of 100 mM, and incubated for 20 minutes at 37^{o}C with 0.1 U alkaline phosphatase to remove the phosphate groups. Contaminants were precipitated by adding three volume equivalents of ice-cold acetonitrile and centrifugation at 4^{o}C for 30 minutes at 13,200 rpm. Supernatants were transferred to new tubes and vacuum centrifuged until dry at room temperature (RT). Finally, samples were dissolved in 100 µl of 10% acetonitrile in water.

2.5.2.3. Preparative purification of dUrd

Prior to LC-MS/MS, a precursory HPLC step was performed to purify dUrd and to separate from dCyd. The process was executed using a reverse phase column with weak acidic ion-pairing groups (Primesep 200, 2.1 mm x 150 mm, 5 µM, SIELC technologies, Prospect Heights, IL) and an Agilent 1100 series HPLC system, with a multiple wavelength

detector (Agilent Technologies, Waldbronn, Germany). Samples were stored at 4^{o}C prior to injection. Each sample was injected with an injection volume of 30 µl in triplicates. The gradient buffer consisted of solution A (0.1% formic acid in water) and solution B (0.1% formic acid in methanol), that uses the following program with a flow rate of 400 µl/min: start ramping with 10% of solution B, ramping of 60% B for 2 minutes, again ramping of 60% B for 2.5 minutes, holding this setup for 7.5 minutes and re-equilibrating with 10% B for 2.55 minutes. Fractions, containing dUrd and dU-IS, were collected which were centrifuged in vacuum at RT until dry. Dry pellets were re-dissolved in 25 µl of 5% methanol and water solution.

2.5.2.4. Instrumentation and reaction conditions of LC-MS/MS

Deoxyuridine was quantified by a LC-20AD high performance liquid chromatography system coupled (Shimadzu Corporation, Kyoto, Japan) to an API5000 MRM- tandem mass spectrometer (Applied Biosystems, Carlsbad, CA, USA). HPLC consisted of a Zorbax SB-C18 reverse phase column (2.1mm x 150mm, 3.5µM, Agilent Technologies, Santa Clara, CA, USA), protected by a Zorbax Reliance guard column (4.6mm x 12.5mm, Agilent technologies). Each sample was injected with an injection volume of 20 µl. The gradient buffers consisted of solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in methanol), that uses the following program with a flow rate of 300 µl/min: 0.5 minutes of ramping with 5% of solvent B, ramping of 80% B for over 1 minute, ramping of 30% B for 2 minutes holding this setup for 7 minutes and re-equilibrating with 5% B for 2.1 minutes. Results were analyzed on Analyst v.1.4.2 software (Applied Biosystems).

3. Results and Discussion

3.1. Target selection

3.1.1. Choice of mice

Two mouse strains, *UNG*$^{+/+}$ and *UNG*$^{-/-}$, were used to estimate tissue specific glycosylase activity and to quantify deoxyuridine by LC/MS/MS analysis. The mice selected for analysis were age matched (6 months old). We collected three *UNG*$^{+/+}$ and *UNG*$^{-/-}$ strains and confirmed their genotypes by performing a PCR assay. We chose six proliferative and non-proliferative organs to check the glycosylase level in each of them: brain, small intestine, kidney, liver, lung and muscle. For the LC/MS/MS quantification, we included spleen as well. The spleen could not be used for glycosylase activity because the weight of the organ is less than what was calculated as necessary for protein and DNA isolation at the same time.

3.2. UDG activity measurement

We present UDG activity in two types of measurements. For glycosylase activity measurement by ^{3}H-Ura release, we report the units in the amount of Ura released per milligram of tissue extract (µmol Ura mg^{-1} min^{-1}). In the glycosylase activity assay by oligonucleotide cleavage, specific glycosylase activity was measured and will be reported in nanomoles of uracil excised per minute per milligram of tissue extract (nmol Ura min^{-1} mg^{-1}). First, we analyzed the percentage of substrate cleavage by dividing the intensity of the product band by the total intensity of the product and substrate band. Then, considering the other defined parameters we calculate 100% uracil excision. The value obtained from this calculation serves as a constant to calculate the number of uracil molecules excised by different glycosylase activity. The equation below uses U:G substrate, 35 minutes of incubation time and 15 µg of tissue extract for example:

$$100\ \%\ \text{specific glycosylase activity} = \frac{\dfrac{\text{Total amount of substrate (nmol Ura)}}{\text{Total reaction time (min)}}}{\text{Amount of protein (mg)}}$$

$$= 3.8\text{x}10^{-4}\ \text{nmol Ura min}^{-1}\ \text{mg}^{-1}$$

(where, substrate amount = 0.0002 nmol, reaction time = 35 mins @ 37 ^{o}C, protein amount = 0.015 mg)

For e.g., for 40 % specific glycosylase excision activity

$$= \frac{40 \times 3.8 \times 10^{-4}}{100} \text{ nmol Ura min}^{-1} \text{ mg}^{-1}$$

$$= 1.52 \times 10^{-4} \text{ nmol Ura min}^{-1} \text{ mg}^{-1}$$

This calculation assumes linearity with time, which is not actually the case for this assay.

3.3. ^{3}H Ura release assay to measure glycosylase activity

In this assay, uracil has been introduced into a calf thymus DNA by nick translation that mimics uracil incorporated during replication *in vivo*. The dUMP of this long U:A substrate is radiolabelled (^{3}H-dUMP). We started our experiments using Tris-buffer containing 10 mM of NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$ that was prepared to serve as a universal buffer, for example, uracil excision, hydrolysis, etc. But we experienced low and unstable UDG activity with this buffer each time we measured. So, we tested different salt concentration in Tris-buffer to verify the best salt condition for UDG activity which is discussed in details in section 3.3.1.

3.3.1. Testing of salt concentration in Tris-buffer solution

The salt concentration may influence the enzyme activity. Firstly, it can affect the reaction equilibrium of binding DNA that in turn has an effect on the outcome of enzyme activity. Certain salts enhance or diminish the activity of an enzyme. Most importantly, the salt concentration may have an impact on the stability of the protein: enzymes are usually more stable in higher salt concentrations, but salts frequently either reduce or increase enzyme activity.

We tested four different salt concentrations with increasing ionic strength of monovalent ions (0-60 mM NaCl) and addition of two divalent ions (Mg^{2+} and Ca^{2+}) in one particular Tris-buffer. We used recombinant UNGΔ84 to test different buffers which had been diluted the stock enzyme to 10^{6} with each of the four mentioned buffer compositions varying in salt concentration. While the experiments with UNGΔ84 demonstrate that salt conditions do have an effect on the enzyme activity representing the major UNG, they might also influence SMUG1, TDG and MBD4 differently. This was not investigated due to time limitations. The final salt concentrations of these buffers are given in Table 3-1:

Table 3-1: Different salt concentration in 20 mM Tris-buffer (pH 7.5):

Buffer 1	10 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$
Buffer 2	20 mM NaCl
Buffer 3	No salt
Buffer 4	60 mM NaCl

UNG activity increases with the increase in NaCl concentration of, in our case probably due to Na^+ (Figure 3-1). We observed the highest enzyme activity (2754 µmol Ura mg^{-1} min^{-1}) with reaction buffer 4 containing 60 mM NaCl. The activity was lower with reaction buffers containing less salt or no salt at all (below 850 µmol Ura mg^{-1} min^{-1} for most cases). It has been previously reported that Mg^{2+} increases the activity of full length UNG2 (101) and decreases UNGΔ84 activity 2-fold. As UNG activity increases with the increase in Na^+ concentration, we could have tested the impact of Mg^{2+} on higher NaCl concentration, but this was not performed.

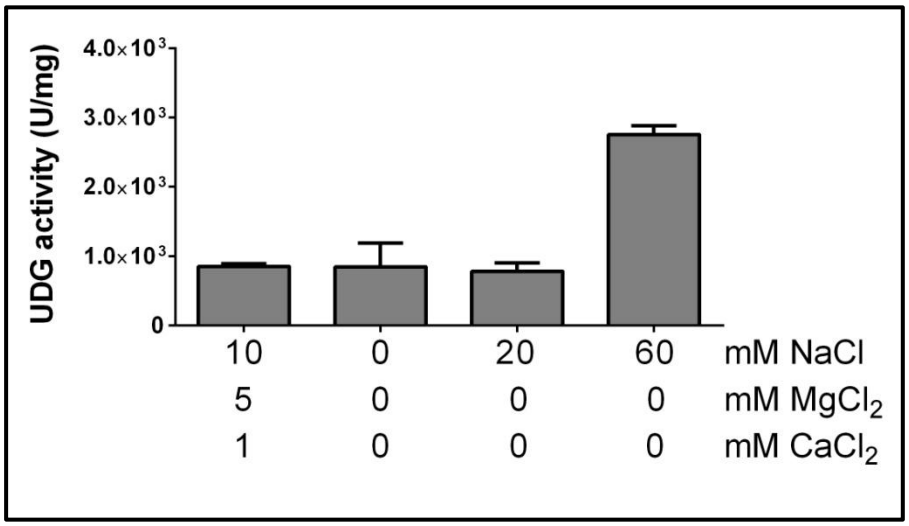

Figure 3-1: UNGΔ84 activity in different salt conditions. The enzyme shows optimal activity in the presence of 60 mM NaCl

It has been reported by Doseth *et al.* that Mg^{2+} does not significantly stimulate the activity of mouse UNG2 (63). Our samples were harvested from mouse and we selected buffer 4 to run the tissue extracts obtained by lysing different organs from mouse. This buffer 4 is referred to as **UDG buffer** in rest of the text.

3.3.2. Measurement of glycosylase activity in different organs

The tissue extracts used in this assay were obtained from the organs harvested from age matched mice that had been stored for three years at -80^{o}C. However, the wild type and knock-out counterparts of all the organs were lost in storage. To obtain statistically significant results, we needed at least three mice from each of *UNG*$^{+/+}$ and *UNG*$^{-/-}$ strains. Initially, we had only selected organs from two *UNG*$^{+/+}$ mice and four *UNG*$^{-/-}$ mice: heart and thymus from one *UNG*$^{+/+}$, and muscle and heart from one *UNG*$^{-/-}$ strain. We measured glycosylase activity of tissue extracts from both wild type and knock-out kidney, liver, heart, lung, and thymus and only from knock-out brain, stomach, intestines and muscle.

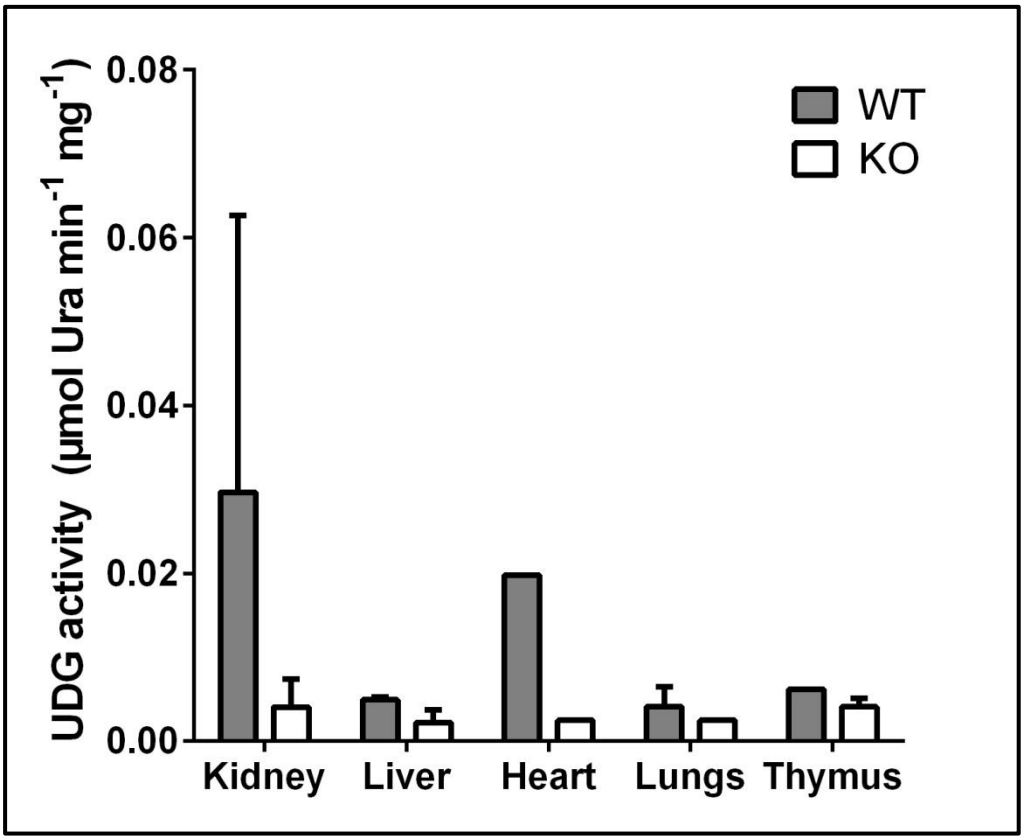

Figure 3-2: **Rate of Ura excision activity by glycosylases on ^{3}H-Ura-DNA (long U:A substrate) in different organs from *UNG* wild-type and knock-out mice.**

Total glycosylase activity in different organs obtained from *UNG*$^{+/+}$ and *UNG*$^{-/-}$ strains is presented in Figure 3-2. The knock-out strains show very little or no glycosylase activity with long U:A substrate indicating that UNG is the major DNA glycosylase measured in this assay. An exception to this case is the thymus which has not investigated further due to time limitations. Therefore, we detected mostly UNG activity in these organs. In wild-type kidney,

liver, thymus and heart, there is relatively high amount of UNG activity whereas it is negligible for the knock-out counterparts. We measured activity in *UNG*$^{+/+}$ lung and *UNG*$^{-/-}$ brain, intestine, muscle and stomach as well, but we will not discuss that further as they lack their counterparts.

Among all these organs, wild-type kidney shows the highest amount of glycosylase activity. We measured very high activity for the knock-out large intestine samples which were far higher than the wild type-kidney. It is supposed that while organ harvest, the large intestines were not properly washed and those were contaminated; it could be that we measured microbial glycosylase activity for these knock-out intestines.

3.4. Oligonucleotide cleavage assay to measure glycosylase activity in new set of mouse organs

The radioactive assay uses a long U:A substrate in which ^{3}H-dUMP is incorporated into DNA instead of dTMP by nick translation. In the previous test we lacked the wild-type and knock-out counterparts of all the organs. In order to exhibit specific results from the activity of different glycosylases on different substrates, we performed the oligonucleotide cleavage assay. Another advantage of oligonucleotide cleavage assay is that it utilizes very small amounts of tissue or cell extract to measure activity.

3.4.1. Protein concentration optimization

We observed in the radioactive assay (section 3.3.3) that the wild-type kidney has the highest amount of glycosylase activity. We chose a wild-type kidney tissue extract as our optimization sample. We utilized reaction conditions that have been previously optimized for cell lines and established in our lab (data not shown): variable amounts of tissue extract of a new kidney sample were incubated with U:G substrate and 1 µg of anti-SMUG1 antibody (PSM1) for 10 minutes at 37 ^{o}C. PSM1 was included to quantify UNG (assuming that TDG and MBD4 contribute little to the total activity). Disappointingly, our experimental set up provided no signal for WT kidney, probably because the amount of extract used here contained too little UNG to act on 20 nM substrate (Figure 3-3).

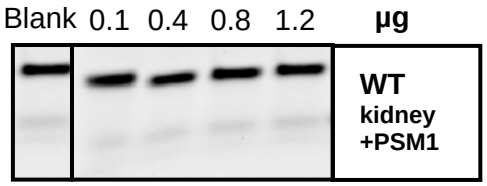

Figure 3-3**: *UNG*$^{+/+}$ kidney shows very low or no glycosylase activity with a low amount of tissue extract.** We measured UNG activity initially.

We increased the amount of extract to 5 µg and 10 µg and incubated each at three time points: 10, 20, and 30 minutes, at 37 ^{o}C and did not add any inhibitor because we wanted to measure the total glycosylase activity from this optimization assay. We tested for three types of substrates: U:G, U:A, and single-stranded-Ura. It has been previously established in our lab that uracil excision activity within 10% to 70% is considered as the quantifiable range of this assay (data not shown). For U:A substrate, 5 µg of tissue extract incubated for the three mentioned time points: 10 minutes, 20 minutes, and 30 minutes, gives excision activity of 13%, 21%, and 27%; and 10 µg of tissue extract excises 17%, 23%, and 30% of uracil at these respective time points. We planned to examine non-proliferative tissues like brain and muscle, which probably contain lower glycosylase activity than rapidly proliferative tissues (49). If we had used 10 µg of tissue extract and incubated for 30 minutes to run all our samples, then wild-type non-dividing tissues would not exhibit any detectable activity. So, we decided to increase the amount of protein to 15 µg and test different incubation times to produce an assay within the quantifiable activity range.

Glycosylases showed a higher excision rate for U:G substrate than for U:A. 5 µg of tissue extract, incubated for the three respective time points, exhibited excision activity of 27%, 44%, and 50% while 10 µg of tissue extract excises 30%, 55%, and 61% of Ura during the same duration. Rapidly dividing tissues such as the intestine likely contain high glycosylase activity. To keep the glycosylase activity of the proliferating tissues within the quantifiable range, we used 15µg of kidney extract and incubated at different time points to achieve the best level of activity that fits within the selected range.

5 µg of kidney extract shows excision activity of 28%, 48%, and 62%, while 10 µg exhibits excision of 33%, 57%, and 70% for 10 minutes, 20 minutes and 30 minutes, respectively (Figure 3-4). Here, we find that 5 µg of protein incubated for 30 minutes fits within the quantifiable range of the assay while 10 µg incubated for the same duration almost reached the edge of saturation. Thus, we successfully optimized the amount of protein and

incubation time for single-stranded-Ura substrate. We continued running extracts from all the organs using 5 µg of protein and 30 minutes incubation time.

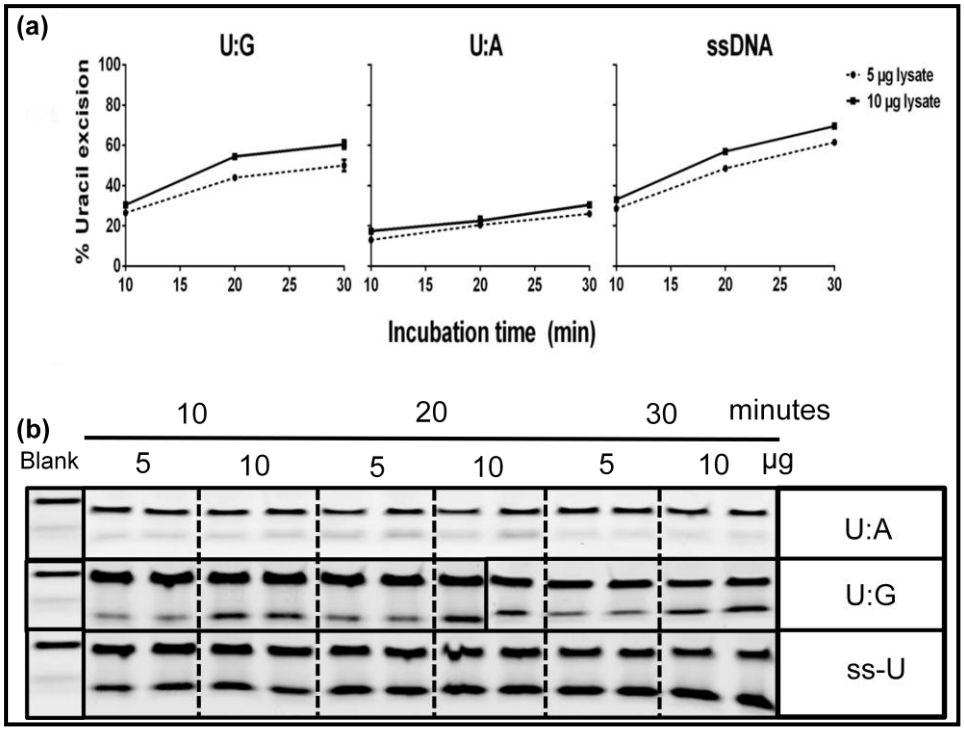

Figure 3-4: Optimization of protein concentration. 5 µg and 10 µg of tissue extract from *UNG* wild-type kidney showing low glycosylase activity for U:G and U:A substrate. (a) is the quantification of (b).

3.4.2. Reaction time optimization

We incubated 15 µg of kidney tissue extract with different DNA substrates at different time points to find an optimal incubation time to run the rest of our samples. Glycosylases exhibit substrate preference of U:G over U:A and are also sequence dependent (63). So, we incubated the kidney extract for longer time with U:A substrate than with U:G. We found that for U:G substrate, 15 µg of kidney extract yields activity of about 40-65% for 15 minutes, 20 minutes, 25 minutes, 30 minutes and 35 minutes respectively. Since, it has been demonstrated in this experiment that excision activity by 15 µg of tissue extract for 35 minutes of incubation time is within the quantifiable range of assay, we employed the mentioned reaction conditions to run all our samples for U:G substrate.

Excision activity for U:A substrate was 30-55% for 30 minutes, 40 minutes, 50 minutes, 60 minutes and 70 minutes of incubation time, respectively (Figure 3-5). Here, it has been showed that with 70 minutes of incubation time 15 µg of extract yield glycosylase

activity within the quantifiable range. Therefore, we ran all our samples according to this mentioned reaction conditions.

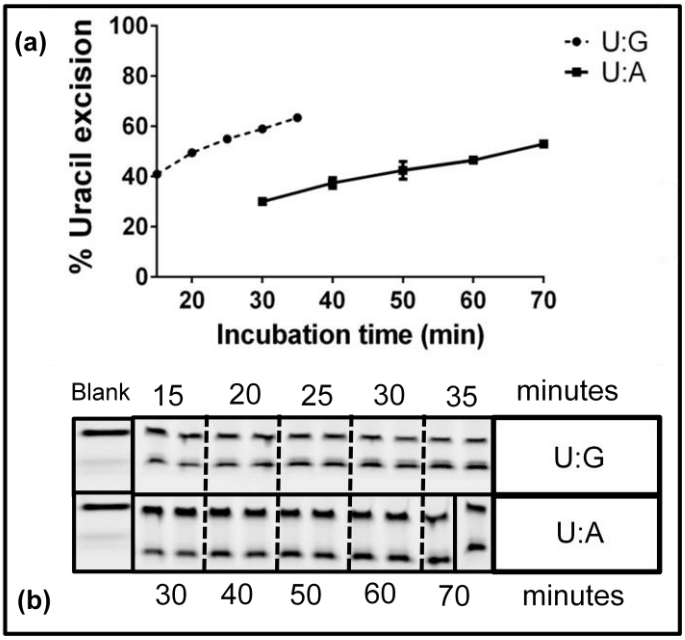

Figure 3-5: Incubation time optimization. (a) 15 µg of kidney tissue extract from *UNG* wild-type kidney shows glycosylase activity within the quantifiable range of the assay for U:G and U:A substrate. (a) is the quantification of (b).

3.4.3. Optimization of amount of inhibitors to be used

UNG2 is the most important glycosylase with SMUG1 thought to be its immediate backup in proliferating cells (54). In order to differentiate between the activities of these two enzymes, we had to inhibit one enzyme at a time to check the activity of the other. To check the UNG and SMUG1 activities, we utilized in-house polyclonal anti-human SMUG1 antibody to inhibit SMUG1 and Ugi (99) to inhibit UNG. The amounts of each of these inhibitors were optimized to ensure complete inhibition of the respective enzymes (Figure 3-6).

Both UNG and SMUG1 were inhibited with the highest amount of inhibitor added to them. Ugi which is obtained from *Bacillus subtilis* phage PBS2, is an irreversible inhibitor (102) that binds to the DNA binding groove of the catalytic domain of UNG thus restricting product binding to the enzyme. PSM1 is a polyclonal antibody that binds to specific epitopes in the active site of SMUG1 thus inactivating its catalytic activity. An increase in the concentration of inhibitors decreased the binding affinity of the enzyme to the substrate. With

increasing inhibitor concentration, more of these molecules irreversibly occupy the catalytic domain of the enzymes thereby decreasing the activity of binding substrates. We finalized the amount of inhibitors to be used for further assay: 4 µg of PSM1 and 1 µg of Ugi.

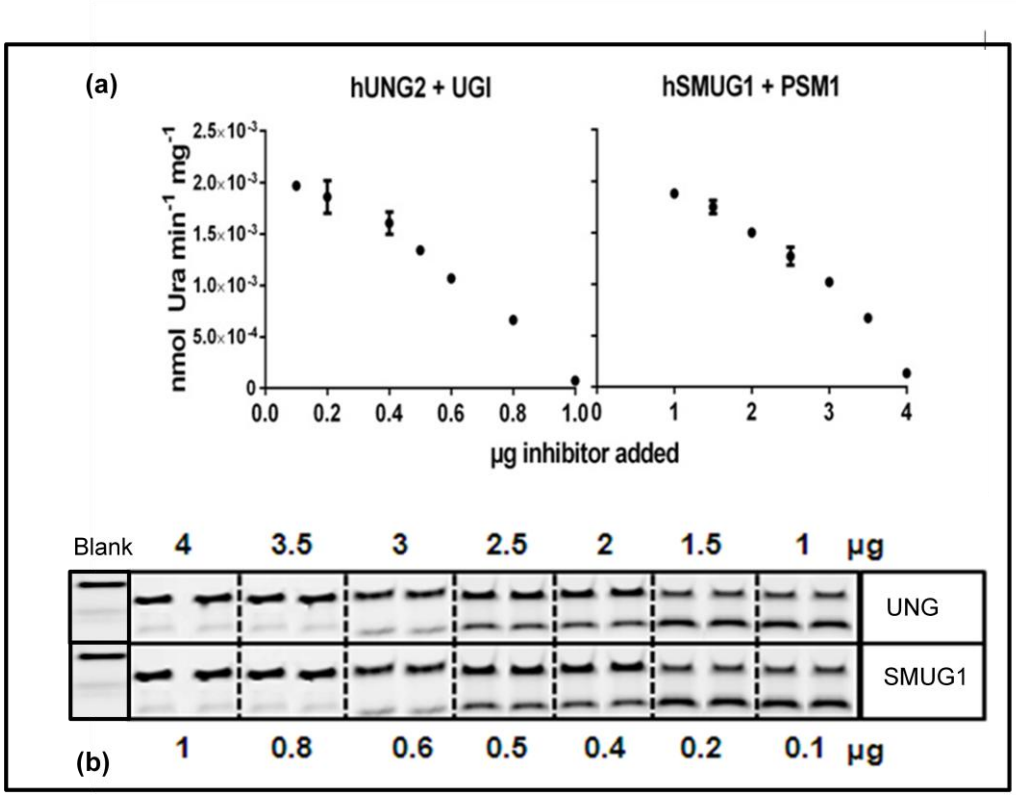

Figure 3-6: Optimization of inhibitor concentration. (A.) With increasing concentration, both the inhibitors, Ugi and PSM1 completely inhibit UNG and SMUG1 activity, respectively.

3.4.4. Glycosylase activity in mouse organs

After initial experiments to establish methods, we tested all the tissue extracts using conditions, based on the initial results, which are summarized in **Table 3-2**.

Table 3-2: Conditions used to run the oligonucleotide cleavage assay:

U:G	U:A	ss-Ura
15 µg tissue lysate	15 µg tissue lysate	5 µg tissue lysate
4 µg PSM1 and 1µg Ugi	4 µg PSM1 and 1 µg Ugi	1 µg Ugi
35 minutes incubation at 37^{o}C	70 minutes incubation at 37^{o}C	30 minutes incubation at 37^{o}C

3.4.4.1. U:G substrate

As mentioned earlier UNG is a major glycosylase in removal of uracil from DNA and at the same time shows some sequence preference of U:G over U:A in most sequence contexts. However, it is even more active towards single stranded DNA substrate (54,100). Our results support these findings.

The *UNG*$^{-/-}$ strain has previously been shown to have no detectable UNG activity and may largely be dependent on SMUG1 for uracil excision. This was, however, not demonstrated in the report on uracil DNA glycosylase activities in the initial paper, as no antibody inhibiting SMUG1 was available at that time (35). It has since been reported that SMUG1 is more abundant in mouse cell lines than human cell lines (63). Generally, we found considerable amounts of SMUG1 activity in the wild-type UNG genotype for U:G substrate. There was no clearly detectable amount of residual activity that could represent TDG or MBD4 activity in any of the wild-type and knock-out tissue extracts. These two enzymes most likely are not as important as UNG and SMUG1 in uracil excision of DNA, or have sequence preferences other than the ones used here, which may be why they could not be detected using this particular U:G substrate.

We have compared the different glycosylase activities in each of the organs. We find that UNG is generally the major uracil-DNA glycosylase with all three substrates (U:G, U:A and ss-DNA) in all organs, but SMUG1 is apparently equally important, perhaps even more important, in some organs (Figure 3-7, Figure 3-8 and Figure 3-9). It is, however, possible that the Ugi concentration was not high enough to inhibit UNG completely in all organs. Generally, uracil-excision activities were roughly equal in the different organs, except small intestine that had high activity that probably saturated the assay system. Interestingly, small intestine appeared to have very high levels of UNG, but apparently no detectable activity of SMUG1 or other uracil-DNA glycosylases, as demonstrated by the lack of detectable substrate cleavage in *UNG*$^{-/-}$ mice. When both inhibitors were present (Ugi and SMUG1 antibody PSM1) essentially all detectable uracil-DNA glycosylase activity was abolished. Below follows a more detailed description for each organ.

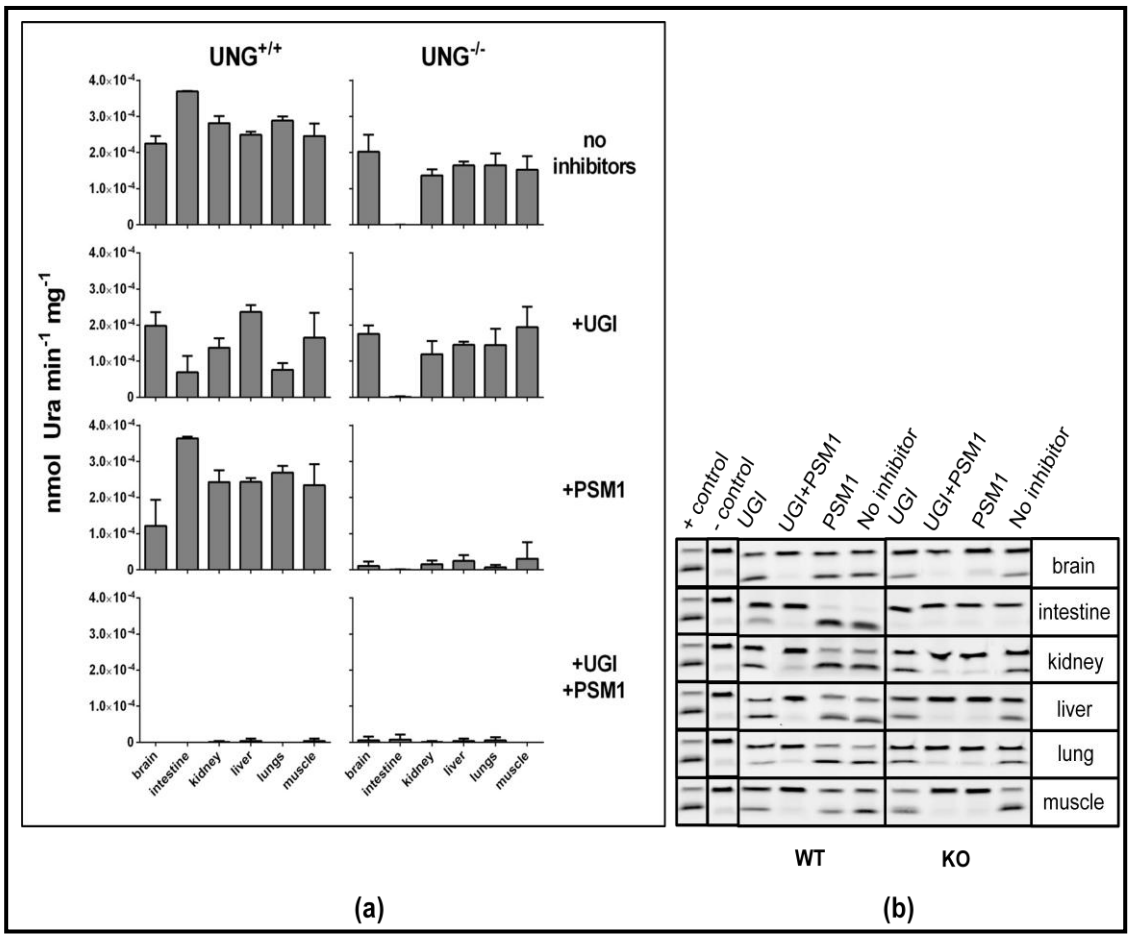

Figure 3-7: Rate of excision activity of different glycosylases, obtained from different proliferative and non-proliferative tissues of *UNG* wild-type and knock-out mice, on U:G substrate. (A.) No inhibitors represent the total glycosylase activity, samples treated with Ugi account for SMUG1 activity, PSM1 samples provide the results for UNG activity while samples treated with both inhibitors show total TDG and MBD4 activity. UNG activity is higher in intestine and lower in brain in wild-type samples. SMUG1 is abundant in brain and not found in intestine. (a) is the quantification of (b) and the error bars represent the standard deviation of three biological replicates.

i> Brain- In principle, activity in samples treated with Ugi should represent SMUG1 activity, along with TDG and MBD4 activities, while samples treated with PSM1 should show mostly UNG activity along with TDG and MBD4. However, we found no significant residual glycosylase activity that could represent TDG and MBD4 when samples were treated with both inhibitors, Ugi and PSM1. Therefore, SMUG1 and UNG are the only detectable uracil DNA glycosylases in brain and account for roughly the same amount of specific excision activity. Both the total glycosylase and SMUG1 activities of the wild-type brain were statistically indistinguishable from the knock-out brain (p= 0.49 and 0.43 respectively). The results show that SMUG1 is an important glycosylase in the brain as total

glycosylase activity in both strains accounts for almost an equal excision rate. We conclude that SMUG1 plays an important role in uracil excision in brain, possibly even more important than UNG (Figure 3-7).

ii> Small intestine- In small intestine, we found no significant residual glycosylase activity that could represent TDG and MBD4 when samples were treated with both inhibitors, Ugi and PSM1. Moreover, there was also no apparent detectable SMUG1 activity which has been indicated by lack of product band formation in $UNG^{-/-}$ mice; a statistical insignificant p value for SMUG1 activity being 0.06 when both the strains were compared. Therefore, it is possible that the resulting product band we are detecting in Figure 3-7 has developed due to residual UNG excision activity. We hypothesize that intestine being a rapidly proliferating tissue contains high amount of UNG and that the Ugi added was not sufficient to inhibit all of it. Here, we conclude that rapidly proliferating tissues mostly depend on UNG to excise uracil lesion.

iii> Kidney, liver and lung- UNG activity is quite higher than that of SMUG1 in all these three organs. In $UNG^{+/+}$ strains, the total glycosylase activity in kidney, liver and lung were statistically distinguishable from their $UNG^{-/-}$ counterparts (p= 0.0005, 0.0004 and 0.003 respectively); while the SMUG1 activity in kidney and lung of the same strains were statistically insignificant with p values 0.6, and 0.07. We found statistically significant p value of 0.001 for the SMUG1 activity in liver of both these strains. The results suggest that UNG is the important enzyme for uracil excision in these organs and that the knock-out genotype employ SMUG1 as the immediate backup for UNG (Figure 3-7).

iv> Muscle- In wild-type muscle, UNG and SMUG1 activity are almost similar. When the muscles of UNG+/+ and UNG-/- strains were compared with each other for total glycosylase activity and SMUG1 activity, the respective statistically indistinguishable p values were found to be 0.6 and 0.56. Results demonstrate that both UNG and SMUG1 are the equally important glycosylases to play roles in uracil excision (Figure 3-7).

In our result, most of the total glycosylase activity does not show synergy between the UNG and SMUG1 activity levels. This has been reported before by Doseth *et al*. This apparent paradox is supposed to be caused by competition between SMUG1 and UNG for the same

substrate (63) . SMUG1 has a high affinity for U:G substrate and a very low turnover. This prevents catalytically efficient UNG from accessing the substrate.

We encountered a large standard deviation for some of our samples, especially the lung and muscle. It is supposed that we faced handling faults during experiment by adding less amount of inhibitor to some of the tubes.

As reported earlier, glycosylase activity is highest in rapidly proliferating tissues, higher in proliferating tissues and lower in non-proliferating ones (10,49). Here, we have intestine as the rapidly dividing tissue which counts for very high levels of UNG activity that the signal reached beyond the quantifiable range of the assay. Lung, kidney and liver, all of which are also proliferating tissues, exhibit high levels of both UNG and SMUG1 activity. But in brain and muscle, we found less UNG and more SMUG1 activity indicating the latter as the important glycosylase in less proliferative tissues.

3.4.4.2. U:A substrate:

We found considerable amounts of UNG activity in the wild-type *UNG* samples for U:A substrate. There was no clearly detectable amount of residual activity that could represent TDG or MBD4 activity in any of the wild-type and knock-out tissue extracts. These enzymes most likely are not as important as UNG in uracil excision of DNA, or have sequence preferences other than the ones used here, which may be why they could not be detected using this particular U:A substrate. We found that UNG is generally the major uracil-DNA glycosylase for U:A substrate in all the organs and its activity is almost identical to the total glycosylase activity (Figure 3-8). SMUG1 has little to no affinity for this substrate in the sequence context mentioned in Table 2-3. So, it could be said that UNG is the chief enzyme in excising Ura from the U:A base-pair in DNA in the sequence context.

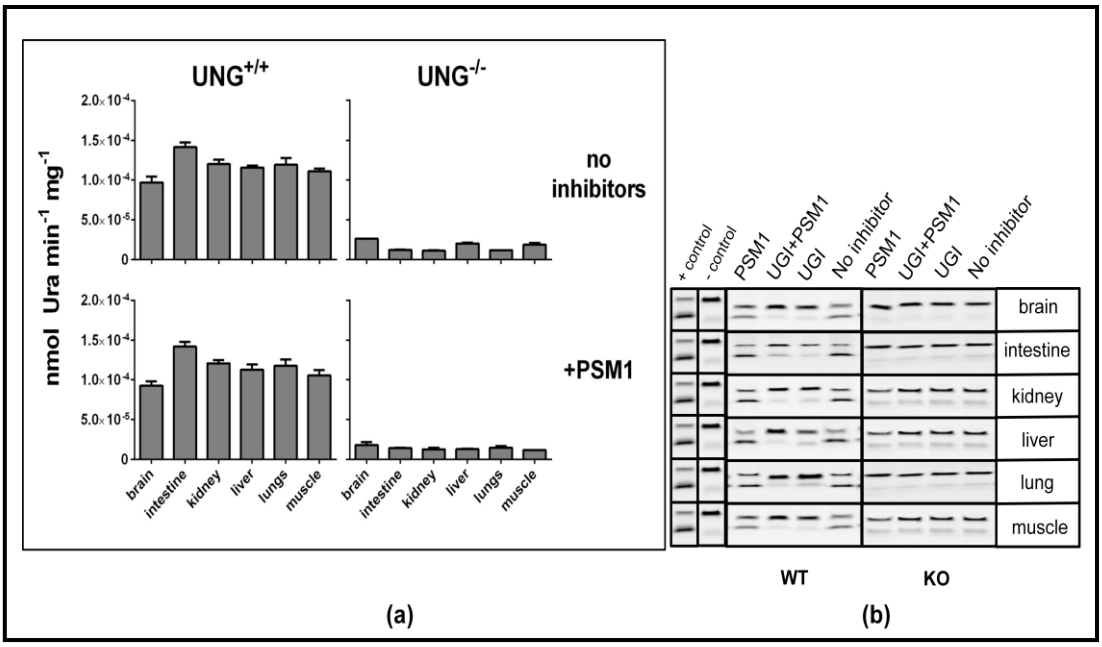

Figure 3-8: **Rate of excision activity of different glycosylases, obtained from different proliferative and non-proliferative tissues of *UNG* wild-type and knock-out mice, on U:A substrate.** (a) No inhibitors represent the total glycosylase activity while samples treated with PSM1 account for UNG activity. UNG activity is higher in intestine and lower in brain in wild-type samples. SMUG1 is abundant in brain and not found in intestine. (a) is the quantification of (b) and the error bars represent the standard deviation of three biological replicates.

The small intestine is a rapidly dividing tissue which counts for very high levels of UNG activity that signal reached beyond the quantifiable range of the assay. Lung, kidney and liver, all of which are moderately proliferating tissues, contain relatively high levels of UNG activity. But, in brain and muscle, we found less extent of uracil excision as they have relatively less capacity of proliferation than the other mentioned tissues.

^{3}H-dUMP release assay has been previously performed by Nilsen *et al.* using U:A substrate and Ugi as UNG inhibitor (35). In our experiment, we additionally introduced PSM1 to test actual UNG activity. Our results confirm that there is no detectable signal for UNG activity in the knock-out kidney and liver.

3.4.4.3. Single-stranded Ura substrate

To measure different glycosylase activity in ss-Ura substrate, we used two sets of tubes one comprising of no inhibitors and the other with Ugi to inhibit UNG activity. It has been previously reported that SMUG1 prefers ss-Ura substrate in presence of divalent cations (63). But we did not add any divalent salt to our buffer mixture. As a result, SMUG1 was

inactive on ss-Ura substrate and we did not check this by pre-incubating our samples with PSM1. TDG or MBD4 are not active on single stranded DNA. We have compared the UNG activity in each of the organs (Figure 3-9).

There was no detectable amount of activity in the Ugi treated and untreated samples from all the organs of *UNG* null mice. Therefore, it has been confirmed that SMUG1 does not act on single strands in presence of only monovalent cations (in our case Na^+). So, these mice are unable to eliminate Ura from single stranded DNA. We did not find any SMUG1 activity either for the wild type samples treated with Ugi.

We encountered variable detectable amounts of glycosylase activities in all the organs of $UNG^{+/+}$ mice. It has been previously reported that glycosylase activity ss-Ura substrate has a considerable amount of advantage over double stranded substrates (54,100). Wild type brain and muscle exhibited least amount of UNG activity than the other tested organs. These two tissues consist of non-dividing cells that likely contain less glycosylase activity in them. Likewise, the small intestine is lined with highly proliferative epithelial cells which show the highest UNG activity. The signal from this organ is in saturation and might have reached beyond the quantifiable range of the assay. The kidney, liver and lung have almost a similar range of glycosylase activity present in them. We encountered very nominal range of standard deviation within the replicates of all the organs (Figure 3-9).

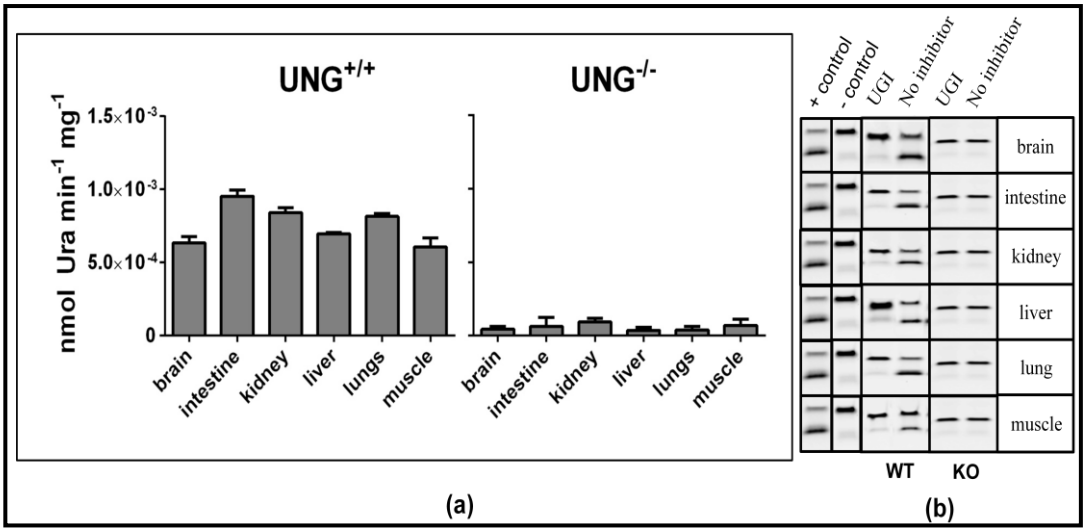

Figure 3-9: **Rate of excision activity of different glycosylases, obtained from different proliferative and non-proliferative tissues of *UNG* wild-type and knock-out mice, on ss-Ura substrate.** (a) UNG activity is higher in intestine and lower in brain for wild type

samples. (a) is the quantification of (b) and the error bars represent the standard deviation of three biological replicates.

3.4.4.4. T:G substrate

It has been previously reported that TDG exhibits preferences for T:G mismatches (67). As we have explained the results of U:G, U:A and single stranded-Ura substrates in the previous sections, we concluded that brain has the low amount of total glycosylase activity while it is highest for small intestine. We measured the TDG activity in 15 µg of tissue extracts of these two organs from wild-type mice using 20 mM of T:G mispair substrate. We selected three incubation time points for this assay: 60 minutes, 90 minutes and 120 minutes at 37^{o}C. No APE1 was added into the reaction mixture.

A weak TDG activity has been illustrated in both wild-type kidney and brain samples (Figure 3-10). The activity is below the quantifiable range for both the organs. As TDG has a low catalytic turnover, it requires APE1 to nick the abasic site that in turn helps the enzyme to dislodge from DNA. A likely explanation for this apparent paradox is that the tissue extract did not contain enough APE1 needed to displace TDG from its substrate. So, the enzymes could not access all of the provided mismatch substrates.

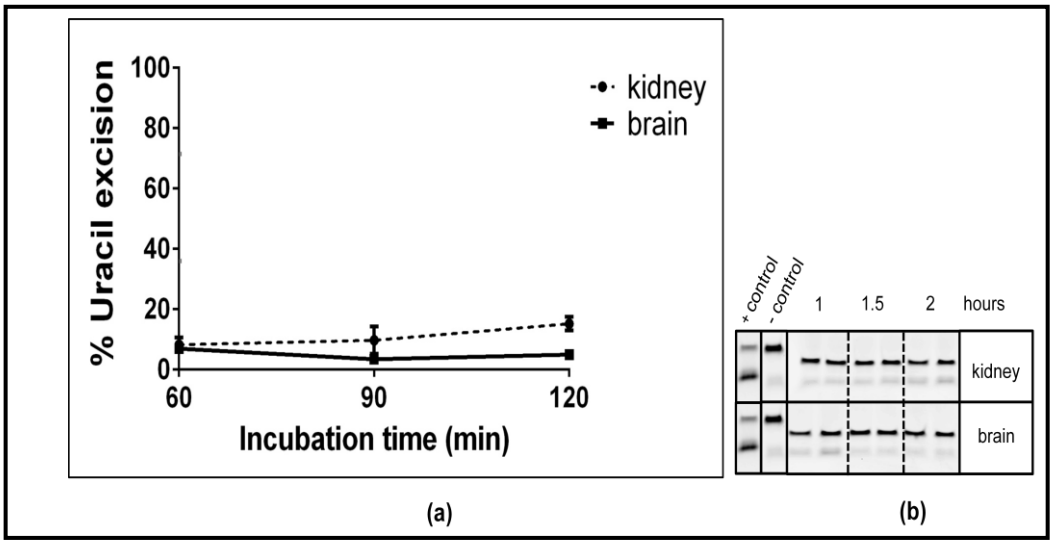

Figure 3-10: **Rate of excision activity of TDG, obtained from different proliferative and non-proliferative tissues of *UNG* wild-type, on T:G substrate.** (a) No activity was found for either of the organs. (a) is the quantification of (b) and the error bars represent the standard deviation of three biological replicates.

3.5. Quantification of genomic uracil in different organs by LC/MS/MS

3.5.1. Quantification results

dUrd measurement in various organs from *UNG* $^{+/+}$ and *UNG* $^{-/-}$ mice were obtained from Sarno A. Here, we report total dUrd per million deoxynucleotides. Referring to the graph in Figure 3-11, it is clear that there is large standard deviation for most of the organs. This was due to the loss of replicates of some organs during quantification procedure. Another probable reason for this high standard variation may be due to error that has arisen from different instruments. Two different LC/MS/MS instruments were used during the assay period. It has been postulated that *UNG* $^{-/-}$ mouse contain a considerably high amount of uracil in DNA that the *UNG* $^{+/+}$ mouse.

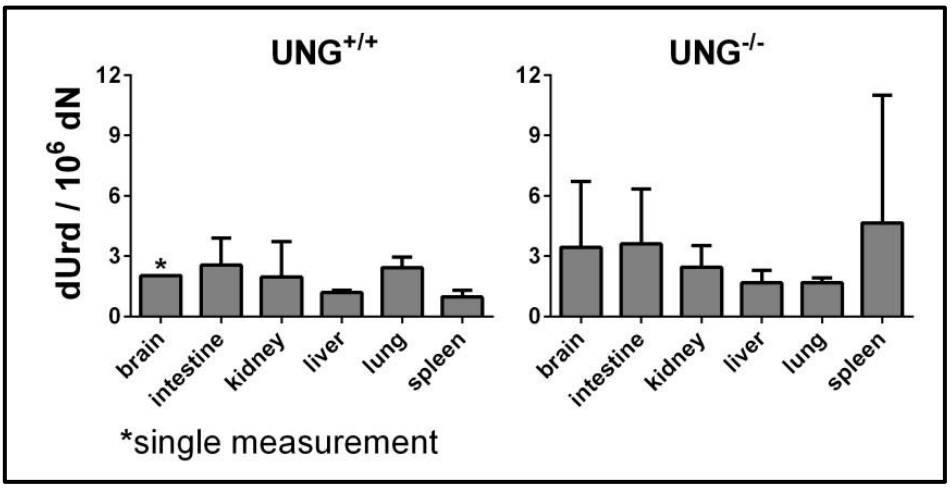

Figure 3-11: **Deoxyuridine measurement in different proliferative and non-proliferative tissues obtained from *UNG* wild-type and knock-out mice.**

3.6. Correlation between oligonucleotide cleavage assay and LC/MS/MS quantification to measure uracil in different organs of mice

We have plotted dUrd/10^6 dNs against nmoles Ura min^{-1} mg^{-1} to correlate between the glycosylase activity and the number of uracil molecules present in DNA. We observed no correlation between these two parameters. There are two main reasons that might explain our unsuccessful correlation graph (Figure 3-12):

i> There was a loss of sample during the dUrd measurement assay due to the error in LC/MS-MS. The replicates show much variation and thus prove inaccurate.

ii> There might not be any biological correlation between the dUrd levels and glycosylase activity.

a. We observed quite high levels of glycosylase activity in all the tested organs from wild-type mice. The values of dUrd are too low, so there must not be any genomic Ura stress in DNA.
b. Glycosylases are responsible for removal of Ura from DNA. But the factors that introduce Ura might be more directly related its levels in DNA. To properly correlate levels of Ura in DNA, we have to test AID, that deaminates Cyt, and test whether increases in nuclear concentration of AID increase Ura levels in DNA. The second most important factor is cellular concentration of dUTPase and TS that aid in misincorporation of Ura in DNA. This obviously depends on the dUTP/dTTP pool of the cell. To plot correlation, we need to check if the levels of these two enzymes are down-regulated that increases Ura levels in DNA.
c. Lastly, we need to test the MMR protein levels in cell. Decrease in the concentration of MMR protein could lead to elevated levels of Ura in the genome.

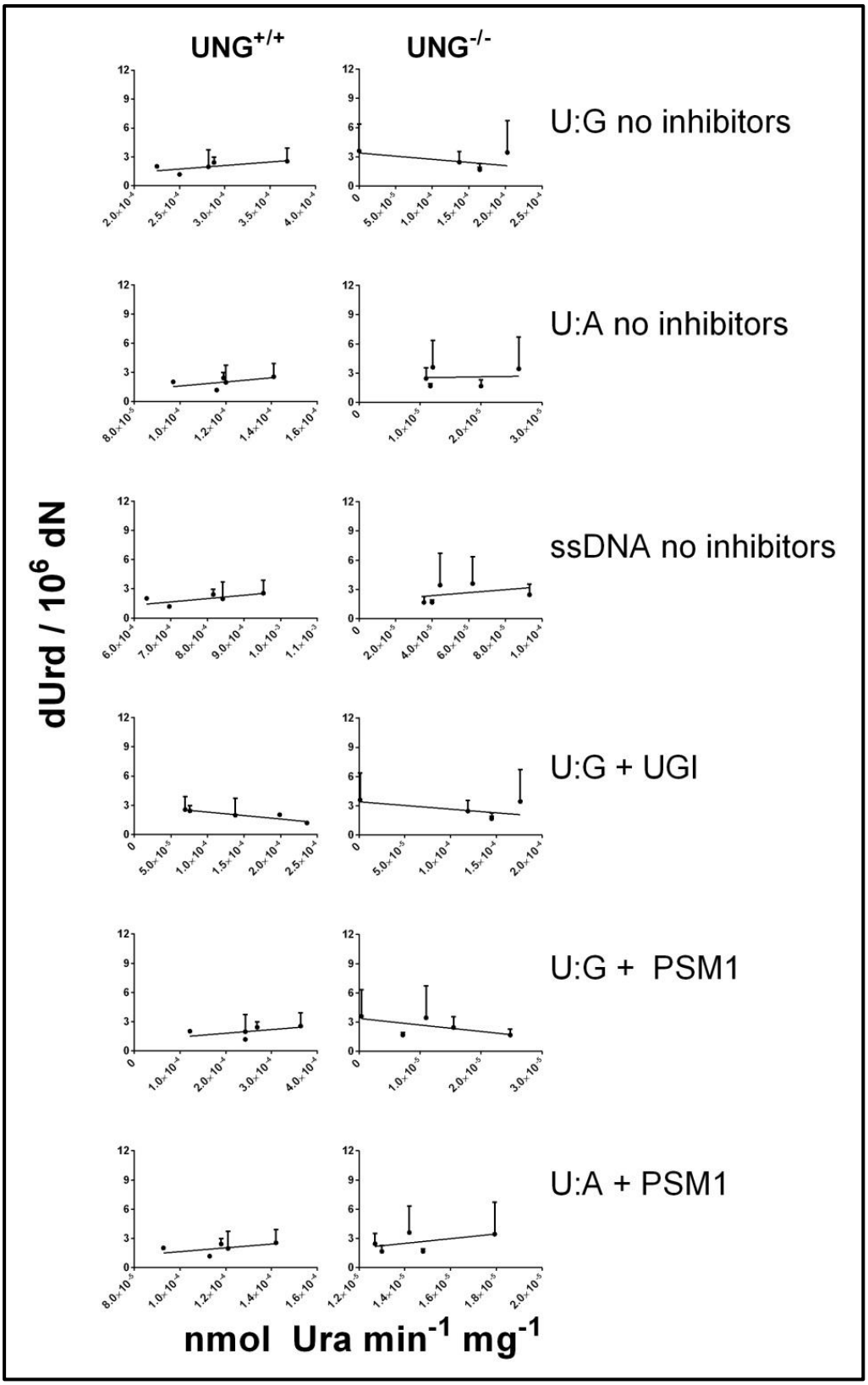

Figure 3-12: **Correlation between deoxyuridine measurement and rate of glycosylase activity in different proliferative and non-proliferative tissues obtained from *UNG* wild-type and knock-out mice.**

4. Conclusion

Activities of uracil-DNA glycosylases depend on the physiological salt concentration, distinctive preferences over various sequences and of course the nature of cell division in the tissue. The relative contribution of UNG, SMUG1, TDG and MBD4 in removal of genomic uracil from different normal mouse organs has not been investigated before. Our results reveal that SMUG1 has a very strong preference for U:G substrate and is not active on U:A or ss-Ura in the presence of optimal amount of monovalent salt (NaCl) . TDG and MBD4 do not exhibit any activity under the given reaction conditions. UNG excises ss-Ura more efficiently U:G mismatches than from U:A base pairs.

Interestingly, the efficiency of the glycosylases varies with the proliferation capacity of the tissue. UNG activity gradually increases with the rapidity of cell division in a particular tissue. Surprisingly, we found no SMUG1 activity in the intestine which has higher proliferation rate than any other tissues in our sample set. Brain, which is a non-proliferative tissue, contains higher amount of SMUG1 activity compared with UNG. Therefore, these results postulate that the extent of glycosylase activity varies between organs and that SMUG1 is more active in non-proliferative tissues than in proliferating ones. In our experiments, we have not examined activities in mitochondria and nuclei separately. UNG is known to be present in both nuclei (UNG2) and mitochondria (UNG1), whereas SMUG1 has only been found in nuclei. It is possible (but not demonstrated) that in non-proliferating tissues rich in mitochondria (e.g. muscle, heart and brain), the UNG-activity we measure could be mainly the mitochondrial form, whereas SMUG1 may be a dominant uracil-DNA glycosylase in the nuclei. This remains to be investigated.

Our efforts to examine the content of genomic uracil in different organs did not demonstrate clear evidence for a relationship to glycosylase activity. These results should be considered preliminary, since we did not have sufficient time to optimize conditions and repeat the experiments the sufficient number of times. There is however, a trend that tissues from *UNG* $^{-/-}$ mice have higher levels of dUrd.

Lastly, to measure specific glycosylase activity, we have used a particular salt condition and four distinct oligonucleotide sequences that do not represent the overall physiological pH and the whole genome, respectively. So, in order to estimate the heterogeneity of different glycosylase activities, we need to obtain more number of tissues,

and further experiment on other oligonucleotide sequences that represent small sets of the genome.

References

1. Friedberg EC, W.G., Siede W. . (1995.) DNA repair and mutagenesis. *ASM Press*.

2. James D. Watson, C.S.H.L., Tania A. Baker, M.I.o.T., Stephen P. Bell, M.I.o.T., Alexander Gann, C.S.H.L., Michael Levine, U.o.C., Berkeley and Richard Losick, H.U. (2004) *Molecular Biology of the Gene, 5/E*. 5 ed. Benjamin Cummings.

3. Sousa MM, K.H., Slupphaug G. . (2007) DNA-uracil and human pathology. *Molecular Aspects of Medicine*, **28**, 276-306.

4. Visnes, T., Doseth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G. and Krokan, H.E. (2009) Uracil in DNA and its processing by different DNA glycosylases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 563-568.

5. Mathews, C.K. (2006) DNA precursor metabolism and genomic stability. *The FASEB journal*, **20**, 1300-1314.

6. Wist, E., Unhjem, O. and Krokan, H. (1978) Accumulation of small fragments of DNA in isolated HeLa cell nuclei due to transient incorporation of dUMP. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, **520**, 253-270.

7. Mosbaugh, D.W. and Bennett, S.E. (1994) Uracil-excision DNA repair. *Progress in nucleic acid research and molecular biology*, **48**, 315.

8. Traut, T.W. (1994) Physiological concentrations of purines and pyrimidines. *Molecular and cellular biochemistry*, **140**, 1-22.

9. Békési, A., Zagyva, I., Hunyadi-Gulyás, É., Pongrácz, V., Kovári, J., Nagy, Á.O., Erdei, A., Medzihradszky, K.F. and Vértessy, B.G. (2004) Developmental regulation of dUTPase in Drosophila melanogaster. *Journal of Biological Chemistry*, **279**, 22362-22370.

10. Andersen, S., Heine, T., Sneve, R., König, I., Krokan, H.E., Epe, B. and Nilsen, H. (2005) Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. *Carcinogenesis*, **26**, 547-555.

11. Verri, A., Mazzarello, P., Biamonti, G., Spadari, S. and Focher, F. (1990) The specific binding of nuclear protein (s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C. *Nucleic acids research*, **18**, 5775-5780.

12. Voet, D. and Voet, J.G. (1995) Biochemistry, 1995. *J. Wiley & Sons, ISBN X*, **47158651**.

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

14. Goulian, M., Bleile, B. and Tseng, B. (1980) Methotrexate-induced misincorporation of uracil into DNA. *Proceedings of the National Academy of Sciences*, **77**, 1956-1960.

15. Shapiro, R. (1982), *Chromosome damage and repair*. Springer, pp. 3-18.

16. Lindahl, T. and Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13**, 3405-3410.

17. Frederico, L.A., Kunkel, T.A. and Shaw, B.R. (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*, **29**, 2532-2537.

18. Kavli, B., Otterlei, M., Slupphaug, G. and Krokan, H.E. (2007) Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA repair*, **6**, 505-516.

19. Tamir, S., Burney, S. and Tannenbaum, S.R. (1996) DNA damage by nitric oxide. *Chemical research in toxicology*, **9**, 821-827.

20. Lonkar, P. and Dedon, P.C. (2011) Reactive species and DNA damage in chronic inflammation: reconciling chemical mechanisms and biological fates. *International Journal of Cancer*, **128**, 1999-2009.

21. You, Y.-H., Lee, D.-H., Yoon, J.-H., Nakajima, S., Yasui, A. and Pfeifer, G.P. (2001) Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. *Journal of Biological Chemistry*, **276**, 44688-44694.

22. Bissonauth, V., Drouin, R., Mitchell, D.L., Rhainds, M., Claveau, J. and Rouabhia, M. (2000) The efficacy of a broad-spectrum sunscreen to protect engineered human skin from tissue and DNA damage induced by solar ultraviolet exposure. *Clinical cancer research*, **6**, 4128-4135.

23. Meador, J.A., Walter, R.B. and Mitchell, D.L. (2000) Induction, Distribution and Repair of UV Photodamage in the Platyfish, Xiphophorus signum¶. *Photochemistry and photobiology*, **72**, 260-266.

24. Tessman, I., Kennedy, M.A. and Liu, S.-K. (1994) Unusual kinetics of uracil formation in single and double-stranded DNA by deamination of cytosine in cyclobutane pyrimidine dimers. *Journal of molecular biology*, **235**, 807-812.

25. Liu, F.-T. and Yang, N. (1978) Photochemistry of cytosine derivatives. 2. Photohydration of cytosine derivatives. Proton magnetic resonance study on the chemical structure and property of photohydrates. *Biochemistry*, **17**, 4877-4885.

26. Jiang, N. and Taylor, J.S. (1993) In vivo evidence that UV-induced C. fwdarw. T mutations at dipyrimidine sites could result from the replicative bypass of cis-syn cyclobutane dimers or their deamination products. *Biochemistry*, **32**, 472-481.

27. Bransteitter, R., Pham, P., Scharff, M.D. and Goodman, M.F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences*, **100**, 4102-4107.

28. Xu, F., Mao, C., Ding, Y., Rui, C., Wu, L., Shi, A., Zhang, H., Zhang, L. and Xu, Z. (2010) Molecular and enzymatic profiles of mammalian DNA methyltransferases: structures and targets for drugs. *Current medicinal chemistry*, **17**, 4052.

29. Shen, J.-C., Rideout III, W.M. and Jones, P.A. (1992) High frequency mutagenesis by a DNA methyltransferase. *Cell*, **71**, 1073-1080.

30. Pham, P., Bransteitter, R. and Goodman, M.F. (2005) Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry*, **44**, 2703-2715.

31. Maul, R.W., Saribasak, H., Martomo, S.A., McClure, R.L., Yang, W., Vaisman, A., Gramlich, H.S., Schatz, D.G., Woodgate, R. and Wilson III, D.M. (2011) Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nature immunology*, **12**, 70-76.

32. Olinski R, J.M., Zaremba T. . (2010) Uracil in DNA-Its biological significance. *Mutat Res* **705**, 239-245.

33. Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y. and Blau, H.M. (2009) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, **463**, 1042-1047.

34. Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E. and Reik, W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, **463**, 1101-1105.

35. 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. *Molecular cell*, **5**, 1059-1065.

36. Visnes, T., Akbari, M., Hagen, L., Slupphaug, G. and Krokan, H.E. (2008) The rate of base excision repair of uracil is controlled by the initiating glycosylase. *DNA repair*, **7**, 1869-1881.

37. Nelson, D.L., Lehninger, A.L. and Cox, M.M. (2008) *Lehninger principles of biochemistry*. Macmillan.

38. Galashevskaya, A., Sarno, A., Vågbø, C.B., Aas, P.A., Hagen, L., Slupphaug, G. and Krokan, H.E. (2013) A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. *DNA repair*, **12**, 699-706.

39. Lindahl, T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proceedings of the National Academy of Sciences*, **71**, 3649-3653.

40. Meira, L.B., Burgis, N.E. and Samson, L.D. (2005), *Genome Instability in Cancer Development*. Springer, pp. 125-173.

41. Krokan, H.E. and Bjørås, M. (2013) Base excision repair. *Cold Spring Harbor perspectives in biology*, **5**, a012583.

42. Holohan, K.N., Lahiri, D.K., Schneider, B.P., Foroud, T. and Saykin, A.J. (2011) Functional microRNAs in Alzheimer's disease and cancer: differential regulation of common mechanisms and pathways. *Frontiers in genetics*, **3**, 323-323.

43. Krokan, H.E., Sætrom, P., Aas, P.A., Pettersen, H.S., Kavli, B. and Slupphaug, G. (2014) Error-free< i> versus</i> mutagenic processing of genomic uracil—Relevance to cancer. *DNA repair*.

44. Gao, Y., Katyal, S., Lee, Y., Zhao, J., Rehg, J.E., Russell, H.R. and McKinnon, P.J. (2011) DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature*, **471**, 240-244.

45. Simsek, D., Furda, A., Gao, Y., Artus, J., Brunet, E., Hadjantonakis, A.-K., Van Houten, B., Shuman, S., McKinnon, P.J. and Jasin, M. (2011) Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. *Nature*, **471**, 245-248.

46. Sung, J.S. and Demple, B. (2006) Roles of base excision repair subpathways in correcting oxidized abasic sites in DNA. *Febs Journal*, **273**, 1620-1629.

47. Fortini, P. and Dogliotti, E. (2007) Base damage and single-strand break repair: mechanisms and functional significance of short-and long-patch repair subpathways. *DNA repair*, **6**, 398-409.

48. Dou, H., Theriot, C.A., Das, A., Hegde, M.L., Matsumoto, Y., Boldogh, I., Hazra, T.K., Bhakat, K.K. and Mitra, S. (2008) Interaction of the Human DNA Glycosylase NEIL1 with Proliferating Cell Nuclear Antigen THE POTENTIAL FOR REPLICATION-ASSOCIATED REPAIR OF OXIDIZED BASES IN MAMMALIAN GENOMES. *Journal of Biological Chemistry*, **283**, 3130-3140.

49. Akbari, M., Peña-Diaz, J., Andersen, S., Liabakk, N.-B., Otterlei, M. and Krokan, H.E. (2009) Extracts of proliferating and non-proliferating human cells display different base excision pathways and repair fidelity. *DNA repair*, **8**, 834-843.

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

51. Dingler, F.A., Kemmerich, K., Neuberger, M.S. and Rada, C. (2014) Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A: T substitutions during somatic mutation. *European journal of immunology*.

52. 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.

53. 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 research*, **25**, 750-755.

54. 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. *Journal of Biological Chemistry*, **277**, 39926-39936.

55. Nilsen, H., Yazdankhah, S.P., Eftedal, I. and Krokan, H.E. (1995) Sequence specificity for removal of uracil from U· A pairs and U· G mismatches by uracil-DNA glycosylase from< i> Escherichia coli</i>, and correlation with mutational hotspots. *FEBS letters*, **362**, 205-209.

56. Krokan, H., STANDAL, R. and SLUPPHAUG, G. (1997) DNA glycosylases in the base excision repair of DNA. *Biochem. J*, **325**, 1-16.

57. Visnes, T., Doseth, B., Pettersen, H.S., Hagen, L., Sousa, M.M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G. and Krokan, H.E. (2009) Uracil in DNA and its processing by different DNA glycosylases. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **364**, 563-568.

58. Haug, T., Skorpen, F., Aas, P.A., Malm, V., Skjelbred, C. and Krokan, H.E. (1998) Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase. *Nucleic acids research*, **26**, 1449-1457.

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

60. Kavli, B., Andersen, S., Otterlei, M., Liabakk, N.B., Imai, K., Fischer, A., Durandy, A., Krokan, H.E. and Slupphaug, G. (2005) B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *The Journal of experimental medicine*, **201**, 2011-2021.

61. 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 research*, **35**, 3879-3892.

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

63. Doseth, B., Ekre, C., Slupphaug, G., Krokan, H.E. and Kavli, B. (2012) Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil. *DNA repair*, **11**, 587-593.

64. 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.

65. Boorstein, R.J., Cummings, A., Marenstein, D.R., Chan, M.K., Ma, Y., Neubert, T.A., Brown, S.M. and Teebor, G.W. (2001) Definitive Identification of Mammalian 5-Hydroxymethyluracil DNAN-Glycosylase Activity as SMUG1. *Journal of Biological Chemistry*, **276**, 41991-41997.

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

67. Neddermann, P. and Jiricny, J. (1994) Efficient removal of uracil from GU mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proceedings of the National Academy of Sciences*, **91**, 1642-1646.

68. Hardeland, U., Steinacher, R., Jiricny, J. and Schär, P. (2002) Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *The EMBO journal*, **21**, 1456-1464.

69. Hardeland, U., Bentele, M., Jiricny, J. and Schär, P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *Journal of Biological Chemistry*, **275**, 33449-33456.

70. Hardeland, U., Kunz, C., Focke, F., Szadkowski, M. and Schär, P. (2007) Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic acids research*, **35**, 3859-3867.

71. Hendrich, B., Hardeland, U., Ng, H.-H., Jiricny, J. and Bird, A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, **401**, 301-304.

72. Krokan, H.E., Kavli, B. and Slupphaug, G. (2004) Novel aspects of macromolecular repair and relationship to human disease. *Journal of molecular medicine*, **82**, 280-297.

73. Begum, N.A. and Honjo, T. (2012) Evolutionary comparison of the mechanism of DNA cleavage with respect to immune diversity and genomic instability. *Biochemistry*, **51**, 5243-5256.

74. 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. *Current Biology*, **12**, 1748-1755.

75. Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M. and Pilch, D.R. (2001) AID is

required to initiate Nbs1/γ-H2AX focus formation and mutations at sites of class switching. *Nature*, **414**, 660-665.

76. Guikema, J.E., Linehan, E.K., Tsuchimoto, D., Nakabeppu, Y., Strauss, P.R., Stavnezer, J. and Schrader, C.E. (2007) APE1-and APE2-dependent DNA breaks in immunoglobulin class switch recombination. *The Journal of experimental medicine*, **204**, 3017-3026.

77. Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E. and Alt, F.W. (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*, **422**, 726-730.

78. Chaudhuri, J. and Alt, F.W. (2004) Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature reviews. Immunology*, **4**, 541-552.

79. Mechtcheriakova, D., Svoboda, M., Meshcheryakova, A. and Jensen-Jarolim, E. (2012) Activation-induced cytidine deaminase (AID) linking immunity, chronic inflammation, and cancer. *Cancer immunology, immunotherapy : CII*, **61**, 1591-1598.

80. Küppers, R. (2005) Mechanisms of B-cell lymphoma pathogenesis. *Nature Reviews Cancer*, **5**, 251-262.

81. Kotani, A., Kakazu, N., Tsuruyama, T., Okazaki, I.-m., Muramatsu, M., Kinoshita, K., Nagaoka, H., Yabe, D. and Honjo, T. (2007) Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proceedings of the National Academy of Sciences*, **104**, 1616-1620.

82. Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.-H., Robbiani, D.F., Di Virgilio, M., San-Martin, B.R., Heidkamp, G. and Schwickert, T.A. (2008) MicroRNA-155 Suppresses Activation-Induced Cytidine Deaminase-Mediated< i> Myc-Igh</i> Translocation. *Immunity*, **28**, 630-638.

83. Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H.C., Nussenzweig, M.C. and Dalla-Favera, R. (2007) AID is required for germinal center–derived lymphomagenesis. *Nature genetics*, **40**, 108-112.

84. Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.-M., Honjo, T. and Chiba, T. (2007) Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nature medicine*, **13**, 470-476.

85. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A. and Børresen-Dale, A.-L. (2013) Signatures of mutational processes in human cancer. *Nature*.

86. Lada, A.G., Dhar, A., Boissy, R.J., Hirano, M., Rubel, A.A., Rogozin, I.B. and Pavlov, Y.I. (2012) AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biol Direct*, **7**, 47.

87. Andersen, S., Ericsson, M., Dai, H.Y., Peña-Diaz, J., Slupphaug, G., Nilsen, H., Aarset, H. and Krokan, H.E. (2005) Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice. *DNA repair*, **4**, 1432-1441.

88. Durandy, A., Revy, P., Imai, K. and Fischer, A. (2005) Hyper-immunoglobulin M syndromes caused by intrinsic B-lymphocyte defects. *Immunological reviews*, **203**, 67-79.

89. Kracker, S., Gardes, P., Mazerolles, F. and Durandy, A. (2010) Immunoglobulin class switch recombination deficiencies. *Clinical Immunology*, **135**, 193-203.

90. Kohn, K.W., Erickson, L.C., Ewig, R.A. and Friedman, C.A. (1976) Fractionation of DNA from mammalian cells by alkaline elution. *Biochemistry*, **15**, 4629-4637.

91. Horváth, A. and Vértessy, B.G. (2010) A one-step method for quantitative determination of uracil in DNA by real-time PCR. *Nucleic acids research*, **38**, e196-e196.

92. Atamna, H., Cheung, I. and Ames, B.N. (2000) A method for detecting abasic sites in living cells: age-dependent changes in base excision repair. *Proceedings of the National Academy of Sciences*, **97**, 686-691.

93. Mashiyama, S.T., Courtemanche, C., Elson-Schwab, I., Crott, J., Lee, B.L., Ong, C.N., Fenech, M. and Ames, B.N. (2004) Uracil in DNA, determined by an improved assay, is increased when deoxynucleosides are added to folate-deficient cultured human lymphocytes. *Analytical biochemistry*, **330**, 58-69.

94. Ren, J., Ulvik, A., Refsum, H. and Ueland, P.M. (2002) Uracil in human DNA from subjects with normal and impaired folate status as determined by high-performance liquid chromatography-tandem mass spectrometry. *Analytical chemistry*, **74**, 295-299.

95. Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N. and Nikas, J.B. (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, **494**, 366-370.

96. Chango, A., Abdel Nour, A., Niquet, C. and Tessier, F.J. (2009) Simultaneous determination of genomic DNA methylation and uracil misincorporation. *Medical Principles and Practice*, **18**, 81-84.

97. 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. *Chemical research in toxicology*, **16**, 1044-1055.

98. Silverstein, R. (1998). John Wiley & Son: New York.

99. 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.

100. Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N.M., Haug, T., Levine, D.W. and Krokan, H.E. (1995) Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry*, **34**, 128-138.

101. Ko, R. and Bennett, S.E. (2005) Physical and functional interaction of human nuclear uracil-DNA glycosylase with proliferating cell nuclear antigen. *DNA repair*, **4**, 1421-1431.

102. Bennett, S.E., Schimerlik, M. and Mosbaugh, D. (1993) Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *Journal of Biological Chemistry*, **268**, 26879-26885.