AID, APOBEC3A and APOBEC3B efficiently deaminate deoxycytidines neighboring DNA damage induced by oxidation or alkylation

By: Cody Diamond

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements

for the degree of Master of Science in Medicine (Immunology and Infectious Diseases)

Division of Biomedical Sciences, Faculty of Medicine

Memorial University of Newfoundland

February 2021

St. John's, Newfoundland and Labrador

Lay Summary:

Processes inside the human body are driven by cellular machines called enzymes. Sometimes, these enzymes are involved in changing genetic material to fulfill a purpose. This paper discusses enzymes involved in changing the genetic material of humans and viruses. These enzymes have recently been found to target genetic material in cancerous tissue. The genetic material of cancerous tissue has unique differences when compared to healthy genetic material. Specifically, cancerous tissue tends to create its own environment in the body that promotes change in the genetic material. In some cases, this can mean the genetic material is more likely to be damaged. Given this fact, the authors of this thesis wanted to know if, the enzymes discussed within, acted differently on damaged genetic material when compared to how they act on healthy genetic material. We found that the enzymes were just as affective at acting on the damaged genetic material as on healthy genetic material, and that in some cases they were better.

Abstract:

AID/APOBEC3s are endogenous enzymes that deaminate deoxycytidine (dC) to deoxyuridine (dU). Although they can mutate any dC, each family member has a unique sequence preference determined by nucleotides immediately surrounding the target dC. This WRC (W=A/T, R=A/G) and YC (Y=T/C) preference is well established for AID and APOBEC3A/B respectively. Common sources of DNA damage such as alkylation and oxidation alter the chemistry of normal base side chains prompting us to examine the activity of AID, APOBEC3A, and APOBEC3B on dCs whose neighboring -2 or -1 position bases are damaged through oxidation or alkylation. We found that all three enzymes efficiently deaminate dC when common damaged bases are present in the -2 and -1 positions. Strikingly, some of these motifs supported even higher catalytic efficiencies by AID, APOBEC3A and APOBEC3B compared to WRC/YC motifs which are their most favored normal DNA sequences. Based on interactions of the surfaces of AID, APOBEC3A and APOBEC3B with DNA as resolved by structural studies, we then modeled interactions with alkylated or oxidized bases. Corroborating the enzyme kinetics experiments, these analyses suggest that the surface regions responsible for recognition of normal bases can also interact with oxidized and alkylated bases. Thus, our data extend the current knowledge of the -2 and -1 position substrate sequence specificity of AID/A3s to include common damaged base motifs.

<u>Acknowledgements:</u>

Contributions:

The contributions to this publication by author are as follows:

Cody Diamond purified and tested experiment specific preparations of AID for this project. He was the first student in the Larijani lab to successfully purify and work with the GST-tagged deaminases APOBEC3A and APOBEC3B. He purified these enzymes from eukaryotic cells using a novel intron-containing plasmid system which he assisted in optimizing (Figure 3). Cody Diamond optimized the standard alkaline cleavage assay to accurately test the catalytic efficiency of the newly purified enzymes A3A and A3B. He was first one to test A3A and A3B on damaged DNA substrates. Cody Diamond designed specific bubble and single-stranded damaged DNA oligonucleotides for testing the activity of AID and the newly purified enzymes on oxidized and alkylated DNA (Figure 4). Cody Diamond and Junbum Im designed and carried out all enzyme experiments and wrote and edited the manuscript with minor contributions from the other authors (Figures 5, 6, 7, 9 and Table 1). Cody Diamond carried out the uracil control experiment and designed the uracil containing oligonucleotides (Figure 8). Erynn A. Button assisted with the enzyme assays. Cody Diamond created and edited figures 1-9 in the final thesis, they are original figures. David N. Huebert and Justin J. King designed and performed the enzyme:substrate docking structural analysis, and created the figures containing the data therein. (Figures 10 and 11). Faeze Borzooee provided technical assistance for the enzyme experiments,

Table of Contents:

List of Figures:

List of Tables:

Abbreviations and Symbols:

AID – Activation-induced cytidine deaminase

APOBEC – Apolipoprotein B mRNA-editing catalytic polypeptide-like

A3 – APOBEC3

BER – Base Excision Repair

DDR – DNA Damage Response

DLR – Direct Lesion Reversal

DNA – Deoxyribonucleic Acid

DSB – Double-Stranded Break

GST – Glutathione S-Transferase

NER – Nucleotide Excision Repair

NHEJ – Non-Homologous End Joining

RNA- Ribonucleic Acid

ROS – Reactive Oxygen Species

SSB – Single-Stranded Break

SSBR – Single-Stranded Break Repair

Y-H2AX – Histone Family Member X

1. Introduction

1.1 AID and the APOBEC family of cytidine deaminases: DNA-editing enzymes mediating diverse biological functions

AID/APOBECs are cytidine deaminases that convert deoxycytidine (dC) to deoxyuridine (dU) [1, 2]. The AID/APOBEC family of proteins represents a group that fulfills diverse physiological functions. The protein family comprises eleven members in humans: AID and APOBEC1 (genes located on chromosome 12), APOBEC2 (gene located on chromosome 6), seven APOBEC3 proteins (APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H; genes located on chromosome 22) and APOBEC4 (gene located on chromosome 1) . AID and APOBEC2 are considered the ancestral members of the AID/APOBEC family, whereas APOBEC1 and APOBEC3 evolved later and are similar to AID [1-20]. APOBEC3 proteins are detectable exclusively in mammals [3]. All members of the AID/APOBEC family share at least one zinc-binding catalytic domain with the consensus amino acid sequence $H\text{-}X\text{-}E\text{-}X_{23-28}\text{-}P\text{-}C\text{-}X_{2-4}\text{-}C$ (X stands for any amino acid) [3]. APOBEC1 was the first to be isolated and is active in mutating C's to U's in mRNA for apolipoprotein B. This results in Apolipoprotein 48 which is involved in lipid uptake from the intestine [3]. APOBEC2 was the next to be discovered. APOBEC2 is expressed in skeletal muscle and the heart but currently has never been shown to be mutagenic in yeast or bacterial assays. Its exact physiological role has yet to be identified, however its expression is crucial for muscle development [1-4]. APOBEC4 is expressed in the testicles. It is thus far inactive in yeast and bacterial cell cultures and its physiological function is not yet known (Figure 1.1) [3].
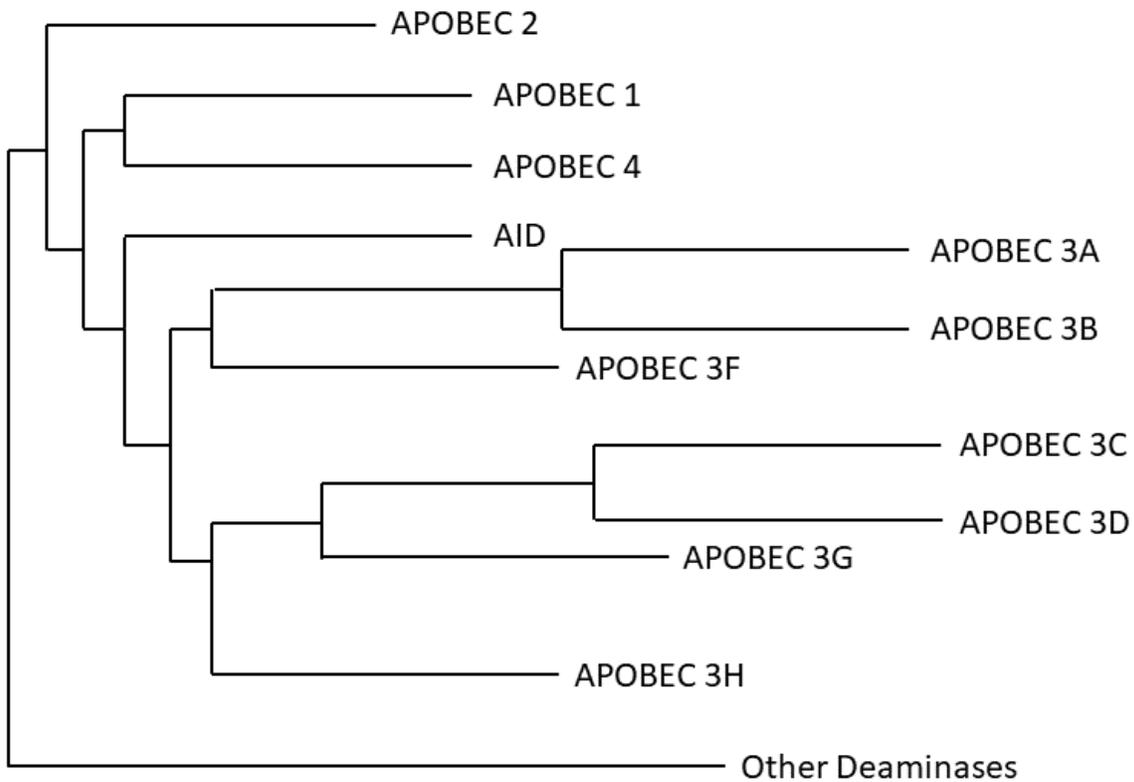
Figure 1.1. The AID/APOBEC family of enzymes. Enzymatic family tree is derived based on comparison and similarities between protein structure of the main catalytic subunit between each enzyme in the family. In the case of double-domain enzymes, comparison is made between the N-terminal domain and single catalytic subunits of each other double-domain enzyme or single-domain enzyme in the family respectively. Structural similarity garnered from information found in Conticello et al. 2005 [11].

The APOBEC3 sub-branch members, including APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3F (A3F), APOBEC3G (A3G) and APOBEC3H (A3H) function as cellular anti-viral machinery by acting on retroviral genomes and mobile endogenous retroelements [5-12]. The APOBEC3 family members can be further classified based on number of zinc-binding domains. A3A, A3C and A3H have single zinc-binding domains. In contrast, A3B, A3D, A3G and A3F have two zinc-binding domains [5-12]. Their specific

physiological functions differ between each sub-branch member. Indeed, as this group of enzymes has only been discovered in the last 18 years, there are still those whose functions are unknown. However, they are all functional deaminases, are widely expressed in multiple tissue types and are associated with antiviral activity [7,8,13].

APOBEC3A acts mostly on ssDNA and seems to be involved in cell-cycle arrest. Its exact cellular role is unknown, but in one experiment, caused clearing via deamination of dC's in 97% of foreign bacterial DNA [8]. Notably, A3A and A3B are extremely similar and share ~95% amino acid sequence [1-3]. They have both been shown to deaminate the lagging-strand in DNA replication bubbles and thus may also be involved in signalling during replication [5]. Both, along with the other A3 family members are involved in cellular defense against many viruses such as Hepatitis B virus, Hepatitis C virus, human papillomavirus, human herpesviruses and T-Cell leukemia virus type-1 [9]. However, A3A/B are the members of the A3 family whose mutational signatures have been more prevalent in human cancer genomes than their A3 family members and thus are of particular interest which will be discussed in more detail later. A3D/F/G/H have been found to more specifically target retroviral genomes and endogenous transposons, specifically Human-Immunodeficiency Virus. APOBEC3G was found to be the most specific endogenous defense factor against Vif-deficient HIV virus. A3G is packaged into HIV virions and mutates the viral genome within, causing the genome to be mis-read and create ineffective translational products [7-9].

AID is primarily expressed in activated B cells and mutates the immunoglobulin (Ig) loci, initiating somatic hypermutation and class switch recombination of antibodies [3, 4]. AID deaminates dC thereby generating dU within the immunoglobulin (Ig) locus [3-8]. Since uracil base pairs with adenine, unrepaired uracils lead to dC>dT transition mutations after DNA

replication, whereas error prone DNA repair of uracil lesions result in mutations at dC:dG as well as dA:dT base pairs. The repair of deaminated cytosines within switch regions, 5′ of the constant regions of antibody genes, leads to the generation of Double-Stranded Breaks (DSBs), which are a prerequisite for class switch recombination [3-8]. Hence, deficiency of functional AID in humans due to homozygous deletions or mutations causes the autosomal recessive form of an immunodeficiency syndrome, hyper-IgM syndrome. Disruption of AID in the chicken B cell lymphoma cell line DT40 abolishes antibody diversification by gene conversion. First studies suggested that the enzyme edits RNA intermediates, before it could be shown *in vitro* that AID deaminates ssDNA during transcription, preferentially the dC within WRC motifs [3-8].

It is well established that off-target AID-induced deamination events are involved in tumorigenesis and cancer genome mutagenesis [13-21]. More recently, it has come to light that expression and mutagenesis footprints of A3A and A3B are also prevalent in tumors arising from a wide range of tissues including bladder, breast, cervix and thyroid cancer, lung adenocarcinoma, B cell lymphomas, multiple myelomas, acute lymphoblastic leukemia and chronic lymphocytic leukemia [7,8,13-34].

Mutations are a basis for evolution and the development of genetic diseases [7,8]. Especially in cancer, somatic mutations in oncogenes and tumor suppressor genes alongside the occurrence of passenger mutations have been observed by recent deep-sequencing approaches. Mutations have long been considered random events induced by DNA-replication errors or by DNA damaging agents [1-5, 7-9]. However, the advent of genome sequencing led to the discovery of non-random mutation signatures in many human cancers. Common non-random mutations comprise DNA strand-biased mutations and mutations restricted to certain DNA motifs, which

4

recently have become attributed to the activity of the AID/APOBEC family of DNA deaminases. Hence, AID/APOBEC enzymes, which have evolved as key players in natural and adaptive immunity, have been proposed to contribute to cancer development and clonal evolution of cancer by inducing adjunct genomic damage due to their DNA deaminating activity [1-15].

AID/A3s act on single-stranded DNA (ssDNA) substrates. They are loosely targeted throughout the genome and are understood to mutate readily accessible and unpaired ssDNA regions, such as those present during transcription and/or replication [35-47]. Once targeted to a stretch of the genome by these ssDNA-generating cellular mechanisms, AID/A3s can mutate any given dC; however, each enzyme exhibits a local target sequence preference based largely on the -2 and -1 nucleotides immediately upstream of the target dC. AID favors WR<u>C</u> (W=A/T, R=A/G) trinucleotides in *in vitro* enzyme assays, when exogenously expressed in bacteria or cell lines, as well as *in vivo* [42, 48-55]. In contrast, A3A/B/C/D/F favor dC in the Y<u>C</u> dinucleotide context [46, 56-61]. A3G and A3H have a more specific preference for CC<u>C</u> trinucleotide sequences, mutating the third cytosine in the trinucleotide [3-8]. Since these signature "hotspots" are unique, they are relied upon to infer the mutational footprints of particular member(s) of the AID/A3 family in the context of sequenced viral genomes and tumor cell genomes. Notably, these substrates have been studied on a variety of substrates *in vivo* such as small ssDNA fragments, "bubble" substrates whereby some of the DNA contains mismatched base pairs to create a bubble of ssDNA within an annealed dsDNA oligonucleotide, and complete sequences of DNA vectors within living prokaryotic and eukaryotic cultures [26,28-29,31,37,40,46]. However, the activity of these enzymes on each of these substrates have never been compared to one another in a formal holistic study to date.

5

Structurally, the core catalytic pocket is relatively conserved amongst AID/A3 enzymes [1, 56, 62-66]. Their differential sequence specificity is mediated by a substrate specificity loop (L7: the β4-α4 connecting loop) that is highly divergent amongst AID/A3 members, and if swapped can transfer the sequence specificity of the donor [67]. In AID, a minor contribution has also been shown for L1 (the α1-β1 connecting loop) [54, 67, 68]. Differences in the length and composition of this substrate specificity loop yield varying surface topologies proximal to the catalytic pocket, which underlies differing preferences for surface interactions with pyrimidine or purine bases at the -2 and -1 positions relative to the target dC [46, 55-57, 61, 65].

In contrast to the enzyme structural basis of sequence specificity in AID/A3s, this issue has also been investigated from the nucleic acid structure perspective. One approach has been probing AID/A3 activity on substrates containing various sequence arrangements including the four canonical DNA bases or naturally occurring analogues. In regards to the chemistry of the -2 and -1 bases, this approach has revealed that AID/A3s can recognize subtle differences in the pyrimidine or purine side chains of the -1 and -2 nucleotide positions relative to dC [42, 56]. In contrast to limited information on the important chemical aspects of the -2 and -1 position bases for AID/A3 recognition, chemical determinants of the target dC itself have been more rigorously investigated. For instance, it has been suggested that some AID/A3 enzymes can function in genome demethylation through deamination of 5-methyl-cytidine(mC) [69-75]. We and others compared AID/A3 activities on substrates containing normal dC, mC, or deoxycytidines with other bulky adducts resulting from oxidation or methylation intermediates. It was found that AID has minimal activity on mC, roughly an order of magnitude weaker compared to dC. A3A and A3B also deaminate mC with less efficiency than dC. A3H

deaminates mC with equal efficiency to the native dC or rC. Overall, A3 activity on mC is significantly more robust compared to that of human AID [76-87]. Due to their function in deaminating methylated cytidines, it is postulated that AID/A3 enzymes are highly involved in gene expression. AID and several of the A3 enzymes such as A3B are localized to the nucleus of the cell, thus giving them greater potential access to methylated genes [76-78]. Further research is required to fully understand the epigenetic role of these enzymes.

1.2 DNA damage and its role in the tumor micro-environment

In addition to off-target and deregulated deamination events, genomes of both somatic and cancer cells are also susceptible to environmental and chemical sources of DNA damage, such as reactive oxygen species (ROS), components of tobacco smoke, or chemotherapeutic drugs. Many of these agents instigate genome damage through alkylating or oxidizing normal DNA bases, thus altering their side chain chemistries [88-96]. Although repaired efficiently in healthy cells, oxidized and alkylated bases are often found in high abundance in cancer genomes due to deficiencies in their associated DNA repair pathways and being induced through chemotherapy treatment [93, 94, 97-115].

As eluded to previously, DNA damage can take the form of oxidation, alkylation, radiation, or deamination. Each type of damage alters the structure of DNA which can cause read-through errors by DNA polymerases and RNA polymerases [90, 92-101]. Some structural changes that may occur are more harmful then others and can range from causing single-base mutations, such as in the case of an 8'-oxo-deoxygaunosine (8oxoG) //Adenosine Hoogsteen base-pair, to the complete stalling of genome replication such as in the case of 1-methyl-adenosine [103-105]. Oxidation occurs when a cellular environment is saturated with ROS. The oxidative free radicals react with DNA nucleotides to create oxygenated derivatives such as the oxygenated

8' carbon in 8oxoG [105-107, 112]. Alkylative damage occurs when an alkylating species containing a reactive methyl group encounters a DNA nucleotide, which causes the donation of the free methyl group to the base. Oxidative and alkylative damage are derived from different sources but cause similar results within the DNA chain. Specifically, the modified nucleotides become molecularly large adducts compared to the non-damaged native nucleotide counterparts. These bulky adducts can cause mispairing between nucleotides, such as an 8oxoG // A mis-pair, or more commonly lead to DNA lesions [90-114]. These DNA lesions are points in the DNA tertiary structure which become 'kinked' due to the presence of a bulky DNA adduct. One of the most impactful results of DNA lesions is known as a double-strand break (DSB). Double-strand breaks occur when the helical backbone of two DNA strands are "snapped" due to stress caused from structural change. Double-strand breaks occur naturally during meiosis in DNA replication, and during VDJ recombination in mature B cells [90-114]. However, they also occur at undesirable points in the cell cycle due to toxin exposure. DSBs can occur due to radiation, oxidation, or alkylation caused by a variety of sources. These sources include chemotherapeutics, radioactive tracing agents, and many common carcinogens such as Asbestos. [112-115]. DSBs are also often preceded by single-strand breaks and often occur at the ends of chromosomes due to defective metabolism of telomeres [90-114].

In order to maintain a healthy genome, each cell must have an active response against DNA damage. An estimated 120,000 DNA lesions per cell occur every day [101]. Thus, each cell must maintain a constant balance between DNA damage and repair [90-114]. If DNA damage accumulates, there are several results that may follow [97, 100-105]:

1. The DNA is repaired as normal and the cell continues normal function.

2. The cell may enter cell cycle arrest causing cellular senescence, eventual apoptosis may occur.

3. The cell may enter a state of rapid mutation and cell cycle amplification, becoming cancerous.

In order to prevent tumorigenesis, there are several mechanisms within a healthy cell that aid in DNA repair and maintenance. The conglomeration of these mechanisms is known as the DNA Damage Response (DDR). In order to respond to DNA damage, cellular machinery must play several key roles. The first of these roles are DNA damage *sensors*. The sensors perform the key role of detecting damage to the genome and providing initiation signals for the DDR. Once damage has been detected and the DDR has been initiated, *signal transducers* serve to transmit and amplify the signal. Finally, *effectors* receive the amplified signal and begin repairing DNA, activating cell cycle arrest, apoptosis, or cellular senescence [97, 100-105].

DNA repair can be subcategorized into several components. For each type of damage, there must be an appropriate response. The first form of repair comes in the form of *base excision repair* (BER). The base excision repair pathway corrects most DNA adducts efficiently. The first step in this pathway is to hydrolyze the DNA adduct and create an abasic site. This function is performed by a variety of cellular DNA glycosylases such as the alkylpurine DNA glycosylases which are responsible for excising large methylated DNA adducts. An example of an alkylpurine glycosylase is N-methylpurine glycosylase (MPG) which is the enzyme responsible for excising 3-methyadenine [100-105]. Many DNA adducts have glycosylases which have evolved specifically to remove them, however there are also glycosylases which can remove a variety of adducts. Once an abasic site has been created, it can be processed by one of two sub-pathways: single-nucleotide BER (SN-BER) and long-patch BER (LP-BER).

SN-BER is initiated by the cleavage of the DNA backbone by apurinic / apyrimidinic (AP) lyase at the 3' end of the abasic site. AP lyase leaves a 3' deoxyribose phosphate and a 5' phosphate which temporarily block DNA polymerase activity. An AP endonuclease then cleaves the 3'deoxyribose phosphate and the 5' phosphate is removed by a polynucleotide kinase/phosphatase. This leaves a single empty DNA slot which is quickly filled by a DNA polymerase. LP-BER is initiated via an AP endonuclease which cleaves the abasic site at the 3' end leaving only a 3' hydroxyl site. The 3' hydroxyl site can act as a direct substrate for DNA polymerases. Once filled, the nick in the DNA is closed by a DNA ligase leaving a fully repaired DNA site. The process that determines whether SN or LP BER is initiated remains a mystery, but it is suspected that the BER scaffold protein X-ray repair cross-compleneting protein 1 (XRCC1) plays a role [100-105, 112-115].

Direct lesion reversal (DLR) is responsible for the repair of many large methylated DNA adducts such as O6-methylgaunine. DLR is a one step pathway that utilizes enzymes which directly catalyze the reverse reaction to that which formed the DNA adduct. For example, if a guanine is methylated to form an O6-methylgaunine by interaction with an alkylating species, the enzyme O6-methylgaunine methyltransferase directly removes the methyl group, restoring the guanine to its original form without a multi-step process [100-105, 112-115].

The last form of single nucleotide repair is *Nucleotide Excision Repair* (NER). This pathway is responsible for repairing lesions formed from radioactive crosslinking such as cyclobutane pyrimidine dimerization (CPD) through photo-reversal. Briefly, a protein known as Phr1 contains two chromophores, Flavin-adenine Dinucleotide (FADH-) and Methylenetetrahydro folate (MTHF). Photoreversal occurs when Phr1 binds to a CPD containing DNA strand. MTHF absorbs a photon and transfers the excitation to FADH-. FADH

then transfers an electron to the CPD which causes unstable dimerization. The DNA reverts to its native duel monomer state and transfers an electron back to the enzyme so that it may be re-used for catalysis [109-115].

The previously described repair mechanisms are constantly active within cells during healthy cellular states. However, when DNA damage burden increases and causes genomic instability, SSBs and DSBs can wreak havoc on the cellular genome. When these breaks occur, several more complicated repair mechanisms are initiated. These repair mechanisms include SSB repair and DSB repair through Homologous Recombination and Non-Homologous End Joining. If these repair mechanisms are unsuccessful or the DNA damage is too severe, cell cycle arrest and apoptosis will be initiated [109-115].

As mentioned previously, in order to initiate the DDR, *sensor* cellular machinery must be activated. These sensors are mainly the MRE11-RAD50-NSB1 (MRN) complex or Poly(ADP-ribose) Polymerase 1 (PARP1). These sensors work by recruiting the apical kinases Ataxia telangiectasia mutated or Ataxia telangiectasia Rad3-related which are the *signal transducers*. These kinases begin the large phosphorylation cascade that will begin the signal amplification and determine which damage response pathway will be activated [97,100-105], they recruit the *effectors*. The first of these response pathways I will discuss is *Single-Strand Break Repair* (SSBR).

First and foremost, SSBR requires that the SSB has 'clean ends'. This means that at the break point there is a 3' Hydroxyl deoxyribose nucleotide monophosphate (dNMP) and a 5' Phosphate dNMP. However, this is often not the case and 'dirty ends', that will not allow repair machinery to act on them, exist instead. However, these 'dirty ends' can be processed through two pathways. 3' ends can be processed by DNA glycosylases/ AP lyases, while 5' ends can

be processed by DNA Polymerases and/ or 5' flap endonuclease activity. Once the ends have been processed and are 'clean', repair can begin. One of the main components of SSB repair is the protein XRCC1. While having no enzymatic activity of its own, XRCC1 is involved in scaffolding for SSB and BER. It has many interactions with enzymes such as OGG1, APE 1 and MPG, glycosylases and endonucleases which were mentioned previously. However, its main function is in recruitment for SSB repair scaffolding. XRRC1 deficient cells show a 3-4 fold reduced SSB rejoining and a 4-6 fold increase in chromosomal breakage [88-115]. After activation by PolyADP-ribose Polymerase 1, XRCC1 complexes with its partners LiG3α and RPA. It can then recognize the damaged DNA and forms scaffolding for DNA Polymerase δ and several DNA Ligases which insert new bases and rejoin the SSB [109-115].

The final two main repair pathways in repairing chromosomal breakage are centred around DSBs. The first and preferred form of DSB repair is *Homologous Recombination* (HR). HR rejoins DSBs using a sister homolog as a template strand. As a result, this method of repair is very high fidelity. Initial steps in this pathway involve the generation of small single-strand regions at both ends of the break (similar to how one would begin to build a bridge). These are formed by DNA polymerases. Meanwhile the damage has been sensed by MRN which signals ATM or ATR (depending on the point of the cell cycle) to begin the phosphorylation cascade. Once these single-stranded regions have been formed, RAD51 recombinase is delivered to the site of the break via Breast Cancer Gene 2 (BRCA2) and Partner and localizer of BRCA 2 (PALB2). Both BRCA2 and PALB2 are key tumor suppressor genes, which will be discussed later. Many members of the Rad protein family are actually ubiquitins, which has implicated ubiquitination in DSB repair [97, 100-105, 112-115]. RAD51 invades the gap in the DNA with the template strand creating what is known as a Holliday junction. XRCC2 and XRCC3 aid in

scaffolding during this process. Once this junction has been created, DNA synthesis occurs using the sister strand as a template. This is followed by branch migration and eventual resolution of the heteroduplex, forming a complete/repaired DNA helix. It should be noted that HR mostly occurs at replication forks and thus most often occurs during S Phase and G2 of the cell cycle in replicating DNA [97, 100-105, 112-115].

Non-Homologous End Joining (NHEJ) occurs at all other areas of DSBs where replication is not involved and is the predominant form of DSB repair in mammals. It is also the form of repair that occurs during VDJ recombination in B-cells. It most often occurs in G1 phase of the cell cycle. NHEJ has much lower fidelity then HR and is somewhat prone to error [111-115]. Since it is often the primary way DSBs are repaired, its low fidelity and error prone state may mean NHEJ errors are often responsible for cell cycle arrest or tumorigenesis. NHEJ starts with signal initiation by MRN. ATM or ATR begin signal transduction much as in HR. The proteins Ku70 and Ku80 dimerize and bind to the end of DNA DSBs. This in turn recruits the DNA-dependant protein kinase to the site of damage. Activated DNA-PKcs complex recruits Artemis, XRCC4, DNA Ligase 4 (LIG4), and DNA Polymerase μ. Artemis has endonuclease activity necessary to provide the DNA ends for ligation and gap filling by LIG4 and DNA Pol μ. This results in a completed DSB repair without requiring a sister chromosome as a template strand [97, 100-105, 112-115]..

Chemotherapeutics are anticancer drugs which can be used either singularly, or in a combination as a regimen. There are many types of chemotherapeutics. Alkylating species, such as cyclophosphamide and cisplatin, are so named because of their ability to alkylate many molecules. Alkylating species act on proteins, RNA and DNA [97-101]. This ability to bind covalently to DNA via their alkyl group is the primary cause for their anti-cancer

effects. DNA is made of two strands and the molecules may either bind twice to one strand of DNA (intra-strand crosslink) or may bind once to both strands (inter-strand crosslink) [98]. If the cell tries to replicate crosslinked DNA during cell division, or tries to repair it, the DNA strands can break. This leads to a form of programmed cell death called apoptosis. Alkylating agents will work at any point in the cell cycle and thus are known as cell cycle-independent drugs [97-101].

Antimetabolites mirror native cellular building blocks. Many of them resemble nucleotides and nucleosides such as deoxycytidine, while others inhibit essential DNA synthesis enzymes such as methotrexate [91-101]. Nucleoside resembling antimetabolites can be incorporated within DNA during synthesis. This causes halting of the duplication of the DNA effected as the cellular machinery for replication can no longer bind to the non-native DNA [98]. Methotrexate acts by inhibiting dihydrofolate reductase (DHFR), an enzyme which provides folate as an essential substrate for folate coenzymes and thus reduce purine synthesis dramatically. As purine synthesis is essential for DNA synthesis and replication, cell-cycle arrest or apoptosis may occur in response to the methotrexate. Antimetabolites are only effective when the cell is undergoing DNA synthesis in the S phase and thus are cell-cycle dependant in contrast to alkylating agents [98].

Other non-DNA base disrupting chemotherapy modalities include anti-microtubules, which act by disrupting microtubule formation, thus preventing mitosis [97,98]. These are once again cell cycle dependant as a cell must be entering a microtubule dependant phase of mitosis for the drug to be effective. Examples of anti-microtubules would be Vincristine, vinflunine, and paclitaxel [97-101]. There are also topoisomerase inhibitors, which act directly to inhibit topoisomerase I and II. Topoisomerase enzymes assist the partner strand of

replicating/transcribing DNA to supercoil, thus protecting its structural integrity during these processes. If these enzymes are inhibited by drugs such as irinotecan and novobiocin, the enzymes will not properly bind to DNA during these processes leading to increased DSB and SSB formation. The increase in DNA damage will lead to apoptosis in the affected cell. Topoisomerase inhibitors are non-cell cycle dependant [97-101]. Finally, we have the cytotoxic antibiotics. This class of chemotherapeutics consist of many groups including the anthracyclines, such as doxorubicin, and bleomycin [98]. The anthracyclines act by binding within two strands of DNA, thus disrupting their tertiary structure leading to instability. Bleomycin acts by binding to metal ions and producing free radicals which damage DNA. These therapeutics may act during multiple phases of the cell cycle [98].

As is illustrated in the information above, cancer and the DDR are intricately connected. For a cell to become cancerous, its DDR must fail. As well, cancer cells use the DDR to their advantage by using it to repair their own DNA or by entering senescence to avoid apoptosis when it has been damaged by chemotherapeutics. Cancer cells want functional repair systems, but they also want lenient DDRs that will allow them to mutate rapidly, thereby avoiding and adapting to the immune system of their host. In this way, cancer plays a game of "cat- and- mouse" with the immune system and with its own repair system [97, 100-105, 112-115].

1.3 AID/A3 enzymes play a potential role in tumor mutagenesis regulation and cancer initiation and progression

Tumor microenvironments contain high levels of oxidative radicals and alkylating species. The pH is often more acidic or basic than the surrounding tissue, meaning that DNA damage is more likely to be induced in these environments [88-115]. As previously stated, it has been shown that AID/A3 enzymes are highly expressed within tumors [2]. Their mutational signatures are prevalent in many different types of cancer tissue, indicating that these enzymes are involved in driving the process of cancer mutation and immune evasion. In addition, AID/A3 enzymes have been found to be active at sites where DNA is undergoing processes requiring single stranded exposure. A3A/B have been found to be active in replication forks, and at the free ends of DNA such as those found in DSBs and SSBs. A3A has been found to activate the DDR and cause cell cycle arrest as previously mentioned. Oxidative species and alkylative species cause the formation of DNA adducts that can ultimately result in SSBs or DSBs in DNA chains. Non-canonical bases such as 8oxoG are known to form lesions within DNA that evolve to G-quadruplexes [109-115]. It is known that the AID/A3 enzymes require a single-stranded area of DNA at least 5 bases in width to act [3-20]. Oxidative and alkylative damage could theoretically form lesions long enough within a chain of dsDNA, or at the ends in the case of helical breaks, whereby AID/A3 enzymes could interact with and mutate DNA. In addition to this, many chemotherapeutics act by disrupting the structure of singular nucleotides in the genome. It may be the case that AID/A3's are active in the regions of chemotherapeutics, as chemotherapeutics often disrupt DNA tertiary structure and form single-stranded regions. Many chemotherapeutics are cell-cycle specific, and thus only act during replication or mitosis [98]. These are the same moments of the cell cycle where AID/A3's have been found to be active. Perhaps these enzymes are involved in recruiting repair machinery to areas which have been damaged by chemotherapeutics. In contrast, they may damage these

affected regions of DNA further and drive the process of cell cycle arrest and apoptosis. This

complex and uncertain process is illustrated within Figure 1.2.
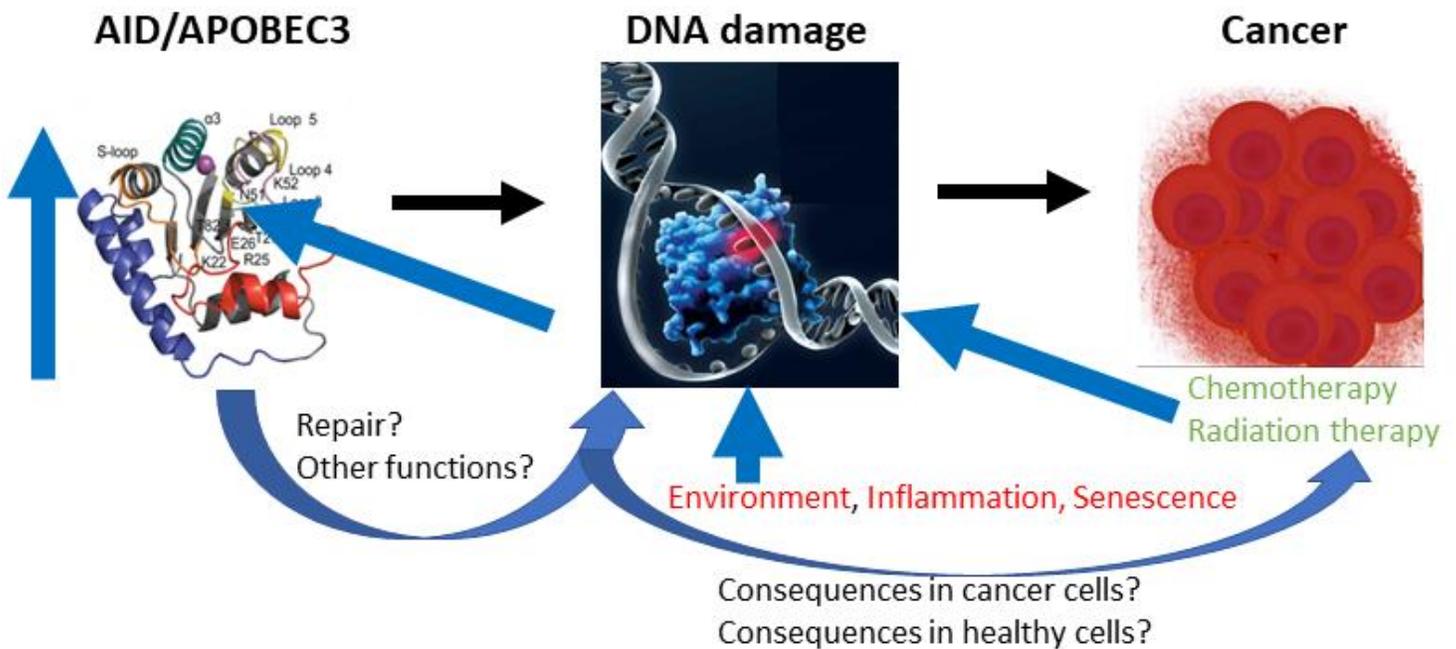
Figure 1.2. Illustration of the complex interactions that may exist between AID/APOBEC3 enzymes and damaged DNA. AID/APOBEC3 enzymes interact with DNA in a destabilized state such as during replication. They act on DNA which is single-stranded. It can be surmised that they may act on DNA in single-stranded states during DNA damage and repair states such as those present in DSBs or SSBs. These ssDNA structure will be more present in cells experiencing stress such as those in oxidative/alkylated environments, radiation exposure, exposure to chemotherapeutics, and inflammation. These processes may lead to cellular senescence or tumorigenesis. Since ssDNA substrates are more actively present in cells experiencing these types of changes, it is reasonable to surmise AID/APOBEC3 enzymes may be active in these settings as well. The exact function of AID/APOBEC3 enzymes in these processes is unclear, but may be one of repair, cancer drive, activation of DDR, or a combination of these. The roles may change depending on whether the cells are cancerous or healthy functioning cells. The black arrows show the natural curse of how DNA-damaging enzymes may lead to cancer. The blue arrows suggest the large amount of interplay that may

exist between these processes and the uncertainty that exists within the current literature regarding the exact chronological order of these processes.

1.4 Overall rationale

Since both AID/A3 enzymes and oxidized/alkylated DNA adducts are more likely to be found within tumor microenvironments, and are indeed even induced through the use of chemotherapeutic drugs, it is reasonable to think that AID/A3 enzymes may encounter such damaged DNA quite frequently in tumor genomes [3-9, 15-20, 109-120]. If this is the case, then it is quite possible that oxidized, or alkylated, DNA adducts have a role to play in the activity levels and regulation of these enzymes within the tumor microenvironment. In a previous paper, it was shown that A3G localized to areas of DSBs within the nucleus [168,169]. A3G was shown to be required for the cells to mount a repair response, and this function was activity dependant [168,169]. It may be that the APOBECs enzymes play a previously undiscovered role in DNA damage and repair. Thus, the work described in this thesis was intended to be the first steps to deciphering whether DNA damage plays a role in the activity regulation of AID, A3A and A3B. We sought to examine whether bases damaged through alkylation or oxidation located proximal to the target dC could impact the catalytic activity of AID, A3A, and A3B.

2. Materials and Methods

2.1 Enzyme expression and purification

Expression and purification of Glutathione S-Transferase (GST) -AID has been previously described [53, 66, 79, 116]. Briefly, DE3 *E. coli* bearing the expression vector pGEX5.3 (GE Heathcare, USA) containing a GST-AID ORF were induced to express GST-AID using 1mM IPTG at 16°C for 16 hours. Cells were lysed using a French pressure cell press (Thermospectronic). The lysate was centrifuged and the supernatant was purified using Glutathione Sepharose high-performance beads (Amersham) as per the manufacturer's instructions. GST-AID was stored at -80°C in 100mM NaCl, 20mM Tris-HCl pH 7.5, and 1mM dithiothreitol. Purity of enzymes in demonstrated in the Coomasie stain shown in Figure 2.1.

Expression and purification of A3A and A3B was carried out using a pcDNA3.1-based eukaryotic expression system, as described [36]. Briefly, EcoRV/KpnI-flanked human GST-A3A or GST-A3B ORF, each containing a single native intron, were inserted into the expression vector pcDNA3.1. The insertion of the intron which is excised upon expression in the eukaryotic host, facilitated the construction of the expression vectors using standard bacteria-based cloning techniques. 25x 10 cm plates, each seeded with 5×10^5 HEK 293T cells, were transfected with 5 μg of expression plasmid per plate using Polyjet transfection reagent (Froggabio). Cells were incubated for 48 hours, collected and resuspended in 500mM NaCl, 50mM Phosphate Buffer pH 8.2, 0.2mM PMSF, and 50ug/ml RNAse A. A3A and A3B expression was confirmed by western blotting using an anti-GST primary (Abcam) followed by secondary Goat anti-Rabbit IgG (SantaCruz). Cells were lysed using a French pressure cell press (Thermospectronic). The Lysate was cleared by ultracentrifugation and the supernatant was purified using Glutathione Sepharose

high-performance beads (Amersham) as per the manufacturer's instructions. GST-A3A and GST-A3B was stored at -80°C in 100mM NaCl, 20mM Tris-HCl pH 7.5, 1mM dithiothreitol (DDT), 5% Glycerol and 50ug/ml BSA. Yields and relative purity of the enzymes were assessed on Coomassie-stained SDS-PAGE using bovine serum albumin (BSA) standard curves, as previously described [69,83]. Briefly, known quantities of purified protein in solution were loaded onto an 8% Acrylamide/SDS resolving gel along with known concentrations of BSA. Once electrophoresis was complete, The gel was imaged using Imagequant software and Kodak Imaging Hardware. Relative quantification of each purified protein was done by comparing the band density of the purified protein to band density of known concentrations of BSA. Relative Quantification was
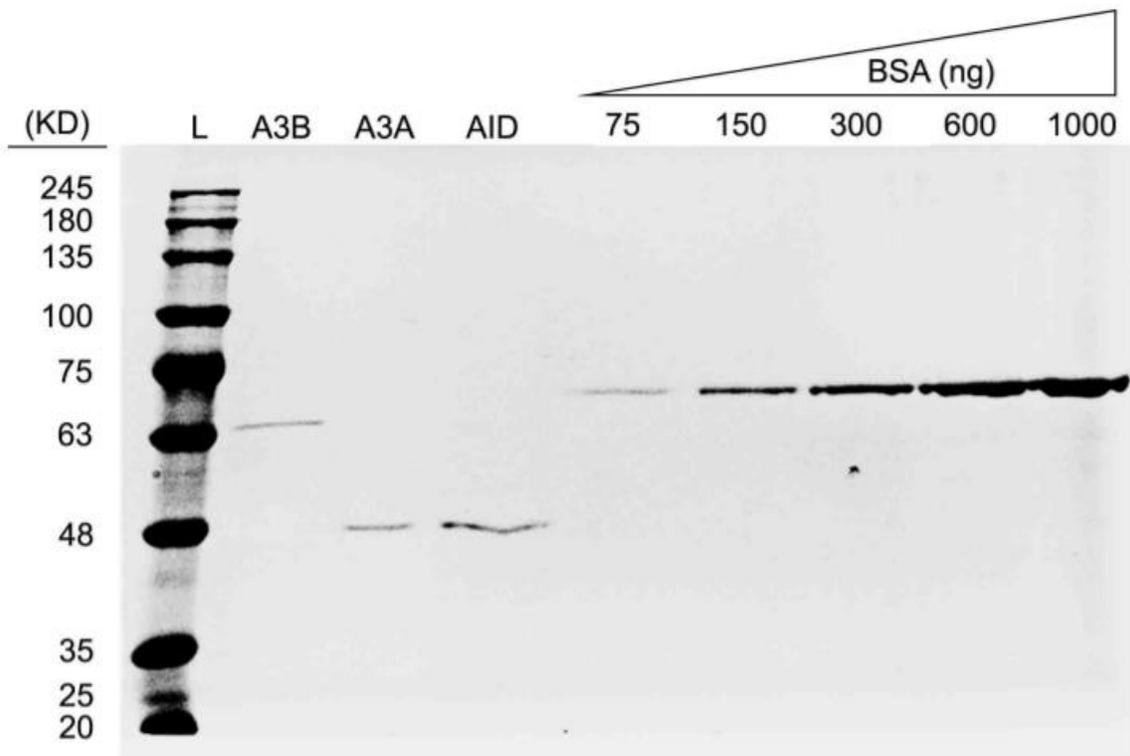
done using three bands of purified protein and 5 concentrations of BSA as shown in Figure 2.1.

Figure 2.1. GST-AID, GST-A3A, and GST-A3B expression and purification. A. Representative Coomassie-stained protein gel of purified GST-tagged A3A, A3B, and AID. A

sample of each purified protein was loaded next to a range of BSA standards used for quantification of enzyme concentrations.

2.2 Preparation of damaged and non-damaged oligonucleotide substrates

Single-stranded and partially single-stranded bubble oligonucleotide substrates for AID, A3A and A3B enzyme assays were prepared as described previously [53, 66, 79, 116]. Briefly, DNA oligonucleotides containing normal and damaged nucleotides (Figure 2.2) were synthesized and HPLC purified (Integrated DNA Technologies, Midland, Trilink). 2.5 pmol of the oligonucleotide bearing the target dC was 5'-labelled with [γ-P32] dATP using T4 polynucleotide kinase (New England Biolabs). After labeling, oligonucleotides were purified using mini-quick spin DNA columns (Roche) and either diluted in TE buffer to 50 fmol to generate single-stranded substrates or annealed with 2-fold excess bottom strand in TE buffer containing 100 mM KCl to generate the bubble substrates. Annealing was carried out through slow cooling from 96°C to 4°C at the rate of 1°C/min. Formation of bubble substrates was verified by electrophoresis on 8% acrylamide native gels alongside unannealed labeled ssDNA oligonucleotides. The designation of substrates which have been described previously [36, 53, 66, 79, 116] indicates their shape and target motif sequence. For instance, Bub7TGC is a bubble type substrate containing a 7-nucleotide long ssDNA region with the target WR<u>C</u> motif TG<u>C</u>. Substrates were concentrated up to 500 fmol/µl using a standard speed vacuum apparatus, followed by serial dilution 250, 100, 75, 50, 40, 25. 12.5, 6.25, 3.125, and 1.6 fmol/ul concentrations for use in enzyme assays.

A

Normal Base	Damaged Base	Brief Description & Rationale of Damaged Base
Gaunine (G)	8-Oxoguanine (8oxoG)	8oxoG is the most abundant damaged nucleoside caused by the oxidation of guanosine by ROS. A common biomarker of endogenous oxidative stress, 8oxoG is responsible for G//A mismatches in DNA due to its resemblance to thymine. Clustered events of guanosine oxidation and also lead to double-stranded breaks (DSBs).
	O6-Methylguanine (O6MeG)	O6MeG is another product of alkylation at the O6 atom of guanosine, resulting in a dual binding affinity for both cytosine and thymine during DNA replication. O6MeG is highly mutagenic and toxic due to its evasion of proofreading through steric mimicry when bound to DNA polymerase. The primary defense against O6MeG is O6-Methylguanine-DNA methyltransferase (MGMT), thus making this a promising target in cancer therapy.
Adenine (A)	8-Oxoadenine (8oxoA)	8oxoA is the most common consequence of adenine oxidation through hydroxyl radicals and DNA irradiation. 8oxoA can induce A>C mutations, although at approximately a four-fold lower frequency than those induced by 8oxoG. When occurring in tandem with an apurinic/apyrimidinic (AP) site, 8oxoA is also capable of inhibiting incision of that AP site by DNA glycosylases and AP endonucleases.
	1-Methyladenine (1MeA)	1MeA is a result of adenosine alkylation at N1. Although not highly mutagenic, 1MeA lacks Watson-Crick (W-C) base pairing capability and thus blocks DNA polymerases during DNA replication. Adenosine alkylation occurs through reactive alkylating agents such as dimethylsulfide and endogenous aldehydes.
Thymine (T)	O4-Methylthymine (O4MeT)	04MeT is caused by alkylation damage at the O4 of thymine. Like 1MeA and O6MeG, this nucleotide is also highly mutagenic and carcinogenic, having lower frequencies of repair than the other alkylated bases. Modification of the O4 results in a shift in favorability from adenine to guanosine, causing T>C transitions.
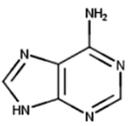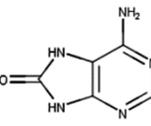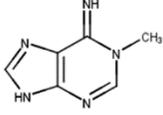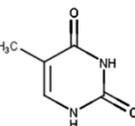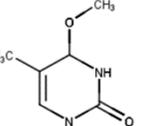
B

Bubble substrates:

Normal:
GCCGCGGGAAGGGACTG   CCGGGCGAAGTGCCGCC
CGGCGCCCTTCCCTGAC   GGCCCGCTTCACGGCGG

Damaged:
GCCGCGGGAAGGGACTG   CCGGGCGAAGTGCCGCC
CGGCGCCCTTCCCTGAC   GGCCCGCTTCACGGCGG

Single-stranded substrates:

Normal: GAAGAGGTGTGTTGATAATATAAGAGA-YYC-AAAAAAGAGAAAGAGGAATGTTAAAT

Damaged: GAAGAGGTGTGTTGATAATATAAGAGA-YXC-AAAAAAGAGAAAGAGGAATGTTAAAT
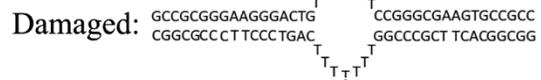
Figure 2.2. Oxidized and alkylated bases and substrates. A. List of all damaged bases used in this study. From left to right, name of each base, the base chemistry relative to the normal version, and a brief description and rationale for selection. The brief statements on the attributes of each base are summarized from cited references (91,92, 95-114, 117-119,128, 136, 147). **B.** Standard bubble and single-stranded substrates used in the alkaline cleavage deamination assay. The upper panel shows an example of the partially single-stranded (bubble) substrates used for testing the enzymatic activity of AID, and the lower panel shows fully single-stranded substrates used for testing the activities of A3A and A3B. The catalytic efficiency of each enzyme on substrates containing the damaged base (denoted by "X") in the -1 or -2 position (right panels) was compared against a control substrate that is identical except for containing the normal undamaged version of the same base in the same -1 or -2 position (left panels).

2.3 Alkaline Cleavage Deamination Assay

The standard alkaline cleavage assay to measure the catalytic activity of AID/APOBECs has previously been described [53, 64, 66, 79, 116]. Briefly, deamination reactions consisted of 1µl of substrate at a substrate concentration ranging from 0.1-10 nM, 6µl reaction buffer (100mM phosphate buffer pH 7.2 for AID, HEPES pH 5.5 for A3A and A3B), and 3µl enzyme (2-10 ng) in a total reaction volume of 10 µl. For initial determination of enzyme activity on normal vs. damaged substrates, a substrate concentration of 2-5 nM was used. For measurement of catalytic efficiency kinetics (Kcat, Km) a range 0.15-100 nM substrate concentrations were used for incubations with AID/A3A/A3B for lengths of time that were determined through time course kinetics (Figure 2.3) to represent the initial velocity phase of the deamination reaction for each enzyme (60 minutes for A3A/A3B, 120 min for AID). Following the AID/A3A/A3B incubation with substrates, AID/A3A/A3B were heat-inactivated by incubation at 80°C for 20 minutes. The reaction volume was then doubled to 20 µl by the addition of Uracil-DNA-Glycosylase (UDG; 1 µl containing 1 unit), 2 µl 10x UDG buffer (New England Biolabs), and 7 µl ddH$_2$O, then incubated at 37°C for 30 minutes to excise deaminated dC's (dU's). Finally, 100mM NaOH was added to the reaction mixture which was incubated at 96°C for 5 minutes to cleave the abasic site left as a

result of dU excision. Each sample was loaded on 14% (for AID) and 16% (for APOBEC3's) Urea

denaturing Polyacrylamide gels for electrophoresis. Gels were exposed to a Kodak Storage

Phosphor Screen and visualized using a phosphorimager (Bio-Rad).
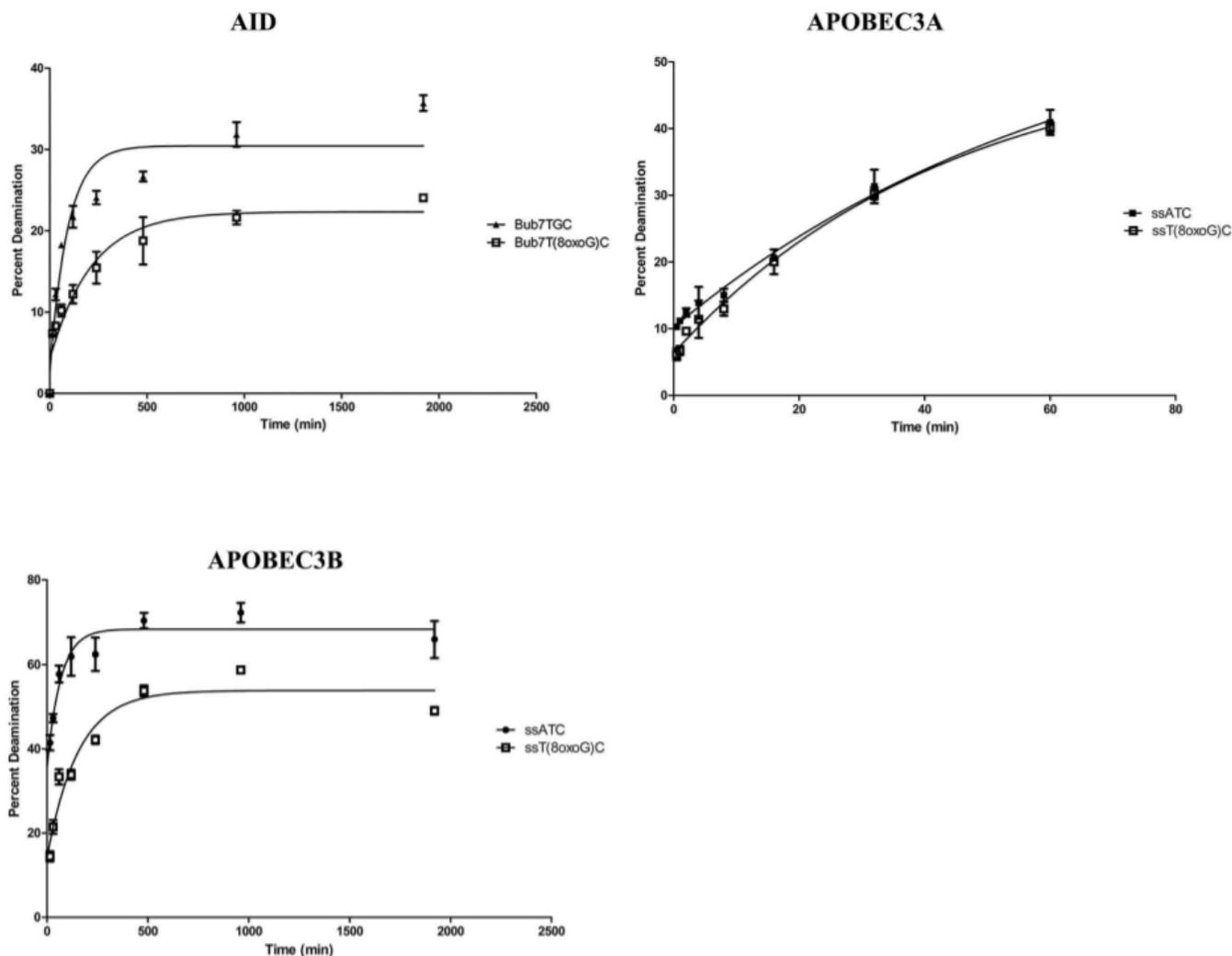
Figure 2.3. Representative time-course kinetics to determine the initial velocity phase of enzyme incubation. Representative time course alkaline cleavage assays of AID (Upper Left), APOBEC3A (Upper Right) and APOBEC3B (Lower Left). Graphs show representative assays for two substrates (one bearing a damaged base, and the corresponding control bearing the normal base version) for each enzyme: Bub7TGC/Bub7T(8oxoG)C for AID, and ssATC/ssT(8oxoG)C for A3A and A3B. Incubation lengths within exponential catalytic phase of each enzyme/substrate were chosen for Michaelis-Menten kinetics.

2.4 Data Collection and Michaelis-Menten Enzyme Kinetics Analysis

Data collection and analysis were performed as previously described for the alkaline cleavage deamination kinetics assay [36, 42, 64, 66]. Briefly, each enzyme kinetics experiment was carried out using 3 preparations of each of AID/A3A/A3B, purified independently but in parallel. Within each experiment, each enzyme:substrate reaction was carried out in triplicate and each triplicate gel lane was analyzed twice independently using the Quantity One 1-D Analysis Software (Bio-Rad) to conduct band densitometry of the substrate and product bands. Using this methodology, the damaged and normal control substrate pairs were always tested in parallel. Data were graphed using GraphPad Prism software and error bars represent standard deviation. Km and Vmax values were obtained by the non-linear regression Michaelis Menten analysis in Graphpad and used to derive Kcat based on the concentration of enzymes used.

2.5 Computational prediction of AID/A3A/A3B surface interactions with damaged bases

Docking of DNA substrates with AID, A3A, and A3B was performed as previously described [64, 65]. Briefly, for AID we used a recently crystalized truncated AID variant (PDB: 5W0U, 5W0R, 5W1C) and a previously-described functional structure of native full-length AID obtained by a combined computational-biochemical method [64]. This functional structure of the AID has been verified by the recent X-ray crystallography studies of partially truncated/mutated AID variants [64, 65, 117, 118]. For A3A, we used 3 recently-published structures: Human A3A X-ray (PDB: 4XXO) [119], Human A3A X-ray (PDB: 5SWW:3.A) [56], Human A3A NMR (PDB: 2m65) [46]. For A3B, we utilized several conformations of the NMR structure of the A3B-CD2 (PDB: 2NBQ) [120]. Using PyMOL we defined catalytic pocket accessibility by ascertaining whether the catalytic glutamate was exposed to the surface of the protein and if the conformation

was able to dock a dC in deamination-feasible conformations as previously described [64]. To

examine the interactions of ssDNA substrates containing the oxidized or alkylated bases with the

surface of AID and A3A we used SwissDock (http://www.swissdock.ch) to ascertain the most

energetically favorable conformations of the substrate across the protein surface [121, 122]. The

creation of substrates began with using ChemDraw Prime v.16.0

(http://www.cambridgesoft.com/software/overview.aspx) to initially make the 2D version of the

substrate, and continued with Marvin Sketch v.5.11.5

(http://www.chemaxon.com/products/marvin/marvinsketch/) to assemble it into 3D. SwissParam

(http://swissparam.ch) was then used to generate the docking parameters [123]. These substrates

were created with a ssDNA template (5'-TTNNNTT'-3) where the central "NNN" motif was either

AGC, TGC, or TTC as normal controls. For damaged DNA, dA in AGC was replaced with (8-

Oxo-dA) or (1-Me-dA), while dT or dG in were replaced with (O4-Me-T), (8-Oxo-dG) or (O6-Me-

dG). UCSF chimera v.1.11.2 (https://www.cgl.ucsf.edu/chimera) was used to view substrate

interactions with AID, A3A and A3B [124]. For each docking simulation, 256 lowest energy

binding modes divided into 32 low-energy clusters were created. Furthermore, each

enzyme:ssDNA binding simulation was carried out multiple times to create 1728 clusters (13824

binding modes) for AID, 5600 clusters (44800 binding modes) for A3A, and 3840 clusters (30720

binding modes) for A3B-CD2. Molecular docking simulations were then qualitatively and

quantitatively analyzed to examine the interactions with the damaged bases in the -2 or -1

positions, in comparison to the normal control nucleotides in the -2 and -1 positions.

3. Results

3.1 Selection of damaged bases and experiment design

The experimental design for comparing AID/A3A/A3B activity on damaged vs. normal substrates is illustrated in Figure 2.2. For damaged bases, we selected 8-oxo-7,8-dihydro-2'-deoxyguanosine (8oxoG), 8-oxo-2'-deoxyadenosine (8oxoA), O6-Methyl-deoxyguanosine (O6MeG), O4-Methyl-deoxythymidine (O4MeT), and 1-Methyl-deoxyadenosine (1MeA). 8oxoG and 8oxoA are the two most common damaged bases in the human genome, because the 8th prime carbon active group of purine bases is easily oxidized due to its low redox potential (Figure 2.2A) [125-129]. Amongst the four bases, dG is most susceptible to oxidation [127-130], making 8oxoG the most common oxidized base form in the human genome and frequently used as a marker of genome senescence [126, 131-134]. O6-Methyl-deoxyguanosine (O6MeG) and O4-Methyl-deoxythymidine (O4MeT) were chosen because they are two frequently found alkylated bases. Alkylation involves addition of alkyl moieties to the ring nitrogen or the exocyclic oxygen by N-nitroso compounds such as the nitrosamines in tobacco smoke [135-137]. O6MeG is one of the most common and toxic forms of DNA alkylation, formed by carcinogenic compounds such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) [135, 138, 139]. O4MeT was selected as another example of alkylation. O4MeT is produced less frequently than O6MeG, but is a more potent source of mutation [138, 140]. 1-Methyl deoxyadenosine (1MeA) is a common alkylated base that is generated by the transfer of a methyl group to the N1 nitrogen atom of deoxyadenosine and can cause critical replication termination [141-144]. Additionally, these damaged bases were chosen due to their clinical relevance, as they are known to be induced in cancer genomes by chemotherapeutic drugs [93, 94, 111, 145].

27

Although AID and A3s exhibit some local sequence specificity beyond the -2/-1 bases that can sometimes include the +1 base [56], the impact of these base positions are generally less than that of the -2/-1 bases. Recently, it has been speculated that the +1 base plays a role in intra-DNA interactions with the -2 nucleotide to conform the U-shape of the substrate. These bases can bind to the enzyme surface as either a purine or a pyrimidine but the -1 and -2 nucleotides play a more prominent role for specificity from [146, 147]. Thus, here we focused on the effect of the -2/-1 base positions. To determine if the aforementioned oxidized or alkylated damaged bases impact the deamination of their neighbouring dC by AID, A3A and A3B, we incorporated them into the -2 and -1 positions relative to the target dC in ssDNA substrates. These substrates have been shown to be optimal targets for deamination activities of these purified form enzymes (Figure 2.2B). A 7 nucleotide-long ssDNA bubble region was selected as the substrate for AID because amongst a panel of different ssDNA forms, this bubble form was found to be the favored substrate for AID [42, 51]. An unpaired single-stranded DNA substrate was used for A3A and A3B since it has been shown to be the optimal substrate for detecting the *in vitro* activity of these enzymes [46, 60, 61].

As a control for each substrate with a damaged base in the -2 or -1 positions, we tested in parallel an identical substrate bearing the corresponding normal base in the same position (Figure 2.2B, left vs. right panels). For instance, substrates with 8oxoA in the -2 position which compose the motif (8oxoA)GC were tested in parallel with control normal substrates containing AGC. We tested the catalytic efficiencies of purified AID, A3A, and A3B on substrates bearing the five alkylated or oxidized bases in the -2 or -1 positions relative to the target dC. To measure relative catalytic efficiencies we used the alkaline cleavage assay to determine deamination rates, as this

is the standard assay that has been widely used to measure the biochemical parameters and substrate preferences of AID/A3 enzymes [51, 53, 79, 80, 148].

3.2 AID can deaminate deoxycytidines neighbouring certain oxidized or alkylated bases with efficiency comparable to native WRC motifs

To evaluate the relative efficiency of AID on substrates bearing damaged bases, we compared initial deamination velocity kinetics of purified AID between substrates containing damaged bases in the -2 or -1 position and control substrates bearing the normal base version in the equivalent position. We first performed time course deamination kinetics experiments to determine enzyme:substrate incubation times for each of the three enzymes (AID, A3A, and A3B) that would be within the initial velocity phase for Michaelis-Menten kinetics (Figure 2.3). To ensure that incubation periods for all substrates fell within the initial velocity phase of the deamination reaction time course kinetics were performed on oligos containing a preferred hotspot sequence motifs of each enzyme and an oligo with a damaged base: TGC/T(8oxoG)C for AID and ATC/T(8oxoG)C for A3A/A3B. Catalytic kinetics were then conducted on a range of substrate concentrations using these incubation times to compare initial velocity kinetics on substrates bearing damaged bases with that of control substrates bearing the normal base in the same position (Figure 2.2B). From the data, catalytic parameters (Km, Kcat, Vmax) were determined for each enzyme:substrate pair, and differences in the catalytic efficiency between damaged and undamaged substrates were quantified through comparison of Kcat/Km ratios (Table 1.)

Table 1. Catalytic parameters/ fold difference of activity of AID, APOBEC3A and APOBEC3B on damaged base motifs compared to normal favored sequence motifs.

Enzyme	Substrate	K_M (μM)	K_{cat} (min^{-1})	K_{cat}/K_M (min$^{-1}\mu$M^{-1})	K_{cat}/K_M fold difference
AID	TG<u>C</u>	0.035±0.0071	1.7×10^{-5}	5×10^{-4}	
	T(8oxoG)<u>C</u>	0.023±0.0023	4.1×10^{-6}	1.8×10^{-4}	0.36
	T(O6MeG)<u>C</u>	0.048±0.0073	2.3×10^{-5}	4.8×10^{-4}	0.96
	(O4MeT)G<u>C</u>	0.026±0.0033	2.2×10^{-5}	8.4×10^{-4}	1.7
	AG<u>C</u>	0.016±0.0023	6.8×10^{-6}	4.2×10^{-4}	
	(8oxoA)G<u>C</u>	0.023±0.0021	1.7×10^{-5}	7.7×10^{-4}	1.8
	(1MeA)G<u>C</u>	0.077±0.0073	2×10^{-5}	2.6×10^{-4}	0.62
APOBEC3A	TG<u>C</u>	0.0024±0.00051	1.5×10^{-4}	0.060	
	T(8oxoG)<u>C</u>	0.019±0.0035	1.5×10^{-3}	0.078	1.3
	T(O6MeG)<u>C</u>	0.030±0.0043	1.2×10^{-3}	0.041	0.68
	TT<u>C</u>	0.023±0.0075	2.6×10^{-3}	0.11	
	T(O4MeT)<u>C</u>	0.053±0.0062	3.2×10^{-3}	0.060	0.55
	TA<u>C</u>	0.10±0.018	4.5×10^{-3}	0.045	
	T(8oxoA)<u>C</u>	0.037±0.0060	2.6×10^{-3}	0.069	1.73
	AT<u>C</u>	0.022±0.0018	3.2×10^{-3}	0.14	
	(8oxoA)T<u>C</u>	0.041±0.0047	7.5×10^{-3}	0.18	1.29
	(1MeA)T<u>C</u>	0.021±0.0019	2.7×10^{-3}	0.13	0.93
APOBEC3B	TG<u>C</u>	0.0028±0.00095	8.1×10^{-4}	0.30	
	T(8oxoG)<u>C</u>	0.026±0.002	9.9×10^{-3}	0.38	1.3
	T(O6MeG)<u>C</u>	0.021±0.002	6.7×10^{-3}	0.32	1.1
	TT<u>C</u>	0.138±0.030	5.9×10^{-2}	0.42	
	T(O4MeT)<u>C</u>	0.080±0.010	1.6×10^{-2}	0.20	0.48
	TA<u>C</u>	0.054±0.0094	1.1×10^{-2}	0.20	
	T(8oxoA)<u>C</u>	0.026±0.0031	1.5×10^{-2}	0.59	3.0
	AT<u>C</u>	0.033±0.0020	1.6×10^{-2}	0.49	
	(8oxoA)T<u>C</u>	0.055±0.0054	4.1×10^{-2}	0.75	1.5

We found that AID was capable of deaminating all damaged trinucleotide motifs that were tested. When the initial velocity of AID activity was compared between damaged and non-damaged control substrate pairs, there was a general trend of robust and comparable deamination of damaged motifs (Figure 6B). Kcat/Km ratios shown in Table 1 (top section: AID) revealed that T(8oxoG)$\underline{C}$ was deaminated with lower efficiency than the control WR$\underline{C}$ motif TG$\underline{C}$ (1.8×10^{-4} vs. 5×10^{-4}) while T(O6MeG)$\underline{C}$ was deaminated with comparable catalytic efficiency (4.8×10^{-4} vs. 5×10^{-4}) and (O4MeT)G$\underline{C}$ was deaminated with higher efficiency (8.4×10^{-4} vs. 5×10^{-4}). (1MeA)G$\underline{C}$ was deaminated by AID with lower efficiency than the control undamaged WR$\underline{C}$ motif AG$\underline{C}$ (2.6×10^{4} vs. 4.2×10^{-4}) whilst (8oxoA)G$\underline{C}$ was deaminated with increased efficiency (7.7×10^{4} vs. 4.2×10^{4}). Of the five damaged base-containing substrates tested, one supported comparable deamination efficiencies to its normal control WR$\underline{C}$-containing and two were deaminated with higher catalytic efficiency. It is noteworthy that all Kcat/Km ratios of AID's deamination of damaged base motifs fell within a relatively comparable 3-fold range of undamaged control WR$\underline{C}$ controls, considering that AID's preference for normal undamaged WR$\underline{C}$ vs. non-WR$\underline{C}$ motifs has been shown to fall within a 3- to 6-fold range using the same assay and substrates employed here, this indicates that AID can efficiently deaminate all tested damaged motifs [42, 50, 53, 66, 149]

AID

A

Bub7TG<u>C</u>   Bub7T(8oxoG)<u>C</u>
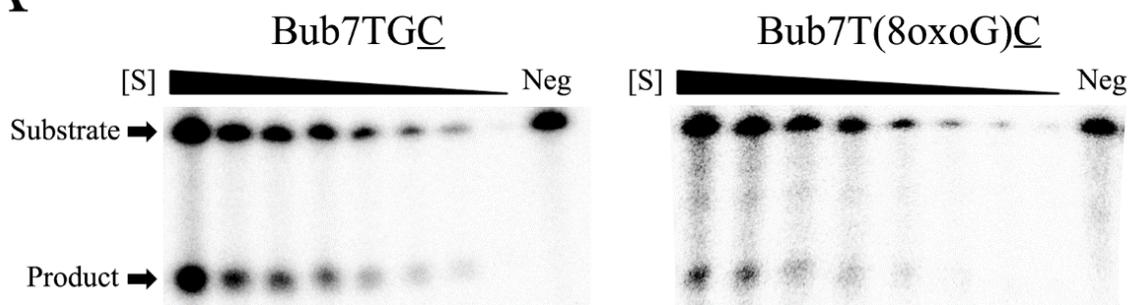
B
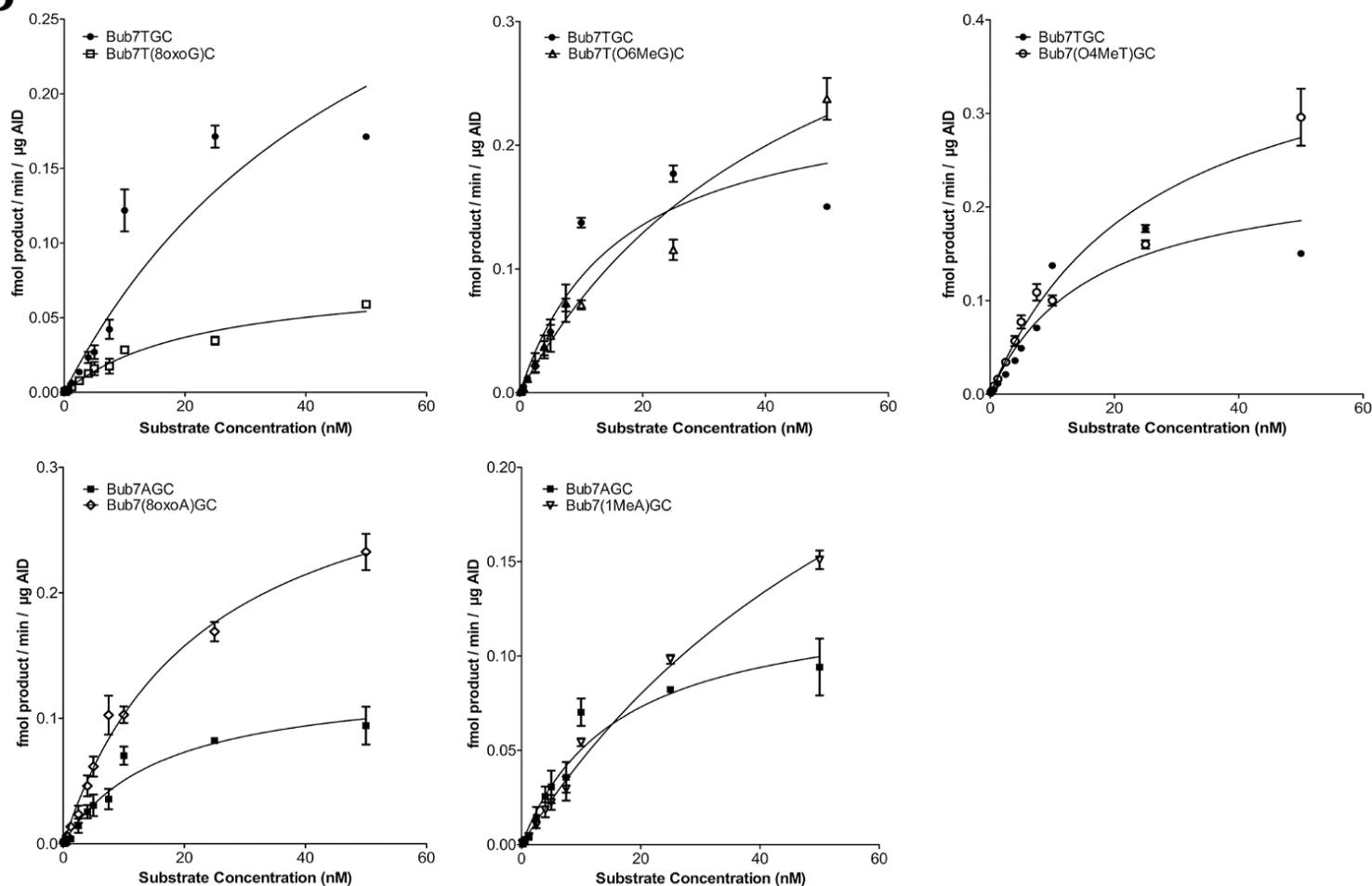
Figure 3.1. AID deaminates substrates with damaged bases at efficiencies nearing that of native WRC motifs. A. Representative enzyme kinetics measuring the initial velocity kinetics of AID deamination of a native WR<u>C</u> motif (Bub7TG<u>C</u>, a seven nucleotide-long bubble substrate containing the motif TGC; left panel), as compared to its damaged motif counterpart bearing the most commonly damaged form of dG 8oxoG in the equivalent (-1) position (Bub7T(8oxo)GC, a seven nucleotide-long bubble substrate containing the motif T(8oxoG)C; right panel), over a range of [Substrate] from 0.15 to 50 nM. Bub7 connotates the seven nucleotide-long bubble substrate

32

used for AID assays. The negative control reaction "Neg" is a no enzyme control that contains buffer in place of AID and is otherwise treated in the same manner with UDG and alkaline cleavage. **B.** Enzyme kinetics of AID-mediated deamination for each damaged base substrate are compared to the control substrates tested in parallel which contain the undamaged native counterpart base in the same -2 or -1 positions; top left: Bub7T(8oxo)G$\underline{C}$ vs. Bub7TG$\underline{C}$, top middle: Bub7T(O6MeG)$\underline{C}$ vs. Bub7TG$\underline{C}$, top right: Bub7(O4MeT)G$\underline{C}$ vs. Bub7TG$\underline{C}$, bottom left: Bub7(8oxoA)G$\underline{C}$ vs. Bub7AG$\underline{C}$, bottom right: Bub7(1MeA)G$\underline{C}$ vs. Bub7AG$\underline{C}$. On each graph, the open symbol represents the substrates containing the damaged base whilst the filled symbol represents the control substrate containing the normal undamaged version of the same sequence motif.

3.3 A3A deaminates deoxycytidines in oxidized or alkylated motifs with comparable or modestly higher efficiency than normal DNA

We then tested the activity of purified A3A on substrates bearing oxidized or alkylated bases. Figure 5A shows an example of A3A's catalytic kinetics for the damaged motif T(8oxoG)$\underline{C}$ compared to its undamaged control counterpart, the TG$\underline{C}$ motif. Like AID, A3A was active on all tested damaged base motifs (Figure 7B). Kcat/Km ratio analyses (Table 1 middle: A3A) revealed that T(8oxoG)$\underline{C}$ supported an increase in catalytic efficiency when compared to its control undamaged substrate TG$\underline{C}$ (0.078 vs. 0.060) whilst T(O6MeG)$\underline{C}$ exhibited a decrease (0.041 vs. 0.060). T(O4MeT)$\underline{C}$ demonstrated reduced catalytic efficiency compared to the favored normal Y$\underline{C}$ motif TT$\underline{C}$ (0.060 vs. 0.11). T(8oxoA)C supported an increase in catalytic efficiency when compared to the undamaged control TA$\underline{C}$ (0.069 vs. 0.045). Of the damaged forms of the favored natural Y$\underline{C}$ motif AT$\underline{C}$, (1MeA)T$\underline{C}$ displayed a comparable catalytic efficiency (0.13 vs. 0.14) whilst (8oxoA)T$\underline{C}$ supported an increase (0.18 vs. 0.14).

Consistent with its demonstrated Y$\underline{C}$ motif preference [46, 56, 58, 80, 83, 148] amongst the four normal control motifs tested, A3A exhibited its highest catalytic efficiencies on TT$\underline{C}$ and AT$\underline{C}$. Remarkably, of the six damaged base motifs tested, two ((1MeA)T$\underline{C}$ and (8oxoA)T$\underline{C}$) were deaminated by A3A with catalytic efficiencies comparable or modestly exceeding that of both of

these most favored normal Y$\underline{C}$ motifs, whilst another two (T(8oxoG)$\underline{C}$ and T(O6MeG)$\underline{C}$) supported catalytic efficiencies that were modestly higher than their own corresponding control undamaged non-Y$\underline{C}$ motifs.

To ensure that the observed deamination preferences for damaged base motifs over their normal control substrates reflect a *bona fide* enzymatic property of the deaminases tested, we conducted an additional control experiment. Since the alkaline cleavage assay for deamination relies on dU excision by Uracil DNA glycosylase (UDG) after dC to dU deamination in the AID/A3:substrate incubation phase, we sought to formally ensure that the more robust product formation on damaged motifs was not due to a sequence preference for the damaged motif at the dU excision stage, though it is known that dU excision by UDG in the context of the alkaline cleavage dC deamination assay is not subject to a local sequence preference. To this end, we incubated dU-converted versions of a normal and damaged substrate pair with UDG and observed no difference in the efficiency of dU excision, and no product in the absence of UDG, as expected (Figure 8). This indicates that the preferences measured in the alkaline cleavage assay reflect the true substrate preferences of AID/A3 enzyme tested.

APOBEC3A

A
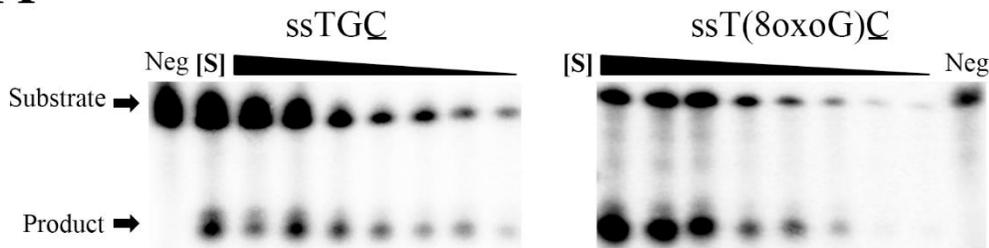
B
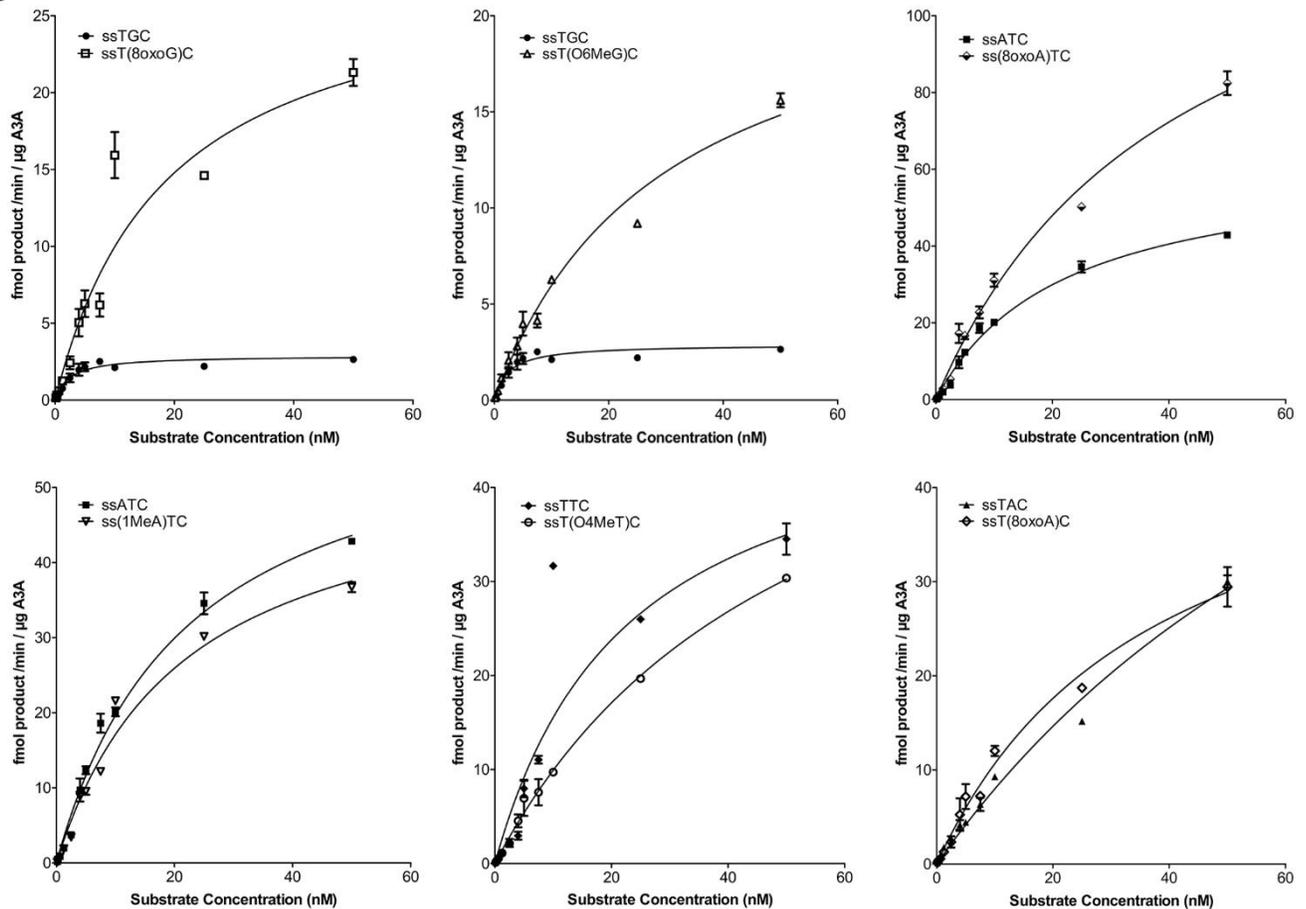
Figure 3.2. APOBEC3A deaminates substrates with certain damaged bases more efficiently

than native sequence motifs. A. Representative enzyme kinetics measuring the initial velocity kinetics of A3A deamination of a damaged and undamaged pair of motifs (ssT(8oxoG)C̲ vs.. ssTGC̲; right panel) over a range of [Substrate] from 0.15 to 50 nM. The negative control reaction "Neg" is a no enzyme control that contains buffer in place of A3A and is otherwise treated in the same manner with UDG and alkaline cleavage. **B.** The initial velocities of A3A mediated

deamination for each damaged base substrate are compared to the control substrates tested in parallel which contain the native normal counterpart base in the same -2 or -1 positions relative to the target dC.; top left: ssT(8oxo)G<u>C</u> vs.vs. ssTG<u>C</u>, top middle: ssT(O6MeG)<u>C</u> vs. ssTG<u>C</u>, top right: ss(8oxoA)T<u>C</u> vs. ssAT<u>C</u>, bottom left: ss(1MeA)T<u>C</u> vs. ssAT<u>C</u>, bottom middle: ssT(O4MeT)<u>C</u> vs. ssTT<u>C</u>, bottom right: ssT(8oxoA)<u>C</u> vs. ssTA<u>C</u>. On each graph, the open symbol represents the substrates containing the damaged base whilst the filled symbol represents the control substrate containing the normal undamaged version of the same sequence motif.

Figure 3.3. Difference in deamination of TGC and T(8oxoG)C is not due to Uracil-DNA Glycosylase. Comparison of uracil excision activity of UDG on the control substrates TGU vs. T(8oxoG)U (Left panel) to demonstrate equivalent UDG activity on both substrates. A representative alkaline cleavage gel depicting the percent excision product produced by the incubation of UDG with 5 nM of GU or (8oxoG)U substrate (right panel).

3.4 A3B can deaminate deoxycytidines located in damaged base motifs with higher efficiency than normal DNA

Figure 7A shows A3B's initial catalytic kinetics on the damaged motif T(8oxoG)<u>C</u> compared to its undamaged control counterpart, the TG<u>C</u> motif. As with AID and A3A, A3B also

exhibited robust deamination activity on the tested damaged motifs (Figure 9B). Kcat/Km ratio analyses revealed that like A3A, A3B also deaminated T(8oxoG)<u>C</u> with higher catalytic efficiency than its undamaged counterpart TG<u>C</u> (0.38 vs. 0.30), whilst T(O6MeG)<u>C</u> was deaminated with comparable catalytic efficiency (0.32 vs. 0.30). T(O4MeT)<u>C</u> was deaminated by A3B less efficiently than its undamaged Y<u>C</u> motif control TT<u>C</u> (0.20 vs. 0.42). On the other hand, A3B deaminated T(8oxoA)<u>C</u> with higher catalytic efficiency as compared to its undamaged control substrate TA<u>C</u> (0.59 vs. 0.20). (8oxoA)T<u>C</u> was deaminated with higher catalytic efficiency than its undamaged Y<u>C</u> control AT<u>C</u> motif (0.79 vs. 0.49) while (1MeA)T<u>C</u> was deaminated with comparable efficiency (0.48 vs. 0.49).

As expected from its favoring of Y<u>C</u> motifs, amongst the normal motifs tested, TT<u>C</u> and AT<u>C</u> exhibited the highest Kcat/Km ratios (0.30 and 0.49, respectively). Of the six damaged base motifs tested, four were favored by A3B over their own undamaged control substrates. Remarkably, of these four damaged motifs, two (T(8oxoA)<u>C</u> and (8oxoA)T<u>C</u>) were deaminated by A3B even more efficiently, and two (T(8oxoG)<u>C</u> and (1MeA)T<u>C</u>) with comparable efficiency, relative to A3B's most favored two normal Y<u>C</u> motifs TT<u>C</u> and AT<u>C</u>.
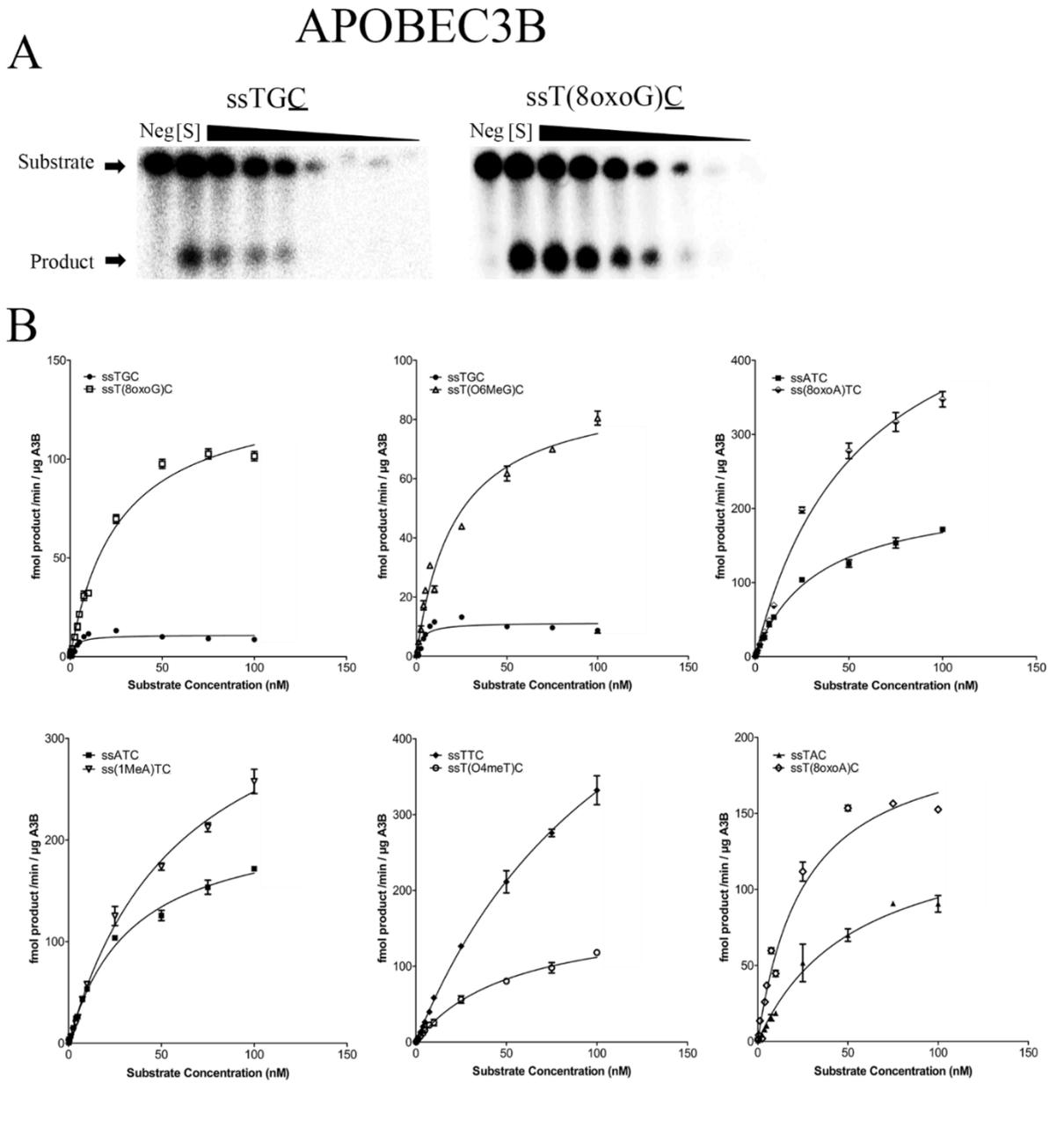
Figure 3.4. APOBEC3B deaminates substrates with certain damaged bases more efficiently than normal motifs. A. Representative enzyme kinetics measuring the initial velocity kinetics of A3B deamination of an undamaged motif (ssTG<u>C</u>; left panel), as compared to its cognate damaged motif (ssT(8oxoG)<u>C</u>; right panel), over a range of [Substrate] from 0.15 to 100 nM. The negative control reaction "Neg" is a no enzyme control that contains buffer in place of A3B and is otherwise treated in the same manner with UDG and alkaline cleavage. **B.** The initial velocities of A3A mediated deamination for each damaged base substrate are compared to the control substrates tested in parallel which contain the native normal counterpart base in the same -2 or -1 positions relative to the target dC. Top left: ssT(8oxo)G<u>C</u> vs. ssTG<u>C</u>, top middle: ssT(O6MeG)<u>C</u> vs. ssTG<u>C</u>, top right: ss(8oxoA)T<u>C</u> vs. ssAT<u>C</u>, bottom left: ss(1MeA)T<u>C</u> vs. ssAT<u>C</u>, bottom middle:

ssT(O4MeT)C̲ vs. ssTTC̲, bottom right: ssT(8oxoA)C̲ vs. ssTAC̲. On each graph, the open symbol represents the substrates containing the damaged base whilst the filled symbol represents the control substrate containing the normal undamaged version of the same sequence motif.

3.5 *In silico* analysis of interactions between the catalytic regions of AID, A3A and A3B with damaged bases

We previously presented AID-DNA docked complexes wherein we highlighted several DNA binding residues and secondary catalytic residues that stabilized dC in the catalytic pocket [64]. Our analysis revealed that the majority of high-frequency DNA binding residues were localized to four secondary catalytic loops: loops 1, 3, 5 and 7 (L1, L3, L5 and L7, respectively). Furthermore, L7, and to a minor extent L1, have been shown to mediate substrate-specificity amongst the AID/APOBEC family [54, 67]. Here, we focused on -1 and -2 base interactions based on previous studies on interactions of AID/APOBECs with ssDNA, the majority of -1 and -2 base binding interactions is likely to be localized to these secondary catalytic loops, with emphasis on L1 and L7 [54, 56, 57, 61, 64, 117, 150-152]. Examining -1/-2 nucleotide positioning in the catalytic region, we observed in certain conformations of AID, the interface between L1 and L7 formed a specificity pocket that frequently bound a purine base from either the -1 or -2 position of TGC or AGC substrates. This finding corroborated previous findings that demonstrated L7 and to a minor extent L1 mediating substrate specificity in AID [54, 55, 67]. The L1/L7 specificity pocket contains a minor contribution from L1 (A21, R24), L5 (W84) and loop 12 (L198) with the majority of residues from L7 (L113, Y114, F115, C116, E117).TG̲C formed the majority of interactions with L1 and L7, specifically within the specificity pocket described above. In contrast to TGC binding, T(8oxoG)C showed an increase in interactions with L1 but a decrease with L7, corresponding to more interactions with A21 and R25 in L1 and decreased interactions with F115, C116, E117 and D118 in L7. When each damaged base was examined alongside the normal control

base in the same position, we found that increased/decreased catalytic activity were associated with increased/decreased predicted interactions with L7, respectively. These analyses suggest that L7 and the L1/L7 interface acts as an anchoring point that stabilizes purines from the -1 and -2 positions, thereby selecting for more catalytically productive AID-DNA complexes (Figure 10A, Figure 11).

In the case of A3A, the damaged motifs that had the most pronounced catalytic effects were in the -1 position. Normal dG in the -1 position was predicted to form many interactions with L5 and L7 on planar and negative residues while 8oxoG in the -1 position interacts largely with L3 on planar and non-charged polar residues. This is seen as a shift from residues F102, and Y132, along with several weaker interactions in L5/L7, towards multiple interactions on Q58, A59, N61, L62, and G68 of L3. This predicted trend of decreased L5/L7 interactions and increased L3 interface usage was consistent across the damaged bases interacting with the surface of A3A. A3B-CD2 exhibits a small extension of α3 such that F285 and W287 and L318 from α4 juxtapose to form a shallow hydrophobic side-pocket. Our docking analysis of A3B-CD2 with TGC showed that when dC was bound in the catalytic pocket, the majority of conformations had the -1 dG of TGC pi-stacked with F285 and W287 along with minimal contact with L318. When comparing T(8oxoG)C, we observed less complexes with 8oxoG bound in the F285/W287/L318 shallow hydrophobic side-pocket. The largest difference involved increased 8oxoG binding to L1 and decreased binding to α4 which corresponded to binding in the L1/L7 region of the ssDNA binding groove (Figure 11C), where the 8oxoG pi- stacked with Y315 and hydrogen bonded from the peptide backbone of R210 and R212 as well as from the side-chains of R211, Y313 and Y315 (Figure 10B, 10C, Figure 11).

A   AID:TGC   AID:T(8oxoG)C

B   A3A:TGC   A3A:T(8oxoG)C

C   A3B-CTD:TGC   A3B-CTD:T(8oxoG)C

41
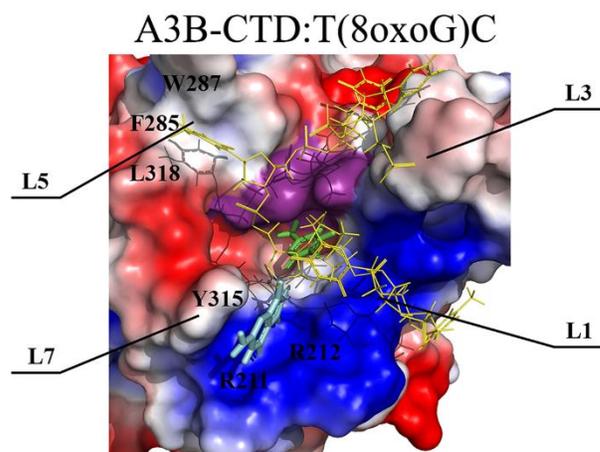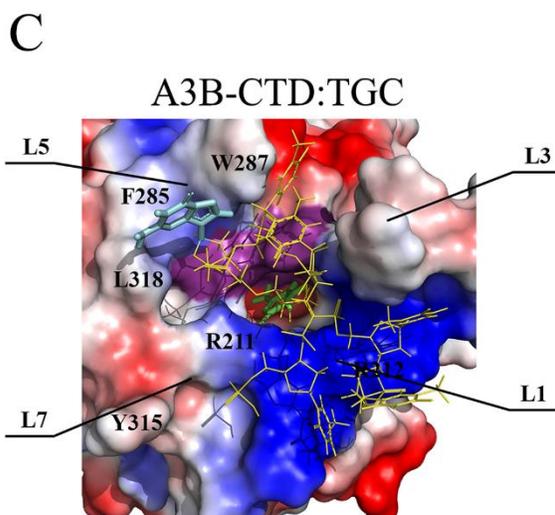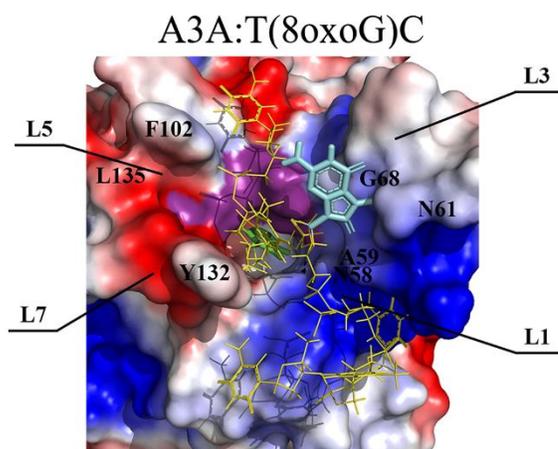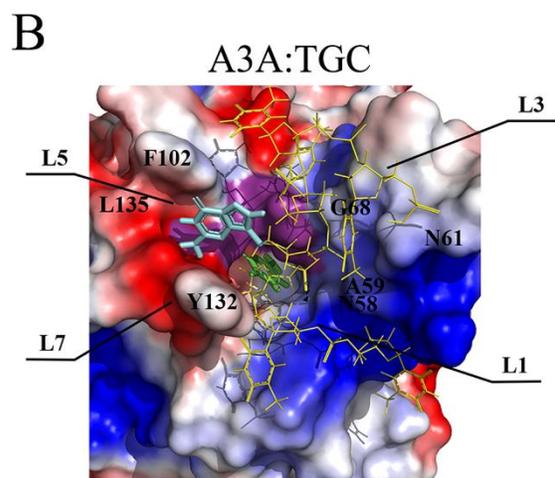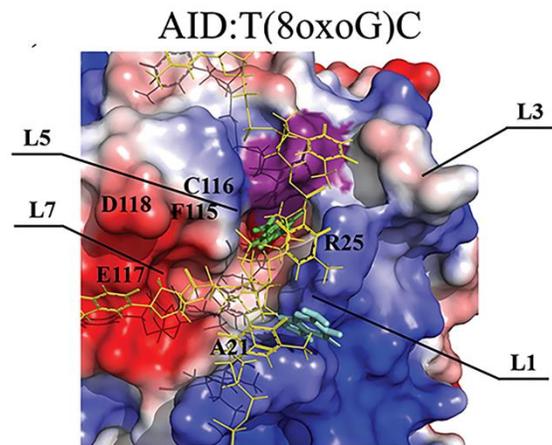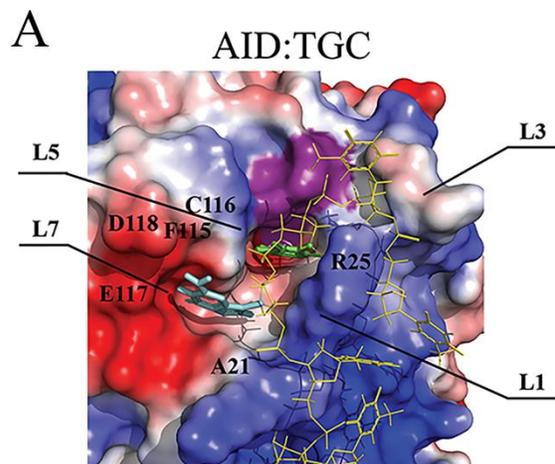
Figure 3.5. Surface regions of AID, A3A and A3B are predicted to interact robustly with damaged bases. AID **(A)**, A3A **(B)** and A3B **(C)** were docked with normal and damaged substrates to examine interactions. Shown are representative analyses of interactions of the catalytic pockets and surrounding regions of AID/A3A/A3B with TG<u>C</u> (Left Panels), and T(8oxoG)<u>C</u> (Right Panels). Residues frequently involved in nucleotide interactions are denoted. Negative and positively charged residues are colored red and blue respectively while purple represents the catalytic pocket residues. The DNA substrate is denoted by the colours yellow, green, and cyan which respectively represent the main DNA chain, the cytidine in the catalytic pocket, and the undamaged nucleotide (-1 dG of TG<u>C</u>; Left Panels) compared to the damaged nucleotide (-1 8oxoG of T(8oxoG)<u>C</u>; Right Panels) .

A

(A)G<u>C</u>
- Loop 1 40%
- Loop 3 10%
- Loop 5 5%
- α-Helix 3 3%
- Loop 7 30%
- α-Helix 4 2%
- Loop 12 10%

(1MeA)G<u>C</u>
- Loop 1 24%
- β-Sheet 2 2%
- Loop 3 39%
- α-Helix 2 6%
- α-Helix 3 14%
- Loop 7 10%
- α-Helix 4 4%
- Loop 12 2%

(8oxoA)G<u>C</u>
- Loop 1 30%
- β-Sheet 2 1%
- Loop 3 1%
- Loop 5 7%
- α-Helix 3 4%
- Loop 7 42%
- α-Helix 4 7%
- Loop 12 8%

(T)G<u>C</u>
- Loop 1 62%
- Loop 3 3%
- Loop 5 7%
- α-Helix 3 3%
- Loop 7 14%
- α-Helix 4 3%
- α-Helix 6 7%

(O4MeT)G<u>C</u>
- Loop 1 26%
- β-Sheet 2 3%
- Loop 3 10%
- α-Helix 3 10%
- Loop 7 44%
- Loop 12 8%

T(G)<u>C</u>
- Loop 1 14%
- β-Sheet 2 3%
- Loop 3 11%
- Loop 5 3%
- Loop 7 56%
- α-Helix 4 6%
- Loop 12 8%

T(8oxoG)<u>C</u>
- Loop 1 33%
- Loop 3 17%
- Loop 5 4%
- α-Helix 3 4%
- Loop 7 29%
- α-Helix 4 4%
- Loop 12 8%

T(O6MeG)<u>C</u>
- Loop 1 6%
- Loop 3 11%
- α-Helix 3 33%
- Loop 7 22%
- α-Helix 4 28%

B

T(T)<u>C</u>
- Loop 1 4%
- Loop 3 47%
- Loop 5 24%
- Loop 7 25%

T(O4MeT)<u>C</u>
- Loop 1 10%
- Loop 3 58%
- Loop 5 4%
- Loop 7 29%

T(G)<u>C</u>
- Loop 3 26%
- Loop 5 39%
- Loop 7 35%

T(O6MeG)<u>C</u>
- Loop 1 3%
- Loop 3 46%
- Loop 5 16%
- Loop 7 35%

T(8oxoG)<u>C</u>
- Loop 1 5%
- Loop 3 62%
- Loop 5 9%
- Loop 7 24%

(A)T<u>C</u>
- Loop 1 3%
- Loop 3 36%
- α-Helix 2 3%
- Loop 5 25%
- α-Helix 3 3%
- Loop 7 28%
- α-Helix 4 2%

(1MeA)T<u>C</u>
- Loop 1 9%
- Loop 3 34%
- α-Helix 2 4%
- Loop 5 24%
- α-Helix 3 3%
- Loop 7 26%

(8oxoA)T<u>C</u>
- Loop 1 24%
- Loop 3 43%
- Loop 5 17%
- α-Helix 2 2%
- α-Helix 3 2%
- Loop 7 11%

T(A)<u>C</u>
- Loop 1 8%
- Loop 3 38%
- Loop 5 10%
- Loop 7 44%

T(8oxoA)<u>C</u>
- Loop 1 5%
- Loop 3 50%
- Loop 5 16%
- α-Helix 3 1%
- Loop 7 28%

C

T(T)<u>C</u>
- Loop 1 9%
- Loop 3 19%
- α-Helix 2 1%
- Loop 5 17%
- α-Helix 3 19%
- Loop 7 16%
- α-Helix 4 19%

T(O4MeT)<u>C</u>
- Loop 1 11%
- Loop 3 22%
- Loop 5 16%
- α-Helix 3 7%
- Loop 7 27%
- α-Helix 4 18%

T(G)<u>C</u>
- Loop 3 25%
- Loop 5 19%
- α-Helix 3 22%
- Loop 7 14%
- α-Helix 4 19%

T(O6MeG)<u>C</u>
- Loop 1 25%
- α-turn 4%
- Loop 3 33%
- Loop 5 10%
- α-Helix 3 6%
- Loop 7 12%
- α-Helix 4 10%

T(8oxoG)<u>C</u>
- Loop 1 19%
- Loop 3 29%
- α-Helix 2 3%
- Loop 5 10%
- α-Helix 3 19%
- Loop 7 16%
- α-Helix 4 3%

(A)T<u>C</u>
- Loop 1 16%
- α-Helix 1 2%
- Loop 3 16%
- α-Helix 2 6%
- Loop 5 14%
- α-Helix 3 29%
- Loop 7 4%
- α-Helix 4 12%

(1MeA)T<u>C</u>
- Loop 1 12%
- α-turn 6%
- Loop 3 41%
- α-Helix 2 4%
- Loop 5 4%
- α-Helix 3 18%
- Loop 6 2%
- Loop 7 6%
- α-Helix 4 6%

(8oxoA)T<u>C</u>
- Loop 1 47%
- α-Helix 4 3%
- Loop 3 38%
- α-Helix 2 3%
- Loop 5 3%
- α-Helix 3 3%
- Loop 7 3%

T(A)<u>C</u>
- Loop 1 9%
- Loop 3 28%
- Loop 5 17%
- α-Helix 3 5%
- Loop 7 19%
- α-Helix 4 22%

T(8oxoA)<u>C</u>
- Loop 1 32%
- α-turn 2%
- Loop 3 47%
- Loop 5 2%
- α-Helix 3 6%
- Loop 7 9%
- α-Helix 4 2%

43

**Figure 3.6. Distribution of the predicted interactions between damaged base-containing substrates and surface regions of AID, A3A and A3B.** AID (A), A3A (B), and A3B (C) were docked with normal and damaged substrates to visualize substrate:enzyme interactions. Each of these pie charts represent the percent interactions of the nucleotide of interest from a 7 nucleotide-long DNA sequence docked on the surface of each enzyme. These sequences contain either a damaged nucleotide in the -1 or -2 position or the undamaged normal form of the same base while the deoxycytidine which enters the catalytic pocket is underlined.

When comparing catalytic activities between other damaged bases and their undamaged controls for A3A and A3B, there was a general structural trend that was in line with the catalytic activity data. We found that T(8oxoG)C, T(8oxoA)C and (8oxoA)TC all showed an increase in catalytic activity relative to their control oligos (Table 1). All oxidized motifs showed an increase in interactions with L1 and L3 and/or a decrease in interaction with L5 and L7 (Figure 11). In contrast, the damaged motifs that resulted in lowered catalytic activities ofA3A and A3B resulted in a similar or greater number of interactions with L1 and L3, but increased interactions with L7. These data suggest that damaged bases -1 or -2 upstream from the target dC that bind more often with L3 or the L1/L3 interface result in a higher catalytic activity. The only exception was T(O6meG)C, which showed an increase on L1/L3 interactions and decreases on L5/L7 interactions, yet exhibited no change in catalytic activity. This is potentially due to the lingering α4 interactions present for A3B and L7 binding for A3A compared with the oxidized -1 nucleotides (Figure 11).

4. Discussion

Although the activities of AID/A3s have been measured on a variety of modified (eg. methylated, oxidized) dC targets, the impact of damaged bases surrounding the target dC has not been studied. To this end, we measured the activity of purified AID, A3A, and A3B on substrates containing several of the most commonly occurring damaged bases proximal to the target dC. Overall, the Michaelis Menten catalytic parameter values determined here for AID, A3A, and A3B were consistent with previous studies, and as established the latter two enzymes exhibited catalytic rates significantly faster than the relatively lethargic AID [46, 53, 65, 80, 153]. We observed that dC positioned proximal to these damaged bases are efficiently deaminated by AID, A3A and A3B, and are sometimes even favored, by several fold in terms of catalytic efficiency, over their native motifs. For instance, when 8oxoA and 8oxoG, two of the most abundant damaged purines in the human genome were placed upstream of the target dC, the catalytic efficiency of A3A and A3B increased compared to the undamaged control sequence, reaching even higher levels than their most preferred normal Y<u>C</u> motif. Thus, we conclude that these enzymes have the inherent catalytic ability to deaminate certain damaged DNA motifs even more efficiently than their most favored sequences yet characterized which consist of normal undamaged WR<u>C</u> or Y<u>C</u> motifs.

When binding the -1 nucleotide of TG<u>C</u>, AID, A3A and A3B all form numerous interactions with L7 and adjacent regions. In AID, the -1 dG nucleotide of TG<u>C</u> binds in the L1/L7 interface (Figure 4C). Due to large structural differences between AID and A3A/A3B, especially at this L1/L7 interface, we found A3A preferred the L5/L7 interface, while A3B bound in a similar region, the L5/α4/L7 interface. Thus, binding of the -1 dG nucleotide of TG<u>C</u> is accomplished through several interactions between L7 in conjunction with neighboring surface regions, dictated

by structural differences between AID, A3A and A3B. Interestingly, interactions formed by the -1 8oxoG nucleotide interactions of T(8oxoG)C also differed amongst the three enzymes:AID bound predominantly with L1, A3A with L3 and A3B with the L1/L7 interface. Comparing the TGC binding with the L5/L7 interface of A3A and the L5/α4/L7 interface of A3B, we found two conserved residues involved: F102 and F285 from A3A and A3B, respectively, as well as L135 and L318 from A3A and A3B, respectively. In agreement with our findings, a structural analysis of A3A found that both F102 and L135 from L5 and L7, as well as several residues in L3 displayed large chemical perturbations upon binding ssDNA [154]. Additionally, it was also discovered that the mutation of F102A, L135A, as well as several L3 residues in A3A lowered catalytic activity but did not completely abrogate it, suggesting these regions are important to fine-tuning catalytic rates in A3A. Because these regions are involved in binding TGC and T(8oxoG)C, our data suggest that the change in interaction with these regions is responsible for the change in catalytic activity on these substrates. This is especially interesting when considering the homology of L1, 3, 5 and 7 amongst AID, A3A and A3B. AID has a shortened loop, thus forming different interactions depending on the substrate [65]. A3A/A3B on the other hand have an extended L3 (8 amino acids longer), which extends outwards from the protein core [56, 147]. Besides the role in secondary Zn coordination [59], L3 also plays a role in ssDNA binding and, from our data, -1/-2 nucleotide interactions with damaged bases. A3A and A3B have 100% identity for L5 and L7, but have near identical homology in the L3 region, except for one amino acid (Q58 in A3A, E241 in A3B). However, the impactful difference appears to reside in L1 which is very different and extended in A3B as compared to A3A. This might explain the differences inT(8oxoG)C binding and activity between A3A vs A3B where the L3 surface is more suitable for binding in A3A, while the L1/L7

interface is more suitable in A3B.

A3 enzymes and AID bind ssDNA substrates which attach to one or two surface grooves in a generally U-shaped manner, with bases that can be flipped to contact the enzyme surfaces and interact with L1, 3, 5, and 7 [46, 56, 57, 64, 65, 117, 146, 155, 156]. Our substrate docking predictions were consistent with this surface binding mode (Figure 10) and also in line with the robust activity of all three enzymes on the oxidized and alkylated motifs, suggested that all three can bind these motifs on their surface in manners that form contacts between the -1 and -2 damaged bases with the substrate specificity loop (L7) and adjacent regions (L1, L5, L3, α3, α4) (Figure 11). These *in silico* observations require definitive confirmation by experimentally examining the enzyme structures bound to damaged vs normal DNA motifs. Nevertheless, the docking predictions support the enzyme assay data showing all three enzymes have the inherent structural capacity to efficiently and specifically recognize and act on damaged base motifs.

Our *in vitro* enzyme assay data indicate that beyond AID/A3 enzymes instigating DNA damage, their catalytic efficiencies can be increased by pre-existing DNA damage proximal to a target dC. AID/A3 enzymes access their genomic targets through various cellular processes that generate ssDNA (bubbles, forks, branched structures). Although we have not examined genome-wide targeting, our data suggests that if AID, A3A, or A3B encounter damaged base motifs through the aforementioned cellular targeting mechanism, they would have the catalytic capacity of deaminating some of these substrates even more efficiently than normal DNA. Under normal conditions, natural oxidization/alkylation events in the genome are repaired with high efficiency, decreasing the likelihood that AID/A3A/A3B will encounter these occurrences. However, DNA damage repair deficiencies are prominent in cancers [105, 157-159]. For instance, 8oxoG which is the most prevalent form of oxidized DNA damage in the genome often causes G:A mismatches

within a cellular genome due to its pyrimidine-like behavior [109, 160], which also correlates with our observation of its favorability for A3A/B as a -1 position base where a pyrimidine is highly favored. 8oxoG:A is typically repaired with high-fidelity in human genome by the glycosylase/AP lyase OGG1 which excises 8oxoG from the DNA duplex or by MYH which removes dA from the 8oxoG:A pair; however, these repair pathways can be perturbed in cancer [109, 161]. O6-methylgaunine (O6meG), which often results from environmental alkylating DNA damage, can be repaired by O6-alkylgaunine-alkyltransferase (MGMT); however, akin to 8oxoG, high levels of O6meG mutations are also found within several different types of cancer cells [108, 115]. Indeed, the most widely used types of chemotherapeutics are also sources of overburdening DNA alkylation and oxidation due to potent compounds [93, 94, 111, 126-128, 137, 145]. In addition, DNA oxidative damage such as 8oxoG and 8oxoA often lead to the stalling of transcriptional elongation, potentially exposing oxidized ssDNA regions known to be targets of AID/A3 enzymes [43, 162]. Oxidation and alkylation of DNA can also induce blockage of replicative polymerases [163, 164], and it has previously been shown that A3A/B can target DNA replication structures [31, 47, 165]. Taken together with our *in vitro* data, the possibility emerges that accumulation of base damage may act through several mechanisms to shift the endogenous patterns of deamination by AID/A3 enzymes. We suggest that this possibility merits further investigation.

Since the AID/A3 enzymes may be active in DNA regions containing damaged DNA, it is also tempting to speculate that they may in this manner play a role in DNA repair, perhaps through instigating recruitment of mismatch repair or base excision pathways which are known to follow their deaminating activities[8, 167, 168]. As, previously mentioned, APOBEC3G (A3G) has been shown to increase efficiency of DSB repair in Lymphoma cells exposed to ionizing radiation [72,174]. A3G multimers were shown to associate with the ssDNA regions and end termini of

resected DSBs. In order for DSB repair rate to be maintained the ability to deaminate the DNA was required, and thus other members of the AID/APOBEC3 family may play similar roles. In a similar vein, A3A is directly involved in activating the DNA damage response in several cell types [175]. A3A was shown to be involved in phosphorylation of several DNA damage response signalling elements such as checkpoint kinase 2 and Replication Protein A (RPA), which are key checkpoint proteins in the cell cycle. In addition, A3A targets the lagging strand of nuclear DNA undergoing replication and may lead to cell cycle arrest when upregulated. Taken together with our results that AID, A3A and A3B act efficiently on the most frequently found forms of base damage, a potential role for these enzymes in repair also merits further investigation.

5. Future Directions

This project has laid the groundwork for many more projects in the future. The fact that these enzymes deaminate these bases *in vitro* does not prove that they do *in vivo*. The natural progression for this project would be to move to cellular assays. First, it would be useful to determine if these enzymes would still be active in cells containing high levels of oxidative or alkylative species. One could test this by incubating cells expressing AID, A3A or A3B with a weak oxidizing agent such as tert-butyl hydroxide, then purifying the enzymes to check and see if they are still active. Another interesting way to study AID/A3 activity in cell culture would be to measure induction of ɣ-H2AX foci in cells expressing these enzymes. ɣ-H2AX histones are indicative of DNA damage occurrence such as double-strand breaks. One can then use a fluorescent antibody to measure the amount of ɣ-H2AX in these cells compared to cells which are not expressing these enzymes. Perhaps the expression of these enzymes in cells cultured in the presence of a DNA oxidizing or alkylating agents would cause an exponential increase in ɣ-H2AX number. Conversely, if these enzymes recognize damaged DNA in order to act as recruiting scaffolds for DNA repair proteins, one is expected to observe fewer ɣ-H2AX foci. This thesis has only scratched the surface of the interactions that may be occurring between AID/APOBEC3 and damaged DNA at a cellular level. As well as the experiments mentioned above, further investigation into the other members of the AID/APOBEC family with these assays may also lead to interesting results. As previously mentioned, the catalytic loop in this family of enzymes is highly conserved and thus it would be interesting to see if the pattern of activity I have discovered with this work holds true within other members of the family. There are many directions to take this project in the future. Hopefully, this project will serve as a basis for unraveling the mystery of how AID/A3 enzymes interact with damaged DNA *in vivo.*

6. Concluding Remarks

Here we have shown that the cytidine deaminases AID, A3A and A3B can efficiently deaminate cytidines neighboring oxidized and alkylated DNA adducts. We have found, through structural modeling, that damaged and non-damaged bases rely on the substrate specificity loop(s) present on the surfaces of these enzymes to regulate access to the catalytic pocket. The alkylated and oxidized groups of the damaged DNA bases seem to show preference for specific regions within the loops compared to their non-damaged counterparts. We have also found that the purine bases in the -1 position tend to align with a second pocket close to the catalytic pocket. As of now, it is unclear what role these enzymes play within cells experiencing genome structural instability. However, it is clear that the relationship between AID/A3 enzymes and DNA damage is not as simple as previously speculated: not only do these enzymes cause DNA damage as previously known, here we have shown that they can sense and act on pre-existing environmental DNA damage.

References:

[1] S.G. Conticello, The AID/APOBEC family of nucleic acid mutators, Genome Biol, 9 (2008) 229.

[2] S. Rebhandl, M. Huemer, R. Greil, R. Geisberger, AID/APOBEC deaminases and cancer, Oncoscience, 2 (2015) 320-333.

[3] P. Revy, T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A.G. Ugazio, N. Brousse, M. Muramatsu, L.D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, A. Durandy, Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2), Cell, 102 (2000) 565-575.

[4] M. Muramatsu, K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, T. Honjo, Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme, Cell, 102 (2000) 553-563.

[5] Y.L. Chiu, W.C. Greene, APOBEC3 cytidine deaminases: distinct antiviral actions along the retroviral life cycle, J Biol Chem, 281 (2006) 8309-8312.

[6] R.S. LaRue, S.R. Jonsson, K.A. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hotzel, V. Andresdottir, T.P. Smith, R.S. Harris, The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals, BMC Mol Biol, 9 (2008) 104.

[7] Rebhandl S, Huemer M, Greil R, Geisberger R. AID/APOBEC deaminases and cancer. *Oncoscience*. 2015;2(4):320–333. Published 2015 Apr 28. doi:10.18632/oncoscience.155

[8] Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol*. 2010;17(2):222–229. doi:10.1038/nsmb.1744

[9] Olson ME, Harris RS, Harki DA. APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem Biol*. 2018;25(1):36–49. doi:10.1016/j.chembiol.2017.10.007

[10] M. Monajemi, C.F. Woodworth, J. Benkaroun, M. Grant, M. Larijani, Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment, Retrovirology, 9 (2012) 35.

[11] Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol. 2005 Feb;22(2):367-77. doi: 10.1093/molbev/msi026. Epub 2004 Oct 20. PubMed PMID: 15496550.

[12] S.U. Siriwardena, K. Chen, A.S. Bhagwat, Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases, Chem Rev, 116 (2016) 12688-12710.

[13] A. Moris, S. Murray, S. Cardinaud, AID and APOBECs span the gap between innate and adaptive immunity, Front Microbiol, 5 (2014) 534.

[14] H.C. Smith, R.P. Bennett, A. Kizilyer, W.M. McDougall, K.M. Prohaska, Functions and regulation of the APOBEC family of proteins, Semin Cell Dev Biol, (2012).

[15] K. Bourara, T.J. Liegler, R.M. Grant, Target cell APOBEC3C can induce limited G-to-A mutation in HIV-1, PLoS Pathog, 3 (2007) 1477-1485.

[16] T.F. Baumert, C. Rosler, M.H. Malim, F. von Weizsacker, Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C, Hepatology, 46 (2007) 682-689.

[17] M. Liu, J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, D.G. Schatz, Two levels of protection for the B cell genome during somatic hypermutation, Nature, 451 (2008) 841-845.

[18] D.F. Robbiani, A. Bothmer, E. Callen, B. Reina-San-Martin, Y. Dorsett, S. Difilippantonio, D.J. Bolland, H.T. Chen, A.E. Corcoran, A. Nussenzweig, M.C. Nussenzweig, AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations, Cell, 135 (2008) 1028-1038.

[19] D.F. Robbiani, S. Bunting, N. Feldhahn, A. Bothmer, J. Camps, S. Deroubaix, K.M. McBride, I.A. Klein, G. Stone, T.R. Eisenreich, T. Ried, A. Nussenzweig, M.C. Nussenzweig, AID produces DNA

double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations, Mol Cell, 36 (2009) 631-641.

[20] A. Ramiro, B. Reina San-Martin, K. McBride, M. Jankovic, V. Barreto, A. Nussenzweig, M.C. Nussenzweig, The role of activation-induced deaminase in antibody diversification and chromosome translocations, Adv Immunol, 94 (2007) 75-107.

[21] L. Pasqualucci, G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H.C. Morse, 3rd, M.C. Nussenzweig, R. Dalla-Favera, AID is required for germinal center-derived lymphomagenesis, Nat Genet, 40 (2008) 108-112.

[22] L. Klemm, C. Duy, I. Iacobucci, S. Kuchen, G. von Levetzow, N. Feldhahn, N. Henke, Z. Li, T.K. Hoffmann, Y.M. Kim, W.K. Hofmann, H. Jumaa, J. Groffen, N. Heisterkamp, G. Martinelli, M.R. Lieber, R. Casellas, M. Muschen, The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia, Cancer Cell, 16 (2009) 232-245.

[23] T. Honjo, M. Kobayashi, N. Begum, A. Kotani, S. Sabouri, H. Nagaoka, The AID dilemma: infection, or cancer?, Adv Cancer Res, 113 (2012) 1-44.

[24] K. Shinmura, H. Igarashi, M. Goto, H. Tao, H. Yamada, S. Matsuura, M. Tajima, T. Matsuda, A. Yamane, K. Funai, M. Tanahashi, H. Niwa, H. Ogawa, H. Sugimura, Aberrant expression and mutation-inducing activity of AID in human lung cancer, Ann Surg Oncol, 18 (2011) 2084-2092.

[25] Y. Matsumoto, H. Marusawa, K. Kinoshita, Y. Niwa, Y. Sakai, T. Chiba, Up-regulation of activation-induced cytidine deaminase causes genetic aberrations at the CDKN2b-CDKN2a in gastric cancer, Gastroenterology, 139 (2010) 1984-1994.

[26] M.B. Burns, L. Lackey, M.A. Carpenter, A. Rathore, A.M. Land, B. Leonard, E.W. Refsland, D. Kotandeniya, N. Tretyakova, J.B. Nikas, D. Yee, N.A. Temiz, D.E. Donohue, R.M. McDougle, W.L. Brown, E.K. Law, R.S. Harris, APOBEC3B is an enzymatic source of mutation in breast cancer, Nature, 494 (2013) 366-370.

[27] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A. Aparicio, S. Behjati, A.V. Biankin, G.R. Bignell, N. Bolli, A. Borg, A.L. Borresen-Dale, S. Boyault, B. Burkhardt, A.P. Butler, C. Caldas, H.R. Davies, C. Desmedt, R. Eils, J.E. Eyfjord, J.A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D.T. Jones, D. Jones, S. Knappskog, M. Kool, S.R. Lakhani, C. Lopez-Otin, S. Martin, N.C. Munshi, H. Nakamura, P.A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J.V. Pearson, X.S. Puente, K. Raine, M. Ramakrishna, A.L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T.N. Schumacher, P.N. Span, J.W. Teague, Y. Totoki, A.N. Tutt, R. Valdes-Mas, M.M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L.R. Yates, I. Australian Pancreatic Cancer Genome, I.B.C. Consortium, I.M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P.A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S.M. Grimmond, R. Siebert, E. Campo, T. Shibata, S.M. Pfister, P.J. Campbell, M.R. Stratton, Signatures of mutational processes in human cancer, Nature, 500 (2013) 415-421.

[28] S.A. Roberts, M.S. Lawrence, L.J. Klimczak, S.A. Grimm, D. Fargo, P. Stojanov, A. Kiezun, G.V. Kryukov, S.L. Carter, G. Saksena, S. Harris, R.R. Shah, M.A. Resnick, G. Getz, D.A. Gordenin, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers, Nat Genet, 45 (2013) 970-976.

[29] S.A. Roberts, D.A. Gordenin, Hypermutation in human cancer genomes: footprints and mechanisms, Nat Rev Cancer, 14 (2014) 786-800.

[30] S. Nik-Zainal, H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L.B. Alexandrov, S. Martin, D.C. Wedge, P. Van Loo, Y.S. Ju, M. Smid, A.B. Brinkman, S. Morganella, M.R. Aure, O.C. Lingjaerde, A. Langerod, M. Ringner, S.M. Ahn, S. Boyault, J.E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J.A. Foekens, M. Gerstung, G.K. Hooijer, S.J. Jang, D.R. Jones, H.Y. Kim, T.A. King, S. Krishnamurthy, H.J. Lee, J.Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporte, X. Pivot, C.A. Purdie, K. Raine, K. Ramakrishnan, F.G. Rodriguez-Gonzalez, G. Romieu, A.M. Sieuwerts, P.T. Simpson, R. Shepherd, L. Stebbings, O.A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G.G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. van't Veer, A. Tutt, S. Knappskog, B.K. Tan, J. Jonkers, A. Borg, N.T. Ueno, C. Sotiriou,

A. Viari, P.A. Futreal, P.J. Campbell, P.N. Span, S. Van Laere, S.R. Lakhani, J.E. Eyfjord, A.M. Thompson, E. Birney, H.G. Stunnenberg, M.J. van de Vijver, J.W. Martens, A.L. Borresen-Dale, A.L. Richardson, G. Kong, G. Thomas, M.R. Stratton, Landscape of somatic mutations in 560 breast cancer whole-genome sequences, Nature, 534 (2016) 47-54.

[31] K. Chan, S.A. Roberts, L.J. Klimczak, J.F. Sterling, N. Saini, E.P. Malc, J. Kim, D.J. Kwiatkowski, D.C. Fargo, P.A. Mieczkowski, G. Getz, D.A. Gordenin, An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers, Nat Genet, 47 (2015) 1067-1072.

[32] B.A. Knisbacher, D. Gerber, E.Y. Levanon, DNA Editing by APOBECs: A Genomic Preserver and Transformer, Trends Genet, 32 (2016) 16-28.

[33] N. McGranahan, C. Swanton, Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future, Cell, 168 (2017) 613-628.

[34] C. Swanton, N. McGranahan, G.J. Starrett, R.S. Harris, APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity, Cancer Discov, 5 704-712.

[35] V.B. Seplyarskiy, R.A. Soldatov, K.Y. Popadin, S.E. Antonarakis, G.A. Bazykin, S.I. Nikolaev, APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication, Genome Res, 26 (2016) 174-182.

[36] B. Leonard, S.N. Hart, M.B. Burns, M.A. Carpenter, N.A. Temiz, A. Rathore, R.I. Vogel, J.B. Nikas, E.K. Law, W.L. Brown, Y. Li, Y. Zhang, M.J. Maurer, A.L. Oberg, J.M. Cunningham, V. Shridhar, D.A. Bell, C. April, D. Bentley, M. Bibikova, R.K. Cheetham, J.B. Fan, R. Grocock, S. Humphray, Z. Kingsbury, J. Peden, J. Chien, E.M. Swisher, L.C. Hartmann, K.R. Kalli, E.L. Goode, H. Sicotte, S.H. Kaufmann, R.S. Harris, APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma, Cancer Res, 73 (2013) 7222-7231.

[37] F. Borzooee, M. Asgharpour, E. Quinlan, M.D. Grant, M. Larijani, Viral subversion of APOBEC3s: Lessons for anti-tumor immunity and tumor immunotherapy, Int Rev Immunol, (2017) 1-14.

[38] G.J. Starrett, E.M. Luengas, J.L. McCann, D. Ebrahimi, N.A. Temiz, R.P. Love, Y. Feng, M.B. Adolph, L. Chelico, E.K. Law, M.A. Carpenter, R.S. Harris, The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis, Nat Commun, 7 (2016) 12918.

[39] U. Storb, Why does somatic hypermutation by AID require transcription of its target genes?, Adv Immunol, 122 (2014) 253-277.

[40] H.S. Abdouni, J.J. King, A. Ghorbani, H. Fifield, L. Berghuis, M. Larijani, DNA/RNA hybrid substrates modulate the catalytic activity of purified AID, Mol Immunol, 93 (2018) 94-106.

[41] L.J. DiMenna, J. Chaudhuri, Regulating infidelity: RNA-mediated recruitment of AID to DNA during class switch recombination, Eur J Immunol, 46 (2016) 523-530.

[42] F.L. Meng, Z. Du, A. Federation, J. Hu, Q. Wang, K.R. Kieffer-Kwon, R.M. Meyers, C. Amor, C.R. Wasserman, D. Neuberg, R. Casellas, M.C. Nussenzweig, J.E. Bradner, X.S. Liu, F.W. Alt, Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability, Cell, 159 (2014) 1538-1548.

[43] S.K. Dickerson, E. Market, E. Besmer, F.N. Papavasiliou, AID Mediates Hypermutation by Deaminating Single Stranded DNA, J Exp Med, 197 (2003) 1291-1296.

[44] S. Zheng, B.Q. Vuong, B. Vaidyanathan, J.Y. Lin, F.T. Huang, J. Chaudhuri, Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA, Cell, 161 (2015) 762-773.

[45] J. Qian, Q. Wang, M. Dose, N. Pruett, K.R. Kieffer-Kwon, W. Resch, G. Liang, Z. Tang, E. Mathe, C. Benner, W. Dubois, S. Nelson, L. Vian, T.Y. Oliveira, M. Jankovic, O. Hakim, A. Gazumyan, R. Pavri, P. Awasthi, B. Song, G. Liu, L. Chen, S. Zhu, L. Feigenbaum, L. Staudt, C. Murre, Y. Ruan, D.F. Robbiani, Q. Pan-Hammarstrom, M.C. Nussenzweig, R. Casellas, B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity, Cell, 159 (2014) 1524-1537.

[46] M. Larijani, A. Martin, Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID, Mol Cell Biol, 27 (2007) 8038-8048.

[47] C. Canugovi, M. Samaranayake, A.S. Bhagwat, Transcriptional pausing and stalling causes multiple clustered mutations by human activation-induced deaminase, Faseb J, 23 (2009) 34-44.

[48] J. Hoopes, L. Cortez, T. Mertz, E.P. Malc, P.A. Mieczkowski, S.A. Roberts, APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication, Cell reports, 14 (2016) 1273-1282.

[49] M.B. Adolph, R.P. Love, L. Chelico, Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates, ACS infectious diseases, 4 (2018) 224-238.

[50] I.J. Byeon, J. Ahn, M. Mitra, C.H. Byeon, K. Hercik, J. Hritz, L.M. Charlton, J.G. Levin, A.M. Gronenborn, NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity, Nat Commun, 4 (2013) 1890.

[51] A.M. Green, S. Landry, K. Budagyan, D.C. Avgousti, S. Shalhout, A.S. Bhagwat, M.D. Weitzman, APOBEC3A damages the cellular genome during DNA replication, Cell Cycle, 15 (2016) 998-1008.

[52] R. Bransteitter, P. Pham, M.D. Scharff, M.F. Goodman, Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase, Proc Natl Acad Sci U S A, 100 (2003) 4102-4107.

[53] P. Pham, R. Bransteitter, J. Petruska, M.F. Goodman, Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation, Nature, 424 (2003) 103-107.

[54] M. Larijani, D. Frieder, W. Basit, A. Martin, The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in ung(-/-)msh2(-/-) mice, Immunogenetics, 56 (2005) 840-845.

[55] A. Sohail, J. Klapacz, M. Samaranayake, A. Ullah, A.S. Bhagwat, Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations, Nucleic Acids Res, 31 (2003) 2990-2994.

[56] K. Yu, F.T. Huang, M.R. Lieber, DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine., J Biol Chem, 279 (2004) 6496-6500.

[57] M. Larijani, A.P. Petrov, O. Kolenchenko, M. Berru, S.N. Krylov, A. Martin, AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner, Mol Cell Biol, 27 (2007) 20-30.

[58] M.A. Carpenter, E. Rajagurubandara, P. Wijesinghe, A.S. Bhagwat, Determinants of sequence-specificity within human AID and APOBEC3G, DNA Repair (Amst), 9 (2010) 579-587.

[59] P. Pham, S.A. Afif, M. Shimoda, K. Maeda, N. Sakaguchi, L.C. Pedersen, M.F. Goodman, Activation-induced deoxycytidine deaminase: Structural basis for favoring WRC hot motif specificities unique among APOBEC family members, DNA Repair (Amst), 54 (2017) 8-12.

[60] K. Shi, M.A. Carpenter, S. Banerjee, N.M. Shaban, K. Kurahashi, D.J. Salamango, J.L. McCann, G.J. Starrett, J.V. Duffy, O. Demir, R.E. Amaro, D.A. Harki, R.S. Harris, H. Aihara, Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B, Nat Struct Mol Biol, 24 (2017) 131-139.

[61] T. Kouno, T.V. Silvas, B.J. Hilbert, S.M.D. Shandilya, M.F. Bohn, B.A. Kelch, W.E. Royer, M. Somasundaran, N. Kurt Yilmaz, H. Matsuo, C.A. Schiffer, Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity, Nat Commun, 8 (2017) 15024.

[62] R.P. Love, H. Xu, L. Chelico, Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A, J Biol Chem, 287 (2014) 30812-30822.

[63] A. Marx, M. Galilee, A. Alian, Zinc enhancement of cytidine deaminase activity highlights a potential allosteric role of loop-3 in regulating APOBEC3 enzymes, Scientific reports, 5 (2015) 18191.

[64] M. Mitra, K. Hercik, I.J. Byeon, J. Ahn, S. Hill, K. Hinchee-Rodriguez, D. Singer, C.H. Byeon, L.M. Charlton, G. Nam, G. Heidecker, A.M. Gronenborn, J.G. Levin, Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties, Nucleic Acids Res, 42 (2013) 1095-1110.

[65] E.C. Logue, N. Bloch, E. Dhuey, R. Zhang, P. Cao, C. Herate, L. Chauveau, S.R. Hubbard, N.R. Landau, A DNA sequence recognition loop on APOBEC3A controls substrate specificity, PLoS One, 9 (2014) e97062.

[66] A.A. Vasudevan, S.H. Smits, A. Hoppner, D. Haussinger, B.W. Koenig, C. Munk, Structural features of antiviral DNA cytidine deaminases, Biol Chem, 394 (2013) 1357-1370.

[67] E.W. Refsland, R.S. Harris, The APOBEC3 family of retroelement restriction factors, Curr Top Microbiol Immunol, 371 (2013) 1-27.

[68] J.J. King, C.A. Manuel, C.V. Barrett, S. Raber, H. Lucas, P. Sutter, M. Larijani, Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity, Structure, 23 (2015) 615-627.

[69] J.J. King, M. Larijani, A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer: Schrodinger's CATalytic Pocket, Front Immunol, 8 (2017) 351.

[70] E.M. Quinlan, J.J. King, C.T. Amemiya, E. Hsu, M. Larijani, Biochemical regulatory features of AID remain conserved from lamprey to humans, Mol Cell Biol, (2017).

[71] R.M. Kohli, S.R. Abrams, K.S. Gajula, R.W. Maul, P.J. Gearhart, J.T. Stivers, A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase, J Biol Chem, 284 (2009) 22898-22904.

[72] A. Rathore, M.A. Carpenter, O. Demir, T. Ikeda, M. Li, N.M. Shaban, E.K. Law, D. Anokhin, W.L. Brown, R.E. Amaro, R.S. Harris, The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution, J Mol Biol, 425 (2013) 4442-4454.

[73] M. Teperek-Tkacz, V. Pasque, G. Gentsch, A.C. Ferguson-Smith, Epigenetic reprogramming: is deamination key to active DNA demethylation?, Reproduction, 142 (2011) 621-632.

[74] N. Bhutani, J.J. Brady, M. Damian, A. Sacco, S.Y. Corbel, H.M. Blau, Reprogramming towards pluripotency requires AID-dependent DNA demethylation, Nature, 463 (2010) 1042-1047.

[75] K. Rai, I.J. Huggins, S.R. James, A.R. Karpf, D.A. Jones, B.R. Cairns, DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45, Cell, 135 (2008) 1201-1212.

[76] E.L. Fritz, F.N. Papavasiliou, Cytidine deaminases: AIDing DNA demethylation?, Genes Dev, 24 (2010) 2107-2114.

[77] D.M. Franchini, K.M. Schmitz, S.K. Petersen-Mahrt, 5-Methylcytosine DNA demethylation: more than losing a methyl group, Annu Rev Genet, 46 (2012) 419-441.

[78] D.M. Franchini, S.K. Petersen-Mahrt, AID and APOBEC deaminases: balancing DNA damage in epigenetics and immunity, Epigenomics, 6 (2014) 427-443.

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

[80] Y. Fu, F. Ito, G. Zhang, B. Fernandez, H. Yang, X.S. Chen, DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B, Biochem J, 471 (2015) 25-35.

[81] M.A. Carpenter, M. Li, A. Rathore, L. Lackey, E.K. Law, A.M. Land, B. Leonard, S.M. Shandilya, M.F. Bohn, C.A. Schiffer, W.L. Brown, R.S. Harris, Methylcytosine and Normal Cytosine Deamination by the Foreign DNA Restriction Enzyme APOBEC3A, J Biol Chem, 287 (2012) 34801-34808.

[82] M. Larijani, D. Frieder, T.M. Sonbuchner, R. Bransteitter, M.F. Goodman, E.E. Bouhassira, M.D. Scharff, A. Martin, Methylation protects cytidines from AID-mediated deamination, Mol Immunol, 42 (2005) 599-604.

[83] H. Abdouni, King, JJ., Suliman, M., Quinlan, M., Fifield, H., Larijani, M., Zebrafish AID is capable of deaminating methylated deoxycytidines, Nucleic Acids Res, In press (2013).

[84] E.K. Schutsky, C.S. Nabel, A.K.F. Davis, J.E. DeNizio, R.M. Kohli, APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA, Nucleic Acids Res, (2017).

[85] P. Wijesinghe, A.S. Bhagwat, Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G, Nucleic Acids Res, 40 (2012) 9206-9217.

[86] H.D. Morgan, W. Dean, H.A. Coker, W. Reik, S.K. Petersen-Mahrt, Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming, J Biol Chem, 279 (2004) 52353-52360.

[87] R. Suspene, M.M. Aynaud, J.P. Vartanian, S. Wain-Hobson, Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase, PLoS One, 8 (2013) e63461.

[88] F. Ito, Y. Fu, S.A. Kao, H. Yang, X.S. Chen, Family-Wide Comparative Analysis of Cytidine and Methylcytidine Deamination by Eleven Human APOBEC Proteins, J Mol Biol, 429 (2017) 1787-1799.

[89] C.S. Nabel, H. Jia, Y. Ye, L. Shen, H.L. Goldschmidt, J.T. Stivers, Y. Zhang, R.M. Kohli, AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation, Nat Chem Biol, 8 (2012) 751-758.

[90] C.S. Nabel, S.A. Manning, R.M. Kohli, The Curious Chemical Biology of Cytosine: Deamination, Methylation,and Oxidation as Modulators of Genomic Potential, ACS Chem Biol, (2012).

[91] L. Budzko, P. Jackowiak, K. Kamel, J. Sarzynska, J.M. Bujnicki, M. Figlerowicz, Mutations in human AID differentially affect its ability to deaminate cytidine and 5-methylcytidine in ssDNA substrates in vitro, Scientific reports, 7 (2017) 3873.

[92] J.M. Soll, R.W. Sobol, N. Mosammaparast, Regulation of DNA Alkylation Damage Repair: Lessons and Therapeutic Opportunities, Trends Biochem Sci, 42 (2017) 206-218.

[93] K. Kino, M. Hirao-Suzuki, M. Morikawa, A. Sakaga, H. Miyazawa, Generation, repair and replication of guanine oxidation products, Genes Environ, 39 (2017) 21.

[94] M. Morikawa, K. Kino, T. Oyoshi, M. Suzuki, T. Kobayashi, H. Miyazawa, Analysis of guanine oxidation products in double-stranded DNA and proposed guanine oxidation pathways in single-stranded, double-stranded or quadruplex DNA, Biomolecules, 4 (2014) 140-159.

[95] J. Cadet, K.J.A. Davies, M.H. Medeiros, P. Di Mascio, J.R. Wagner, Formation and repair of oxidatively generated damage in cellular DNA, Free Radic Biol Med, 107 (2017) 13-34.

[96] C.H. Tang, W. Wei, L. Liu, Regulation of DNA repair by S-nitrosylation, Biochim Biophys Acta, 1820 (2012) 730-735.

[97] N. Kondo, A. Takahashi, K. Ono, T. Ohnishi, DNA damage induced by alkylating agents and repair pathways, J Nucleic Acids, 2010 (2010) 543531.

[98] Malhotra V, Perry MC (2003). "Classical chemotherapy: mechanisms, toxicities and the therapeutic window". Cancer Biology & Therapy. doi:10.4161/cbt.199. PMID 14508075.

[99] K. Cheung-Ong, G. Giaever, C. Nislow, DNA-damaging agents in cancer chemotherapy: serendipity and chemical biology, Chem Biol, 20 (2013) 648-659.

[100] D. Fu, J.A. Calvo, L.D. Samson, Balancing repair and tolerance of DNA damage caused by alkylating agents, Nat Rev Cancer, 12 (2012) 104-120.

[101] C.J. Lord, A. Ashworth, The DNA damage response and cancer therapy, Nature, 481 (2012) 287-294.

[102] P. Mao, A.J. Brown, E.P. Malc, P.A. Mieczkowski, M.J. Smerdon, S.A. Roberts, J.J. Wyrick, Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity, Genome Res, 27 (2017) 1674-1684.

[103] J. Bartkova, P. Hamerlik, M.T. Stockhausen, J. Ehrmann, A. Hlobilkova, H. Laursen, O. Kalita, Z. Kolar, H.S. Poulsen, H. Broholm, J. Lukas, J. Bartek, Replication stress and oxidative damage contribute to aberrant constitutive activation of DNA damage signalling in human gliomas, Oncogene, 29 (2010) 5095-5102.

[104] J. Bartkova, C.E. Hoei-Hansen, K. Krizova, P. Hamerlik, N.E. Skakkebaek, E. Rajpert-De Meyts, J. Bartek, Patterns of DNA damage response in intracranial germ cell tumors versus glioblastomas reflect cell of origin rather than brain environment: implications for the anti-tumor barrier concept and treatment, Molecular oncology, 8 (2014) 1667-1678.

[105] A. Frick, V. Khare, G. Paul, M. Lang, F. Ferk, S. Knasmuller, A. Beer, G. Oberhuber, C. Gasche, Overt Increase of Oxidative Stress and DNA Damage in Murine and Human Colitis and Colitis-Associated Neoplasia, Molecular cancer research : MCR, 16 (2018) 634-642.

[106] A. Maciag, G. Sithanandam, L.M. Anderson, Mutant K-rasV12 increases COX-2, peroxides and DNA damage in lung cells, Carcinogenesis, 25 (2004) 2231-2237.

[107] M. Romanowska, A. Maciag, A.L. Smith, J.R. Fields, L.W. Fornwald, K.D. Kikawa, K.S. Kasprzak, L.M. Anderson, DNA damage, superoxide, and mutant K-ras in human lung adenocarcinoma cells, Free Radic Biol Med, 43 (2007) 1145-1155.

[108] N.F. de Miranda, R. Peng, K. Georgiou, C. Wu, E. Falk Sorqvist, M. Berglund, L. Chen, Z. Gao, K. Lagerstedt, S. Lisboa, F. Roos, T. van Wezel, M.R. Teixeira, R. Rosenquist, C. Sundstrom, G. Enblad, M. Nilsson, Y. Zeng, D. Kipling, Q. Pan-Hammarstrom, DNA repair genes are selectively mutated in diffuse large B cell lymphomas, J Exp Med, 210 (2013) 1729-1742.

[109] C. Lahtz, G.P. Pfeifer, Epigenetic changes of DNA repair genes in cancer, Journal of molecular cell biology, 3 (2011) 51-58.

[110] F. Dietlein, L. Thelen, H.C. Reinhardt, Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches, Trends Genet, 30 (2014) 326-339.

[111] E. Mambo, S.G. Nyaga, V.A. Bohr, M.K. Evans, Defective repair of 8-hydroxyguanine in mitochondria of MCF-7 and MDA-MB-468 human breast cancer cell lines, Cancer Res, 62 (2002) 1349-1355.

[112] A.R. Trzeciak, S.G. Nyaga, P. Jaruga, A. Lohani, M. Dizdaroglu, M.K. Evans, Cellular repair of oxidatively induced DNA base lesions is defective in prostate cancer cell lines, PC-3 and DU-145, Carcinogenesis, 25 (2004) 1359-1370.

[113] P.A. Crosbie, G. McGown, M.R. Thorncroft, P.N. O'Donnell, P.V. Barber, S.J. Lewis, K.L. Harrison, R.M. Agius, M.F. Santibanez-Koref, G.P. Margison, A.C. Povey, Association between lung cancer risk and single nucleotide polymorphisms in the first intron and codon 178 of the DNA repair gene, O6-alkylguanine-DNA alkyltransferase, Int J Cancer, 122 (2008) 791-795.

[114] S. Delaney, D.A. Jarem, C.B. Volle, C.J. Yennie, Chemical and biological consequences of oxidatively damaged guanine in DNA, Free radical research, 46 (2012) 420-441.

[115] M.T. Esposito, C.W. So, DNA damage accumulation and repair defects in acute myeloid leukemia: implications for pathogenesis, disease progression, and chemotherapy resistance, Chromosoma, 123 (2014) 545-561.

[116] G.P. Margison, M.F. Santibanez Koref, A.C. Povey, Mechanisms of carcinogenicity/chemotherapy by O6-methylguanine, Mutagenesis, 17 (2002) 483-487.

[117] T.L. Scott, S. Rangaswamy, C.A. Wicker, T. Izumi, Repair of oxidative DNA damage and cancer: recent progress in DNA base excision repair, Antioxid Redox Signal, 20 (2014) 708-726.

[118] C.G. Smith, H. West, R. Harris, S. Idziaszczyk, T.S. Maughan, R. Kaplan, S. Richman, P. Quirke, M. Seymour, V. Moskvina, V. Steinke, P. Propping, F.J. Hes, J. Wijnen, J.P. Cheadle, Role of the oxidative DNA damage repair gene OGG1 in colorectal tumorigenesis, J Natl Cancer Inst, 105 (2013) 1249-1253.

[119] D. Woods, J.J. Turchi, Chemotherapy induced DNA damage response: convergence of drugs and pathways, Cancer biology & therapy, 14 (2013) 379-389.

[120] A.C. Povey, C.N. Hall, D.P. Cooper, P.J. O'Connor, G.P. Margison, Determinants of O(6)-alkylguanine-DNA alkyltransferase activity in normal and tumour tissue from human colon and rectum, Int J Cancer, 85 (2000) 68-72.

[121] A.M. Dancyger, J.J. King, M.J. Quinlan, H. Fifield, S. Tucker, H.L. Saunders, M. Berru, B.G. Magor, A. Martin, M. Larijani, Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding, Faseb J, (2012).

[122] Q. Qiao, L. Wang, F.L. Meng, J.K. Hwang, F.W. Alt, H. Wu, AID Recognizes Structured DNA for Class Switch Recombination, Mol Cell, 67 (2017) 361-373 e364.

[123] P. Pham, S.A. Afif, M. Shimoda, K. Maeda, N. Sakaguchi, L.C. Pedersen, M.F. Goodman, Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification, DNA Repair (Amst), 43 (2016) 48-56.

[124] M.F. Bohn, S.M. Shandilya, J.S. Albin, T. Kouno, B.D. Anderson, R.M. McDougle, M.A. Carpenter, A. Rathore, L. Evans, A.N. Davis, J. Zhang, Y. Lu, M. Somasundaran, H. Matsuo, R.S. Harris, C.A. Schiffer, Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain, Structure, 21 (2013) 1042-1050.

[125] I.J. Byeon, C.H. Byeon, T. Wu, M. Mitra, D. Singer, J.G. Levin, A.M. Gronenborn, Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity, Biochemistry, 55 (2016) 2944-2959.

[126] A. Grosdidier, V. Zoete, O. Michielin, SwissDock, a protein-small molecule docking web service based on EADock DSS, Nucleic Acids Res, 39 (2011) W270-277.

[127] A. Grosdidier, V. Zoete, O. Michielin, Fast docking using the CHARMM force field with EADock DSS, J Comput Chem, (2011).

[128] V. Zoete, M.A. Cuendet, A. Grosdidier, O. Michielin, SwissParam: a fast force field generation tool for small organic molecules, J Comput Chem, 32 (2011) 2359-2368.

[129] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, J Comput Chem, 25 (2004) 1605-1612.

[130] M.S. Cooke, M.D. Evans, M. Dizdaroglu, J. Lunec, Oxidative DNA damage: mechanisms, mutation, and disease, Faseb J, 17 (2003) 1195-1214.

[131] M.S. Cooke, M.D. Evans, 8-Oxo-deoxyguanosine: reduce, reuse, recycle?, Proc Natl Acad Sci U S A, 104 (2007) 13535-13536.

[132] P. Fortini, B. Pascucci, E. Parlanti, M. D'Errico, V. Simonelli, E. Dogliotti, 8-Oxoguanine DNA damage: at the crossroad of alternative repair pathways, Mutat Res, 531 (2003) 127-139.

[133] M. Dizdaroglu, Oxidatively induced DNA damage and its repair in cancer, Mutat Res Rev Mutat Res, 763 (2012) 212-245.

[134] I. Talhaoui, S. Couve, A.A. Ishchenko, C. Kunz, P. Schar, M. Saparbaev, 7,8-Dihydro-8-oxoadenine, a highly mutagenic adduct, is repaired by Escherichia coli and human mismatch-specific uracil/thymine-DNA glycosylases, Nucleic Acids Res, 41 (2013) 912-923.

[135] P.H. Aguiar, C. Furtado, B.M. Repoles, G.A. Ribeiro, I.C. Mendes, E.F. Peloso, F.R. Gadelha, A.M. Macedo, G.R. Franco, S.D. Pena, S.M. Teixeira, L.Q. Vieira, A.A. Guarneri, L.O. Andrade, C.R. Machado, Oxidative stress and DNA lesions: the role of 8-oxoguanine lesions in Trypanosoma cruzi cell viability, PLoS neglected tropical diseases, 7 (2013) e2279.

[136] H. Kasai, Analysis of a form of oxidative DNA damage, 8-hydroxy-2'-deoxyguanosine, as a marker of cellular oxidative stress during carcinogenesis, Mutat Res, 387 (1997) 147-163.

[137] A. Viel, A. Bruselles, E. Meccia, M. Fornasarig, M. Quaia, V. Canzonieri, E. Policicchio, E.D. Urso, M. Agostini, M. Genuardi, E. Lucci-Cordisco, T. Venesio, A. Martayan, M.G. Diodoro, L. Sanchez-Mete, V. Stigliano, F. Mazzei, F. Grasso, A. Giuliani, M. Baiocchi, R. Maestro, G. Giannini, M. Tartaglia, L.B. Alexandrov, M. Bignami, A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer, EBioMedicine, 20 (2017) 39-49.

[138] B. Halliwell, Can oxidative DNA damage be used as a biomarker of cancer risk in humans? Problems, resolutions and preliminary results from nutritional supplementation studies, Free radical research, 29 (1998) 469-486.

[139] P. Rai, T.T. Onder, J.J. Young, J.L. McFaline, B. Pang, P.C. Dedon, R.A. Weinberg, Continuous elimination of oxidized nucleotides is necessary to prevent rapid onset of cellular senescence, Proc Natl Acad Sci U S A, 106 (2009) 169-174.

[140] P.J. Abbott, R. Saffhill, DNA-synthesis with methylated poly(dA-dT) templates: possible role of O4-methylthymine as a pro-mutagenic base, Nucleic Acids Res, 4 (1977) 761-769.

[141] B. Kaina, G. Fritz, S. Mitra, T. Coquerelle, Transfection and expression of human O6-methylguanine-DNA methyltransferase (MGMT) cDNA in Chinese hamster cells: the role of MGMT in protection against the genotoxic effects of alkylating agents, Carcinogenesis, 12 (1991) 1857-1867.

[142] E.M. Noonan, D. Shah, M.B. Yaffe, D.A. Lauffenburger, L.D. Samson, O6-Methylguanine DNA lesions induce an intra-S-phase arrest from which cells exit into apoptosis governed by early and late multi-pathway signaling network activation, Integr Biol (Camb), 4 (2012) 1237-1255.

[143] G.T. Pauly, S.H. Hughes, R.C. Moschel, Comparison of mutagenesis by O6-methyl- and O6-ethylguanine and O4-methylthymine in Escherichia coli using double-stranded and gapped plasmids, Carcinogenesis, 19 (1998) 457-461.

[144] P. Pourquier, J.L. Waltman, Y. Urasaki, N.A. Loktionova, A.E. Pegg, J.L. Nitiss, Y. Pommier, Topoisomerase I-mediated cytotoxicity of N-methyl-N'-nitro-N-nitrosoguanidine: trapping of topoisomerase I by the O6-methylguanine, Cancer Res, 61 (2001) 53-58.

[145] K.B. Altshuler, C.S. Hodes, J.M. Essigmann, Intrachromosomal probes for mutagenesis by alkylated DNA bases replicated in mammalian cells: a comparison of the mutagenicities of O4-methylthymine and O6-methylguanine in cells with different DNA repair backgrounds, Chem Res Toxicol, 9 (1996) 980-987.
[146] K. Larson, J. Sahm, R. Shenkar, B. Strauss, Methylation-induced blocks to in vitro DNA replication, Mutat Res, 150 (1985) 77-84.
[147] J.C. Delaney, J.M. Essigmann, Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli, Proc Natl Acad Sci U S A, 101 (2004) 14051-14056.
[148] H. Yang, S.L. Lam, Effect of 1-methyladenine on thermodynamic stabilities of double-helical DNA structures, FEBS Lett, 583 (2009) 1548-1553.
[149] H. Yang, Y. Zhan, D. Fenn, L.M. Chi, S.L. Lam, Effect of 1-methyladenine on double-helical DNA structures, FEBS Lett, 582 (2008) 1629-1633.
[150] Y. Ding, H. Wang, J. Niu, M. Luo, Y. Gou, L. Miao, Z. Zou, Y. Cheng, Induction of ROS Overload by Alantolactone Prompts Oxidative DNA Damage and Apoptosis in Colorectal Cancer Cells, Int J Mol Sci, 17 (2016) 558.
[151] T.V. Silvas, S. Hou, W. Myint, E. Nalivaika, M. Somasundaran, B.A. Kelch, H. Matsuo, N. Kurt Yilmaz, C.A. Schiffer, Substrate sequence selectivity of APOBEC3A implicates intra-DNA interactions, Scientific reports, 8 (2018) 7511.
[152] J.R. Wagner, O. Demir, M.A. Carpenter, H. Aihara, D.A. Harki, R.S. Harris, R.E. Amaro, Determinants of Oligonucleotide Selectivity of APOBEC3B, Journal of chemical information and modeling, 59 (2019) 2264-2273.
[153] M.B. Adolph, R.P. Love, Y. Feng, L. Chelico, Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B, Nucleic Acids Res, 45 (2017) 11925-11940.
[154] R. Bransteitter, P. Pham, P. Calabrese, M.F. Goodman, Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase, J Biol Chem, 279 (2004) 51612-51621.
[155] X. Xiao, S.X. Li, H. Yang, X.S. Chen, Crystal structures of APOBEC3G N-domain alone and its complex with DNA, Nat Commun, 7 (2016) 12193.
[156] S.J. Ziegler, C. Liu, M. Landau, O. Buzovetsky, B.A. Desimmie, Q. Zhao, T. Sasaki, R.C. Burdick, V.K. Pathak, K.S. Anderson, Y. Xiong, Insights into DNA substrate selection by APOBEC3G from structural, biochemical, and functional studies, PLoS One, 13 (2018) e0195048.
[157] X. Yan, W. Lan, C. Wang, C. Cao, Structural Investigations on the Interactions between Cytidine Deaminase Human APOBEC3G and DNA, Chemistry, an Asian journal, 14 (2019) 2235-2241.
[158] S.U. Siriwardena, T.A. Guruge, A.S. Bhagwat, Characterization of the Catalytic Domain of Human APOBEC3B and the Critical Structural Role for a Conserved Methionine, J Mol Biol, 427 (2015) 3042-3055.
[159] M. Mitra, K. Hercik, I.J. Byeon, J. Ahn, S. Hill, K. Hinchee-Rodriguez, D. Singer, C.H. Byeon, L.M. Charlton, G. Nam, G. Heidecker, A.M. Gronenborn, J.G. Levin, Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties, Nucleic Acids Res, 42 (2014) 1095-1110.
[160] K. Shi, O. Demir, M.A. Carpenter, J. Wagner, K. Kurahashi, R.S. Harris, R.E. Amaro, H. Aihara, Conformational Switch Regulates the DNA Cytosine Deaminase Activity of Human APOBEC3B, Scientific reports, 7 (2017) 17415.
[161] J.D. Salter, H.C. Smith, Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands, Trends Biochem Sci, 43 (2018) 606-622.
[162] Y.K. Chae, J.F. Anker, B.A. Carneiro, S. Chandra, J. Kaplan, A. Kalyan, C.A. Santa-Maria, L.C. Platanias, F.J. Giles, Genomic landscape of DNA repair genes in cancer, Oncotarget, 7 (2016) 23312-23321.
[163] N.J. Curtin, DNA repair dysregulation from cancer driver to therapeutic target, Nat Rev Cancer, 12 (2012) 801-817.

[164] P.A. Jeggo, L.H. Pearl, A.M. Carr, DNA repair, genome stability and cancer: a historical perspective, Nat Rev Cancer, 16 (2016) 35-42.

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

[166] N. Al-Tassan, N.H. Chmiel, J. Maynard, N. Fleming, A.L. Livingston, G.T. Williams, A.K. Hodges, D.R. Davies, S.S. David, J.R. Sampson, J.P. Cheadle, Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors, Nat Genet, 30 (2002) 227-232.

[167] I. Kuraoka, M. Endou, Y. Yamaguchi, T. Wada, H. Handa, K. Tanaka, Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis, J Biol Chem, 278 (2003) 7294-7299.

[168] E.C. Minca, D. Kowalski, Replication fork stalling by bulky DNA damage: localization at active origins and checkpoint modulation, Nucleic Acids Res, 39 (2011) 2610-2623.

[169] E. Markkanen, Not breathing is not an option: How to deal with oxidative DNA damage, DNA Repair (Amst), 59 (2017) 82-105.

[170] N. Kanu, M.A. Cerone, G. Goh, L.P. Zalmas, J. Bartkova, M. Dietzen, N. McGranahan, R. Rogers, E.K. Law, I. Gromova, M. Kschischo, M.I. Walton, O.W. Rossanese, J. Bartek, R.S. Harris, S. Venkatesan, C. Swanton, DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer, Genome Biol, 17 (2016) 185.

[171] B. Pilzecker, H. Jacobs, Mutating for Good: DNA Damage Responses During Somatic Hypermutation, Front Immunol, 10 (2019) 438.

[172] S.P. Methot, J.M. Di Noia, Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination, Adv Immunol, 133 (2017) 37-87.

[173] R. Nowarski, O.I. Wilner, O. Cheshin, O.D. Shahar, E. Kenig, L. Baraz, E. Britan-Rosich, A. Nagler, R.S. Harris, M. Goldberg, I. Willner, M. Kotler, APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair, Blood, 120 (2012) 366-375.

[174] R. Nowarski, M. Kotler, APOBEC3 cytidine deaminases in double-strand DNA break repair and cancer promotion, Cancer Res, 73 (2013) 3494-3498.

[175] S. Landry, I. Narvaiza, D.C. Linfesty, M.D. Weitzman, APOBEC3A can activate the DNA damage response and cause cell-cycle arrest, EMBO Rep, 12 (2011) 444-450.