Substrate Specificity and RNA-based Regulation of AID Orthologs

by

Mina Khoddami

Pharm.D., Shahid Sadoughi University of Medical Sciences, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2023

Declaration of Committee

Name:	**Mina Khoddami**
Degree:	**Master of Science**
Title:	**Substrate specificity and RNA-based regulation of AID orthologs**
Committee:	**Chair:** **Nancy Hawkins** Associate Professor, Molecular Biology and Biochemistry
	Mani Larijani Supervisor Professor, Molecular Biology and Biochemistry
	Lisa Craig Committee Member Professor, Molecular Biology and Biochemistry
	Dipankar Sen Committee Member Professor Emeritus, Molecular Biology and Biochemistry
	Peter Unrau Examiner Professor, Molecular Biology and Biochemistry

Abstract

Activation-induced cytidine deaminase (AID) is a DNA editing enzyme which facilitates antibody maturation through somatic hypermutation (SHM) and class switch recombination (CSR) in activated B cells. Lack of functional AID in human and mouse models impairs class switch recombination resulting in hyper IgM syndrome. Furthermore, it is thought that dysregulated activity of AID can contribute to the onset and progression of various types of cancer. AID binds ssDNA and deaminates deoxycytidine to deoxyuridine at the Ig loci. To date, several cofactors have been proposed to bind AID, but unlike the well-established role of AID itself in CSR and SHM, whether there is any potential regulation on its deamination activity has remained unclear. In addition, AID has some conserved and non-conserved biochemical features among species, of which substrate specificity for different topologies of DNA/DNA and DNA/RNA is not identified yet. In chapter 3, we focused on the characterization of AID orthologs from divergent species, aiming to find the structural determinants that regulate the enzyme's targeting of specific genes. The results of this study reveal variations in activity and thermal sensitivity among AID orthologs. In addition, we showed that the activity of AID orthologs varies to different extents by removing the bound RNAs from the enzyme. In chapter 4, we discovered that specific RNA fragments can inhibit human AID activity, with stronger inhibition observed in sequences found in class switch regions (AID’s target for CSR), as opposed to variable regions (AID’s target for SHM). This study presents novel insights into the potential regulation of AID activity through RNA interactions.

Keywords: Activation induced cytidine deaminase (AID); RNA binding specificity; AID regulation; AID orthologs; Class switch recombination (CSR)

Dedication

To my family who have been a constant source of support and encouragement throughout my academic journey. Their unwavering belief in my abilities and their unconditional love has been helping me to continue even in the darkest moments.

To my professors, colleagues, and friends (in Canada and Iran) whose guidance and expertise have helped shape my academic and professional development. Their passion for science and commitment to excellence has inspired me to strive for greatness.

Acknowledgements

I owe a great deal of gratitude to my supervisor, Dr. Mani Larijani, who was willing to give me the opportunity, when I was new to the country and desperately looking for a chance to pursue my research interests, for his guidance, support, and encouragement throughout this program. His expertise and dedication to teaching helped me to shape my research skills and learn that luck can sometimes play a role in scientific discovery. I am extremely grateful for his support, which has been crucial in the successful completion of this thesis.

I am also thankful to my supervisory committee members, Dr. Lisa Craig and Dr. Dipankar Sen, for their insightful comments, constructive feedback, and valuable suggestions that have contributed to the overall quality of this work. I am also thankful to my examiner Dr. Peter Unrau for his time investment and precious comments on my work.

I would like to acknowledge the invaluable assistance and support of all the current and previous lab members and my friends, Ying, Atefeh, Faezeh, Mahdi, Reyhaneh, David, Justin, Priyanka, Ysabel, Hanadi and specially Sarah, who has contributed to this research work. I am deeply grateful to everyone who made my graduate experience both enriching and memorable.

Additionally, I would like to express my sincere appreciation to the staff and faculty members in the MBB department for their unwavering support and guidance throughout the past few years.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AID	Activation-induced cytidine deaminase
APC	Antigen presenting cell
APOBEC	Apolipoprotein B mRNA-editing catalytic component
APS	ammonium persulfate
Amp	Ampicillin
BCL6	B-cell lymphoma 6
BCR	B cell receptor
BER	Base excision repair
Bp	base pair
BSA	Bovine serum albumin
CDA ½	Cytidine deaminase like ½
CDR	Complementarity-determining region
CSR	Class switch recombination
CLL	Chronic lymphocytic leukemia
dC	Deoxycytidine
dU	Deoxyuridine
DAMPs	Damage-associated molecular patterns (DAMPs)
Dox	Doxycycline
Dr	Danio rerio
DSB	Double-stranded breaks
dsDNA	double-stranded deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
Fab region	Fragment antigen binding region
Fc region	Fragment constant region
Gc	Ginglymostoma cirratum
Gg	Gallus gallus
GST	Glutathione S-transferase

HEK	Human embryonic kidney
His	Histidine
hnRNP	heterogeneous nuclear ribonucleoproteins
hr	Hour
hrs	Hours
Hs	Homo sapiens
Ig	Immunoglobulin
ILCs	Innate lymphoid cells
Ip	Ictalurus punctatus
IPTG	Isopropyl-beta-D-thiogalactopyranoside
KD	Kilodaltons
LB	Luria broth
MAC	Membrane attack complex
Me	Macropus eugenii
MCF7	Michigan Cancer Foundation-7, breath cancer cell line
MHC	Major histocompatibility complex
Min	Minutes
mL	Millilitre
MMR	Mismatch Repair
mRNA	messenger ribonucleic acid
Ng	Nanogram
NHEJ	Non-homologous end joining
NK	Natural killer (cell)
nM	Nanomolar
ORF	Original reading frame
Oc	Oryctolagus curriculus
Ol	Oryzias latipes
PAF1	RNAPII associated factor I
PAMPs	Pathogen associated molecular patterns
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction

pI	Isoelectric point
Pm	Petromyzon marinus
Pmol	Picomole
PMSF	Phenylmethylsulphonyl fluoride
PNK	Polynucleotide 5’-hydroxyl-kinase
Pt	Pan troglodytes
PTBP2	polypyrimidine tract binding protein 2
Pv	Pteropus vampyrus
RAG	Recombination activating gene products
RMS	Root mean square
RNP	Ribonucleoprotein
RNAP	RNA polymerase
RPA	Replication protein A
RSS	Recombination Signal Sequence
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SHM	Somatic hypermutation
ssDNA	single-stranded deoxyribonucleic acid
Ss	Salmo salar
TCR	T cell receptors
TEMED	Tetramethylethylenediamine
UDG	Uracil DNA glycosylase
UNG	Uracil-N-glycosylase (UNG)
Xl	Xenopus laevis

Chapter 1. Introduction

1.1. Overview of the immune system

The immune system is an interrelated network of physical barriers, cells, and cellular pathways that safeguard against pathogens and abnormal cells. The immune system comprises two principal arms: innate and adaptive immunity. Innate immunity is a rapid, non-specific immune response including physical and chemical barriers against pathogens, such as skin, stomach acid, mucous membranes, and lysozyme. Innate immunity also includes cellular components such as phagocytic macrophages, which can recognize pathogen-associated molecular patterns (PAMPs) and eliminate pathogens during the early stages of an infection. On the other hand, adaptive immunity is a more specific yet slower immune response involving B and T lymphocytes. B cells mediate the antibody response known as humoral immunity, while T cells catalyze cell-mediated immunity (1).

1.2. Initial immune response: Innate immunity

The innate immune system is an immediately responsive system that is activated when any pathogens, external or internal, are encountered. The epithelial layer, which is the initial defence structure, creates a barrier that separates the inside of the body from the outside environment. Additionally, chemical barriers like mucus, stomach acid, sebum, lactoferrin, lysozyme, surfactant proteins, and other substances with antimicrobial activity play a crucial role in innate immunity (1).

Pathogens that evade the initial physical and chemical barriers of the innate immune system will encounter its cellular component. The innate immune system's initial reaction is set off by specific types of cells, such as myeloid cells, including macrophages and granulocytes, and innate lymphoid cells (ILCs), such as natural killer cells (NKs). The activation of the innate immune response is triggered by the recognition of pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs) by pattern recognition receptors (PRRs) on the surface of immune cells. This activation leads to the release of antimicrobial molecules, cytokines and chemokines, as well as the initiation of phagocytic and cytotoxic activities by immune

cells. These innate cellular responses constitute a secondary line of defence against pathogen invasion. The innate immune system tries to disrupt the pathogen during the first few days of infection while the adaptive immune system is getting activated (1, 2).

1.3. Long-term immune response: Adaptive immunity

Adaptive immunity, also known as acquired immunity, is the slower arm of the immune system, allowing it to provide long-lasting protection against future infections with the same pathogen. This type of immunity is in contrast to innate immunity, which is the first line of defence and provides a more general, non-specific response to pathogens (1).

The main players in adaptive immunity are B cells and T cells, which are produced in the bone marrow and thymus, respectively. T cells, or T lymphocytes, can be either CD4+ T cells (also known as T helper cells) or CD8+ T cells (also known as cytotoxic T cells). T helper cell receptors can detect foreign antigens, which are presented by Major Histocompatibility Complexes (MHC) II on professional antigen-presenting cells (APC). The activated T cells destroy virally infected cells by releasing cytokines or inducing other immune cells. Helper T cells stimulate the production of antibodies by B cells and the activation of other immune cells, while cytotoxic T cells directly kill infected cells. B cells, or B lymphocytes, produce antibodies, which are proteins that specifically target and neutralize pathogens. Once activated, T cells and B cells undergo clonal expansion, in which they proliferate and differentiate into memory cells and effector cells. Memory cells remain in the body and provide long-lasting protection against future infections with the same pathogen. Effector cells, on the other hand, carry out the immune response and neutralize the pathogen (3).

There are two main types of adaptive immunity: cell-mediated immunity and humoral immunity. Cell-mediated immunity involves the activation of T cells and is effective against intracellular pathogens, such as viruses and some bacteria. Humoral immunity, on the other hand, involves the production of antibodies by B cells and is effective against extracellular pathogens, such as bacteria and viruses. B cells have their specific receptor to recognize free antigens. When B cells get activated, they transform into plasma cells and release B cell receptors (BCR) in the form of antibodies which could bind and nuteralize antigens via several mechanisms (1).

1.4. Antigen-specific B cell response: Antibody

Activated B cells differentiate into two types: plasma cells and memory B cells. The plasma cells produce antibodies for the current infection, while memory cells retain the memory of the initial infection and can trigger a faster response upon future exposure to similar pathogens. The antibodies produced bind to antigens located on the surface of cells or in the extracellular space. They neutralize toxins or mark the microbes, infected cells, and cancer cells for removal by phagocytes or go through a process called antibody-dependent cell-mediated cytotoxicity (ADCC) by another type of white blood cells known as natural killer cells (4, 5).

Antibodies, or B cell receptors, contain four polypeptide chains (two heavy and two light immunoglobulin protein chains). Each heavy and light chain divides into a constant and variable region. The variable region (Fab) is the antigen-binding fragment that is unique to each antibody. The constant (Fc), or the effector region, is consistent among the same class of antibody (6) (Figure 1. Antibody structure). Antibodies, also called immunoglobulin (Ig), are categorized into five classes according to their heavy chain constant region, including alpha (IgA), delta (IgD), epsilon (IgE), gamma (IgG), and mu (IgM). IgM and IgD are generated early in the adaptive immune response, while IgA, IgG, and IgE are produced following a genomic alteration in B cells. During the course of the immune response, activated B cells receive environmental differentiating signals from T cells and undergo a process called alternative RNA splicing and class switch recombination (CSR) to switch antibodies' effector function and generate IgD, or IgA, IgG and IgE (6). IgM forms large pentameric or hexameric molecules, which have high avidity for antigens and can stimulate complements. Monomeric IgG and IgE, or dimeric IgA, can pass into the extravascular space and do their functions (7). IgE is generally effective against large extracellular parasites and involved in allergic reactions (8). IgA mainly acts against pathogens in the mucous membranes, and IgG is effective against viruses and most bacterial infections (9, 10).

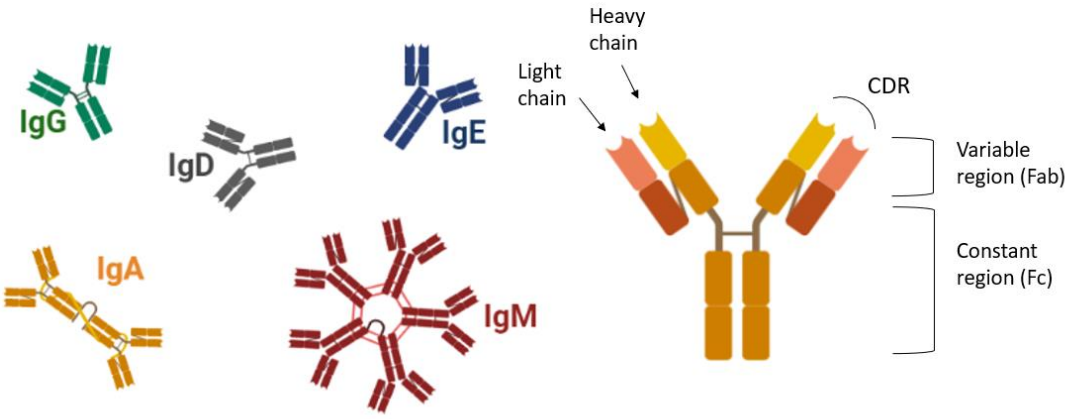


Figure 1. Antibody structure.
The structure of an antibody (secreted B cell receptor) class IgG is on the right. The heavy chain and light chain are shown in different colors. The variable regions generate CDR (complementary determining region). Different immunoglobulin classes, IgG, IgD, IgA, IgE, and IgM, are shown on the left side.

1.5. Antibody genes and primary diversification

The adaptive immune response can generate millions of unique antibodies to specifically recognize different antigens. Therefore, BCR genes have multiple gene segments that can rearrange in various orders, resulting in particular Ig genes in mature B cells and translated to specific antibodies. BCR immunoglobulin locus comprised variable (V), diversity (D), and joining (J) gene segments, which are book-ended by Recombination Signal Sequence (RSS). Blunt double-strand breaks between the coding region and RSS triggers by Recombination Activating Gene (RAG) products, RAG1 and RAG2, leading to a process called V(D)J recombination in this region. In this process, two gene segments form a synapsis with the help of non-homologous end joining (NHEJ) repair mechanisms, and the intermediate DNA is removed (1). The first layer of antibody diversification occurs by transcription through different combinations of V(D)J products resulting in unique IgM and/or IgD antibodies. IgM and IgD are created by modifying RNA transcripts that generated from the Igh locus. This modification process is known as alternative splicing. As a result, IgM and IgD have a similar N-terminal antigen-binding variable region. However, they differ in their C-terminal constant region, which is encoded by either Cμ or Cδ exons for IgM or IgD, respectively. After V(D)J recombination, those immature B cells which complete an efficient self-antigen tolerant

combination of their BCRs migrate to the spleen for further differentiation; they will then migrate to lymph nodes. Mature B cells, which are exposed to the antigen in lymph nodes, get activated and start proliferation and produce antibody secreting plasma cells. These plasma cells secrete the initial response IgM antibodies or undergo secondary antibody diversification through somatic hypermutation (SHM) to produce higher affinity antibodies, and class switch recombination (CSR) to generate specific antibody isotype including Ig A, Ig G and Ig E (1, 11).

1.6. Affinity maturation of antibodies through SHM and CSR

SHM is a process of antibody diversification that helps the generation of high-affinity antibodies. The variable regions of light and heavy chains of activated BCR undergo a high rate of point mutations by an enzyme called activation-induced cytidine deaminase (AID). Mutated BCRs are evaluated in the germinal centers of the lymph nods, and those with higher affinity to the recognized antigen survive. Environmental factors driven by T cells and other surrounding cells can also trigger activated B cells to switch antibody effector functions through CSR. Constant regions of the Ig heavy chain are located downstream of the VDJ segment, and switch regions located between the constant regions as there is a switch region upstream of each constant gene (1). (Figure2)

The Ig constant region exons (Cμ, Cδ, Cγ3, Cγ1, Cα1, Cγ2, Cγ4, Cε, Cα2) transcribe to antibodies' Fc region and determine antibodies' isotype. The non-coding regions, referred to as "switch regions" (S regions), are located upstream of each constant region. These regions contain tandem sequences that are particularly rich in sequences targeted by AID on both the sense and non-sense strands. AID mutations in the switch regions can lead to double-stranded breaks in DNA, which are repaired by synapsis to another switch region and replacement of a constant region downstream with the Cμ region, resulting in the removal of the intervening DNA segment. This leads to the replacement of another constant region downstream of VDJ genes, resulting in a different Fc region (IgA, IgE, or IgG) with the same Fab area specific for the same antigen. Without AID-mediated processes such as SHM and CSR, the immune system produces low-affinity antibodies or only generates IgM and IgG, respectively. AID is crucial for both SHM and CSR and for production of high affinity antibodies (1, 12).

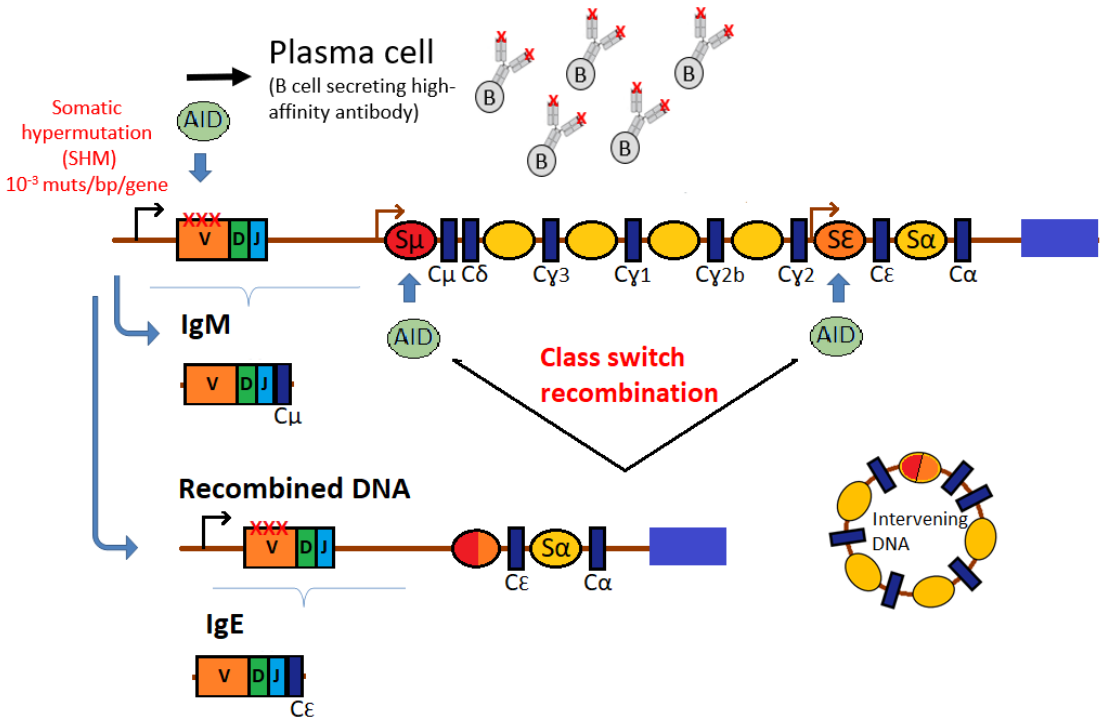


Figure 2. SHM and CSR in Ig heavy chain locus.
The immunoglobulin heavy chain locus was shown. The VDJ region assembled from the recombination of variable(V), diversity(D) and joining(J) segments coding the variable region. AID induces a high rate of single point mutations on variable regions of both heavy chain and light chain through somatic hypermutation (SHM). In the constant region, segments of repetitive sequences called switch regions (ovals) are located before each constant gene (blue rectangles). AID activity on switch regions facilitates the substitution of Cμ with downstream constant region genes through a process called class switch recombination (CSR). This reformation results in the synthesis of secondary antibody isotypes (as shown for IgE).

1.7. Activation induced cytidine deaminase (AID): expression and biochemical function

AID is a member of the Apolipoprotein B mRNA-editing catalytic component (APOBEC) family of DNA/RNA editing enzymes. The AID/APOBEC family includes AID, APOBEC1, APOBEC2, APOBEC3(A-H), and APOBEC4 (13). AID expression in B cells can occur through two pathways: T cell dependent activation and T cell independent activation. T cell dependent activation occurs through the interaction of the peptide/MHCII complex and CD40 on B cells with T cell receptors (TCR) and CD40L on T helper (TH) cells. T cell independent activation takes place by dual engagement of B-cell receptors and Toll-like receptors on B cells with antigens like lipopolysaccharides (LPS) (14, 15). Both pathways result in similar AID expression and CSR induction, with

the T cell dependent pathway having a peak increase of 100-fold within 24-48 hours of activation with TLR ligands in mouse B cells. This suggests that T cell independent activation of B cells occurs early in the immune response and before TH cell assistance is available (15).

AID expression has been primarily identified in activated B cells in the lymph node and spleen, which are the main sites of T cell dependent B cell activation. T cell dependent B cell activation mainly takes place in the conventional germinal centers located in the lymph nodes of mammals and in the spleens of birds(16, 17). In reptiles and amphibians, T cell dependent activation of B cells occurs in the spleen, while melano-macrophage clusters have been identified as the main site of AID expression in early gnathostome vertebrates and most fish species (18, 19). Low levels of AID expression have also been reported in various tissues such as the thymus, pancreas, kidney, liver, and lungs of mammals (16).

AID binds and deaminates single-stranded DNA but does not act on double-stranded DNA or RNA. It favors small bubble structures of 5-9 nucleotides (20, 21). Despite the strong binding affinity of AID for ssDNA (nM amount), the catalytic rate of AID is slow, resulting in only one mutation every few minutes (22, 23). This enzyme is mostly expressed in activated B cells to initiate secondary antibody diversification (24). It can access single-stranded DNA generated during the transcription of select genes in the Ig locus. AID prefers WRC motifs (W = T/A and R = G/A) and deaminates deoxycytidine(dC) to deoxyuridine(dU). The activity of uracil-N-glycosylase (UNG) deletes the uracil base and leaves an abasic site which could be repaired by Base Excision Repair (BER) or Mismatch Repair (MMR), known as error-prone mechanisms are employed to mutate antigen binding regions. This process causes C to T mutations of the variable region during SHM (25). However, in CSR, AID targets two closely spaced regions in both DNA strands (both strands are AGC-rich in switch regions), creating abasic nicks through UNG activity. These nicks are then processed by exonucleases, resulting in a double-stranded break (DSB) at the proximal sites. Alternative-end joining repair mechanisms then connect previously unlinked gene segments and excise the intermediate DNA (25).

The AID deamination activity initiates by zinc activated water molecules in form of zinc hydroxide directly attacking the carbon 4 of the deoxycytidine pyrimidine ring and

break the double bond. This nucleophilic attack is facilitated by the protonation of N3 (of the pyrimidine ring) interacting a glutamate in the AID active site. The side chain of glutamate transports protons from zinc hydroxide to the leaving ammonia group (NH3), which results in the substitution of a carbonyl group on the pyrimidine ring. Another water molecule which coordinates zinc ion protonates the glutamate carboxyl group and restores the catalytic site (13, 21, 26). (Figure 3)

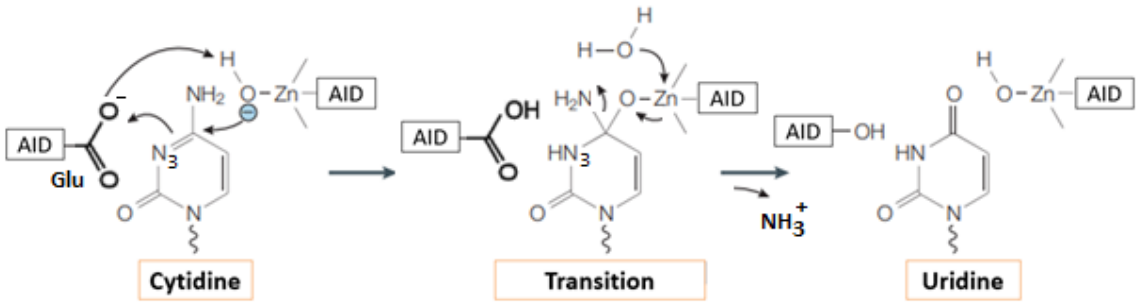


Figure 3. AID cytidine deamination.
AID's catalytic zinc ion is bonded to catalytic residues and a water molecule. The water molecule charged by the zinc ion facilitates a nucleophilic attack on C4 of the pyrimidine ring producing a transitional tetrahedral. The catalytic glutamic acid deprotonates the hydroxyl group resulting in the formation of a double-bonded oxygen. The amino group is removed by deprotonation of glutamic acid once more and recovers the AID.

1.8. AID's structure reflects its function

The AID/APOBEC enzymes have three primary functional elements in common. One is the catalytic domain, which includes the pocket that participates in the substrate binding surface. Some of these enzymes also have a region for cofactor interaction that allows for multimerization. Lastly, there are elements in the sequence that determine where the protein is located within the cell. Changes in the sequence or structure of any of these elements can impact the preferred nucleic acid, either through slight modifications in the substrate binding groove or through limited subcellular localization as a result of cofactor interactions or self-oligomerization (27).

AID is a small enzyme composed of ~200 amino acids (198 amino acids in human) and a zinc ion in its catalytic site. The catalytic site contains two cysteines and one histidine that coordinate a zinc ion with the help of a water molecule and a proton-donor glutamic acid. This globular protein has an overall positive surface charge and strong binding affinity (~nM) for ssDNA, which has a negative charge (22, 23). AID's

structure was found to have a bifurcated substrate binding surface, rather than the typical U-shaped substrate binding groove observed in A3A and A3G enzymes (21, 28). Loops 1, 3, and 7 are crucial in creating the channel for the dCMP to coordinate, which is linked to a second groove referred to as the assistant patch. The substrate channel, also known as DNA binding groove 1, is primarily composed of the l2 (α1-β1 loop), l4 (β2-α2 loop), and l8 (β4-α4 loop). The assistant patch refers to the group of positively charged amino acids in α6. Basic, positively charged residues in these channels form a binding surface that is separated by negatively charged residues in loop 7 near their point of convergence (21). It has been predicted that the DNA binding groove 2 starts at the junction of l2 and l4, crosses the catalytic pocket, and then proceeds between α2 and α3 (29). The substrate specificity in the AID/APOBEC family is established by the interaction of l2 with the +1 position and l8 with the bases at -1 and -2 positions relative to dC (30). The residues in the grooves are conserved in AID proteins from different species but not in other APOBECs, which demonstrates that this structure is unique to AID (28).

The shape of the substrate binding region in AID is consistent with recent studies that suggest G-quadruplex structures may play a role in guiding and directing AID in the context of immunoglobulin class switch recombination (21, 31). Forming an oligomeric AID complex is critical for clustered mutations in CSR. The mammalian Ig S region sequences could form G4 structures which are preferred for AID binding and cooperative oligomerization (21, 23). This information highlights the significance of the substrate binding groove's structure in allowing AID/APOBECs to distinguish between substrates based on their secondary structure.

AID can bind RNA, particularly in RNA-DNA hybrids, but cannot modify it, indicating that binding alone is not enough to initiate catalysis (32). In the crystal structure of AID/dCMP (PDB: 5W0U), there is an interaction between R25 and the 5' phosphate, hydrogen bonding between N51 and the 3'-OH, and interaction between Y114 and O5'. When RNA is replaced by DNA, there is steric hindrance between R25 and the 2'-OH in ribose, preventing AID from acting on RNA despite its ability to bind to it. This highlights the importance of the proper interaction with the 5' phosphate for placing the dC into the catalytic pocket for effective AID activity (21). Collectively, AID contains five substrate binding motifs, including ssDNA binding groove 1 and groove 2, which pass through the catalytic site, an assistant patch on the back (identified by

structural analysis), and two RNA binding motifs. One RNA binding region was identified with mutagenesis assays, and one RNA binding region at the dimeric interface was proposed by computational simulation studies (33). Numerous combinations of these ssDNA and RNA binding motifs can be used to yield substrate binding. (Figure 4)

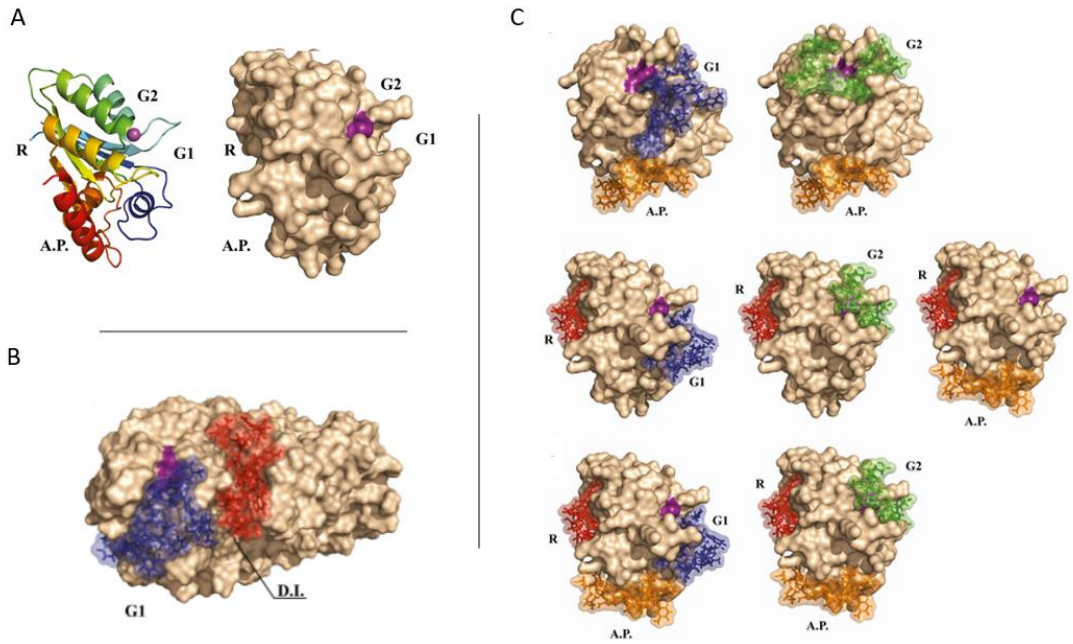


Figure 4. AID structure and DNA, RNA binding motifs.
A. The structure of human AID structure showed in ribbon diagram (left) and surface structure. DNA binding groove 1 (G1), groove 2 (G2), assistant patch (A.P.), and the RNA binding groove (R) are displayed. The purple surface shows the catalytic site. B. Second RNA binding region in AID dimeric interface (D.I.). ssRNA (red) binding dimeric interface while ssDNA (blue) binds groove 1. C. Different binding modes of ssDNA and ssRNA in AID. ssDNA is colored blue, green, and orange, binding G1, G2, and A.P. respectively. ssRNA was colored red, binding to the RNA binding groove. Figure from King et al., 2020 and used with publisher permission.

1.9. AID regulation and cancer

Since the discovery of AID, many factors have been defined for AID targeting, including the target sequences, transcription, cis elements, chromatin markers, and protein co-factors (23). The role of co-factors has been challenged due to the small size of AID, its positively charged surface, and the distribution of proposed co-factors at Ig genes (34). King et al. (2015) used a combination of computational simulations and biochemical experiments to predict AID structure and explain its regulation as an internally built-in pocket occlusion. They proposed two grooves passing the catalytic

pocket that contain positively charged residues. The majority of AID and ssDNA bindings occur randomly on AID's positively charged surface, which does not pass over this catalytic pocket and is, therefore, non-productive. They proposed dynamic pocket closure as an internal regulator of AID activity such that only less than 1% of bindings to ssDNA result in cytidine deamination (34, 35). This predicted structure was later confirmed in another study that produced a crystal structure of a truncated near-native AID and revealed its preference for recognition and deamination of structured G quadruplex substrates (21).

Although AID expression and activity are tightly regulated, it may act off-target on non-Ig genes and cause somatic mutations, cell transformation, tumor development, and chromosomal translocation (13, 36). AID mutations on other genes may cause DSB similar to what happens in CSR and connect the Ig gene to unrelated genes, resulting in chromosomal translocations. Gene translocation can cause oncogenes like C-Myc to link to the regulatory factors of highly active genes and generate novel oncogenic genes. AID activity has been identified as a significant contributing factor in oncogene translocations such as IgH-cMyc translocations in Burkitt's lymphomas (37). AID mutations were correlated to B cell lymphomas, T cell lymphomas, non-Hodgkin lymphomas, multiple myeloma, and chronic lymphocytic leukemia (CLL) (38-43).

1.10. The role of transcription and R-loop formation in AID activity

The early studies after AID discovery suggested the importance of active transcription for AID activity since ssDNA is provided during transcription (44, 45). In addition, many studies proposed that transcription contributes to AID targeting through the association of RNA polymerase (RNAP) complex or transcription machinery cofactors such as RNAPII, the transcription elongation factor Spt5, spliceosome-associated factor CTNNBL1, the ssDNA binding protein Replication protein A (RPA), RNAPII associated factor I (PAF1), splicing factor SRSF1-3, splicing regulator polypyrimidine tract binding protein 2 (PTBP2), RNA binding heterogeneous nuclear ribonucleoproteins (hnRNP) (46-53). However, none of these co-factors could truly explain AID targeting to Ig loci considering its non-specific genome-wide activity. Further investigations indicated the unique features of transcription itself, which explains AID genome targeting. AID's preference for targeting Ig loci is associated with strong

transcription enhancers and convergent transcription (sense and anti-sense transcription at the same time) of this region, which also overlaps with AID's off-target activities (54, 55). Transcription not only generates ssDNA directly but also facilitates AID targeting through de-chromatinization, which provides accessible naked DNA. AID can target and bind short bubbles of ssDNA generated on breathing dsDNA upstream of the traversing RNA polymerase as well as DNA/RNA hybrid formed during the transcription (32, 56). Nucleosomes, physically prevent AID binding to DNA, clear out by unwinding the DNA for initiation of transcription (57, 58).

Transcription through the V and S regions of the Ig locus plays an essential role in AID targeting, which leads to SHM and CSR. Switch regions in mammalians include tandem repeats of G-rich sequences. Transcription through switch regions forms a stable RNA: DNA hybrid including the template strand and transcribed RNA and a non-template single strand DNA. This secondary structure, called R-loop, is more stable in regions containing a highly G-rich non-template strand (59). However, R-loop formation can also initiate by a single-stranded DNA nick upstream of a G-rich sequence (60). The long stable R-loop formation in the switch regions provides single-stranded DNA to attract AID, therefore making a more productive CSR (59). The ssDNA generated during transcription could yield transient secondary structures due to torsional strain and chromatin remodeling such as bubbles, stem-loop, supercoiling on both downstream or upstream traversing RNA polymerase or even more complex G-quadruplex like structures (61-63). These structures could provide a more stable ssDNA target or potentially pause RNA polymerase (when encountering altered structures) and extend DNA accessibility (64, 65). Branton et al. have developed a novel coupled transcription-AID activity assay called degenerate PCR assay, allowing for the precise observation of all AID-mediated mutations without any selectable functional consequences of the presence, absence, load, or location of the mutations. Using this assay, they confirmed that AID could effectively target supercoiled dsDNA (dechromatinated) without the need for transcription to occur. Additionally, they found that AID can also act on relaxed linear dsDNA when transcription is not present and that even the optimal transcription conditions only lead to a slight improvement in AID activity on supercoiled dsDNA (56). Thus, AID is able to mutate both DNA stands (transcribed and non-transcribed strands), targeting different DNA topologies generated during transcription as well as targeting

supercoiled and relaxed linear dsDNA in the absence of transcription (58, 66, 67). (Figure 5)

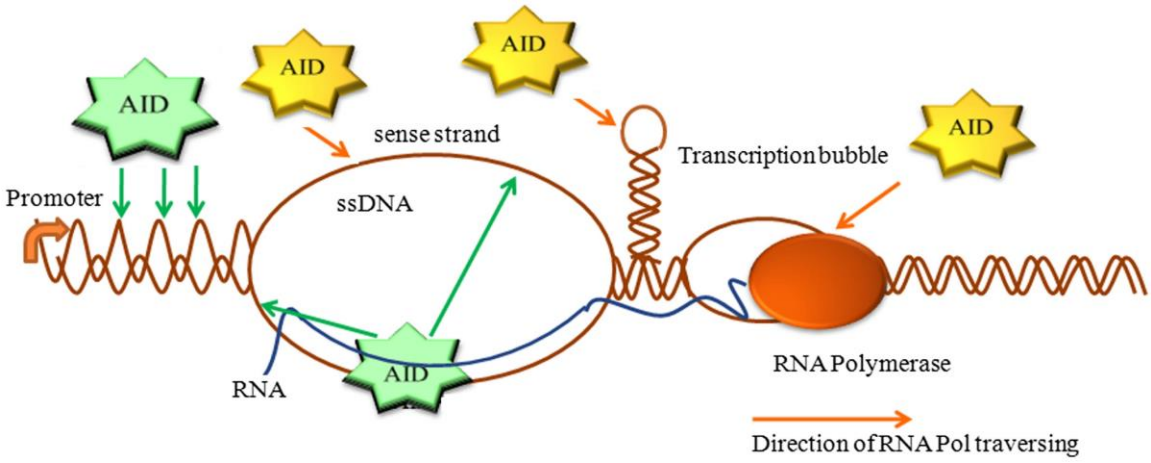


Figure 5. AID activity in the context of transcription.
AID targets accessible ssDNA in the forms of R-loops, bubbles, or stem-loops generated during the transcription (AID is colored yellow). AID targets the breathing dsDNA following transcription, which is stripped of chromatin and has an extremely high affinity for DNA/RNA hybrid of switch region sequences found at the Ig loci (AID was colored in green). Figure from Branton et al. 2020 and used with publisher permission.

Although transcription enhances AID targeting, the germline switch transcripts do not code for proteins, and their exact role in CSR must be identified. Previous studies proposed that the intronic switch RNA is implicated by RNA debranching enzyme, DBR1, and eventually forms G-quadruplex structures. These G-quartet structures bind AID and mediate targeting to the complementary switch region (31). Our previous study demonstrated that AID could act on different sequences of DNA/DNA or DNA/RNA hybrid forming a bubble substrate including a G-rich sequence of Ig switch region. In spite of reduced deamination activity on substrates similar to the S regions sequence, AID bounds to DNA/RNA hybrids of the S region with the highest affinity (32).

1.11. AID in non-human species demonstrates different biochemical features

It is believed that the AID/APOBEC family appeared in the vertebrate class through the evolution of AID-like and APOBEC4-like enzymes which have a water molecule or a cysteine residue as their fourth zinc coordinating agent (68, 69). PmCDA1 and PmCDA2 (AID-like) evolved in jawless vertebrates such as lamprey. However, AID-like enzymes gave rise to AID and APOBEC2 in the root of jawed vertebrates and

diverged into APOBEC3 and APOBEC1 in tetrapods and mammals, respectively (69). PmCDA1 is involved in the diversification of lamprey immune receptors. The fact that AID plays the same role in jawed vertebrates as in lamprey, shows that AID-like enzymes have had this function since before the diversification of vertebrate species (69, 70).

Previous studies on AID orthologs of divergent species demonstrated that AID exhibits certain conserved structural and biochemical characteristics, such as a slower catalytic rate when compared to its strong binding affinity to single-stranded DNA. However, it has also been observed that there are notable variations in the catalytic rate and thermosensitivity of AID among different species (29, 70, 71). In particular, studies have shown that the AID in amphibians and bony fish is more adaptable to cold temperatures compared to that in mammalian and avian species which might be a result of adaptation to their habitat temperature. Additionally, various AID orthologs have been found to have different ranges of catalytic activity at their optimal temperatures (69-73). The AID in zebrafish, for instance, is found to be more catalytically active at its optimal temperature than human AID, while catfish AID is substantially less active (71, 74). Furthermore, Zebrafish AID have the ability to deaminate 5-methyl cytosine (5mC), although it does so less efficiently than it does for cytosine (dC). Additionally, AID-mediated deamination of 5-mC has been observed in various cell types and organisms, such as induced pluripotent stem cells, primordial germ cells, B cells, cancer cell lines, bovine embryos, and zebrafish embryos. This observation led to the suggestion that AID may participate in other cellular processes, such as epigenetics and genetic reprogramming. However, this hypothesis has been called into question by a growing amount of contrary evidence afterwards (74-78). Given the role of AID in antibody diversification, the biochemical variations in AID may account for the different immune responses among species.

AID is considered a crucial component in the process of diversification of antibodies in jawed vertebrates. Despite this, exceptions have been found in the anglerfish and pipefish, where the AICDA gene that codes for AID is missing. This is believed to be related to their unique reproductive behaviors, which require reducing immunity to prevent tissue rejection (79, 80). In addition, a new case has been found in the bony fish clade, where AID activity has compromised but is not related to reproductive behaviors. Instead, these species have changes in their innate and cell-

based immunity. The immune response of Atlantic cod (a member of the Gadiformes group) was found to be intriguing in functional analyses, as it showed high levels of low-affinity IgM and an absence of strong antigen-specific antibody response after immunization. However, other studies have claimed that the cod's response to certain pathogens, such as Aeromonas salmonicida and different serotypes of Vibrio anguillarum, was comparable to that of salmon and resulted in specific antibody responses. These antibodies were mainly specific to lipopolysaccharides (LPS), and some of the serum pools reacted to O-polysaccharides, suggesting B-cell activation through the T-independent pathway. Despite the loss of central genes required for T-dependent B-cell activation, evidence of TI B-cell activation has been reported in this species. In general, T-dependent activated B cells produce highly pathogen-specific antibodies, and the lack of the TD pathway in cod is consistent with the significant remodeling of its immune genes. It seems the loss of AID function in Gadiformes is linked to adaptation to cold temperatures, where the gene for AID is present, but the protein compromised its activity (69, 81).

1.12. Co-evolution of AID with Ig gene

Research has found that the process of antibody affinity maturation, where antibodies become more effective at targeting specific antigens, emerged as far back as in cartilaginous fish. This was determined through the analysis of antibody repertoire in various vertebrates and by identifying AID-mediated mutations in the CDRs of the Ig genes (82-86). A study on frogs showed that after being immunized with DNP-KLH (a highly immunogenic antigen), there was 5-10 times increase in antibody affinity after four weeks. Point mutations were also found in the VH1 region of the frogs but at a lower rate than that observed in mice (84, 87). In another study, the role of AID-mediated mutations in the diversification of antibodies in zebrafish (*Danio rerio*) was verified through a mutational analysis of a healthy individual's IgL cDNA library. This analysis found that the WRCH/DGYW motifs in the CDR regions were the main target of mutations (88). Similarly, high rates of somatic mutations in the Ig genes of nurse sharks have been reported, which can result in a 10-fold increase in antibody affinity (83). The occurrence of SHM in the Ig genes of vertebrates and its contribution to antibody affinity maturation is conserved to some extent, although the degree of Ig mutations varies among the studied vertebrates.

Biochemical analysis and in vivo studies have identified that AID prefers to target WRC motifs, particularly AGCT, in the Ig variable region. Mammalian Ig variable regions are quietly enriched with WRCH/DGYW sequences, identified as AID hotspots, and the recurrence of these motifs is associated with SHM frequency (86, 89, 90). Therefore, it was suggested that AID substrate specificity has co-evolved with the Ig variable region among species (23). Studies on mammals, avians, amphibians, and fish species demonstrated this co-evolution (71, 73, 91, 92). The human CDR1 and 2 regions contain a higher frequency of overlapping AID and Polη hotspot motifs than other framework regions of IGHV3-23. Additionally, the substitution of AID hotspots with neutral sequences decreases mutations in the entire V region, including CDR1 and 2 (91). Furthermore, 4 out of 7 of the residues that mainly contribute to antibody-antigen binding are coded by mutated AGY(WRC) codons in the CDRs (93). The co-evolution of Ig genes and AID substrate specificity might help to explain the AID targeting and regulations towards Ig genes.

1.13. Objectives:

In this thesis, we aim to study the structural plasticity of AID by a multi-dimensional approach to enzyme characterization using both computational techniques and biochemical assays on broad evolutionary dimensions. This thesis includes two main hypotheses discussed in separate chapters.

In chapter 3, we hypothesized that despite conserved core structure, AIDs from divergent species reveal specific structural features which could result in specificity for a different type of DNA substrate.

In chapter 4, which is a tangent of chapter 3's outcome, we proposed that AID has RNA binding specificity, and these RNAs potentially play a regulatory role in its cytidine deaminase activity in CSR.

The material and methods used in these two chapters are fully explained in the next section.

Chapter 2. Material and methods

2.1. Structure prediction of AID orthologs

The AID structures were predicted in the same manner described in King et al. (2015), four available PDBs of recently resolved human AID structure, including MBP fused human AID (PDB: 5W0Z), MBP fused AID in complex with dCMP (PDB: 5W0U), MBP fused AID in complex with cacodylic acid (PDB: 5W0R), and MBP fused AID in complex with cytidine (PDB: 5W0C) were chosen as the template for homology modeling using default parameter setting of I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (29, 94). The PDB structures were downloaded from http://www.rcsb.org. A total of 240 models were created by generating five models for each of four template structures resulting in 20 models of each AID ortholog (12 AID orthologs). The predicted structures were evaluated visually to assess their maintenance of the core structure and the presence of the catalytic pocket using PyMOL v1.7.6 (http://www.pymol.org). The catalytic pocket was defined based on the zinc coordinating space and catalytic residues equal to H56, C87, C90, and E58 in Human AID. The open conformations with a higher confidence score (C-Score >0.1) of each AID were evaluated using a Ramachandran plot to measure the quality of the predicted structure based on the stereochemical angles of each individual residue (95). The models with 75-90% residues in the favored regions were selected for each ortholog and subjected to PROPKA 3.0 (PARSE force field) to calculate the isoelectric point and protein surface charge (96).

2.2. AID expression and purification

The sequence of original reading frames (ORFs) of AID orthologs was synthesized in the pBluescript II plasmid (Genscript, US) containing EcoRI fragment in the 5’ end and NotI fragment in the 3’ end. The GST-AID expression vector of Human (Homo sapiens, Hs), zebrafish (Danio rerio, Dr), channel catfish (Ictalurus punctatus, Ip), salmon (Salmo salar, Ss), and Nurse Shark (Ginglymostoma cirratum, Gc) were obtained from previously published studies (70, 81).

The EcoRI and NotI restriction enzymes (1 mg DNA, ~10 units of EcoRI and NotI (New England Biolabs Inc. (NEB)), 1x reaction buffer (NEB) were incubated at 37 °C for

1 hour to digest ORFs from their respective plasmids. A home-made loading dye (60% glycerol (Fishersci), 60 mM EDTA, 10 mM Tris-HCL, pH 7.6, 0.03% Bromophenol Blue, and 0.03% Xylene Cyanol) was added to the digested DNA and run on a 1% agarose gel (1% agarose (ThermoFisher), TAE buffer, 5% SYBR safe (Invitrogen)). The ORF band was separated at 600–700 bp, which was visualized via blue light. A sterilized scalpel was used to cut and excise the ORF band and, the DNA was purified using Qiaquick Gel Extraction Kit (Qiagen). The DNA fragments containing ORFs were inserted into the pGEX-5X-3 (GE Healthcare, USA) to produce a GST-AID expression construct. The vector and ORF were mixed in 1:5 ratio with 400 units of T4 DNA Ligase (NEB), 1x ligase reaction buffer (NEB), and incubated at 16°C for 16 hours. Five µl of the ligation reaction transformed to 50 µl of DH5α competent cells, incubated on ice for 20 minutes, heat shocked in 42°C water bath for 35-50 seconds and restored on ice for 1 minute. Then 250 ml of Luria broth (LB) was added and restored for 1 hour in a 37°C shaker at 225 rpm. Half of the suspension was plated on LB agar + ampicillin (AMP, 100 µg/mL) and grown overnight at 37°C. The growing colonies that contain insert fragments were verified by PCR screening using PGEX5 and MB101 as forward and reverse primers. The PCR reactions were run on a 1% agarose gel, and colonies of the correct size were selected. A bacterial suspension of the chosen colonies (producing correct PCR product) was then inoculated into 5 mL of LB + AMP and incubated at 37°C with 225 rpm for 16 hours. The Geneaid High-Speed Plasmid Mini kit was used to extract the DNA plasmids. Plasmids with the correct size and direction of insert sequences were verified by Sanger sequencing (Genewize).

Three correct plasmids of each AID ortholog were transformed into *E. coli* BL21 (DE3) cells for expression. Two correct expression constructs of each AID ortholog were then subjected to expression. The expression was induced by adding 1 mM IPTG (isopropyl β-D thiogalactopyranoside), and the cells were incubated for 16 hours at 16°C. The cells were harvested by centrifugation (7000xg for 10 minutes) and resuspended in 1xPBS. Two ng/µl of RNase A was added when present (only for some of the Hs-AID samples used in Chapter 4) and incubated for 30 minutes at room temperature. The cells were lysed using a French pressure cell press (Thermospectronic) at 4°C twice. After centrifugation (12 k xg for 10 minutes), the supernatant solution containing the expressed protein was applied to Glutathione Sepharose high-performance beads (Amersham), and the flowthrough was reapplied to

the column again and washed at least twice with 20 mL of 1x PBS (phosphate-buffered saline) (Sigma). Thereafter, the GST-tagged protein was eluted in a reduced glutathione solution (0.05 M Tris, pH 8.0, and 0.01 M L-glutathione) in small fractions. The amount of protein in each fraction was measured on NanoDrop (ThermoFischer NanoDrop 2000 spectrophotometer), and the highest 3-4 fractions were selected for dialysis. The sample was dialyzed through a SnakeSkin 22mm Dialysis Tubing (ThermoFisher Scientific) in 1L of storage buffer (20 mM Tris-Cl, pH 7.5, 100 mM NaCl, and 1 mM dithiothreitol (DTT)) for 16 hours at 4°C, and the fresh dialysis buffer was used for another 2 hours dialysis. The samples were aliquoted, flash frozen by liquid nitrogen and stored in -80 freezer.

For the preparation of eukaryotic AID from HEK 293T cells, we inserted the GST-AID construct into the pcDNA3.1-Topo vector and transiently transfected 40 plates of HEK 293T cells (seeded with 5×10^5 cells) using 5 g of plasmid and the Polyjet reagent (Froggabio). After three days of incubation at 37°C, the cells were collected and lysed using a French Press. The cell lysate were purified using Glutathione Sepharose high-performance beads (Amersham), as previously explained for the prokaryotic system, in 20 ml of 1xPBS (phosphate-buffered saline, Sigma).

For AID purification from yeast (Pichia pastoris), the GST-AID was inserted in the pPICZ vector. The vector transformed in Pichia pastoris cells using electroporation and spread the cells on YPDS plus Zeocin plates (1% yeast extract, 2% peptone, 2% dextrose, 1 M sorbitol, 2% agar, 100 µg/ml Zeocin) and grow for two days at 28°C. Colonies were screened to select those with the highest expression level of GST-AID. Colonies were cultured in BMGY medium (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% YNB, 4×10^5 % biotin, 1% glycerol) overnight. The cells were harvested, and the media was replaced with BMMY (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% YNB, 4×10^5 % biotin, 0.5% methanol) to induce expression. The GST-AID expression was evaluated with western blots (polyclonal rabbit anti-V5 antibody and goat-anti-rabbit antibody (Abcam) were used as the primary and secondary antibodies for AID, respectively). The highest level of GST-AID expression was identified after 27 hrs incubation of the selected colonies, as shown in Figure 6 B. The selected colony was cultured in BMMY media at 28-30°C, and the cells were harvested after 27 hrs. The French pressure cell press was used for lysing

the cells, and the glutathione Sepharose beads column was used for the purification of GST-AID, as explained earlier.

The quality and concentration of each purified GST-AID were evaluated on Coomassie blue stained SDS-PAGE gel (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) compared to bovine serum albumin (BSA) standard. To prepare the SDS-gel, 8% acrylamide resolving gel containing 8% acrylamide/bis solution, 19:1 (BioRad); 0.1% SDS; 0.375 M Tris-HCl, pH 8.8; 0.066% APS (ammonium persulfate); and 0.04% TEMED all from Sigma-Aldrich and a 4% stacking gel was added to the top. The stacking gel was made of 4% acrylamide/bis acrylamide, 19:1; 0.1% SDS; 0.125 M Tris-HCl, pH 6.8; 0.066% APS; and 0.08% TEMED. Four different concentrations of purified samples were mixed with 4x SSB (4% SDS, 0.1 M Tris-HCl, pH 6.6, 20% glycerol, 10% Beta Mercaptoethanol, and bromophenol blue) and heat denatured in a dry hot bath for 10 minutes at 99 °C then shortly spun down and loaded on the gel. A protein ladder from Thermoscientific was loaded on each gel. 1x SDS running buffer in the BioRad apparatus was applied for running the gels on 60 V for 30 minutes to pass the stacking gel, followed by 100V for approximately 1 hour till the dye was close to the bottom of the gel. The gels were stained with Coomassie blue and imaged (Figure 6 A and C).

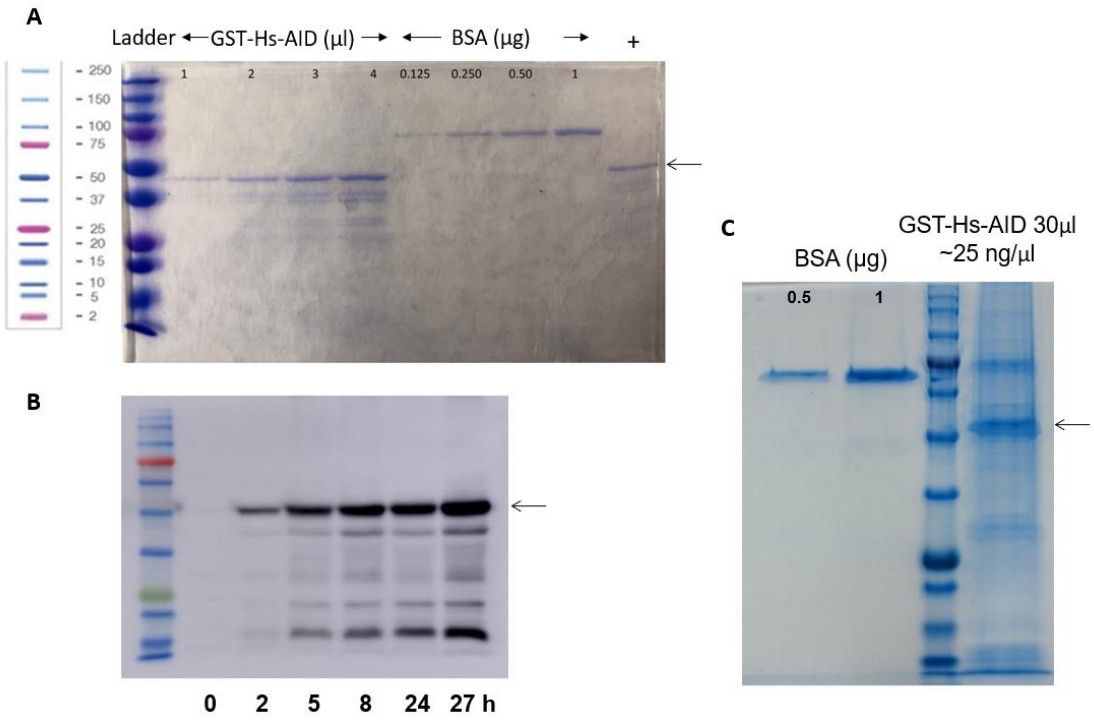


Figure 6. Purification gels of human GST-AID.
The A.SDS-phage gel of purified GST-tagged Hs-AID expressed in *E.coli*. Different amounts of purified samples were loaded to evaluate the purity and concentration of purified protein. 0.125 µg, 0.25 µg, 0.5 µg, and 1 µg of BSA (Bovine Serum Albumin) were used as standard concentrations compared to 4 dilutions of purified human GST-AID (1,2,3, and 4 µL). Positive control in the end right side is from previous preparations. B. Western blot showing GST-AID expression time-course in yeast. C. SDS-phage gel showing purified GST-AID using yeast expression system. The size of GST-tagged AID is ~ 50 kDa, denoted by the black arrow in all gels.

2.3. DNA substrates for deamination assay

The substrates used in this thesis to test AID activity with alkaline cleavage assay are illustrated in Figure 7A. All the oligonucleotides (oligos) were ordered from Integrated DNA Technologies (IDT) via HPLC grade purification. For radioactive labelling, 2.5 pmol of the target strand was labelled with 10 µCi of [γ-32P] ATP using 10 units of T4 polynucleotide kinase (PNK) in 1x PNK Buffer (NEB). It was incubated for 1 hour at 37°C following 65 °C for 10 min to deactivate the enzyme. The 10 µL reaction was diluted with 10 µL of TE buffer and filtered through a mini-Quick spin DNA column (Roche, USA) to purify labelled oligo by 2 min centrifuged at 1000x g. For single stranded substrate, the flowthrough was diluted using TE buffer to 50 fmol/µL. For double stranded substrate, KCl (100 mM final concentration) and three times of the complementary DNA or RNA oligonucleotide (7.5 pmol) were added to the reaction and

then diluted to 50 fmol/μL with TE Buffer. To generate a bubble structure (the sequence is not complementary in the bubble region), the double strand substrate reaction was incubated at 96 °C for 5 min; then, the temperature decreased every 30 sec by 1°C until it reached 4°C. For annealing each of the stem-loop and stem bubble substrates with their complementary sequence, we first incubate them at 98°C for 12 min, then 25°C for 5 min, and then on ice for 2 min. The prepared radiolabeled oligos were then stored at -20 °C. To check the correct formation and integrity of the substrates, ssDNA, DNA/DNA and DNA/RNA bubble substrates were subjected to electrophoresis on native polyacrylamide gel (8%) (Figure 7. B). To check the formation of any of the complex structures (stem loop, stem bubble mix and 2 bubbles), we incubate them with proper digestive enzymes, BamHI and EcoRI, which are able to cut the double stranded complementary DNA section only in case of correct annealing. Then, we ran them on the gel for validation (Figure 7.D).

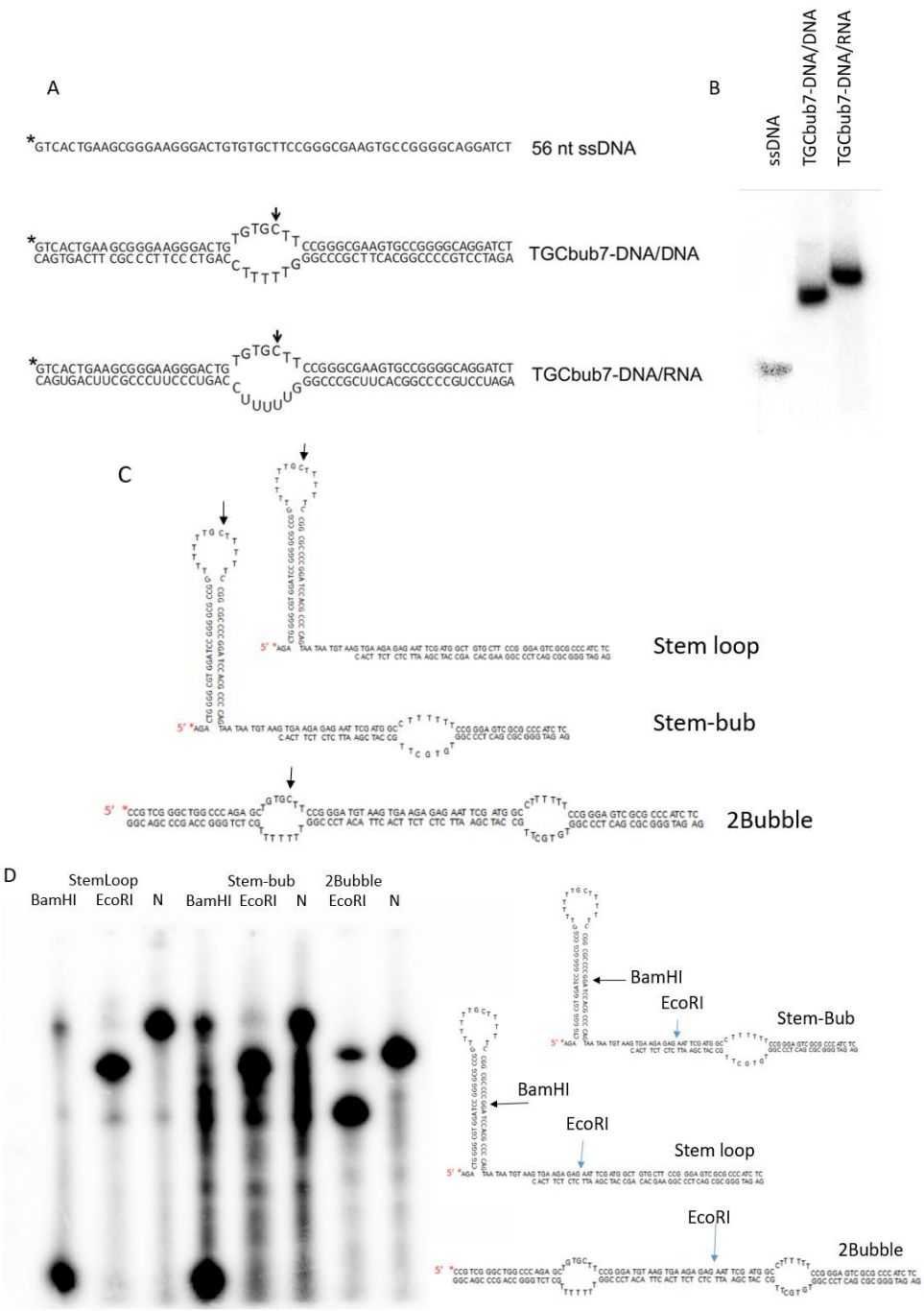


Figure 7. Substrates were used to test AID activity.
A. DNA/DNA and DNA/RNA hybrids formed a 7-nucleotide long bubble containing the target dC that was noted by arrows in a WRC motif (TGC). The 5' labeled strand (shown by *) annealed to the equivalent sequence of DNA or RNA. B. Migration of ssDNA, DNA/DNA and DNA/RNA hybrids were compared on native gel electrophoresis. C. Various DNA structures were used in this study. Stem loop, Stem-bub and 2Bubble structure contain TGC target motif. The gel image shows the running of cleaved substrates after incubation with digestive enzymes.

2.4. RNA extraction from cells

Two plates of HEK 293T cells (seeded with 2×10^5 cells), one transiently transfected to express GST-AID as explained before and one transfected with empty plasmid as the control condition. Also, two plates of MCF7 cells were previously transduced with pInducer21 to express Hs-AID by adding doxycycline to the media, one induced with 1 µg/µl doxycycline and the other one kept as control. All plates were incubated in a 37°C incubator for 2 days.

To extract whole cell RNA from the cell culture, the media was aspirated, and cells were washed gently with PBS. Eight hundred µL of TRIzol (Invitrogen) was added to each plate to lyse the cells. The cells were collected in a 1.5 ml tube, and 160 µL of chloroform was added. After 2-3 min at room temperature, it was centrifuged at 12k xg for 15 min at 4°C. The upper phase, containing RNA, was transferred into a new tube and ethanol precipitated by 400 µL cold isopropanol. Briefly, the tube was stored for 10 min at -20°C and then centrifuged at 12k xg for 10 min at 4°C, the supernatant was removed, and the RNA pellet was air dried for a few minutes. The RNA was finally dissolved in DNA/RNA free water. The RNA amount was measured with Nanodrop and stored in -20°C freezer.

2.5. RNA and DNA oligonucleotides preparation

The RNA oligos of random sequences and different sizes (22-75 bp), including AT-rich (RNA 3), GC-rich (RNA 1) and a sequence of switch region (RNA 4), were ordered from IDT via HPLC grade purification condition. All sequences are shown in table 1. The equivalent DNA sequences of the RNA fragments which we synthesized using the method described in the following section were also ordered from IDT (table 2). All the oligos were diluted to 100 µM, and the final concentration was checked by NanoDrop, adjusted, and stored in -20 freezer.

Table 1. RNA sequences ordered from IDT.

RNA name	Sequence	Length (bases)	Tm °C	GC%
RNA 1	CCUUCGGUACCGGGGGCCGGGCGGGGUAUGCAUACAUGG CGGGGGCCUAUGCAUGGGGGCGUUGCAUCACUCCUC	75	74.6	68
RNA 2	AGAUCCUGCCCCGGCACUUCGCCCGGGUUUUUCCAGUCCC UUCCCGCUUCAGUGAC	56	66.7	62.5
RNA 3	CUCUUCUAUACGAUUAUACUAGAACAUUCUGCAUACAUAAA UAUAAAUUAUGCAAUUAUUUUUGCACUAGUCAUC	75	56.5	26.7
RNA 4	GGGCUAACCUGGGCUGGGCUGAGGGAGGAGGGGUUGGGG CAGGGCUGGGCUGGGCU	56	85.8	73.2
RNA 5	UCACCUCACCAUCCAAACCAAACCAAAAAACCACACAUCUAC CAA	45	67.3	42.2
RNA 6	AUCCACACCUAAACCAACACAAAAAACACCAUCCAAUUCACA AUC	45	63.8	37.8
RNA 7	AGGGUAGGGAGAAUGGGAUGGGAUUUGCUUGUGUGGGAU GGAUGG	45	70.4	53.3
RNA 8	UGAUGGUAAGGUGGAUGUGUUUGCUUGGAGUGAAUGAGGG UUGGA	45	64.8	46.7
RNA 9	UGAUUAGGAUCUGCCAACCGUG	22	-	50

2.6. RNA synthesis

The DNA sequences for the different switch regions (Sα, Sγ1, Sμ), variable regions (VDJ1, VDJ2, VDJ3) and oncogenes (BCl2, BCl6, C-Myc) were obtained from Genbank (https://www.ncbi.nlm.nih.gov/genbank/). For switch regions, the shortest repeat elements of Sα (80bp), Sγ1 (48bp) and Sμ (20bp) and for other genes, a 70 bp sequence downstream of the transcription promotor was selected. Table 2 shows the selected DNA sequences.

In order to produce double strand DNA of each region with a T7 promotor upstream of the target sequence, we design a forward and reverse primer which has 17-20 nucleotides complementary region and forward primer attached to a T7 promotor upstream (5′ TAATACGACTCACTATAG 3′) (Table 3). Each forward and reverse primer was subjected to primer extension using Taq DNA polymerase for 0, 2, 4, 6, 8 and 14 cycles. The reactions were run on a 1.5% agarose gel 100 V for 30 min to find the optimal cycling number and annealing temperature for each construct. Then, 250 μL total of each reaction was produced with the optimal cycling condition. About 6-10 μL of

the reaction ran on an agarose gel to test the existence of desired dsDNA sequence. The rest of the PCR reaction was subjected to ethanol precipitation to extract the dsDNA. This PCR product was used for transcription with T7 RNA polymerase and for producing RNA. 200 µL total reaction containing PCR product, 5xNTP (40 mM GTP, 25 mM CTP, 25 mM ATP, 10 mM UTP), T7 RNA polymerase (NEB) and T7 buffer was prepared and incubated in 37°C water bath for 2-10 hours. 200 µL dye (20% formamide, 20 mM EDTA, 0.25% xylene cyanol, 0.25% bromophenol blue) was added to the reaction, and after 5 min 95 °C incubation, it was loaded on 5% acrylamide denaturing gel (made of 10 mL of 25% 19:1 acrylamide:bisacrylamide, 35 mL 6.667 M urea, 5 mL 10xTBE buffer) and run 28 w (~400 V) roughly 1 hour. The gel was visualized with UV light to find the RNA and cut it from the gel. The size of RNA was checked according to the movement of bromophenol blue and xylene cyanole on the gel.

To extract RNA, the gel was soaked in 0.3 M NaCl solution overnight. 2.5 times ethanol was added to the solution and incubated for 30 min in -20 freezer. Then, the solution was centrifuged at 13000 xg for 30 min, the supernatant was discarded, and the precipitated RNA was dissolved in 20 µl DNA/RNA free water. The concentration of RNA was measured by NanoDrop, diluted to 100 µM and stored in a -20 freezer.

Table 2. Selected DNA and RNA sequences were used in this study.

Source gene	Sequences		Length (bases)	GC%
Sγ1	DNA	GGTGACCCAGCAGAGCAGCTCCAGGGGCGCCAGGACAGGTGGAAGTGT	48	66.7
	RNA	GGUGACCCAGCAGAGCAGCUCCAGGGGCGCCAGGACAGGUGGAAGUGU	48	66.7
Sα	DNA	ATGAGCTGGGATGRRCTGAGCTAGGCTGGAATAGGCTGGGCTGGGCTGGTGTGAGCTGGGTTAGGCTGAGCTGAGCTGGA	80	60
	RNA	AUGAGCUGGGAUGRRCUGAGCUAGGCUGGAAUAGGCUGGGCUGGGCUGGUGUGAGCUGGGUUAGGCUGAGCUGAGCUGGA	80	60
Sµ	DNA	GAGACTGAGCTGGGGTAGCT	20	60
	RNA	GAGACUGAGCUGGGGUAGCU	20	60
VDJ1	DNA	GGGGGGTCCCTGCGACTCTCCTGTGCAGCCTCTGGATTCACTTTTAGCCGCTATGCCATGAGCTGGGTCC	70	61.4
	RNA	GGGGGGUCCCUGCGACUCUCCUGUGCAGCCUCUGGAUUCACUUUUAGCCGCUAUGCCAUGAGCUGGGUCC	70	61.4
VDJ2	DNA	GGAGTGGGTCTCATATATCAGTAGTAATAGTCTTGCTATATATTACAAAGACTCCATGAAGGGCCGCTTT	70	40

	RNA	GGAGUGGGUCUCAUAUAUCAGUAGUAAUAGUCUUGCUAUA UAUUACAAAGACUCCAUGAAGGGCCGCUUU	70	40
VDJ3	DNA	GAGACTGACTGGCTATAAGTTATCGTACTTTGACCTCTGGG GCCAGGGAACCCTGGTCACCGTCTCCTCA	70	54.3
	RNA	GAGACUGACUGGCUAUAAGUUAUCGUACUUUGACCUCUGG GGCCAGGGAACCCUGGUCACCGUCUCCUCA	70	54.3
BCL2	DNA	TGCATCTCAGCGTTTTTTTGTTTTTAATTGTATTTAGTTATGG CCTATACACTATTTGTGAGCAAAGGTG	70	32.9
	RNA	UGCAUCUCAGCGUUUUUUUGUUUUUAAUUGUAUUUAGUUA UGGCCUAUACACUAUUUGUGAGCAAAGGUG	70	32.9
BCL6	DNA	GGAACCTCCAAATCCGAGACGCTCTGCTTATGAGGACCTCG AAATATGCCGGCCAGTGAAAAAATCTTGT	70	48.6
	RNA	GGAACCUCCAAAUCCGAGACGCUCUGCUUAUGAGGACCUC GAAAUAUGCCGGCCAGUGAAAAAAUCUUGU	70	48.6
c-Myc	DNA	CCGCCACCGCCGGGCCCCGGCCGTCCCTGGCTCCCCTCCT GCCTCGAGAAGGGCAGGGCTTCTCAGAGGC	70	77.1
	RNA	CCGCCACCGCCGGGCCCCGGCCGUCCCUGGCUCCCCUCC UGCCUCGAGAAGGGCAGGGCUUCUCAGAGGC	70	77.1

Table 3. DNA primers used for RNA synthesis

Gene	Forward primer	Reverse primer
Sγ1	CTTTAATACGACTCACTATAGGGGTGAC CCAGCAGAGCAGCTCCAG	ACACTTCCACCTGTCCTGGCGCCCCTGGA GCTGCTCTGCTG
Sα	CTTTAATACGACTCACTATAGGATGAGCT GGGATGRRCTGAGCTAGGCTGGAATAG GCTGGG	TCCAGCTCAGCTCAGCCTAACCCAGCTCAC ACCAGCCCAGCCCAGCCTATTCCAGCCTA GC
Sμ	CTTTAATACGACTCACTATAGGGAGACT GAGCT	AGCTACCCCAGCTCAGTC
VDJ1	CTTTAATACGACTCACTATAGGGGGGGG TCCCTGCGACTCTCCTGTGCAGCCTCTG G	GGACCCAGCTCATGGCATAGCGGCTAAAA GTGAATCCAGAGGCTGCACAGGAGAG
VDJ2	CTTTAATACGACTCACTATAGGGGAGTG GGTCTCATATATCAGTAGTAATAGTCTTG	AAAGCGGCCCTTCATGGAGTCTTTGTAATA TATAGCAAGACTATTACTACTGATA
VDJ3	CTTTAATACGACTCACTATAGGGAGACT GACTGGCTATAAGTTATCGTACTTTGAC	TGAGGAGACGGTGACCAGGGTTCCCTGGC CCCAGAGGTCAAAGTACGATAACTTAT
BCl2	CTTTAATACGACTCACTATAGGTGCATCT CAGCGTTTTTTTGTTTTTAATTGTATTT	CACCTTTGCTCACAAATAGTGTATAGGCCA TAACTAAATACAATTAAAAACAAAA
BCl6	CTTTAATACGACTCACTATAGGGGAACC TCCAAATCCGAGACGCTCTGCTTATGA	ACAAGATTTTTTCACTGGCCGGCATATTTC GAGGTCCTCATAAGCAGAGCGTCTCG
C-Myc	CTTTAATACGACTCACTATAGGCCGCCA CCGCCGGGCCCCGGCCGTCCCTGGCTC CC	GCCTCTGAGAAGCCCTGCCCTTCTCGAGG CAGGAGGGGAGCCAGGGACGGCCGGG
RNA 4	CTTTAATACGACTCACTATAGGGGGGCT AACCTGGGCTGGGCTGAGGGAGG	AGCCCAGCCCAGCCCTGCCCCAACCCCTC CTCCCTCAGCCCAGCCC

2.7. Alkaline cleavage deamination assay

The standard alkaline cleavage assay, which was applied in previous papers, was used to measure the deamination activity. We incubated 3-5 µL of purified GST-AID (~1 µg) with radioactive-labelled substrate (mostly 50 fmol, a range of 1.5–300 fmol for enzyme kinetic measurements) in 100 mM phosphate buffer (pH 7.13, a range of 5.6–8.0 for pH sensitivity measurement) at different temperatures (4–40 °C depends of the species, 31 °C for Hs-AID) for a range of incubation times regards the enzyme activity (1-3 hours for Hs-AID), following heat deactivation of the enzyme (85 °C for 20 min). Then, 0.5 units of Uracil DNA Glycosylase (UDG) (NEB) was added to each reaction in UDG Buffer and incubated at 37 °C for 30 min. NaOH (200mM) and loading dye (95% Formamide, 0.25% Bromophenol blue, 0.25% Xylene cyanol) were added (10 µl total) and heated at 96 °C for 10 min to induce alkali-labile abasic site cleavage and complete denaturation of the DNA.

In this assay, AID deaminates the target cytidine (bub7 bubble substrate with TGC sequence) into uridine. Then UDG enzyme removes uridine and generates an abasic site which cleaves by adding NaOH and heat. This cleavage and denaturing results in a smaller labeled DNA and an un-deaminated substrate which will separate in an acrylamide gel. The ratio of the product volume (smaller band) to the total volume of product plus substrate was calculated as percentage values indicating cytidine deaminase activity (Figure 8). The negative control was treated in the same manner, except that dialysis buffer was added instead of a purified AID sample. The 14% denaturing acrylamide gel consisting of 14% acrylamide:bisacrylamide, 1x Tris-Borate-EDTA (TBE), 25% formamide and 45% urea, was used for electrophoresis in 1x TBE running buffer. After a pre-run at 50 V for 30 min, the wells were cleaned with running buffer, and samples were loaded. A ladder containing labeled oligos with different sizes (56, 35, 29, 24, 22, 18, 15, 12, 8 nts) was treated with loading dye and loaded alongside samples. The gel was run at 250-300 V for ~1 hour. Then exposed to a Phosphor Screen GP (Bio-Rad) for 3-16 hours and imaged using Amersham Typhoon PhosphorImager.

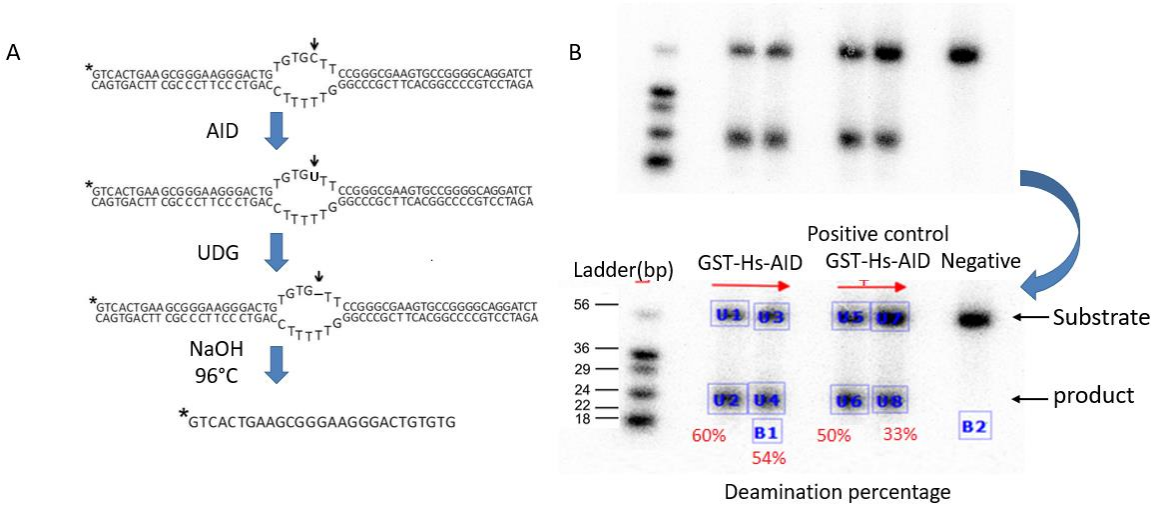


Figure 8. Alkaline cleavage assay of labelled TGC bubble substrate.
A. Representative of alkaline cleavage assay using a DNA/DNA TGC bub7 substrate. AID deaminates the target cytidine (shown by the black arrow) into uridine. UDG removes the uridine, and high temperature incubation with NaOH cleaves the abasic site and denatures the dsDNA. The labeled product is smaller than the uncleaved single stranded substrate. B. Separation of substrate and product on acrylamide gel. Comparing the volumes of product to the substrate by densitometry analysis provides percentage deamination activity which represents AID catalytic activity.

2.8. Determination of optimal temperature and pH

For the preliminary activity test, the standard alkaline cleavage assay for cytidine deamination, which was explained earlier, was used. 3 µL(~1 µg) of each purified GST-AID ortholog were incubated with TGC bub7 DNA substrate for various intervals (1 hour, 3 hours, and overnight) at three different temperatures (18, 25, 31°C) to get an estimation of their activity level. The reactions were treated with UDG and NaOH and loaded on the acrylamide gel as described for alkaline cleavage assay. According to the measured deamination activity for each species, an approximate incubation time was determined for the optimal deamination activity of each AID ortholog.

In order to assess the optimal temperature for each GST-AID ortholog, the same alkaline cleavage assay was used while incubating the reactions in a range of temperatures (5, 10, 15, 20, 25, 30 and 35°C) in proper incubation time based on the previous estimation for each AID ortholog (Table 6. Chapter 3). All the reactions were treated the same as described before, and the same phosphate buffer (pH 7.13) was used for all of them. Two reactions for each of the two individual, purified AID preparation were performed. The results were plotted by the percentage deamination

against temperature which the peak shows the optimum temperature of each AID ortholog.

To determine the optimal pH for each AID ortholog, the radioactively labelled TGC bub7 substrate was incubated with the enzyme in phosphate buffers with a range of pH 6.33 to 8.2 measured as effective pH. Then the deamination activity percentage was plotted against pH for each AID ortholog. The final activity reaction is 10 µL with a 1:3:6 ratio of TE buffer (in which the DNA substrates are diluted), dialysis buffer (in which the GST-AID are suspended) and pH buffer (phosphate buffer saline). Therefore, the phosphate buffer was scaled up with the same ratio to measure the effective pH.

2.9. Michaelis-Menten enzyme kinetics experiment

To compare the initial velocity of different AID orthologs due to Michaelis-Menten kinetics, about 0.3 µg of each AID was incubated with a range of 3 to 300 fmol of the radiolabeled substrate at its optimal temperature for the incubation time that does not pass the initial linear phase of deamination to avoid saturated activity (different incubation times were used for each species based on the previous results). Then the reactions were incubated at 85 °C for 20 min for heat deactivation and subjected to incubation with UDG, following NaOH and heat denaturing as described previously. At least 2-3 individual reactions were tested of 2 independent purified enzymes, and the average deamination percentage was used to calculate velocity. Velocity (fmol of deaminated product/min incubation/µg AID) was plotted against substrate concentration (nM).

2.10. Deamination activity assay in the presence of RNA or DNA competitor

The same standard alkaline cleavage assay was used for these experiments in various conditions as follows:

- Preincubation of cold RNA with RNA-free enzyme: The AID enzyme was pretreated by 1 µL of serially diluted 100 µM to 6.25 µM RNA oligonucleotides. Three µL (~500 ng of prokaryotic expression system or ~125 ng of eukaryotic expression system) of purified AID (treated with RNase

A during the purification) was preincubated with RNA for 30 min on ice, and then 1 µL of radioactively labeled TGC-bub7 DNA substrate (50 fmol) and phosphate buffer (pH 7.13) was added to the reaction.

- Preincubation of cold RNA with purified AID enzyme which was not treated with RNase A during purification: 3 µL (~500ng of prokaryotic expression system) of purified AID (purification without RNase A or any other nucleases) was preincubated with 1 µL of serially diluted RNA oligonucleotides for 30 min on ice, and then the radiolabeled DNA substrate and phosphate buffer was added to the reaction.

- Incubation with cold RNA and radiolabeled DNA substrate at the same time: 1 µL of serially diluted RNA was added to 1 µL of radioactive-labeled TGC-bub7 DNA substrate (50 fmol) and 5 µL phosphate buffer (pH 7.13) for each reaction. Then 3 µL (~500 ng of prokaryotic expression system) of purified AID was added to each reaction.

- Comparison of the effect of cold RNA and DNA of the same sequence: The AID enzyme was pretreated with 1 µL of serially diluted 100 µM to 6.25 µM RNA or DNA oligonucleotides. Three µL (~125 ng of eukaryotic expression system) of purified AID (treated with RNase A during the purification) was preincubated with the same amount of RNA or DNA for 30 min on ice and then radioactively labeled TGC-bub7 substrate (50 fmol) and phosphate buffer (pH 7.13) was added to the reaction.

The total volume of each reaction was 10 µL containing a range of 100 pmol to 6.25 pmol cold RNA or DNA, 50 fmol labelled DNA substrate, about 2.5 pmol GST-AID, 100 mM KCl, 6 mM Tris-HCl, 30 mM NaCl, 0.3 mM DTT, 1mM EDTA, 100 mM phosphate buffer plus 3 µL of GST-AID (~500 ng of prokaryotic expression system or ~125 ng of eukaryotic expression system). The reactions containing human GST-AID purified from *E.coli* were incubated at 31°C for 2 hours; however, the human GST-AID samples purified from eukaryotic cells were subjected to longer incubation time (5 hours) due to a lower concentration of the enzyme in reactions. All the reactions were done in duplicate, and a reaction with 1 µL water instead of RNA was used as negative RNA control.

2.11. Statistical significance and data analysis

The quantification of alkaline cleavage gels was performed as reported in previous publications (71, 74). Image Lab 6.1 was used to measure the intensity of bands on imaged gels. The average value of three quantifications per lane in alkaline cleavage gels was represented as a single data point. Typically, the experiments were repeated 2-3 times, and all reactions were performed in duplicates. Duplicate reactions showed less than 5% variation. Two independently purified samples of each AID ortholog were tested for the pH and temperature experiments in chapter 3, and the graphed data points are the means of 8-12 values (duplicate reactions of 2-3 times each experiment using 2 AID preparations). For kinetic measurement using various DNA substrates, one purified AID sample of each species were tested in duplicates, resulting in 2-3 values per data point. The graphs presented in chapter 4, section 4.2.7, were generated from 2-4 values per data point using one purified sample of human GST-AID at the time of writing this thesis. The aim is to repeat these experiments and also test another purified sample to validate the data. All the graphs were created using GraphPad Prism software (GraphPad, San Diego, CA, USA), and error bars represent the standard error (SEM).

Chapter 3. Evolutionary analysis of AID's structure and function

3.1. Introduction

AID from divergent species has a conserved core structure and catalytic site residues (two cysteines, a histidine, and a glutamic acid) as well as conserved biochemical properties such as slow catalytic rate compared to strong binding affinity (70). Nevertheless, distinctive enzymatic properties have been discovered in purified AIDs from divergent species. Dancynger et al. reported that a single amino acid change of D to G in the c-terminus of bony fish AID modulates its activity. Other studies demonstrated divergent and conserved properties in early evolved AID orthologs (71). The fish AID activity is cold-adapted, possibly due to higher ssDNA binding affinity in low temperatures or other structural features (70). Zebrafish AID has a higher catalytic rate than human AID. It is also unique among all AID orthologs in that it can deaminate methylated deoxycytidine (74). Moreover, an unexpected expansion of AID/APOBEC-like enzymes was discovered in lamprey. Remarkably, different lamprey individual represents unique AID/APOBEC-like genotype (97). Emma et al. reported a unique substrate sequence specification and temperature preferences for a few earliest diverged AID orthologs; however, they have structural and functional conserved properties (70). Previous studies have characterized a substrate sequence specificity loop for AID in different species, yet binding site interaction with different DNA substrate structures is not determined.

We hypothesize that despite AID's conserved architecture among vertebrates, it has a plastic structure and adopts different functionality on the DNA of divergent species. AID surface topology varies across species and indicates a range of positive surface charges at pH 7.1 (+5.49 in Ss-AID to +18.16 in Ol-AID). It could be due to the need to target different types of DNA topologies. Although AID has a high binding affinity to ssDNA, more investigation is needed to determine its preferences for specific shapes of ssDNA (e.g. bubble, linear, stem-loop, etc...) in divergent species. To do so, we sought to study 12 selected AID orthologs of divergent vertebrates according to their surface topologies and surface charges. We tested their biochemical features and initial velocity and aimed to continue biochemical tests using different structures of DNA.

Moreover, we evaluated AID interaction with DNA/RNA hybrids as well as RNA itself in different species.

3.2. Results

3.2.1. Selection of AID orthologs

First, a list of identified AID sequences (about 100) of different vertebrate species was aligned and 20 of the most divergent species with the lowest identity percentage were selected from different lineages. In order to get structural insights, the I-TASSER server was used to model AID's 3D structure using crystalized MBP fused human AID (PDB: 5W0Z), MBP fused AID in complex with cytidine (PDB: 5W0C), MBP fused AID in complex with dCMP (PDB: 5W0U), and MBP fused AID in complex with cacodylic acid (PDB: 5W0R) as a template (10). The best predicted structure of each AID ortholog was selected based on the lowest energy and conserved core structure of the AID/APOBEC family. Selected models were submitted to propka3.0 to calculate the surface charge and Isoelectric point (PI) of folded proteins (Table 4) (96). Remarkably, the predicted structures represent various distribution patterns of positively charged residues proximal to the binding grooves and over their entire surface (Figure 9). We selected 12 AID orthologs representative of different lineages considering their surface topology and covering a range of surface charges for the following assessments. Table 4 lists the selected species and their abbreviations.

Figure 10 shows the phylogenetic tree of selected species. Zebrafish, salmon, medaka fish, and channel catfish are from various clads of ray-finned fish species which among those, salmon and medaka have exceptionally lower (~+5.5) and higher (~+18) surface positive charges, respectively, compared to other AID orthologs. Zebrafish AID demonstrated extraordinary biochemical properties, such as high activity rate and deamination of methylated cytidine in previous studies (74). Nurse shark was selected from Chondrichthyes (cartilaginous fish), which is considered the earliest diverged jaw vertebrates' lineage. Additionally, they are among the earliest species that evolve adaptive immune systems (98). We selected Xenopus laevis from amphibians and Gallus gallus from bird species. Divergent mammalian species are included in this study: wallaby of non-placental mammals (Metatheria), megabat of bat species, rabbit of supraprimates, and two close primates, human and chimp. Bats possess a unique

immune system and suppress inflammatory responses while their APOBEC genes have expanded (99). The chimp was selected as the closest AID ortholog to human.

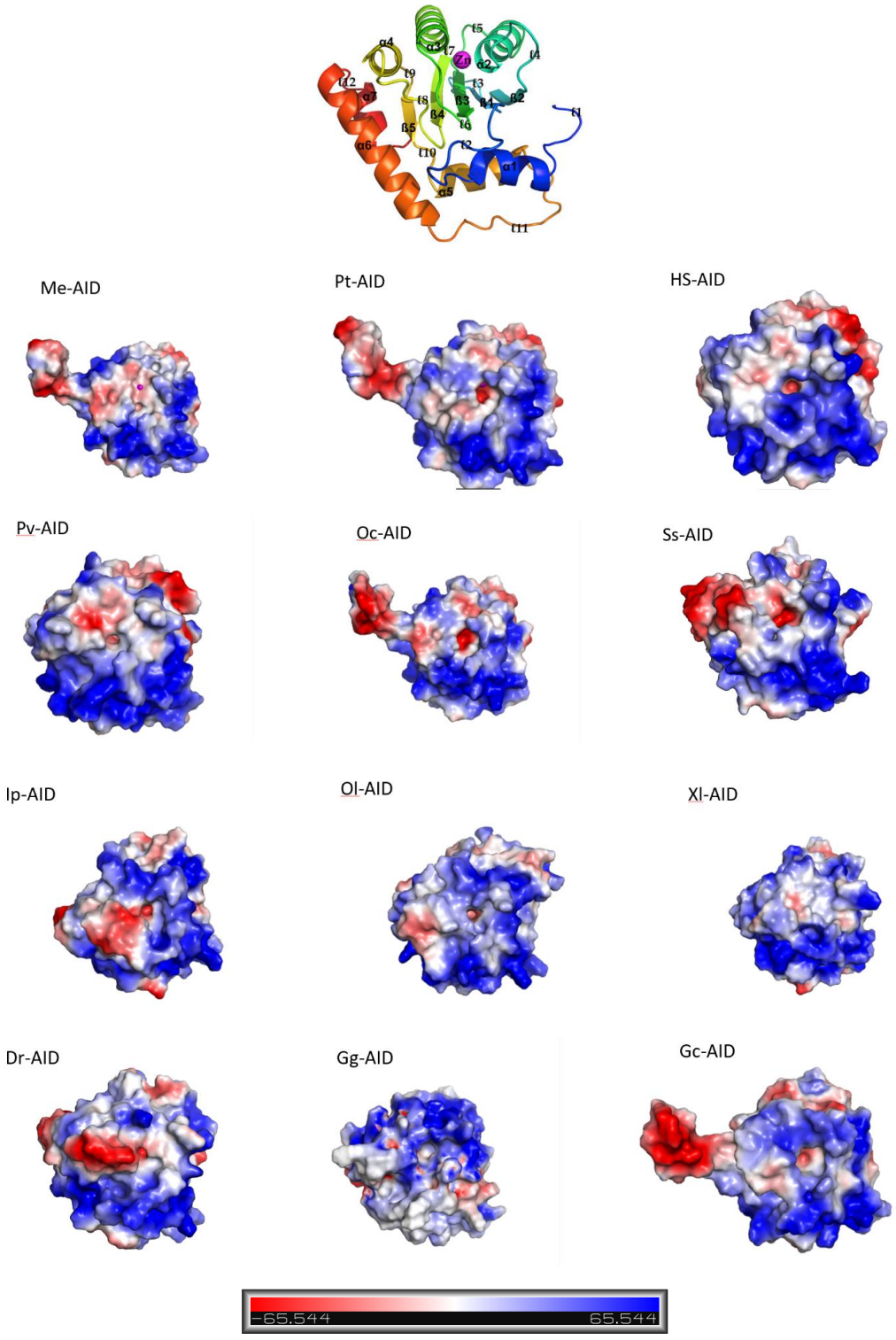


Figure 9. Comparison of the predicted surface topology of AID orthologs.
The AID structure of Hs-AID in cartoon is shown (top). Surface electrostatics of catalytic site of all predicted AID orthologs are shown with the same orientation as cartoon structure. Positive, negative, and neutral residues are colored blue, red, and white, respectively. The electrostatic bar is shown at the bottom.

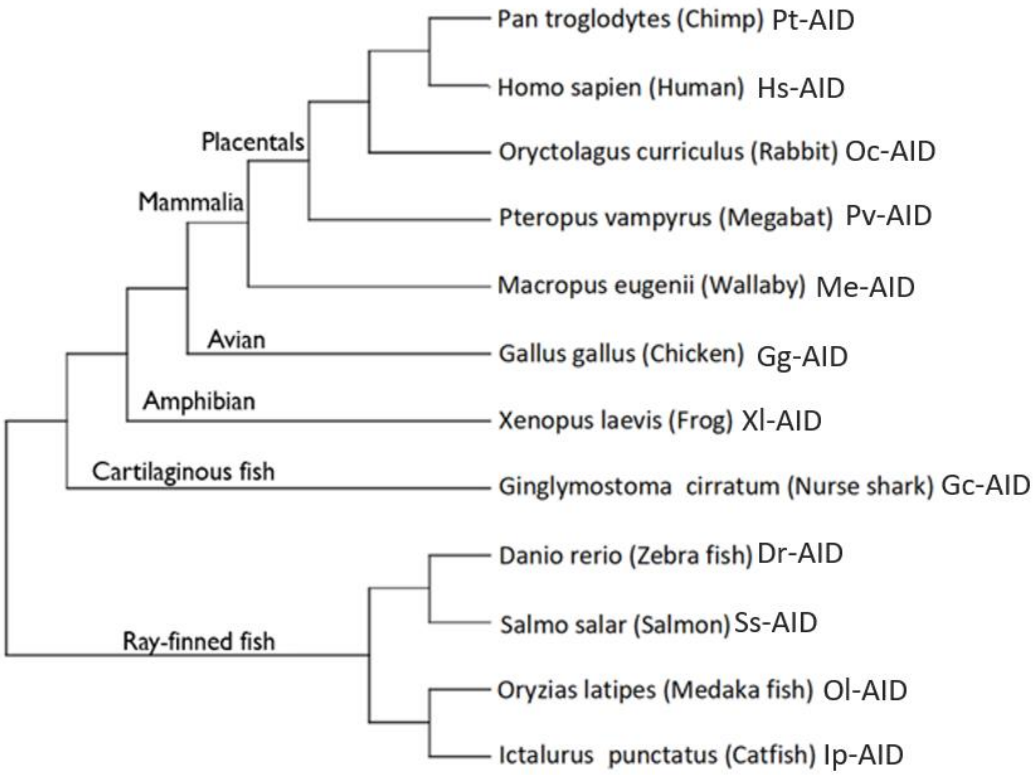


Figure 10. Phylogenetic tree of selected species.
This phylogenetic tree indicates the evolutionary relationship of diverged species selected in this study.

Models generated in this study includes AID orthologs of a range of various homology percentages to Hs-AID, which can shed light on the development of biochemical properties in AID through evolution (Table 5). The Ss-AID has the lowest identity (55%) and similarity (63%) to Hs-AID, and Pt-AID is the closest to Hs-AID with 99% identity and 92% similarity. Figure 11 shows the alignment of selected species. The AID signature, SWSPC, and zinc coordinating residues, including H56, E58, C87 and C90 in human AID, are conserved among AID orthologs. However, there are multiple non-conserved sites such as bony fish insert and N terminal extension in fish AIDs denoted in Figure 11. Also, the residues containing substrate specificity loop (the loop between α4 and β4) are varied among AID orthologs.

As structures of AID for non-human species obtained through empirical methods were not available to use as templates for homology modeling, the predicted structures

using I-TASSER require further refinements. To generate better structural models and facilitate further analysis, we are using Alpha Fold, an artificial intelligence software that predicts proteins' 3D structure using a deep learning system, to regenerate these AID orthologs for future studies on these AID orthologs (100).

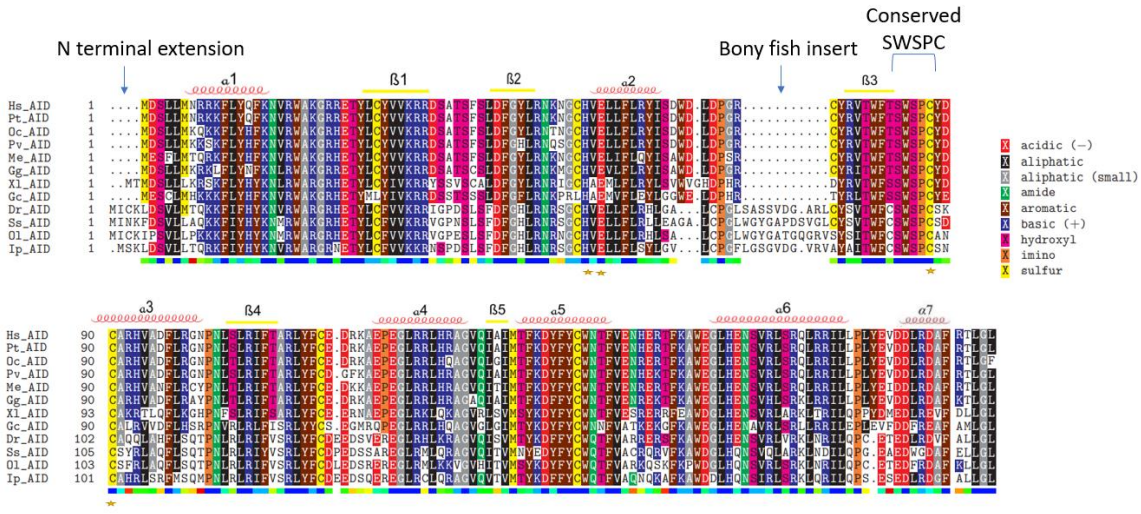


Figure 11. Alignment of amino acid sequences of the AID orthologs selected in this study.

The conserved catalytic residues, Histidine, Glutamate, and two Cytidines, were shown with stars. The highly conserved signature of SWSPC among AID orthologs was denoted, as well as non-conserved N terminal extension and bony fish insert.

Table 4. The PI and protein charge of folded state at pH 7

Selected AID orthologs	Abbreviation	Surface protein charge (+) at pH7	PI
Ginglymostoma_cirratum (Nurse shark)	Gc-AID	10.66	10.23
Danio rerio (Zebra fish)	Dr-AID	10.61	9.94
Oryzias latipes (Medaka fish)	Ol-AID	18.16	10.60
Ictalurus_punctatus (Catfish)	Ip-AID	9.00	9.77
Xenopus laevis (Frog)	Xl-AID	11.97	10.13
Salmo salar (Salmon)	Ss-AID	5.49	9.05
Gallus gallus (Chicken)	Gg-AID	12.91	10.31
Macropus eugenii (Wallaby)	Me-AID	11.74	9.88
Pteropus vampyrus (Megabat)	Pv-AID	9.63	9.84
Oryctolagus curriculus (Rabbit)	Oc-AID	8.80	9.77
Pan troglodytes (Chimp)	Pt-AID	9.96	10.13
Homo sapiens (Human)	Hs-AID	10.55	10.38

Table 5. Comparison of AID amino acid identity and similarity in selected species

Similarity

	Gc-AID	Dr-AID	Ol-AID	Ip-AID	Xl-AID	Ss-AID	Gg-AID	Me-AID	Pv-AID	Oc-AID	Pt-AID	Hs-AID
Gc-AID	100%	63.42%	61.57%	61.11%	68.05%	58.79%	70.83%	70.83%	70.37%	71.75%	71.29%	71.29%
Dr-AID	52.31%	100%	81.48%	81.48%	65.27%	78.24%	69.44%	70.37%	69.44%	68.98%	69.44%	69.44%
Ol-AID	51.38%	78.7%	100%	77.31%	63.88%	81.94%	66.2%	66.66%	65.74%	65.27%	66.66%	66.66%
Ip-AID	51.38%	78.24%	74.07%	100%	64.35%	75%	65.74%	67.12%	66.2%	65.74%	66.2%	66.2%
Xl-AID	63.42%	58.79%	56.94%	56.48%	100%	62.5%	74.53%	74.07%	74.07%	73.61%	74.53%	74.53%
Ss-AID	48.14%	76.38%	79.16%	70.83%	56.01%	100%	63.88%	64.81%	63.88%	63.42%	63.88%	63.88%
Gg-AID	67.12%	58.79%	55.09%	55.55%	68.51%	54.62%	100%	85.64%	85.64%	85.64%	87.96%	87.96%
Me-AID	69.44%	60.64%	55.55%	56.94%	67.59%	56.48%	87.5%	100%	84.72%	86.57%	87.03%	87.03%
Pv-AID	69.9%	61.57%	57.4%	58.33%	69.44%	57.4%	88.42%	89.35%	100%	87.5%	88.42%	88.42%
Oc-AID	70.37%	60.64%	56.01%	56.01%	70.83%	56.48%	89.35%	91.66%	93.05%	100%	88.88%	88.88%
Pt-AID	68.98%	60.18%	56.01%	56.01%	69.9%	56.01%	91.2%	91.2%	94.44%	94.9%	100%	92.12%
Hs-AID	68.51%	59.72%	55.55%	56.48%	69.9%	55.55%	90.74%	91.66%	94.44%	94.44%	99.53%	100%

Identity

3.2.2. Divergent AID orthologs have different rates of cytidine deamination activity

We generated expression constructs expressed in prokaryotic systems and purified two separate preparations of GST-tagged AID of all the aforementioned species using the previously described method in Chapter 2. The concentration of purified samples was evaluated on denaturing SDS gels, and the enzyme activity on the ssDNA substrate was checked using the standard alkaline cleavage assay, which was previously established (32, 70, 81). The cytidine deamination activity of each purified AID on TGCbub7 substrate was tested at three different intervals, including long overnight incubation in three temperatures (18, 25, and 31°C). TGC-bub7 has a WRC motif (TGC) located in a 7-nucleotide bubble, recognized as the most favored substrate for human AID. Long overnight incubation, as well as testing different temperatures, would allow even the lowest activity levels to reveal themselves. On the other hand, shorter incubations (1 hour) provide an estimation of saturated activity. Those AIDs with saturated activity (Ol-AID and Dr-AID) were incubated for shorter times. Ol-AID and Dr-AID indicated measurable enzyme activity in as short as 10 minutes of incubation. Table 6 presents the incubation time used for each AID ortholog in the following experiments.

Table 6. The incubation time of AID orthologs

AID orthologs	Incubation time (minutes)
Gc-AID	120
Dr-AID	10
Ol-AID	10
Ip-AID	600
Xl-AID	180
Ss-AID	180
Gg-AID	60
Me-AID	60
Pv-AID	120
Oc-AID	120
Pt-AID	60
Hs-AID	120

3.2.3. AID orthologs display a wide range of temperature preferences

Previous outcomes of in vitro enzyme assays demonstrated unique thermosensitivity among AID orthologs. Many fish AID orthologs indicated cold adopted AID enzymes showing as low as 8°C optimal temperature for cod AID (69, 70). To have a complete assessment of the temperature preferences of each GST-AID ortholog, we measured their activity on TGC-bub7 at temperatures ranging from 5°C to 35°C in various incubation times corresponding to their activity level (Table 6).

As illustrated in Figure 12, the optimum temperature for AID deamination activity varies among the studied species. Collectively, the higher optimal temperature for deamination activity was observed for mammalian species (Hs-AID and Pt-AID) at 31°C and lower optimal temperature for fish species at 12°C (Ss-AID) and 15°C (Ip-AID). However, there are some exceptions, such as Ol-AID, which has the highest activity at 32°C.

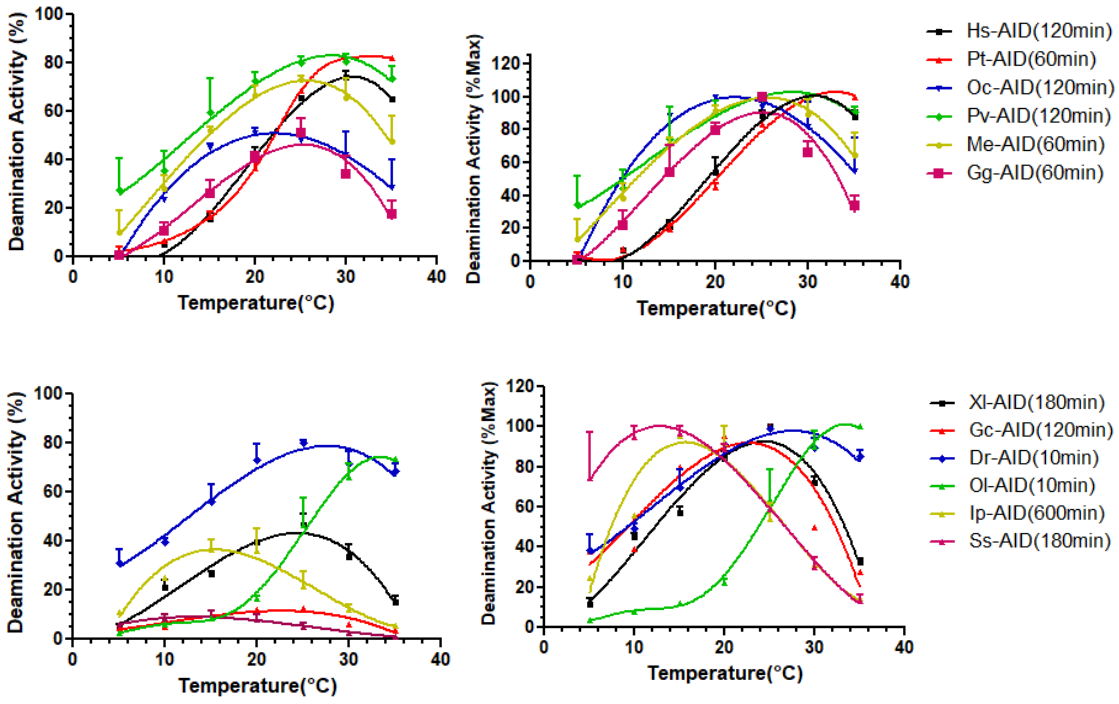


Figure 12. The optimal temperature of AID orthologs.
The optimal temperature of AID orthologs was compared at a 5°C temperature increment (5-35°C). Two independent protein preparations of each AID ortholog were examined. Results are plotted as deamination activity percentage (left panel) and percentage of maximum deamination activity (right panel). (n=4-6)

3.2.4. Divergent AID orthologs demonstrate efficient enzyme activity near neutral pH

According to previous studies, in optimal in vitro conditions most APOBECs function in acidic environments, while AID prefers neutral pH (~7.3) and could maintain its activity at more basic pH (~8) (101). To evaluate the pH sensitivity of AID orthologs, the deamination activity of each AID ortholog was measured using the same alkaline cleavage method explained before in phosphate buffer with an effective pH range of 6.33 to 8.2. The percentage of AID deamination activity against pH is plotted in Figure 13. Although AID orthologs exhibit different percentages of their maximum activity by changing pH, they were all efficient in pH 7 to 8.

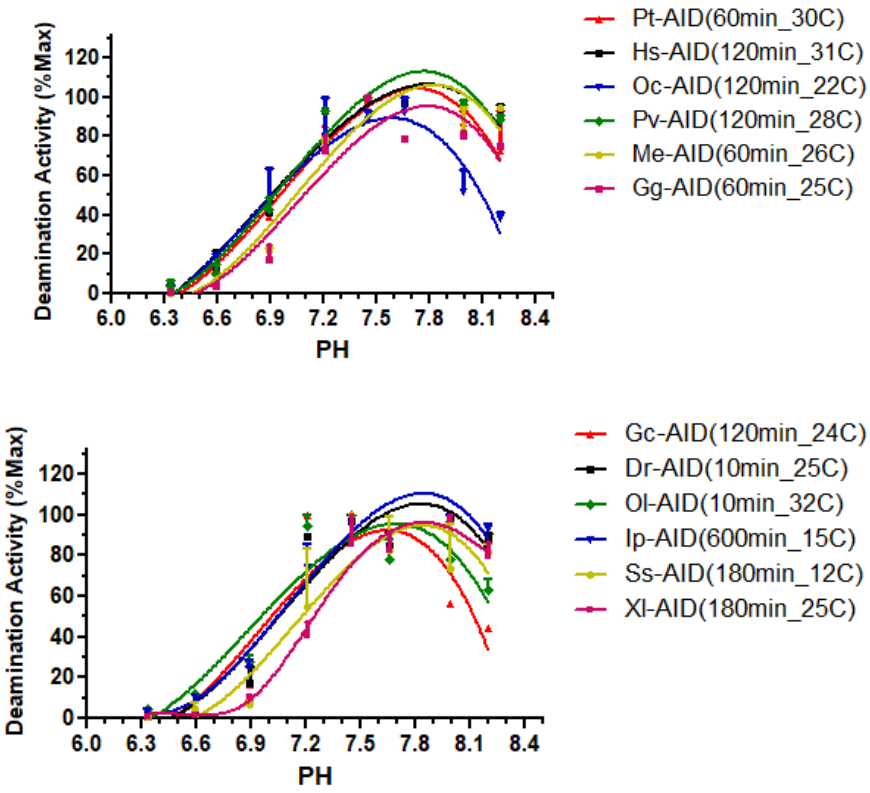


Figure 13. Optimal pH of AID orthologs.
Two independent protein preparation of each AID ortholog were tested in phosphate buffer with an effective pH range of 6.33 to 8.2. Results were plotted as the percentage of maximum deamination activity. (n=4-6)

3.2.5. AID of diverged species favors different DNA substrate structures

AID orthologs of 12 diverged species were incubated with a range of substrate concentrations using four different substrate structures, including TGC bub7, 2bubble, stem-bub, and stem loop, with the same TGC target motif for AID. As shown in Figure 14, AID orthologs of human, chimp, wallaby, frog, zebrafish and medaka favor small bubble substrate as TGC-bub7. There is not any specific substrate among our substrates that chicken-AID favors the most, as the velocity graph is almost similar for all the tested substrates. Human and chimp, which are closely related evolutionary, demonstrate comparable trends for the substrates tested in this study, except that chimp-AID has a higher activity rate. Rabbit-AID showed over 8-fold higher velocity on stem bubble mix structure (stem-bub). It is worth mentioning that stem loop structure was identified as the less favored substrate for most species tested here. A different pattern of substrate preference was shown among fish species, as nurse shark and salmon favor stem bubble mix substrate (about 3-fold higher than bubble structure), while zebrafish and medaka have the highest velocity rate for bubble structures (TGC-bub7 and 2bubble). Catfish shows the lowest velocity rate among the species included in this study, mentioning that its activity was evaluated after as long as 10 hours of incubation. However, it indicates enhanced activity on the 2bubbles substrate structure. Frog-AID demonstrates approximately 5-fold higher velocity for TGC-bub7 structure, and it has the lowest activity on stem loop similar to other AID orthologs.

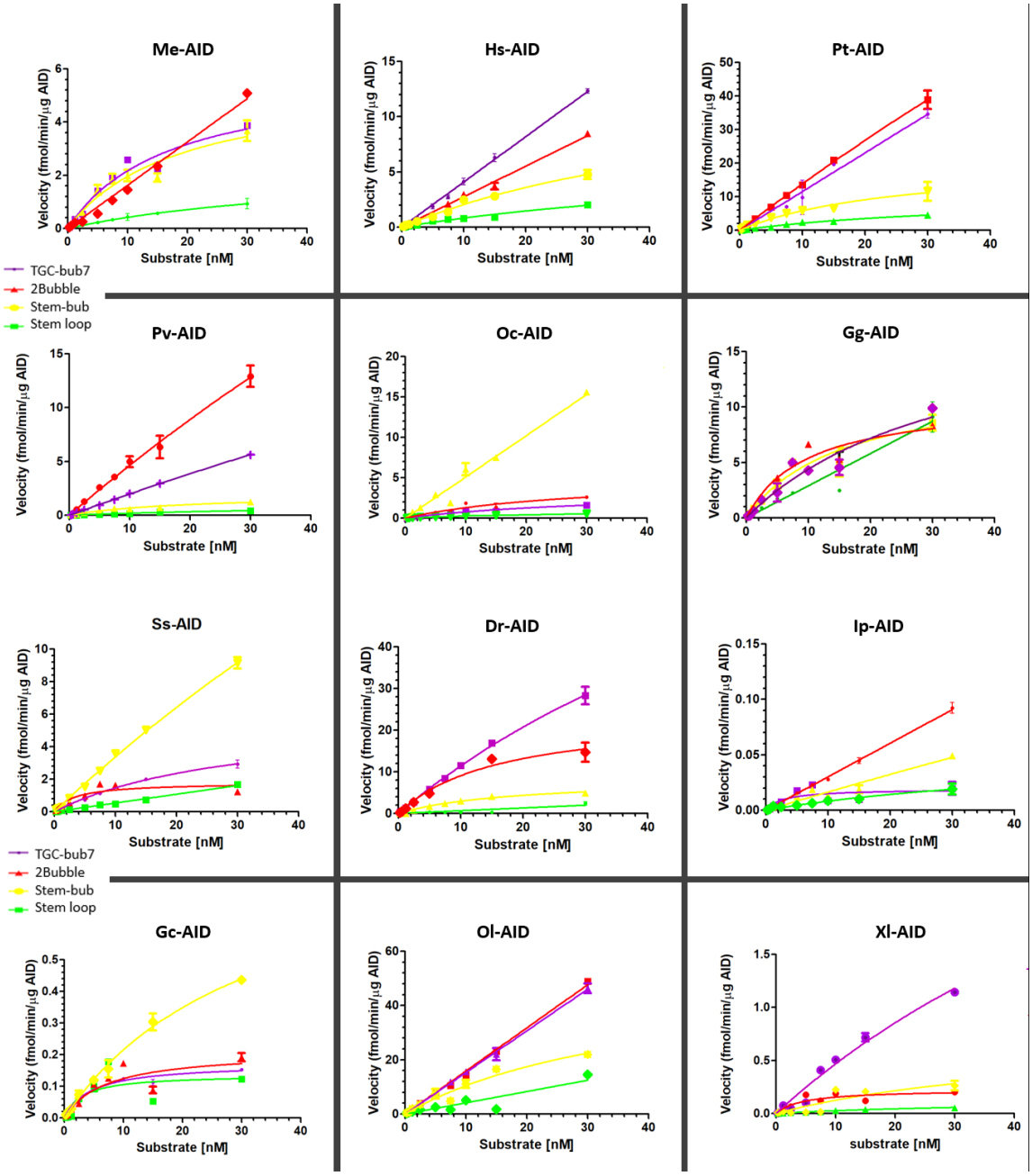


Figure 14. The initial velocity of AID orthologs on bubble type DNA/DNA and DNA/RNA substrate.

Initial velocity for deamination activity of GST-AID of different species on four different structures of TGC-bub7 (purple), 2Bubble (red), stem-bub (yellow), and stem loop(green) are shown. The optimal temperature and incubation time of each AID ortholog were used. All reactions were done in duplicates. (n=2-3) Results are plotted as velocity (fmol product/ incubation time/ enzyme) to the substrate concentration.

3.2.6. AID orthologs exhibit different initial velocities for DNA/RNA substrates compared to DNA/DNA substrates

Human AID is capable of binding and acting efficiently on DNA/RNA hybrid. Although AID has reduced deamination activity on the DNA/RNA hybrid of random sequences (similar sequence to TGCbub7), DNA/RNA hybrid enhances AID activity on sequences of Ig switch regions (32). We evaluated the initial velocity of AID orthologs on DNA/DNA compared with DNA/RNA of the same sequence using the alkaline cleavage assay. Purified GST-AID orthologs of 12 species were incubated with a range of concentrations of TGC-bub7 substrate, including 0.31, 0.62, 1.25, 2.5, 5, 7.5, 10, 15, 30 nM in its optimal temperature and incubation time as mentioned in table 7. Since the purified AID samples were not treated with RNase A during purification, we expect that a small amount of cellular RNA will remain bound to the positively charged surface of AID even after purification. Therefore, to test if these RNAs interfere with AID activity, RNase A was added to the samples. Overall, three conditions were used in this section: the DNA/DNA TGC-bub7 substrate, the DNA/RNA TGC-bub7 substrate, and the DNA/DNA TGC-bub7 substrate plus 2ng/µl RNase A. Buffer and salt concentration were equal among all reactions.

As shown in Figure 15, the majority of AID orthologs in this study exhibit a decreased deamination activity on DNA/RNA hybrid substrates compared to DNA/DNA substrates. Gg-AID, Ss-AID, and Gc-AID displayed similar initial velocities on both types of substrates. Additionally, the addition of RNase A to the reactions resulted in varying increases in the deamination activity of AID orthologs and did not affect the initial velocity of Pt-AID, Pv-AID, Dr-AID, and Ol-AID. This suggests that RNA bound to the surface can interfere with AID activity in some species but not in others. We conclude that AID orthologs have different interactions with RNA or DNA/RNA hybrids, likely due to differences in surface topology.

3.2.7. Treatment with RNase A increases the deamination activity of some AID orthologs, and this effect is not dependent on their surface charge

As depicted in Figure 15, treating AID from different species with RNase A had varying effects on their activity. While the enzyme activity rate increased significantly for Hs-AID, Me-AID, Gg-AID, Ss-AID, Gc-AID, and Xl-AID, it did not affect the activity of Ol-

AID, Dr-AID, Pt-AID, and Pv-AID. Given that all AID orthologs studied in this thesis had an overall positive surface charge, as reported previously (Table 7), it was expected that random RNA bindings would affect all AID orthologs equally. However, this was not the case, as Ol-AID, which had the highest predicted positive charge (around +18), was not affected by RNA bindings, while the activity of Ss-AID, which had a lower positive charge (around +5.5), was increased by the removal of bound RNAs. Additionally, Hs-AID and Dr-AID had similar predicted surface charges (approximately +10.5) but displayed different effects in response to RNase A treatment. These observations suggest that AID orthologs interact with RNA independent of their overall electrostatic surface charge and that these interactions may be unique for each AID ortholog.

Table 7. Optimal temperature and incubation time of AID orthologs for maximum deamination activity.

Selected AID orthologs	Abbreviations	Optimal temperature (°C)	Incubation time (minutes)	Surface charge
Ginglymostoma_cirratum (Nurse shark)	Gc-AID	24	120	10.66
Danio rerio (Zebra fish)	Dr-AID	25	10	10.61
Oryzias latipes (Medaka fish)	Ol-AID	32	10	18.16
Ictalurus_punctatus (Catfish)	Ip-AID	15	600	9.00
Xenopus laevis (Frog)	Xl-AID	25	180	11.97
Salmo salar (Salmon)	Ss-AID	12	180	5.49
Gallus gallus (Chicken)	Gg-AID	25	60	12.91
Macropus eugenii (Wallaby)	Me-AID	26	60	11.74
Pteropus vampyrus (Megabat)	Pv-AID	28	120	9.63
Oryctolagus curriculus (Rabbit)	Oc-AID	22	120	8.80
Pan troglodytes (Chimp)	Pt-AID	30	60	9.96
Homo sapiens (Human)	Hs-AID	31	120	10.55

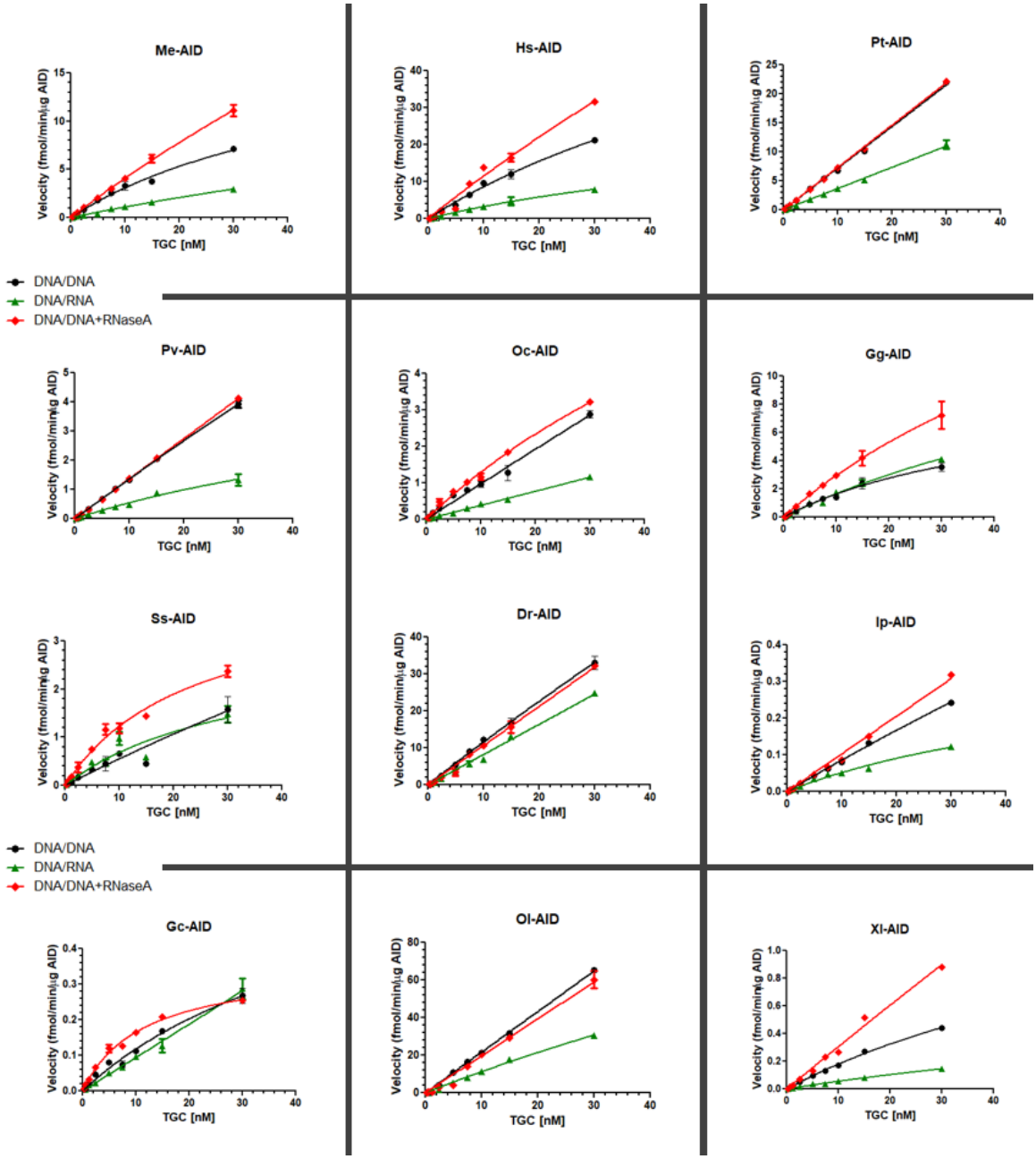


Figure 15. The initial velocity of AID orthologs on bubble type DNA/DNA and DNA/RNA substrates.

Initial velocity for deamination activity of GST-AID of different species on DNA/DNA substrate (black), DNA/RNA substrate (green) and DNA/DNA substrate with 2 ng/μl RNase A (red) are shown. The optimal temperature and incubation time of each AID ortholog were used. All reactions were done in duplicates. (n=4-6) Results are plotted as velocity (fmol product/ incubation time/ enzyme) to the substrate concentration.

3.3. Discussion

In this chapter, we aimed to investigate the enzymatic properties of AID across a broad range of species in order to determine its target specificity. To achieve this, we expressed and purified AID orthologs from 12 different species, including divergent lineages. Based on the predicted 3D structures and surface electrostatic charge, these AID orthologs possess a variety of surface topologies, which may result in distinct biochemical properties and target preferences. Our findings reveal that the optimal temperature and incubation time required for the deamination activity of each AID ortholog differ among evolutionarily divergent species, as well as the rate of catalytic activity. In general, AID displayed increased activity among the mammal species that we studied, with an optimal temperature above 22°C. Human and chimp AID demonstrated optimal activity at 31°C, while AID orthologs from all mammals (human, chimp, rabbit, wallaby, megabat) and chicken AID were able to maintain activity at 37°C. Conversely, the deamination activity of most fish AIDs (salmon, catfish, shark) and one amphibian that we studied (frog) decreased significantly at higher temperatures. Previous studies have also exhibited a broad range of activity and preferred temperatures among AID from divergent species (70, 81).

We demonstrated that certain AID orthologs exhibit different preferences for DNA substrate structure, which may be due to co-evolution with their immunoglobulin repertoire genes or other biological functions. Earlier studies that compared the predicted structure and catalytic activity of Dr-AID and Ip-AID with Hs-AID found various biochemical characteristics among AID orthologs. Interestingly, the catalytic rate and optimal temperature of these AIDs are primarily maintained by a single amino acid in their C-terminal region (71). Substitution of positively charged R25 and H29 in Hs-AID and Dr-AID respectively with neutral N28 in Ip-AID decreased its activity. For all tested substrate structures, Ip-AID showed a significantly lower catalytic rate compared to Hs-AID. This finding is consistent with previous studies that have demonstrated Ip-AID's reduced ability to facilitate CSR when it is transfected into AID-deficient B cells (71). In contrast, Dr-AID showed several fold higher catalytic rates for bubble substrates compared to Hs-AID, which is in line with previous studies; however, it showed significantly lower activity on stem loop substrates. In addition, zebrafish AID (Dr-AID) is notably more efficient at deaminating 5m-C compared to other similar enzymes, such as

human, catfish, medaka, and tetraodon AID. This unique ability sheds light on its distinct role in embryonic development and explains why its suppression leads to hypermethylated CpG motifs (74). From structural aspects, both Hs-AID and Dr-AID form a catalytic pocket that can accommodate dC residues and contains a triad of Zn-coordinating residues and a catalytic glutamic acid. However, the catalytic pocket of Dr-AID has a more extended and flexible loop than that of Hs-AID, allowing it to provide more space for the larger 5mC substrate (69, 102). Thus, studying the enzymatic characteristics of diverse AID orthologs could lead to the discovery of novel AID features and their structural determinants.

During immune responses in bony fish, affinity maturation occurs, but it is not as efficient as in mammals, possibly due to the lack of GCs or the lower catalytic rate of AID, as shown in this study and previous publications (81, 103). Although teleost fish express AID, they cannot undergo class switch recombination (CSR) due to the structure of their IgH gene. However, AID from teleost fish, such as zebrafish, Japanese puffer, and catfish, can catalyze CSR in vitro in mammalian AID-deficient lymphocytes, suggesting that teleost AID has the full catalytic functions required for CSR (104, 105). Xenopus (frog) antibodies also undergo SHM and CSR; however, the switch regions in Xenopus are AT-rich instead of G-rich, which may impact switching efficiency (69, 106). Xenopus AID has shown CSR activity and is expressed in hematopoietic tissues, suggesting a role in ontogeny (106, 107). However, neither Xenopus AID nor other amphibian AID has been extensively characterized previously. Our results indicate that frog AID (Xl-AID) exhibits significantly lower activity and adapts to colder temperatures than Hs-AID, likely due to its colder habitat temperature, while it has a preference for small ssDNA bubble structures similar to Hs-AID.

Cartilaginous fish (Chondrichthyes) are the earliest jawed vertebrates (gnathostomes), and their descendants, including nurse sharks, were among the first animals to develop an adaptive immune system based on V(D)J and Ig (103). It has been reported that affinity maturation occurs in the nurse shark, as evidenced by a correlation between somatic mutations and increased binding affinity in IgNAR clones from the immune tissues of a hyperimmunized nurse shark (83). Although it was once believed that cartilaginous fish were also incapable of CSR due to the organization of their genes, it is now known that they can undergo an "unconventional" type of CSR among different IgM clusters and between IgW and IgM clusters (98, 108). Interestingly,

shark AID(Gc-AID) demonstrated several-fold higher activity on the stem-loop bubble mix DNA structure in this study. The same trend was observed in rabbit AID(Oc-AID). In rabbits, a limited diversity is generated by VDJ rearrangement through a process known as somatic gene conversion. During somatic gene conversion, sequences from upstream pseudogenes replace sequences within the functional V region gene, leading to a greater range of diversity (109). Compared to other mammals studied here, rabbit AID displayed an exceptional ability to act on stem-loop bubble mix substrate (about 8-fold higher than TGC-bub7). Further investigations are needed to identify whether these findings are linked to the rudimentary type of CSR, or gene conversion mechanism in sharks and rabbits, respectively. From a structural point of view, these AID orthologs have a different surface topology and slightly lower surface charge; nevertheless, a detailed structural comparison is required to find the responsible residues for their odd behaviour.

Furthermore, we discovered that while all AID orthologs act on DNA/RNA hybrids, they exhibit a wide range of preferences for DNA/DNA or DNA/RNA. Additionally, some species displayed significantly elevated activity when RNase A was added to the reaction. These findings prompted us to investigate AID-RNA binding specificity and evaluate this hypothesis using human AID in the following chapter. The increased activity in the presence of RNase A for some species also suggests that RNA may play a role in regulating AID activity within cells.

Given the various surface topologies and differences in activity on DNA/RNA hybrids among AID orthologs, further docking studies on nucleic acid- AID interactions are necessary to identify the structural determinants of these results. In addition to the structural determinants, it is also important to consider the functional implications of these results. The diversity in activity and preference among AID orthologs suggests that they may have evolved to perform specific functions within their respective organisms. It is possible that certain AID orthologs have evolved to exhibit a preference for acting on specific DNA/DNA or DNA/RNA structures.

Furthermore, these findings raise the question of how these variations in activity and preference among AID orthologs may affect their role in the immune system. AID is known to be involved in the generation of SHM and CSR, which are essential for the

adaptive immune response. Therefore, the variations in activity and preference among AID orthologs may have implications for the immune response of different species.

Finally, this study has demonstrated that subtle structural differences lead to various functions of these enzymes, and further studies are needed to determine the activity and preference among AID orthologs. Additionally, future studies should also assess the activity of AID orthologs on various DNA structures, which would provide deeper insight into AID's mechanism of action and the structural basis of its enzymatic activity. This chapter was left to be continued in future research, as the AID-RNA interactions proved to be an intriguing tangent of this study, which will be discussed in the subsequent chapter.

Chapter 4. Determination of RNA binding specificity and its potential regulatory role on Activation-induced cytidine deaminase (AID)

4.1. Introduction

Activation-induced cytidine deaminase (AID) is a member of the AID/APOBEC family of DNA/RNA editing enzymes, which contributes to antibody diversification by causing mutations at the immunoglobulin (Ig) loci, thereby mediating SHM and CSR of antibodies (110). AID is needed for good antibody response; however, it has substantial off-target mutational activity, which can cause and exacerbate cancers by mediating drug-resistance and aggressiveness-promoting mutations in tumors (111, 112). AID, like other APOBECs, targets single-stranded DNA (ssDNA) for mutation. Although AID is somewhat targeted to Ig loci, it can bind to any accessible and unpaired ssDNA generated as a result of transcription. The transient opening of DNA duplex upon transcription can lead to higher AID activity on the accessible non-transcribed DNA strand. This enzyme can deaminate deoxycytidine (dC) to deoxyuridine (dU) within DNA, but it does not have detectable activity on RNA (113, 114). Branton et al. proposed that features of DNA inherent to a gene, including its structure and topology, could affect AID recruitment. AID acts on breathing dsDNA transcription independently, as well as several topologies of ssDNA in the context of transcription, including transcription bubbles, R-loops formed between long-lived DNA/RNA hybrid, stem-loops, and DNA/RNA hybrids found at the Ig loci (56). Structural-based studies of human AID elucidated that AID possesses five binding motifs, three ssDNA binding channels, and two RNA binding sites. These five binding channels can make different combinations of substrate binding modes (33).

Our previous results (Chapter 3.2.7) on structural plasticity and different functionality of AID orthologs of divergent species indicated an elevated deamination activity of most AID orthologs while adding RNase A to the purified enzyme (Figure 15). Since AID has a positive surface charge, it can potentially bind RNA during purification, but whether these RNA bindings are specific or randomly covering the surface needs to be clarified. We aim to study whether and how RNA may modulate AID activity. To do so, two different approaches were used in this study. First, RNase A was added to the

purified AID to degrade any bound RNA and evaluate whether AID activity changes (this section is in common with the previous chapter). Second, we evaluate the activity of Human GST-AID on a preferred DNA bubble substrate in the presence of free RNA oligonucleotides of different sequences. Particularly, RNA sequences of three switch regions, variable regions, and oncogenes were included in the study. We hypothesize that RNA binding may contribute to the regulation of AID deamination activity in the S and V regions during CSR and SHM.

4.2. Results

4.2.1. RNase A increases deamination activity of purified AID.

As was explained in the previous chapter, the enzyme activity of purified AID from 12 species was evaluated while treated with RNase A. RNase A degraded any RNA bound to the enzyme surface and provided naked AID for activity evaluation. The initial velocity of 12 AID orthologs incubated with TGCbub7 substrate (0.3 to 30 nM) in optimal temperature and incubation time of each AID, plotted in Figure 15. The red line shows the effect of RNA removal, which elevates enzyme activity in Hs-AID, Me-AID, Gg-AID, Ss-AID, Gc-AID, Xl-AID and to a lower extent in Oc-AID and Ip-AID comparing to the black line. Consequently, RNA binding decreases the initial velocity of some AID orthologs, including human AID. Therefore, we decided to treat human GST-AID with RNase A during the purification (described in material and methods) and then add RNA extracted from the cells or different RNA oligonucleotides to identify whether random RNA binding decrease the activity or there are specific RNAs which only form in biologic context.

4.2.2. Pure RNA oligos interfere with AID activity more than whole cell RNA extract

We examined the effect of RNA extraction from two different cell lines on AID activity in the context of alkaline cleavage assay in order to determine whether RNA extraction from cells expressing AID has a distinct impact on AID activity compared to RNA from control cells.

1- The RNA was extracted from HEK 293 T cells transfected with pcDNA plasmid containing human GST-AID to express the enzyme after two days of transfection. A control plate without transfection was also extracted using the same method for comparison

2- The RNA was extracted from MCF7 breast cancer cells that stably express human AID through the addition of doxycycline (Dox). Extractions were performed two days after AID expression was induced (+Dox) and from a control plate where AID expression was not induced (-Dox).

The concentration of RNA extracted from the cells was measured using NanoDrop. One µL of RNA at three different concentrations (1, 0.5, and 0.2 µg/µL) was diluted with RNA/DNA-free water and incubated with human GST-AID purified from *E. coli*. The enzyme was treated with RNase A and rigorously washed on the purification column to remove RNase A. The deamination activity of AID pre-incubated with these RNA extractions was assessed using the alkaline cleavage assay, as described in the methods section. Additionally, two RNA oligos were pre-incubated with AID and treated in the same manner to evaluate the randomness of RNA binding to the positive surface of AID. As shown in Figure 15, there were no significant differences between the RNA extraction from cells expressing AID and those that did not. Notably, the RNA extraction did not affect AID activity in the context of this assay. However, oligonucleotide RNA 1 reduced AID activity significantly compared to the other RNA oligo (RNA 2) and the negative RNA control.

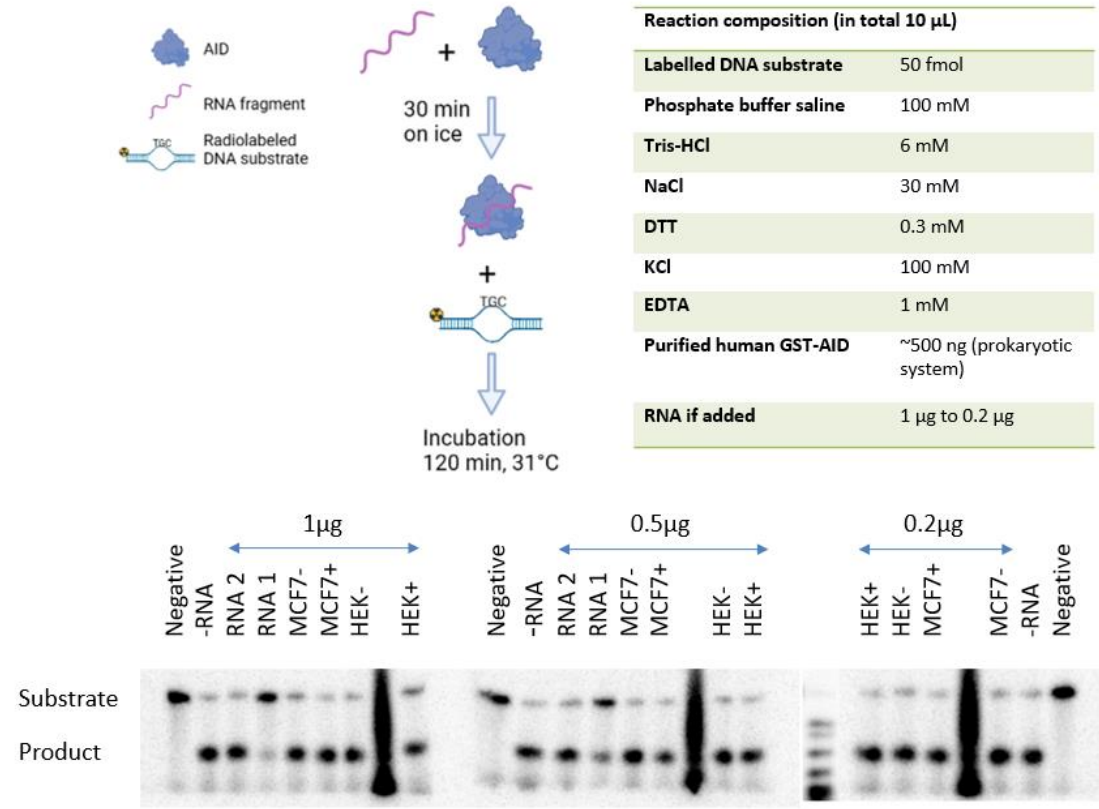


Figure 16. The effect of extra RNA on deamination activity.
GST-AID expressed in *E.coli* was treated with RNase A during the purification and was preincubated with RNA extraction from cells or RNA oligos for 30 min on ice, then subjected to activity assay, 120 min 31 °C incubation with TGC bub7 substrate and amount of RNA added to each reaction is labeled, +: RNA extracted from HEK293 or MCF7 expressing AID after 2 days, - : RNA extracted from control HEK293 cells or MCF7, the negative reaction does not contain AID or RNA, -RNA is positive control of maximum AID activity in this experiment. All reactions have the same salt compositions mentioned in the table (upright).

4.2.3. Selective RNA oligos diminished AID activity to various extents

Nine different RNA oligos (Table 1) were incubated with human GST-AID, and the percentage of deamination activity was measured using the standard alkaline cleavage assay mentioned before. The RNA1 (75 bp), RNA4 (56 bp), RNA8 (45 bp) and RNA7 (45 bp), to a lower extent, reduced AID activity, while the same concentrations of other RNA oligos does not affect the deamination compared to the - RNA (without any RNA oligo) reactions (Figure 17). All the sequences which could decrease activity have a higher percentage of guanidine; However, G-richness might not be the only influential factor since RNA4 and RNA8 has higher inhibitory effect compared to RNA1 and RNA7, respectively. These RNAs might form secondary structures favored for AID binding, such

as G-quartet or include a sequence pattern which is preferred by AID with higher binding affinity than TGC bub7 substrate. These free RNA structures might bind to the AID surface and repel DNA or wrap around the protein and block its catalytic site. Further structural studies on AID mutants are needed to find out the binding mechanism of decoy sequences. Table 8 includes all the compositions used in the final reaction for the alkaline cleavage assay in this chapter.

Table 8. Final reaction composition in the alkaline cleavage assay.

Reaction composition (in total 10 µL)	
Labelled DNA substrate	50 fmol
Phosphate buffer saline	100 mM
Tris-HCl	6 mM
NaCl	30 mM
DTT	0.3 mM
KCl	100 mM
EDTA	1 mM
Purified human GST-AID	~120 ng (2.5 pmol) eukaryotic / ~500ng (10 pmol) prokaryotic
RNA if added	100 µM to 6 µM (100 pmol to 6 pmol)

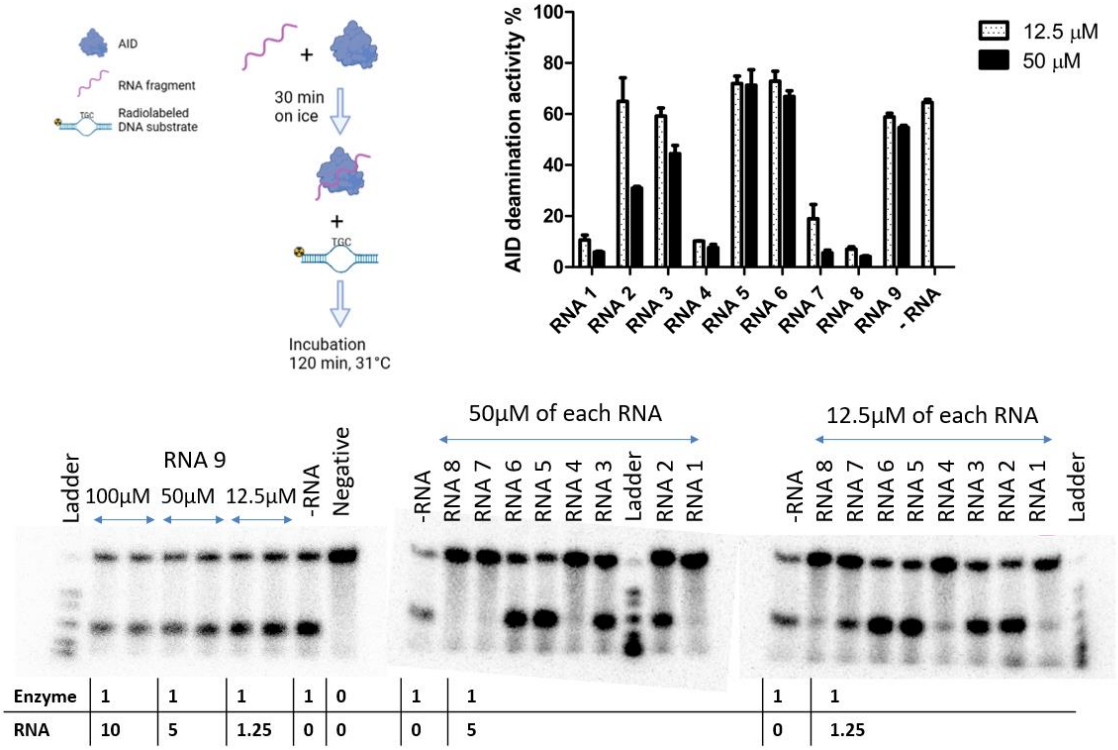


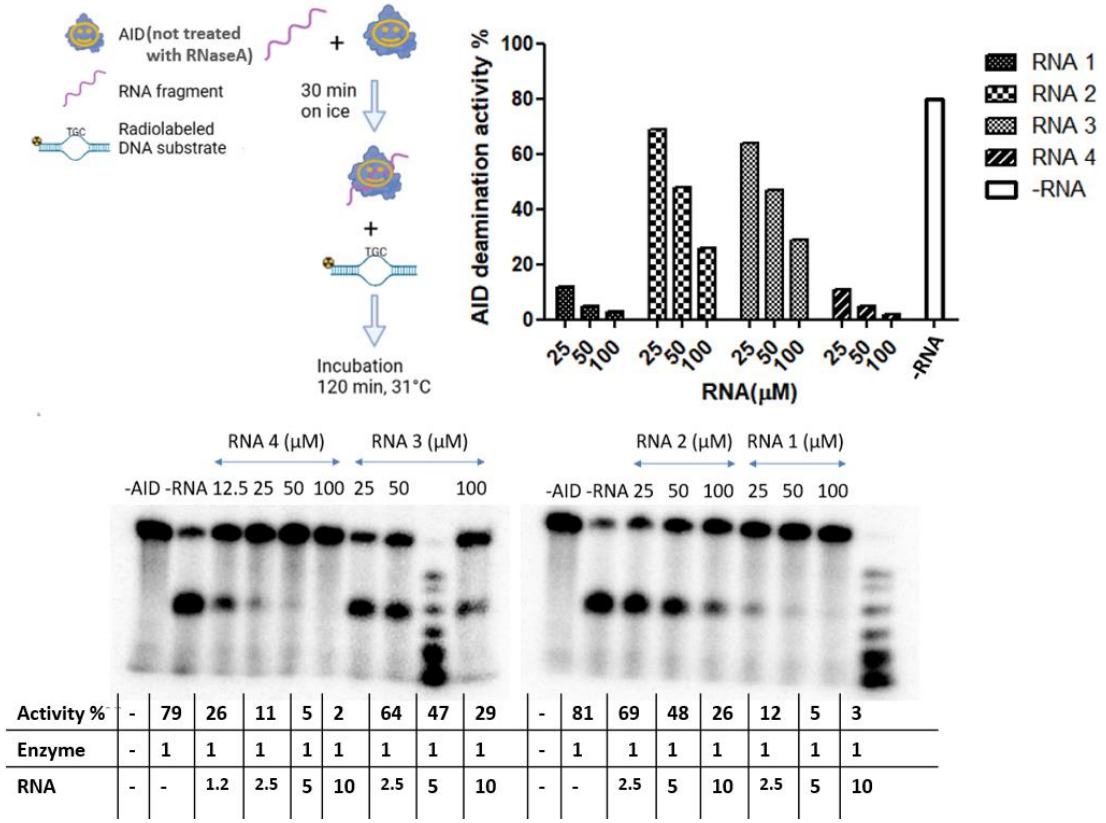
Figure 17. Deamination activity of human GST-AID incubated with RNA oligos of 9 different sequences.

Representative alkaline cleavage gels of 12.5μM cold RNA oligo (RNA 1 to 8) added to the reactions (bottom right), or 50μM cold RNA oligo added to the reactions (bottom middle), or 100, 50, 12.5 μM of RNA 9 added to the reactions (bottom left). A comparison of AID activity in the presence of 12.5 or 50 μM of 9 RNA oligos compared to the control condition without RNA graphed by an average of 3 values for each point (upright). Human GST_AID (prokaryotic) treated with 1μl of different RNA concentrations incubated with 50fmol of radioactively labeled TGC bub7 substrate in a total of 10μl reactions, 120min at 31°C. Nine different RNA oligos were used in this experiment mentioned in Table 1. The salt composition of all samples was similar, as mentioned in Table 8. The molar AID: RNA ratio for each reaction is shown at the bottom.

4.2.4. Specific RNA oligos diminish AID activity independent of RNase A treatment

GST-tagged Hs-AID was incubated with RNase A to remove any RNA bound to the AID. Despite the fact that GST-AID was washed extensively with 1xPBS on glutathione sepharose bead column for RNase A removal, any remaining RNase A could interfere with the result. Therefore, we decided to test the activity of purified AID samples without RNase A treatment in the presence of RNA oligos. High (100 μM) and low (25 μM) concentration of RNA 1, RNA 2, RNA 3, and RNA 4 was tested on human GST-AID, expressed in a prokaryotic system, purified without RNase A treatment. The results of this experiment are consistent with previous findings; RNA 1 and RNA 4 reduced AID

activity significantly, whereas RNA 2 and RNA 3 diminished the AID activity slightly and only at higher concentrations (100 μM)(Figure 18). Thus, RNA1 and RNA 4 bind to and decrease AID activity, even when other RNAs are bound to the AID surface. This inhibitory effect is consistent with our observations on AID that have been treated with RNase A (Figure 16, 17).



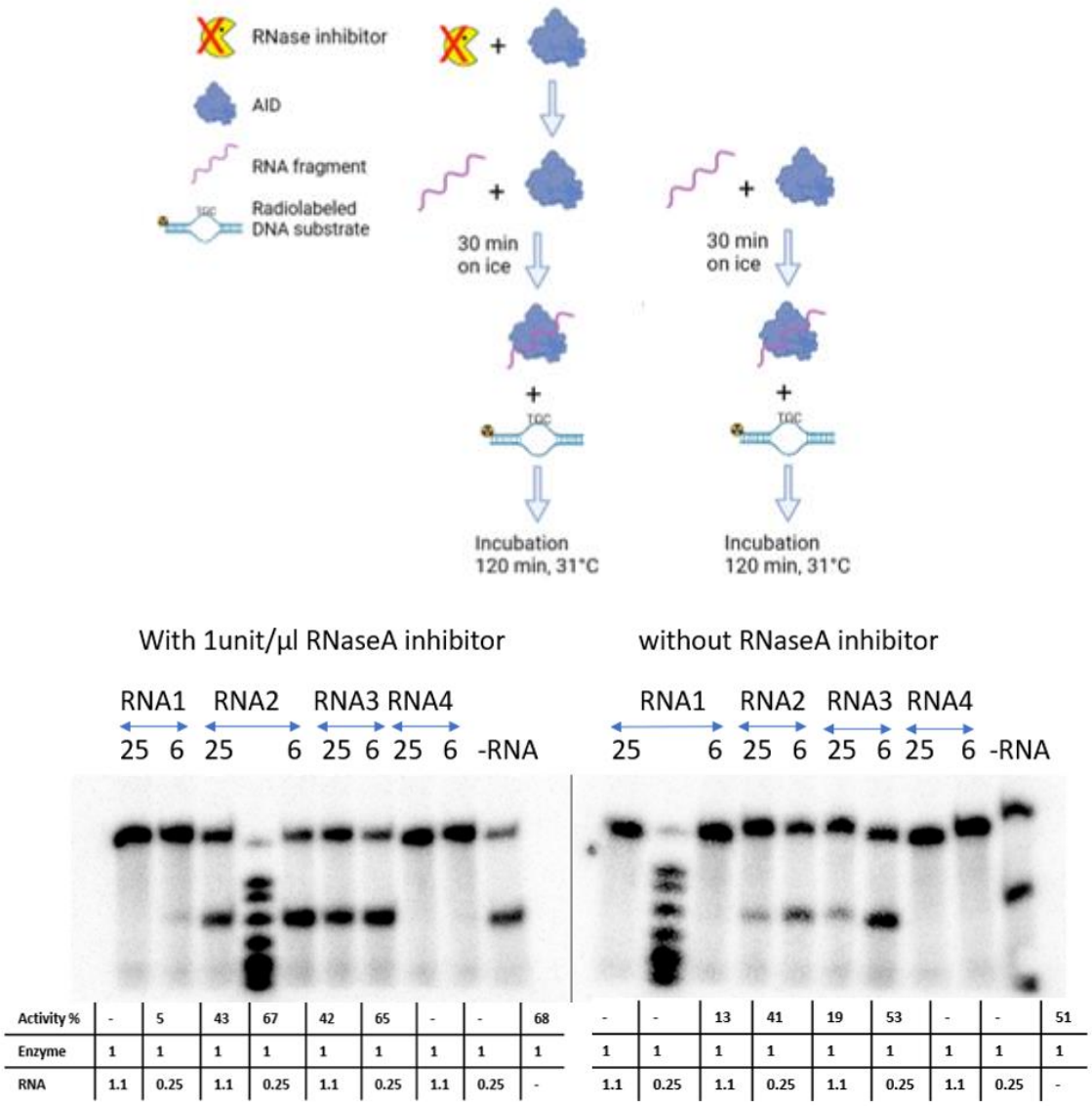
Activity %	-	79	26	11	5	2	64	47	29	-	81	69	48	26	12	5	3
Enzyme	-	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1
RNA	-	-	1.2	2.5	5	10	2.5	5	10	-	-	2.5	5	10	2.5	5	10

Figure 18. Specific RNAs diminish the deamination activity of AID not treated with RNase A.

Human GST-AID (prokaryotic) not treated with RNase A incubated with 1 μl of serially diluted RNA concentrations (100 μM, 50 μM, 25 μM and 12.5 μM for RNA 4) in 10 μl reactions. 50fmol of radioactively labeled TGC bub7 was used as substrate. Four different RNA oligos were used in this experiment, mentioned in Table 1. Reactions were incubated for 120min at 31°C for optimum Hs-AID activity. The salt composition of all samples was similar, as mentioned in Table 8. The molar AID: RNA ratio for each reaction is shown at the bottom.

To completely remove the effect of probable RNA nucleases that might exist in the reaction, we compared two sets of reactions with or without RNase inhibitor (1 unit/μl), each including two diluted concentrations (25 μM and 6.25 μM) of 4 RNAs incubated with the same amount of purified AID. Although the AID activity increased by more than 10% in the reactions containing the RNase inhibitor, there is not any

significant difference in the inhibitory effect of RNA 1 and RNA 2 in the presence of the RNase inhibitor, as shown in Figure 19.



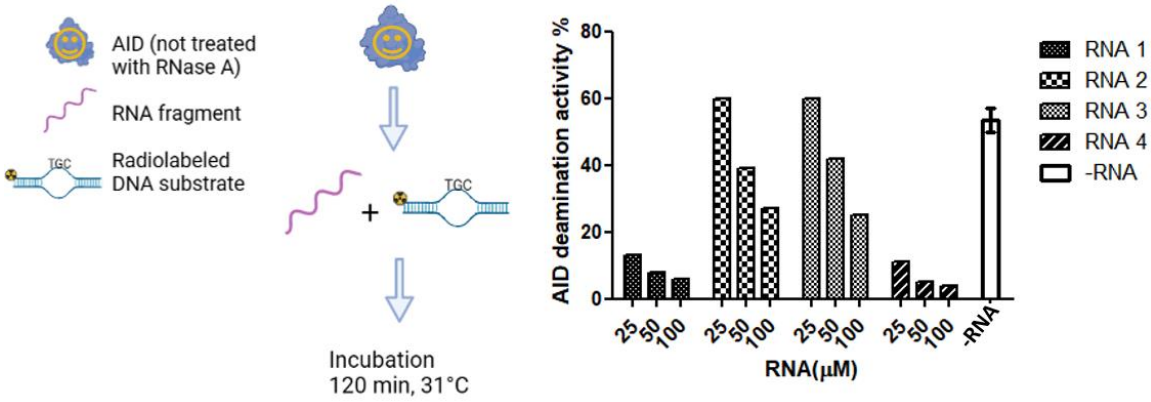
Activity %	-	5	43	67	42	65	-	-	68
Enzyme	1	1	1	1	1	1	1	1	1
RNA	1.1	0.25	1.1	0.25	1.1	0.25	1.1	0.25	-

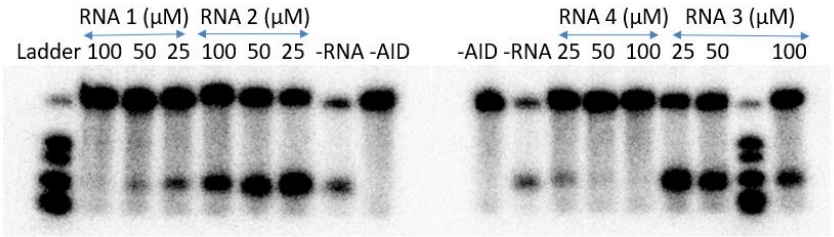
-	-	13	41	19	53	-	-	51	
1	1	1	1	1	1	1	1	1	
1.1	0.25	1.1	0.25	1.1	0.25	1.1	0.25	-	

Figure 19. The effect of RNase inhibitor on AID activity.
Two concentrations (25 and 6 µM) of RNA1, RNA2, RNA3, and RNA4 incubated with 112 ng purified GST-Hs-AID (prokaryotic) with or without RNase inhibitor (1 unit/µL) for 30 min on ice and then incubated with the radioactively labeled substrate 120 min at 31°C. All samples had similar salt compositions, as listed in Table 8. The molar AID: RNA ratio for each reaction is shown at the bottom.

4.2.5. Specific RNA oligos inhibit AID activity when competing with DNA substrates concurrently

Previously, we pre-incubated purified AID with RNA and then tested its activity on DNA substrates. This resulted in the RNA being bound to the enzyme, which either repelled the DNA substrate or prevented its binding. We later decided to add AID to the mixture of RNA and DNA substrate in a phosphate buffer to evaluate the effect of RNA on activity while concurrently competing with DNA. The findings of this experiment were consistent with previous ones, showing that RNA1 and RNA4 had a strong inhibitory effect, while RNA2 and RNA3 had little effect on AID activity at high concentrations (100 μM) (Figure 20). Therefore, the inhibitory effect of RNA1 and RNA4 is not simply due to them being accessible to AID before DNA substrate exposure.





Activity %	6	8	13	27	39	60	57	-	-	50	11	5	4	60	42	25
Enzyme	1	1	1	1	1	1	1	-	-	1	1	1	1	1	1	1
RNA	10	5	2.5	10	5	2.5	-	-	-	-	2.5	5	10	2.5	5	10

Figure 20. The effect of RNA oligos competing with DNA substrate at the same time.

The 100, 50, and 25 μM of 4 RNA oligonucleotides and 5 nM radioactive labeled TGC bub7 DNA substrates were incubated with human GST-AID (prokaryotic) not treated with RNase A for 120 min at 31°C. All samples had similar reaction compositions, as listed in Table 8. The AID: RNA ratio for each reaction is shown at the bottom.

4.2.6. The activity of human AID expressed in eukaryotic cells is hindered by specific RNA oligos

The results mentioned thus far have been obtained using human GST-AID expressed in a prokaryotic system. To demonstrate that these findings are not specific to the *E. coli* expression system, we employed two different eukaryotic systems (*Pichia pastoris* and HEK293 cells) to express and purify human AID. The effect of 25 μM RNA1, RNA2, RNA3, and RNA4 on AID preparations expressed in eukaryotic systems confirmed the previous results. The yield of protein synthesis in eukaryotic systems is much lower (<1/4 of prokaryotic system) than that in the bacterial expression system. Therefore, we increased the incubation time to 5 hours to achieve detectable enzyme activity. A comparison of the effect of added RNA to the negative RNA control revealed inhibitory effects of RNA1 and RNA4 for both AID expressed in yeast or the HEK293 system. Taken together, these observations confirm the inhibitory effect of specific RNA oligonucleotides (RNA1 and RNA4) on AID activity.

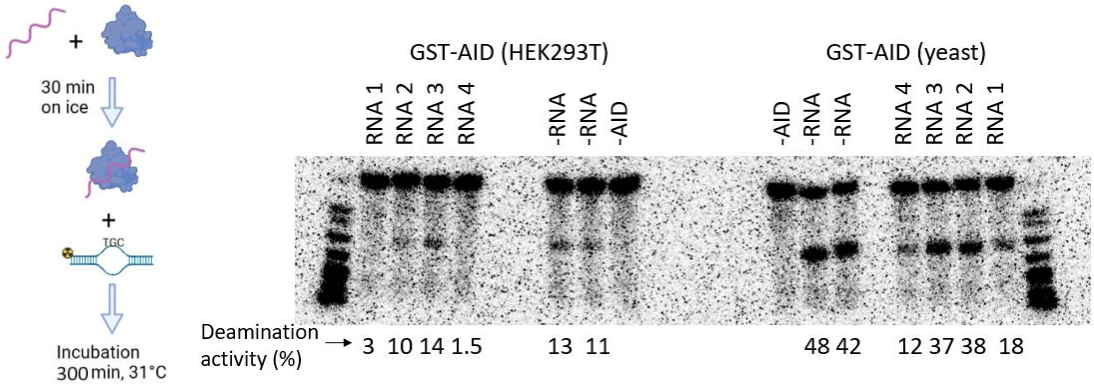


Figure 21. The effect of specific RNA oligos on AID purified from eukaryotic expression systems.

The 25 μM of RNA 1, RNA 2, RNA 3, and RNA 4 was preincubated for 30 min with AID expressed in yeast (about 125 ng protein, right) or HEK293 cells (protein amount not detectable on the gel, left) and tested on 5 nM TGC bub7 DNA substrate for 5 hours incubation at 31 °C. All samples had similar reaction compositions, as listed in Table 8.

4.2.7. RNA sequences of Ig Switch regions differentially impact AID activity compared to variable regions and oncogenes

The Ig Switch regions are GC-rich sequences and are frequently transcribed during CSR while not coding any proteins. Ig heavy and light chain variable gene undergo SHM leading to synthesis of antibody variable region with higher affinity. RNA4,

which strongly inhibits AID activity, has a similar sequence as Ig switch regions. Considering the G-richness of switch regions and the essential role of transcription in CSR and AID activity, RNA transcripts of S regions might interact with AID and involve in regulatory mechanisms. Thus, we decided to test RNA sequences of three different switch regions: Sα (80 bp), Sγ1 (48 bp), Sμ (20 bp); RNA sequences of three variable regions: VDJ1, VDJ2, VDJ3 (70 bp); and three oncogenes which contain high AID mutation rates based on previous studies: BCl2, BCl6, C-Myc (70 bp).

The RNA oligos of these different sequences were generated by T7 RNA polymerase using a DNA sequence with a T7 promotor upstream, as described in the methods section. The inhibitory effect of the synthesized RNA 4 (the same sequence) was compared with the previously ordered RNA 4 from IDT at a range of concentrations (100 μM to 6.25 μM) on GST-Hs-AID (Figure 22). The effect of the synthesized RNA fragments is comparable to that of the IDT-ordered RNA in the context of our assay.

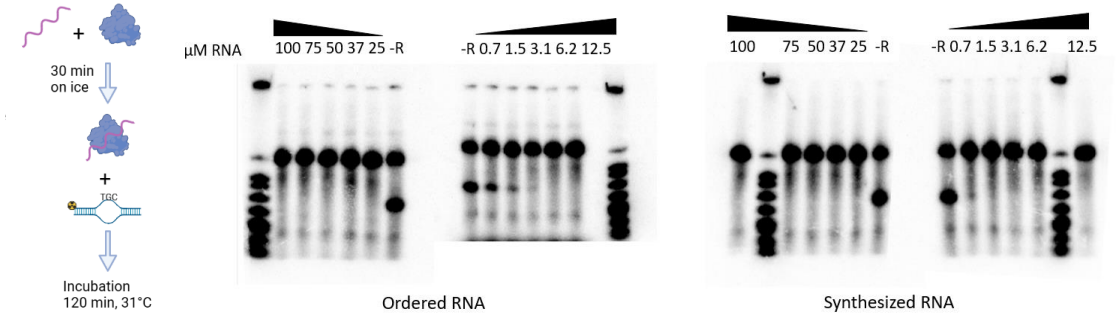


Figure 22. The effect of synthesized RNA is comparable with IDT-ordered RNA.
A range of concentrations (100 to 0.7 μM) of RNA4 (synthesized or ordered) incubated with 112 ng purified GST-Hs-AID, 30 min on ice and then 120 min at 31°C with TGC-bub7 substrate. A similar reaction composition, as listed in Table 8, was used for all samples.

The deamination activity of human GST-AID purified from yeast treated with serially diluted RNA oligos (100 μM to 6.25 μM) of the aforementioned regions or DNA oligos of the same sequence were measured. As shown in Figure 23, the transcribed RNA of S regions inhibits AID activity with lower IC50 (1.3, 1.8, and 1.6 for Sα, Sμ, and Sγ1, subsequently) compared to its equivalent DNA sequence (8.5, 55, and 8.6 for Sα, Sμ, and Sγ1, subsequently). This difference is notably higher (30-fold) for Sμ and approximately 6.5 and 5 fold for Sα and Sγ1 regions, respectively. In contrast, sequences from variable regions yielded comparable inhibitory effects between RNA and DNA sequence or even lower IC50 for the DNA sequence. From an evolutionary

perspective, this difference between RNAs of switch regions and variable regions might be due to the modulations which is needed for tighter regulation of CSR while allowing ample SHM to proceed. However, to get a more accurate IC50, we need to repeat this experiment with more concentration points and try lower concentrations (decrease to 1:1 ratio of RNA: DNA) as the amount of free RNA or DNA is more than 100-fold higher than the DNA substrate in the final reactions. However, the molar RNA: AID ratio is 2.5 for the lowest RNA concentration which is reasonable. In addition, for further assessment we need to evaluate at least two AID preparations to be more confident about the results.

As illustrated in Figure 24, there isn't a unique trend for the three oncogene sequences that we tested here. BCL2 and C-Myc showed several fold lower IC50 for RNA compared to DNA of the same sequence, but there aren't any significant differences between RNA and DNA for BCl6.

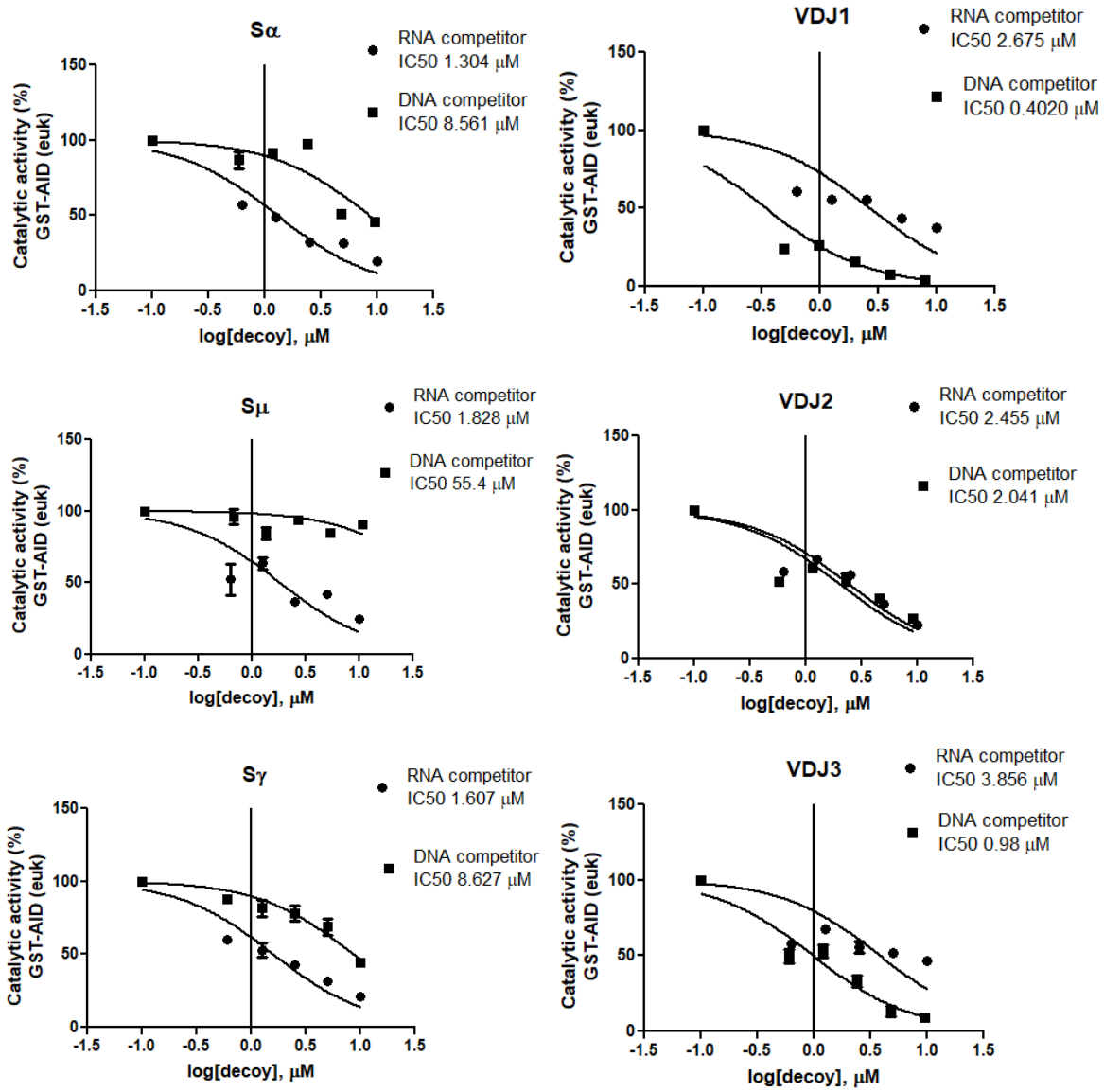


Figure 23. The inhibitory effect of RNA sequences from switch regions vs variable regions.

Results are plotted as the percentage of AID catalytic activity to the maximum using 5 nM of the standard bubble oligonucleotide substrate TGCbub7 treated with RNA from Ig switch regions or DNA of the same sequence as a function of log inhibitor concentration (left), and RNA from Ig variable regions or DNA of the same sequence (right). A negative control reaction without RNA was considered 100% AID activity. Eukaryotic-expressed and purified GST-AID was used for all reactions. All reactions were incubated at 31°C for 2hrs at pH 7.1. AID was preincubated with RNA for 30 minutes on ice. The salt composition was similar for all reactions, as mentioned in Table 8.

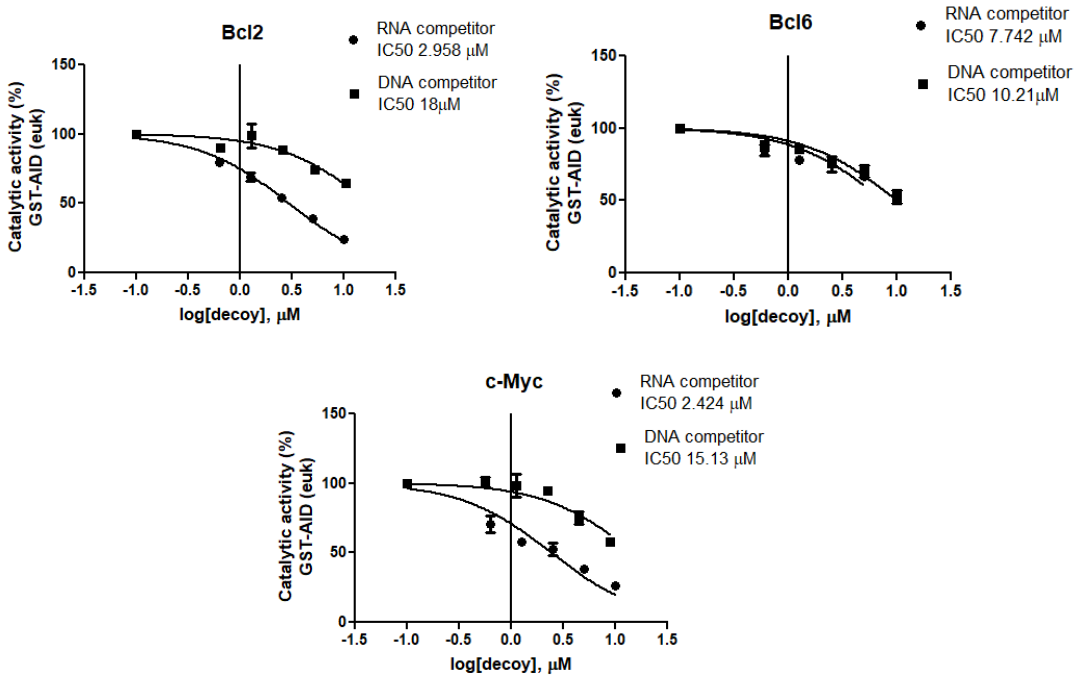


Figure 24. The inhibitory effect of RNA sequences from oncogenes.
Results are plotted as the percentage of AID catalytic activity to the maximum using 5 nM of the standard bubble oligonucleotide substrate TGCbub7 treated with RNA or DNA of the same sequence as a function of log inhibitor concentration. The salt composition was similar for all reactions, as mentioned in Table 8.

4.3. Discussion

Previous studies have identified structural and regulatory features of AID, such as catalytic pocket inaccessibility and a high overall positive surface charge, that contribute to non-productive binding to single-stranded DNA and slow catalytic activity. These characteristics are conserved among AID orthologs and play an important role in enzyme self-regulation (34, 69, 70). Moreover, in this study, we propose another possible regulatory mechanism. Here, we presented the first discovery that specific RNAs can decrease AID activity, providing new insight into the regulation of this enzyme. Two RNA binding motifs have been suggested for human AID in previous studies: first, amino acids 130–138. This RNA binding domain does not pass near the catalytic site but is located on the opposite side of the AID, similar to the assistant patch (31). It is believed that G-quadruplex RNAs of the transcribed switch regions bind to this AID-RNA binding motif to guide AID targeting the complementary DNA of the switch region and enhance CSR (31). Furthermore, AID G133V and G133P mutants, two identified mutations in the AID-RNA binding motif of hyper-IgM patients, lack the ability to associate with CSR while retaining deamination activity (115, 116). Second, an RNA binding groove forms at the dimeric interface of the AID quaternary structure and is located near the catalytic site (33). Considering the dimeric RNA binding region, AID oligomerization may contribute to RNA inhibitory effect.

It was considered that Hs-AID has rampant non-specific RNA binding because of its highly positive surface charge at pH 7.1 (about +10.5); and due to the massive non specific binding with other proteins and aggregation, it was extremely challenging to purify soluble active AID (21). However, the binding specificity of these RNAs and their inhibitory effect were not identified previously. The AID inhibition was observed in high concentrations of RNA oligonucleotides compared to the DNA substrate (100 to 2000 fold), which could resemble the microenvironment of the Ig locus during the transcription. Although the RNA amount was at the lowest 100-fold higher than the DNA substrate, the molar RNA: AID ratio was in a range of 40 to 1.25 fold which is acceptable considering non-specific binding to the enzyme surface. In addition, Previous studies have shown the importance of transcription for AID activity during CSR or SHM. Particularly, the unique features of transcription in the Ig region, such as bi-directional transcription and super-enhancer genes, which overlapped the majority of AID’s off-target activities,

decisively explained AID targeting (55, 117). Super enhancers frequently enforce gene expression; therefore, a massive amount of RNA is produced in the Ig region. As the RNA transcripts of the switch regions do not code protein, they might play a role in recruiting AID to target switch regions during CSR, as proposed by Zheng et al. and also modulating AID activity in higher concentrations by mediating a feedback mechanism (Figure 25).

It was shown that AID has a high affinity for the DNA/RNA hybrid of switch regions (32). Although the majority of the studies in the field restrict AID targeting on accessible non transcribed ssDNA produced in R-loop during the transcription, previous studies of our lab demonstrated AID activity on any single stranded naked DNA generated in transcription such as stem-loop, R-loop and the breathing dsDNA following traversing RNA polymerase as well as DNA/RNA hybrid (56). Hala et al. tested AID activity on a range of different single stranded and bubble type DNA substrates containing DNA/DNA and DNA/RNA hybrid. According to their findings, AID can mutate the DNA strand of a DNA/RNA hybrid, and it has the highest affinity measured so far for the DNA/RNA hybrid of the switch region compared to any other substrate. However, the mutation rate in those regions is lower than the TGC motif in a random DNA sequence (32). These findings are in accordance with our results showing a stronger inhibitory effect for GC-rich RNA sequences (e.g. RNA1) and particularly RNA sequences of switch region (e.g. RNA 4). Another recent study confirmed our findings by proposing an mRNA tether model for AID activity, which explains the concurrent binding of AID to RNA transcript and non-transcribed DNA strand during transcription. They suggest that AID binds to RNA through its assistant patch and deaminate the cytosine of ssDNA passing through the catalytic site (118).

In this study, we showed that the RNA inhibitory effect is sequence specific and observed more significantly by RNA sequences of switch region. Given that high concentrations of DNA and RNA may inhibit enzyme activity through interfering interactions, we measured the IC50 for DNA or RNA of selected sequences. The estimated IC50 for RNAs of the S region was lower than the same sequence of DNA, which demonstrates AID’s preference for binding to RNA or its higher binding affinity for RNA than DNA of the switch region sequence. On the contrary, RNA sequences of the variable regions have slightly higher IC50 than DNA sequences. In fact, RNA oligonucleotides of variable regions do not show sequence specific binding to AID

leading inhibition. Altogether, RNA binding may regulate human AID activity on the switch regions during CSR, which is different from its role in the variable regions.

RNA sequences which were used for switch regions were the smallest repetitive units of each switch region and had different lengths of Sα (80bp), Sγ1 (48bp), and Sμ (20bp). As the inhibitory effect was consistent and even more outstanding for Sμ (20bp), this effect is not dependent on the length of RNA or any structure formed by more than 20 nucleotides. It is worth noting that different RNA sequences of the same length (RNA1 and RNA 3, RNA 2 and RNA4) showed various effects on the inhibition of AID activity, as demonstrated in Figure 17. Despite this evidence, the effect of RNA sequences of the switch regions with the same length needs to be measured in order to exclude the length variable and have a better understanding of their sequence-specific effects on AID activity.

Additionally, the contribution of G-quadruplex DNA structures in AID oligomerization and accumulation in the Ig regions has been suggested previously (21). One important avenue of investigation would be to assess the formation of G-quadruplex or any other secondary structures with RNA sequences that inhibit AID activity, as these sequences are known to form complex structures, and K ions are present in the reaction composition (Table 8).

Although the inhibitory effect of RNA oligos was consistent among the reactions of purified AID treated with RNase A or not, we avoided the impact of inevitable remaining nucleases in the purified sample by adding an RNase inhibitor (Figure 19). Nevertheless, considering that RNA is not as stable as DNA, we should also test the integrity of RNA in the context of this assay. Therefore, the radioactively labelled RNA should be incubated with AID preparation to identify if RNA is degraded and the extent of this degradation. Running labeled RNA on the native gel will also indicate the formation of secondary structures and can be applied to electromobility shift assays to measure AID binding affinity to various RNAs or the effect of RNA cold inhibitor on the binding of AID to its DNA substrate.

Consequently, we proposed a model for AID regulation in CSR through transcribed noncoding RNA sequences, which do not affect SHM. RNA sequences in switch regions may recruit AID to target these regions and enhance its activity to

promote CSR. However, a high amount of transcribed switch region RNA, which is produced due to the existence of super enhancer genes in the Ig locus, could hinder AID activity through negative feedback regulation to prevent extensive double-strand breaks and translocations. Conversely, in the variable regions where excess mutation helps generate high-affinity antibodies, transcribed RNA does not have such an inhibitory interaction with AID. (Figure 25)

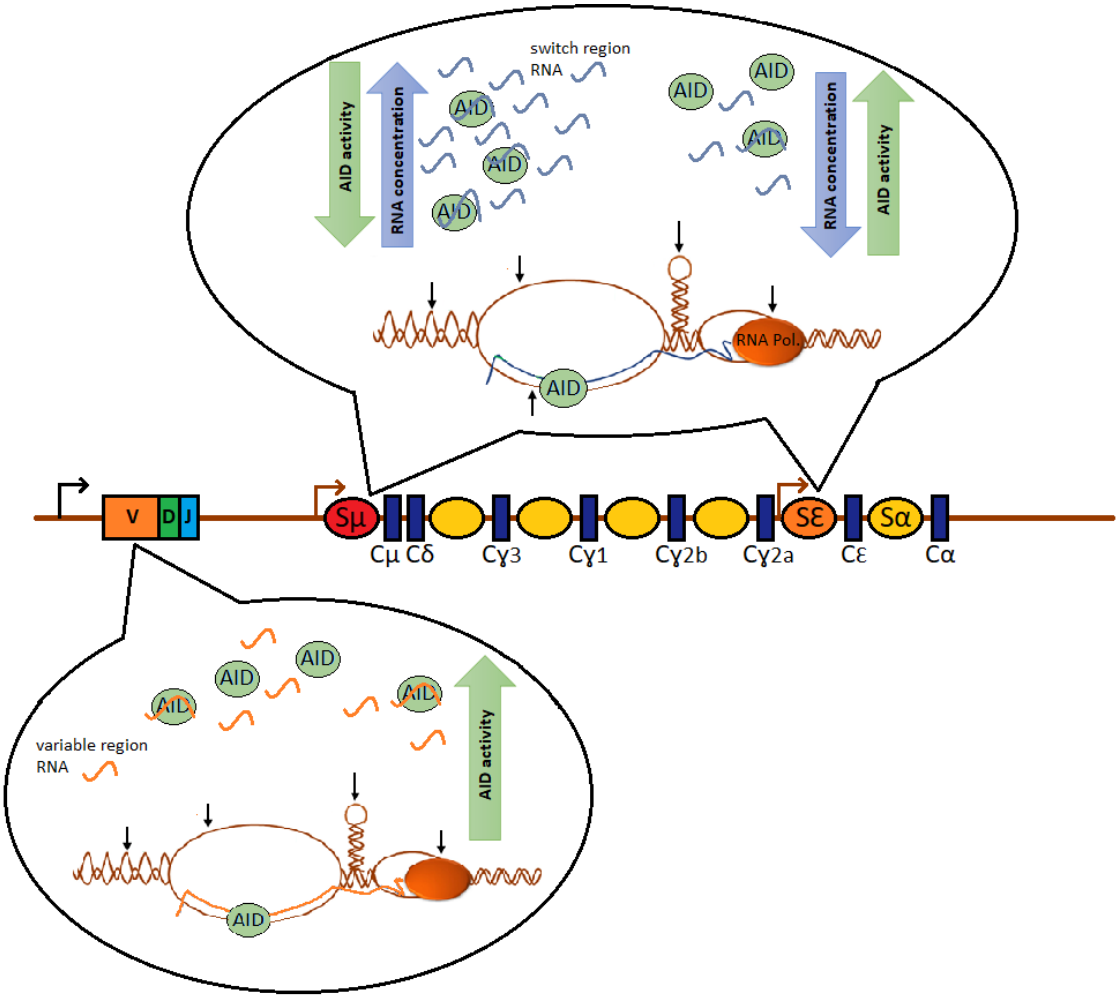


Figure 25. A proposed model of RNA dependent AID regulation.

Overall, we propose a novel role for RNA transcripts in regulating AID during CSR that requires confirmation through further experiments, including cellular and in vivo assays. One approach is to transfect stable cell lines expressing AID with shRNA containing specific inhibitory RNAs identified in this study and measure AID activity in those cells. Another in vitro method to evaluate the effect of transcribed RNAs on AID

activity is the degenerate PCR assay, as previously published by Branton et al.(56). Briefly, we can introduce a T7 promoter upstream of the variable or switch region sequences in a plasmid and incubate it with AID in the presence of T7 polymerase transcription. Then, amplify the target sequence using degenerate primers or AT-rich primers designed to anneal with both ends of the target sequence, even if it is mutated. Subsequently, sequence the PCR amplicon using next-generation sequencing or transform it to bacteria and send each colony for Sanger sequencing. Finally, computational docking studies of RNA-AID interactions could provide insight into the structural features that result in RNA binding and the potential mechanism for its regulatory effect.

References

1. Owen JA, Punt J, Stranford SA. Kuby immunology: WH Freeman New York; 2013.
2. Roh JS, Sohn DH. Damage-associated molecular patterns in inflammatory diseases. Immune network. 2018;18(4).
3. Abbas AK, Lichtman AH, Pillai S. Cellular and molecular immunology E-book: Elsevier Health Sciences; 2014.
4. Nimmerjahn F, Ravetch JV. Antibody-mediated modulation of immune responses. Immunological reviews. 2010;236(1):265-75.
5. Sun JC, Beilke JN, Lanier LL. Adaptive immune features of natural killer cells. Nature. 2009;457(7229):557-61.
6. Schroeder Jr HW, Cavacini L. Structure and function of immunoglobulins. Journal of Allergy and Clinical Immunology. 2010;125(2):S41-S52.
7. Considine DM, Considine GD. Van Nostrand's scientific encyclopedia: Springer Science & Business Media; 2013.
8. Gould HJ, Sutton BJ. IgE in allergy and asthma today. Nature Reviews Immunology. 2008;8(3):205-17.
9. Cerutti A, Chen K, Chorny A. Immunoglobulin responses at the mucosal interface. Annual review of immunology. 2011;29:273.
10. Mond JJ, Lees A, Snapper CM. T cell-independent antigens type 2. Annual review of immunology. 1995;13(1):655-92.
11. Pieper K, Grimbacher B, Eibel H. B-cell biology and development. Journal of Allergy and Clinical Immunology. 2013;131(4):959-71.
12. Staudt LM, Dave S. The biology of human lymphoid malignancies revealed by gene expression profiling. Advances in immunology. 2005;87:163-208.
13. Silvas TV, Schiffer CA. APOBEC3s: DNA-editing human cytidine deaminases. Protein Science. 2019;28(9):1552-66.
14. Stavnezer J, Schrader CE. IgH chain class switch recombination: mechanism and regulation. The Journal of Immunology. 2014;193(11):5370-8.
15. Pone EJ, Zhang J, Mai T, White CA, Li G, Sakakura JK, et al. BCR-signalling synergizes with TLR-signalling for induction of AID and immunoglobulin class-switching through the non-canonical NF-κB pathway. Nature communications. 2012;3(1):767.
16. Muto T, Muramatsu M, Taniwaki M, Kinoshita K, Honjo T. Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics. 2000;68(1):85-8.
17. Muramatsu M, Sankaranand V, Anant S, Sugai M, Kinoshita K, Davidson NO, et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. Journal of Biological Chemistry. 1999;274(26):18470-6.
18. Boehm T, Hess I, Swann JB. Evolution of lymphoid tissues. Trends in immunology. 2012;33(6):315-21.
19. Saunders HL, Oko AL, Scott AN, Fan CW, Magor BG. The cellular context of AID expressing cells in fish lymphoid tissues. Developmental & Comparative Immunology. 2010;34(6):669-76.
20. Larijani M, Martin A. Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. Molecular and cellular biology. 2007;27(23):8038-48.

21. Qiao Q, Wang L, Meng F-L, Hwang JK, Alt FW, Wu H. AID recognizes structured DNA for class switch recombination. Molecular cell. 2017;67(3):361-73. e4.
22. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. Molecular and cellular biology. 2007;27(1):20-30.
23. Choudhary M, Tamrakar A, Singh AK, Jain M, Jaiswal A, Kodgire P. AID Biology: A pathological and clinical perspective. International Reviews of Immunology. 2018;37(1):37-56.
24. Zan H, Casali P. Regulation of Aicda expression and AID activity. Autoimmunity. 2013;46(2):83-101.
25. Feng Y, Seija N, Di Noia JM, Martin A. AID in antibody diversification: there and back again. Trends in immunology. 2020;41(7):586-600.
26. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome biology. 2008;9(6):1-10.
27. Salter JD, Bennett RP, Smith HC. The APOBEC protein family: united by structure, divergent in function. Trends in biochemical sciences. 2016;41(7):578-94.
28. Pecori R, Di Giorgio S, Paulo Lorenzo J, Nina Papavasiliou F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. Nature Reviews Genetics. 2022;23(8):505-18.
29. King J, Manuel C, Barrett C, Raber S, Lucas H, Sutter P, et al. Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. Structure. 2015;23(4):615-27.
30. Gajula KS, Huwe PJ, Mo CY, Crawford DJ, Stivers JT, Radhakrishnan R, et al. High-throughput mutagenesis reveals functional determinants for DNA targeting by activation-induced deaminase. Nucleic acids research. 2014;42(15):9964-75.
31. Zheng S, Vuong BQ, Vaidyanathan B, Lin J-Y, Huang F-T, Chaudhuri J. Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. Cell. 2015;161(4):762-73.
32. Abdouni HS, King JJ, Ghorbani A, Fifield H, Berghuis L, Larijani M. DNA/RNA hybrid substrates modulate the catalytic activity of purified AID. Molecular immunology. 2018;93:94-106.
33. King JJ, Larijani M. Structural plasticity of substrate selection by activation-induced cytidine deaminase as a regulator of its genome-wide mutagenic activity. FEBS letters. 2021;595(1):3-13.
34. King JJ, Larijani M. A novel regulator of activation-induced cytidine deaminase/APOBECs in immunity and cancer: Schrödinger's CATalytic pocket. Frontiers in immunology. 2017;8:351.
35. King JJ, Manuel CA, Barrett CV, Raber S, Lucas H, Sutter P, et al. Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. Structure. 2015;23(4):615-27.
36. Lindley RA, Humbert P, Larner C, Akmeemana EH, Pendlebury CR. Association between targeted somatic mutation (TSM) signatures and HGS-OvCa progression. Cancer medicine. 2016;5(9):2629-40.
37. Takizawa M, Tolarová H, Li Z, Dubois W, Lim S, Callen E, et al. AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. The Journal of experimental medicine. 2008;205(9):1949-57.
38. Burns A, Alsolami R, Becq J, Stamatopoulos B, Timbs A, Bruce D, et al. Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHVmut and IgHVunmut subgroups. Leukemia. 2018;32(2):332-42.

39. Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, et al. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell. 2008;135(6):1028-38.
40. Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nature communications. 2015;6(1):1-12.
41. Pasqualucci L, Guglielmino R, Houldsworth J, Mohr J, Aoufouchi S, Polakiewicz R, et al. Expression of the AID protein in normal and neoplastic B cells. Blood. 2004;104(10):3318-25.
42. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proceedings of the National Academy of Sciences. 2012;109(10):3879-84.
43. Nakamura M, Sugita K, Sawada Y, Yoshiki R, Hino R, Tokura Y. High levels of activation-induced cytidine deaminase expression in adult T-cell leukaemia/lymphoma. British Journal of Dermatology. 2011;165(2):437-9.
44. Fukita Y, Jacobs H, Rajewsky K. Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity. 1998;9(1):105-14.
45. Storb U, Peters A, Klotz E, Kim N, Shen H, Kage K, et al. Somatic hypermutation of immunoglobulin genes is linked to transcription. Somatic Diversification of Immune Responses. 1998:11-9.
46. Singh AK, Tamrakar A, Jaiswal A, Kanayama N, Agarwal A, Tripathi P, et al. Splicing regulator SRSF1-3 that controls somatic hypermutation of IgV genes interacts with topoisomerase 1 and AID. Molecular immunology. 2019;116:63-72.
47. Nowak U, Matthews AJ, Zheng S, Chaudhuri J. The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. Nature immunology. 2011;12(2):160-6.
48. Hu W, Begum NA, Mondal S, Stanlie A, Honjo T. Identification of DNA cleavage-and recombination-specific hnRNP cofactors for activation-induced cytidine deaminase. Proceedings of the National Academy of Sciences. 2015;112(18):5791-6.
49. Conticello SG, Ganesh K, Xue K, Lu M, Rada C, Neuberger MS. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. Molecular cell. 2008;31(4):474-84.
50. Nambu Y, Sugai M, Gonda H, Lee C-G, Katakai T, Agata Y, et al. Transcription-coupled events associating with immunoglobulin switch region chromatin. Science. 2003;302(5653):2137-40.
51. Chaudhuri J, Khuong C, Alt FW. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature. 2004;430(7003):992-8.
52. Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, et al. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell. 2010;143(1):122-33.
53. Willmann KL, Milosevic S, Pauklin S, Schmitz K-M, Rangam G, Simon MT, et al. A role for the RNA pol II–associated PAF complex in AID-induced immune diversification. Journal of Experimental Medicine. 2012;209(11):2099-111.
54. Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon K-R, Resch W, et al. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell. 2014;159(7):1524-37.
55. Meng F-L, Du Z, Federation A, Hu J, Wang Q, Kieffer-Kwon K-R, et al. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. Cell. 2014;159(7):1538-48.

56. Branton SA, Ghorbani A, Bolt BN, Fifield H, Berghuis LM, Larijani M. Activation-induced cytidine deaminase can target multiple topologies of double-stranded DNA in a transcription-independent manner. The FASEB Journal. 2020;34(7):9245-68.
57. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. Annual review of biochemistry. 2009;78(1):273-304.
58. Shen HM, Poirier MG, Allen MJ, North J, Lal R, Widom J, et al. The activation-induced cytidine deaminase (AID) efficiently targets DNA in nucleosomes but only during transcription. Journal of Experimental Medicine. 2009;206(5):1057-71.
59. Yu K, Chedin F, Hsieh C-L, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nature immunology. 2003;4(5):442-51.
60. Roy D, Zhang Z, Lu Z, Hsieh C-L, Lieber MR. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Molecular and cellular biology. 2010;30(1):146-59.
61. Kouzine F, Gupta A, Baranello L, Wojtowicz D, Ben-Aissa K, Liu J, et al. Transcription-dependent dynamic supercoiling is a short-range genomic force. Nature structural & molecular biology. 2013;20(3):396-403.
62. Kuetche VK. Ab initio bubble-driven denaturation of double-stranded DNA: self-mechanical theory. Journal of Theoretical Biology. 2016;401:15-29.
63. Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nature Reviews Immunology. 2004;4(7):541-52.
64. Canugovi C, Samaranayake M, Bhagwat AS. Transcriptional pausing and stalling causes multiple clustered mutations by human activation-induced deaminase. The FASEB Journal. 2009;23(1):34-44.
65. Rajagopal D, Maul RW, Ghosh A, Chakraborty T, Khamlichi AA, Sen R, et al. Immunoglobulin switch μ sequence causes RNA polymerase II accumulation and reduces dA hypermutation. Journal of Experimental Medicine. 2009;206(6):1237-44.
66. Basu U, Meng F-L, Keim C, Grinstein V, Pefanis E, Eccleston J, et al. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. Cell. 2011;144(3):353-63.
67. Boursier L, Su W, Spencer J. Analysis of strand biased'G'· C hypermutation in human immunoglobulin Vλ gene segments suggests that both DNA strands are targets for deamination by activation-induced cytidine deaminase. Molecular immunology. 2004;40(17):1273-8.
68. Iyer LM, Zhang D, Rogozin IB, Aravind L. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic acids research. 2011;39(22):9473-97.
69. Ghorbani A, Quinlan EM, Larijani M. Evolutionary Comparative Analyses of DNA-Editing Enzymes of the Immune System: From 5-Dimensional Description of Protein Structures to Immunological Insights and Applications to Protein Engineering. Frontiers in Immunology. 2021;12:642343.
70. Quinlan EM, King JJ, Amemiya CT, Hsu E, Larijani M. Biochemical regulatory features of activation-induced cytidine deaminase remain conserved from lampreys to humans. Molecular and cellular biology. 2017;37(20):e00077-17.
71. Dancyger AM, King JJ, Quinlan MJ, Fifield H, Tucker S, Saunders HL, et al. 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. The FASEB Journal. 2012;26(4):1517-25.

72. Ichikawa HT, Sowden MP, Torelli AT, Bachl Jr, Huang P, Dance GS, et al. Structural phylogenetic analysis of activation-induced deaminase function. The Journal of Immunology. 2006;177(1):355-61.
73. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. Molecular biology and evolution. 2005;22(2):367-77.
74. Abdouni H, King JJ, Suliman M, Quinlan M, Fifield H, Larijani M. Zebrafish AID is capable of deaminating methylated deoxycytidines. Nucleic acids research. 2013;41(10):5457-68.
75. Bhutani N, Decker MN, Brady JJ, Bussat RT, Burns DM, Corbel SY, et al. A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. The FASEB Journal. 2013;27(3):1107-13.
76. Ao X, Sa R, Wang J, Dao R, Wang H, Yu H. Activation-induced cytidine deaminase selectively catalyzed active DNA demethylation in pluripotency gene and improved cell reprogramming in bovine SCNT embryo. Cytotechnology. 2016;68(6):2637-48.
77. Moon S-Y, Eun H-J, Baek S-K, Jin S-J, Kim T-S, Kim S-W, et al. Activation-induced Cytidine Deaminase induces DNA Demethylation of Pluripotency genes in bovine differentiated cells. Cellular Reprogramming (Formerly" Cloning and Stem Cells"). 2016;18(5):298-308.
78. Dominguez PM, Teater M, Chambwe N, Kormaksson M, Redmond D, Ishii J, et al. DNA methylation dynamics of germinal center B cells are mediated by AID. Cell reports. 2015;12(12):2086-98.
79. Roth O, Solbakken MH, Tørresen OK, Bayer T, Matschiner M, Baalsrud HT, et al. Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes. Proceedings of the National Academy of Sciences. 2020;117(17):9431-9.
80. Swann JB, Holland SJ, Petersen M, Pietsch TW, Boehm T. The immunogenetics of sexual parasitism. Science. 2020;369(6511):1608-15.
81. Ghorbani A, Khataeipour SJ, Solbakken MH, Huebert DN, Khoddami M, Eslamloo K, et al. Ancestral reconstruction reveals catalytic inactivation of activation-induced cytidine deaminase concomitant with cold water adaption in the Gadiformes bony fish. BMC biology. 2022;20(1):1-32.
82. Malecek K, Brandman J, Brodsky JE, Ohta Y, Flajnik MF, Hsu E. Somatic hypermutation and junctional diversification at Ig heavy chain loci in the nurse shark. The Journal of Immunology. 2005;175(12):8105-15.
83. Dooley H, Stanfield RL, Brady RA, Flajnik MF. First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. Proceedings of the National Academy of Sciences. 2006;103(6):1846-51.
84. Wilson M, Hsu E, Marcuz A, Courtet M, Du Pasquier L, Steinberg C. What limits affinity maturation of antibodies in Xenopus--the rate of somatic mutation or the ability to select mutants? The EMBO Journal. 1992;11(12):4337-47.
85. Mehr R, Edelman H, Sehgal D, Mage R. Analysis of mutational lineage trees from sites of primary and secondary Ig gene diversification in rabbits and chickens. The Journal of Immunology. 2004;172(8):4790-6.
86. Yang F, Waldbieser GC, Lobb CJ. The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. The Journal of Immunology. 2006;176(3):1655-67.
87. Kojima F, Frolov A, Matnani R, Woodward JG, Crofford LJ. Reduced T Cell–Dependent Humoral Immune Response in Microsomal Prostaglandin E Synthase-1 Null

Mice Is Mediated by Nonhematopoietic Cells. The Journal of Immunology. 2013;191(10):4979-88.
88. Marianes AE, Zimmerman AM. Targets of somatic hypermutation within immunoglobulin light chain genes in zebrafish. Immunology. 2011;132(2):240-55.
89. Hackney JA, Misaghi S, Senger K, Garris C, Sun Y, Lorenzo MN, et al. DNA targets of AID: evolutionary link between antibody somatic hypermutation and class switch recombination. Advances in immunology. 2009;101:163-89.
90. Larijani M, Frieder D, Basit W, Martin A. The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in ung−/− msh2−/− mice. Immunogenetics. 2005;56(11):840-5.
91. Wei L, Chahwan R, Wang S, Wang X, Pham PT, Goodman MF, et al. Overlapping hotspots in CDRs are critical sites for V region diversification. Proceedings of the National Academy of Sciences. 2015;112(7):E728-E37.
92. Oreste U, Coscia MR. Specific features of immunoglobulin VH genes of the Antarctic teleost Trematomus bernacchii. Gene. 2002;295(2):199-204.
93. Detanico T, Phillips M, Wysocki LJ. Functional versatility of AGY serine codons in immunoglobulin variable region genes. Frontiers in immunology. 2016;7:525.
94. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics. 2008;9(1):1-8.
95. Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, et al. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics. 2003;50(3):437-50.
96. Olsson MH, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. Journal of chemical theory and computation. 2011;7(2):525-37.
97. Holland SJ, Berghuis LM, King JJ, Iyer LM, Sikora K, Fifield H, et al. Expansions, diversification, and interindividual copy number variations of AID/APOBEC family cytidine deaminase genes in lampreys. Proceedings of the National Academy of Sciences. 2018;115(14):E3211-E20.
98. Hsu E. Assembly and expression of shark Ig genes. The Journal of Immunology. 2016;196(9):3517-23.
99. Clayton E, Munir M. Fundamental characteristics of bat interferon systems. Frontiers in cellular and infection microbiology. 2020;10:527921.
100. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.
101. Ghorbani A, King JJ, Larijani M. The optimal pH of AID is skewed from that of its catalytic pocket by DNA-binding residues and surface charge. Biochemical Journal. 2022;479(1):39-55.
102. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nature chemical biology. 2012;8(9):751-8.
103. Smith NC, Rise ML, Christian SL. A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish. Frontiers in immunology. 2019;10:2292.
104. Wakae K, Magor BG, Saunders H, Nagaoka H, Kawamura A, Kinoshita K, et al. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. International immunology. 2006;18(1):41-7.
105. Barreto VM, Pan-Hammarstrom Q, Zhao Y, Hammarstrom L, Misulovin Z, Nussenzweig MC. AID from bony fish catalyzes class switch recombination. The Journal of experimental medicine. 2005;202(6):733-8.

106. Barreto VM, Magor BG. Activation-induced cytidine deaminase structure and functions: a species comparative view. Developmental & Comparative Immunology. 2011;35(9):991-1007.
107. Marr S, Morales H, Bottaro A, Cooper M, Flajnik M, Robert J. Localization and differential expression of activation-induced cytidine deaminase in the amphibian Xenopus upon antigen stimulation and during early development. The Journal of Immunology. 2007;179(10):6783-9.
108. Zhu C, Lee V, Finn A, Senger K, Zarrin AA, Du Pasquier L, et al. Origin of immunoglobulin isotype switching. Current Biology. 2012;22(10):872-80.
109. Ratcliffe MJ. B cell development in gut associated lymphoid tissues. Veterinary immunology and immunopathology. 2002;87(3-4):337-40.
110. Honjo T, Muramatsu M, Fagarasan S. AID: how does it aid antibody diversity? Immunity. 2004;20(6):659-68.
111. Pavri R, Nussenzweig MC. AID targeting in antibody diversity. Advances in immunology. 2011;110:1-26.
112. Diamond CP, Im J, Button EA, Huebert DN, King JJ, Borzooee F, et al. AID, APOBEC3A and APOBEC3B efficiently deaminate deoxycytidines neighboring DNA damage induced by oxidation or alkylation. Biochimica et Biophysica Acta (BBA)-General Subjects. 2019;1863(11):129415.
113. Smith HC, Bennett RP, Kizilyer A, McDougall WM, Prohaska KM, editors. Functions and regulation of the APOBEC family of proteins. Seminars in cell & developmental biology; 2012: Elsevier.
114. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proceedings of the National Academy of Sciences. 2003;100(7):4102-7.
115. Mondal S, Begum NA, Hu W, Honjo T. Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. Proceedings of the National Academy of Sciences. 2016;113(11):E1545-E54.
116. Yewdell WT, Kim Y, Chowdhury P, Lau CM, Smolkin RM, Belcheva KT, et al. A hyper-IgM syndrome mutation in activation-induced cytidine deaminase disrupts G-quadruplex binding and genome-wide chromatin localization. Immunity. 2020;53(5):952-70. e11.
117. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-19.
118. Liu D, Goodman MF, Pham P, Yu K, Hsieh C-L, Lieber MR. The mRNA tether model for activation-induced deaminase and its relevance for Ig somatic hypermutation and class switch recombination. DNA repair. 2022;110:103271.