





**Biochemical analysis of Activation-induced cytidine deaminase (AID) from bony fish: insights into the mechanisms of AID**

by

Alex Dancyger (B. Sc Honours)

© Alex Dancyger

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Masters of Science from the department of BioMedical Sciences of Memorial University of Newfoundland

May 2012

Immunology and Infectious Diseases Research Program

Division of BioMedical Sciences

Memorial University of Newfoundland

St. John's, NL A1B 3V6

## Abstract

The properties of Activation-induced cytidine deaminase (AID) necessary for its role in immunity and lymphoma generation are currently unknown. AID mutants and species-variants, as compared to wild-type human AID, were identified so as to map out responsible residues/domains. Variant AIDs were analyzed by alkaline cleavage and EMSA. AID from bony fish species (*Ictalurus punctatus* and *Danio rerio*) were identified as the most divergent from human AID. Dr-AID acted enzymatically similar to H-AID, while Ip-AID was less mutagenic. Point mutations within Ip AID were generated, substituting the human counterpart amino acid. One mutant rescued Ip-AID enzyme activity to levels seen in H-AID and Dr-AID. The reverse mutation in H-AID almost fully abolished the deaminating ability of H-AID. DNA-binding kinetics were tested at different temperatures to analyze if the disparity in activity levels was due to binding of DNA. The single mutation in IP-AID increases the affinity of AID to bind ssDNA and this effect is substantially more pronounced at lower temperatures. These results have provided novel information on AID domains/residues necessary for its function.

### **Acknowledgments**

I would like to take this opportunity to thank my supervisor, Dr. Mani Larijani. I could not have asked for a better teacher and mentor and cannot thank him enough for his support, encouragement and patience. Dr. Larijani is responsible for sparking my interest in immunology and showing me how dynamic the field can be. Thank you for bringing me to Newfoundland, teaching me how to be a better scientist and helping me grow as a person.

I would also like to show gratitude to the members of my committee, Dr. Rodney Russell and Dr. Sheila Drover. They have been an essential part of my Masters experience, not only for their scientific teaching, insight and advice but for their moral support and guidance. I believe they have truly gone above and beyond their duties as committee members.

I would like to thank everyone in the Larijani Lab for their technical and moral support. I am thankful to the staff and students both past and present, for their help, encouragement and friendship. I am also appreciative to the Canadian Institutes of Health Research (CIHR) for honouring me with a Master's Award: Frederick Banting and Charles Best Canada Graduate Scholarship.

I would like to express my appreciation for my family and friends. Their continuous love and encouragement have been an essential part in accomplishing my goals. Thank you for listening, for inspiring me, helping me through the hard times and being there to celebrate the good times. Lastly, a special thank you goes to my mother and

father who have been there every step of the way helping and guiding me; I could not have done this without you and only wish I can somehow return the favour one day.

**Note:** The work in this thesis has been accepted for publication in FASEB J. as the following article:

“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”

Alex M Dancyger, Justin J King, Matthew J Quinlan, Heather Fifield, Stephanie Tucker, Holly L Saunders, Maribel Berru, Bradley G. Magor, Alberto Martin & Mani Larjani

I, Alex Dancyger, completed 90% of the work described in the paper noted above, however Justin J King and Matthew J Quinlan carried out some experiments, approximately 10%, under my supervision.

## Table of Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Acknowledgements</b> .....	<b>iii</b>
<b>Tables of Contents</b> .....	<b>v</b>
<b>List of Tables</b> .....	<b>vii</b>
<b>List of Figures</b> .....	<b>viii</b>
<b>List of Abbreviations and Symbols</b> .....	<b>x</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1 The immune system .....	1
1.2 Primary antibody diversification .....	2
1.3 Secondary antibody diversification: the role of Activation-induced cytidine deaminase (AID) .....	3
1.4 The role of AID in cancer .....	12
1.5 The mechanisms behind the mutagenic activities of AID .....	13
1.6 AID is a member of the APOBECS family of cytidine deaminases .....	19
1.7 Structural features and enzymatic properties of human AID .....	20
1.8 Evolution of AID amongst species .....	24
1.9 Purpose of this study: discovering the structure function relationship within AID .....	26
<b>2 Materials and Methods</b> .....	<b>29</b>
2.1 Plasmids and constructs used for AID expression .....	29
2.2 Site directed mutagenesis .....	29

2.3 AID protein production and purification .....	31
2.4 Substrate preparation .....	33
2.5 Deamination assay .....	34
2.6 Electrophoretic mobility shift assay .....	35
2.7 Quantification of data .....	36
2.8 Statistical analysis .....	37
<b>3 Results .....</b>	<b>38</b>
3.1 The catfish AID sequence is most divergent from other AID species .....	38
3.2 Differences in the activity levels of H-AID, Dr-AID and Ip-AID .....	45
3.3 The activity of Dr-AID and Ip-AID are temperature dependent .....	53
3.4 The basis of the low activity of Ip-AID at 37°C .....	59
3.5 The effects of G176D mutation on Ip-AID activity .....	72
3.6 The D176 residue in Ip-AID suppresses DNA binding activity .....	78
<b>4 Discussion .....</b>	<b>86</b>
4.1 Overall rational and most significant findings .....	86
4.2 The mechanisms of activity modulation by D176 .....	88
4.3 The cold adaptation of bony fish AID .....	89
4.4 The significance of our findings with respect to the evolution of AID .....	91
4.5 Conclusions and future work .....	94
<b>5 References Appendices .....</b>	<b>96</b>

## List of Tables

<b>Table</b>	<b>Title</b>	<b>Page</b>
1	$K_d$ values of H-AID, Ip-AID and Ip-AID D176G binding to ssDNA at various temperatures.	84

## List of Figures

Figure	Title	Page
1	A simplified schematic representation of the structure of an IgG antibody molecule.	6
2	V(D)J-recombination during B-lymphocyte development	8
3	Secondary antibody repertoire generation by Somatic Hypermutation (SHM) and Class Switch Recombination (CSR)	10
4	The DNA-deamination processes of Somatic Hypermutation (SHM)	17
5	A schematic representation of AID.	22
6	Sequence alignment of AID from representative species across classes	41
7	Percent identity and similarity between 3 human enzymes.	43
8	Typical substrate used in the <i>in vitro</i> alkaline cleavage activity assay.	47
9	<i>In vitro</i> activity of human and bony fish AID.	49
10	Comparison of activity kinetics between Hs-AID, Dr-AID and Ip-AID.	51
11	Temperature and substrate preference of bony fish AID.	55
12	Comparison of deamination kinetics between H-AID, Dr-AID and Ip-AID under optimal conditions.	57
13	Predicted structures of Hs-AID, Dr-AID and Ip-AID monomers.	62
14	Analysis of hybrid AIDs.	64
15	Alignment of the carboxyl terminal of Hs-AID, Dr-AID and Ip-AID.	66
16	The activity of Ip-AID mutants.	68

17	Ip-AID D176G is identified as a rescue mutation.	70
18	The effects of the D176G mutation on Ip-AID activity.	74
19	Comparison of the deamination activities of various AIDs at 37°C, 25°C and 18°C.	76
20	Surface structure models of Hs-AID, Dr-AID, Ip-AID and APOBEC3G.	80
21	EMSA measuring the binding affinity of AID to ssDNA at different temperatures.	82

## List of Abbreviations and Symbols

7	A gene knockout, where an organism's gene is made inoperative
µg	Microgram
µl	Microlitre
°C	Degrees celsius
A	Adenine
A3F	Apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3F
A3G	Apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3G
Ab	Antibody
AID	Activation-induced cytidine diaminase
Amp	Ampicillin
APE1	Apurinic/aprimidinic endonuclease
ApoB	Apolipoprotein B
APOBEC	Apolipoprotein B mRNA-editing, enzyme-catalytic polypeptide-like
BCL-6	B-cell lymphoma 6
BER	Base excision repair
Bp	Base pair
BSA	Bovine serum albumin
C	Cytidine
C	Constant exon
C region	Constant region of Ig

CSR	Class switch recombination
Cys	Cysteine
DCDT	dCMP diaminase
DLBCL	Diffuse large B cell lymphoma
DNA	Deoxyribonucleic acid
DSB	Double stranded break
DTT	Dithiothreitol
EDTA	Ethyldiamine tetraacetic acid
EMSA	Electrophoretic mobility shift assay
<i>g</i>	Gravitational force
g	Gram
G	Guanine
GST	Glutathione S-transferase
Ig	Immunoglobulin
IPTG	Isopropyl- $\beta$ thiogalactopyranoside
kDa	Kilodalton
LB	Lysogeny broth
M	Molar
ml	Millilitre
MMR	Mismatch repair
mRNA	Messenger ribonucleic acid
MSH2	MutS homolog 2
MSH6	MutS homolog 6

NES	Nuclear export signal
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
nt	Nucleotide
OD	Optical density
ORF	Open reading frame
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
RHOH	Ras homolog gene family member H
RAG	Recombination activating genes
RNA	Ribonucleic acid
RPA	Replication protein A
S	Switch region of Ig
SHM	Somatic hypermutation
SOC	Super optimal broth with catabolite repression
SSB	Single stranded break
ssDNA	Single stranded DNA
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
T	Thymidine
TBE	Tris borate/EDTA electrophoresis buffer
T <sub>c</sub>	cytotoxic T cell
TCR	T cell receptor
TEMED	Tetramethylethylenediamine

T <sub>H</sub>	T helper cell
U	Uracil
UNG	Uracil-DNA glycosylase, also abbreviated UDG in non-human species
WRC	a ssDNA hotspot motif targeted by AID where W=A/T and R=A/G
V	Variable region of Ig
V(D)J	Recombination of variable (V), diversity (D) and joining (J) gene segments
VLR	Variable lymphocyte receptors

## **1. Introduction**

### **1.1 The immune system**

The immune system is the body's natural defence, tasked with the recognition and clearance of pathogens, such as fungi, bacteria and viruses, within a host. It responds to a myriad of pathogens by being adaptive and self-tolerant. The innate immune system is a host's first line of defence. This system is comprised of non-specific mechanisms that provide immediate rather than long lasting protection. The adaptive immune response is the second line of defence and is composed of two types of leukocytes: T lymphocytes, whose function is to destroy infected cells and eradicate intracellular pathogens, and B lymphocytes, cells that secrete antibodies and attack extracellular pathogens. Bone marrow represents the dominant site for haemopoiesis where the majority of cells mature. However, T lymphocytes mature in the thymus, undergoing a series of positive and negative selections with only the former progressing to maturity and the latter being marked for termination by apoptosis (Fleisher & Tomar, 1997; Warrington et al., 2011; Lanasa and Weinberg, 2011).

There are three types of lymphocytes: T, B and NK cells. However, only T and B lymphocytes exhibit antigen-specific receptors and memory. T cells can be broadly divided into T helper ( $T_H$ ) and cytotoxic ( $T_C$ ) T cells. B cells are lymphocytes that produce antibodies and are responsible for humeral mediated immunity. Once B cells are activated by a foreign antigen, proliferation and maturation begins producing plasma cells that secrete antibodies. Not all B cells develop into plasma cells. Some remain as memory B cells conferring protection upon re-exposure to the same pathogen. Similarly, T cells undergo clonal selection generating memory T cells. T and B cells are constantly

circulating around the body, migrating to and from blood and tissues (Fleisher & Tomar, 1997; Warrington et al., 2011; Lanasa & Weinberg, 2011).

Antibodies have two major roles in the immune system: first, to bind antigen and second, to interact with host tissues and effector systems to eradicate foreign pathogens. The human immune system possesses five isotypes of antibodies: IgM, IgG (which contains 4 subclasses: IgG1-4), IgA, IgE and IgD. The basic structure of an antibody contains four polypeptide chains, two light chains and two heavy chains, which roughly form a “Y” shaped molecule (Figure 1). One end of the immunoglobulin, known as the Fab portion, contains the variable (V) regions of each of the light and heavy chains. This portion gives the antibody its specificity and is the region that binds the antigen. The Fc portion contains the constant (C) region of the heavy chains, is far less variable and determines the isotype and effector function of the antibody (Fleisher & Tomar, 1997).

### **1.2 Primary antibody diversification**

An essential step in the development of the vertebrate immune system is V(D)J recombination. This process of DNA rearrangement occurs early in T and B cell development generating functional immunoglobulins (Ig) and T cell receptors (TCR). This mechanism of genetic recombination is the only identified site-specific recombination in higher eukaryotes. Vertebrate Ig and TCR genes are arranged in a linear fashion and require recombination to be joined into a functional gene. Mammals possess seven antigen receptor loci: Ig H,  $\kappa$  and  $\lambda$  loci, and TCR  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  loci. Each locus contains a set of V and J segments, with Ig H and TCR $\beta$  and  $\delta$  possessing additional D segments flanked by V and J segments. During lymphoid cell maturity, various V-J

and V-D-J segments become linked to form variable and constant region exons that are spliced together (Figure 2). Failure of a cell to create a successful product that does not self-react results in apoptosis. However, with multiple types of each segment, there are many combinatorial possibilities. This phenomenon is responsible for generation the primary (pre-immune response) diversity of Ig and TCRs (Gellert, 2002).

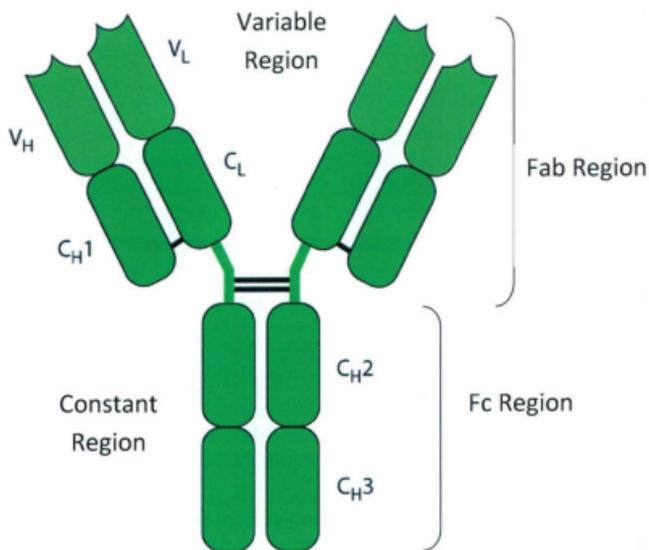
### **1.3 Secondary antibody diversification: the role of Activation-induced cytidine deaminase (AID)**

As mentioned previously, the vertebrate immune system has evolved to respond to a diverse array of pathogens by generating a seemingly infinite number of antigen receptors. The primary diversity of B cell antigen receptors is first generated through V(D)J recombination which is the process of rearrangement of gene segments encoding receptors. The outcome is a naïve (pre-immune) repertoire of antibodies covering broad specificities but of generally low affinity for a specific antigen. Upon exposure to an immune challenge, B lymphocytes are activated, migrate to germinal centres within lymph nodes and begin the process of proliferating and refining their antibody repertoire through secondary antibody diversification processes (Muramatsu et al., 2000; Revy et al., 2000; Martin et al., 2002; Okazaki et al., 2002). The molecular processes responsible for fine tuning the antibody response are mediated by somatic hypermutation (SHM) and class switch recombination (CSR) (Figure 3) (Martin et al., 2002; Muramatsu et al., 2000; Okazaki et al., 2002; Revy et al., 2000). SHM is the means whereby random point mutations are introduced in the variable (V) antigen-recognition exons of immunoglobulin (Ig) genes. Some mutations may result in an enhanced affinity to a

specific pathogen. B lymphocytes harbouring these mutations will be at an advantage for interaction with a limited amount of antigens resulting in cellular selection for survival and proliferation. This yields a pool of antibodies with higher affinities for antigens. In addition to affinity maturation, antibodies also switch isotypes during an immune response. During this process, the constant region of the antibody heavy chain is changed from the initial IgM to IgG, IgA or IgE, but the same variable region is retained. Therefore, class switching does not affect antigen specificity; instead it generates the ability for the antibody molecule to interact with different effector molecules and carry out a variety of effector functions necessary for the efficient clearance of pathogens (de Yebenes & Ramiro, 2006). At the molecular level isotype switching is mediated by class-switch recombination (CSR). The process of CSR is initiated by double stranded breaks (DSBs) which are created in the switch (S) regions upstream of the constant (C) exons in the Ig heavy chain. C $\mu$  is exchanged for a downstream C region and the intervening DNA is excised producing antibodies with new isotypes. While SHM and CSR are mediated by distinct DNA lesions and recruit different DNA repair machineries, both processes are initiated by the enzyme Activation-induced cytidine deaminase (AID) (Martin et al., 2002; Muramatsu et al., 2000; Okazaki et al., 2002; Revy et al., 2000; Barreto et al., 2005; Chaudhuri & Frederick, 2004; Longrich et al., 2006; Neuberger et al., 2005).

Human patients with genetic defects in AID as well as AID deficient mouse models suffer from Hyper-IgM syndrome, an immunodeficiency characterized by the inability to clear specific pathogens and carry out CSR leading to an absence of antibodies other than IgM. The result is recurrent bacterial infections, lymphoid

hyperplasia, autoimmune hepatitis, haemolytic anaemia, thrombocytopenia and an imbalance in a myriad of immune functions. Therefore AID is essential for secondary antibody diversification (Dooley et al., 2006; Durandy, Peron, & Fischer, 2006; Revy et al., 2000).

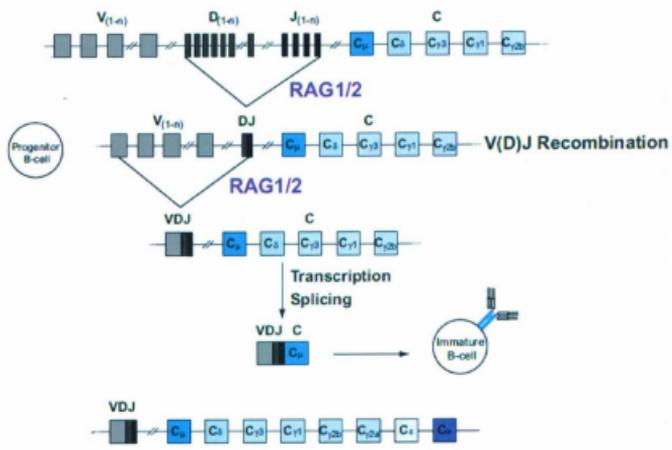


(modified from Roux, 1999)

**Figure 1. A simplified schematic representation of the structure of an IgG antibody molecule.**

**Figure 1. A simplified schematic representation of the structure of an IgG antibody molecule.**

The antibody molecule is composed of three globular regions that form a Y-shaped molecule. The two antigen-binding sites are at the tip of each arm. The antibody molecule is composed of two heavy (H) chain and two light (L) chains. The separate domains comprising each chain have been identified. The Fab fragment contains the variable (V) regions of the light and heavy chains and binds antigen. The Fc fragment contains the constant (C) regions of the immunoglobulin molecule.

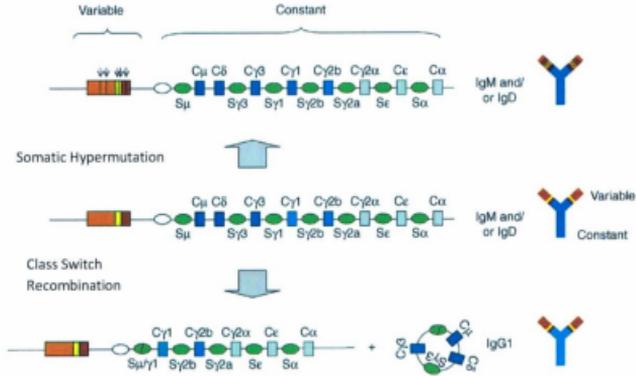


(modified from Ganesh and Neuburger, 2011)

Figure 2. V(D)J recombination during B lymphocyte development.

**Figure 2. V(D)J recombination during B lymphocyte development.**

V(D)J recombination is the genetic recombination process that occurs in the early development of immunoglobulins and T cell receptors. During B lymphocyte development V, D and J segments in the heavy chain and V and J segments in the light chain are rearranged by the lymphocyte specific enzymes recombination activating genes -1 and -2 (RAG1 and RAG2). RAG1 and RAG2 splice the variable and constant regions together. Many different combination are possible generating a diverse primary antibody repertoire.



(modified from de Yebenes & Ramiro, 2006)

**Figure 3. Secondary antibody repertoire generation by somatic hypermutation (SHM) and class switch recombination (CSR).**

**Figure 3. Secondary antibody repertoire generation by somatic hypermutation (SHM) and class switch recombination (CSR).**

SHM and CSR are the two molecular mechanisms responsible for generating the secondary diversification of antibodies. These processes occur in the germinal centres of lymph nodes once a B lymphocyte has been activated by exposure to an immune challenge. SHM introduces point mutations into the antigen-binding region of the antibody by nucleotide substitutions (arrows). As a result, rearrangement of the variable Ig genes occurs and antibodies with an increased affinity are generated. CSR is a region-specific recombination reaction taking place between the switch regions (green ovals). The  $C_{\mu}$  constant region is exchanged with a downstream constant region (blue boxes). The resulting antibody has a new effector function.

#### **1.4 The role of AID in cancer**

While the action of AID, inserting mutations into nucleic acids, is for the most part targeted to the immunoglobulin loci, this honing is not flawless. As a result AID acts promiscuously, mutating and creating DSBs throughout the B cell genome. Prior to the discovery of AID, it was proposed that a link existed between genes mutated following secondary antibody diversification and certain B-cell lymphomas with a major hallmark of disease being the chromosomal translocations between Ig genes and a proto-oncogene (Gordon et al., 2003; Müschen et al., 2000; Pasqualucci et al., 1998; Pasqualucci et al., 2001; Shen et al., 1998). Many of the most aggressive B-cell lymphomas are initiated in germinal-centre B cells and contain translocations in V or S regions of Ig genes (Küppers & Dalla-Favera, 2001; Willis & Dyer, 2000). Specific examples include the BCL-1-IgH translocation found in mantle-zone lymphoma, the BCL-2-Ig H translocation in follicular lymphomas and the C-MYC-IgH translocation in Burkitt's lymphoma. Proto-oncogenes *PIMI*, *MYC*, *RHOH* and *PAX5* are also subject to translocation in diffuse large B-cell lymphoma (DLBCL) (Pasqualucci et al., 2001). Translocation results in the control of proto-oncogene expression under the robust IgH enhancer and/or promoters and thus over-expression and/or constitutive expression (Küppers & & Dalla-Favera, 2001; Küppers, 2005). Several lines of evidence from both murine and human models have identified AID as being responsible for the development of lymphomas. First, many of these proto-oncogene translocations occur at or near WRC hotspot motifs, a ssDNA trinucleotide motif specifically targeted by AID with a preference for C (W=A/T, R=A/G), implicating AID involvement (Pasqualucci et al., 2001). Second, transgenic mice over-expressing AID were found to develop tumours (Okazaki et al., 2003). Third,

AID is often found expressed in cancer tissues and cell lines where it is not normally expressed (Bachl et al., 2001). Fourth, studies on transgenic IL-6 mice, a model for Burkitt's lymphoma, showed that AID is required for the *c-MYC/IgH* translocations when AID<sup>-/-</sup> IL-6 transgenic mice lacked these specific translocations (Ramiro et al., 2004). Fifth, mice constitutively expressing the proto-oncogene B-cell lymphoma 6 (*BCL-6*) demonstrated clonal splenic lymphomas, however when AID expression in these mice was downregulated lymphoma production decreased (Pasqualucci et al., 2008). Finally strong correlations have been observed between the severity of human chronic lymphocytic leukemia, DLBCL and AID expression (McCarthy et al., 2003; Reiniger et al., 2006). This growing body of evidence strongly supports the implications of AID activity in the transformation of activated B cells into leukemia and lymphomas as well as in disease aggressiveness.

### **1.5 The mechanisms behind the mutagenic activities of AID**

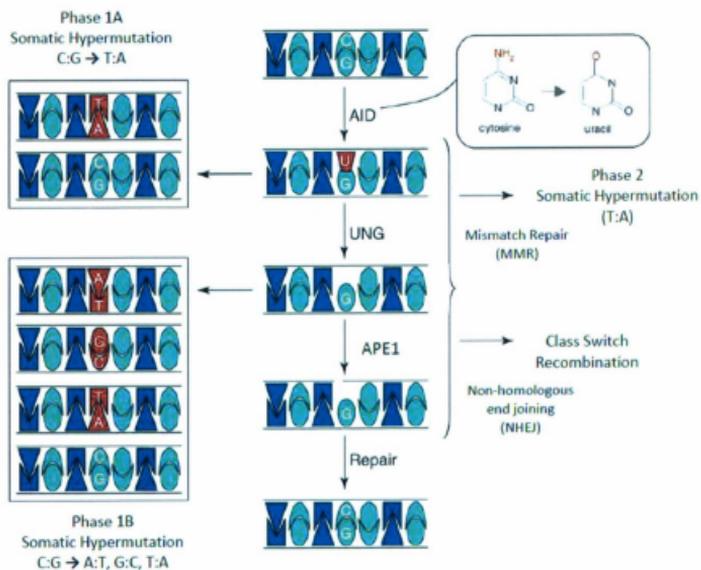
Upon its discovery in 1999 AID was initially thought to act as an RNA editing enzyme which through the modification of mRNA would result in the expression of a mutagenic protein that could in turn carry out SHM and CSR (Muramatsu et al., 1999; Muramatsu et al., 2000). This idea was largely based on AID's homology to the family member apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (APOBEC1), an mRNA editing enzyme. Upon further examination it was identified that AID is the only factor required for secondary antibody diversification processes (Conticello et al., 2005; Conticello et al., 2007). Support for this notion was provided when AID was heterogeneously expressed in fibroblasts and shown to be the only requirement necessary

to support SHM and CSR on transcribed substrates (Okazaki et al., 2002; Yoshikawa et al., 2002). Furthermore, interactions between AID and B cell S region DNA going through CSR was observed (Nambu et al., 2003). Thirdly, *in vitro* studies have shown that AID acts directly on ssDNA and not RNA, dsRNA or dsDNA (Bransteitter et al., 2003). These findings fit well with the previous observations that genes that are more highly transcribed are mutated more frequently by AID presumably because the process of transcription unwinds DNA making it more accessible (Yang & Schatz, 2007). Experiments show that AID specifically targets ssDNA with a preference for C within WRC hotspot motifs (W=A/T, R=A/G) (Beale et al., 2004; Pham et al., 2003). This preferential targeting of trinucleotide motifs parallels the pattern of mutations observed in the Ig loci (Larijani et al., 2005; Yang & Schatz, 2007). It is extremely unlikely that AID would act on an mRNA that currently remains unidentified, resulting in the expression of an AID mutator with the same WRC specificity as AID itself. Thus, AID directly mutates ssDNA at the Ig locus. The current paradigm most widely accepted is that AID deaminates C to U on ssDNA sequences of Ig genes. Since AID mediates only one type of mutation, C to U, other mutagenic processes are required to explain the pattern of SHM mutations observed at the Ig loci (Figure 4). These mutations could be from any one nucleotide (A, T, C, G) to any of the other four possible nucleotides. The initial AID mutagenesis creates a U:G mismatch in DNA. Previous work in the field of DNA repair has shown that this mismatch can be processed by a multitude of different DNA repair pathways. The U:G mismatch may be a substrate for either the base-excision repair (BER) or the mismatch repair (MMR) pathways of DNA damage repair. As a result, said U:G mismatch generated by AID has three possible fates. The first is that the U remains

unrepaired and is replicated across during cell division, resulting in a T:A transition mutation during replication. Alternatively the U:G mismatch could be subjected to one or both of two aforementioned DNA repair processes. The first is the removal of the U base by uracil-DNA glycosylase (UNG) generating an abasic site which could be prone to strand breakage by apurinic/apyridinic endonuclease (APE1) or a base substitution during replication resulting in the excision of several nucleotides surrounding the lesion. This overall process is known as BER. Alternatively, the U:G mismatch may be recognized by the mismatch repair (MMR) machinery consisting of MSH2/MSH6 molecules which physically recognizes the U:G mismatch bulge in the double helix, again resulting in the recruitment of downstream nucleases and excision of bases surrounding the U:G lesion. Under normal circumstances these lesions would be filled by high fidelity polymerases leading to repair back to the original unmutated sequence. Through mechanisms that are not yet understood, AID-induced mutations within germinal centre B cells are acted upon by error-prone polymerases. Through these mechanisms the initial C to U conversion by AID can ultimately result in the full range of mutations on any of the four nucleotides to any of the other three possible nucleotides (Chaudhuri & Frederick, 2004; Neuberger et al., 2005).

Examining the profile of SHM mutations in  $UDG^{-/-}$  or  $MSH2^{-/-}$  cell lines as well as mice showed that the pattern of mutations were consistent within these 2 pathways acting downstream of AID and the processing of AID and UDG generated lesions as described above (Chaudhuri & Frederick, 2004; Neuberger et al., 2005). Further support that AID acts as a DNA mutator came from studying UNG. UNG is responsible for excising U from the DNA backbone and was found to act downstream of AID.

Furthermore, downregulating UNG limits CSR while knocking out both UNG and MSH2, a mismatch repair protein, and eradicates both SHM and CSR (Rada, Di Noia, & Neuberger, 2004).



(modified from de Yebenes & Ramiro, 2006)

Figure 4. The DNA deamination processes of Somatic Hypermutation (SHM).

**Figure 4. The DNA deamination processes of Somatic Hypermutation (SHM).**

The process of SHM is proposed to occur in two phases. The first phases being the deamination of C to U by AID generating a U:G mismatch pairing which can have three potential outcomes: i) replicating over the U:G mismatch pairing leads to a transition mutation (Phase 1A SHM); ii) the U:G mismatch pairing can be subject to uracil excision by UNG creating an abasic site leading to either a transition or transversion mutation (Phase 1B SHM); or iii) the U:G mismatch pairing is recognized by MMR machinery generating an A:T mutation (Phase 2). The abasic site created by UNG excising the U residue may act as a substrate for APE1 leading to a nick in the DNA strand. This DNA nick can be repaired by the BER pathway or processed by non-homologous end joining (NHEJ) resulting in CSR.

### **1.6 AID is a member of the APOBEC family of cytidine deaminase enzymes**

AID is a member of the Apolipoprotein B mRNA-editing, enzyme-catalytic polypeptide-like family (APOBEC). The AID/APOBEC family members deaminate cytidine (in RNA) or deoxy-cytidine (in DNA) to uracil thus generating mutations in either DNA or RNA. The AID/APOBEC family has eleven members found in humans: AID, APOBEC-1, -2, -3 (A to H) and -4 (Navaratnam et al., 1993; Teng et al. 1993). The first family member identified was APOBEC1, an enzyme expressed in the human small intestine. The family is thus named because APOBEC1 edits the apolipoprotein B (ApoB) pre-mRNA by converting a glutamine codon, more specifically deaminating a cytidine at position 6666, into a stop codon generating a truncated form of ApoB (Navaratnam et al., 1993; Teng et al., 1993). Other family members within the AID/APOBEC family, such as the APOBEC3s, act as DNA mutators. These are involved in restricting retrovirus reproduction with the capability of limiting endogenous retroelements as well (Liddament et al., 2004; Sheehy et al., 2002; Simon et al., 2005). Besides AID, APOBEC3G (A3G) and APOBEC3F (A3F) have also been the subject of intense investigation because of their ability to mutate and act as restriction host factors to the HIV-1 genome. A hallmark of AID, A3G and A3F is their preference for deaminating the 3' cytidines in specific trinucleotide motifs (Sheehy et al., 2002). These trinucleotide motifs include CCC, TTT or WRC (W=A/T, R=A/G) (Conticello et al., 2005). This sequence specificity is an inherent property of these enzymes reflecting their footprint *in vivo* (Coker & Petersen, 2007; Larijani et al., 2005; Pham et al., 2003). Currently the functions of APOBEC2 and APOBEC4 have not been characterized. Thus

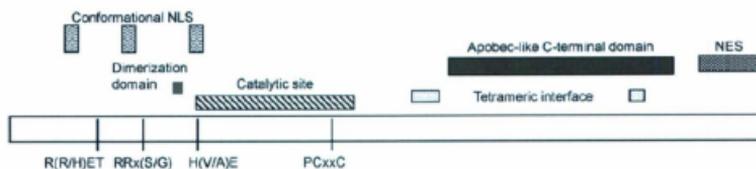
far, only the structures of APOBEC2 and the catalytic domain of A3G have been resolved (Chen et al., 2008; Prochnow et al., 2007).

### **1.7 Structural features and enzymatic properties of human AID**

The structure of AID has not yet been determined largely due to difficulties with its purification to homogeneity. However, sequence homology between AID and other APOBEC members have allowed for identification of putative domains and important residues necessary for its function (Figure 5). Among the functional domains identified thus far are a putative nuclear localization signal (NLS) and dimerization domains found within the N terminus, a zinc-coordinated catalytic site and a terminal carboxyl region with both an APOBEC-like domain and a nuclear export signal (NES) (Barreto & Magor, 2011). It has been suggested that AID forms a quaternary structure, like other AID/APOBEC family members (Holden et al., 2008; Prochnow et al., 2007). This notion was further supported by the discovery of a dimerization domain between residues 46-53 and the findings that when mutations were made within this region a decrease in cytidine deaminase activity was observed (Prochnow et al., 2007). The putative catalytic domain is located in the middle of the enzyme ranging from positions 56-90 and is characterized by the H[AV]E-X<sub>(24-36)</sub>-PCxxC motif (x: any amino acid). Based on homology with other deaminases it has been suggested that histidine (H) and two cysteines (Cys) together with a zinc cation form the catalytic core of the enzyme (Conticello, 2008). The zinc atom together with a nearby glutamate, acting as a proton donor, coordinates the deamination of the bound C to U by releasing a water molecule and the ammonium group from the fourth carbon atom of the C (Conticello, 2008). Our structural modeling of AID based on

the solved structure of the catalytic domain of A3G as well as our mutagenesis studies (Dancyger et al., 2011) support this model of deamination catalysis (Chen et al., 2008; Holden et al., 2008). Within the carboxyl terminus of AID, the APOBEC-like domain can be found between residues 119-183, this is another noted similarity to APOBEC1 due to its homology and alignment; however it remains largely uncharacterized (Barreto & Magor, 2011). A NES at position 183-198 has also been identified. NES together with NLS and the many hydrophobic residues within the carboxyl terminus may help to explain why AID shuttles between the nucleus and the cytoplasm. These hydrophobic residues have been proposed to exclude AID from the nucleus and help to explain AID's localization in the cytoplasm (Barreto & Magor, 2011; Rada et al., 2004).

Our laboratory has previously described the enzymatic characteristics of human AID and observed several unique features such as slow catalytic rates and high affinity binding to ssDNA (Larijani et al., 2007). We suggested that these attributes fit well with the *in vivo* functions of AID in that they limit its genome-wide mutating activities. In addition, these observations may explain the enzymatic processivity of AID *in vitro* (Coker & Petersen, 2007; Pham, Chelico, & Goodman, 2007). We proposed that the enzymatic properties of AID are important determinants of its biological functions. Support for this notion has been provided by recent studies of human AID mutants that showed altered catalytic rates and/or WRC specificity *in vitro*, and a corresponding change in the pattern of activity *in vivo* (Kohli et al., 2010; Wang et al., 2009; Wang, Rada, & Neuberger, 2010).



(modified from Barreto & Magor, 2011)

**Figure 5. A schematic representation of AID.**

**Figure 5. A schematic representation of AID.**

A to date schematic representation of the AID motifs (boxes above) and residues (listed below) currently identified. In the N terminus, a putative nuclear localization signal (NLS) and a dimerization domain, characterized by the H[AV]E-X<sub>[24-36]</sub>-PCxxC motif (x: any amino acid) have been characterized. A zinc-coordinated catalytic site was discovered flanked by the N and carboxyl terminal regions. The carboxyl terminus was found to have an APOBEC-like domain and a nuclear export signal (NES).

### **1.8 Evolution of AID amongst species**

The Gnathostomata (jawed vertebrates) superclass of vertebrates are the first species to show primary and secondary antibody diversification through V(D)J recombination and SHM. Since then, these processes have been shown in a variety of species including fish, frogs, lizards, birds and mammals (Barreto & Magor, 2011; Flajnik & Kasahara, 2010; Pancer & Cooper, 2006). Although Chondroichthyes (cartilaginous fish), such as sharks, are the earliest class of jawed vertebrates known to carry out SHM and thus possess AID, Osteichthyes (bony fish) are to date the earliest species identified to express full-length AID (Flajnik & Kasahara, 2010). Of the teleost fishes examined, AID from each of the pufferfish fugu (*Takifugu rubripares*), zebrafish (*Danio rerio*) and catfish (*Ictalurus punctatus*) have been examined and studied for functionality by transfection into AID<sup>-/-</sup> primary B cells and/or non-lymphocyte cell lines (Conticello et al., 2005; Conticello, 2008; Saunders & Major, 2004). Previous studies found that these bony fish are able to mount an effective adaptive immune response but are unable to undergo CSR, a mechanism that diverged after bony fish (Barreto et al., 2005). Bony fish can perform affinity maturation but it appears to be carried out in cell clusters that are distinct from mammalian germinal centers (Saunders et al., 2010). Furthermore, while fish can change isotypes, this does not appear to be mediated by AID and is instead initiated by an alternative form of V(D)J recombination (Danilova et al., 2005; Hansen, Landis, & Phillips, 2005). Interestingly, fish AID can rescue CSR in AID<sup>-/-</sup> B cells despite a lack of CSR in these organisms, suggesting that AID gained this function prior to the evolution of mammals (Barreto & Magor, 2011). The region of AID that is least conserved among these lower vertebrates is the catalytic domain with each teleost having

eight or nine extra residues within the zinc-coordinating motif as compared to higher order species (Barreto & Magor, 2011; Conticello et al., 2005).

AID has also been identified in amphibians like the frog (*Xenopus tropicalis*), Iberian ribbed newt (*Pleurodeles waltl*) and the clawed toad (*Xenopus laevis*), all of which are able to produce a diverse antibody repertoire (Bascove & Frippiat, 2010; Zhao et al., 2005). The type and extent of SHM and CSR in amphibians is currently not well characterized although it does appear that they are the first species to carry out *bona fida* CSR (Bascove & Frippiat, 2010; Musmann et al., 1997; Stavnezer & Amemiya, 2004). Studies have found that these amphibians possess DGYW tetramer SHM motifs within Ig S regions, such as AGCT (D= A/G/T, Y=C/T and W=A/T) demonstrating a preferred DNA target site for AID binding. SHM induced affinity maturation is rather poor in these species however, antibody gene rearrangement mimics that of mammals, with AID sharing a similar gene and protein structure to higher order vertebrates (Ichikawa et al., 2006; Musmann et al., 1997; Ohtani, Miyadai, & Hiroishi, 2006). While *X. laevis* is able to undergo CSR, this process is impaired in *P. waltl* (Ohtani, Miyadai, & Hiroishi, 2006), this significant difference in activity between the two species is worthy of further investigation.

Mammalian and bird species have the closest adaptive immune systems to humans. Species whose AID sequence has been identified include mouse, rat, chicken, duck, rabbit, dog, cat, cow, chimpanzee and human (Barreto & Magor, 2011; Bascove & Frippiat, 2010; Zhao et al., 2005). Amongst land dwelling tetrapods, the AID sequence appears to be more highly conserved. This is highlighted by the conservation of all 15 residues whose mutations in humans can lead to Hyper-IgM syndrome among chicken

and mammalian species (Zhao et al., 2005). This suggests an important structure-function relationship with respect to the primary sequence of AID.

### **1.9 Purpose of this study: discovering the structure function relationships within AID**

#### **Rationale:**

Since its discovery in 1999, much has been revealed about AID however, major gaps regarding our understanding of the enzyme still exist. Specifically, details concerning AID's mechanism of action, the importance of its properties, the specific domains/residues responsible for initiating these effects as well as the enzyme's evolutionary context remain unknown. As mentioned above, previous work has identified the slow catalytic rate and high affinity binding of human AID to ssDNA (Larijani et al., 2007). It was suggested that these biochemical and enzymatic characteristics fit well with the *in vivo* functions of AID in that they explain its enzymatic processivity and means to limit genome-wide mutating activities. In part because the structure of AID has not yet been solved, two underlying questions will be examined: first, what are the roles of individual domains and/or residues responsible for the biochemical properties of AID? Second, what is the relationship between the different properties of AID? For instance, it is not understood how differences in ssDNA binding influence enzymatic activity.

One approach to gain insight into these issues is a comparative study of the biochemical properties of AID from different species. We reasoned that differing AIDs likely possess altered biochemical properties due to the biological and evolutionary

divergence between organisms. These may include homeothermy *versus* poikilothermy, possible variances in the cellular context of affinity maturation, and the types and extent of secondary antibody diversification processes that are present in these organisms (i.e. SHM and CSR) (Saunders et al., 2010). Our approach is based on observation that AID is highly conserved in mammals, birds and amphibians suggesting an intimate structure-function link (Barreto & Magor, 2011). On the other hand, bony fish AID are the most divergent from their human counterpart and therefore we hypothesized that these are likely to have different biochemical properties (Barreto & Magor, 2011; Conticello et al., 2005; Wakae et al., 2006). Characterizing the distinct biochemical properties of AID from divergent species and identifying the mechanisms of these differences can shed light on the molecular basis of the biochemical parameters of AID.

Several earlier studies have examined the functionality of AID from pufferfish fugu (*Takifugu rubripares*), zebrafish (*Danio rerio*) and channel catfish (*Ictalurus punctatus*). These studies showed that fish AID are capable of supporting SHM and CSR when expressed in bacteria and cell lines (Barreto et al., 2005; Barreto, Ramiro, & Nussenzweig, 2005; Ichikawa et al., 2006; Wakae et al., 2006). Others have demonstrated that bony fish AID displayed differences in the levels of mutation and/or CSR activity when compared amongst them and to human AID, suggesting that these AID enzymes may have inherent catalytic differences between species. However, these earlier studies used bacterial systems in which AID mutations create antibiotic resistance and in eukaryotic cell lines where GFP reversion is measured as a function of AID-mediated mutations. These experiments have examined the functionality of fish AID *ex vivo* yet no direct structure-function information has been gained due to several

limitations of the experimental systems. First, the presence of multiple downstream DNA repair pathways that process AID-mediated deaminations can alter the levels and spectra of final mutations; second, the effect of varying conditions (e.g. temperatures) on bacterial or eukaryotic cell growth may unequally influence the catalytic efficiencies of different AIDs; third, disparity in the DNA substrates (e.g. transgenes and their integration sites, specific bacterial genes, immunoglobulin genes) used to quantify AID activity can influence the efficiency of AID activity; and fourth, discrepancies in the expression levels and/or solubility of various AIDs in the bacteria or eukaryotic cells could result in differences in mutational activity.

To directly compare AID of bony fish and humans in a quantifying manner, we have utilized a cell-free system to characterize the biochemical properties of purified AID from bony fish in parallel with human AID in order to relate differences in their properties to differences in their primary amino acid sequences. We report that purified AID from dissimilar species have inherently diverse deamination efficiencies, that the catalytic rates of bony fish AID are cold adapted, and that these discrepancies are mediated through variable ssDNA binding that is influenced by a single residue in the carboxyl terminus.

**Objective 1:** To determine the roles of individual domains and/or residues responsible for the biochemical properties of AID.

**Objective 2:** To determine the relationship between different properties of AID, such as ssDNA binding and enzymatic activity.

## **2. Materials and Methods**

### **2.1. Plasmids and constructs use for AID expression**

EcoRI fragments encoding the ORF of Ip-AID and Dr-AID were cloned into the pGEX-5x-3 vector (GE Healthcare, Baie d'Urfe, Quebec, Canada), a Glutathione S transferase (GST) gene fusion vector used for detection and purification of GST fusion proteins in bacteria, to generate GST-AID expression constructs analogous to the Hs-AID which we have described (Larijani et al., 2007; Saunders & Major, 2004; Zhao et al., 2005). These constructs were used as templates for gene-SOEing PCR to generate hybrid AID molecules using primers that contained dual complementarity overlapping at the junctions (Figure 6; arrows). Point mutants of Ip-AID were generated using standard PCR site-directed mutagenesis.

### **2.2 Site directed mutagenesis**

Site directed mutagenesis was performed to generate H-AID G164D. The reaction was done in duplicate for accuracy. MilliQ H<sub>2</sub>O was mixed with 10µl High Fidelity 5X PCR buffer, 1.5µl 10mM dNTP mixture, 125ng of H-AID G164D primer (CTTTCAAAGCCTGGGAAGACCTGCATGAAAATTCAG), 125nl of H-AID G164D compliment primer (CTGAATTTTCATGCAGGTCTTCCCAGGCTTTGAAAG), 25ng of pGEX-5x-3 GST H AID plasmid and 1µl of High Fidelity DNA Polymerase (New England BioLabs, Pickering, Ontario, Canada) to a total volume of 50µl. After a hot start at 98°C for 1 minute, 30 cycles of a 3 step PCR (98°C for 30 seconds, 50°C for 45

seconds and 72°C for 6 minutes) was followed by 72°C for 10 minutes. The samples were held at 15°C for 1 hour and stored at 4°C overnight. The following morning 1µl of the restriction endonuclease DPN1 was added and the reaction was incubated at 37°C for 2 hours. An aliquot of 45µl of the PCR product was pipetted onto membrane filters (Millipore, Etobicoke, Ontario, Canada) suspended in MilliQ H<sub>2</sub>O and left undisturbed for 1 hour at room temperature to remove unwanted salt, transferred to clean, labelled 1.5ml tubes and spun in a SpeedVac Concentrator (Savant, Napean, Ontario, Canada) to remove excess water. A 5µl aliquot of the 5 Kb PCR product was added to 100µl Top10 DH5α cells (Invitrogen, Burlington, Ontario, Canada). The reaction was gently mixed and incubated on ice for 25 minutes. The reaction was then heat shocked at 42°C for 1 minute and 30 seconds and placed back on ice. To the reaction, 200µl of Super Optimal broth with Catabolite repression (SOC) media was added and the culture was shaken at 37°C for 1 hour. After incubation, a 250µl aliquot of the culture was plated on Lysogeny broth (LB) ampicillin (37g LB + agar, 1L MilliQ dH<sub>2</sub>O, 100µg/ml ampicillin) and incubated at 37°C overnight. The following day, five colonies were selected and grown in 5ml LB medium with 50mg/ml ampicillin culture at 37°C and left to grow overnight, approximately 18 hours. Each culture was centrifuged and the pellet collected in a 1.5ml tube. The plasmid DNA was then purified using a QIAprep Spin Miniprep Kit (Qiagen, Toronto, Ontario, Canada). A restriction digest using EcoRI and sequencing was performed to verify the products had been cloned into the vector and that the inserts were of the correct orientation (Larijani et al., 2007).

### **2.3 AID protein production and purification**

Prior to protein purification, 1 $\mu$ l of GST-AID DNA was added to 50 $\mu$ l *Escherichia coli* BL21 (DE3) cells. The reaction was gently mixed and incubated on ice for 25 minutes, followed by heat shock at 42°C for 1 minute and 30 seconds and then placed back on ice. To the reaction, 200 $\mu$ l of SOC media was added and the culture was shaken at 37°C for 1 hour. After incubation, 100 $\mu$ l of the culture was plated on LB ampicillin (37g LB + agar, 1L MilliQ dH<sub>2</sub>O, 100 $\mu$ g/ml ampicillin) and incubated at 37°C overnight. The following day, one colony from each plate or a sterile loop of a glycerol stock, approximately 5 $\mu$ l, was selected and grown in 250ml LB medium with 50mg/ml ampicillin culture, at 37°C and left to grow to log phase in an orbital shaker (Forma Scientific, Marietta, Ohio, USA). To induce protein production 1mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) and 50mg/mL of ampicillin were added. The culture was incubated at 16°C overnight, approximately 16 hours, in a PsycroTherm Controlled Environment Incubator Shaker. The culture was poured into a 250ml Nalgene polypropylene bottle and centrifuged at 3836g for 20 minutes at 4°C in a Sorvall Evolution RC Ultra Centrifuge. The supernatant was drained and the pellet was re-suspended in 20mL of ice cold 1X PBS. The re-suspended pellet was placed in a 50mL falcon tube.

Cells were lysed in a French pressure cell press (Thermospectronic, Napean, Ontario, Canada) twice, washing the cylinder with ice cold 1X PBS in between each cycle. The sample was transferred to a Nalgene polypropylene falcon tube and centrifuged at 3836g for 10 minutes at 4°C in a Sorvall Evolution RC Ultra Centrifuge.

The supernatant was purified on Glutathione sepharose beads (Amersham, Baie d'Urfe, Quebec, Canada) and the protein was eluted as per the manufacturer's recommendations. Multiple fractions, approximately 15 – 30, of 500µl of eluted GST fusion protein were aliquoted into clean, labeled 1.5ml tubes kept on ice. The optical density (OD) of each collected fraction was measured using a nanodrop (Thermo Scientific, Napean, Ontario, Canada) at 260nm and 280nm to determine concentration and purity. Aliquots with a protein concentration greater than 1mg/ml were pooled and dialyzed against 1L of buffer (20mM Tris-HCl, pH 7.5, 100mM NaCl, and 1 mM DTT) using SnakeSkin® Pleated Dialysis Tubing (Thermo Scientific, Napean, Ontario, Canada) overnight, approximately 16 hours. The following morning, new dialysis buffer was added and the protein was left to dialyze for an additional 3 to 5 hours. Aliquots of 20 – 50µl of purified protein were dispensed into clean, labelled 1.5ml tubes kept on ice. The protein was dipped in liquid nitrogen for 30 seconds and stored at -80°C. Each batch of GST fusion protein was purified in parallel and prepared 3-4 times independently for accuracy. Protein expression levels, purity and concentration were quantified and equalized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) using bovine serum albumin (BSA) standard and Coomassie blue staining. AID proteins were assayed for activity and complex formation using deamination assays and electrophoretic mobility shift assays (EMSA) to ensure acceptable integrity of the purified proteins (Larijani et al., 2007).

## **2.4 Substrate preparation**

Three bubble substrates containing a single-stranded target region 7 nucleotides long were generated. All the oligonucleotides used in cleavage and binding assays were prepared by Integrated DNA Technologies (Toronto, Ontario, Canada) and were 56 base pairs long. The top strand (BHagtop) was desalted while the bottom strands (TGClub7, TACclub7, AGClub7) were HPLC purified. The three bottom strands TGClub7, TACclub7, AGClub7 only differed from each other by the two nucleotides before the target cytidine. The sequences of these strands are shown below with the target cytidine highlighted with an asterisk (\*):

BHagtop:

5'-AGA TCC TGC CCC GGC ACT TCG CCC GGG TTT TTC CAG TCC CTT CCC  
GCT TCA GTG AC-3'

AGClub7:

5'-GTC ACT GAA GCG GGA AGG GAC TGT GAG C\*TT CCG GGC GAA GTG  
CCG GGG CAG GAT CT-3'

TACclub7:

5'-GTC ACT GAA GCG GGA AGG GAC TGT GTA C\*TT CCG GGC GAA GTG  
CCG GGG CAG GAT CT-3'

TGClub7:

5'-GTC ACT GAA GCG GGA AGG GAC TGT GTG C\*TT CCG GGC GAA GTG  
CCG GGG CAG GAT CT-3'

To label the oligonucleotides 4 $\mu$ L of autoclaved MilliQ dH<sub>2</sub>O, 1 $\mu$ L 10X PKN buffer, 2.5pmol/ $\mu$ L of the bottom strand, 15 $\mu$ Ci/ $\mu$ L ATP-[ $\gamma$ -<sup>32</sup>P] and 1 $\mu$ L of PNK enzyme were mixed and incubated at 37°C for 53 minutes followed by 65°C for 10 minutes to

deactivate the PNK enzyme then returned to 4°C. The reaction was quickly centrifuged and 10µl of TE buffer was added. The 20µl reaction is added to a mini Quick Spin Column (Roche, Laval, Quebec, Canada) and spun at 393.12g for 3 minutes to remove any unlabelled oligonucleotides. The final volume was adjusted to 50µl by adding 5µl 1M KCl, 3µl 2.5pmol/µl top strand and autoclaved MilliQ H<sub>2</sub>O. The reaction was mixed and incubated at 96°C for 2 minutes. The incubation temperature decreased by 1°C every minute until 82°C was reached, then decreased by 1°C every 30 seconds to 35°C and finally decreased by 2°C every 30 seconds to 6°C. The labeled reaction tube was stored at -20°C. A typical substrate with the WRC motif TGC in a 7 nt bubble (TGCub7) is shown in Figure 8 (Larijani et al., 2007).

### **2.5 Deamination assay**

The deamination assay was used to quantify the enzymatic activity of pure AID. Labelled substrate, ranging in concentration of 0.15 to 50fmol, was incubated with 0.1 to 0.25µg of GST-AID in 100mM Phosphate buffer pH 7.28 to a total volume of 10µl, for time periods ranging from 30 minutes to 3 hours at 15°C, 18°C, 25°C or 37°C. For time sensitive or activity sensitive reactions, AID was deactivated at 85°C for 20 minutes and 37°C for 10 minutes. The volume of the reaction was increased to 20µl by adding 7.8µl autoclaved MilliQ H<sub>2</sub>O, 2µl 10X UDG buffer and 0.2µl UDG enzyme (New England BioLabs, Pickering, Ontario, Canada). A 30 minute incubation period at 37°C followed to excise the uracil generated by the deamination of the target cytidine. Finally, 200mM

NaOH was added and the sample was heated to 95°C for 10 minutes to cleave the alkali-labile abasic site in the DNA backbone. Formamide loading dye (10µl of 95% formamide, 0.25% Bromophenol Blue) was added to the samples which were heated again to 95°C for 10 minutes. Samples were electrophoresed at room temperature on a 14% denaturing acrylamide gel (1X TBE, 25% Formamide, 14% acrylamide:bisacrylamide [19:1], 7M Urea) with a running buffer of 1X Tris-Borate-EDTA (TBE) at 300 – 350V for 2 – 3 hours. The gels were exposed to a Kodak Storage Phosphor Screen GP (BioRad, Mississauga, Ontario, Canada) overnight, for approximately 18 hours, and visualized using a PhosphorImager (Molecular Dynamics, Sunnyvale, California, USA) (Larijani et al., 2007).

## **2.6 Electrophoretic mobility shift assay**

The electrophoretic mobility shift assay (EMSA) was used to quantify the enzymatic binding affinity of pure AID. Labelled substrate, ranging in concentration from 0.15 to 50fmol, was incubated with 0.1 to 0.25µg of GST-AID in EMSA binding buffer (50mM Tris, pH 7.5, 2µM MgCl, 50mM NaCl, and 1mM DTT) in a final volume of 10µl for time periods ranging from 2 – 3 hours at 18°C, 25°C or 35°C. Samples were optimally UV crosslinked (Stratagene, Santa Clara, California, USA) on ice, twice. Samples were electrophoresed at 4°C on an 8% native gel (6% glycerol, 8% acrylamide:bisacrylamide [19:1], 0.5X TBE) with a running buffer of 0.5X Tris-Borate-EDTA (TBE) at 300 – 350V for 2 – 3 hours. The gels were dried on chromatography paper (Whatman, Kent, United

Kingdom), exposed to a Kodak Storage Phosphor Screen GP (BioRad, Mississauga, Ontario, Canada) overnight, for approximately 18 hours, and visualized using a PhosphorImager (Molecular Dynamics, Sunnyvale, California, USA) (Larijani et al., 2007).

### **2.7 Data collection and quantification of data**

All quantifications were done using Quantity One 1-D Analysis Software (BioRad, Mississauga, Ontario, Canada). Data from deamination assay gels were plotted as substrate and product fractions. Data from EMSA gels were plotted as bound and free fractions of the substrate. Deamination assay gels and EMSA gels were quantified 2-3 times independently by two people and the results were averaged. Sister repeated gels were averaged thus obtaining an average value for the lane. This single data point represents the mean of the values obtained in this manner from all experiments. Graph averages represent 4-6 independent experiments. Error bars represent the standard deviation between independent experiments and are not a range of values in one experiment but the range across all parallel experiments. The data were graphed using GraphPad Prism Software (La Jolla, California, USA).

Densitometry was done using Quantity One 1-D Analysis Software (BioRad, Mississauga, Ontario, Canada). Deamination or EMSA gels contained duplicate independent reactions for each AID: substrate combination. Experiments were repeated 5-12 times to ensure that all independent AID purifications where all species variants,

hybrid enzymes and mutants were purified in parallel were represented. For each experiment, an individual lane of an alkaline cleavage or EMSA gel was quantified 3 times, thus obtaining an average value for the lane, representing a single data point (typically the triplicate measurements of each lane as well as the duplicate reaction lanes within each experiment contained a variability of < 5%). The data points on the graphs are the mean of the values obtained in this manner from all experiments (5-12 experiments yielding 10-24 values for each data point). Error bars represent standard deviation between independent experiments. The data were graphed using GraphPad Prism Software (La Jolla, California, USA).

## **2.8 Statistical analysis**

Statistical significance of differences between the activity levels of two types of AID was ascertained by performing analysis of variance and two-tailed student tests. Data from all experiments were compared (typically n=10-24). Analysis was performed using GraphPad Prism Software (La Jolla, California, USA).

### 3. Results

#### 3.1. The catfish AID sequence is most divergent from other AID species

Cartilaginous and bony fish are the earliest species to express AID (Barreto & Magor, 2011; Flajnik & Kasahara, 2010). While AID is necessary for both CSR and SHM, the earliest evidence of CSR only appears after the divergence of bony fish, from amphibians onward. Bony fish are able to mount an effective adaptive immune response but unable to undergo CSR (Barreto et al., 2005; Barreto & Magor, 2011; Conticello et al., 2005). In order to identify the species which possesses the most divergent form of AID from humans, we aligned the published sequences of representative organisms from various classes throughout evolution including: Osteichthyes (bony fish), Amphibia, Aves, and Mammalian. Figure 6 displays the sequence alignment of bony fish fugu (*Takifugu rubripes*, Tr-AID), channel catfish (*Ictalurus punctatus*, Ip-AID) and zebrafish (*Danio rerio*, Dr-AID), with xenopus (*Xenopus laevis*, Xl-AID), chicken (*Gallus gallus*, Gg-AID), human (*Homo sapiens*, Hs-AID) and mouse (*Mus musculus*, Mm-AID). Residues underlined in blue indicate those that are identical between species, residues underlined in green identify amino acids within a similar group while those underlined in red indicate unlike residues. Comparing the various forms of AID from bony fish to amphibians to birds to mammals our results show, in agreement with previous work, that AID is a relatively conserved enzyme over evolution possessing a high proportion of identical and similar residues (Barreto & Magor, 2011). The three domains of the enzyme, the N-, catalytic, and carboxyl terminals, are shown using arrows (Figure 6). Among chicken (Gg), mouse (Mm) and human (Hs) AIDs, there is a remarkably high level of conservation with chicken (Gg) and mouse (Mm) having seven and eight non-

similar residues to Hs-AID, respectively. Less residue preservation is seen between xenopus (XI) and bony fish (Tr, Ip and Dr). The AID sequence in bony fish does include extra residues in all 3 termini, with the most being present in the catalytic domain, indicating that over time the enzyme has deviated from its ancestral form.

To gain additional perspective, a residue identity and similarity analysis was conducted where the percentage of residues that were identical and similar in various species were compared to its human counterpart. Figure 7 shows the catalytic domain of Seryl tRNA synthetase, a housekeeping enzyme, weighed against RAG-1/2, an enzyme involved in the generation of primary antibody diversity, and AID (left, middle, and right panels respectively) where each point in the graph represents the emergence millions of years ago (mya) of the species mentioned above. In agreement with its function *in vivo*, Seryl tRNA synthetase shows little divergence over evolution. RAG-1/2 and AID, which are enzymes involved in diversity generation in the immune system are thus under heightened evolutionary pressure, show much greater residue variance over time. Amongst the land species human (Hs), mouse (Mm) and chicken (Gg) AID were found to be more highly conserved compared to RAG-1/2 (96% vs. 74% similarity). In contrast, a notable level of deviation, as seen by the drastic drop in amino acid similarity and identity, was observed between zebrafish (Dr) and pufferfish (Tr) as compared to channel catfish (Ip) AID with the latter being considerably less conserved with Hs-AID (126 and 128 *versus* 126 identical residues). Unlike AID, neither Seryl tRNA synthetase nor RAG-1/2 showed the same level of divergence between the three bony fish. From these results, it was concluded that amongst bony fish AID, Ip-AID was the most divergent form from

its human counterpart. Therefore, we hypothesized that it may possess altered enzymatic and biochemical characteristics.



**Figure 6. Sequence alignment of AID from representative species across classes.**

The comparison of AID sequences from representative species across classes. Residues are colored according to side-chain. Blue, green and red underlines indicate identical, similar and unlike residues respectively. Arrows denote the borders of the N-, catalytic and carboxyl domains. The sequence alignment is based on the published sequences of AID (*Mus musculus*: NP033775, *Homo sapien*: AAM95406.1, *Gallus gallus*: XP416483, *Xenopus laevis*: NP001089181, *Takifugu rebripes*: AAU95747, *Danio rerio*: NP001008403, *Ictalurus punctatus*: AAR7544).

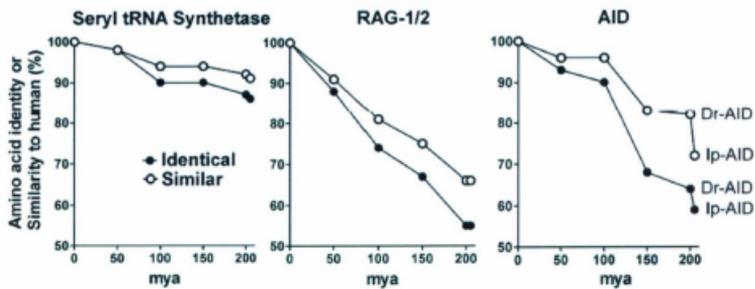


Figure 7. Percent identity and similarity between 3 human enzymes.

**Figure 7. Percent identity and similarity between 3 human enzymes.**

The percent of amino acid identity or similarity between 3 human enzymes and their bony fish counterparts. Based on the sequence alignment of the published sequences for Seryl tRNA synthetase (*Homo sapiens*: BAA95602.1, *Mus musculus*: NP034559.1, *Gallus gallus*: NP001026563.1, *Xenopus laevis*: NP001080487.1, *Danio rerio*: BA144434.1, *Ictalurus punctatus*: AAY86956.1), RAG-1/2 (*Homo sapiens*: AAA60248.1, *Mus musculus*: NP033045.2, *Gallus gallus*: AAA49051.5, *Xenopus laevis*: AAX68652, *Danio rerio*: O13033.1, *Ictalurus punctatus*: ABG47664.1) and AID (as shown previously in Figure 4). The left panel shows the sequence comparison of the essential house-keeping enzyme, Seryl tRNA synthetase catalytic domain. The middle panel shows the sequence comparison of RAG-1/2, an enzyme involved in antibody diversity. The right panel shows the sequence comparison of AID. Approximate periods of species emergence (millions of years ago) are shown on the x-axis.

### **3.2 Differences in the activity levels of Hs-AID, Dr-AID and Ip-AID**

To measure the activity of bony fish and human AID, we used a previously established *in vitro* experimental system where Hs-AID, Ip-AID, and Dr-AID were independently cloned into an expression vector, expressed in E.coli and column purified to produce GST-tagged AID (Larijani et al., 2007). All three AIDs were expressed and purified in parallel multiple independent times with similar AID yields. The amount of AID used was equalized by semi-quantitative SDS PAGE. We used the previously established alkaline cleavage assay to measure the enzymatic activity of each enzyme. In this assay, AID is incubated with the radioactively labeled bubble substrate (Figure 8). We used a 7 nt bubble substrate containing various WRC motifs (AGC, TAC or TGC) which were previously shown to be an optimal target for deamination by AID *in vitro* (Larijani & Martin, 2007). AID converts the target C to U, the U is then removed by the addition of UNG generating an abasic site which is alkali-labile (Larijani & Martin, 2007). Alkaline-cleavage at this site produces a cleaved product, the amount of which we can use to accurately measure and quantify AID activity. We used partially single-stranded substrates with a bubble of 7 nucleotides in length, which we previously found to be an optimal substrate of AID (Larijani & Martin, 2007).

Figure 9A shows the deamination activity of equal amounts of Hs-AID, Dr-AID and Ip-AID incubated with the TGClub7 WRC substrate for one hour. We observed that Hs-AID and Dr-AID mediated comparable levels of deamination (34 vs 28%). Conversely, Ip-AID appeared to have little to no activity, similar to levels of the negative control which lacked AID. This observation was repeated using three independent preparations of Ip-AID starting with independently cloned expression vectors. To

decipher if Ip-AID was a catalytically-inactive enzyme or whether it was significantly less active than AIDs from the other species tested, the incubation period of Ip-AID with the labeled bubble substrate as well as exposure time of the gel was increased. Figure 9B shows the overnight incubation and overexposed assay. The gel revealed that indeed Ip-AID is not catalytically inactive, but rather has approximately one order of magnitude less activity as compared with Hs-AID and Dr-AID. Some catalytic activity is present in Ip-AID, though very little (2.6%).

To compare the deamination kinetics of each AID enzyme, we measured the production of deaminated product on the TGCbub7 substrate by Hs-AID, Dr-AID and Ip-AID as a function of time, ranging from 5 minutes to 420 minutes. The reaction was terminated by initiating alkaline cleavage. As shown in Figure 10, with a representative quantifying gel in the inset and consistent with our previous findings, Hs-AID and Dr-AID exhibited robust enzymatic activity detectable at 5 minutes which increased linearly for 1 hour (62 vs 56 fmol product, respectively). In contrast, Ip-AID produced very little product (6 fmol) even after a prolonged period of incubation of 420 minutes, carrying out the same level of deamination as seen by Hs-AID and Dr-AID in 5 minutes. Considering the amount of deamination in a given period of time and the length of time taken to generate a set amount of deaminated product, we conclude that Ip-AID is approximately 10 fold less active and therefore less efficient than both Hs-AID and Dr-AID.

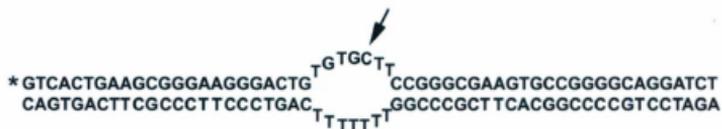


Figure 8. Typical substrate used in the *in vitro* alkaline cleavage activity assay.

**Figure 8. Typical substrate used in the *in vitro* alkaline cleavage activity assay.**

The target cytidine is indicated by the arrow and the radioactively labeled strand is denoted by “\*”. Dinucleotide upstream of the target cytidine can be varied among different substrates to generate a WRC or non-WRC motif.

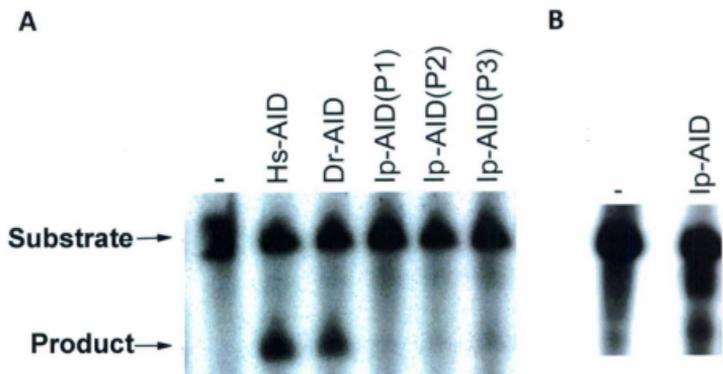


Figure 9. *In vitro* activity of human and bony fish AID.

**Figure 9. *In vitro* activity of human and bony fish AID.**

**A.** Oligonucleotide alkaline-cleavage assay showing the activity of human (Hs), zebrafish (Dr) and catfish (Ip) GST-AID incubated with TGCbub7 substrate for 1 hour. Three independently purified preparations of catfish GST-AID were tested, Ip-AID (P1), Ip-AID (P2), Ip-AID (P3).

**B.** Activity of catfish GST-AID (Ip-AID) in the alkaline cleavage assay upon prolonged overnight incubation and over-exposure of the assay gel.

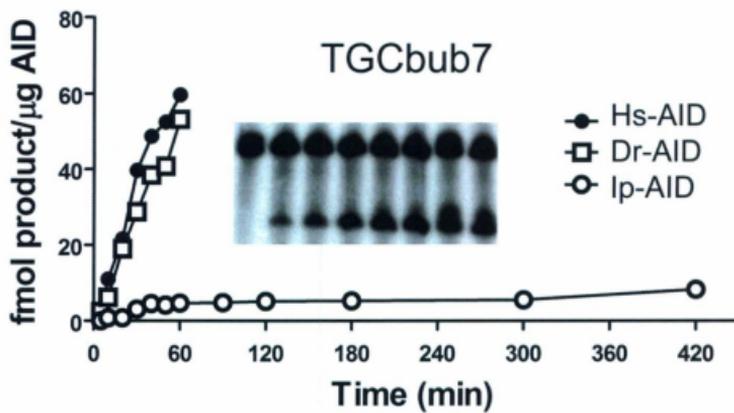


Figure 10. Comparison of activity kinetics between Hs-AID, Dr-AID and Ip-AID.

**Figure 10. Comparison of activity kinetics between Hs-AID, Dr-AID and Ip-AID.**

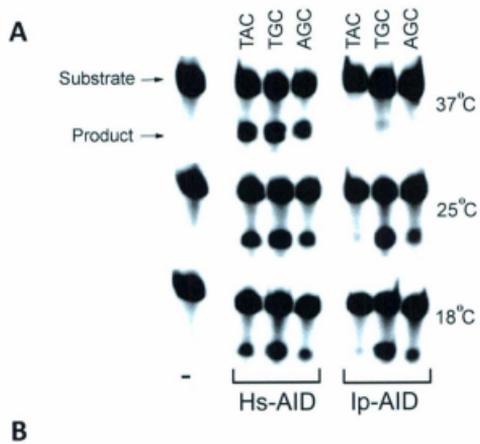
The TGCbub7 substrate, 10 fmol, was incubated with human (Hs), zebrafish (Dr) and catfish (Ip) GST-AID for various times ranging from 5 to 420 minutes. The reactions were subsequently terminated by initiating alkaline cleavage. Inset shows a typical gel used to measure product generated at each time point and to calculate deamination velocity.

### **3.3 The activity of Dr-AID and Ip-AID are temperature-dependent**

The assays described above were carried out at 37°C, the optimal temperature for Hs-AID; however, the physiological temperatures for catfish and zebrafish are approximately 14°C to 25°C. To test whether the catalytic activity of purified Ip-AID and Dr-AID are adapted to lower temperatures, we carried out deamination assays on three WRC substrates (TGCbub7, TACbub7 and AGCbub7) at 18°C, 25°C, and 37°C. Whereas Hs-AID showed generally decreasing activity with decreasing temperatures, the enzymatic activities of both Ip-AID and Dr-AID increased appreciably as the temperature was decreased (Figure 11A and B). Regardless of temperature or substrate, Ip-AID activity levels were always lower than Hs-AID, only approaching it at 18°C (20% vs. 26% for TGC and 5.7% vs. 7.4% for AGC). In contrast, Dr-AID was up to 4.8 fold more active than the highest level reached by Hs-AID on the same substrate (80% vs. 38% on TGC, 33% vs. 24% on TAC and 63% vs. 13% on AGC).

Given that Dr-AID, Ip-AID and Hs-AID showed optimal activity under different conditions, we tested the maximal enzymatic potential of each under its own optimal conditions (Figure 12). Dr-AID at 25°C exhibited the greatest velocity of product formation reaching nearly twice that of Hs-AID at the point of maximal difference. In agreement with our previous results, Ip-AID was almost inactive at 37°C but gained activity at lower temperatures of 18°C and 25°C with a velocity approaching at best half that of Hs-AID. To examine if temperatures at lower than 18°C would further increase Ip-AID activity the deamination kinetics of Ip-AID were tested at 15°C; however, no difference in activity was found in comparison to that of our results at 18°C. In

conclusion, we found that at the points of maximal difference Dr-AID was approximately 2 fold more active than Hs-AID while Ip-AID (at 25°C or 18°C) was approximately 2 fold less active than Hs-AID.



**B**

Figure 11. Temperature and substrate preference of bony fish AID.

**Figure 11. Temperature and substrate preference of bony fish AID.**

**A.** The activity of Hs-AID and Ip-AID were compared on the three WRC substrates (TGCbub7, TACbub7 and AGCbub7) at three difference temperatures, 37°C (top panel), 25°C (middle panel), and 18°C (bottom panel). Hs-AID reactions are shown on the left of each gel, while Ip-AID reactions are shown on the right. A negative control reaction is also shown on each gel. Twenty fmol of substrate was incubated with GST-AID for 1 hour.

**B.** The percent deamination products were quantified and plotted for each substrate and temperature combination for Hs-AID, Ip-AID and Dr-AID. For TGCbub7, comparing Dr-AID to Hs-AID at 18°C:  $p=0.004$ ; at 25°C  $p=0.01$ ; at 37°C  $p=0.001$  and comparing Hs-AID to Ip-AID at 25°C  $p=0.04$ . For AGCbub7, comparing Dr-AID to Hs-AID at 18°C:  $p=0.004$ ; at 25°C  $p=0.009$ ; at 37°C  $p=0.02$ . For TACbub7, comparing Dr-AID to Hs-AID at 18°C:  $p=0.008$ ; at 25°C  $p=0.02$ ; at 37°C  $p=0.006$ .

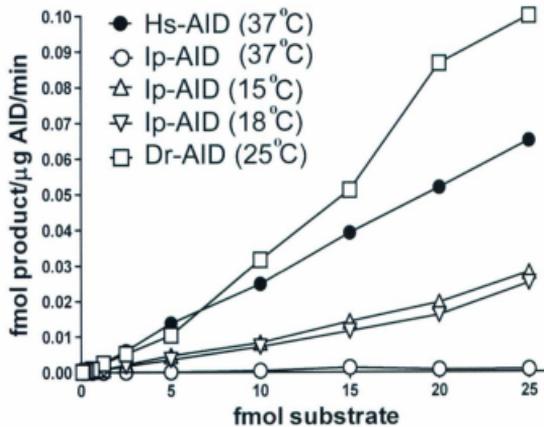


Figure 12. Comparison of deamination kinetics between Hs-AID, Dr-AID and Ip-AID under optimal conditions.

**Figure 12. Comparison of deamination kinetics between Hs-AID, Dr-AID and Ip-AID under optimal conditions.**

The deamination kinetics of Hs-AID, Dr-AID and Ip-AID were compared under their optimal substrate (TGCBub7) and temperatures conditions. Reaction velocities are shown as a function of substrate concentration over a 100 fold dilution range (0.025 - 2.5 nM). Comparing Dr-AID (25°C) and Hs-AID (37°C)  $p=0.03$ , and comparing Hs-AID (37°C) and Ip-AID (18°C)  $p=0.01$ .

### **3.4 The basis of the low activity of Ip-AID at 37°C**

Alignment of the AID proteins of our test species revealed considerable divergence between the fishes and mammals, particularly when compared with the divergence of another protein involved in adaptive immunity (RAG-1) or a housekeeping protein (seryl-tRNA-synthetase; Figure 7). To determine if these differences might result in gross structural differences in the enzymes, we modeled all three based on the determined crystal structure of the catalytic domain of family member APOBEC3G (Chen et al., 2008; Holden et al., 2008; Shandilya et al., 2010). The progression of domains from the N- to carboxyl-terminus is shown from blue to red in Figure 13. Residues involved in the Zn-coordinating catalytic domain are shown as red sticks. For Hs-AID, these include H56, E58, C87, C90. For Dr-AID, these include: H60, E62, C99, C102. For Ip-AID these include: H59, E61, C98, C101. All three were predicted to form the canonical APOBEC family arrangement of alpha helices surrounding a beta sheet, with the catalytic (Zn-coordinating) residues forming a catalytic pocket that was indistinguishable (Figure 13) (Conticello et al., 2007). As we did not observe major structural differences, we conclude that the disparity in activity is likely due to subtle distinctions in amino acid residues.

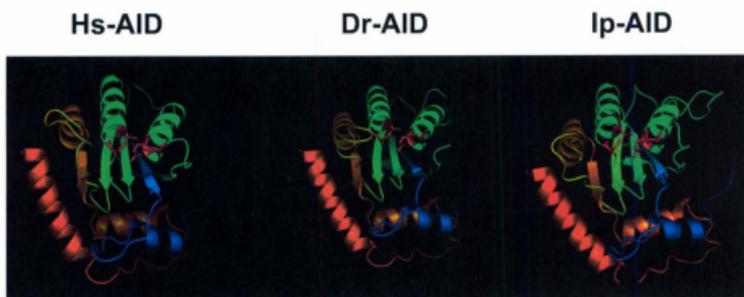
Based on the placement of the Zn-coordinating catalytic residues and other motifs identified in AID by homology to other deaminases, AID can be divided into three segments: The N-, catalytic and carboxyl- terminals (see arrows in Figure 6). To determine which of the three domain(s) are responsible for the change of the Ip-AID lethargy at 37°C, gene-SOEing was used to construct various hybrid GST-AID enzymes

where segments from each species were swapped. Linear representations of the hybrids generated are displayed in Figure 14A. Figure 14B shows the test of the activities of the various domain-swapped hybrid AID enzymes with TGCbub7 at 37°C. As previously revealed in Figure 9, Ip-AID shows little activity while Hs-AID and Dr-AID are potent deaminators. Replacing the catalytic domain of Hs-AID with either the Dr (HDH) or Ip (HIH) did not diminish AID activity at 37°C. In opposition, replacing the catalytic domain of Ip-AID with the Hs segment (IHI) did not rescue enzyme activity. These experiments exclude the possibility that the catalytic domain of AID is responsible for its deaminating ability. Interchanging the N terminus of Ip-AID with the Hs portion (HII) also failed to recover activity, ruling out the N terminus as well as solely being accountable for the change in enzyme activity. However, substituting the carboxyl terminus of Ip-AID with the Hs (IIH) rescued AID activity at 37°C. These results identify the carboxyl terminus as responsible for the poor activity of Ip-AID at 37°C.

Next, we sought to determine which residue(s) within the carboxyl terminus of AID were specifically responsible for the rescued effect previously noted. Figure 15 shows an alignment of the carboxyl terminal domains of all three species. We identified twelve residues (denoted by an asterisk) as being divergent in Ip-AID from both Hs-AID and Dr-AID. We mutated each of these residues to its Hs-AID equivalent. Thus, twelve point mutants and a triple mutant (N167H, Q168E, A170T) were generated and tested for activity at 37°C on all three WRC substrates (TGCbub7, TACbub7 and AGCbub7) (Figure 16A). The deamination activity of each point mutation was quantified on each of the three substrates. All the mutants demonstrated a preference for deaminating TGCbub7

over AGCub7 and to a lesser extent TACub7. Interestingly only one mutant, Ip-AID D176G, rescued deamination activity to levels comparable to those seen in Hs-AID. Figure 16B shows the rescue of Ip-AID activity by the D176G mutation. We conclude that the poor enzymatic activity of Ip-AID at 37°C is due to a single non-conservative amino acid difference with Hs-AID and Dr-AID in the carboxyl-terminus.

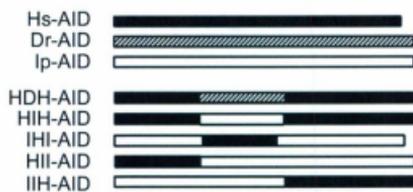
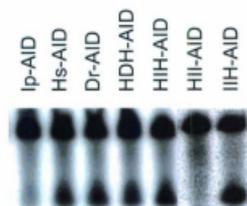
To confirm this result we introduced the reverse mutation in Hs-AID. Figure 17 shows the deamination kinetics for Hs-AID, Ip-AID, Ip-AID D176G and Hs-AID G164G with the TGCub7 substrate at 37°C. We found that the D176G mutation almost completely rescued the velocity of Ip-AID to levels of HS-AID at 37°C (0.39 vs 0.49 fmol/μg AID/min), indicative of an almost complete recovery of enzymatic activity. We confirmed that Hs-AID G164D had diminished enzymatic velocity at 37°C by 5.4 fold (0.49 vs 0.09 fmol/μg AID/min). We note that this G is conserved in AID from all species analyzed to date, with the exception of a few bony fish (including Tr-AID) that like Ip-AID have an acidic residue at this position (Figure 6, arrow) (Ichikawa et al., 2006).



**Figure 13. Predicted structures of Hs-AID, Dr-AID and Ip-AID monomers.**

**Figure 13. Predicted structures of Hs-AID, Dr-AID and Ip-AID monomers.**

The predicted ribbon structures of human (Hs), zebrafish (Dr) and catfish (Ip) AID monomers, modeled based on the published structure of the APOBEC3G catalytic region. N to carboxyl-terminus progression is shown as blue to red. The putative Zn-coordinating catalytic residues are shown as sticks in red (Hs-AID: H56, E58, C87, C90; Dr-AID: H60, E62, C99, C102; Ip-AID: H59, E61, C98, C101).

**A****B**

**Figure 14. Analysis of hybrid AIDs.**

**Figure 14. Analysis of hybrid AIDs.**

**A.** Various hybrid GST-AIDs containing the N-, catalytic and carboxyl-terminus domains from human (Hs), zebrafish (Dr) and catfish (Ip) AID were constructed and tested for activity. The three letter designation of each hybrid denotes the N-, catalytic and carboxyl-terminus domains, respectively.

**B.** Alkaline cleavage assay showing the activity of Ip-AID, Hs-AID, Dr-AID and the hybrid GST-AIDs. Twenty fmol of the TGCbub7 substrate was incubated with each GST-AID for 1 hour at 37°C.



**Figure 15. Alignment of the carboxyl terminal of Hs-AID, Dr-AID and Ip-AID.**

The twelve residues divergent in Ip-AID from both Hs-AID and Dr-AID are denoted by an “\*”. These residues were targeted for point mutation and mutated to their Hs-AID equivalent. In addition, a triple mutant (N167H, Q168E, A170T) was also generated. The arrow denotes the Ip-AID D176G mutation that explains Ip-AID lethargy at 37°C.

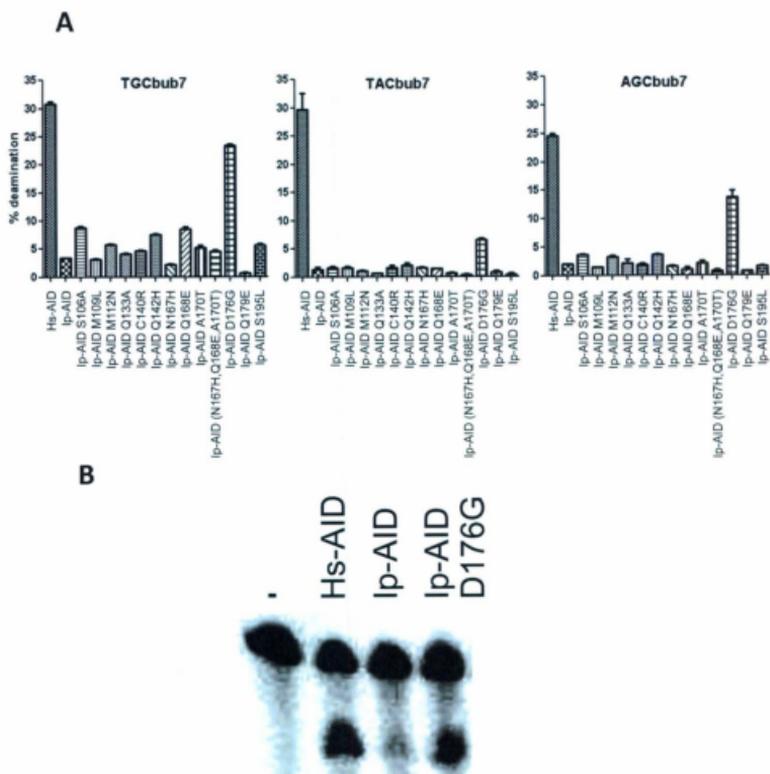


Figure 16. The activity of Ip-AID mutants.

**Figure 16. The activity of Ip-AID mutants.**

**A.** The activity of Hs-AID, Ip-AID and 13 Ip-AID mutants were tested on the three WRC substrates, TGCbub7, TACbub7 and AGCbub7 by alkaline cleavage assay. Fifty fmol of substrate was incubated with GST-AID for 1 hour and the percent deamination products are plotted.

**B.** The deamination assay showing the rescue of Ip-AID activity at 37°C by the D176G mutation. Twenty fmol of TGCbub7 substrate was incubated with GST-AID for 1 hour and the deamination assay is visualized.

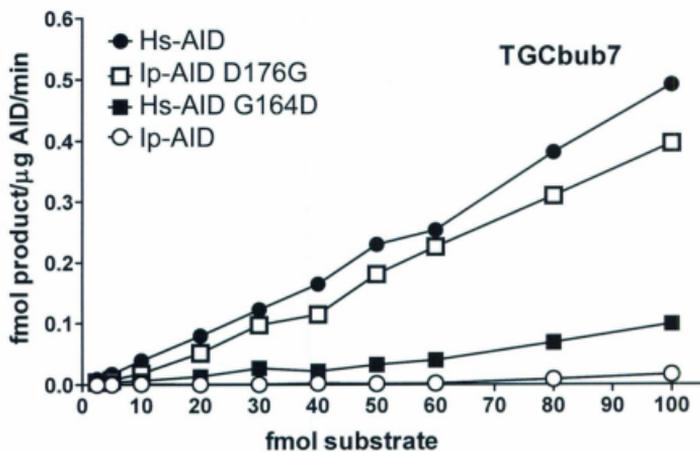


Figure 17. Ip-AID D176G is identified as a rescue mutation.

**Figure 17. Ip-AID D176G is identified as a rescue mutation.**

The deamination kinetics comparing the product formation of Hs-AID, Ip-AID D176G, Hs-AID G164D and Ip-AID using a substrate concentration ranging 100 fold (1 fmol to 100 fmol). The TGClub7 substrate was incubated with each GST-AID for 1 hour at 37°C. Comparing Ip-AID and Ip-AID (D176G)  $p=0.009$ , and comparing Hs-AID and Hs-AID (G164D)  $p=0.008$ .

### **3.5 The effects of G176D mutation on Ip-AID activity**

To examine the effects of the Ip-AID D176G and Hs-AID G164D mutations on sequence and temperature dependence we tested their activity on all three WRC substrates, TGCbub7, TACbub7 and AGCbub7 and at all three conditions, 37°C, 25°C and 18°C in comparison to Hs-AID and Ip-AID (Figure 18, Figure 11). Ip-AID D176G showed a temperature profile similar to that of Ip-AID and Dr-AID previously observed, deaminating more efficiently at 18°C and 25°C rather than 37°C. D176G was able to increase the enzymatic efficiency of Ip-AID at all temperatures, displaying approximately 10 fold more activity at 37°C and 6 fold more activity at 18°C and 25°C. The activity of Ip-AID D176G was 3.4 fold higher than Hs-AID (at 18°C: 78% vs 23% on TGCbub7 and 23% vs 7% on AGCbub7). Conversely, the Hs-AID G164D reduced the activity of Hs-AID only at 37°C. From these results we conclude that the D to G mutation has a profound effect on Ip-AID activity, increasing activity at 37°C to levels comparable to Hs-AID, and increasing activity at 18°C and 25°C to levels several fold higher than Hs-AID, reminiscent of Dr-AID. Thus, we sought to compare Ip-AID G176D and Dr-AID directly.

Figure 19A shows a comparison of the deamination activities of Dr-AID, Ip-AID D176G, Hs-AID and Ip-AID on the preferred WRC substrate TGCbub7 (right panel) with a linear representation of the enzymes (left panel). Remarkably, we found the activity levels of Dr-AID and Ip-AID D176G and their temperature dependence profiles to be identical. These results show that the D176G mutation is a potent modulator of activity and that this single residue divergence between Dr-AID and Ip-AID accounts for the difference in their catalytic potential. However, the temperature preference pattern of Ip-

AID remains unaltered by the mutation; therefore, this effect must also be dependent on other residues.

The higher activity of Ip-AID D176G as compared to Hs-AID at 18°C and 25°C suggests that residue(s) other than G164 of Hs-AID exert an activity-suppressing effect at lower temperatures. We postulated that these residues likely lie in the carboxyl terminus since we observed that the velocity of Ip-AID D176G, in which one residue was replaced with its Hs-AID counterpart, was higher relative to that of Hs-AID as compared to IIH-AID, in which the entire carboxyl-terminus was replaced with that of Hs-AID (Figure 17). To test this notion directly, we examined Ip-AID D176G, IIH-AID and Hs-AID in parallel (Figure 19B). We found IIH-AID was up to approximately 4-fold less active at 18°C and 25°C than Ip-AID D176G. Thus, residue(s) other than G164 in the carboxyl-terminal domain of Hs-AID account for its reduced enzymatic activity.

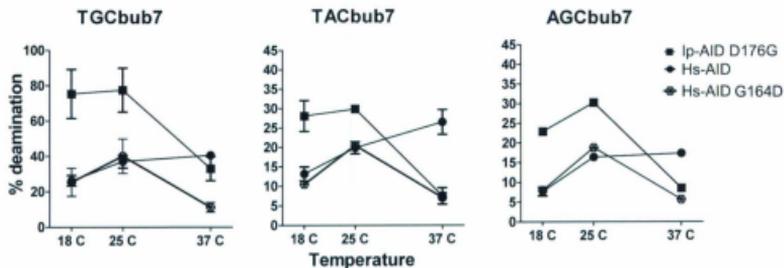


Figure 18. The effects of the D176G mutation on Ip-AID activity.

**Figure 18. The effects of the D176G mutation on Ip-AID activity.**

The deamination activities of Hs-AID, Ip-AID D176G and Hs-AID G164D were compared on all three WRC substrates (TGCbub7, TACbub7 and AGCbub7) at 37°C, 25°C and 18°C. Twenty fmol of substrate was incubated with each GST-AID for 1 hour and the percent of product formation was plotted. For TGCbub7, comparing Ip-AID D176 to Hs-AID at 18°C:  $p < 0.0001$ , at 25°C:  $p < 0.0001$ , at 37°C:  $p < 0.0001$ . For TACbub7 comparing Ip-AID D176 to Hs-AID at 18°C:  $p = 0.006$ , at 25°C:  $p = 0.01$ , at 37°C:  $p = 0.02$ . For AGCbub7 comparing Ip-AID D176 to Hs-AID at 18°C:  $p = 0.007$ , at 25°C:  $p = 0.002$ , at 37°C:  $p = 0.02$ .

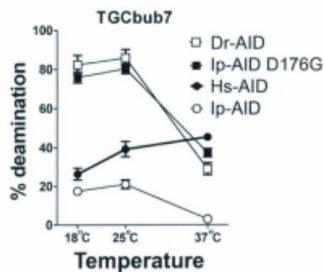
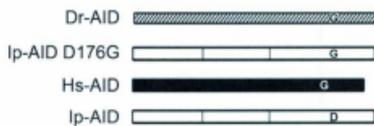
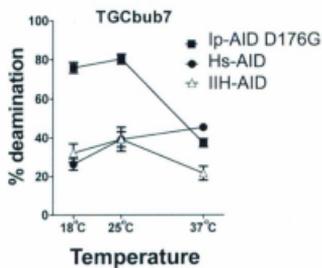
**A****B**

Figure 19. Comparison of the deamination activities of various AIDs at 37°C, 25°C and 18°C.

**Figure 19. Comparison of the deamination activities of various AIDs at 37°C, 25°C and 18°C.**

**A.** A comparison of the deamination activities of Dr-AID, Ip-AID D176G, Hs-AID and Ip-AID on the TGCbub7 substrate at 37°C, 25°C and 18°C. Twenty fmol of substrate was incubated with each GST-AID for 1 hour and the percent of product formation was plotted. The left panel shows the linear representation of these enzymes while the right panel compares the deamination efficiencies of each at three temperatures, 37°C, 25°C and 18°C. Comparing Dr-AID/Ip-AID G176D to Hs-AID at 18°C:  $p=0.003$ , at 25°C:  $p=0.01$ , at 37°C:  $p=0.001$ .

**B.** A comparison of the deamination activities of Ip-AID D176G, Hs-AID and IIH-AID on the TGCbub7 substrate at 37°C, 25°C and 18°C. The left panel shows the linear representation of these enzymes while the right panel compares the deamination efficiencies of each at three temperatures, 37°C, 25°C and 18°C. Comparing Ip-AID D176G and IIH-AID at 18°C:  $p=0.002$ , at 25°C:  $p=0.009$ , at 37°C:  $p<0.0001$ .

### **3.6 The D176 residues in Ip-AID suppresses DNA binding activity**

We next examined how the D176 residue exerts its effects on Ip-AID activity. To locate D176, we modeled the surface of Hs-AID, Dr-AID, and Ip-AID based on the resolved structure of the APOBEC3G catalytic unit (Figure 20) (Chen et al., 2008; Holden et al., 2008; Shandilya et al., 2010). All AID models show topology similarity to APOBEC3G. The D176 residue on Ip-AID (G177 on Dr-AID, and G164 on Hs-AID) was located on the surface and proximal to residues in APOBEC3G that have been shown to contact ssDNA (Figure 11 arrow, bottom right panel) (Chen et al., 2008; Holden et al., 2008). To determine if the D176 residue on Ip-AID leads to a reduction in ssDNA binding, we examined ssDNA binding properties of Ip-AID, Hs-AID and Ip-AID D176G. We used quantifying EMSA that we previously developed to determine the binding affinity of Hs-AID to ssDNA (Larijani & Martin, 2007; Larijani et al., 2007). Figure 21 shows the half-saturation kinetics of AID and ssDNA binding over a range of substrate concentrations at 18°C, 25°C, and 37°C.  $K_d$  values were derived using non-linear regression analysis (Table 1). For Hs-AID, the  $K_d$  values at the three temperatures were similar (Table 1: 0.10-0.31 nM). Although the  $K_d$  for Ip-AID was similar to Hs-AID at 18°C and 25°C (Table 1: 0.07-0.14 nM), the  $K_d$  for Ip-AID was approximately 7-fold higher than that for Hs-AID at 37°C (Table 1: 1.1 nM vs. 0.15 nM, respectively). These values correlate well with the relatively comparable levels of Ip-AID and Hs-AID activity at 18°C and 25°C and with the severely diminished activity of Ip-AID at 37°C. On the other hand, Ip-AID D176G had  $K_d$  values that were 2 to 6-fold lower than Hs-AID and Ip-AID at 18°C and 25°C (0.03-0.05 nM range), consistent with the increased activity of this mutant at these temperatures. The  $K_d$  value of Ip-AID D176G at 37°C was between

that of Hs-AID and Ip-AID (0.69 nM), which would explain the elevated activity of this mutant compared to Ip-AID, and the slightly reduced activity of this mutant compared to Hs-AID at 37°C. We conclude that the D176G mutation increases the affinity of Ip-AID for ssDNA, and that this effect is significantly more pronounced at lower temperatures.

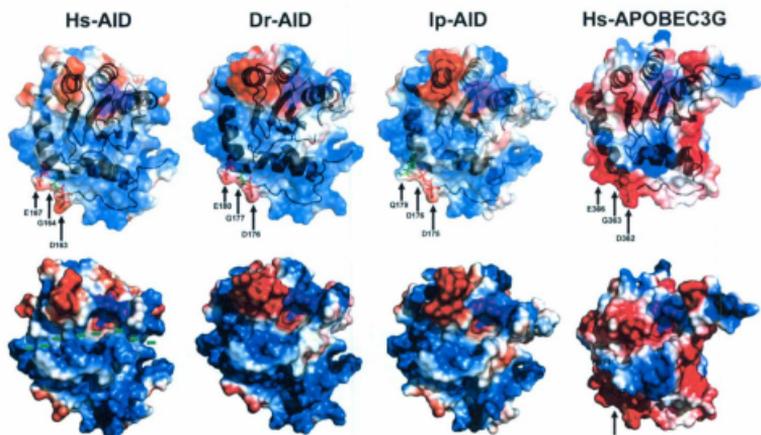


Figure 20. Surface structure models of Hs-AID, Dr-AID, Ip-AID and APOBEC3G.

**Figure 20. Surface structure models of Hs-AID, Dr-AID, Ip-AID and APOBEC3G.**

The surface topology and charge of Hs-AID, Dr-AID, and Ip-AID were modeled on the basis of the solved structure of the APOBEC3G catalytic unit. The top panel shows a transparent superimposition of the ribbon diagram within the surface. The bottom panel shows a solid representation of the surface of each model. Positively and negatively charged residues are coloured red and blue, respectively. The putative catalytic pocket housing the Zn-coordinating residues are shown in magenta. Arrows mark the position of D176 and the surrounding residues in Ip-AID as well as the corresponding residues in Hs-AID and Dr-AID. Several proximal residues that bind ssDNA in APOBEC3G are indicated (H367, H374). A putative ssDNA binding groove in Hs-AID is shown by green dashes.

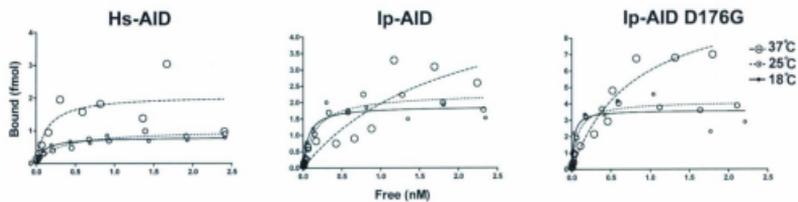


Figure 21. EMSA measuring the binding affinity of AID to ssDNA at different temperatures.

**Figure 21. EMSA measuring the binding affinity of AID to ssDNA at different temperatures.**

EMSA with Hs-AID, Ip-AID and Ip-AID D176G binding to the TGCbub7 substrate at 37°C, 25°C and 18°C. Reactions were carried out with increasing concentrations of TGCbub7 substrate over a range of 100 fold (from 0.025 nM to 25nM). The fraction of shifted (bound) and un-shifted substrate was quantified at each substrate concentration and plotted. Curves with large, medium and small circles represent binding at 37°C, 25°C and 18°C, respectively. Each data point represents the mean of values obtained from 5-10 independent experiments. The half-saturation values ( $K_d$ ) and the corresponding 95% confidence upper limit values for each interaction was determined based on the plotted curves and shown in Table 1. Comparing Ip-AID binding at 37°C to 18°C:  $p=0.008$ . Comparing Ip-AID D176G binding at 37°C to 18°C:  $p=0.005$ .

**Table 1.  $K_d$  values of Hs-AID, Ip-AID and Ip-AID D176G binding to ssDNA at various temperatures.**

	37 °C		25 °C		18 °C	
	$K_d$ (nM) <sup>a</sup>	Upper limit (nM) <sup>b</sup>	$K_d$ (nM) <sup>a</sup>	Upper limit (nM) <sup>b</sup>	$K_d$ (nM) <sup>a</sup>	Upper limit (nM) <sup>b</sup>
<b>Hs-AID</b>	0.15 ± 0.06	0.37	0.31 ± 0.09	0.52	0.10 ± 0.02	0.16
<b>Ip-AID</b>	1.1 ± 0.80	2.4	0.14 ± 0.03	0.18	0.07 ± 0.02	0.12
<b>Ip-AID (D176G)</b>	0.69 ± 0.38	1.6	0.05 ± 0.01	0.08	0.03 ± 0.01	0.06

<sup>a</sup>  $K_d$  (nM) as determined by EMSA

<sup>b</sup> Upper limit value in the 95% confidence range of  $K_d$  (nM)

**Table 1.  $K_d$  values of Hs-AID, Ip-AID and Ip-AID D176G binding to ssDNA at various temperatures.**

The half-saturation values ( $K_d$ ) and the corresponding 95% confidence upper limit values for each interaction was determined based on the plotted curves shown in Figure 21.

## 4. Discussion

### 4.1 Overall rationale and most significant findings

AID is a ssDNA mutating enzyme that initiates secondary antibody diversification processes, such as SHM and CSR, by catalyzing the deamination of C to U in the immunoglobulin genes of B lymphocytes. The downstream pathways following this conversion initiate mutations and double stranded breaks in DNA that result in altered effector functions of antibodies as well as the diversification and increased affinity of antibody receptors. However, AID action is not limited and specifically targeted to immunoglobulin loci. With the potential to produce an effect genome-wide, AID has been implicated in causing the transformation of activated B cells into leukemia and lymphoma, disease aggressiveness and drug resistance. The structure of AID is currently unresolved. As a result, little is known regarding its enzymatic and biochemical properties. For this reason we investigated the characteristics of human AID in comparison to the evolutionarily divergent forms from bony fish in order to elucidate structure-function relationships.

We previously reported that Hs-AID binds to ssDNA with high affinity and a long complex half-life (Larijani et al., 2005). However, partly because the structure of AID has not been determined, the molecular basis for this and other enzymatic properties are unknown. Mutants of human AID with altered WRC specificity or catalytic rates have been described in the past (Kohli et al., 2010; Wang et al., 2009; Wang et al., 2010). While these studies have identified critical residues for AID activity, the mechanisms through which the mutations influence catalysis have not been elucidated. Furthermore,

the biological relevance of laboratory-generated mutations is unclear. Here we examined the biochemical properties of purified wild type AID from different species. We chose to study bony fish AID from zebrafish (*Danio rerio*) and channel catfish (*Ictalurus punctatus*) because their primary AID structure is, to date, the most distant from humans and therefore likely to possess altered characteristics (Barreto & Magor, 2011; Conticello et al., 2005; Wakae et al., 2006). Previous work in the field had shown that AID from these bony fish are functional enzymes when transfected into eukaryotic cells or transformed into bacteria. However, quantifying measurement of the enzymatic activity of AID was not possible due to limitations of the experimental systems (Barreto et al., 2005; Conticello et al., 2005; Ichikawa et al., 2006). Said limitations include the selection imposed by the measurement of AID-mediated antibody resistance gain or GFP reversion mutations in bacteria and eukaryotic cells, respectively, the presence of downstream DNA repair pathways altering the levels and patterns of AID-induced mutations, and variation of AID expression levels. Here we use purified AID in an *in vitro* free system, without the potential for selection or DNA repair, thus allowing for direct comparison of AID activity amongst species.

We hypothesized that individual sequence differences could manifest in different biochemical properties, yielding insights into the molecular mechanisms of AID action. Indeed, we found that wild type bony fish and human AID have different catalytic efficiencies and that these differences are mediated by a single residue in the C-terminus of AID.

#### **4.2 The mechanisms of activity modulation by D176**

In principle, several mechanisms may account for the catalytic differences mediated by the single residue between Ip-AID, Ip-AID D176G and Dr-AID. It is unlikely that the D176G substitution influences deamination catalysis directly since this residue is not proximal to the putative Zn-coordinating and catalytic residues (in Hs-AID: H56, E58, C87, C90, Ip-AID: H59, E61, C98, C101 and Dr-AID: H60, E62, C99, C102) in the primary sequence or the surface model of AID structure (Figure 6, Figure 13 and Figure 20). On the surface charge model of AID that we generated based on the determined structure of APOBEC3G (Figure 20), we identified a putative ssDNA binding groove (Figure 20 bottom left; putative groove denoted by green line, putative catalytic pocket shown in magenta). This groove is in a similar position to the ssDNA binding groove of APOBEC3G which has been suggested based on NMR shifts and mutational studies (Chen et al., 2008; Holden et al., 2008). Although it is unclear whether this groove in AID is involved in binding ssDNA, it is lined with a high number of positively charged amino acids in Ip-AID, Dr-AID and Hs-AID (Figure 20; shown in blue), which may provide a basis for our previous finding that Hs-AID binds ssDNA with high affinity (Larjani & Martin, 2007). However, we noted that in the surface models, the D176 is located in the middle of a 3 residue-ridge identifiable in AID from all three species (D175, D176 and Q179 in Ip-AID) that is not located in the putative ssDNA binding groove (Figure 20, arrows). We thus wondered how this residue influences ssDNA binding. We found that the D176 residue is located proximal to several positively charged residues that are in an equivalent position to residues on APOBEC3G (H367 and H374)

that have been suggested to bind DNA in two studies. One study reported that these residues change positions upon ssDNA by NMR shift analysis, while another study also identified the binding groove and found targeted mutation of these residues diminished APOBEC3G activity, consistent with a role in ssDNA binding (Chen et al., 2008; Holden et al., 2008). Thus, we suggest that by virtue of its proximity to the end of the putative binding groove, this region of AID plays a role in ssDNA binding. In this regard, D176 may exert its influence indirectly through influencing the conformation of a nearby portion of the putative ssDNA binding groove. This model would be consistent with the greater negative effect of this residue on ssDNA binding at 37 °C, when the overall protein structure is more flexible. Whatever the mechanism(s) may be, our data indicate that variations in ssDNA binding affinity result in differences in the catalytic activity of AID amongst species, and that the carboxyl-terminal region of AID contains residues that influence ssDNA binding and thus regulate the enzymatic efficiency of AID.

#### **4.3 The cold adaptation of bony fish AID**

The reaction rate and catalytic efficiency of temperate enzymes decreases as temperature is lowered. Decreased temperatures prevent the enzyme from changing shape during the catalysis process. However, cold-adapted organisms have evolved by having modified proteins that function efficiently under such conditions. Increased temperatures have been shown to affect cold-adapted enzyme by rendering the enzyme more flexible thus reducing thermal stability (D'Amico et al., 2002; Feller & Gerday, 2003). This is a commonality shared among thermosensitive enzymes. Cold adaptation appears in

poikilotherms, those whose body temperatures match that of the environment in which they reside, and not in homeotherms such as mammals, whose body temperature is maintained independently from the external environment (D'Amico et al., 2002). Studies have shown that the fundamental kinetic properties of enzymes are conserved from species to species when they are within the normal physiological range of temperatures encountered by that species (Feller & Gerday, 2003). We chose to explore this notion in the context of AID from bony fish in comparison to that from humans.

In an antibiotic resistance screen in *E. coli*, pufferfish AID was found to be inactive at 37°C, but was functional at 18°C (Conticello et al., 2005). While this result may reflect increased activity of AID at lower temperatures, an alternative explanation is that bacterial expression at lower temperatures leads to more soluble AID (Coker, Morgan, & Petersen, 2006; Larijani & Martin, 2007). In a separate report measuring the activity of Ip-AID and Dr-AID in a mammalian cell line (NIH-3T3) no increase in activity was found when the temperature was lowered from 37°C to 28°C, though the range of temperatures tested were limited presumably by constraints on cell survival (Wakae et al., 2006). Here, we demonstrated that different forms of the AID enzyme from various species, Hs-AID, Dr-AID and Ip-AID, showed a preference to deaminate the TGCBub7 WRC motif sequence but functioned optimally at different temperatures. In this study and in agreement with previous work, Hs-AID was able to deaminate preferentially at 37°C (Larijani et al., 2007). Interestingly, Hs-AID demonstrated decreasing activity with lessening temperatures (Figure 12). Conversely, we report that purified Ip-AID and Dr-AID exhibit high levels of activity at low temperatures and high

thermosensitivity, hallmarks of cold adaptation. Thus, bony fish AID are cold adapted at the enzymatic level.

Two general types of strategies for cold adaptation of enzymes have been reported; the first is overall increased structural flexibility (Gerday et al., 1997; Marshall, 1997; Závodszy et al., 1998). Low temperatures limit “breathing” and thus cold-adapted enzymes ought to be more flexible in order to retain the structural mobility required for catalysis at lower temperatures. The second is increased flexibility or accessibility of substrate binding grooves, possibly by modifying the electrostatic environment in the proximity of grooves (D’Amico et al., 2002; Feller & Gerday, 2003). Our data suggest that temperature-sensitive ssDNA binding affinity of AID determines catalytic efficacy, and together with structural models of the position of D176, are supportive of the latter mechanism. Based on our data comparing the thermosensitivity of hybrid AIDs (Figure 19), we suggest that residues located in the N-terminus are likely important for this attribute of fish AID.

#### **4.4 The significance of our findings with respect to the evolution of AID**

Adaptive immunity and the genetic mechanisms responsible for antibody diversification, such as V(D)J recombination and SHM, are cellular anatomical features that first arose in the Gnathostomata, the superclass representing all jawed vertebrates. The earliest Gnathostomata species are cartilagenous fish (e.g. sharks) (Flajnik & Kasahara, 2010; Pancer & Cooper, 2006). Sharks possess antibody affinity maturation,

though to date only a fragment of shark AID has been cloned (Conticello et al., 2005; Dooley et al., 2006). Within the Gnathostomata superclass, Osteichthyes, characterizing bony fish, appears to be the earliest species identified to date to undergo SHM, contain full known sequences of AID, and are the earliest vertebrates for which AID function relative to the protein structure has been examined. Bony fish are the earliest divergent jawed vertebrates from which AID has been cloned and compared to humans. From the representative species of each evolutionary class examined, including Osteichthyes, Amphibia, Aves, and Mammalian, zebrafish and catfish are the earliest diverged bony fish species with sequenced AID and they differ at residue 176. In more recently derived bony fish there have been a mixture of outcomes with sticklebacks and the two pufferfishes (*Takifugu* & *Tetraodon*) all having E176, and Medaka and Ayu sharing the G176 with zebrafish (unpublished observations). Consistent with our findings, replacement of this residue in Tr-AID also improved the enzyme's activity in a bacterial papillation assay (Wang et al., 2009).

It is possible that the differences in the enzymatic efficiency of human and bony fish AID are mere byproducts of sequence divergence. Alternatively, such differences may be of significance with respect to the evolutionary path of AID. We propose the latter for several reasons: first, although zebrafish and channel catfish belong to the sister taxa of Cypriniformes and Siluriformes respectively, there is evidence to suggest that Siluriformes arose earlier (Nakatani et al., 2011; Steinke, Salzburger, & Meyer, 2006). Second, Dr-AID is more closely related to Hs-AID than Ip-AID is to Hs-AID (Figure 6, Figure 7). Third, the G residue of Dr-AID is conserved in all land-dwelling species.

Fourth, APOBEC2 which is an ancestral member of the APOBEC family (expressed in cartilaginous fish) contains a D residue similar to Ip-AID, while APOBEC3 enzymes which are the most recently diverged members (primate specific) contain a G residue, in the equivalent position. Taken together, these observations suggest that D176 was the ancestral residue in AID and that its alteration into a G residue during the evolution of AID resulted in a more active enzyme. It is tempting to speculate that this event reflects an increase in evolutionary pressure on adaptive immunity due to environmental pathogens.

The AID/APOBEC family is believed to have evolved from the tRNA adenosine deaminases, present from bacteria to metazoans (Coticello, 2008). Partial AID sequences have been identified in the shark, a member of the cartilaginous fishes that arose prior to bony fish. Sharks and rays immunoglobulin genes possess a single V-element that may recombine with a single J- or DJ-element, believed by many to be a very early immunoglobulin gene (Greenberg et al., 1995). Some have argued that sharks can develop a considerable primary antibody repertoire without the need for further diversification through AID by clustering multiple Ig loci and combining up to three D-elements into a single VDDDJ exon (Greenberg et al., 1995; Zhao et al., 2005). Conversely, others have stated that sharks use AID for secondary repertoire development, supported by observations of accumulated mutations on Ig light chain genes expressed in secondary lymphoid organs representative of antigen driven selection (Dooley et al., 2006). Limited studies of elasmobranchs and their use of somatic mutation have been performed and future studies examining the homology of cartilaginous AID to that of

bony fish as well as similarities and differences in enzymatic properties will shed light on the functionality of the earliest forms of AID.

#### **4.5 Conclusions and future work**

In this study, we examined the biochemical properties of purified wild type AID from different species including bony fish channel catfish, zebrafish and human. We hypothesized that AID from divergent species were likely to possess altered properties which could be explained by sequence differences, thus, yielding insight into the molecular mechanism of AID action. We found that wild type bony fish and human AID have distinct catalytic efficiencies. Cross-species differences in AID expression and post-translational modifications may counterbalance the inherent catalytic differences between Ip-AID and Dr-AID, *in vivo*. For instance, phosphorylation of Hs-AID at S38 has been shown to modulate its activity by promoting interaction with replication protein A (RPA) (Chaudhuri, Chan, & Frederick, 2004). In this regard, one notable difference is that like Hs-AID, Ip-AID has a S38 residue. In contrast, Dr-AID lacks this residue, but has a phosphomimetic D44 residue that has been shown to interact with RPA (Basu, Wang, & Frederick, 2008). Given these differences, it will be of interest to compare the *in vivo* levels of AID-induced mutations and double-strand breaks between the two fish.

Ongoing and future studies will investigate the biochemical properties of AID from various species including cartilaginous fish, multiple other bony fish and tetrapods. Structural models will be generated to guide and assist data interpretation, in a similar

manner as described in this thesis. Once AID from various species have been characterised *in vitro*, it will be of interest to determine whether the differences in activity reflect differences in DNA-damage levels induced by AID *in vivo*.

## 5. References Appendices

- Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M. & Olsson, C. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol*, 166 (8), 5051-7.
- Barreto, V. M., Ramiro, A. R. & Nussenzweig, M. C. (2005). Activation-induced deaminase: Controversies and open questions. *Trends Immunol*, 26 (2), 90-6.
- Barreto, V. M. & Magor, B. G. (2011). Activation-induced cytidine deaminase structure and functions: A species comparative view. *Dev Comp Immunol*, 35 (9), 991-1007.
- Barreto, V. M., Pan-Hammarstrom, Q., Yaofeng, Z., Hammarstrom, L., Misulovin, Z. & Nussenzweig, M. C. (2005). AID from bony fish catalyzes class switch recombination. *J Exp Med*, 202 (6), 733-8.
- Bascove, M. & Fripiat, J.P. (2010). Molecular characterization of pleurodeles waltl activation-induced cytidine deaminase. *Mol Immunol*, 47 (7-8), 1640-9.
- Basu, U., Wang, Y. & Frederick, W. A. (2008). Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase. *Mol Cell*, 32 (2), 285-91.
- Beale, R. C. L., Petersen, S. K., Watt, I. N., Harris, R. S., Rada, C. & Neuberger, M. S. (2004). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: Correlation with mutation spectra in vivo. *J Mol Biol*, 337 (3), 585-96.
- Bransteitter, R., Phuong, P., Scharff, M. D. & Goodman, M. F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A*, 100 (7), 4102-7.

- Chaudhuri, J. & Frederick, W. A. (2004). Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol*, 4 (7), 541-52.
- Chaudhuri, J., Chan, K. & Frederick, W. A. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature*, 430 (7003), 992-8.
- Chen, K., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y. & Shindo, K. (2008). Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature*, 452 (7183), 116-9.
- Coker, H. A., Morgan, H. D., & Petersen, S. K. (2006). Genetic and in vitro assays of DNA deamination. *Methods Enzymol*, 408, 156-70.
- Coker, H. A. & Petersen, S. K. (2007). The nuclear DNA deaminase AID functions distributively whereas cytoplasmic APOBEC3G has a processive mode of action. *DNA Repair (Amst)*, 6 (2), 235-43.
- Coticello, S. G. (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol*, 9 (6), 229.
- Coticello, S. G., Thomas, C. J. F., Petersen, S. K. & Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol*, 22 (2), 367-77.
- Coticello, S. G., Langlois, M. A., Yang, Z. & Neuberger, M. S. (2007). DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol*, 94, 37-73.

- D'Amico, S., Claverie, P., Collins, T., Georgette, D., Gratia, E., Hoyoux, A., Meuwis, M., Feller, G. & Gerday, C. (2002). Molecular basis of cold adaptation. *Phil Trans R Soc Lond B*, 357, 917-25.
- Dancyger, A.M., King, J.J., Quinlan, M.J., Fifield, H., Tucker, S., Saunders, H.L., Burry, M., Major, B.G., Martin, A. & Larijani, M. (2011). Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding. *FASEB J*, 26 (4), 1517-25.
- Danilova, N., Bussmann, J., Jekosch, K. & Steiner, L. A. (2005). The immunoglobulin heavy-chain locus in zebrafish: Identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat Immunol*, 6 (3), 295-302.
- de Yebenes, V. G. & Ramiro, A. R. (2006). Activation-induced deaminase: Light and dark sides. *Trends Mol Med*, 12 (9), 432-439.
- Dooley, H., Stanfield, R. L., Brady, R. A. & Flajnik, M. F. (2006). First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. *Proc Natl Acad Sci U S A*, 103 (6), 1846-51.
- Durandy, A., Peron, S. & Fischer, A. (2006). Hyper-IgM syndromes. *Curr Opin Rheumatol*, 18 (4), 369-76.
- Feller, G. & Gerday, C. (2003). Psychrophilic enzymes: Hot topics in cold adaptation. *Nat Rev Microbiol*, 1 (3), 200-8.
- Flajnik, M. F. & Kasahara, M. (2010). Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nat Rev Genet*, 11 (1), 47-59.
- Fleisher, T.A. & Tomar, R.H. (1997). Introduction to diagnostic laboratory immunology. *JAMA*, 278 (22), 1823-57.

- Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem*, 71, 101-32.
- Ganesh, K. & Neuburger, M. (2011). The relationship between hypothesis and experiment in unveiling the mechanisms of antibody gene diversification. *FASEB J*, 25 (4), 1123-32.
- Gerday, C., Aittaleb, M., Arpigny, J. L., Baise, E., Chessa, J. P. & Garsoux, G. (1997). Psychrophilic enzymes: A thermodynamic challenge. *Biochim Biophys Acta*, 1342 (2), 119-31.
- Gordon, M. S., Kanegai, C. M., Doerr, J. R. & Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). *Proc Natl Acad Sci U S A*, 100 (7), 4126-31.
- Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C. & Flajnik, M. F. (1995). A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*, 374 (6518), 168-73.
- Hansen, J. D., Landis, E. D. & Phillips, R. B. (2005). Discovery of a unique ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. *Proc Natl Acad Sci U S A*, 102 (19), 6919-24.
- Holden, L. G., Prochnow, C., Y, P. C., Bransteitter, R., Chelico, L. & Sen, U. (2008). Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*, 456 (7218), 121-4.

- Ichikawa, H. T., Sowden, M. P., Torelli, A. T., Bachl, J., Pinwei, H. & Dance, G. S. C. (2006). Structural phylogenetic analysis of activation-induced deaminase function. *J Immunol*, 177 (1), 355-61.
- Kohli, R. M., Maul, R. W., Gurmink, A. F., McClure, R. L., Gajula, K. S. & Husein, S. (2010). Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J Biol Chem*, 285 (52), 40956-64.
- Küppers, R. & Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*, 20 (40), 5580-94.
- Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer*, 5 (4), 251-62.
- Lanasa, M. C. & Weinberg, J. B. (2011). Immunoglobulin class switch recombination in chronic lymphocytic leukemia. *Leuk Lymphoma*, 52 (7), 1398-1400.
- Larijani, M. & Martin, A. (2007). Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol Cell Biol*, 27 (23), 8038-48.
- Larijani, M., Petrov, A. P., Kolenchenko, O., Berru, M., Sergey, N. K. & Martin, A. (2007). AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol Cell Biol*, 27 (1), 20-30.
- Larijani, M., Frieder, D., Basit, W. & Martin, A. (2005). The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in *ung(-/-)msh2(-/-)* mice. *Immunogenetics*, 56 (11), 840-5.

- Larijani, M., Frieder, D., Sonbuchner, T. M., Bransteitter, R., Goodman, M. F. & Bouhassira, E. E. (2005). Methylation protects cytidines from AID-mediated deamination. *Mol Immunol*, 42 (5), 599-604.
- Liddament, M. T., Brown, W. L., Schumacher, A. J. & Harris, R. S. (2004). APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr Biol*, 14 (15), 1385-91.
- Longerich, S., Basu, U., Alt, F. & Storb, U. (2006). AID in somatic hypermutation and class switch recombination. *Curr Opin Immunol*, 18 (2), 164-74.
- Marshall, C. J. (1997). Cold-adapted enzymes. *Trends Biotech*, 15 (9), 359-64.
- Martin, A., Bardwell, P. D., Woo, C. J., Manzia, F., Shulman, M. J. & Scharff, M. D. (2002). Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature*, 415 (6873), 802-6.
- McCarthy, H., Wierda, W. G., Barron, L. L., Cromwell, C. C., Jing, W. & Coombes, K. R. (2003). High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia. *Blood*, 101 (12), 4903-8.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, 102 (5), 553-63.
- Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K. & Davidson, N. O. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem*, 274 (26), 18470-6.

- Müschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K. & Küppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J Exp Med*, 192 (12), 1833-40.
- Mussmann, R., Courtet, M., Schwager, J. & Pasquier, L. D. (1997). Microsites for immunoglobulin switch recombination breakpoints from xenopus to mammals. *Eur J Immunol*, 27 (10), 2610-9.
- Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K., & Nishida, M. (2011). Evolutionary history of otophysi (teleostei), a major clade of the modern freshwater fishes: Pangaeian origin and mesozoic radiation. *BMC Evol Biol*, 11, 177.
- Nambu, Y., Sugai, M., Gonda, H., Chung-Gi Lee, Katakai, T. & Agata, Y. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science*, 302 (5653), 2137-40.
- Navaratnam, N., Morrison, J. R., Bhattacharya, S., Patel, D., Funahashi, T. & Giannoni, F. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem*, 268 (28), 20709-12.
- Neuberger, M. S., Di Noia, J. M., Beale, R. C. L., Williams, G. T., Yang, Z. & Rada, C. (2005). Somatic hypermutation at A.T pairs: Polymerase error versus dUTP incorporation. *Nat Rev Immunol*, 5 (2), 171-8.
- Ohtani, M., Miyadai, T. & Hiroishi, S. (2006). Identification of genes encoding critical factors regulating B-cell terminal differentiation in torafugu (takifugu rubripes). *Comp Biochem Physiol Part D Genomics*, 1 (1), 109-14.
- Okazaki, I., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M. & Kinoshita, K. (2003). Constitutive expression of AID leads to tumorigenesis. *J Exp Med*, 197 (9), 1173-81.

- Okazaki, I., Kinoshita, K., Muramatsu, M., Yoshikawa, K. & Honjo, T. (2002). The AID enzyme induces class switch recombination in fibroblasts. *Nature*, 416 (6878), 340-5.
- Pancer, Z. & Cooper, M. D. (2006). The evolution of adaptive immunity. *Annu Rev Immunol*, 24, 497-518.
- Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A. & Baldini, L. (1998). BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside ig loci. *Proc Natl Acad Sci U S A*, 95 (20), 11816-21.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R. S., Kuppers, R. & Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*, 412 (6844), 341-6.
- Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P. & Muramatsu, M. (2008). AID is required for germinal center-derived lymphomagenesis. *Nat Gen*, 40 (1), 108-12.
- Pham, P., Bransteitter, R., Petruska, J. & Goodman, M. F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, 424 (6944), 103-7.
- Pham, P., Chelico, L. & Goodman, M. F. (2007). DNA deaminases AID and APOBEC3G act processively on single-stranded DNA. *DNA Repair (Amst)*, 6 (6), 689-92.
- Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*, 445 (7126), 447-51.

- Rada, C., Di Noia, J. & Neuberger, M. S. (2004). Mismatch recognition and uracil excision provide complementary paths to both ig switching and the A/T-focused phase of somatic mutation. *Mol Cell*, 16 (2), 163-71.
- Ramiro, A. R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S. & Muramatsu, M. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*, 118 (4), 431-8.
- Reiniger, L., Bodor, C., Bogнар, A., Balogh, Z., Csomor, J. & Szepesi, A. (2006). Richter's and prolymphocytic transformation of chronic lymphocytic leukemia are associated with high mRNA expression of activation-induced cytidine deaminase and aberrant somatic hypermutation. *Leukemia*, 20 (6), 1089-95.
- Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A. & Sanal, O. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*, 102 (5), 565-75.
- Roux, K H. (1999). Immunoglobulin structure and function as revealed by electron microscopy. *Int Arch Allergy Immunol* 120 (2), 85-99.
- Saunders, H. L., Oko, A. L., Scott, A. N., Chia, W. F. & Major, B. G. (2010). The cellular context of AID expressing cells in fish lymphoid tissues. *Develop Comp Immunol*, 34 (6), 669-76.
- Saunders, H. L. & Magor, B. G. (2004). Cloning and expression of the AID gene in the channel catfish. *Develop Comp Immunol*, 28 (7-8), 657-63.
- Shandilya, S. M. D., Madhavi, N. L. N., Nalivaika, E. A., Gross, P. J., Valesano, J. C. & Shindo, K. (2010). Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure*, 18 (1), 28-38.

- Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral vif protein. *Nature*, 418 (6898), 646-50.
- Shen, H. M., Peters, A., Baron, B., Zhu, X., & Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of ig genes. *Science*, 280 (5370), 1750-2.
- Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D. D., & Bieniasz, P. D. (2005). Natural variation in vif: Differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog*, 1 (1), e6.
- Stavnezer, J. & Amemiya, C. T. (2004). Evolution of isotype switching. *Seminars in Immunol*, 16 (4), 257-75.
- Steinke, D., Salzburger, W. & Meyer, A. (2006). Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J Mol Evol*, 62 (6), 772-84.
- Teng, B., Burant, C. F. & Davidson, N. O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science*, 260 (5115), 1816-9.
- Wakae, K., Magor, B. G., Saunders, H., Nagaoka, H., Kawamura, A. & Kinoshita, K. (2006). Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. *Int Immunol*, 18 (1), 41-7.
- Wang, M., Rada, C. & Neuberger, M. S. (2010). Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J Exp Med*, 207 (1), 141-53.

- Wang, M., Yang, Z., Rada, C. & Neuberger, M. S. (2009). AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. *Nat Struct Mol Biol*, 16 (7), 769-76.
- Warrington, R., Wade, W., Kim, H. L. & Antonetti, F. R. (2011). An introduction to immunology and immunopathology. *Allergy Asthma Clin Immunol*, 7 Supp 1, S1.
- Willis, T. G. & Dyer, M. J. (2000). The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. *Blood*, 96 (3), 808-22.
- Yang, S. Y. & Schatz, D. G. (2007). Targeting of AID-mediated sequence diversification by cis-acting determinants. *Adv Immunol*, 94, 109-25.
- Yoshikawa, K., Okazaki, I. M., Eto, T., Kinoshita, K., Muramatsu, M. & Nagaoka, H. (2002). AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science*, 296 (5575), 2033-6.
- Závodszy, P., Kardos, J., Svingor & Petsko, G. A. (1998). Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc Natl Acad Sci U S A*, 95 (13), 7406-11.
- Zhao, Y., Pan-Hammarstram, Q., Zhihui, Z. & Hammarstram, L. (2005). Identification of the activation-induced cytidine deaminase gene from zebrafish: An evolutionary analysis. *Develop Comp Immunol*, 29 (1), 61-71.

