RNA enhances the activity of purified AID on DNA sequences of Immunoglobulin switch regions

by

Hala Abdouni (B.Sc.)

A thesis submitted to the school of Graduate Studies in

partial fulfillment of the requirements for the degree of

Master of Science in Medicine

Graduate program in Immunology and Infectious disease,

Faculty of Medicine

Memorial University of Newfoundland

October 2017

St. John's, Newfoundland

Abstract

Activated B lymphocytes in peripheral lymphoid organs undergo class switch recombination and somatic hypermutation of antibody genes to produce antibodies with higher antigen-binding affinity. These processes are initiated by the DNA mutator enzyme activation–induced cytidine deaminase (AID) which mutates cytidine to uridine at Immunoglobulin loci. AID activity also results in mutations across the genome which in turn causes and exacerbates cancers. *In vitro*, AID mutates single stranded but not double-stranded oligonucleotide DNA substrates whereas *in vivo*, it acts preferentially on loci that are robustly transcribed or contain unusual transcriptional features. Taken together, AID is thought to target single stranded DNA liberated during transcription. In addition, AID has been shown to interact with factors associated with transcription, including the RNA exosome and spliceosome components, and with non-coding RNA itself. Here, we examined the behavior of purified AID on DNA/RNA hybrid substrates designed to simulate structures found at Immunoglobulin loci. We found that when composed of Immunoglobulin switch sequences, DNA/RNA hybrids significantly boost the catalytic activity of AID. We also identified numerous AID mutants which are catalytically inactive on DNA, whose activity is restored by proximity to RNA. Given the prevalence of RNA:DNA hybrids at Immunoglobulin switch sequences, our finding that RNA directly enhances the catalytic activity of AID on DNA suggest a novel role for RNA in regulating AID activity.

Acknowledgments

Firstly, I would like to express my outmost gratitude to my supervisor and mentor Dr. Mani Larijani for providing me with the opportunity of being a part of your lab team, first as a science technician and then as a master's student. Your continuous support, encouragement and advice throughout my journey in graduate school motivated me to work hard and aspire to be the best researcher that I can be.

Secondly, I want to extend the sentiments towards my Committee Supervisory Members, Dr. Hirasawa and Dr. Drover, for providing me with guidance and insight towards my research project as well as my performance during seminars and classes.

Thirdly, I would like to give my thanks to the lab team, students and staff alike. Without your professional advice and emotional support, I wouldn't be able to navigate through graduate school with such ease.

Lastly and most importantly, my love and appreciation goes to my family and my parents specially, for getting me to where I am today. It wouldn't have been possible without your unceasing encouragement, care and constant sacrifices that go above and beyond what a normal person can handle. Having role models like you made the hard work and the long hours easier to manage. You've helped me realize and reach my goal day in day out and never allowed me to doubt myself. I'm forever in your debt. *Thank you. Thank you. Thank you.*

Note: *The work described in this thesis has been submitted for publication in Scientific Reports as the following manuscript:*

"RNA enhances the activity of purified AID on DNA sequences of Immunoglobulin switch regions"

Hala S. Abdouni, Justin J. King, Atefeh Ghorbani, Heather Fifield, Lesley Berghuis, Mani Larijani

Author Contribution Statement: HA carried out the work described in figures 6-25, except for figure 15, 16, 19 and 20. JK carried the work described in 19 and 20 with contribution to figure 21. AG carried out the work described in figures 15 and 16. HF and LB contributed to expression and purification of AID used in all experiments. HA, JK and ML prepared the figures and the manuscript.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

µg	Microgram
µl	Microliter
°C	Degrees Celsuis
A	Adenine
AID	Activation-Induced Cytidine Deaminase
Amp	Ampicillin
APOBEC	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like
BCR	B-cell receptor
BSA	Bovine serum albumin
C	Cytidine
C region	Constant region of Ig
CDR	Complementarity-determining regions
CSR	Class switch recombination
DE3	*Escherichia coli* BL21
DNA	Deoxyribonucleic acid
DSB	Double stranded break
DTT	Dithiothreitol
EDTA	Ethyldiamine tetraacetic acid

EMSA	Electrophroretic mobility shift assay
Fab	Fragment antigen-binding
FAS	Fas cell surface death receptor
Fc	Fragment crystallizable region
g	Gram
G	Guanine
GC	Germinal Center
GST	Glutathione S-transferase
IDT	Integrated DNA technologies
Ig	Immunoglobulin
IgH	Immunoglobulin heavy chain
IPTG	Isopropyl-β thiogalactopyranoside
kDa	Kilodalton
LB	Lysogeny broth
M	Molar
MBN	Mung bean Nuclease
ml	Milliliter
Myc	v-Myc myelocytomatosis viral oncogene homolog (avian)

NEB	New England Biolabs
nt	Nucleotide
ORF	Open reading frame
PAMP	Pathogen-associated molecule pattern
PBS	Phosphate buffer saline
PAX-5	Paired box 5
PCR	Polymerase chain reaction
PPR	Pattern recognition receptors
RAG1/2	Recombination activation genes
RHOH	Ras homolog family member H
RNA	Ribonucleic acid
RSSs	Recombiantion signal sequences
S	Switch region of Ig
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Standard error of mean
SHM	Somatic hypermutation
SOC	Super optimal broth with catabolite repression
SR	Switch Region

ssDNA	Single stranded DNA
T	Thymidine
TBE	Tris borate/EDTA electrophoresis buffer
T_C	cytotoxic T cell
TCR	T cell receptor
TEMED	Tetramethylethylenediamine
T_H	T helper cell
U	Uracil
U-CLL	Unmutated Chronic lymphocytic leukemia
UDG	Uracil-DNA glycosylase
WRC	a ssDNA hotspot motif targeted by AID where W=A/T and R=A/G
wt	wild-type activation-induced cytidine deaminase
V(D)J	Recombination of variable (V), diversity (D) and joining (J) genes
Zn	Zinc

Introduction

1.1 The immune system

The immune system across species varies from basic and rudimentary in bacteria to complex subsystems in multi cellular organisms. In humans, for example, the immune system consists of two main subsystems: innate and adaptive immunity. Such a defence system ensures a thorough protection for the host in parallel with their surrounding environment and constantly evolving pathogens (Beck et al., 1996; Fliesher, 1997, Warrington et al., 2011).

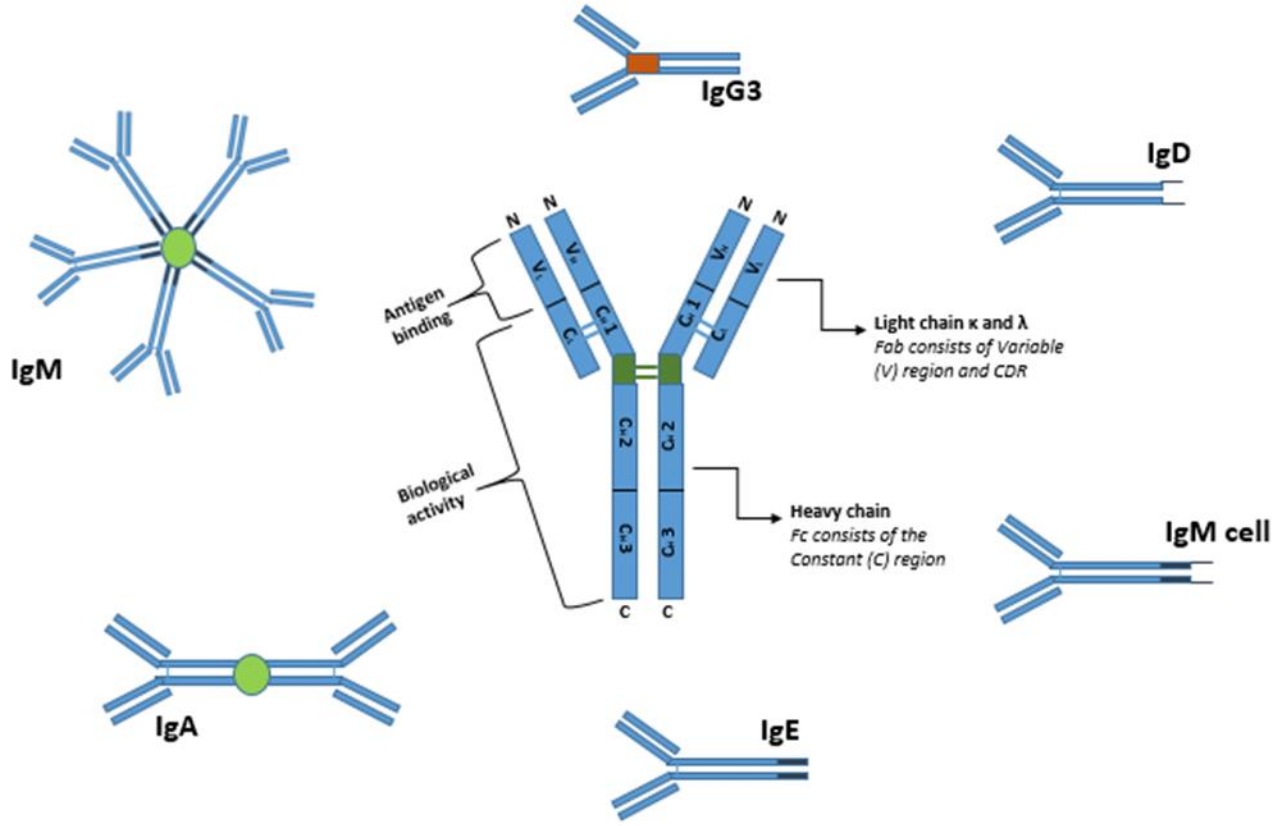
During an innate immune response which is considered the first line of antigen and toxin resistance in the organism, an immediate protection against foreign invaders, be it through physical barriers such as the skin, or more specific chemical barriers that involve secretion of mucus, saliva, tears and several gastrointestinal tract enzymes is provided. This subsystem also encompasses specific types of cells that are responsible for capturing pathogens and destroying them with the help of the complement system. These cells present *pattern recognition receptors* (PPRs) that recognize pathogen-associated molecular patterns (PAMPs) thereby differentiating pathogens from host molecules (Akira et al., 2006).

The adaptive immune response on the other hand provides more specific and long-lasting protection to the organism by creating immunological memory from an initial response. The cells involved in this subsystem are known as lymphocytes which are highly specialized and specific to particular pathogens. These lymphocytes consist of B-cells which mature in the bone

marrow and T-cells that mature in the thymus. T-cells contribute to cell mediated immune responses through its various subsets. T helper (T_H) cells, for example, regulate the active immune response while cytotoxic (T_C) cells wipe out virus-infected cells and tumor cells. B-cells are responsible for antibody responses and include antibody secreting plasma cells and memory cells that allow protection from reoccurring foreign pathogens. B-cells specifically, play an important role in the adaptive immune response in that they interact with a wide range of antigens that activates them to further specialize under affinity maturation.

1.2 Antibodies in the context of the immune system

To achieve this great influence on host protection, naïve (inactivated) B-cells are covered with low affinity antibodies that are also called surface-expressed receptors (BCRs) or immunoglobulins (Ig). Antibodies are "Y" shaped proteins made up of two light and heavy polypeptide chains that are connected by disulphide bonds (Fig. 1). The polypeptide chains are further divided in to variable (V) regions and constant (C) regions. The V regions of the light chains and the highly variable loops of the heavy chain located at the "arms" of the "Y" shaped antibody are responsible for identifying and binding to antigens. The C region on the other hand serves as an anchor in the B-cell membrane as well as a basis for the isotope and effector functions of the antibody by binding to receptors on other immune cells (Barreto et al., 2005; Fleisher, 1997).



(Adapted from abcam protocols)

Figure 1. Antibody structure and isotypes

Antibodies are made of 2 light chains and 2 heavy chains shaped in a loose "Y" form. Fab, which is located at the N-terminus or the arm of the molecule is an antigen-binding region that is highly variable and specific. Fc on the other hand is located at the C-terminus and encompasses the tail of the antibody. It is responsible for interacting with cell surface receptors and complement proteins. Antibodies can be differentiated in to five isotypes (IgM, IgG, IgD, IgA and IgE) that vary in function.

1.3 Primary and Secondary antibody diversification

Functional antibodies are first developed through the rearrangement of their genes in developing B lymphocytes in the bone marrow. This primary diversification process of antibodies is unique to lymphocytes and is also known as V(D)J recombination where V stands for "Variable", D for "Diversity" and J for "Joining". This process involves recombining signal DNA sequences (RSSs) that target RAG1 and RAG2 recombinases to the V, D and J genes to be cut and then joined back together by DNA repair enzymes (Gellert, 2002). The heavy chain genes are rearranged first and include the V, D and J segments that in turn account for the highly variable antigen-binding region of antibodies. This is followed by rearranging a paired "surrogate" (κ or λ) light chain through VJ recombination. Secondary antibody diversification processes are initiated in mature B-cells once they are activated by binding to an antigen, thus activating B-cells to migrate to secondary lymphoid organs (lymph nodes) to form germinal centers (GC) where they interact with T cells and become antibody-producing plasma cells that produce soluble antibodies (Gellert, 2002; Victora and Nussenzweig et al., 2012). Here, activated B-cells will undergo antibody affinity maturation, a process that will result in production of higher affinity and more effective antibodies thereby allowing a stronger immune response against the specific antigen. Affinity maturation requires class switch recombination (CSR) and somatic hypermutation (SHM), followed by cellular selection of B-cells. In CSR, the mature B-cells which produce IgM switch their antibody isotype to IgG, IgA and IgE isotypes. This occurs by

changing the antibody gene's constant regions (C_H) from C_μ to other C_H genes, while retaining the same variable regions (Fig. 3) (Lanasa et al., 2011). This allows the antibody to preserve antigen specificity while assuming different effector functions from opsonisation, complement activation to antibody-dependent cell-mediated cytotoxicity (ADCC) and degranulation. On the other hand, SHM involves introducing numerous mutations in the form of single base substitutions in the genes that encode the variable region of antibodies. (Fig. 4). More specifically, it results in mutations in the hypervariable regions that correlate with the complementarity determining regions (CDRs) which encompass antigen recognition and binding on the antibody (Longerich, Basu, Alt, & Storb, 2006; Teng & Papavasiliou, 2007).

SHM and CSR are initiated by a common DNA-specific modifying enzyme, activation induced cytidine deaminase (AID) (Longerich, Basu, Alt, & Storb, 2006). AID, a major player in antibody diversity, belongs to the Zn-dependent Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like (APOBEC) superfamily of deaminases that mutate cytidine (C) to uridine (U) in RNA or deoxycytidine (dC) to deoxyuridine (dU) in DNA (Conticello, 2008) and is found and expressed in activated B-cells (Honjo, 2008). It is 24 kDa in size and in humans it's 198 amino acids long (Fig. 2). In Somatic Hypermutation, AID acts by introducing point mutations in the antigen-binding coding sequences of the variable regions in immunoglobulin genes therefore creating higher affinity antibodies (Martin et al., 2002). However, in Class switch recombination, AID mutates G-C rich switch region sequences downstream of the constant

region genes thereby initiating double strand breaks that mediate recombination of the constant region genes (Okazaki et al., 2002).

The essential role AID plays in these secondary antibody processes can be reflected in the absence of AID (Revy et al., 2000). It has been shown that while B-cell maturation or activation is not affected, AID deficient mice and humans lack CSR and SHM processes and thereby they suffer from Hyper IgM syndrome type immunodeficiency (HIGM) characterized by elevated IgM levels and diminished levels of switched immunoglobulin isotypes IgG, IgA and IgE causing a lack of high affinity antibodies and recurrent infections (Muramatsu et al., 2000). Hyper IgM syndrome consists of five subtypes that include defects in CD40LG (HIGM 1), AID (HIGM 2), CD40 (HIGM 3), UNG (HIGM 5) genes and class switch recombination downstream of the AID gene (HIGM 4). The most common defects are X-linked and autosomal recessive found in CD40L and AID genes and represent 70% and 25% cases of HIGM syndrome respectively (Cabral- Marques et al, 2014).To date, 47 AID mutations that include missense mutations, point mutations or frame shifts that cause premature stop codons, deletions and mutations that introduce improper splicing of the RNA have been reported (Table 1) (Imai et al., 2005, Durandy et al., 2006, Ouadani, et al., 2015, Caratão et al., 2013, Trotta et al, 2016). While all these mutations result in CSR deficiency, some also contribute to SHM impairment.

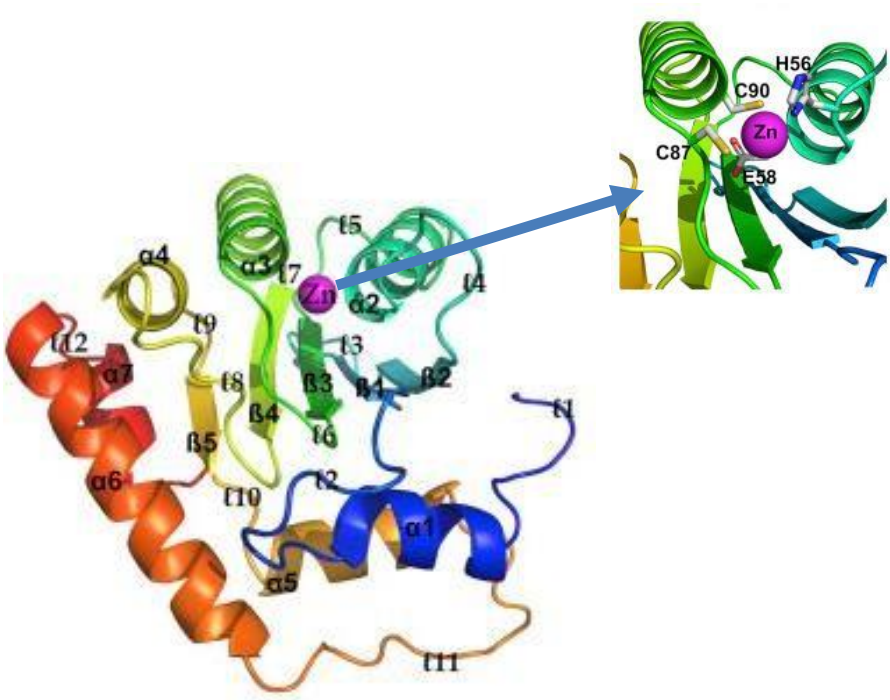
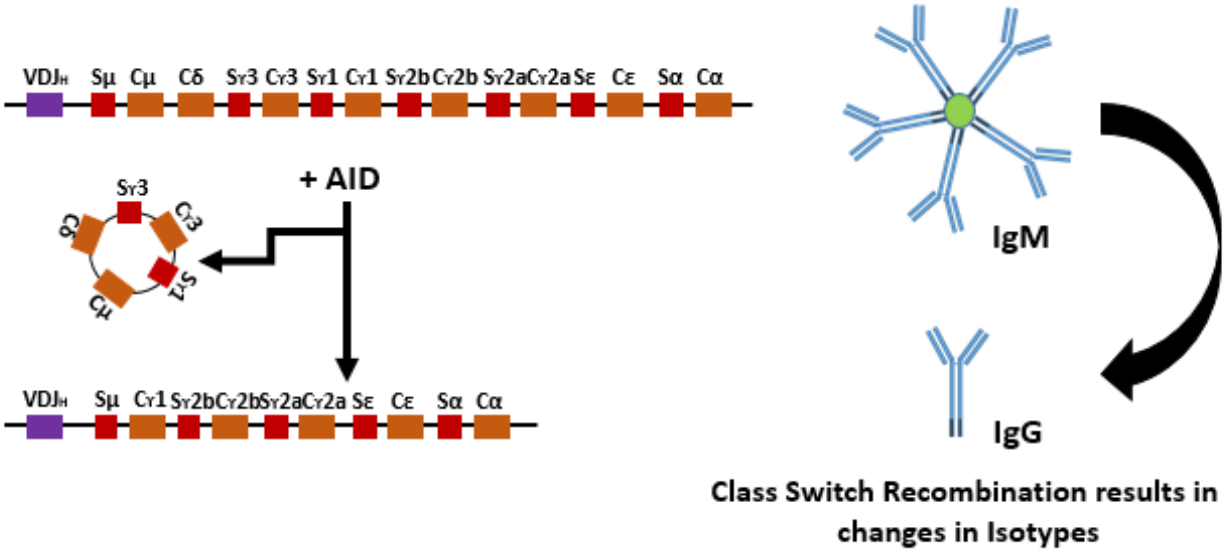


Figure 2. A ribbon model of a human wild-type AID

A representative ribbon model of AID was predicted using I-TASSER and visualized using Pymol v1.30 based on the APOBEC family X-ray structures. Negatively charged residues shown in red (N-terminal) progress towards blue for positively charged residues (C-terminal). The Zn binding catalytic pocket is shown in purple and is characterized by Zn-coordinating (H56, C87, and C90) and catalytic (E58) residues.

Table 1. Updated list of AID mutations in HIGM 2 patients

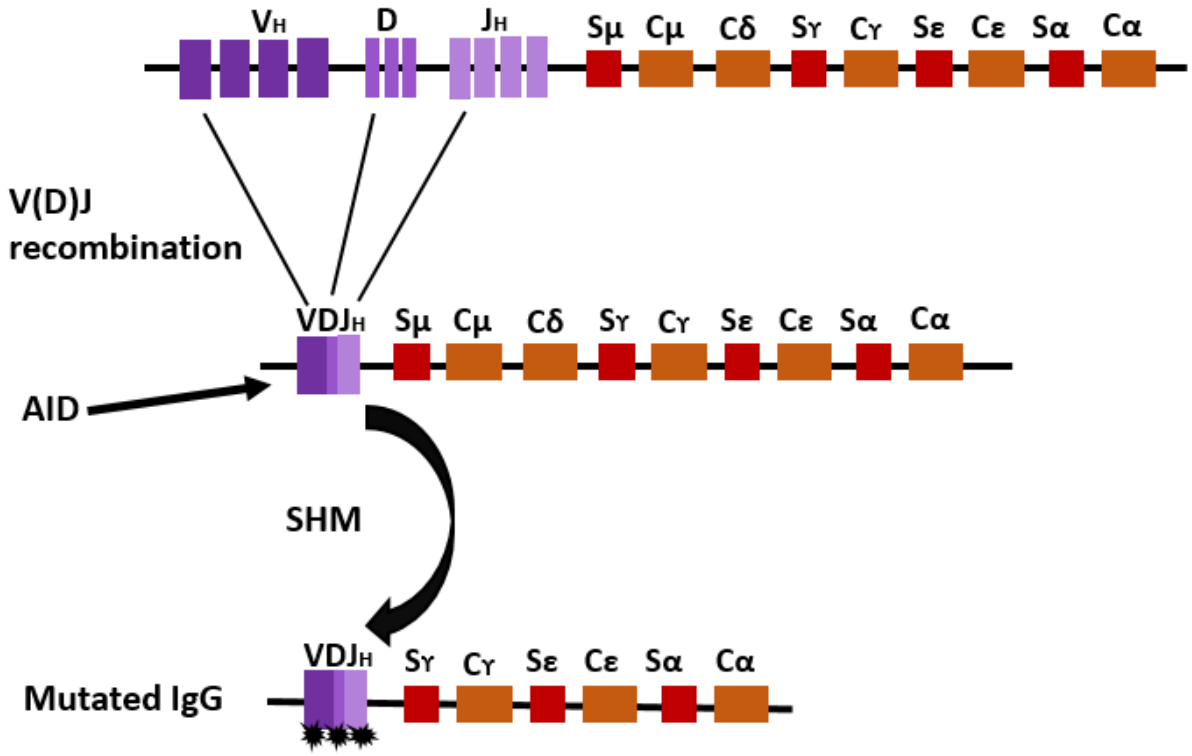
AID mutations library					
1	S3G\N168S	1 + α 6	**25**	L113P	ℓ 8
2	K10R	α 1	**26**	E122X	α 4
3	M6T	ℓ 1	**27**	R112C	ℓ 8
4	R8-Y13del +F15X	ℓ 1 + α 1	**28**	R112H	ℓ 8
5	L22X	ℓ 2	**29**	G133V	ℓ 9
6	F11V	α 1	**30**	C116X	ℓ 8
7	R24W	ℓ 2	**31**	C147X	α 5
8	Y31X	β 1	**32**	I136K	β 5
9	Y31H	β 1	**33**	G125E	α 4
10	A25C	ℓ 2	**34**	M139T	ℓ 10
11	S43P	ℓ 3	**35**	F151S	α 5
12	H56Y	α 2	**36**	D143-L181delinsAfsX4	α 5 + α 6
13	H56-E58del + L59F	α 2	**37**	M139V	ℓ 10
14	W68X	ℓ 5	**38**	S169X	α 6
15	W79-F81del	β 3	**39**	R174S	α 6
16	E58K	α 2	**40**	L181_P182ins31	α 6 + α 7
17	W84X	ℓ 6	**41**	P182fsX208	α 7
18	S85N	ℓ 6	**42**	R190X	α 7
19	L98R	α 3	**43**	L181delinsCfsX26	α 6
20	H56X	α 2	**44**	H130P	α 4
21	C87S	α 3	**45**	F15L	α 1
22	C87R	α 3	**46**	W80R	β 3
23	S83P	ℓ 6	**47**	L106P	β 4
24	A111E	ℓ 8			



(Adapted from Maizels, 2000)

Figure 3. Class switch recombination entails DNA deletion

Class switch recombination process takes place in the germinal centers of the lymph nodes and is initiated by AID which introduces double stranded breaks in the switch (S) regions resulting in the removal of the Cμ constant region and joining of a new constant region. This allows for different antibody isotypes with different effector functions to be expressed.



(Adapted from Bergmann, 2013)

Figure 4. Schematic illustration of Somatic Hypermutation

Another secondary antibody diversification process that occurs in the germinal center of the lymph nodes is Somatic Hypermutation. It targets antigen binding regions in the variable sequences of both heavy and light chains and increases their affinity to specific antigens by introducing point mutations in the V(D)J exon. Stars indicate mutations in the DNA

1.4 AID's impact on cancer

In contrast to its beneficial and necessary functions for antibody diversification, AID was also shown to mistarget and mutate genome-wide thereby creating DSBs on non-Ig genes which can lead to chromosomal translocations, causing and exacerbating cancers. In fact, AID was revealed to deaminate more than half of all transcribed non-Ig genes in germinal center B-cells (Liu et al., 2008). As another example, chromosomal translocations with several oncogenes in Ig switch region DNA in B lymphocytes that are induced by AID deamination, were also found to associate with extremely high percentages of lymphoid leukemias and lymphomas (Ramiro et al., 2004). A study has shown that during CSR, AID produces double stranded DNA breaks in immunoglobulin heavy-chain locus IgH which leads to c-myc/IgH reciprocal translocations and the subsequent formation of Burkitt's lymphoma in humans and plasmacytomas in mice (Robbiani et al., 2008). Other proto-oncogenes that also underwent translocation/mutation due to AID activity include BCL2/IgH in Follicular lymphoma (FL), BCL-6/IgH in diffuse B-cell lymphoma (DLBCL), PAX-5 (Paired box 5) in small lymphocytic lymphoma, FAS (Fas cell surface death receptor) and RHOH (Ras homolog family member H) in non-Hodgkin's lymphomas (Jiang et al., 2002; Kramer et al., 1998; Torlakovic et al., 2002; Pasqualucci et al., 2001). Furthermore, not only was AID shown to initiate somatic hypermutation in Chronic Lymphocytic Leukemia cells but highly expressed AID in patients with U-CLL more specifically was also found to be directly linked to increased class switching and cell

proliferation levels (Palacios et al., 2010).

1.5 AID targeting mechanisms

When AID was first discovered, it was proposed to act on RNA, like its homologue APOBEC1(Muramatsu et al., 1999; Muramatsu et al., 2000; Muramatsu et al., 2007). However, subsequent evidence suggested that AID mutates single-stranded DNA (ssDNA). First, although purified AID, like all other AID/APOBEC enzymes can bind RNA as well as ssDNA, it strictly mutates (ssDNA) but has no apparent activity on RNA or double stranded DNA (dsDNA) oligonucleotides (Bransteitter et al., 2003; Sohail el al., 2003; Dickerson et al., 2003; Pham et al., 2003; Larijani et al., 2007; Yu et al., 2004; Larijani et al., 2007; Nabel et al., 2013). Second, AID is physically found at the switch region (S-region) of the Ig loci in B-cells undergoing CSR (Chaudhuri et al., 2007; Ranjit et al., 2011). Third, there is a strong association between AID and dU in DNA at loci undergoing SHM and CSR (Maul et al., 2011; Di Noia et al., 2002; Rada et al., 2002; Rada et al., 2004; Di Noia et al., 2007). Fourth, the WRC (W= A/T, R=A/G) sequence preference pattern of purified AID on ssDNA *in vitro*, or of AID-mediated mutations in the DNA genomes of bacteria, yeast or human cell lines made to express exogenous AID, are identical to that of SHM *in vivo* (Rada et al., 2004; Petersen-Mahrt et al., 2002; Larijani et al., 2007).

On the other hand, there has also been a significant amount of evidence in support of RNA

targeting by AID, namely that AID-dependent targeting of Ig loci is associated with certain nonconventional and non-catalytic aspects of uracil glycosylase activity (Begum et al., 2004; Yousif et al., 2014; Yousif et al., 2014; Begum et al., 2009), and that CSR requires *de novo* protein synthesis after AID expression (Begum et al., 2004). In addition, AID robustly binds RNA (Dickerson et al., 2003), can mutate viral RNA (Liang et al., 2013), and APOBEC-1, which is known to be an RNA-editing enzyme, also has the capacity to mutate DNA when expressed exogenously (Petersen-Mahrt et al., 2003). Based on these lines of evidence, it has been argued that SHM and CSR are not directly mediated by AID mutating Ig genes, but are rather by repair factors/polymerases and other DNA cleaving agents whose expression is modulated through RNA editing by AID (Begum et al., 2012).

The DNA targeting model is also consistent with the finding that the absolute requirement for mutation of a gene by AID is active transcription, and that highly transcribed loci are preferred targets of mutation by AID (Peters et al., 1996; Storb et al., 1998; Storb et al., 1998; Fukita et al 1998), since transcription generates and liberates ssDNA. For instance, at GC-rich sequences such as the Ig switch regions, the newly synthesized RNA transcript can associate with the DNA template strand creating a transient non-template strand R loop. Transcription also generates stem loops (SL), cruciforms and bubbles, due to relaxation of supercoiling in the wake of the traversing RNA polymerase (Dayn et al., 1992; Krasilnikov et al 1999). Indeed, such transient ssDNA structures, such as R-loops have been detected

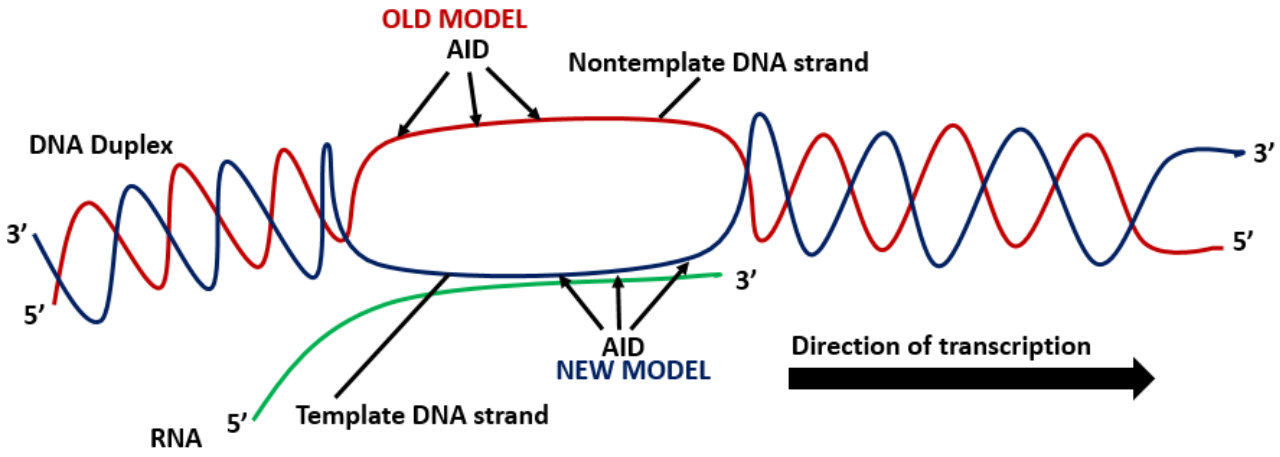
abundantly at Ig switch sequences, where AID- induced mutations lead to double strand breaks (DSB) that initiate CSR (Yu et al., 2003; Yu et al., 2005; Stavnezer, 2011; Wright et al., 2004; Huang et al., 2007). We and others showed that these ssDNA substrates simulating these structures are efficiently mutated by AID in *in vitro* enzyme assays, or if made available through *in vitro* transcription of dsDNA (Yu et al., 2004; Larijani et al., 2007; Besmer et al., 2006; Shem et al., 2009; Canugovi et al., 2009; Storb, 2014).

It has also been suggested that AID interacts with various components of the transcriptional machinery including RNA polymerase II, Spt5, RNA polymerase II associated factor I (PAF1), spliceosome-associated factor CTNNBL1, RNA binding heterogeneous nuclear ribonucleoproteins (hnRNP) and splicing regulator polypyrimidine tract binding protein 2 (PTBP2), and many of these interactions are dependent on the presence of RNA itself (Basu et al., 2011; Conticello et al., 2008; Pefanis and Basu, 2015; Pefanis et al., 2014; Hu et al., 2015; Mondal et al., 2016; Nowak et al., 2011; Pavri et al., 2010; Willmann et al., 2012; Taylor et al., 2014; DiMenna and Chaudhuri, 2016; Zheng et al., 2015). Thus, both the process of transcription as well as interactions between AID and the transcription complex involve RNA (Fig. 5).

1.6 Hypothesis

Here, we examined the behavior of purified AID on DNA/RNA hybrid substrates designed to simulate structures found at immunoglobulin loci. We hypothesized that beyond

facilitating recruitment of AID, RNA may play a direct role in modulating its catalytic activity. To address this, we studied the activity of purified AID on DNA substrates bearing various sequences, including Ig region switch region sequences, in the presence or absence of hybridized RNA. We found that RNA can significantly boost the enzymatic activity of AID on ssDNA. We observed specificity, both in regards to DNA/RNA hybrid sequences able to elicit this effect, and surface residues of AID involved in this recognition.



(modified from Molecular Biology: Principles and Practice, 2012)

Figure 5. RNA might play an important role in AID recruitment

Earlier research suggests that the availability of ssDNA released during transcription attracts AID action since AID only mutates DNA. However, not only has AID been shown to bind RNA but defects in RNA processing during transcription contribute to altered AID activity. This suggests that while RNA itself is not a substrate for AID, it may play a role in AID targeting to the Ig locus.

Materials and Methods

2.1 AID expression and purification

Several types of highly purified AID were used in this study including bacterially expressed glutathione s-transferase (GST) tagged AID and eukaryotically expressed GST-AID or AID- His. We have previously described the expression and purification of GST-AID from bacteria (Dancyger et al., 2012; Abdouni et al., 2013; Larijani et al., 2007). Briefly, 50µl *Escherichia coli* BL21 (DE3) cells were transformed with 5µl of sequenced AID plasmid harboring the expression vector pGEX5.3 (GE Healthcare, USA) with GST-AID and incubated on ice for 20 min followed by heat shock for 45sec at 42°C. 200 µl of LB broth was supplemented before incubation at 37 °C and 225 rpm for 1 hr. 250 µl of the culture was plated on LB-ampicillin and incubated at 37 °C overnight. One colony of the glycerol stock of the transformed DE3 cells was picked to inoculate a 250 mL LB-ampicillin (100 ~g/mL) culture and then incubated at 37 °C and 225 rpm until the culture reached the log phase of growth and induced to express AID with 250 µl 1 mM IPTG at 16 °C for 16 hrs. The culture was then pelleted using a Sorval Evolution RC Ultra Centrifuge at 4°C, SOOO rpm for 10 min and the pellet was re-suspended in 20 mL of cold 1 X Phosphate buffered saline (PBS) and kept on ice. Cells were lysed twice using the French pressure cell press (Thermospectronic) and rinsed with cold 1 X PBS. The lysate was ultra-centrifuged at 8000 rpm and 4 °C for 10 min to remove cell waste and applied to Glutathione sepharose beads (Amersham) and AID protein was eluted as per the manufacturer's recommendations. Elution

fractions were collected at 4°C in 10 aliquots with 500 µl volumes each and kept on ice. Protein concentration and purity of each protein was tested using nanodrop spectrophotometry (Thermo Scientific) at 260 and 280 nm wavelengths. Protein fractions exceeding 1 mg/mL were dialysed using Snakeskin R Pleated Dialysis Tubing (Thermo Scientific) in Dialysis Buffer (1mM Dithiothreitol, 20mM Tris-HCl and 100 mM NaCl) twice at 4 °C. 50 µl of the purified protein was aliquoted in 1.5 mL micro centrifuge, flash frozen in liquid nitrogen for 30 sec and stored at -80 °C.

For eukaryotic expression of AID, an EcoRV/KpnI-flanked human AID fragment was cloned into pcDNA3.1-V5-6xHis- Topo which was modified with the addition of 2 extra His-residues to encode a 8x C-terminal His-tag. 50 x 10 cm plates, each seeded with 5 $x10^5$ HEK 293T cells were transfected with 5 µg of plasmid per plate using Polyjet transfection reagent (Froggabio). Cells were incubated 48 hours at 37 °C, followed by resuspension in the lysis buffer 50 mM phosphate buffer pH 8.2 + 500 mM NaCl, 0.2 mM PMSF, 50 µg/ml RNAse A and lysis using a French pressure cell press and batch binding on nickel sepharose beads (Amersham) a per manufacturer's recommendations. Beads were serially washed with lysis buffer containing increasing concentrations (1 mM and 30 mM) of Imidazole, and eluted in lysis buffer containing 500 mM of Imidazole. Alternatively, washed beads with bound AID-His were stored in AID storage buffer as a source of bead-bound AID. Eukaryotic GST-AID was likewise produced through the aforementioned transient transfection protocol with a pcDNA3.1-V5-6xHis-Topo

vector into which we cloned a previously described GST-AID encoding fragment (Larijani et al., 2007). Cells were resuspended in PBS and lysed using a French Pressure cell, followed by binding to Glutathione Sepharose high performance beads (Amersham) as per manufacturer's recommendations. Beads were washed with PBS and resuspended in AID storage buffer.

All purifications were subject to analysis by coomassie staining to test for protein concentration, purity, and expression levels using comparison to Bovine Serum Albumin (BSA) standards. Briefly, purified GST -AID proteins were run on 8% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) against BSA titrations followed by coomassie staining. Gels were stored by first drying onto chromatography paper (Whatman) using a slab dryer, at 80 °C for 2 hrs.

To verify the relative yield and purity of eukaryotic expression of AID, western blotting was used. Western blots were probed with anti-V5 (Abcam) at a 1:5000 dilution or anti-GST (SantaCruz) antibodies at a 1:500 dilution, followed by the secondary detection by Goat anti-Rabbit IgG (SantaCruz) at a 1:5000 dilution. Site directed mutagenesis was used to generate GST-AID mutants. Briefly, the GST-fusion expression construct Pgex-5x-3 containing human AID gene (50ng/µl) was PCR-amplified using Phusion® High-Fidelity DNA Polymerase (2 units/µl) (NEB) and forward and reverse primers (62.5ng/µl for each) that contained the indicated point mutations. Samples were incubated at 96°C for 1 min followed 35 cycles of 96°C for 30 sec, 58°C for 30 sec and 72°C for 5 min and ending with 72°C for 10 min. To eliminate the

original template that was subsequently replicated, the hemi-methylated DNA was cleaved with the addition of DPN1 (20 units/ul) and incubating at 37°C for 4 hours followed by an overnight hold at 15°C. The PCR products were then dialyzed on millipore disks (0.025 um) in autoclaved water for 20 min, then reduced to 10ul in a heated speed vac machine. This was followed by transformation into XL blue *E.coli* bacteria, and sequence verification of correct clones. Two correct clones were then transformed into DE3 *E.coli* bacteria for protein production in the same manner described as wild type bacterially expressed GST-AID. For each form of AID (GST-AID, AID-His or GST-AID mutants), 2-6 independently purified preparations were used in this study.

2.2 Preparation of substrates

Partially single-stranded bubble substrates for AID enzyme assays were prepared as described (Larijani et al., 2007; Larijani and Martin, 2007). Briefly, DNA and RNA oligonucleotides were synthesized and subject to FPLC purification (IDT). All oligonucleotides used are listed in Table 2. 2.5 pmol of the target oligonucleotide was 5'-labeled with [γ-32P] dATP using polynucleotide kinase (New England Biolabs, USA). For simple bubble structures labelled, oligonucleotides were purified through mini-Quick spin DNA columns (Roche) and annealed with 3-fold excess cold strand (7.5 pmol) in a total volume of 50 µl by incubation starting at 96 °C and slow cooling at 1 °C/min to 6 °C. Triplex bubble structures were prepared by mixing the labeled strand, the cold strand of the same length and shorter complimentary strand at a ratio of 1:3:9 and incubation at

96 °C and slow cooling at 1 °C/30 sec to 6 °C. Following annealing reactions and throughout the experimental phase, the expected formation and integrity of substrates were verified by electrophoresis on native (8-10%) polyacrylamide gels (6 % Glycerol, 8 % acrylamide:bisacrylamide 19:1) to confirm size and shape, as well as by MBN and RNase mapping of single-stranded regions. For MBN digestion, 50 fmol substrates were incubated with 0.1 or 1 units of MBN in 10 µl containing MBN buffer (New England Biolabs) at 37 °C for 5 min followed by snap cooling and addition of 10 µl of 0.5 EDTA Na^+ pH 8.0 and 2 µl of 0.1% SDS prior to electrophoresis on a 14% denaturing gel. For RNase digests, 2 µl of RNase A (1mg/ml) was incubated with 50 fmol substrate in 20 µl in the reaction buffer (50 mM Tris, pH 7.5; 50 mM NaCl) at 37 °C for one hour, prior to electrophoresis on an 8% native polyacrylamide gel. Gels were exposed to a Kodak Storage Phosphor Screen GP (Bio-Rad, Hercules, CA, USA) and visualized by PhosphorImager (Bio-Rad, Hercules, CA, USA) using Quantity One software (Bio-Rad, Hercules, CA, USA).

2.3 Alkaline cleavage deamination assay for AID activity

We and others have previously described the alkaline cleavage assay for measuring the activity and catalytic rate kinetics of AID on various oligonucleotide-based substrates (Dancyger et al., 2012; Abdouni et al., 2013; Larijani et al., 2007). Briefly, 5 nM radioisotope labelled substrates were incubated with 1-300 ng AID (depending on expression system for obtaining AID) in an Eppendorf Mastercycler Epigradient PCR thermal cycler (Fisher Scientific, Ontario, Canada) for

time periods ranging from 60 min to several hours in 100 mM phosphate buffer pH 7.21. Enzyme kinetics were performed using a range of substrate concentrations including 7.5, 5, 4, 2.5, 1.25, 0.625, 0.315 and 0.15 nM. The 10 µl reaction volume was doubled by the addition of Uracil DNA glycosylase and buffer (NEB) for 30 min at 37 °C to excise the dU. 2 µl of 2 mM NaOH was added to the reaction followed by 8μl of formamide-loading dye solution (95% formamide, 0.25% Bromophenol Blue) and incubated at 96 °C for 10 mins to cleave the abasic site. The gels were electrophoresed using 14% denaturing urea-formamide-acrylamide gel (1X TBE, 25% formamide, 14% acrylamide:bisacrylamide, 7M urea) at 350V for 4.5-8 hours in TE buffer. This was followed by exposure to a Kodak Storage Phosphor Screen GP (Carestream Health Inc., Rochester, NY, USA) overnight and visualization by Phosphorimager Scanner (Bio-Rad). The gels were quantified using Image Lab (Bio-Rad).

2.4 Transcription-associated AID activity assay

The plasmid construct used as substrate is pcDNA3.1 V5-6xHIS containing the human APOBEC3DE gene under T7 promoter control. Plasmid was isolated by maxiprep (Qiagen), followed by purification of the supercoiled fraction through Cesium chloride gradient centrifugation. The plasmid was transcribed *in vitro* using either the MEGAscript T7 Transcription Kit (Ambion) or purified T7 RNA polymerase (New England Biolabs). For MEGA script T7 Transcription Kit, 0.25 pmol of supercoiled plasmid was transcribed in 1x transcription buffer, 2 µl of enzyme mix in 3.75 mM rNTPs. Reactions were incubated at 32 °C for 30 min. For

reactions using purified T7 RNA polymerase, 0.25 pmol of supercoiled plasmid was transcribed in the transcription buffer (40 mM Tris-HCl, 6 mM MgCl2, 2 mM spermidine, 1 mM dithiothreitol, pH 7.9) at 25 °C using 100 units of T7 polymerase in presence of 0.5 mM rNTPs (final volume of 20µl). Where added, 1.5 µg of human GST-AID, 10 unit of RNase H (Invitrogen,), and 2 µg of RNase A (sigma) were used in the transcription reactions. To detect AID-mediated mutations, 1µl of each reaction was amplified by deamination-specific nested PCR using Taq DNA polymerase and mutation-specific primers, as previously described (Larijani et al., 2005; Larijani et al; 2005). PCR products were analysed on an 1% agarose gel and band intensity was quantified using Image J 1.50i software. Data were submitted to independent samples T-test using IBM SPSS Statistics version 20 software. To confirm AID mediated mutations, we used PCR with AT-rich primers that bind to wild-type as well as AID mutated sequences with equal preference and thus do not select for mutated or unmutated plasmids. The 1.2 kb-long PCR amplicons was subsequently TA-cloned, and 100-200 amplicons from each reaction were sequenced. AID-mutated mutations (C to T, or G to A on the sense and non-sense strands, respectively) were divided by total number of sequenced nucleotides to obtain the AID mediated mutation frequency.

2.5 Electrophoretic mobility shift assay for AID binding

Electromobility shift assay (EMSA) was used to measure the binding affinity of AID to substrates as previously described (Larijani et al., 2007; King et al., 2015). Briefly, a range of substrate

dilutions (7.5, 5, 4, 2.5, 1.25, 0.625, 0.315, 0.15, 0.075, 0.03 nM) were incubated with 1 μg of GST-AID in binding buffer (50 mM Tris-HCl pH 7.5; 2 μM MgCl; 50 mM NaCl; and 1 mM DTT) in a final volume of 10 μl incubated at 25 °C for 1 hour, followed by UV cross-linking on ice using the optimum setting of a UV-crosslinker (Stratagene) and rotated 180 degrees. EMSA loading dye (0.25 % Bromophenol Blue dye, 0.25 % Xylene Cyanol dye, 49.75 % Glycerol, and 49.75 % autoclaved MilliQ water) was added to each samples and reactions were electrophoresed at 4 °C on an 8% acrylamide native gel (6 % Glycerol, 8 % acrylamide:bisacrylamide 19:1), in 0.5X TBE buffer for 3 h, at 300 V and 4 °C. Gels were exposed to a Kodak Storage Phosphor Screen GP (Kodak Storage, BioRad) and visualized using a PhosphoImager Scanner (Molecular Dynamics). Described. K_d values were derived from relative amounts of bound *vs.* free substrate at each substrate concentration.

2.6 AID activity and binding assay data quantification and statistical analysis

Band densitometry was performed as previously described (Abdouni et al., 2013). Briefly, The Quantity One 1-D Analysis Software (Bio-Rad) was used to quantify band intensities of imaged gels. Typically, each experiment was repeated 4-10 times using 2-6 independently purified preparations of each type of purified AID. Quantification of each individual lane of an alkaline cleavage or EMSA gel was done 3 independent times to obtain an average value, representing a single-data point. Typically, triplicate measurements of each lane as well as duplicate reaction

lanes within each experiment exhibited a variation of <5%. Each graphed data point is thus the mean of all values obtained (12-30 values per data point) for a given AID:substrate combination. Data were graphed using GraphPad Prism software (GraphPad, San Diego, CA, USA). Error bars represent standard error. Paired T-test in GraphPad Prism was used to analyze data to scrutinize the significance of AID activity differences between each DNA/RNA hybrid and corresponding DNA/DNA substrate.

2.7 AID structure and AID:nucleic acid complex formation modeling

AID structure modeling and substrate docking was performed as described previously (King et al., 2015). Briefly, five APOBEC template structures were chosen as templates: mouse A2 NMR (PDB: 2RPZ), A3A NMR (PDB: 2M65), A3C X-ray (PDB: 3VOW), A3F-CTD X-ray (PDB: 4IOU) and A3G-CTD X-ray (PDB: 3E1U). All APOBEC templates were obtained from the PDB (http://www.rcsb.org) and visualized using PyMOL v1.7.6 (http://pymol.org). Using the default parameters of I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010; Zhang, 2008) full-length human AID (Hs-AID) were modeled from the APOBEC templates to generate 24 lowest-energy conformations. In each conformation, the entire structure was homology modeled except for the non-homologous 18 C terminal amino acids, which were modeled *ab initio*. ssDNA (5'-TTTTGCTT-3') and RNA (5'-AUGCAGC-3') were constructed in Marvin Sketch v.5.11.5 (http://www.chemaxon.com/products/marvin/marvinsketch/), while surface topology and docking parameters were generated using Swiss-Param

(http://swissparam.ch) (Zoete et al., 2011). These output files served as docking ligands. Docking was performed using AutoDock Vina (http://vina.scripps.edu/). 9 catalytically accessible conformations of AID, as described previously (King et al., 2015) were docked with ssDNA and RNA, generating 9 low-energy clusters for each conformation, totaling 81 clusters. Residues within 3 Å of docked ssDNA or RNA were considered surface contact residues.

Results

3.1 DNA/RNA hybrid bubble structures are stably formed and efficiently mutated by AID

We and others have previously shown that purified AID can deaminate ssDNA but not RNA or dsDNA oligonucleotides *in vitro* (Sohail et al., 2003; Dickerson et al., 2003; Pham et al., 2003; Larijani et al., 2007; Yu et al., 2004; Larijani et al., 2007; Nabel et al., 2013). However, the activity of AID on DNA/RNA bubbles which simulate transient breathing of hybrid regions such as R loops, has not been examined. To investigate the role of RNA on AID activity, we designed oligonucleotide substrates in which a given 5'-labeled DNA oligonucleotide DNA sequence was used as ssDNA, or annealed to fully or partially complementary DNA or RNA strands to form dsDNA, a DNA/DNA bubble (TGCbub7-DNA/DNA) or a DNA/RNA hybrid bubble (TGCbub7- DNA/RNA) (Fig. 6a). We chose these initial substrates, since we have previously shown that TGCbub7DNA/DNA is a favoured substrate of human AID in the alkaline cleavage assay, amongst a comprehensive panel of ssDNA shapes and sequences (Larijani et al., 2007).

To confirm that DNA/RNA bubbles of the same sequence form similar structures as DNA/DNA bubbles we subjected substrates to native electrophoresis (Fig. 6b). We observed that DNA/DNA and DNA/RNA bubbles migrated slower than dsDNA. Since these fully dsDNA and DNA/DNA bubbles have near identical molecular weights, the slower migration of the

bubble *vs.* the fully ds DNA confirms the additional size footprint of the bubble region. In addition, DNA/RNA bubbles migrated slightly slower than their pure DNA counterparts of the same length and shape, as expected due to heavier weight of RNA (Table 2). To examine stability in the presence of AID, we incubated ssDNA, DNA/DNA and DNA/RNA bubbles with purified AID and observed that DNA/RNA bubble structures remained intact (Fig 6c). We further verified the expected structures of DNA/RNA bubbles by RNase and mung bean nuclease (MBN) digestions which degrade the RNA and only the ssDNA in the bubble regions, respectively (Fig 7). We conclude that DNA/RNA hybrid bubbles are formed in a similar manner to DNA/DNA bubbles, and are stable in the presence of AID.

We then compared the activity of AID on DNA/DNA bubbles *vs.* DNA/RNA hybrids, with ssDNA and dsDNA substrates as controls. As expected, AID deaminated multiple dC located along the ssDNA target whereas activity was not detected on dsDNA (Fig. 8a). We found bacterially-expressed GST-AID can efficiently deaminate both DNA/DNA bubbles and DNA/RNA hybrids (Fig. 8b left panel, 53 vs. 22%). To ensure that activity on DNA/RNA hybrids is a *bona fide* property of highly purified AID, we tested AID purified in a different system using a different fusion tag. As shown in Fig 8b, we observed that no matter if AID was expressed in bacteria *vs.* eukaryotic cells, or purified using N terminal GST vs. C-terminal His tag, it mutated DNA/RNA bubbles efficiently (16 vs. 5.4%, 43 versus 35% and 8.2 versus 10%). Based on these data we conclude that AID can efficiently mutate DNA/RNA bubbles.

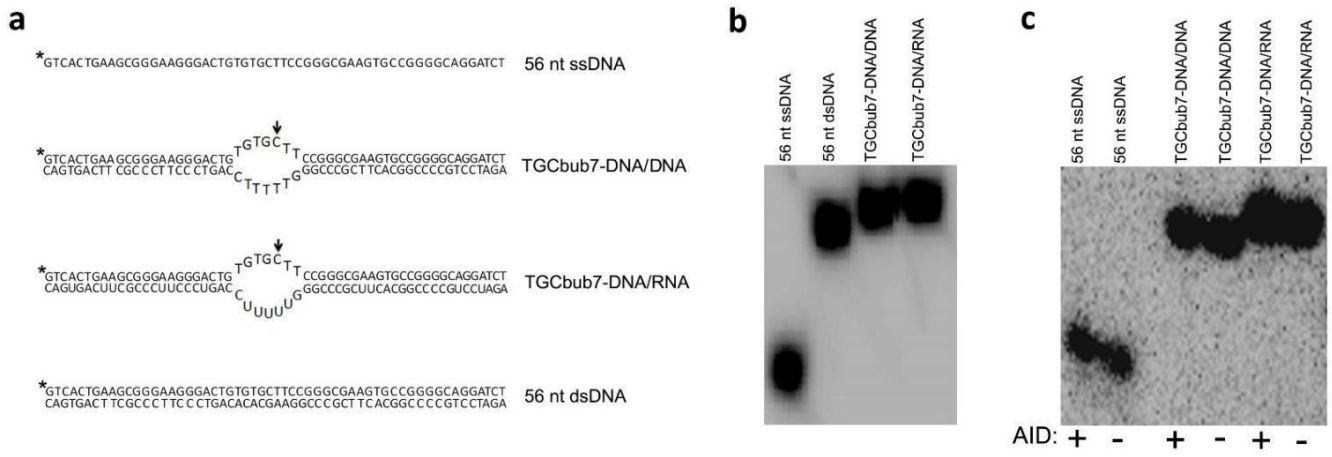


Figure 6. DNA/RNA hybrid bubbles are stably formed.

(a) Substrates designed to test AID activity on DNA/RNA hybrids include a 7 nt-long DNA/RNA bubble of random but pyrimidine-rich sequence and the equivalent DNA/DNA bubble, both containing the target dC (denoted by arrows) in the context of AID's favoured trinucleotide WRC motif TGC. All four substrates (ssDNA, bubbles and fully dsDNA) contain the same 5' labeled DNA target strand (denoted by *) annealed to either the equivalent sequence of partially complementary DNA or RNA, or a fully complementary DNA. **(b)** Native electrophoresis to compare migration of ssDNA, dsDNA, DNA/RNA bubbles and DNA/DNA bubbles (n=5) **(c)** Native electrophoresis of ssDNA, DNA/DNA and DNA/RNA bubbles to examine stability of DNA/RNA bubbles in the presence of AID. Substrates were incubated with purified AID followed by electrophoresis on native gels (n=5).

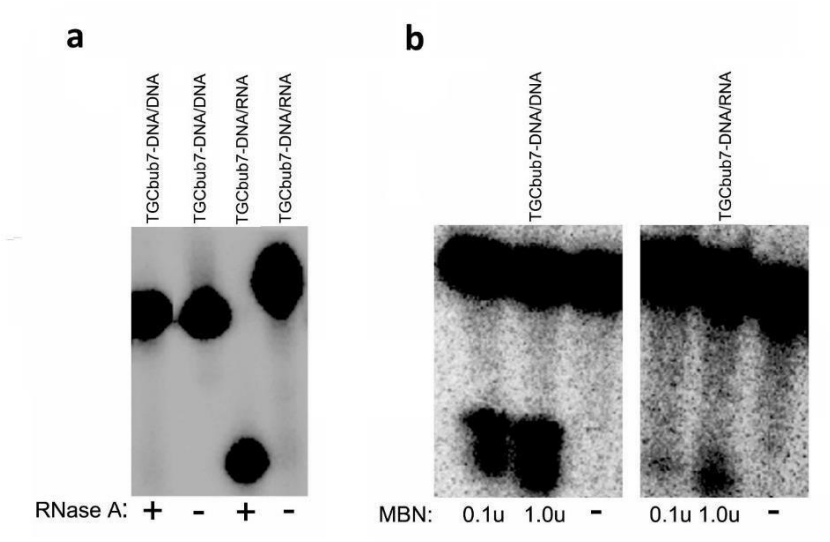


Figure 7. RNase and Mung Bean Nuclease mapping of DNA/DNA and DNA/RNA bubbles.

(a) RNase digest of DNA/RNA and DNA/DNA bubbles to verify the formation of DNA/RNA bubbles, followed by native electrophoresis resulted in degradation of the RNA strand in DNA/RNA bubbles, whilst the DNA/DNA bubble was unaffected. **(b)** Mung Bean Nuclease (MBN) digestion of ssDNA target regions of DNA/RNA and DNA/DNA bubbles, followed by denaturing electrophoresis yielded a ~ 23 (+/- 2) nt product band indicative of digestion up to the edge of the double-stranded region within the DNA and DNA/RNA bubbles (n=8).

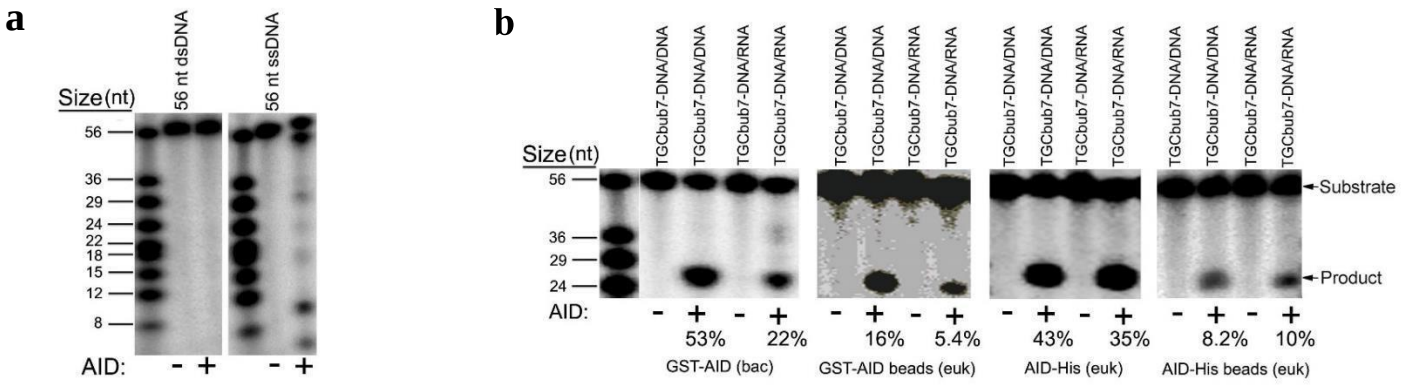


Figure 8. DNA/RNA hybrid bubbles are efficiently mutated by AID.

(a) Representative alkaline cleavage gel demonstrating that purified AID mutates ssDNA but not dsDNA. 50 fmol ssDNA or dsDNA were incubated with AID. **(b)** Purified AID can mutate DNA/RNA hybrid bubbles efficiently. To ensure that activity on DNA/RNA hybrids is a *bona fide* property of AID and not a function of a particular expression system or purification method, we tested highly purified AID that was expressed in either bacterial or eukaryotic cells and purified using either an N-terminal GST- or C-terminal His-tag. The four gels are representative alkaline cleavage experiments where 50 fmol TGCbub7-DNA/DNA or TGCbub7-DNA/RNA were incubated with bacterially expressed GST-AID (in solution), eukaryotically expressed GST-AID (bead-attached), eukaryotically expressed AID-His (in solution), and eukaryotically expressed AID-His (bead-attached) from left to right, respectively (n=3). Percent deamination for each reaction is shown below the lane.

Table 2. Molecular weights and sizes of DNA and RNA substrates

Substrate name	Oligonucleotide Molecular Weight	Total Molecular weight
56 nt ssDNA	DNA top strand (17458.3)	17458.3
56 nt dsDNA	DNA top strand (17458.3)	34461.28
	DNA Complementary bottom strand (17002.98)	
TGCbub7-DNA/DNA	DNA top strand (17458.3)	34430.3
	DNA bottom strand (16972)	
TGCbub7-DNA/RNA	DNA top strand (17458.3)	35125.8
	RNA bottom strand (17667.5)	
56 nt ssDNA (GC-rich)	DNA (GC-rich) top strand (17368.1)	17368.1
TGCbub7-DNA/DNA (GC-rich)	DNA (GC-rich) top strand (17368.1)	34605.2
	DNA (GC-rich) bottom strand (17237.1)	
TGCbub7-DNA/RNA (GC-rich)	DNA (GC-rich) top strand (17368.1)	35501.2
	RNA (GC-rich) bottom strand (18133.1)	
56 nt ssDNA (SR)	DNA (SR) top strand (16908.9)	16908.9
TGCbub7-DNA/DNA (SR)	DNA (SR) top strand (16908.9)	34601.3
	DNA (SR) bottom strand (17692.4)	
TGCbub7-DNA/RNA (SR)	DNA (SR) top strand (16908.9)	35371
	RNA (SR) bottom strand (18462.1)	
93 nt ssDNA	DNA top strand (29331.9)	29331.9
TGCbub45-DNA	DNA top strand (29331.9)	58075.5
	DNA bottom strand (28743.6)	
TGCbub45-DNA/DNA	DNA top strand (29331.9)	71606.4
	DNA bottom strand (28743.6)	
	Short DNA bottom strand (13530.9)	
TGCbub45-DNA/RNA	DNA top strand (29331.9)	72256.2
	DNA bottom strand (28743.6)	
	Short RNA bottom strand (14180.7)	

3.2 DNA/RNA hybrids boost catalytic activity of AID in a sequence dependentmanner

We compared catalytic kinetics of bacterially-expressed GST-AID and eukaryotically-expressed AID-His on DNA/RNA and DNA/DNA bubbles. Both versions of highly purified AID exhibited a modest (1.3-fold) preference at the highest measured initial deamination velocity, for TGCbub7-DNA/DNA over the TGCbub7-DNA/RNA (Fig. 9). DNA/DNA and DNA/RNA bubble substrates are composed of a random sequence with a GC content of 63%, but a pyrimidine- rich target bubble region. To examine the impact of sequence, we designed substrates with a higher GC content (80%) and purine-rich bubble region. In sharp contrast to the aforementioned set, both GST-AID and AID-His exhibited substantial preference at highest measured initial velocity (13 and 4-fold) for the DNA/RNA hybrid TGCbub7- DNA/RNA (GC-rich) over its DNA/DNA counterpart TGCbub7-DNA/DNA (GC-rich) which supported lower AID activity than the non- GC-rich DNA/DNA substrate TGCbub7-DNA/DNA (Fig. 10a). We then designed DNA/RNA hybrids incorporating a *bona fide* switch region (SR) sequence from the human Sμ locus (Mills et al., 1990). Similar to substrates with a random GC-rich sequence, AID activity was diminished on the DNA/DNA bubble TGCbub7-DNA/DNA (SR), but at maximal velocity, GST-AID and AID- His both were significantly more active (3.5- and 4-fold, respectively) on the DNA/RNA hybrid TGCbub7-DNA/RNA (SR) (Fig. 10b).

To ascertain whether these are reflective of recognition of DNA/RNA hybrids by AID,

we performed several additional experiments. First, we confirmed the same trend of AID preference on DNA/DNA *vs*. DNA/RNA bubbles on two additional versions of purified AID (Fig. 11). Second, to ensure that the preference of AID for DNA/RNA bubbles reflects recognition of the hybrid bubble itself rather than a boost in AID activity due to the presence of free RNA, we added excess free RNA to DNA/DNA bubbles but observed no change in AID activity (Fig. 12). We then designed longer substrates in which a third complementary strand of either DNA or RNA was used to form triplex structures (Fig. 13a). As with bubble substrates, we ensured proper formation of these by native electrophoresis (Fig. 13b). Both GST-AID and AID-His exhibited comparable catalytic kinetics on DNA/DNA/RNA and DNA/DNA/DNA triplex bubbles (Fig. 14). We conclude that RNA annealed to the target DNA strand can significantly boost AID activity in a sequence-dependent manner, but RNA that is free or annealed to complement of the target DNA strand does not have an appreciable effect.

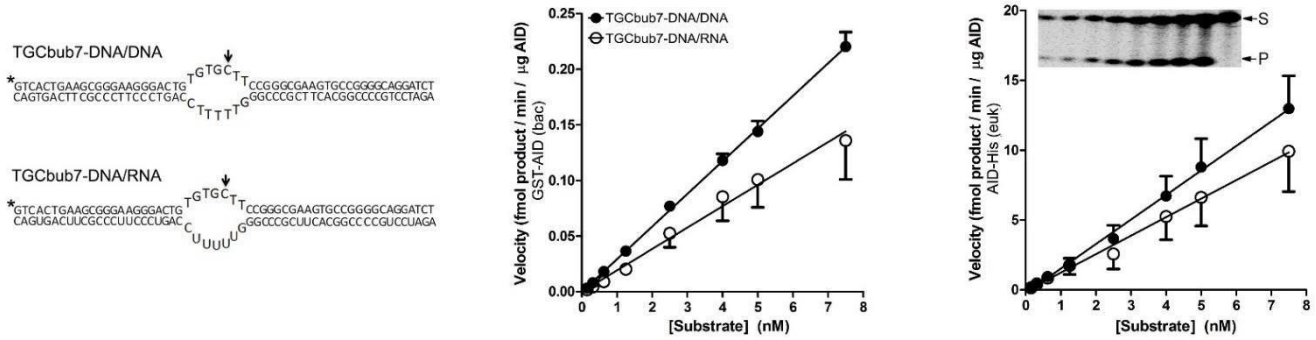


Figure 9. DNA/RNA hybrids is active on AID across different substrate concentrations DNA/DNA and DNA/RNA bubbles with a random sequence but pyrimidine-rich DNA target regions were incubated with bacterially expressed GST-AID (left panel) or eukaryotically expressed AID-His (right panel). 1µg GST-AID or 5 ng AID-His were incubated with substrate concentrations ranging 0.1 to 8 nM. A representative alkaline cleavage gel is shown as inset in the right panel. P values for TGCbub7-DNA/DNA *vs.* TGCbub7-DNA/RNA were 0.05 and not significant in the left and right panels respectively (n=6).

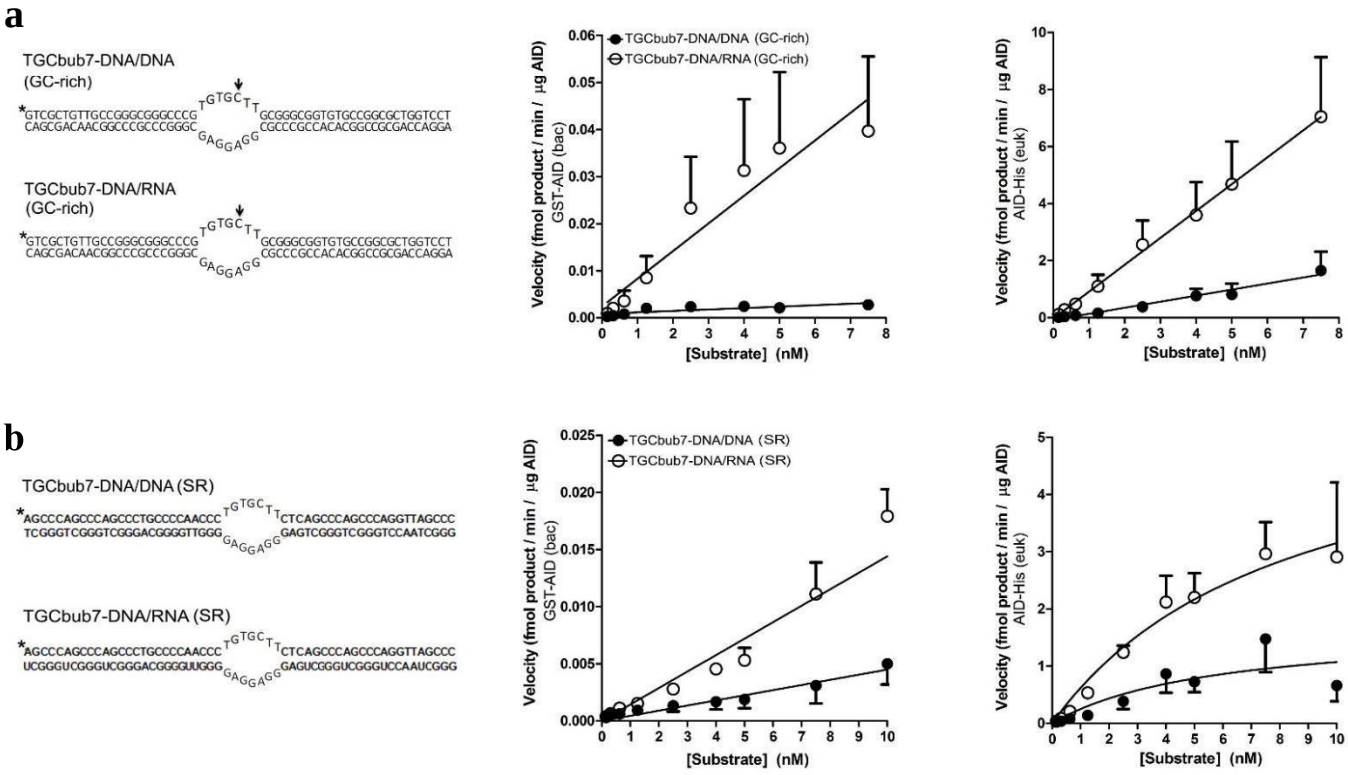


Figure 10. DNA/RNA hybrids enhance AID activity in a sequence dependent manner.

(a) Comparison of purified AID's catalytic rate on DNA/DNA and DNA/RNA bubbles containing a GC-rich sequence with a purine-rich bubble region. P values were 0.004 and 0.01 for the left and right panels, respectively. **(b)** Comparison of purified AID's catalytic rate on DNA/DNA and DNA/RNA bubbles containing human Sμ region (SR) sequences from. P values were <0.001 and 0.005 for the left and right panels, respectively.

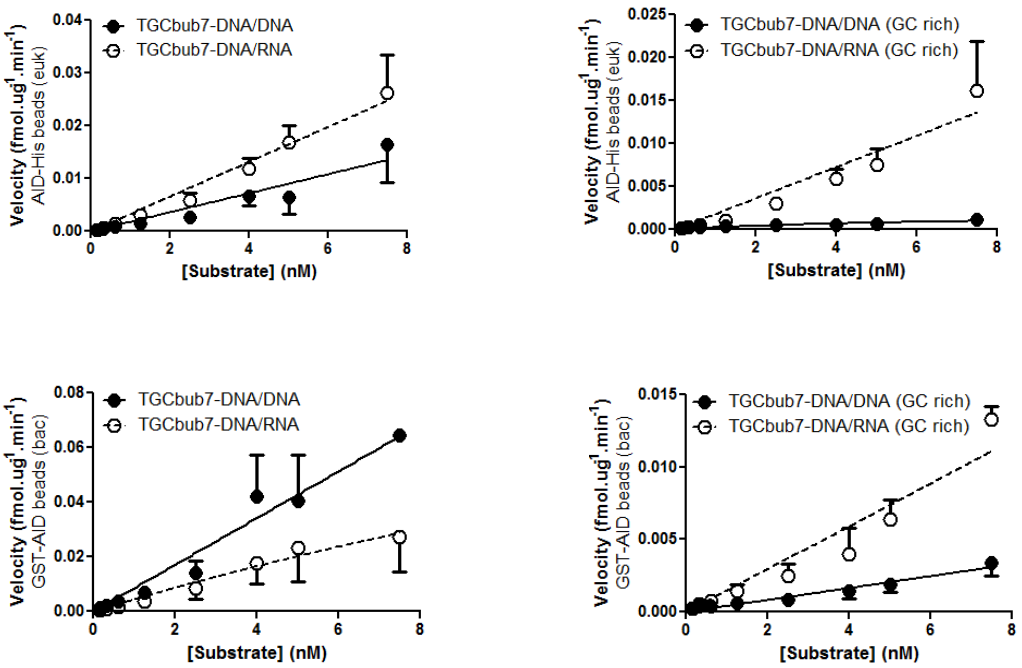


Figure 11. Two additional versions of purified AID also exhibit a similar pattern of relative catalytic kinetics on DNA/DNA and DNA/RNA bubbles.

DNA/DNA and DNA/RNA bubbles of random or GC-rich makeup were incubated with eukaryotically expressed AID-His that was bead-attached (top row) or with bacterially expressed GST-AID that was bead-attached (bottom row). 1µg equivalent of GST-AID beads or 5 ng equivalent of AID-His beads were incubated with substrate concentrations ranging 0.1 to 8 nM. P values were < 0.05 and 0.01, respectively, for TGCbub7-DNA/DNA *vs.* TGCbub7-DNA/RNA, and for TGCbub7-DNA (GC-rich) *vs.* TGCbub7-DNA/RNA (GC-rich).

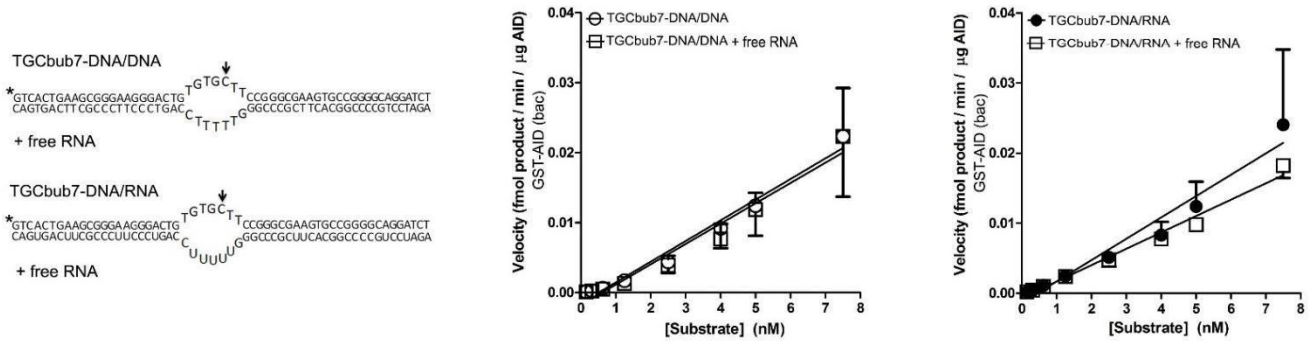


Figure 12. AID activity on DNA/RNA hybrids is not due to free RNA

Equivalent amounts of excess free RNA, as used in the generation of DNA/RNA hybrid substrates, were added to reactions containing AID and DNA/DNA or DNA/RNA hybrids to examine the impact of non-annealed RNA on AID activity.

Figure 13. ***In Vitro*** **R-loop model can be formed in a stable manner**

(a) Three substrates were used to measure the activity of purified AID on triplex bubble structures. A 93 nucleotide-long substrate containing a 45 nucleotide-long bubble region with the target dC (denoted by arrows) positioned in the WRC motif TGC is shown on top. Annealing of a 45 nucleotide long DNA (middle) or RNA (bottom) results in the generation of triplex DNA/DNA/DNA or DNA/DNA/RNA structures, with either DNA or RNA annealed to the complementary strand of AID's target strand. **(b)** Electrophoresis of substrates on native gels to confirm proper substrate formation as intended.

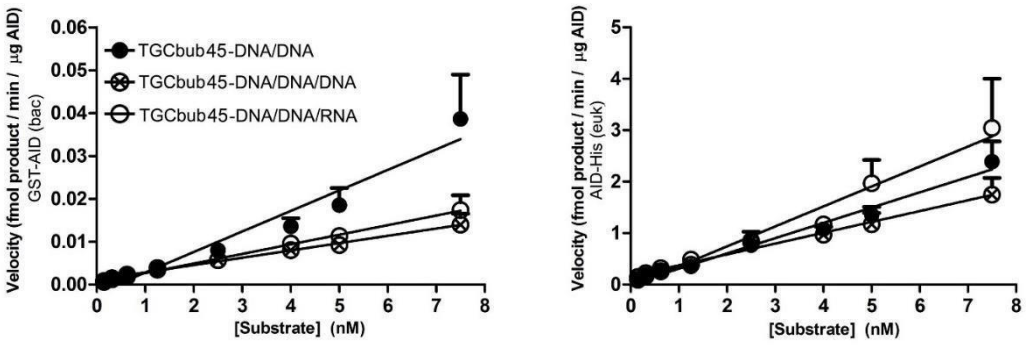


Figure 14. RNA-dependent enhancement of AID activity requires annealing to target DNA strand.

Comparison of the catalytic kinetics of GST AID (left panel) or AID-His (right panel), on DNA/DNA/DNA, DNA/DNA/RNA, and DNA/DNA triplex bubble structures (Not Significant).

3.3 DNA/RNA hybrids enhance transcription-associated AID activity

Since the generator of DNA/RNA hybrids is transcription, we sought to examine the role of DNA/RNA in modulating AID activity in the context of transcription. To this end, we employed a cell-free coupled transcription-deamination assay in which a purified supercoiled plasmid containing a T7 promoter is transcribed *in vitro* using T7 polymerase in the presence of AID (Fig. 15, left panel). This assay has previously been employed in several studies to model the transcription dependence of AID activity *in vitro* (Shen et al., 2009; Canugovi et al., 2009). Said studies embedded an AID-favored WRC motif within the stop codon of an antibiotic-resistance gene on a plasmid, and measured transcription-dependent AID activity by evaluating the number of antibiotic-resistant bacterial colonies after transformation with substrate plasmid. We used a deamination-specific PCR assay which can be used to reliably compare relative amounts of plasmid DNA mutated by purified AID (Larijani et al., 2005; Larijani and Martin, 2012). As expected, in the presence of transcription and AID we detected highly mutated plasmid DNA by PCR (Fig. 15, right panel). With the addition of RNase H which catalyzes the cleavage of RNA in RNA/DNA substrates, we observed on average a 6-fold reduction in abundance of mutated plasmid DNA (Fig. 15 and 16a). This RNase H mediated reduction was consistent whether transcription was carried out using an *in vitro* transcription kit (Fig. 15, top right panel), or highly purified T7 RNA polymerase (Fig. 15, bottom right panel). To further verify these results, we amplified the target region of the substrate DNA using degenerate AT-rich

primers that anneal to AID-mutated and unmutated substrates without preference. We observed that indeed the reactions containing RNase H yielded reduced frequency of AID-mediated mutations (Fig. 16b). Although this assay lacks the complexities of eukaryotic transcription, the finding that degradation of DNA/RNA hybrids diminishes transcription dependent AID activity is consistent with the biochemical data performed earlier with the RNase A digestion.

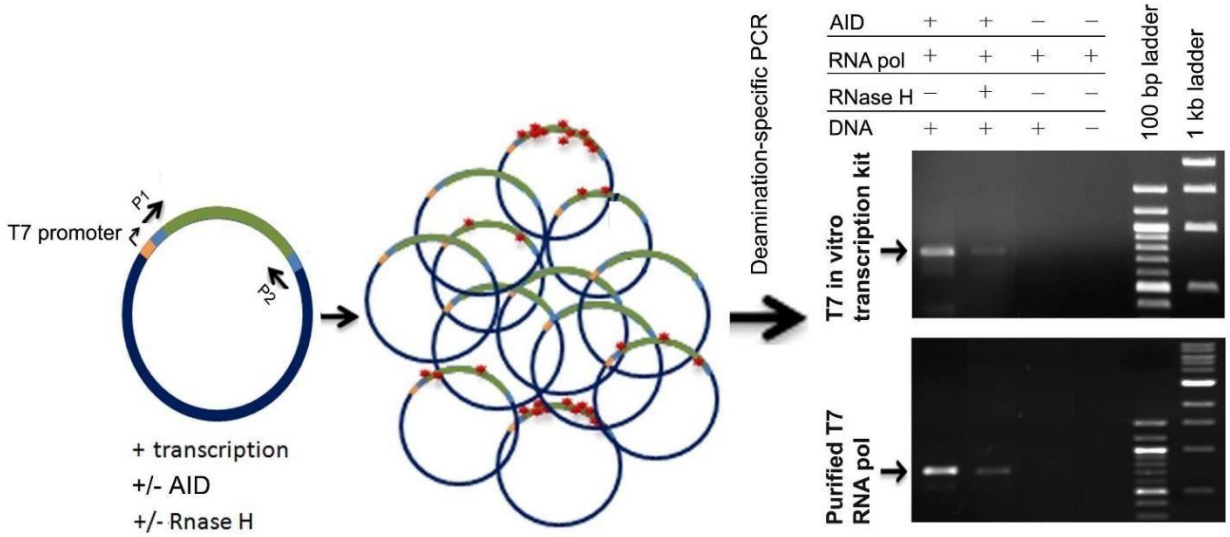


Figure 15. DNA/RNA hybrids enhance transcription-associated AID activity.

Schematic of the cell-free transcription-associated AID activity assay: a plasmid substrate containing the T7 promoter upstream of a 1.5 kb target sequence is transcribed *in vitro* in the presence of AID, with or without the addition of RNase A or RNase H. Some plasmids are mutated by AID (depicted as red dots) at different levels, whilst others are unmutated. The reactions are then subject to one of two types of PCR. The first type employs deamination-specific primers and thus selectively amplifies AID-mutated DNA; representative gels of this PCR assay are shown in the right panel. The second type of PCR employs degenerate AT-rich primers devoid of any selectivity for AID mutated or unmutated DNA, thus providing an unbiased measure of AID activity in the reaction.

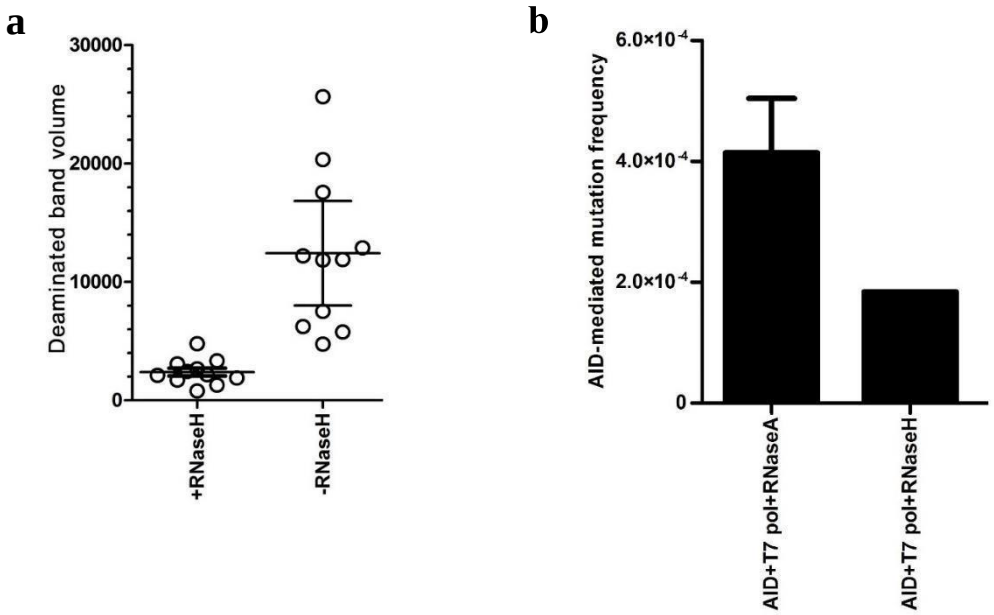


Figure 16. Presence of RNase H decreases AID-mediated mutations

(a) Quantification of relative amounts of AID-mutated DNA as a result of transcription and in the presence or absence of RNase H. Results are reflective of band intensity quantitation of multiple independent experiments, using either purified T7 polymerase or a T7 polymerase-based *in vitro* transcription system. 11 agarose gels from independent deamination-specific PCR assays were quantified, representatives of which are shown in figure 15. P-value = 0.0005. **(b)** Sequencing of non-selective PCR-amplified DNA from transcription-associated AID activity assays, in the presence or absence of RNase H. For each reaction ~200 cloned amplicons, each ~ 1 kb in length, were sequenced. AID-mediated mutations (C to T or G to A) were divided by total nucleotide coverage to obtain AID-mediated mutation frequencies. Results are the average of 3 independent experiments, with a total coverage of 430,196 nucleotides sequenced. P value was < 0.0001.

3.4 AID binds to DNA/RNA hybrids efficiently

We sought to determine whether differential activity of AID on DNA/RNA *vs.* DNA/DNA bubbles is reflective of ability to bind AID. We previously showed that AID binds ssDNA with high affinity in the nM range, and that AID's preference for binding ssDNA is independent of sequence but dependent on the structure shape, with optimal targets being bubbles with 5-7 nt-long ssDNA regions. We found that AID bound TGCbub7-DNA/RNA as efficiently as ssDNA, and with ~ 25- fold lower affinity than TGCbub7-DNA/DNA (Fig. 17a, average K_d values of 8.6 *vs.* 0.3 nM). In contrast, when we examined AID binding to GC-rich bubbles, we found the DNA/RNA hybrid bound AID with a slightly higher affinity as compared to the DNA/DNA bubble (Fig. 17b, average K_d values of 0.40 *vs.* 0.61 nM), and with ~ 20-fold higher affinity than the DNA/RNA bubble of lower GC content (average K_d values of 0.40 *vs.* 8.6 nM). One possible explanation for the difference in AID's affinity to DNA/RNA hybrids depending on GC content could be that GC-rich hybrids form duplex structures with different structural conformations; however, we did not observe evidence of this through perturbed migration in native gels in conditions sensitive enough to detect the change in shape caused by protrusion of a 7 nt-long bubble (Fig. 6b, Table 2, Fig. 18). We conclude that AID is capable of binding DNA/RNA hybrids at least as efficiently as ssDNA of the same length, but depending on the sequence even more efficiently than DNA/DNA bubbles. Furthermore, differences between binding affinities of AID to DNA/RNA hybrids of varying GC content correlate with relative

levels of AID activity on these substrates.

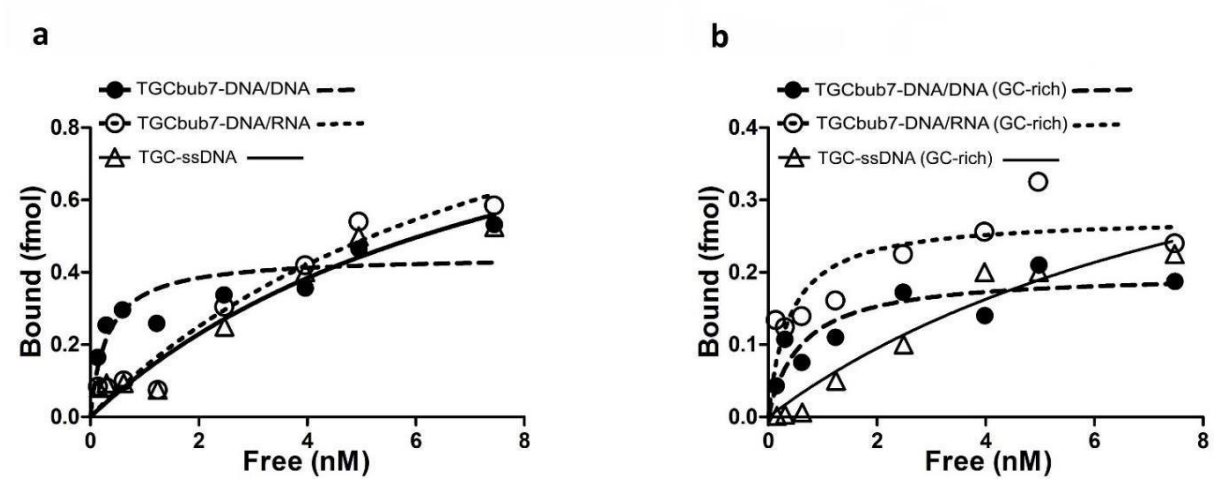


Figure 17. AID binds DNA/RNA hybrids efficiently.

(a) EMSA was performed to compare binding affinity of purified GST-AID to ssDNA, DNA/DNA or DNA/RNA bubbles. **(b)** EMSA was performed to compare binding affinity of purified GST- AID to ssDNA (GC-rich), TGCbub7-DNA/DNA (GC-rich), or DNA/RNA (GC-rich) bubbles. Binding reactions contained 1μg AID incubated with a substrate concentration range of 0.1 to 8 nM, followed by native electrophoresis to measure the relative proportions of free and AID-bound DNA substrate, from which Kd values were derived.

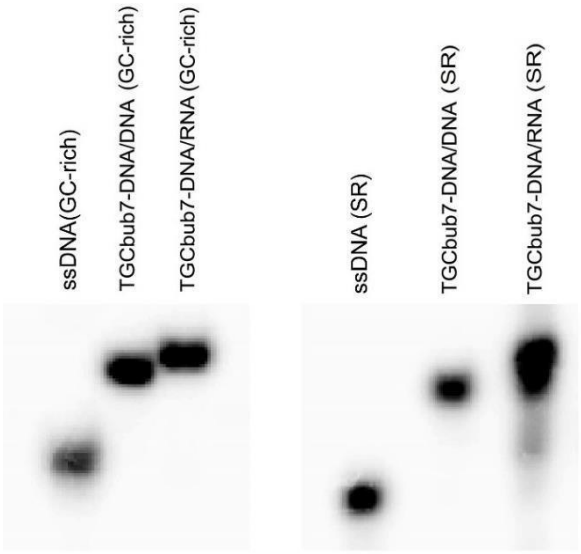


Figure 18. Native electrophoresis demonstrating proper formation of GC-rich and SR-set DNA/RNA hybrid bubbles.

Native electrophoresis to compare relative migration of the DNA/RNA bubble relative to DNA/DNA bubble and ssDNA of the same length, for the GC rich (left panel) and SR sets (right panel) of DNA/DNA and DNA/RNA bubbles.

3.5 Surface residue mutations of AID impact relative preference for DNA/RNA hybrids

Using a combination of *in silico* modeling supported by biochemical work we recently presented a possible functional structure for AID (King et al., 2015). We showed that due to its unusually highly positively charged surface, AID binds ssDNA sporadically across its surface with only a small proportion of AID:ssDNA binding events positioning the target ssDNA in one of two ssDNA binding grooves that pass over the catalytic pocket (Fig. 19a,b). We found that this sporadic binding of ssDNA on the surface, in addition to closure of AID's catalytic pocket, are responsible for AID's unusually low catalytic efficiency. To gain insight into binding of DNA/RNA hybrids, we simulated AID:RNA complexes in the same manner. We found that in contrast to ssDNA binding, RNA binding was more focused on a specific portion of AID's surface: 56% (45/81) of AID:RNA complexes bound RNA in a groove distal from the catalytic pocket (Fig. 19c), whilst 17% (14/81) bound in one of the ssDNA binding grooves and 23% (19/81) bound elsewhere on the surface. This putative RNA binding groove is composed of residues contributed by ⌡3, ⌡4, ⌡5, α3, α4, α5, loops 7 and 9 (Table 3). We noted that the distance between the putative RNA binding groove and ssDNA-binding groove 2 is such that DNA/RNA hybrid bubbles are predictedto bind the surface of AID using these two grooves (Fig. 20).

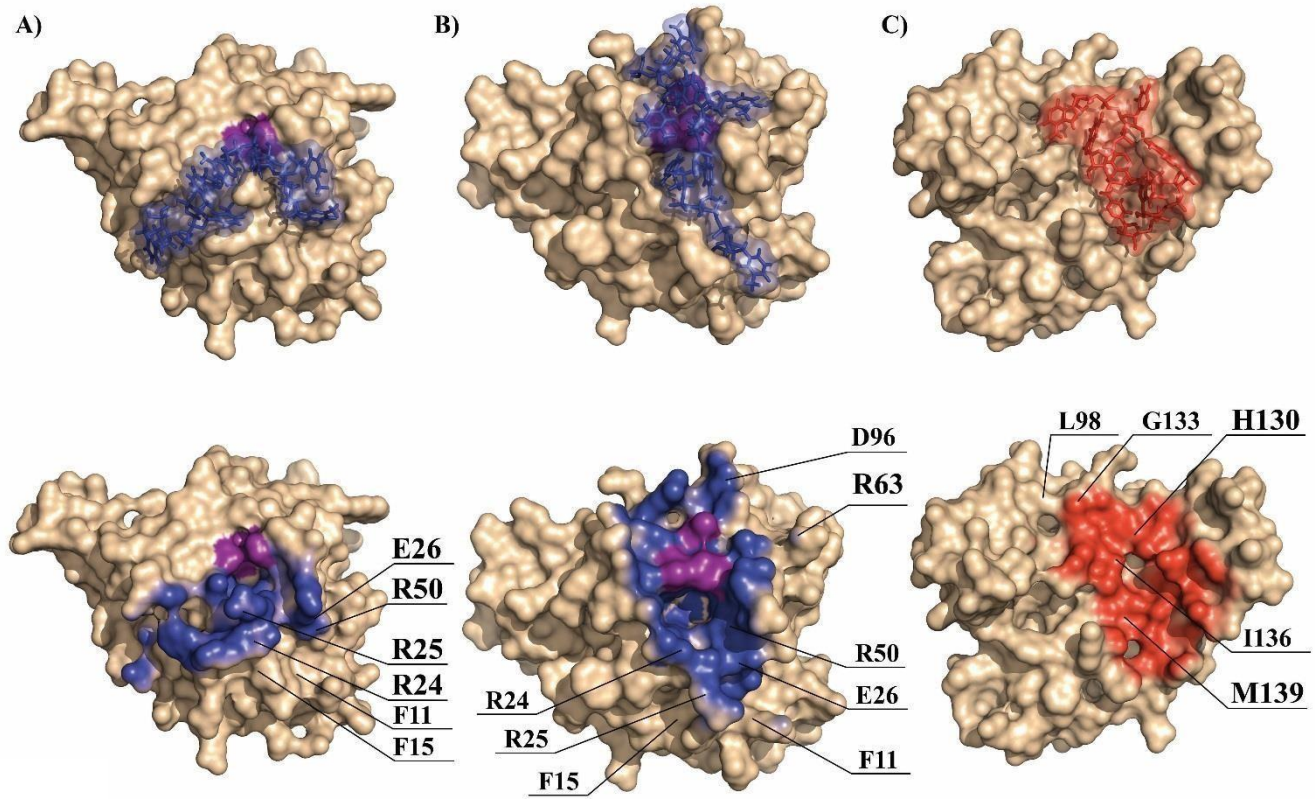


Figure 19. Simulated AID complexes with DNA, RNA or DNA/RNA hybrids.

Docking of ssDNA on the surface of AID illustrating the primary ssDNA binding groove **(a)**, the secondary ssDNA binding groove **(b)**, and a putative RNA binding groove **(c)**. Top panels are representative conformations of ssDNA or RNA bound to the respective groove. Bottom panels show residues lining each groove in blue (for ssDNA binding) or red (for RNA binding) without the bound ssDNA or RNA. Purple residues represent the three Zn-coordinating residues (H56, C87 and 90) and catalytic glutamic acid (E58), constituting the catalytic pocket. Key residues that determine relative efficacy of AID binding to DNA/RNA *vs.* DNA/DNA bubbles are labeled.

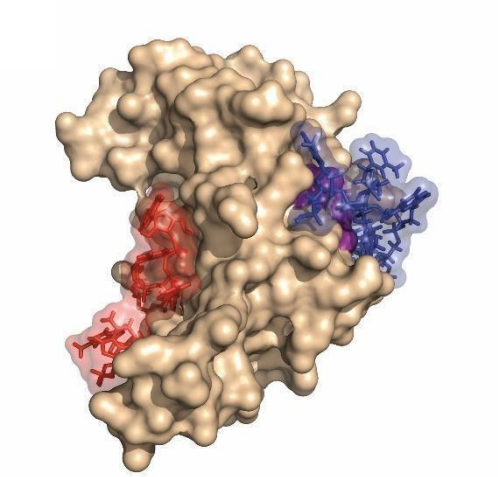


Figure 20. Prediction of DNA/RNA hybrids binding to the surface of AID

A probable conformation of a DNA/RNA hybrid bubble bound on the surface of AID. RNA in red is located in the RNA binding groove whilst the DNA strand in blue is bound along the secondary DNA binding groove due to the relative proximity of the two grooves.

Table 3. RNA contact frequencies of residues lining the putative RNA binding groove.

Residue	RNA binding frequency (%)
Q135	89
G133	82
H130	82
R107	76
S105	69
R77	67
R131	64
L104	62
L98	60
R177	60
T79	60
A132	58
F109	58
K142	58
P102	58
D143	56
I136	56
Y146	56
V134	53
A137	51
L106	51
T140	49
V32	49
R99	44
I138	40
L198	40
S173	40
L180	38
F81	36
Y184	36
G197	33
L176	29
P182	29
K34	27
L181	27
N101	27
T195	27
D187	22
E185	22
R36	22

We reasoned that the ssDNA binding grooves and putative RNA binding groove are likely to play a role in the recognition of DNA/RNA hybrids. To examine this notion, we tested a library of 41 AID mutants including mutants of the most probable ssDNA contact regions and Hyper IgM mutants of surface residues. Of these 41 mutants, 28 were residues from the DNA binding grooves, 11 were from the RNA binding groove, and 11 were Hyper IgM mutants (Mu et al., 2012; Durandy et al; 2006; Ouandani et al., 2016). We measured the relative activities of each mutant on the 6 different bubble substrates used in our study (DNA/DNA and DNA/RNA bubbles with random, GC-rich and SR sequence). Based on relative levels of activity on DNA/DNA *vs.* DNA/RNA bubbles, the 41 mutants tested fell broadly into 4 categories (Fig. 21). The most prominent, category I, contained 17/41 mutants that behaved generally like wild type (wt) AID in that they were active on both DNA/DNA and DNA/RNA bubbles but more active on the former. Category II contained 8/41 mutants which were quite weakly active (< 2% deamination or < 5% of wt) on both DNA/DNA and DNA/RNA bubbles. Category III contained 5/41 mutants that were active on both DNA/DNA and DNA/RNA bubbles, but exhibited significantly more activity on the latter, the opposite preference to that of wt AID. Category IV was the second largest and contained 11/41 mutants, which had weak to undetectable activity on DNA/DNA but retained activity on DNA/RNA hybrids.

Within the DNA binding grooves, we found 12 mutants (R25H, R25N, E26R, G54A, R63A, D89A, D89R, D96R, E117A, R119A, R171Y, R178D) in category I, 5 mutants (R50D,

A91K, R112C, K52D and R24D-R63D) in category II, 4 mutants (F11I, R24A, R50A and R63D) in category III, and 7 mutants (F15L, R24D, R24W, R25A, R25D, R25del and E26A) incategory IV. Within the RNA binding groove, we found 4 mutants (R107A, F109A, Q135A and R190X) in category I, 2 mutants (L106P and D143A) in category II and 5 mutants (L98R, H130P, G133V, I136K and M139T) in category IV. 3/11 Hyper IgM mutants (L106P, R112C and G125E) were nearly inactive (category II). Interestingly, 7/8 enzymatically active Hyper IgM mutants (F15L, R24W, L98R, H130P, G133V, I136K and M139T) fell into category IV. The only Hyper IgM mutant that did not ablate activity on DNA/DNA bubbles was R190X. We confirmed the specific preference of category IV mutants for DNA/RNA over DNA/DNA bubbles by carrying out catalytic kinetics (Fig. 22). We also tested each mutant on the GC-rich and SR-set of DNA/DNA and DNA/RNA bubbles, and observed similar results (Fig. 23, 24 and 25). Interestingly, the sole Hyper IgM mutant (R190X) which fell into category I on the DNA/DNA and DNA/RNA substrates of random sequence, also fell into category 4 when tested on the GC-rich and SR-set. Taken together, since the putative RNA binding groove and the two ssDNA-binding grooves house most residues whose mutations perturbed AID activity on DNA/DNA relative to DNA/RNA hybrids, we conclude that specific residues on AID's surface mediate RNA recognition thus playing a role in targeting DNA/RNA hybrids.

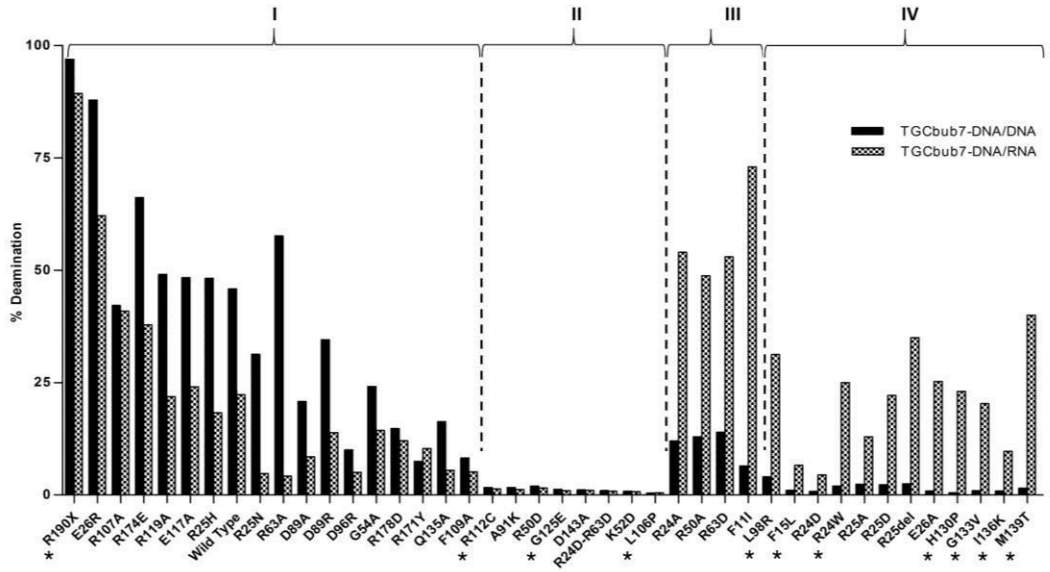


Figure 21. Surface residue and Hyper IgM mutants of AID differentially impact activity on DNA/DNA and DNA/RNA bubbles.

A panel of 41 AID mutants were tested for activity on TGCbub7-DNA/DNA and TGCbub7-DNA/RNA to determine relative preference. Mutants included mutations of surface residues located proximal to, or within one of AID's two putative ssDNA binding grooves or putative RNA binding groove, as well as known Hyper IgM mutants of surfaceresidues (indicated by the *).

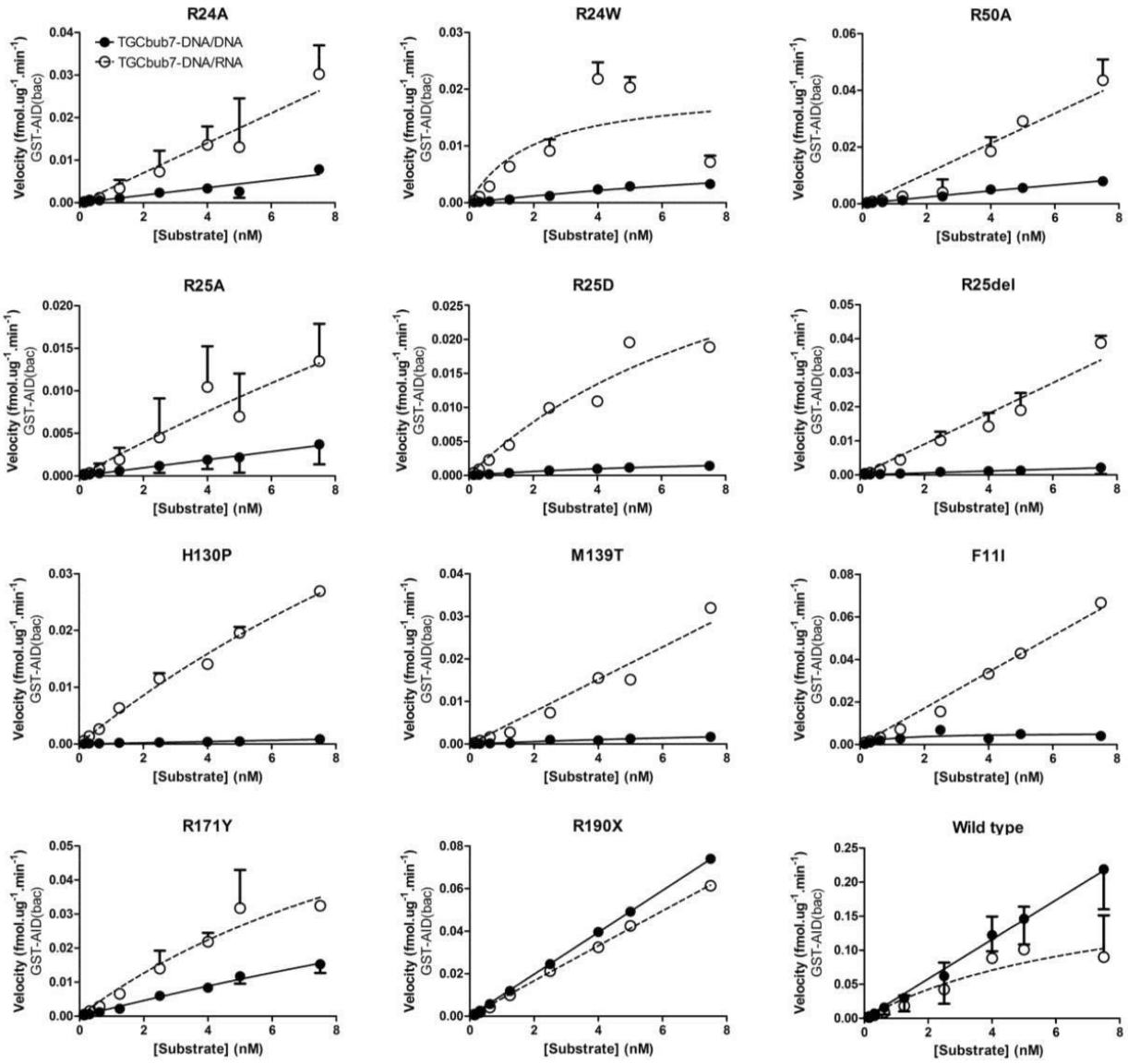


Figure 22. Enzymatic kinetics of surface residue and Hyper IgM mutants of AID on DNA/RNA hybrids

Catalytic rates of selected category IV mutants on TGCbub7-DNA/DNA and TGCbub7-DNA/RNA, to confirm the significantly enhanced preference for the latter. A range of 0.1-8nM substrate concentration were incubated with 1μg GST-AID. P values for all shown graphs were <0.05. For each mutant, two independently purified preparations were tested.

a

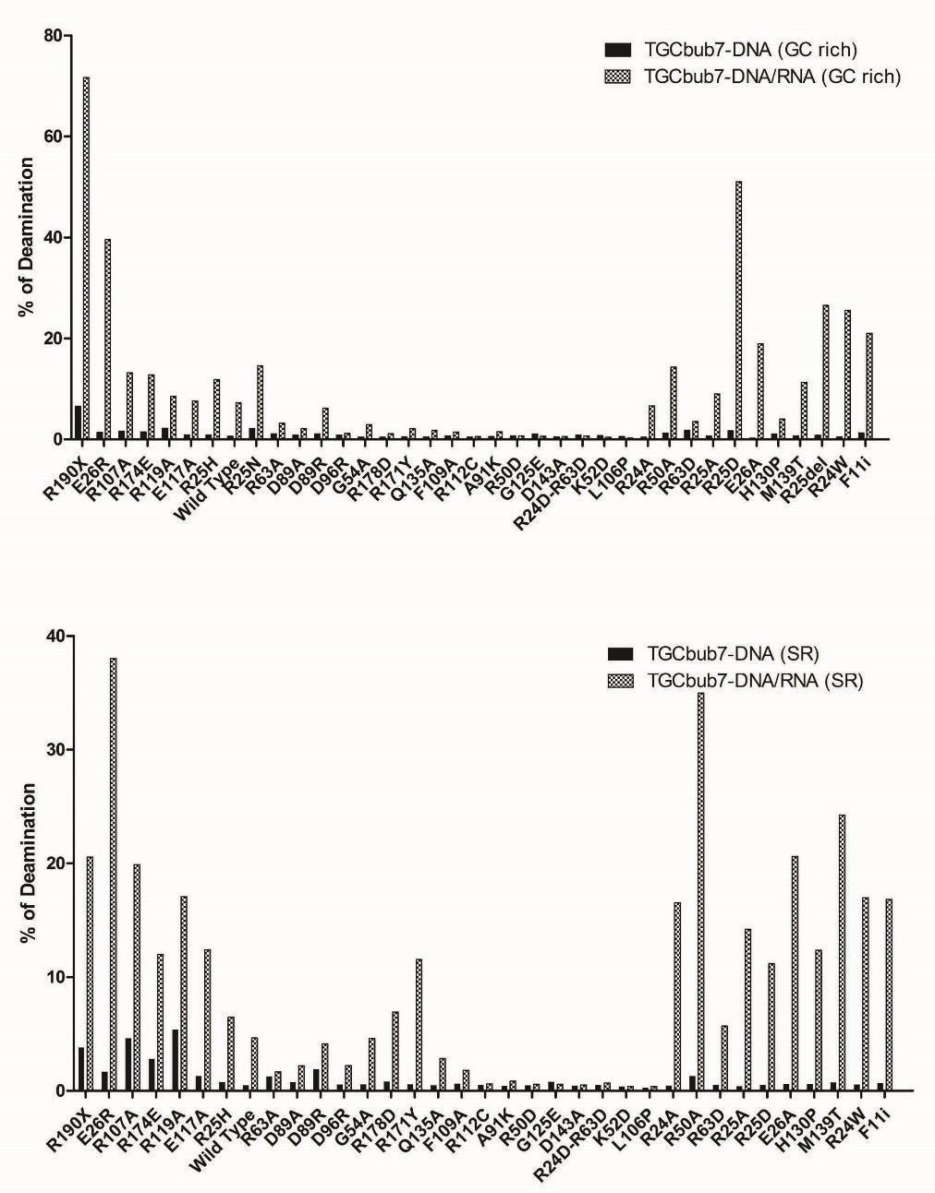


Figure 23. Surface residue and Hyper IgM mutants of AID exhibit differential preference for GC-rich and SR DNA/DNA *vs.* DNA/RNA bubbles.

Top shows a panel of 36 AID mutants tested for activity on TGCbub7-DNA/DNA (GC-rich) *vs.* TGCbub7-DNA/RNA (GC-rich) to determine relative preference. Bottom panel shows a comparison of the activity of the set of AID mutants on TGCbub7-DNA/DNA (SR) *vs.* TGCbub7-DNA/RNA (SR).

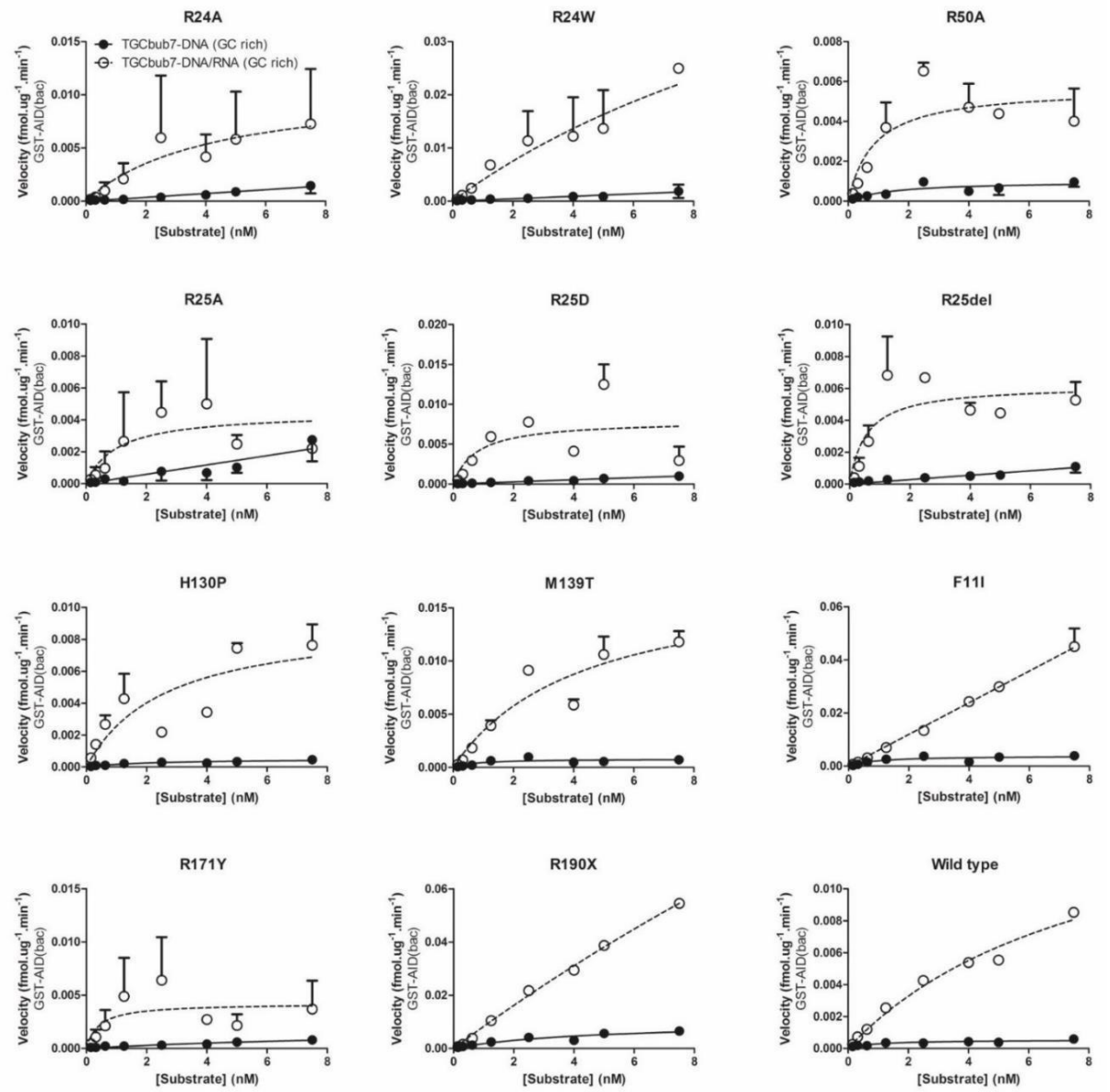


Figure 24. Enzymatic kinetics of Surface residue and Hyper IgM mutants of AID on GC-rich DNA/RNA hybrids

Catalytic rates of selected category IV mutants on TGCbub7-DNA/DNA (GC-rich) and TGCbub7-DNA/RNA (GC-rich), to confirm the significantly enhanced preference for the latter. A range of 0.1-8nM substrate concentration were incubated with 1μg GST-AID. P values for all shown graphs were <0.05. For each mutant, two independently purified preparations were tested.

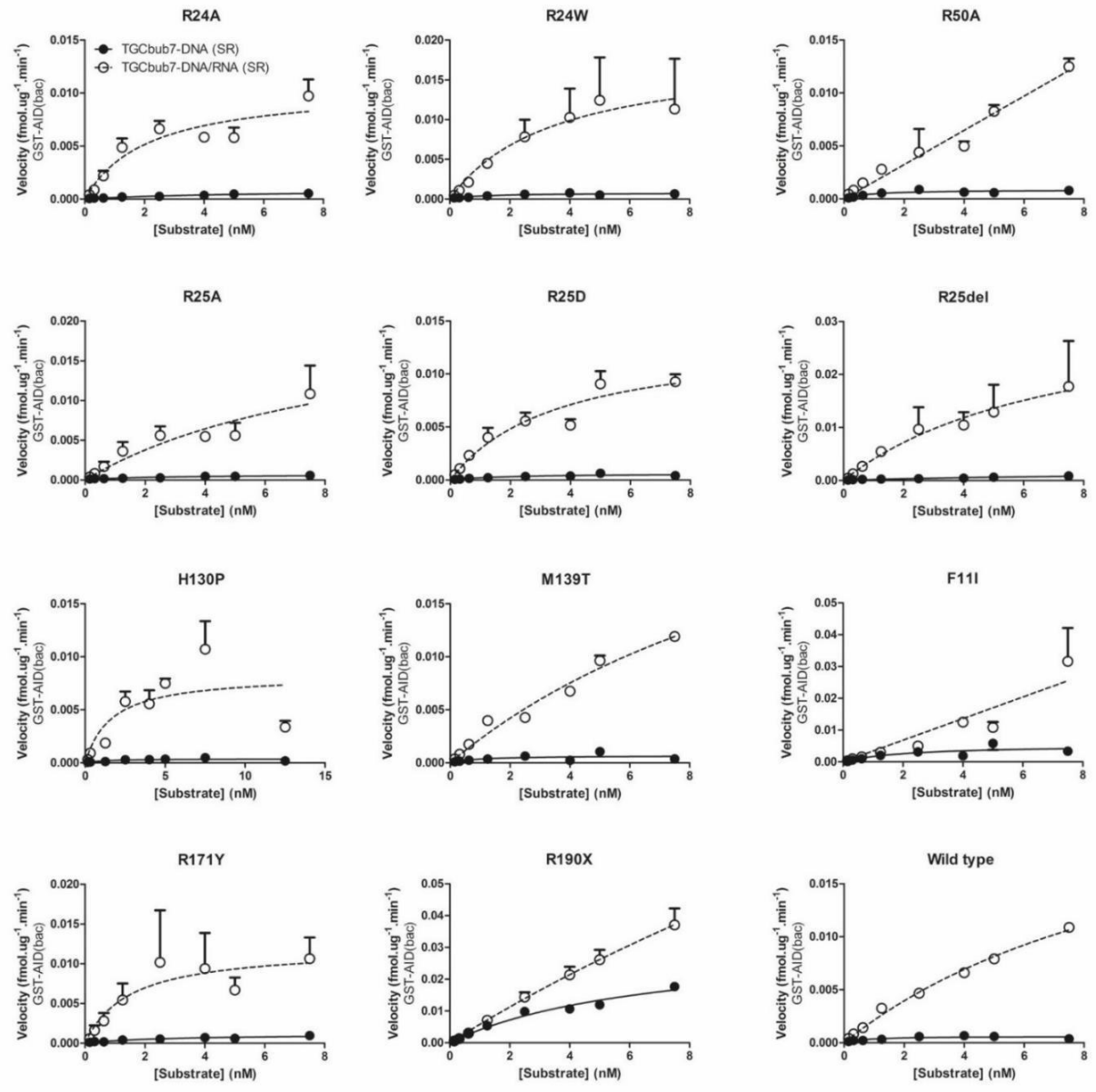


Figure 25. Enzymatic kinetics of Surface residue and Hyper IgM mutants of AID on SR DNA/RNA hybrids

Catalytic rates of selected category IV mutants on TGCbub7-DNA/DNA (SR) and TGCbub7-DNA/RNA (SR), to confirm the significantly enhanced preference for the latter. A range of 0.1-8nM substrate concentration were incubated with 1μg GST-AID. P values for all shown graphs were <0.05. For each mutant, two independently purified preparations were tested.

Summary

AID mutates antibody genes to initiate antibody diversification by preferentially targeting single stranded DNA structures *in vitro* and *in vivo* such as stem-loops, bubbles and R-loop regions that are released during the process of transcription.

In this study, we sought to investigate the role of RNA itself in attracting AID activity. We designed stable substrates that simulate transient breathing of DNA/RNA hybrids during transcription and we confirmed the structural integrity of these substrates through several assays and digests. We then measured the activity of purified AID on DNA/RNA hybrids and discovered that not only does AID bind to DNA/RNA hybrids but if the DNA/RNA hybrids contain a higher GC content, as in the case in immunoglobulin switch region sequences, AID favours these over pure DNA sequences. We confirmed these results using AID purified in a variety of different manners, to ensure our results reflect a bona fide property of AID itself. Moreover, we used a cell-free coupled transcription-AID activity assay to measure transcription dependent AID activity *in vitro* in the presence of RNaseH and found that RNA also stimulates transcription-dependent AID activity.

We then studied how AID can interact with DNA/RNA hybrids by performing in silico docking of RNA and AID. We discovered a putative RNA binding groove on the surface of AID which can accommodate the RNA strand of DNA/RNA hybrids, whilst allowing the DNA

strand to be situated in one of two DNA binding grooves that we previously identified. To validate the insights gained through the in silico structural and docking work, we used 41 mutants of AID targeting surface residues predicted to interact with ssDNA or RNA. We found that DNA/RNA hybrids restored activity of several AID mutants including those found in Hyper IgM patients that otherwise lack activity on pure DNA substrates. Given the prevalence of DNA/RNA hybrids at Immunoglobulin switch sequences, our findings provide evidence of a novel role for RNA in regulating AID activity.

Discussion

The link between AID activity and transcription is well established. AID's targeting of highly transcribed genes together with its specificity for ssDNA *in vitro* have led to the model that transcription liberates ssDNA substrates for AID to act on. The long-standing puzzle within this model has been why AID preferentially targets Ig loci whilst seemingly sparing other highly transcribed genes. Three facets of the transcription process appear to play a role in this regulation: firstly, it was recently shown that AID targets Ig and non-Ig loci at sites that are marked by a unique convergence of promotors, enhancers and bidirectional transcription, and it has long been appreciated that Ig loci are home to numerous robust and bidirectional promoters and enhancers (Buerstedde et al., 2014; Meng et al., 2014; Qian et al., 2014; Alinikula and Schatz, 2014). Secondly, several factors associated with the transcriptional machinery, including the RNA exosome and spliceosome, have been implicated in recruiting AID to sites of transcription (Basu et al., 2011; Conticello et al., 2008; Pefanis et al., 2014; Hu et al., 2015; Mondal et al., 2016; Nowak et al., 2011; Pavri et al., 2010; Willmann et al., 2012; Taylor et al., 2014; DiMenna and Chaudhuri, 2016; Zheng et al., 2015). Thirdly, intronic RNA has been suggested to directly bind AID and contribute to its targeting to switch regions (DiMenna and Chaudhuri, 2016; Zheng et al., 2015). Thus, RNA has been implicated in recruitment of AID both through direct interaction, and indirectly through association with transcription and RNA-processingfactors.

Our data suggest an additional role for RNA in that in the absence of any other proteins, RNA itself boosts the catalytic activity of AID on switch region-like sequences. Like dsDNA, DNA/RNA hybrids *in vivo* are of course fully complementary in sequence. It has previously been shown that AID can efficiently target supercoiled fully complementary dsDNA, presumably due to natural strand breathing liberating transient ssDNA regions in the form of bubbles and stemloops (Storb, 2014; Shen, 2007). In the same manner that DNA/DNA bubble oligonucleotide substrates simulate this transient breathing, the DNA/RNA hybrid structures studied here were intended to simulate transiently single stranded bubble or stemloop DNA/RNA hybrids formed due to natural breathing of multi kb-long GC-rich R loop dsDNA/RNA known to be abundantly present at Ig loci (Yu et al., 2003; Yu el al., 2005; Stavnezer, 2011; Wright et al., 2004; Huang et al., 2007).

There are several possible mechanisms to explain this RNA-mediated enhancement of AID's catalytic activity. We found that AID can bind DNA/RNA hybrids with high nM range affinities, and depending on the sequence, even more efficiently than DNA/DNA substrates. Structural modeling of AID in complex with ssDNA or RNA hybrids hinted at a putative RNA binding groove present on the surface of AID, distal to the catalytic pocket, in addition to the two ssDNA-binding groove previously identified. Recent work suggests that sporadic ssDNA binding on the surface of AID is a bottleneck for activity, since only a low proportion of ssDNA bound in one of said ssDNA binding grooves pass over the catalytic pocket and thus can

potentially position a dC for mutation (King et al., 2015). Simulations suggested that when binding DNA/RNA hybrids AID likely binds the RNA strand in said RNA binding groove and the DNA strand in ssDNA- binding groove 2 since it is proximal to the RNA binding groove (Fig. 20) relative to the separation between ssDNA-binding grooves 1 and 2. In contrast, when binding DNA/DNA, the target ssDNA strand has an equal probability of binding in either ssDNA-binding grooves 1 or 2 (Fig. 19a,b). Because ssDNA binding groove 2 forms fewer deamination-conducive complexes than groove 1 (King et al., 2015), this may partly explain reduced activity on DNA/RNA substrates. Indeed, mutations that perturb ssDNA binding groove 1 (F11I, F15L, R24D, R24W, R25A, R25D, R25del, E26A and R50A) hinder/ablate AID activity on DNA/DNA bubbles. Interestingly, these mutations retain, or even enhance activity on DNA/RNA substrates. Partial rescue of these mutants by DNA/RNA hybrids may indicate that DNA/RNA hybrids guide the target ssDNA preferentially to ssDNA binding groove 2 and limit binding to the disrupted DNA binding groove 1. Non-disruptive DNA binding groove 1 mutants (e.g. R25H and R25N) which reflect variations naturally present in orthologous AID and other APOBECs, did not disrupt activity on DNA/DNA activity and thus exhibited comparable activity to wild type AID. Likewise, disruption of residues specific to ssDNA binding groove 2 (R63A, R63D and D96R) did not ablate activity on DNA/DNA substrates, presumably because the dominant ssDNA binding groove 1 was left intact and was capable of positioning ssDNA for deamination.

Since GC-rich and SR DNA/RNA hybrid and DNA/DNA bubbles contain highly GC-rich bubble regions which are likely less flexible due to inter-strand hydrogen bonding, it is possible that the two ssDNA strands of the bubble are less able to bind in both ssDNA binding grooves 1 and 2, as compared to bubbles of lower GC content, thus accounting for generally lower activity of AID on GC-rich and SR DNA/DNA bubbles as compared to DNA/DNA bubbles of a random sequence. In contrast, in GC-rich or SR DNA/RNA hybrids, RNA fit in the RNA binding groove may guide the target DNA strand to ssDNA-binding groove 2 because of the closer proximity of these two grooves as described above, thus increasing the likelihood of the target ssDNA passing over the catalytic pocket. In support of this, mutation of residues specific to DNA binding groove 2, but not groove 1, hinder or ablate activity on GC-rich or SR DNA/RNA substrates. R190X in which the C-terminal α7 domain is truncated is of interest as we and others have shown that the highly flexible α7 domain likely adopts several conformations on the surface of AID but favors binding to a region distal from the catalytic pocket (King et al., 2015; Methot et al., 2015). We noted that this region also encompasses many of the residues in the putative RNA binding groove and thus α7 has a high likelihood of at least partially blocking RNA binding. In R190X, the absence of α7 likely increases the availability of the RNA binding groove thus resulting in significant activity on GC-rich DNA/RNA substrates. Notably, when examining point mutants localized to the RNA binding region, none ablated activity on DNA/RNA bubbles. This suggests that the RNA binding groove contains highly redundant RNA-

contact residues; a notion that is also supported by AID:RNA complex simulation data (Table 3). This is in contrast to point mutations of ssDNA binding grooves which more frequently impact AID activity, and may be taken as evidence for the biological importance of RNA binding to the surface of AID.

In addition to the role played by specific surface residues of AID located along the ssDNA and RNA binding grooves, it is possible that non-B form duplex structures are assumed by DNA/RNA hybrids of a certain GC content, which may enhance binding to the surface of AID. Indeed, GC-rich DNA/RNA hybrids have been shown to frequently adopt an intermediate conformation between an A-form of a dsRNA structure and a B-form DNA duplex (Shaw et al., 2008). Whatever the case may be, human and mouse SR sequences targeted by AID during CSR are particularly GC-rich. It appears that this is an evolutionary adaptation given that SRs of earlier- diverged species such as amphibians are less GC-rich (Roy and Lieber, 2009; Zarrin et al., 2004; Roy et al., 2008). It remains to be determined if DNA/RNAhybrids directly boost AID activity *in vivo*, as we have observed for purified AID *in vitro*. If so, the inherent catalytic preference for GC- rich and SR DNA/RNA hybrids could be viewed as an example of enzyme:substrate co-evolution.

Future directions

While this study proposes an important role for RNA in recruiting AID, future experiments will be conducted to further explore this hypothesis. This includes expanding the list of mutants that were tested on the DNA/RNA hybrids. Currently 41 mutants were used, 28 of which are residues from the DNA binding grooves, 11 are from the RNA binding groove, and 11 are Hyper IgM mutants. The following experiments should include all the Hyper IgM mutants identified in literature as well as designing multiple surface residue mutations further targeting the putative RNA binding region on the surface of AID, or mutants targeting combinations of the RNA as well as ssDNA binding grooves. Since single point mutants in the RNA binding groove didn't ablate AID activity due to the redundancy of the RNA-contact residues, a more comprehensive library can be composed by adding double and triple surface residue mutants that targets the RNA and DNA binding grooves separately or together.

In addition, further work with the transcription-dependent AID activity assay using DNA sequences of various GC contents, or actual gene sequences from the Ig loci will be conducted to probe for differential role of RNA, according to sequence context. In the future, novel assays can be designed using yeast or mammalian cell lines to study the role of DNA/RNA hybrids in recruitment of AID *in vivo.*

References

Abdouni, H., King, JJ., Suliman, M., Quinlan, M., Fifield, H., Larijani, M. (2013) Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Res*earch, 41(10), 5457-5468.

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen Recognition and Innate Immunity. *Cell*, 124(4), 783-801.

Alinikula, J. and Schatz, D.G. (2014) Super-enhancer transcription converges on AID. *Cell*, **159**, 1490-1492.

Barreto, V., Pan-Hammarstrom, Q., Zhao, Y., Hammarstrom, L., Misulovin, Z., & Nussenzweig, M. (2005). AID from bony fish catalyzes class switch recombination. *The Journal Of Experimental Medicine*, 202(6), 733-738.

Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R. *et al.* (2011) The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*, **144**, 353-363

Beck, G. & Habicht, G. (1996). *Immunity and the Invertebrates. Scientific American*, 275(5), 60-66.

Begum, N.A., Kinoshita, K., Kakazu, N., Muramatsu, M., Nagaoka, H., Shinkura, R., Biniszkiewicz, D., Boyer, L.A., Jaenisch, R. and Honjo, T. (2004) Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. *Science*, **305**, 1160-1163aUG 1120.

Begum, N.A., Kinoshita, K., Muramatsu, M., Nagaoka, H., Shinkura, R. and Honjo, T. (2004) De novo protein synthesis is required for activation-induced cytidine deaminase-dependent DNA cleavage in immunoglobulin class switch recombination. *Proc Natl Acad Sci U S A*, **101**, 13003-13007

Begum, N.A., Stanlie, A., Doi, T., Sasaki, Y., Jin, H.W., Kim, Y.S., Nagaoka, H. and Honjo, T. (2009) Further evidence for involvement of a noncanonical function of uracil DNA glycosylase in class switch recombination. *Proc Natl Acad Sci U S A*, **106**, 2752-2757

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

Besmer, E., Market, E. and Papavasiliou, F.N. (2006) The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. *Mol Cell Biol*, **26**, 4378-4385.

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

Buerstedde, J.M., Alinikula, J., Arakawa, H., McDonald, J.J. and Schatz, D.G. (2014) Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *PLoS Biol*, **12**, e1001831.

Cabral-Marques, O., Klaver, S., Schimke, L.F. et al. (2014) First Report of the Hyper-IgM Syndrome Registry of the Latin American Society for Immunodeficiencies: Novel Mutations, Unique Infections, and Outcomes *J Clin Immunol* 34: 146.

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

Caratão, N., Cortesão, C., Reis, P., Freitas, R., Jacob, C., & Pastorino, A. et al. (2013). A novel activation-induced cytidine deaminase (AID) mutation in Brazilian patients with hyper-IgM type 2 syndrome. *Clinical Immunology* 148(2), 279–286

Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan,

R.T., Datta, A. *et al.* (2007) Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv Immunol*, **94**, 157-214.

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

Conticello, S.G., Ganesh, K., Xue, K., Lu, M., Rada, C. and Neuberger, M.S. (2008) Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol Cell*, **31**, 474-484.

Dancyger, A.M., King, J.J., Quinlan, M.J., Fifield, H., Tucker, S., Saunders, H.L., Berru, M., Magor, B.G., Martin, A. and Larijani, M. (2012) Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single668 stranded DNA binding. *Faseb J*.

Dayn, A., Malkhosyan, S. and Mirkin, S.M. (1992) Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res*, **20**, 5991-5997.

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

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

Di Noia, J. and Neuberger, M.S. (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*, **419**, 43-48.

Di Noia, J.M., Williams, G.T., Chan, D.T., Buerstedde, J.M., Baldwin, G.S. and Neuberger, M.S. (2007) Dependence of antibody gene diversification on uracil excision. *J Exp Med*, **204**, 3209-3219.

Durandy, A., Peron, S., Taubenheim, N. and Fischer, A. (2006) Activation-induced cytidine deaminase: structure-function relationship as based on the studyof mutants. *Hum Mutat*, **27**, 1185-

1191.

Fleisher, T. (1997). Introduction to Diagnostic Laboratory Immunology. JAMA: *The Journal Of The American Medical Association*, 278(22), 1823.

Fukita, Y., Jacobs, H. and Rajewsky, K. (1998) Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity*, **9**, 105-114.

Gellert, M. (2002). V(D)J Recombination: RAG Proteins, Repair Factors, and Regulation. Annual Review Of Biochemistry, 71(1), 101-132.

Honjo, T. (2008) A memoir of AID, which engraves antibody memory on DNA. *Nat Immunol*, **9**, 335-337.

Huang, F.T., Yu, K., Balter, B.B., Selsing, E., Oruc, Z., Khamlichi, A.A., Hsieh, C.L. and Lieber, M.R. (2007) Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. *Mol Cell Biol*, **27**, 5921-5932.

Hu, W., Begum, N.A., Mondal., S., Stanlie, A. and Honjo, T. (2015) Identification of DNA cleavage- and recombination-specific hnRNP cofactors for activation-induced cytidine deaminase. *Proc Natl Acad Sci U S A*, **112**, 5791-5796.

Imai, K., Zhu, Y., Revy, P., Morio, T., Mizutani, S., Nonoyamac, S., Durandy A. & Fischer, A. et al. (2005). Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. *Clinical Immunology*, 115(3), 277-285.

Jiang, F., Lin, F., Price, R., Gu, J., Medeiros, L., & Zhang, H. et al. (2002). Rapid Detection of IgH/BCL2 Rearrangement in Follicular Lymphoma by Interphase Fluorescence in Situ Hybridization with Bacterial Artificial Chromosome Probes. *The Journal Of Molecular Diagnostics*, 4(3), 144-149.

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

Kramer MH, Hermans J, Wijburg E, Philippo K, Geelen E, van Krieken JH, et al. (1998) Clinical relevance of BCL2, BCL6, and MYC rearrangements in diffuse large B-cell lymphoma. *Blood* **92**(9):3152–62.

Krasilnikov, A.S., Podtelezhnikov, A., Vologodskii, A 606. and Mirkin, S.M. (1999) Large-scale effects of transcriptional DNA supercoiling *in vivo*. *J Mol Biol*, **292**, 1149-1160.

Lanasa, M. & Weinberg, J. (2011). Immunoglobulin class switch recombination in chronic lymphocytic leukemia. Leukemia & Lymphoma, 52(7), 1398-1400.

Larijani, M., Frieder, D., Basit, W. and 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**, 840-845

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

Larijani, M., Petrov, A.P., Kolenchenko, O., Berru, M., Krylov, S.N. and 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**, 20-30.

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

Larijani, M. and Martin, A. (2012) The biochemistry of activation-induced deaminase and its physiological functions. *Semin Immunol*, **24**, 255-263.

Liang, G., Kitamura, K., Wang, Z., Liu, G., Chowdhury, S., Fu, W., Koura, M., Wakae, K., Honjo, T. and Muramatsu, M. (2013) RNA editing of hepatitis B virus transcripts by activation588 induced cytidine deaminase. *Proc Natl Acad Sci U S A*, **110**, 2246-2251.

Liu, M. et al. (2008) Two levels of protection for the B cell genome during somatic hypermutation. Nature 451, 841–845.

Longerich, S., Basu, U., Alt, F., & Storb, U. (2006). AID in somatic hypermutation and class switch recombination [Current Opinion in Immunology 2006, 18:164–174]. Current Opinion In Immunology, 18(6), 769.

Martin, A., Bardwell, P., Woo, C., Fan, M., Shulman, M., & Scharff, M. (2002). Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. Nature, 415(6873), 802-806.

Maul, R.W., Saribasak, H., Martomo, S.A., McClure, R.L., Yang, W., Vaisman, A., Gramlich, H.S., Schatz, D.G., Woodgate, R., Wilson, D.M., 3rd *et al.* (2011) Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol*, **12**, 70-76.

Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D. *et al.* (2014) Convergent transcription at intragenic super enhancers targets AID-initiated genomic instability. *Cell*, **159**, 1538-1548.

Methot, S.P., Litzler, L.C., Trajtenberg, F., Zahn, A., Robert, F., Pelletier, J., Buschiazzo, A., Magor, B.G. and Di Noia, J.M. (2015) Consecutive interactions with HSP90 and eEF1A underlie a functional maturation and storage pathway of AID in the cytoplasm. *J Exp Med*, **212**, 581-596.

Mills, F.C., Brooker, J.S. and Camerini-Otero, R.D. (1990) Sequences of human immunoglobulin switch regions: implications for recombination and transcription. *Nucleic Acids Res*, **18**, 7305-7316.

Mondal., S., Begum, N.A., Hu, W. and Honjo, T. (2016) Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. *Proc Natl Acad Sci U S A*, **113** E1545-1554.

Mu, Y., Prochnow, C., Pham, P., Chen, X.S. and Goodman, M.F. (2012) A structural basis for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination- defective hyper-IgM-2 mutants. *J Biol Chem*, **287**, 28007-28016.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O. and Honjo, T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase familyin germinal center B-cells. *J Biol Chem*, **274**, 18470-18476

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

Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N.A. and Honjo, T. (2007) Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Adv Immunol*, **94**, 1-36.

Nabel, C.S., Lee, J.W., Wang, L.C. and Kohli, R.M. (2013) Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. *Proc Natl Acad Sci U S A*, **110**, 14225-14230.

Nowak, U., Matthews, A.J., Zheng, S. and Chaudhuri, J. (2011) The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat Immunol*, **12**, 160-166.

Okazaki, I., Kinoshita, K., Muramatsu, M., Yoshikawa, K., & Honjo, T. (2002). The AID enzyme induces class switch recombination in fibroblasts. *Nature*, 416(6878), 340-345.

Ouadani, H., Ben-Mustapha, I., Ben-ali, M., Ben-khemis, L., Larguèche, B., & Boussoffara, R. et al. (2015). Novel and recurrent AID mutations underlie prevalent autosomal recessive form of HIGM in consanguineous patients. *Immunogenetics*, 68(1), 19-28.

Ouadani, H., Ben-Mustapha, I., Ben-Ali, M., Largueche, B., Jovanic, T., Garcia, S., Arcangioli, B., Elloumi-Zghal., H., Fathallah, D., Hachicha, M. *et al.* (2016) Activation induced cytidine deaminase mutant (AID-His130Pro) from Hyper IgM 2 patient retained mutagenic activity on SHM artificial substrate. *Mol Immunol*, **79**, 77-82.

Palacios F, Moreno P, Morande P, Abreu C, Correa A, Porro V, et al. High expression of AID and active class switch recombination might account for a more aggressive disease in unmutated CLL patients: link with an activated microenvironment in CLL disease. *Blood*. 2010;115(22):4488–96.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R., Küppers, R., & Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*, 412(6844), 341-346.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V. *et al.* (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*, **143**, 122-133.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N. and Basu,U. (2014) Noncoding RNA transcription targets AID to divergently transcribed loci in B-cells. *Nature*, **514**, 389-393.

Pefanis, E. and Basu, U. (2015) RNA Exosome Regulates AID DNA Mutator Activity in the BCell Genome. *Adv Immunol*, **127**, 257-308.

Peters, A. and Storb, U. (1996) Somatic hypermutation of immunoglobulin genes is linked to

transcription initiation. *Immunity*, **4**, 57-65.

Petersen-Mahrt, S.K., Harris, R.S. and Neuberger, M.S. (2002) AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature*, **418**, 99-104.

Petersen-Mahrt, S.K. and Neuberger, M.S. (2003) *In vitro* deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem*, **278**, 19583-19586.

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

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-704 Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathe, E., Benner, C. *et al.* (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell*, **159**, 1524-1537.

Quartier, P., Bustamante, J., Sanal, O., Plebani, A., Debré, M., & Deville, A. et al. (2004). Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. *Clinical Immunology*, 110(1), 22-29.

Rada, C., Williams, G.T., Nilsen, H 556 ., Barnes, D.E., Lindahl, T. and Neuberger, M.S. (2002) Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG558 Deficient Mice. *Curr Biol*, **12**, 1748-1755.

Rada, C., Di Noia, J.M. and 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**, 163-171.

Ramiro, A., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., & Muramatsu, M.

et al. (2004). AID Is Required for c-myc/IgH Chromosome Translocations *In Vivo*. Cell, 118(4), 431-438.

Ranjit, S., Khair, L., Linehan, E.K., Ucher, A.J., Chakrabarti, M., Schrader, C.E. and Stavnezer, J. (2011) AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus. *J Immunol*, **187**, 2464-2475

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal., O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A. *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565-575.

Robbiani, D., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., & Difilippantonio, S. et al. (2008). AID Is Required for the Chromosomal Breaks in c-myc that Lead to cmyc /IgH Translocations. Cell, 135(6), 1028-1038.

Roy, D., Yu, K. and Lieber, M.R. (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol*, **28**, 50-60.

Roy, D. and Lieber, M.R. (2009) G clustering is important for the initiation of transcription induced R-loops *in vitro*, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol*, **29**, 3124-3133.

Roy, A., Kucukural., A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, **5**, 725-738.

Shaw, N.N. and Arya, D.P. (2008) Recognition of the unique structure of DNA:RNA hybrids. *Biochimie*, **90**, 1026-1039.

Shen, H.M. (2007) Activation-induced cytidine deaminase acts on double-strand breaks in vitro. *Mol Immunol*, **44**, 974-983.

Shen, H.M., Poirier, M.G., Allen, M.J., North, J., Lal., R., Widom, J. and Storb, U. (2009) The activation-induced cytidine deaminase (AID) efficiently targets DNA in nucleosomes but only during transcription. *J Exp Med*, **206**, 1057-1071.

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

Stavnezer, J. (2011) Complex regulation and function of activation-induced cytidine deaminase. *Trends Immunol*, **32**, 194-201

Storb, U., Peters, A., Klotz, E., Kim, N., Shen, H.M., Hackett, J., Rogerson, B., O'Brien, R. and Martin, T.E. (1998) Immunoglobulin transgenes as targets for somatic hypermutation. *Int J Dev Biol*, **42**, 977-982.

Storb, U., Peters, A., Klotz, E., Kim, N., Shen, H.M., Kage, K., Rogerson, B. and Martin, T.E. (1998) Somatic hypermutation of immunoglobulin genes is linked to transcription. *Curr Top Microbiol Immunol*, **229**, 11-19.

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

Taylor, B.J., Wu, Y.L. and Rada, C. (2014) Active RNAP pre-initiation sites are highly mutated by cytidine deaminases in yeast, with AID targeting small RNA genes. *eLife*, **3**, e03553.

Teng, G. & Papavasiliou, F. (2007). Immunoglobulin Somatic Hypermutation. Annual Review Of Genetics, 41(1), 107-120.

Torlakovic E, Torlakovic G, Nguyen PL, Brunning RD, Delabie J (Oct 2002). "The value of anti-pax-5 immunostaining in routinely fixed and paraffin-embedded sections: a novel pan pre-B and B-cell marker". *The American Journal of Surgical Pathology*. **26** (10): 1343–50.

Trotta, L., Hautala, T., Hämäläinen, S., Syrjänen, J., Viskari, H., & Almusa, H. et al. (2016). Enrichment of rare variants in population isolates: single AICDA mutation responsible for hyper-IgM syndrome type 2 in Finland. *European Journal Of Human Genetics*, 24(10), 1473-1478.

Victora, G. & Nussenzweig, M. (2012). Germinal Centers. Annual Review Of Immunology, 30(1), 429-457.

Warrington, R., Watson, W., Kim, H., & Antonetti, F. (2011). An introduction to immunology and immunopathology. Allergy, Asthma & Clinical Immunology, 7(Suppl 1), S1.

Willmann, K.L., Milosevic, S., Pauklin, S. 655, Schmitz, K.M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V. *et al.* (2012) A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *J Exp Med*, **209**, 2099-2111.

Wright, B.E., Schmidt, K.H. and Minnick, M.F. (2004) Mechanisms by which transcription can regulate somatic hypermutation. *Genes Immun*, **5**, 176-182.

Yousif, A.S., Stanlie, A., Mondal., S., Honjo, T. and Begum, N.A. (2014) Differential regulation of S-region hypermutation and class-switch recombination by noncanonical functions of uracil DNA glycosylase. *Proc Natl Acad Sci U S A*, **111**, E1016-1024

Yousif, A.S., Stanlie, A., Begum, N.A. and Honjo, T. (2014) Opinion: uracil DNA glycosylase (UNG) plays distinct and non-canonical roles in somatic hypermutation and class switch recombination. *Int Immunol*, **26**, 575-578.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E. and Lieber, M.R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B-cells. *Nat Immunol*, **4**, 442-451.

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

Yu, K., Roy, D., Bayramyan, M., Haworth, I.S. and Lieber, M.R. (2005) Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol Cell Biol*, **25**, 1730-1736.

Zarrin, A.A., Alt, F.W., Chaudhuri, J., Stokes, N., Kaushal., D., Du Pasquier, L. and Tian, M. (2004) An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol*, **5**, 1275-1281.

Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.

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

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