Structure-function analysis of AID reveals intramolecular catalytic inhibition and a novel functional domain required for targeting to chromatin.

Anil Kumar Eranki (McGill ID: 260219694)

Division of Experimental Medicine

Department of Medicine

McGill University, Montreal

September 2012

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master's in Science (*M.Sc.*)

TABLE OF CONTENTS

1. Abstract – English ……..…Page 03
2. Abstract – French……..Page 05
3. Acknowledgments…….……Page 07
4. Introduction……..…….....Page 08
5. Methods……...……..Page 20
6. Results……..…….……Page 29
7. Discussion and conclusions ..…….....Page 39
8. References……..……..…Page 47

ABSTRACT- ENGLISH

The B-cell specific enzyme Activation-induced deaminase (AID) is responsible for the diversification of rearranged antibody genes in activated B cells during humoral immune responses. AID deaminates dC to dU at the Immunoglobulin (Ig) loci and the uracil thereby introduced is processed in an error-prone fashion by specific DNA repair enzymes leading to somatic hypermutation (SHM), which underpins the affinity maturation of the antibody response, and to the DNA breaks necessary for class-switch recombination (CSR), which changes the isotype class of the antibody from IgM to IgG, IgA or IgE.

Structure-function analysis of the AID protein was undertaken to better refine the 3-dimensional model of the protein previously published (Patenaude et al., 2009) which was based on a similar protein APOBEC2 whose structure has been previously elucidated (Prochnow et al., 2007). However, the C-terminal domain of AID (residues 182-198) is unique to AID and its structural details are lacking. This region is further interesting because the C-terminal domain, encoded by exon 5 of *Aicda* gene, has been shown to be essential for several of its biological functions: CSR, cytoplasmic retention, and nucleo-cytoplasmic shuttling. To better understand this region, and to define and identify potential functional domains, structure-function analysis was undertaken by creating truncations and point mutants and systematically assessing them for catalytic, SHM, and CSR activities. From such studies, it was identified that alpha-helix 6 (α6) is structurally important for AID catalytic activity while the C-terminal domain is not just dispensable for catalytic activity, but mutants lacking a portion of or all of the C-terminal domain are hyperactive for both catalytic activity and somatic hypermutation. We speculated that the hyperactivity of these mutants could account for the apparent

dominant negative effect observed in a subset of patients with Hyper IgM type II (HIGM-II) syndrome, an immunodeficiency characterized by a normal or elevated SHM and lack or reduced switched isotypes in serum. While most loss of function mutations in AID are transmitted as autosomal recessive traits resulting in AID deficiency, in a subset of patients suffering from HIGM-II, *Aicda* mutations are inherited in an autosomal dominant (AD) manner. While the heterozygous carriers of the recessive alleles are asymptomatic, AD patients, who express one normal protein of AID along with a C-terminally truncated protein, show a similar HIGM-II immunodeficiency to AID-deficient patients and AID-null mice. This would suggest that the truncated AID protein has a dominant negative effect on the full-length protein. Two non-mutually exclusive possible explanations for this effect have been proposed: the truncated protein could form oligomers with full-length AID protein and render them non-functional, or the truncated protein could compete with the full-length protein for cellular factors involved in nucleo-cytoplasmic shuttling and/or other factors. We have investigated these possibilities and propose the alternative explanation that the hyperactivity of the truncated proteins leads to increased off-target mutations, DNA double-strand breaks, clonal expansion defects, and presumably p53-dependant apoptosis in primary B-lymphocytes. Our results correlate well with the clinical symptoms presented by AD patients: B cells with normal SHM and absence of the lymphadenopathies characteristic of AID-deficient patients. Furthermore, we identify a novel domain in alpha-helix 6 containing positively charged arginines that seem to be necessary for the enzyme's biological function, by mediating the interaction with a DNA targeting factor.

ABSTRACT- FRENCH

L’enzyme Activation-Induced Deaminase (AID), qui se trouve uniquement dans des lymphocytes B, est responsable pour la diversification des gènes d'anticorps réarrangés dans les cellules B actives au cours de la réponse immunitaire humorale. AID est responsible pour la déamination des dC à dU dans les gènes d’immunoglobulines (Ig). L'uracile ainsi introduit est traitée d'une manière erreurs par des enzymes de réparation conduisant l’hypermutation somatique (SH), promouvant la maturation de l'affinité de la réponse immunitaire, et des l'ADN brisées nécessaire pour la commutation isotypique (CS), qui change la classe isotype de l'anticorps d'IgM à IgG, IgA ou IgE. Analyse structure-fonction de la molécule AID a été réalisée afin d’ améliorer la modèle en 3-D de la protéine, précédemment publiée (Patenaude et al., 2009). Cependant, le domaine C-terminale de l'AID (les résidus 182-198) est unique à l'AID, mais ses détails structurels font absents. Cette région est plus intéressant parce que le domaine C-terminale, codée par le cinquième l'exon du gène AICDA, a été révélée essentielle pour plusieurs de ses fonctions biologiques: CS, la rétention cytoplasmique, et pour le transport nucléo-cytoplasmique. Afin de mieux comprendre cette région et à définir et à identifier des nouveaux domaines fonctionnels, un rapport structure-fonction a été réalisée par la création de troncatures et des mutants ponctuels et les évaluer systématiquement pour les activités catalytiques, SH, et CS. De ces études, on a constaté que l'hélice alpha 6 (α6) est structurellement importante pour l'activité catalytique d’AID alors que le domaine C-terminale n'est pas seulement superflu pour l'activité catalytique, mais des mutants manquant une partie ou de la totalité du domaine C-terminale sont hyperactifs pour l'activité catalytique et hypermutation somatique. Nous

avons supposé que l'hyperactivité de ces mutants pourrait expliquer l'effet apparent dominante négative observée dans un sous-ensemble de patients avec syndrome d'Hyper-IgM de type II (HIGM-II), une immunodéficience caractérisée par l'hypermutation somatique normale ou élevée et l'absence ou réduit d'anticorps de types IgG, IgA, et IgE. Alors que la plupart des mutations affectant le gène de la protéine AID sont transmises comme récessive autosomique résultant de carences de la protéine AID, dans un sous-ensemble de patients souffrant de HIGM-II, les mutations sont héritées dans une manière autosomique dominante (AD). Alors que des porteurs hétérozygotes des mutations récessifs sont asymptomatiques, les patients qui héritent des mutations dans une manière autosomique dominante expriment une protéine normale AID avec une protéine tronquée de côté C-terminale souffrent un déficit immunitaire. Cela suggère que la protéine AID tronquée a un effet dominante négative sur la protéine normale. Deux explications possibles ont été proposées: la protéine tronquée pourrait former des oligomères avec protéine AID normale et les rendre non-fonctionnelles, soit la protéine tronquée pouvait rivaliser avec la protéine normale pour des facteurs cellulaires impliqués dans le transport nucléo-cytoplasmique ou pour des autres facteurs. Nous avons étudié ces possibilités et on propose que l'hyperactivité des protéines tronquées conduit à une augmentation des mutations dans les gènes d'immunoglobulines, résultant en les défauts expansion clonale à cause d'apoptose dépendante sur p53 dans les lymphocytes B primaires. Nos résultats sont bien corrélés avec les symptômes cliniques présentés par les patients atteints de HIGM-II: les cellules B avec SHM normale et l'absence contrairement à les patients atteints du syndrome HIGM récessif autosomique avec d'adénopathies lymphocytiques. Nous aussi identifions

une domaine nouvelle en hélice alpha 6 contenant arginines chargés positives qui semblent être nécessaires pour la fonction biologique de l'enzyme.

ACKNOWLEDGMENTS

I would like to extend my humble gratitude to my supervisor, Javier M di Noia, for giving me the oppurtunity to work in his lab and for guiding me through a summer research stint and through my Master's studies. Javier was principally responsible for designing the experiments, while I was performed the experiments and analyzed the data. I would also like to the thank the members of my committee, who provided me with valuable feedback on the direction of my research: Dr. Jean-François Côté (Institut de Recherche Cliniques de Montréal, IRCM, Montréal), Dr. Alain Lamarre (Institit Nationale de Recherche Scientifique, INRS, Montréal), Dr. Marko Horb (Institut de Recherche Cliniques de Montréal, IRCM, Montréal).

I am also indebted to the following members of the lab, who helped me throughout my research: Anne-Marie Patenaude (Research Assistant), Stephen Methot (Ph. D candidate), Astrid Zahn (Research Associate), Alexandre Orthwein (Ph. D candidate). Anne-Marie Patenaude is credited with generating the data for the following figures: figure 4 A, B, and D, while I analyzed the data. Astrid Zahn generated and analyzed the data for the following figure: figure 6A. Stephent Methot generated the data for supplementary figure 3. I generated and analyzed the data for all the other figures. I would also like to thank the flow cytometry team at IRCM, Eric Massicotte and Julie Lord, for their patience and guidance. I would also like to thank Mylene Cawthorne for taking care of the mice used in my project.

INTRODUCTION

The vertebrate immune system is composed of two complementary arms – innate and adaptive. The innate system is characterized by non-specific immune responses to foreign antigens, is rapid, requires no priming, and forms the primary barrier against foreign antigens. The adaptive immune system, on the other hand, is very specific, requires priming, and is slow to respond. B- and T-cells are the two key players of the adaptive system that interact dynamically with each other and with the innate immune system to mount a successful immune response. Although both the cell lineages originate in the bone marrow, T cell precursors migrate to the thymus to further develop and mature, while the B cells remain in the bone marrow throughout their development and exit the bone marrow as mature B cells expressing immunoglobulins IgM and IgD on their surface after the primary gene diversification process. These B cells migrate to the spleen and other peripheral lymphoid organs to further mature into either B1-a/b cells, marginal zone (MZ) B-cells, or follicular (Fo) B-cells, the last of which can further develop into short-lived or long-lived antibody-producing plasma cells (LeBien, 2008; Corfe and Paige, 2009).

Development of B cells

The development of B-cells in the bone marrow involves several stages beginning with the hematopoietic stem cells (HSCs), early pro-B cell, late pro-B cell, large pre-B cell, small pre-B cell, and immature cells and mature B cells, in chronological order. During the early ProB cell stage, RAG 1 and RAG 2 enzymes initiate the primary antibody gene diversification processes, which involves the recombination of V, J, and C gene segments to form the variable and constant regions of the λ and κ light chains, and the

recombination of V, D, J, and C segments to form the variable and constant regions of the heavy chain (Schatz and Ji, 2011). The DNA cleavage activity of RAG1, aided by RAG2, is responsible for creating the DSBs necessary to initiate the recombination events. The DSBs are repaired and resolved by the non-homologous end-joining (NHEJ) machinery, which includes signaling and transducer molecules such as the ATM/ATR complex, Ku70/80 complex, and DNA-PKCs (Schatz and Ji, 2011). The result is the loss of the intervening double-stranded DNA, and the joining of the distal DNA strands (Figure I).

The immature B cells expressing IgM on their cell surface will be tested in the bone marrow for self-reactivity, a process termed central tolerance, to prevent autoimmunity. Self-tolerant B cells will be allowed to mature and express both IgM and IgD on the cell surface. These mature cells circulate throughout the lymphatic system and drain into the secondary lymphoid organs, of which spleen is the largest. The B cells enter the lymph nodes via the afferent lymphatic vessels and can either transition into follicular B cells or, in the spleen, into marginal zone B cells (Batista and Harwood, 2009). The follicular B cells are distinguished as IgM_{med} IgD_{hi} $CD21_{med}$ $CD23_{hi}$, while the marginal zone B cells are marked by IgM_{hi} IgD_{lo} $CD21_{hi}$ $CD23_{lo}$. The follicular B cells form follicle-like structures under the subcapsular sinus that are rich in professional antigen-presenting cells: macrophages and follicular dendritic cells (FDCs). It has been postulated that the mature B cells in the follicles encounter paracortical T cells and this interaction, given the appropriate signals and antigen stimulation, could initiate the germinal center reaction (Batista and Harwood, 2009) and AID-dependent secondary antibody gene diversification.

Antibody gene organization and protein structure

The basic structure of an antibody molecule consists of two heavy chains and two light chains. The heavy chains are bound to each other by disulfide bridges, which in turn are each bound to a light chain by additional disulfide bridges. The heavy and light chains each consist of a constant region and a variable region. The variable regions of one heavy and one light chain form the antigen-binding site and their corresponding exons are the target of somatic hypermutation. While the light chain constant region has no known functional role apart from its structural importance, the heavy chain constant region determines the effector function of the antibody isotype classes: IgM, IgD, IgA, IgG, and IgE (IgG has a few additional sub-isotypes). In humans, the immunoglobulin heavy chain locus is located on chromosome 14. The approximately 39 functional V_H, 27 D_H gene segments and 6 J_H gene segments are arranged in clusters (Figure I). On the same chromosome, downstream to the VDJ loci are the 9 functional C_H genes. Each C gene encodes for a specific constant domain, and is preceded by its own switch region (see section below). There are two immunoglobulin light chain loci in mammals: λ and κ. In humans, they are located on chromosomes 22 and 2 respectively. The Ig λ chain locus consists of about 32 functional V_λ genes and 4 alternating pairs of J_λ and C_λ genes. The Ig κ locus consists of 40 functional V_κ genes, 5 J_κ genes and a single C_κ constant gene. Although there is no functional difference between the λ and the κ locus genes, there seems to be a differential preference for one type over the other in different species, with a usage ratio of approximately 65% κ to 35% λ in humans (Murphy et. al, 2008). The diversity of the primary immunoglobulin repertoire, initiated by the RAG1/2 as mentioned above, is achieved through different combinations of the V(D)J segments

(termed combinatorial diversity) and through the deletion and insertion of nucleotides (termed junctional diversity) at the junctions of the recombined gene segments (Fanning et al., 1996). While the primary immunoglobulin repertoire offers sufficient flexibility in the initial recognition of foreign antigens, it is insufficient for the high affinity recognition needed for a successful clearance of the antigen and for mounting an efficient secondary response. The vertebrate immune system has evolved mechanisms to strengthen the initial interaction and to modulate the effector function of the immunoglobulin through a secondary gene diversification at the site of antigen encounter in the peripheral lymph nodes.

Mature B cells exiting the bone marrow express both IgM and IgD on their surface through alternative splicing of mRNA encoding Cμ and Cδ exons (Maki et al., 1981). These cells circulate through the peripheral lymph nodes, sampling free antigens and antigens presented by the antigen-presenting cells (APCs) like macrophages and dendritic cells. Once a B cell identifies its cognate antigen via the surface receptor (BCR) in peripheral lymph nodes, the cell is activated in either a T-cell dependent or independent manner. While the role of sIgD is unclear, it has been recently proposed that mature B cells preserve the ability to switch from IgM to IgD, and that IgD plays a crucial role in regulating respiratory immunity through basophils (Chen et al., 2009). However, it is generally the surface IgM that recognizes the antigen in the periphery through its variable region. A T-cell dependent immune response initiates the germinal center reaction where two forms of secondary Ig gene diversification processes - somatic hypermutation and class-switch recombination – take place. Both, SHM and

CSR are dependent on the germinal-center B cell specific protein Activation Induced Deaminase (AID).

Activation Induced Deaminase

AID is an approximately 24-KDa enzyme encoded by the *Aicda* gene, located on chromosome 12. The five exons of the human *Aicda* gene encode a 198 amino acid protein. The enzyme contains a zinc coordination motif that catalyzes the deamination reaction, which involves a nucleophilic attack on the deoxycytidine side chain's amine group, and its replacement with a hydroxyl group, converting the deoxycytidine to deoxyuridine. (AID) was first identified as an inducible protein in a subtractive hybridization screen for cDNAs in cytokine activated mouse B cell line CH12F3-2 (Muramatsu et al., 1999). Subsequently, Revy and colleagues identified that AID deficiency causes a hyper-IgM-like syndrome, similar to CD40L (CD40 ligand) X-linked deficiency, in eighteen patients from twelve unrelated families. Soon, strong evidence suggested that AID is necessary for class switch recombination, somatic hypermutation (Muramatsu et al., 2000) and gene conversion (Arakawa et al., 2002); the last of which is a form of genetic diversification used by chicken, rabbits, cats and pigs where AID-induced DNA breaks are resolved through homologous recombination with non-functional pseudogenes serving as donors (Maizels, 2005). While it was first hypothesized that AID might be an RNA-editing enzyme given its similarity to APOBEC-1, it was shown that AID could deaminate cytidine residues on DNA in *Escherichia coli* (Petersen-Mahrt et al., 2002) and cytidine residues *in vitro* on an artificial DNA substrate (Bransteitter et al., 2003). The same group also showed that AID can only deaminate cytidine residues on single-stranded DNA, but not on double-stranded DNA or on

RNA/DNA hybrids. Further evidence of AID's ability to deaminate cytidine residues in DNA came from the inhibition of the base-excision repair enzyme uridine DNA glycosylase (UNG) in DT40 chicken B cell line (Di Noia and Neuberger, 2002). Ung is responsible for the removal of the uracil base, which normally belong in the RNA, from the DNA. Di Noia and colleagues found that in the absence of Ung, the pattern of AID-induced mutations in the V_λ region of the DT40 genome changed from a mixture of transitions and transversions to predominantly transversions (more on this below, Figure II).

AID-induced somatic hypermutation, as attested to before, increases the affinity of antibodies to their cognate antigens during the germinal center reaction. AID deaminates deoxycytidine residues predominantly in the hot-spot motif WRCY (W = A or T, R = purine, Y = pyrimidine) in the recombined V(D)J of the heavy and light chain genes. Replication over the G:U mismatch can result in transition mutations to A:T, as the deoxyuridine is treated as a thymidine by the replication (Petersen-Mahrt et al., 2002). Mutations outside the G:U mismatch, i.e. at the A:T pairs can occur when the G:U mismatch is processed by the mismatch repair pathway (Di Noia and Neuberger, 2007). However, the deoxyuridine can also be processed by the base-excision repair pathway, in which uridine DNA glycosylase (UNG) first removes the uridine creating an abasic site. Further processing by error-prone polymerases at the abasic site can lead to transversions and transitions at the C:G pairs (Di Noia and Neuberger, 2007). B-cells in the germinal center reaction undergoing clonal expansion are under selection pressure, most likely driven by limited availability of antigen and pro-survival cytokines. Only those B cell clones with the highest possible affinity for a given antigen have the survival

advantage to out-proliferate the other B cells. It is this continuous process of AID-induced deamination and subsequent mutations at the immunoglobulin loci (mutations that start 150 nucleotides downstream of the IgV promoter and extend over about 2 Kb) that allows the B cells to improve the affinity of the expressed antibodies. The constant genes of either the heavy chain or the light chain region are not targeted by AID.

If the deoxycytidine residues are deaminated in the switch regions, instead of the recombined variable genes, class-switch recombination results. The switch regions, made of repeats which are upstream of the genes encoding for the constant region of the heavy chain, extend for a few kilobases containing multiple AID hotspot motifs and serve as targets for AID induced deoxycytidine deamination to deoxyuridine. The base excision repair (BER) machinery translates these mutations into single-strand breaks (SSBs) and then double strand breaks (DSBs). The DSBs created in this manner in the "donor" switch region of the Cμ and a downstream "acceptor" switch region (one of Cα1, Cα2, Cγ1, Cγ2, Cγ3, Cγ4, or Cε) are then resolved by the non-homologous end joining (NHEJ) machinery, which results in the excision of the intervening ds-DNA. The recombined VDJ is now transcribed in frame with the switched constant gene (reviewed in (Stavnezer et al., 2008), Figure III).

Regulation of Activation Induced cytidine Deaminase

AID is the only known enzyme in the human genome whose function is to mutate the DNA. The danger posed by such a mutagenic enzyme is obvious and the etiological role of AID in B cell lymphoma has been demonstrated (Ramiro et al., 2004; Robbiani et al., 2008). The oncogenic potential of AID stems from its preferential, but not exclusive,

targeting to the Ig locus. Indeed, in normal and in cancerous B cells, AID mediated deamination can lead to DSBs, of which some initiate chromosomal translocations at a large subset of genes outside the Ig loci including tumor suppressor and proto-oncogenes (Robbiani et al., 2009). On the other hand, AID deficiency leads to immunodeficiency (Revy et al., 2000), indicating the need and presence of mechanisms to maintain a balance between the two outcomes.

Research has shown that AID is tightly regulated at various levels. Expression is regulated at the transcriptional level by both enhancers and repressors located in four regions between 9 kb upstream of the first exon and 18 kb downstream of the last exon of *Aicda* gene (Tran et al., 2009). These regions form the binding sites for at least 19 transcription factors that regulate AID expression in response to extracellular signals like cytokines during a T-cell dependent immune response. Expression also seems to be regulated post-transcriptionally by at least two micro RNAs: miRNA 155 and miRNA 188b ((Teng et al., 2008), (de Yébenes et al., 2008)).

While the pre-and post-transcriptional regulation of AID might control the temporal expression of AID protein, of particular interest are post-translation regulatory mechanisms present to control the activity of the actual protein: phosphorylation, regulation of AID protein levels and the regulation of accessibility to genomic DNA. Phosphorylation of four specific amino acid residues has been identified on Ser3, Ser38, Thr140 and Tyr184. Although shown to be phosphorylated, no known function exists for phosphorylation of tyrosine at position 184. Phosphorylation of serine at 38 and threonine at 140 seem to be critical for both CSR and SHM (McBride et al., 2008), partially because S38 is required for the binding of AID to replication protein A (RPA), a

protein that binds and protects the AID substrate ssDNA (Wold, 1997). On the other hand, phosphorylation of serine 3 seems to reduce AID induced CSR and SHM, albeit marginally (Gazumyan et al., 2011), suggesting that B cells might employ phosphorylation as a method of regulating AID activity, either positively or negatively. It has been shown previously that reducing AID protein levels through haploinsufficiency can impair CSR, SHM, and chromosomal translocations (Sernández et al., 2008), suggesting that levels of AID protein is critical for maintaining the balance between the physiological and pathological effects.

Another manner in which AID seems to be regulated is by limiting its access to DNA by modulating the nucleo-cytoplasmic localization of the protein. Early reports showed that under physiological conditions AID is sequestered in the cytoplasm, and by shutting off the CRM1 mediated nuclear export machinery, AID accumulates in the nucleus ((Brar et al., 2004), (Ito et al., 2004)). Later studies showed that as-yet unknown mechanisms exist to actively sequester AID in the cytoplasm, perhaps as a regulatory mechanism to limit the enzyme's access to DNA (Patenaude et al., 2009). These mechanisms probably complement those that are involved in targeting AID to the appropriate loci. While physiological levels of AID does not seem to induce IgH-cMyc translocations, a hallmark of human Burkitt's lymphoma (T(8;14)), overexpression of AID (Takizawa et al., 2008) or knockout of genes involved in the recognition and repair of DNA double-strand breaks certainly increase the frequency of these chromosomal translocations (Ramiro et al., 2006). Indeed, it was found that approximately 25% of a subset of genes expressed, apart from the immunoglobulin loci, in primary mouse B cells isolated from the Peyer's patches (B cells from the patches are in a state of activation) are targets of AID

deamination, albeit at a significantly lower frequency than at the immunoglobulin loci. On the other hand, it was found that a vast majority of the genes targeted by AID are repaired efficiently (Liu et al., 2008). Since not all of the genes transcribed in B cells during germinal center reactions are targeted by AID, it is likely that certain protein-protein interactions target AID to specific loci. Furthermore, the efficient repair of most of the AID gene targets adds another level of protection from aberrant AID activity. One candidate targetting protein, Spt5, has been shown to bind to AID at stalled RNAPol II and to be necessary for its biological activity (Pavri et al., 2010). Nevertheless, it is not clear if Spt5 targets AID to certain loci more than others. And in any case, it seems that only ~150 of the 5000 genes targeted by AID are mutated and broken to a sufficient extent to produce chromosomal translocations (Hakim et al 2012). Thus the question ensues of what mechanism defines whether AID will mutate efficiently or not once targeted to a certain locus. One possibility are 14-3-3 class of adaptor proteins, which have several redundant functions in cells, that have been shown to bind to the C-terminus of AID, stabilize the enzyme and target it to the switch regions to initiate class-switch recombination while increasing its enzymatic activity (Xu et al., 2010).

Clearly, several questions about the regulation of AID remain to be answered. It is still not clear how the AID-mediated DNA deamination is regulated by the interplay of the mechanisms outlined above: phosphorylation, nucleo-cytoplasmic shuttling, and specific DNA targeting. It is possible to conceive that all the mechanisms work together to maintain sufficient levels of AID protein in the nucleus at the right time during the germinal center reaction and target it to the appropriate genetic loci. For instance, as mentioned above, phosphorylation at serine 38 is mediated by protein kinase A, and this

phosphorylation could occur while AID is at the immunoglobulin loci, leading to increased stability of the enzyme and/or recruitment of additional factors. Therefore, the enzyme itself needs to be better studied to identify and resolve domains involved in its regulation. As mentioned above, the C-terminus of AID has been shown to be necessary for CSR, nuclear export, and cytoplasmic retention, but the interplay between these domains is yet to be understood to better account for some of the phenotypic effects observed both *in vitro* and *in vivo*. For instance, artificial (Wang et al., 2009) and natural mutants (Barreto et al., 2003) with a high catalytic activity also show increased SHM, c-myc-IgH translocations (at least in the case of artificial mutants, (Wang et al., 2009)) and toxicity (Wu et al., 2005), again suggesting the presence of domains in the enzyme that are involved in regulating its activity; regulation that is perturbed by these mutations.

We undertook a structure-function analysis of the enzyme to resolve these issues. Early insight into AID structure-function came from the analysis of natural AID mutants found in patients with Hyper IgM immunodeficiency syndrome type 2 (HIGM2). Most of these mutations produce catalytically inactive enzymes, and are inherited as recessive traits (Revy et al., 2000). Compared to age-matched controls, these AID-deficient patients showed severely depleted levels of switched isotypes in serum and compromised SHM activity. However, a small proportion of HIGM2 patients acquire an autosomal dominant mutant of AID. They carry one normal allele and one AID allele that encodes a variant lacking the C-terminal 8 or 12 amino acids. Despite expressing a copy of the normal full-length allele of AID (Imai et al., 2005), these patients are immunodeficient, presumably from lack of switched isotypes in serum since they still have nearly normal levels of SHM in their antibodies. The fact that these truncated AID variants were transmitted in an

autosomal dominant fashion suggests that they behave as dominant negatives, specifically for CSR (Imai et al., 2005), (Ta et al., 2003); but this has never been experimentally addressed. *In vitro*, the C-terminus of AID was found to be necessary for CSR while dispensable for SHM (Ta et al., 2003). Since the truncations analyzed from residue 187 eliminated AID nuclear export signal (Brar et al., 2004; Imai et al., 2005), the requirement of nuclear export for CSR has been explored. The results show that nuclear export is not necessary per se and suggest that some CSR-specific partner must interact with that region ((Doi et al., 2009), (Geisberger et al., 2012)). Our structure-function analysis now suggests a role for the AID C-terminus in allosterically modulating the enzyme's catalytic activity, which would provide a mechanism for activating the enzyme *in situ*. Furthermore, our findings suggest an explanation for the immunodeficiency and autosomal dominant transmission in patients carrying mutations that produce truncations of this domain. In addition, we identified a novel domain within alpha helix 6 that is required for AID's ability to mutate eukaryotic, but not prokaryotic, cells. The combination of targeting and activating domains suggests a new mechanism to ensure AID specificity.

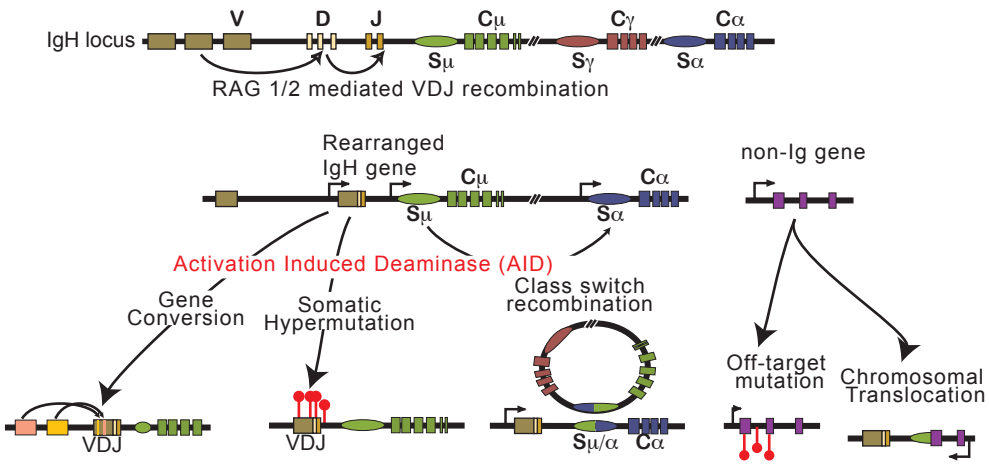


Figure I: The immunoglobulin heavy chain locus is located on chromosome 14. The approximately 39 functional VH, 27 DH gene segments and 6 JH gene segments are arranged in clusters (for clarity only a few are shown in the figure). The primary gene diversification in the bone marrow during the B cell development is initiated by RAG 1/2. Secondary gene diversification in the peripheral lymphoid tissues is mediated by AID. AID mediated deamination in the non-immunoglobulin gene could lead to off-target mutations and chromosomal translocations.

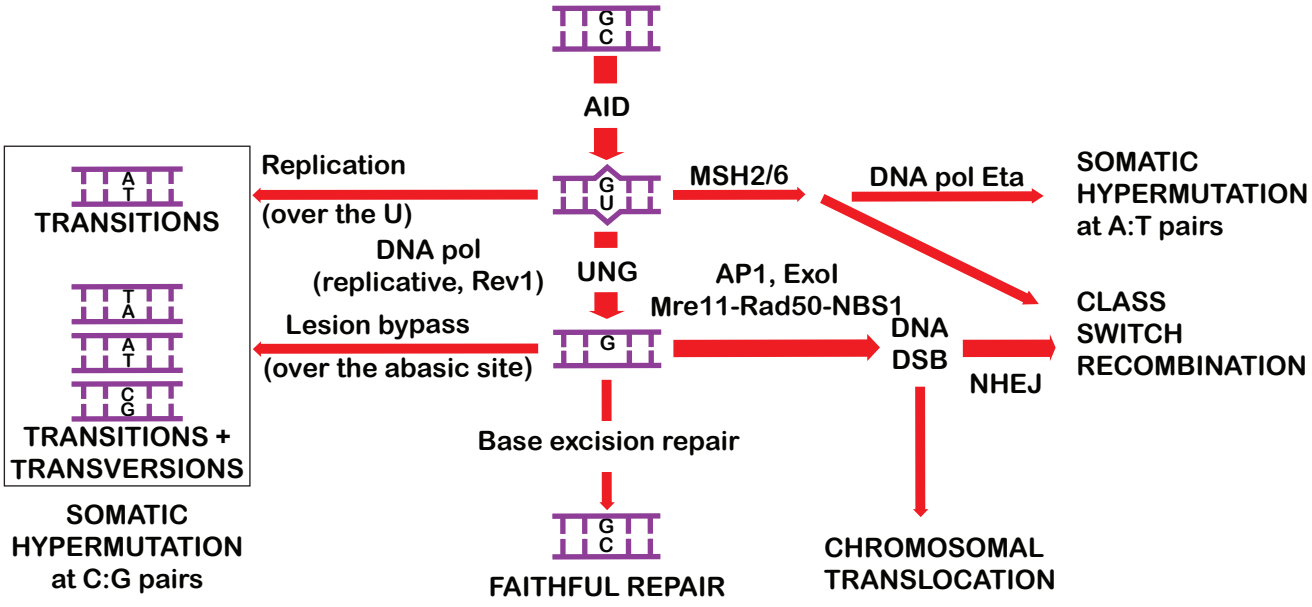


Figure II: AID-mediated deamination of deoxycytidine to deoxyuridine in ssDNA, substrate for AID, can be processed by several pathways: replication over the G:U mismatch, removal of the uridine by the glycosylase UNG, mis-match repair of the G:U mismatch. The removal of deoxyuridine by Ung is absolutely necessary for class-switch recombination.

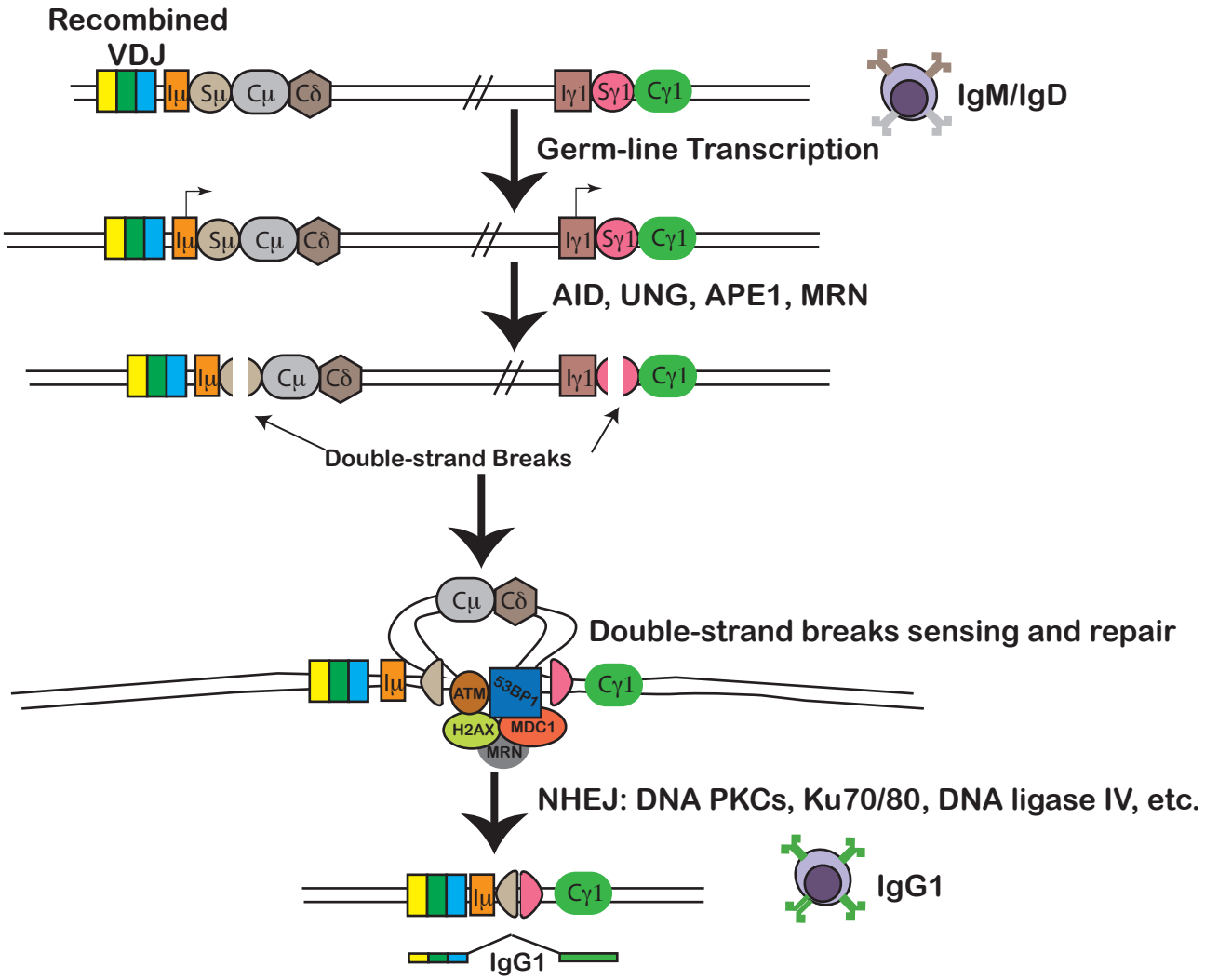


Figure III: Class switch recombination: the cascade of events that begins with AID mediated deamination of dC and leads to double strand breaks and resolution by the NHEJ machinery.

METHODS

Constructs

Quickchanges to introduce point mutations, with primer pairs listed in table 1, were performed with Pfu Turbo according to the manufacturer's instructions (Agilent Inc. Cat.# 600410). PCR amplifications of the *Aicda* open reading frame to create deletion mutants were performed with KOD DNA polymerase according to the manufacturer's instructions (Toyobo Inc. Cat.#KOD-201) The primer pairs used for the PCR reactions are listed in TABLE 1 below. The forward primers used for the PCR was either OJ60 or OJ166. C-terminal GFP tagged wild-type AID and mutants were subcloned from pEGFP-N3 (Takara Bio Inc.) with the C-terminally fused EGFP into pTrc23-24 using Nhe1 and Not1 restriction sites. N-terminal 6XHis-tagged wild-type AID and mutants were first amplified by PCR, and then cloned into the prokaryotic vector pTrcHis-A (Invitrogen). For expression of genes retrovirally, coding regions were cloned into pMx-Ires-GFP (Kitamura et al., 2003) with BamHI-XhoI sites for R190X, 1-181, and L198S with or without a Kozak sequence, which was introduced by a quickchange reaction using oligos OJ345 and OJ346. AID 1-187 was subcloned BamHI/EcoRI from pTrcHisA into pMx-Ires-GFP. AID(A2α6)ΔE5 (originally a gift from S. Conticello (Patenaude et al., 2009)) was amplified using the primer pair OAM6 and OAM13, with the Kozak and cloned into pMx-Ires-GFP. Rat APOBEC2-pEGFP-C3 (used as control in co-immunopreciptation assay) was originally a gift from Silvo (Orthwein et al., 2010))

Mice and Cell Lines

CH12, Plat-E, primary B-lymphocytes, and K562 were cultured in complete Roswell Park Memorial Institute (RPMI) medium: RPMI 1X (Wisent), 2% Penicillin/Streptomycin

(Wisent), and 0.1 mM 2-mercaptoethanol (Bioshop) in a 37°C incubator supplied with 5% CO_2. DT40 cells were cultured in complete RPMI with 1% chicken serum (Wisent) in a 41°C incubator supplied with 5% CO_2. HEK293T cells were cultured in complete Dulbecco's Modified Eagle's Medium (DMEM): DMEM (Wisent), 2% Penicillin/Streptomycin (Wisent), and 0.1 mM 2-mercaptoethanol (Bioshop) in a 37°C incubator supplied with 5% CO_2.

Table 1: List of oligonucleotides used in this study for PCR and mutagenesis (refer to methods section for more information). The primer pairs used for PCR amplifications are shaded while the pairs used for mutagenesis are not.

Mutation	**Oligos**
R171E	GGGAAGGGCTGCATGAAAATTCAGTTGAGCTCTCCAGACAGCT
	AGCTGTCTGGAGAGCTCAACTGAATTTTCATGCAGCCCTTCCC
R174S	GTTCGTCTCTCCAGCCAGCTTCGGCGCAT
	ATGCGCCGAAGCTGGCTGGAGAGACGAAC
R178A	GACAGCTTCGGGCCATCCTTTTGCCCCTG
	CAGGGGCAAAAGGATGGCCCGAAGCTGTC
R178E	CTCTCCAGACAGCTTCGGGAGATCCTTTTGCCCCTGTAT
	ATACAGGGGCAAAAGGATCTCCCGAAGCTGTCTGGAGAG
R171Y	GAAGGGCTGCATGAAAATTCAGTTTATCTCTCCAGACAGCTT
	AAGCTGTCTGGAGAGATAAACTGAATTTTCATGCAGCCCTTC
R174E	GAAAATTCAGTTCGTCTCTCCGAGCAGCTTCGGCGCATCCTTTTG

	CAAAAGGATGCGCCGAAGCTGCTCGGAGAGACGAACTGAATTTTC
R178D	CTCCAGACAGCTTCGGGACATCCTTTTGCCCCTG
	CAGGGGCAAAAGGATGTCCCGAAGCTGTCTGGAG
R177A	GTTCGTCTCTCCAGACAGCTTGCGCGCATCCTTT
	AAAGGATGCGCGCAAGCTGTCTGGAGAGACGAAC
P182A	GGCGCATCCTTTTGGCCCTGTATGAGGTTG
	CAACCTCATACAGGGCCAAAAGGATGCGCC
E185A	CCTTTTGCCCCTGGCTGAGGTTGATGACTTACG
	CGTAAGTCATCAACCTCAGCCAGGGGCAAAAGG
Y184A	CCTTTTGCCCCTGGCTGAGGTTGATGACTTACG
	CGTAAGTCATCAACCTCAGCCAGGGGCAAAAGG
Y184D	CCTTTTGCCCCTGGATGAGGTTGATGACTTACG
	CGTAAGTCATCAACCTCATCCAGGGGCAAAAGG
Y184F	GCGCATCCTTTTGCCCCTGTTTGAGGTTGATG
	CATCAACCTCAAACAGGGGCAAAAGGATGCGC
E185A	GCCCCTGTATGCGGTTGATGACTTACGAG
	CTCGTAAGTCATCAACCGCATACAGGGGC
D187A	CCTGTATGAGGTTGCTGACTTACGAGACGC
	GCGTCTCGTAAGTCAGCAACCTCATACAGG
D188A	GTATGAGGTTGATGCCTTACGAGACGCATTTC
	GAAATGCGTCTCGTAAGGCATCAACCTCATAC
D187E	CCCCTGTATGAGGTTGAGGACTTACGAGACGCATT
	AATGCGTCTCGTAAGTCCTCAACCTCATACAGGGG

D188E	CTGTATGAGGTTGATGAGTTACGAGACGCATTTCG
	CGAAATGCGTCTCGTAACTCATCAACCTCATACAG
Kozak	OJ 345: GCTAGTTAATTAAGGATCCACCATGGACAGCCTCTTGATG OJ 346: CATCAAGAGGCTGTCCATGGTGGATCCTTAATTAACTAGC
OJ60	GGATCCcaaATGGACAGCCTCTTGATGAA (BamHI site)
L198S (reverse primer)	OJ60 + CTCGAGTCAAGATCCCAAAGTACGAAATGCGT (XhoI site with stop codon)
5-198	OJ237: GGATCCCAAATGAACCGGAGGAAGTTTCTT (BamHI) OJ166: CGAATTCCCAAGTCCCAAAGTACGAAATGC (EcoRI)
27-198	OJ250: AAGGATCCCAAATGACCTACCTGTGCTACGTAGT (BamHI) OJ166: CGAATTCCCAAGTCCCAAAGTACGAAATGC (EcorRI)
1-142	OJ60 + OJ235: GAATTCCCTTTGAAGGTCATGATGGCTAT (EcoRI)
1-160	OJ60 + OJ236: GAATTCCCTTTGAAAGTTCTTTCATGGTT (EcoRI)
1-187	OJ60 + OJ211: CGAATTCCCATCAACCTCATACAGGGGCA (EcoRI)
1-181	OJ60 + OJ181: GAATTCCCCAAAAGGATGCGCCGAAGCT (EcoRI)
AID(A2α6)ΔE5	OJ60 + OJ197: GAATTCCCCTTCAGGATGTCTGCCAACTTCT (EcoRI)
1-181 (with stop codon)	OJ60 + OJ327: CTCGAGTCACAAAAGGATGCGCCGAAGCT (XhoI)
V186X (with stop codon)	OJ60 + OJ332:CTCGAGTCACTCATACAGGGGCAAAAGGAT (XhoI)
R190X (with stop codon)	OJ60 + OJ338: CTCGAGTCATAAGTCATCAACCTCATACA (XhoI)

AID(A2α6)ΔE5 (with stop codon)	OJ60 + OJ207: AAGCTTTCACTTCAGGATGTCTGCCAACTTCT (HindIII)
L198S (stop)	OJ60 + OJ 202: CTCGAGTCAAGATCCCAAAGTACGAAATGCGT (XhoI)
OAM6	AGATCTACCATGGACAGCCTCTTGATGAA
OAM13	GAATTCTCACTTCAGGATGTCTGCCAACTT

AID assays.

We assayed the catalytic activity of AID by a mutation assay described previously (Petersen-Mahrt et al) using the ung-deficient BW310 bacterial strain expressing N-terminal His-tagged full-length AID and its mutants subcloned into pTrcHisA (Invitrogen). Briefly, for each experiment, 5 independent overnight cultures each for the controls and AID mutants were grown in 2xTY medium with 100 μg/mL carbenicllin and 1 mM IPTG for 16 hours, and plated at 10-6 dilution onto 100 μg/mL ampicillin plates and undiluted onto 100 μg/mL ampicillin/rifampicin plates. Mutation frequencies were measured as the median number of colony forming units that survived rifampicin selection per 109 viable cells plated. Somatic hypermutation activity was assayed in sIgM+ DT40 line ΨV- AIDR with ablated IgV pseudogenes (gift from H.Arakawa and J.-M. Buerstedde, Institute for Molecular Radiobiology, Neuherberg, Germany; Arakawa et al., 2004) that now can only undergo somatic hypermutation when reconstituted with full-length AID. HEK293T producer cells were used to produce retroviruses to infect DT40 cells. The producer cells were seeded at 5×105 cells/well in a 6-well plate. 24 hours later, the cells were transfected with 2.5 μg pMx-Ires-GFP vectors encoding full-length and mutant AID and

vectors encoding MLV Gag-Pol and VSV-G (4:1:1 ratio) with Trans-IT LT-1 reagent (Mirus Bio LLC). 24 hours post-transfection, the medium was replaced with fresh medium and 10 mM HEPES, and 48 hours post-transfection, 1 mL supernatant from the producer cells was used to spin-infect 106 DT40 cells seeded in a 24-well plate at 600g for 60 minutes at 37°C. After the spin, the cells were incubated with the virus for 4 hours, after which the medium was replaced with fresh medium. Infected DT40 cells that were GFP+ and IgM+ were sorted and seeded at 50 cells/well in a 96-well round bottom plate (Corning), and expanded for 4-weeks. The clones were stained for IgM with mouse anti-chicken IgM-RPE (Southern Biotech) and IgM expression measured by flow cytometry. Isotype class-switching was assayed by retrovirally delivering untagged full-length AID and mutants into in-vitro activated resting B cells derived from *Aicda* +/+ and *Aicda* +/- mice, as described previously (di Noia et al. 2007). Briefly, CD43-depleted (Miltenyi Biotech) primary lymphocytes were seeded at 106 cells/well in a 24 well plate in 1 mL RPMI with 10% FBS and 0.02 mM 2-mercaptoethanol with 25 µg/mL lipopolysaccharide (LPS). Retroviruses were produced by transfecting 2.5 µg of pMx-Ires-GFP vectors into Plat-E packaging cells, seeded at 4×105 cells/well in a 6-well plate, with Trans-IT LT-1 reagent on the day of splenic extraction. 24h post-transfection, the medium was replaced with fresh medium and 10 mM HEPES. 48 hours after splenic extraction, B cells were spin-infected with 1 mL of supernatant from PLAT-E at 400g for 90 minutes with 16 µg/mL polybrene and 10 mM HEPES. After a 4 hour incubation with the virus, the medium was replaced with fresh medium and the B cells activated with 50 ng/mL IL-4 (to induce switching to IgG1) or 3 ng/mL anti-IgD-Dextran (to induce switching to IgG3). Isotype switching to IgG1 was analyzed 4 days after activation by

flow cytometry with biotin-anti-IgG1/IgG3-biotin (BD) followed by APC-conjugated anti-biotin antibody (Miltenyi Biotech) and PI on FACSCalibur (BD).

Western Blots

Protein expression of N-terminal histidine-tagged and C-terminal GFP-tagged full-length AID and mutants in BW310 strain of *E. coli* was determined from whole cell extracts (WCEs). Overnight 2xTY cultures were induced with 1mM IPTG and grown at 37°C, and WCEs obtained 12 hours later

PLAT-E lysates were obtained 48 hours post-transfection by scraping the cells off the 6-well plates, washing them twice in cold 1X PBS, and resuspending them in 1X PBS with complete protease inhibitors and boiling them for 10 minutes in SDS-loading buffer. DT40 lysates from GFP+ sorted cells were similarly washed twice in cold 1X PBS, and resuspending them in 1X PBS with complete protease inhibitors and boiling them for 10 minutes in SDS-loading buffer. AID expression from the cell lines were detected using a mouse anti human N-terminal AID primary antibody at 1:5000 dilution and a polyclonal goat anti-mouse Immunoglobulins secondary tagged to HRP (Dako). The lysates were analyzed by Western blot developed with SuperSignal West Pico Chemiluminescent substrate (Thermo Fisher Scientific).

Detection of γH2AX foci

2×105 HeLa cells were plated on sterile cover slips and incubated for 24-h. The cells were then transfected with pMx-ires-GFP retroviral vector encoding full-length AID and mutants in the presence of a consensus Kozak sequence. 48-h post-transfection, the cells were fixed in 4% PFA/PBS and permeabilized with 0.5% triton-X100/PBS. After

blocking with 5% goat serum/PBS for 1 hour, the fixed cells were incubated overnight with anti-γH2AX (Upstate JBW301 / Milipore) at 1:1000 dilution and then for 1 hour with a secondary antibody (Molecular Probes anti-mouse, Alexa 680) at 1:500 dilution. The slides were then incubated with PI (10ug/ml) and RNase A (0.2mg/ml) for 10 minutes, mounted, left overnight, and imaged with Carl Zeiss LSM510 microscope.

Imatinib resistance assay

The imatinib resistance assay was done as previously described (Orthwein et al, 2010). BCR-ABL+CML cell line K562 (a gift from Dr. T Moroy) infected with and stably expressing GFP, Kozak AID-IRES-GFP or Kozak AID 1-181-IRES-GFP were mixed with the parental cell line at a fixed ratio, and seeded in triplicates in a 24-well plate and the treated population of K562 cells were placed under 0.4 μM Imatinib. The ratio of GFP+ to GFP- was measured every 2 days by flow cytometry.

Clonal survival

CH12 or DT40 cells transiently infected with GFP, full-length AID-IRES-GFP or mutants were sorted for GFP expression and seeded at 50 cells/well into four 96-well round-bottom plates (Corning) per construct, cultured for 15 days and the number of clones that appear are scored.

Co-Immunoprecipitation

3×106 HEK-293T cells were plated onto a 10-cm dish and transfected a day later with 6 μg of DNA (pEGFP-N3 vectors encoding GFP-tagged full-length AID or mutants). Whole-cell lysates were prepared as previously described (Pavri et al., 2010). Briefly,

the cells were first lysed in a low salt buffer (20 mM HEPES, 10 mM KCl, 1mM MgCl2, 10% glycerol and 1% NP-40 alternative) with 100 µg/mL of DNaseI and RNase, followed by sonication with a probe sonicator (2 cycles at 30% amplitude, 10 second ON and 30 seconds OFF with samples on ice). The supernatant was transferred to a fresh tube and a high-salt buffer (20 mM HEPES, 10 mM KCl, 1 mM MgCl2, 10% glycerol, 1% NP-4, 400 mM NaCl) was added such that the final salt concentration was 200 mM. The pellet was subjected to another round of low-salt extraction, and the pellet from this extraction was subjected to a high-salt extraction with an appropriate volume such that the final salt concentration was 200 mM when the three supernatants were pooled.
Approximately 5% of this pooled supernatant was used as the loading control. To the rest, 30 µl of MACS magnetic anti-GFP beads (Cat. #) were added and the mixture was incubated at 4°C for 3 hours. The incubated beads were loaded into prepared magnetic columns according to the manufacturer's instructions
(MACS Cat. #130-091-125). After washing, the elute was collected and loaded into a 7% SDS-polyacrylamide gel with the input.

Statistical Analysis

Graphpad Prism 5.0d was used for statistical analysis of data using student's T test, where indicated.

RESULTS

Different structural requirements for AID mutagenicity in *E. coli* versus B cells

To identify AID domains that were important for its B cell functions, we performed structure-function analysis using several AID variants (Figure 1 A) that we compared for their ability to mutate *E. coli* and the Ig genes. We designed truncations using a structural model of AID (Patenaude et al., 2009), but also used natural AID mutants from HIGM2 patients, and a catalytically active chimeric protein of AID and APOBEC2 (an AID paralog without DNA deaminase activity (Harris et al., 2002)). We used nested truncations to determine first the minimal AID fragment that could mutate the *E. coli* genome. The frequency of rifampicin-resistant cfu that arise as a consequence of mutations in rpoB serves as a faithful reporter of the AID DNA deaminase activity (Coker et al., 2006). We used AID-GFP fusions in order to compare the expression of the truncated proteins by Western blot (no available anti AID would recognize all the variants); and to use the GFP signal, measured by flow cytometry, as a proxy for the solubility of each AID variant in live bacteria (precipitated AID would inactivate the fused GFP). Deletion of the first 26 N-terminal amino acids completely ablated AID mutagenic activity in *E. coli* (Figure 1 B). In contrast, deleting the C-terminal nuclear export signal (NES) or the whole exon 5-encoded C-terminal domain (Supplementary Figure S1A, hereafter referred to as E5) increased this activity. Longer truncations such as AID 1-160 were inactive (Figure 1 B). Amino acids 160 to 181 harbor no catalytic site residues but contain a predicted α-helix 6. Substituting APOBEC2 α6 into AID 1-181 produced a highly mutagenic AID(A2α6)-ΔE5 enzyme (Figure 1 B). Given the low sequence conservation between AID and APOBEC2 in this region (Supplementary Figure S1A) this likely reflects a structural requirement for α6 in allowing AID activity. Differences in

protein expression or solubility did not account for the relevant differences in activity (Supplementary Figure S1B, C). Moreover, the 3-fold increase in *E. coli* mutation for AID 1-181 nicely correlated with the 3-fold increase in specific activity of this truncated protein over full length AID that was previously reported (Kohli et al., 2009). We conclude that the region between residues 6 and 181 is the minimal AID enzymatic domain with truncation of the C-terminus leading to higher *E. coli* mutagenic activity and a region homologous to APOBEC2 α6 being structurally important.

The *E. coli* assay bypasses AID regulation by subcellular localization and post-translational modifications and therefore the extent to which an AID variant mutates *E. coli* may not directly correlate with its capacity for SHM. We therefore tested those AID variants that were hyperactive in *E. coli* for their ability to mutate the Ig genes in B cells. To eliminate possible effects of the C-terminal GFP tag on AID's activity, we did the *E. coli* mutation assay using a small N-terminal His-tag. We extended the analysis to those AID variants found in HIGM2 patients that were mutated within E5 or a6 because the molecular basis of the defect in these patients is unknown (see below). In parallel, untagged versions of the same AID variants were expressed into a SHM reporter line of the DT40 chicken B cell lymphoma. This DT40 line lacks AID and the IgVλ pseudogenes that act as gene conversion donors so that complementation with AID leads to the accumulation of SHM in the IgVλ (Arakawa, 2004). A proportion of these mutations inactivate the Ig gene and therefore the median proportion of IgM-loss cells appearing over time in multiple populations is proportional to the SHM activity (Barreto et al., 2003). The AID variants could be divided into two different categories according to these assays (compare Figures 1 C and D). On one side, AID 1-181 and the autosomal dominant AID variants R190X and V186X, which have a truncated E5 (Patenaude et al.,

2009) showed higher *E. coli* mutagenic activity than wt AID and a proportionally higher SHM activity. On the other side, AID(A2α6)-ΔE5, which is 4-8-fold more mutagenic in *E. coli* than wt AID, and AID R174S, an autosomal recessive variant with a single amino acid change in the α6 that retains ~70% of wt AID activity in *E. coli*, were extremely inefficient for SHM. These results are yet to be confirmed by sequencing the IgVλ region from GFP+ sIgM- DT40 cells and the *rpoB* gene from rifampicin-resistant BW310 colonies (Coker et al., 2006).

Our results revealed increased ability to mutate *E. coli* and B cells for AID partial or complete C-terminal truncations and a striking difference between the structurally similar AID 1-181 and AID(A2α6)-ΔE5 in their ability to perform SHM ability that can neither be ascribed to differences in catalytic activity, nor to nuclear access (Patenaude et al., 2009). This suggests an inhibitory role for the AID C-terminus and the presence of residues within α6 that are important for AID B cell functions, which is supported by the similar phenotype of AID R174S.

Dominant negative effect on CSR requires enzymatic activity

We went on to compare the behavior of the AID variants in other antibody diversification-related processes. Unfortunately, AID variants with C-terminal truncations, including AID R190X and V186X, are unable to perform CSR ((Barreto et al., 2003), (Ta et al., 2003)). However, the postulate that they may behave as dominant negatives for CSR (Ta et al., 2003) was never experimentally tested. We developed an assay to test this suggestion in which AID variants were retrovirally transduced into LPS-activated B cells from mice either wild-type or heterozygous for AID. The infected cells were then induced to undergo CSR either to IgG1 (by adding IL-4), or to IgG2b (by

adding TGFβ1), so as to synchronize the activity of endogenous and transduced AID at the same downstream switch region. We normalized the switching obtained in the infected and the uninfected cells to that of the GFP control; any reductions in switching in the GFP+ population presumably indicating dominant negative effects mediated by the mutated AID (Figure 2A). In these conditions, AID was saturating for IgG1 switching even in *Aicda*+/- cells as transduced AID did not further increase CSR (Figure 2 B). However, AID was limiting for switching to IgG2b in the same cells and AID overexpression increased CSR in the infected cells (Figure 2 C). In any case, and regardless of the system, a similar dominant negative effect as before could be observed in these assays. For both IgG1 and IgG2b, all C-terminally truncated AID variants reduced the efficiency of CSR in the infected cells while neither control infection, full-length AID, R174S or AID(A2α6)-Δ5 variants had this effect (Figure 2 B and C and Supplementary Figure S2). However, the effect was lost when the AID 1-181 was made catalytically inactive by mutating E58A. Importantly, E58A does not cause any major structural perturbations since it does not prevent Zn++ coordination (Uchimura et al., 2011), nor does it interfere with DNA binding (Figure 2 D). This suggested that we were not seeing a classical dominant negative effect (i.e. due to competition for substrate or other factors) but rather it was a consequence of the enhanced mutagenic capacity of the truncated mutants.

We asked whether any artificial AID variants with higher mutagenic activity could produce this apparently dominant negative effect. For this, we compared AID L198S and AID D187A, both of which produce similarly higher levels of SHM than wt AID (Figure 2 E), presumably as a consequence of their nuclear enrichment (Supplementary Figure S3). AID L198S has a non-functional NES ((McBride et al., 2004), (Patenaude et al.,

2009)) and AID D187A has a mutation in the cytoplasmic retention motif (Patenaude et al., 2009). They differ in their ability to perform CSR; AID L198A has been shown to be inactive for CSR ((Doi et al., 2009), (Geisberger et al., 2012)) and so is L198S (this study) while D187A is active for CSR (Häsler et al., 2011). Interestingly, only AID L198S, but not D187A, showed the apparent dominant negative effect on CSR (Figure 2 F). However, they also differed in their ability to mutate *E. coli*, with AID L198S being 2-fold more, and AID D187A ~25% less, active than AID in the rifampicin assay (Figure 2 F). Nevertheless, AID 7.3, an artificial mutant selected for its enhanced ability to mutate *E. coli* (Wang et al., 2009), did not behave as a dominant negative construct. Interestingly, AID 7.3 is at least 3-fold more active than wild-type AID for SHM but unlike AID L198S, it is still proficient for CSR (Wang et al., 2009). We conclude that the ability to confer dominant negative effects in the CSR assays requires an AID variant with an enhanced ability to mutate both *E. coli* and the Ig genes and deficient for CSR.

Increased off-target mutations by C-terminally truncated AID and α6 requirement

We further explored the possible causes behind the dominant negative effect as well as the role of the AID α6 region by testing the ability of AID variants to mutate non-Ig genes. We used the BCR-ABL oncogene, which encodes an oncogenic kinase present in most chronic myelogenous leukemia (CML) patients, as a reporter. In CML patients, AID mutates BCR-ABL thereby generating variants that are resistant to the inhibitor imatinib (Klemm et al., 2009). We transduced the BCR-ABL+ CML cell line K562 with the different variants using a Kozak AID-IRES-GFP vector. Infected cells were mixed 1:1 with non-infected K562 cells and the cultures expanded for 30 days in the presence of imatinib (Figure 3A). As previously demonstrated ((Orthwein et al., 2010),(Klemm et al.,

2009)), expression of AID led to an increase in the GFP+/GFP- cell ratio starting at day ~18 due to the selective advantage of imatinib-resistant cells. Cells expressing AID 1-181 showed a much earlier increase, at day ~9. However, in this case the AID 1-181 GFP+ cells stopped increasing one week later and disappeared over time (Figure 3 A). This decrease was also visible in cells expressing AID but occurred much later and allowed for a much higher GFP+/GFP- ratio. Performing the competitive assay in the absence of imatinib selection showed that AID 1-181 had a strong negative effect on cell fitness resulting in rapid loss of GFP+ cells over time (Figure 3 B). This was consistent with the observed expression instability of AID 1-181 in mouse or chicken B cell lines, which was again a consequence of its enzymatic activity since AID 1-181 E58A was easily tolerated by the cells (Figures 3 C and D). In stark contrast, AID(A2α6)-Δ5 expression had no effect on imatinib resistance in K562 cells, nor did it had any impact on competitive cell growth (Figure 3 B) and its expression in B cells was completely stable (Fig 3 E). We conclude that AID variants with CSR dominant negative capacity, but not the enzymatically hyperactive AID(A2α6)-Δ5, show higher off-target mutation frequency and decrease cell fitness.

CSR dominant negative AID variants are genotoxic and impair clonal expansion

Higher levels of off-target mutation by AID correlate with more DNA damage ((Robbiani et al., 2009),(Hakim et al., 2012)). We compared the capacity of AID variants to produce DNA double strand breaks by monitoring the ability to induce γH2AX foci when expressed in HeLa cells. Once again we could distinguish the same two groups: AID 1-181 and R190X produced significantly more DNA breaks than AID while neither AID R174S nor AID(A2α6)-Δ5 showed any increase (Figure 4 A). Underlining the correlation

between dominant negative effect for CSR and DNA damage, AID L198S increased the number of γH2AX foci while AID D187A did not (Figure 4 B). As expected, genotoxicity correlated with a detrimental effect on growth kinetics. This was evidenced as delayed cell division of retrovirally transduced primary B cells (Figure 4 C). It was also visible from growth curves of DT40 stable cell transfectants when started at low cellular densities (Figure 4 D), but not at higher cell densities (not shown). Since immune responses require clonal expansion of individual activated B cells we mimicked this more physiological situation by performing clonal growth assays using mouse and chicken B cell lines expressing the different AID variants. All natural and artificial dominant negative AID variants, but not the autosomal recessive R174S, severely compromised cloning efficiency (Figure 4 E).

Autosomal dominant HIGM2 AID variants do not cause lymphadenopathy

AID-deficient B cells show increased proliferation and reduced apoptosis (Zaheen et al., 2009) thus explaining at least in part the large germinal centers and probably the B cell and lymphoid hyperplasia observed in AID-deficient mice and humans ((Muramatsu et al., 2000),(Revy et al., 2000)) and even in AID haploinsufficient mice (Sernández et al., 2008). However, our results would suggest that the genotoxic effect of AID autosomal dominant variants could counteract this hyperproliferation. This was indeed consistent with the fact that only 2 out of 11 (18%) HIGM2 patients carrying the autosomal dominant AID variants showed lymphadenopathies, compared to 75% incidence in autosomal recessive patients (Table 2) ((Quartier et al., 2004), (Minegishi et al., 2000), and personal correspondence with Dr. Anne Durandy). AID-deficient HIGM2 patients have no CSR or SHM ((Revy et al., 2000), (Ta et al., 2003)). However, autosomal

dominant HIGM2 patients have impaired CSR but show normal levels of SHM (Imai et al., 2005). Interestingly, both autosomal dominant HIGM2 patients presenting lymphadenopathies were also defective for SHM, suggesting that they may have an additional defect associated with the truncated AID and that the mutagenic activity of the truncated AID forms may prevent the occurrence of lymphadenopathies. Thus, in line with our results *in vitro*, the absence of lymphadenopathies in autosomal dominant HIGM2 patients correlates with the ability to perform SHM. All together our results indicate that the dominant transmission of truncated AID variants in HIGM2 patients is caused by their detrimental effect on B cell growth rather than by a classical dominant negative effect on the co-expressed wild-type AID.

The C-terminal region of AID modulates its catalytic activity

One possible explanation for the increased activity of C-terminally truncated AID was that this is a self-inhibitory domain. Interestingly, several single point mutations within AID E5 were able to either increase or decrease the ability of AID to mutate *E. coli*, depending on the residue mutated and the amino acid introduced (Figure 5). Since, E5 is dispensable for activity and presumably far away from the active site, we hypothesize that the mutations may be displacing this domain, causing conformational changes that affect the enzymatic activity.

AID α6 is functionally important

Finally, we explored the basis for the difference in biological activity between AID 1-181 and AID(A2α6)-ΔE5. It was surprising that while both were highly mutagenic in *E. coli* (Figure 1) and abundantly nuclear ((Brar et al., 2004), (Patenaude et al., 2009)),

AID(A2α6)-ΔE5 showed little ability to mutate or damage the eukaryotic genome in our multiple assays. As a further demonstration of their different behavior in antibody diversification we used another sensitive assay for SHM, in which AID variants were transduced into LPS + IL-4-stimulated *Aicda*-/- *Ung*-/- primary B cells. The UNG-deficiency prevents switching to a large extent but allows substantial accumulation of mutations at the Sμ switch region ((Longerich, 2005), (Liu et al., 2008)). In this assay, AID 1-181 produced twice as many mutations than wt AID while AID(A2a6)-DE5 introduced very few mutations (Figure 6A). Despite their predicted structural similarity, which is supported by the fact that substituting APOBEC2 α6 for AID α6 restores catalytic activity (Figure 1), comparison of α6 between AID and APOBEC2 showed a clear difference in surface charge (Figure 6 B-D). Modeling AID α6 predicts 4 solvent-exposed Arg residues, including R174 that is mutated in some autosomal recessive HIGM2 patients (Quartier et al., 2004), (Revy et al., 2000). In contrast, APOBEC2 exposes mostly acidic residues over the same region. Alanine substitution or charge inversion mutation of R171, R174 or R178 preserved the capacity to mutate *E. coli* but severely reduced CSR and SHM (Figure 6 C). This functional deficiency could not be explained by defects in nuclear accumulation (Supplementary Figure S3). Similar results were obtained when mutating these Arg residues to the corresponding APOBEC2 residue (Figure 6 F-H). Importantly, there was some specificity as only Arg 171, 174 and 178 were required for AID biological activity while mutations at Ser 173 and Arg 177 maintained *E. coli*, SHM and CSR activity (Figure 6 F-H).

AID α6 is contiguous to the E5 region of AID, which is required for CSR ((Barreto et al., 2003), (Ta et al., 2003), (Geisberger et al., 2012)). The reason for this is unclear and truncation of the C-terminus still allows SHM but it was nevertheless conceivable that E5

and α6 could be part of the same structural domain so we explored this possibility. We made a new construct, AID(A2a6), which bore the α6 replacement but contained the E5 region (Figure 6 E); which we compare to its parental variants AID(A2a6)-ΔE5 and AID 1-181. Unlike the hyperactive AID(A2α6)-ΔE5, adding back the E5 region reduced the mutagenic activity in *E. coli* back to wt AID levels but it did not rescue CSR activity (Figure 6 F). Of note, AID 1-181 was very inefficient but still able to elicit some CSR, while exchanging the α6 prevented even this small amount of switching. We conclude that the E5 region is necessary but not sufficient for CSR and that the AID mutants affecting α6 identify a novel functional domain that is required for both CSR and SHM. Furthermore, as it is also required for off-target mutations and to produce DNA damage, but dispensable to mutate *E. coli*, we conclude that this domain is likely required for targeting AID to the eukaryotic genome.

AID α6 is necessary to bind to Spt5

It has been shown previously that AID can only deaminate cytidine residues on ssDNA substrates (Bransteitter et al., 2003). In support of this, it was subsequently shown that AID associates with RNA Poll II, suggesting that AID perhaps accesses ssDNA during transcription (Nambu et al., 2003). More recently, however, Spt5 was identified as a possible factor responsible for targeting AID to the stalled RNA Pol II (Pavri et al., 2010). Given the importance of α6 for AID-mediated SHM, we tested if this helix is necessary for mediating AID binding to Spt-5. We expressed GFP-tagged full-length AID, AID 1-181 and the chimeric protein AID(A2α6)ΔE5 in HEK293T cells. We show that while full-length AID and AID 1-181 co-immunoprecipitated with Spt5, AID(A2α6)ΔE5 could not (Figure 7).

DISCUSSION AND CONCLUSIONS

We have performed a structure-function analysis of AID and report two major findings. First, we systematically explored the role of the AID C-terminal domain in physiological and pathological AID contexts and suggest that it possibly acts as a self-inhibitory domain for AID catalytic and mutagenic activity. Second, we characterized a hitherto unnoticed domain in α6, which allows AID to access the eukaryotic genome. Both findings together provide an insight into the mechanism of action of this mutagenic enzyme in eukaryotic cells and have implications for AID targeting to the Ig loci.

AID C-terminal domain can be structurally defined as starting at proline 182, immediately after the predicted α6. This region is encoded by *Aicda* exon 5 (Figure 1A) and is absent in APOBEC2 and APOBEC3, which end with α6 (Bransteitter et al., 2009). Interestingly, APOBEC1, which normally edits RNA but has been shown be active on DNA (Harris et al., 2002), has a completely different and functionally unique C-terminal region, encoded by exon 5, which was acquired during APOBEC1 evolution (Severi et al., 2011). In AID, this region has long been known to be required for CSR but is very well conserved in the gnathostome vertebrates down to at least osteichthyes (in which CSR is largely absent) suggesting it evolved to also perform other functions.

The structure-function analysis revealed that AID mutants lacking the C-terminal domain displayed a significantly higher catalytic activity and somatic hypermutation activity (Figure 1B, C, and D), in line with what has been described before (Barreto et al., 2003). We then tested the spontaneous mutants transmited in an autosomal dominant manner by a subset of HIGM2 patients (Muramatsu et al., 2000), V186X and R190X, and

indeed, similar to truncation of the entire C-terminal domain, these patient mutations displayed high catalytic and somatic hypermutation activity. On the other hand, patient-derived AID mutants that are inherited in a recessive manner, namely R174S (Quartier et al., 2004), display reduced, but not absent, catalytic and somatic hypermutation activity. Our analysis also revealed that α6 is important to maintain the catalytic activity of the enzyme since the variants lacking this helix are inactive in *E. coli* (Figure 1B). Furthermore, AID's α6 can be replaced with that from APOBEC2 resulting in the chimeric protein AID(A2α6)ΔE5, which replicates the increased catalytic activity observed with the truncated AID variants (Figure 1 B and C). Given the sequence dissimilarity between the AID's and APOBEC2's α6 (Figure 5E), the need for α6 is presumably to maintain the structural integrity of the enzyme allowing it to deaminate the DNA. Surprisingly however, this chimeric protein, as opposed to the AID truncations, is comparably much less active for somatic hypermutation (SHM). This suggests that the α6 of AID is important for both maintaining the structure of AID and perhaps also to target AID to the appropriate genomic loci. These results are particularly useful in addressing a long-standing question regarding the autosomal dominant inheritance of these two mutants, R190X and V186X. Plausible explanations that have been proposed include the possibilities that they act as dominant negatives and/or that they are cytotoxic ((Ta et al., 2003), (Barreto and Magor, 2011)) but this had not been experimentally addressed.

To better understand this phenomenon *in vitro*, we devised an experimental method where inactive CD43- B cells from mice either wild-type or heterozygous for AID are

infected with several AID mutants and activated to switch to IgG1 or IgG2b. We observe that when compared to the uninfected cells and cells infected with full-length AID, cells infected with truncated AID mutants (AID 1-181, V186X, and R190X) show a significantly reduced levels of CSR (Figure 2A, B). The endogenous AID in the cells infected with the truncated variants are somehow inhibited from performing their activity, replicating the apparent dominant negative effect of the truncated variants observed in HIGHM-2 patients. Since the truncated variants are inactive for CSR ((Ta et al., 2003), (Shinkura et al., 2004)), one possibility is that there could be an "insufficient dose" of AID in cells infected with the truncated variants and not in those infected with full-length AID. However, cells infected with the full-length AID bearing the catalytically inactive mutation E58A, R174S or the chimeric protein AID(A2α6)ΔE5 (all of which are inactive for CSR) do not have a dominant negative effect, suggesting that this effect is due to an inherent property of the truncation mutants. Given the high activity displayed by the truncation mutants (Figure 1), we then extended our analysis to other artificial mutants with high somatic hypermutation activity and/or catalytic activity, to see if we could make similar observations. We selected four mutants with varied catalytic and hypermutation activities. The D187A point mutant displayed ~50% catalytic activity relative to wt AID, while the other three mutants (Y184A, L198S, and the upmutant AID 7.3 (Wang et al., 2009)) showed a 2-fold higher catalytically activity than wt AID (Figure 2G); comparable to the activities of the truncation mutants. Also, all the mutants, except for Y184A, had a relatively high SHM activity, (Figure 2E and (Wang et al., 2009) for AID 7.3). However, only the point mutant L198S, whose protein product is fully nuclear (Patenaude et al., 2009), displayed an apparent dominant negative effect. On the other hand, AID upmutant 7.3 which has been shown to have a high catalytic (Figure 2G) and SHM

activity, is also active for class switch recombination (Wang et al., 2009), although its localization has not been studied.

Therefore, only mutants that are significantly nuclear with high catalytic and somatic hypermutation activities along with an inability to perform class switch recombination (AID 1-181, V186X, R190X, and L198S) seem to display an apparent dominant negative effect in the CSR assay. Several hypotheses have been previously offered to explain this effect such as by titrating out of the full length AID at the Ig locus by the truncated protein due its enhanced nuclear localization ((Ito et al., 2004), (Patenaude et al., 2009)) of the truncated AID variants and/or by interfering for instance with oligomerization (Durandy et al., 2007). But, given that inactivating the catalytic activity of AID 1-181 by the E58A point mutation abrogated the effect (Figure 2 B and 2C), we suggest that the high catalytic activity and SHM activity, and possibly the resulting genotoxicity, which has been previously shown for AID 1-181 (Wu et al., 2005), of these mutants could explain this effect at least in part, and also provide a reason for the autosomal dominant inheritance of the truncation mutants.

The genotoxic potential of AID 1-181 is displayed when it is expressed in the CML-cell line K562. Upon treatment with imatinib, cells infected with AID 1-181 show a much earlier resistance than those infected with AID, courtesy of the increased off-target activity of the truncation mutant. The genotoxicity is better observed in the absence of any selection pressure. Over time, K562 cells infected with AID 1-181 lose the GFP signal more rapidly than those infected with the full-length protein, a direct result of its cytotoxic effect (Figure 3A, B). This effect is replicated in other cell lines as well (Figure

3B-D). The true question, however, is if the increased off-target activity (Figure3A) and resulting double-strand breaks (Figure 4A,B) translates into increased apoptosis, as has been suggested for AID 1-181 before (Wu et al., 2005). We could not detect any significant difference in apoptosis between mouse primary B cells infected with full-length AID and AID 1-181 (not shown). However, the significantly high background of apoptosis of the primary B cells might perhaps might have masked any true differences. Once a circulating mature B cell in the peripheral lymphoid organs recognizes its cognate antigen, signaling events, both endogenous and exogenous, prompt this single B cell to proliferate and undergo a secondary gene diversification process. We show here that DT40 cells infected with the autosomal dominant mutants have a severe defect in clonal expansion, while the recessive mutants do not (Figure 4D). Taken together, the apparent dominant negative effect and the autosomal dominant inheritance of the truncated mutants observed in the HIGM-II patients can be attributed to the high catalytic, SHM, and off-target activities and the resulting DNA damage and clonal expansion defects of the B cells expressing these mutants. Furthermore, the clinical data support this conclusion (Supplementary Figure 4). (Zaheen et al., 2009) showed that AID-/- mice have more splenic B cells and larger germinal centers than wild-type mice, and they suggested that AID predisposes the B-cells to apoptosis. Although we could not observe increased apoptosis in cells expressing AID 1-181 over wild-type AID *in vitro*, it can still be hypothesized that the cytotoxicity of the truncated protein could result in increased apoptosis of cells *in vivo* compared to wild-type B cells. Indeed, while HIGM-II patients with autosomal recessive mutations in *Aicda* develop lymphoid hyperplasia ((Quartier et al., 2004), (Minegishi et al., 2000)), similar to AID null mice, HIGM-II patients with autosomal dominant mutations do not (Table 2).

It was shown in (Barreto et al., 2003) and in this study that removal of the C-terminal domain in AID increases its activity in both *E. coli* and DT40 cells. However, no clear explanation for this has been offered. Given that the truncated protein has a significantly higher specific activity than the full-length protein (Kohli et al., 2009), we suggest that the increased catalytic and SHM activity (Figure 1B, C, and D) we observed points to an inherent modulatory role for the C-terminal domain. This idea is supported by the unexpected result that point mutations within E5 affect the catalytic activity, with some point mutations increasing and others decreasing the enzyme's catalytic and switching activity (Figure 6) while the complete loss of the domain does not abrogate its activity. Furthermore, when the C-terminal domain of AID is fused to the highly catalytically active (Figure 1B and C) chimeric protein AID(A2α6)ΔC the activity of the resulting protein is reduced to wild-type AID levels, supporting the idea that the C-terminal domain has a modulatory role (Figure 6E).

It was surprising that AID(A2α6)ΔC, which displayed a relatively high mutagenic activity in *E. coli* (Figure 1), had much reduced somatic hypermutation activity. On the other hand, mutants lacking part or all of the C-terminus and L198S were highly active in both *E. coli* and DT40. This suggested that AID α6 holds functional information that is crucial for efficient SHM. However, one major difference between AID(A2α6)ΔE5 and AID 1-181 is that the former has none of AID subcellular localization determinants but is able to diffuse through the nuclear pores and is therefore homogeneously distributed throughout the cell and abundant in the nucleus (Patenaude et al., 2009). So, it is not clear, at the moment, if active nuclear import, which AID(A2α6)ΔE5 lacks, is necessary

for proper “licensing” of the protein to target DNA, or if the residues on α6 are important for binding to some factor that targets AID to the proper loci in activated B cells (Hakim et al., 2012). This is in line with the fact that patients homozygous for the α6 point mutation R174S have no SHM or CSR (Quartier et al., 2004), which would suggest that this variant is enzymatically inactive, but here we show that this is not the case (Figure 1C). It can be hypothesized that the negatively charged arginines on this helix could be involved in AID targeting, because while AID 1-181 can mutate the switch region at a rate higher than the full-length AID, AID(A2α6)ΔE5 cannot (Figure 6A). We explore this possibility by changing the positive charge of the arginines on α6 to a negative charge. Interestingly, the charge of the three arginines, R171, R174 and R178, all seem to be important for CSR and SHM activities, but not for catalytic activity (Figure 6 B-D). Analysis of the subcellular localization at the steady state, and of nuclear import after treatment with leptomycin B, of AID bearing R171E, R178E or R174S did not show any visible difference with wild-type AID (Supplementary Figure 3) We cannot rule out small differences in kinetics for nuclear transport but it would be surprising that these would cause the major defect we observe in CSR in retroviral complementation assays, which can bypass much stronger AID nuclear import defects (Shinkura et al., 2004). We then proceeded to mutate the same arginines to the corresponding residues in APOBEC2, based on sequence alignment (Figure 6E). Again, mutating R171,R174 or R178, resulted in the loss of SHM and CSR (Figure 6F-H). Importantly, mutating the residues R177 or S173, both on α6, did not affect CSR or SHM, supporting the idea that the three critical residues alone perhaps form a surface that could be important for mediating AID interaction with another factor necessary for targeting. One such possible factor is Spt5, which has been shown previously to associate with stalled RNA polymerase-II and

necessary for recruiting AID to both Ig and non-Ig targets (Pavri et al., 2010). Indeed, we show here that while GFP-tagged AID and AID 1-181 can co-immunoprecipitate with Spt-5, AID(A2α6)ΔE5 does not (Figure 7).

Based on our results, we propose a novel model in which the targeting of AID involves two steps: first, AID is recruited to chromatin enriched in certain factors, like Spt5, that could interact with AID α6 region. Second, in situ activation (through phosphorylation, for instance) or enhancement of AID mutagenic capacity by proteins interacting with and/or modifying AID C-terminus initiates the deamination process, and subsequently SHM or CSR. Potential C-terminal domain-interacting proteins are members of the14-3-3 class of proteins. It was previously demonstrated that 14-3-3 proteins directly bind to the C-terminus of AID and recruit it to the switch regions (Xu et al., 2010).The loss of such interaction(s) by the autosomal dominant mutants lacking part of the C-terminus could potentially result in constitutive excessive activity of these mutants at unwanted genomic locations leading to increased DNA damage and negatively affecting the cell fitness and ability to undergo clonal expansion, which could at least in part explain the phenotypes observed both clinically and *in vitro*.

REFERENCES

Arakawa, H. (2004). Immunoglobulin gene conversion: Insights from bursal B cells and the DT40 cell line - Arakawa - 2004 - Developmental Dynamics - Wiley Online Library. Developmental Dynamics.

Barreto, V., Reina San-Martin, B., Ramiro, A.R., McBride, K.M., and Nussenzweig, M.C. (2003). C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Molecular Cell 12, 501–508.

Barreto, V.M., and Magor, B.G. (2011). Activation-induced cytidine deaminase structure and functions: a species comparative view. Dev. Comp. Immunol. 35, 991–1007.

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.

Bransteitter, R., Prochnow, C., and Chen, X.S. (2009). The current structural and functional understanding of APOBEC deaminases. Cell. Mol. Life Sci. 66, 3137–3147.

Brar, S.S., Watson, M., and Diaz, M. (2004). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. J. Biol. Chem. 279, 26395–26401.

Coker, H.A., Morgan, H.D., and Petersen-Mahrt, S.K. (2006). Methods in Enzymology (Elsevier).

de Yébenes, V.G., Belver, L., Pisano, D.G., González, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine

deaminase in B cells. Journal of Experimental Medicine 205, 2199–2206.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular Mechanisms of Antibody Somatic Hypermutation. Annu. Rev. Biochem. 76, 1–22.

Doi, T., Kato, L., Ito, S., Shinkura, R., Wei, M., Nagaoka, H., Wang, J., and Honjo, T. (2009). The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. Proceedings of the National Academy of Sciences 106, 2758–2763.

Durandy, A., TAUBENHEIM, N., PERON, S., and Fischer, A. (2007). Advances in Immunology. 94, 275–306.

Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E.M., Chait, B.T., Reina San-Martin, B., Nussenzweig, M.C., and McBride, K.M. (2011). Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. Mol. Cell. Biol. 31, 442–449.

Geisberger, R., Huemer, M., Gassner, F.J., Zaborsky, N., Egle, A., and Greil, R. (2012). Lysine Residue at Position 22 of the AID Protein Regulates Its Class Switch Activity. PLoS ONE 7, e30667.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.-R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature 484, 69–74.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Molecular Cell 10,

1247–1253.

Häsler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1α (eEF1A). Proceedings of the National Academy of Sciences 108, 18366–18371.

Imai, K., Zhu, Y., Revy, P., Morio, T., Mizutani, S., Fischer, A., Nonoyama, S., and Durandy, A. (2005). Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. Clin. Immunol. 115, 277–285.

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc. Natl. Acad. Sci. U.S.a. 101, 1975–1980.

Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T., and Kumagai, H. (2003). Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. Exp. Hematol. 31, 1007–1014.

Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., Levetzow, von, G., Feldhahn, N., Henke, N., Li, Z., Hoffmann, T.K., Kim, Y.-M., et al. (2009). The B Cell Mutator AID Promotes B Lymphoid Blast Crisis and Drug Resistance in Chronic Myeloid Leukemia. Cancer Cell 16, 232–245.

Kohli, R.M., Abrams, S.R., Gajula, K.S., Maul, R.W., Gearhart, P.J., and Stivers, J.T.

(2009). A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J. Biol. Chem. 284, 22898–22904.

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

Longerich, S. (2005). The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. Journal of Experimental Medicine 202, 1443–1454.

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J. Exp. Med. 199, 1235–1244.

McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. Journal of Experimental Medicine 205, 2585–2594.

Minegishi, Y., Lavoie, A., Cunningham-Rundles, C., Bédard, P.M., Hébert, J., Côté, L., Dan, K., Sedlak, D., Buckley, R.H., Fischer, A., et al. (2000). Mutations in activation-induced cytidine deaminase in patients with hyper IgM syndrome. Clin. Immunol. 97, 203–210.

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.

Nambu, Y., Sugai, M., Gonda, H., Lee, C.-G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. Science 302, 2137–2140.

Orthwein, A., Patenaude, A.-M., Affar, E.B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. J. Exp. Med. 207, 2751–2765.

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat. Struct. Mol. Biol. 16, 517–527.

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.

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, 99–103.

Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., and Chen, X.S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature 445, 447–451.

Quartier, P., Bustamante, J., Sanal, O., Plebani, A., Debré, M., Deville, A., Litzman, J.,

Levy, J., Fermand, J.-P., Lane, P., et al. (2004). Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. Clin. Immunol. 110, 22–29.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature 440, 105–109.

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.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., et al. (2009). AID Produces DNA Double-Strand Breaks in Non-Ig Genes and Mature B Cell Lymphomas with Reciprocal Chromosome Translocations. Molecular Cell 36, 631–641.

Sernández, I.V., de Yébenes, V.G., Dorsett, Y., and Ramiro, A.R. (2008). Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both in vitro and in vivo. PLoS ONE 3, e3927.

Severi, F., Chicca, A., and Conticello, S.G. (2011). Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. Mol. Biol. Evol. 28, 1125–1129.

Shinkura, R., Ito, S., Begum, N., and Nagaoka, H. (2004). Separate domains of AID are

required for somatic hypermutation and class-switch recombination - Nature Immunology. Nature.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and Regulation of Class Switch Recombination. Annu. Rev. Immunol. 26, 261–292.

Ta, V.-T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., et al. (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat. Immunol. 4, 843–848.

Takizawa, M., Tolarova, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W., et al. (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. J. Exp. Med. 205, 1949–1957.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity 28, 621–629.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2009). B cell–specific and stimulation-responsive enhancers derepress *Aicda* by overcoming the effects of silencers. Nat. Immunol. 11, 148–154.

Uchimura, Y., Barton, L.F., Rada, C., and Neuberger, M.S. (2011). REG-γ associates with and modulates the abundance of nuclear activation-induced deaminase. Journal of Experimental Medicine 208, 2385–2391.

Wang, M., Yang, Z., Rada, C., and Neuberger, M.S. (2009). AID upmutants isolated

using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. Nat. Struct. Mol. Biol. 16, 769–776.

Wold, M.S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu. Rev. Biochem. 66, 61–92.

Wu, X., Geraldes, P., Platt, J.L., and Cascalho, M. (2005). The double-edged sword of activation-induced cytidine deaminase. J. Immunol. 174, 934–941.

Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.-R., et al. (2010). 14-3-3 adaptor proteins recruit AID to 5′-AGCT-3′–rich switch regions for class switch recombination. Nat. Struct. Mol. Biol. 17, 1124–1135.

Zaheen, A., Boulianne, B., and Parsa, J. (2009). AID constrains germinal center size by rendering B cells susceptible to apoptosis. Blood.

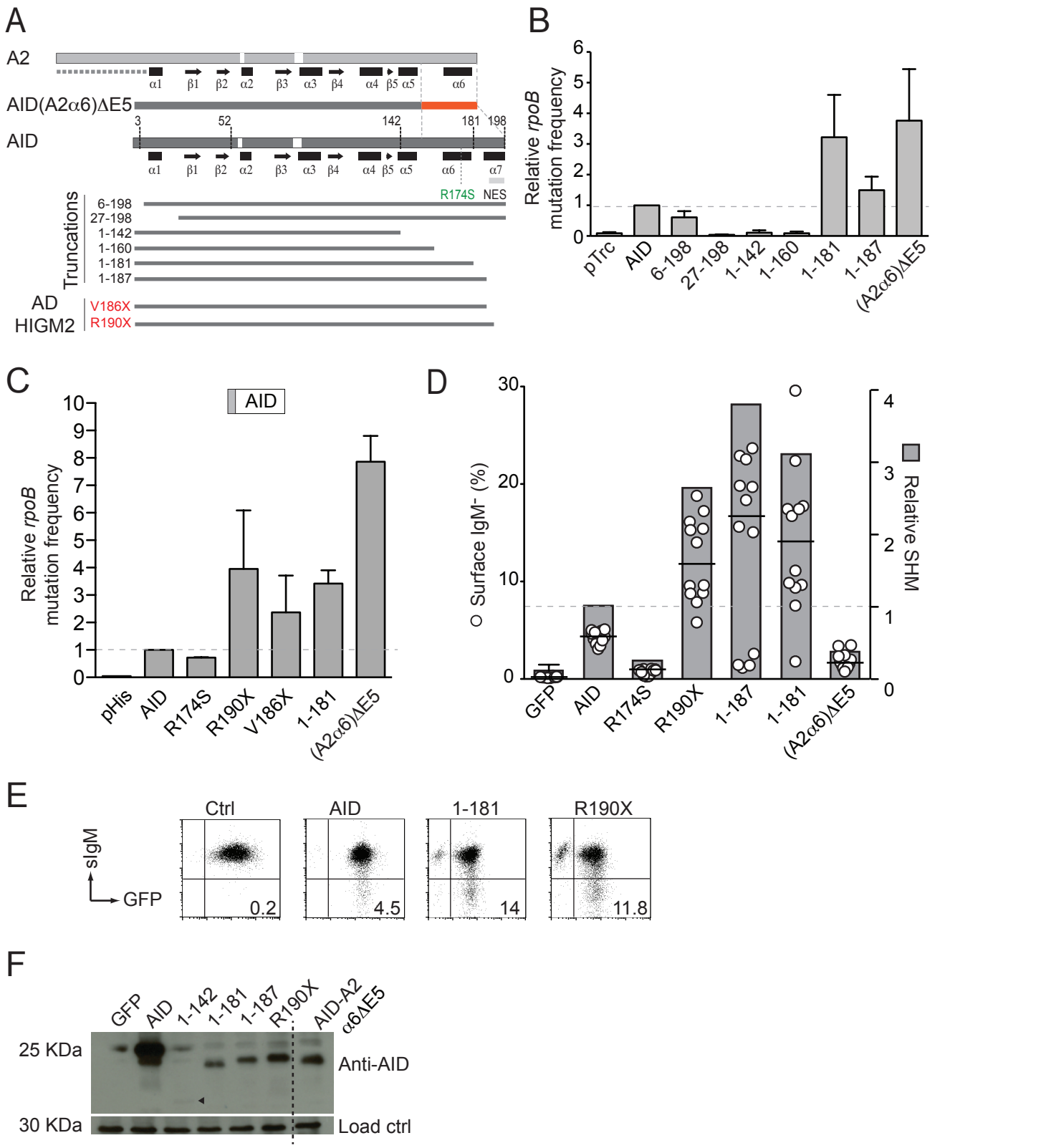

A
A2
α1 β1 β2 α2 β3 α3 β4 α4 β5 α5 α6
AID(A2α6)ΔE5
3 52 142 181 198
AID
α1 β1 β2 α2 β3 α3 β4 α4 β5 α5 α6 α7
R174S NES
Truncations
6-198
27-198
1-142
1-160
1-181
1-187
AD
HIGM2
V186X
R190X
B
Relative rpoB mutation frequency
pTrc AID 6-198 27-198 1-142 1-160 1-181 1-187 (A2α6)ΔE5
C
AID
Relative rpoB mutation frequency
pHis AID R174S R190X V186X 1-181 (A2α6)ΔE5
D
Surface IgM- (%)
Relative SHM
GFP AID R174S R190X 1-187 1-181 (A2α6)ΔE5
E
Ctrl AID 1-181 R190X
sIgM
GFP
0.2 4.5 14 11.8
F
GFP AID 1-142 1-181 1-187 R190X AID-A2 α6ΔE5
25 KDa
Anti-AID
30 KDa
Load ctrl

Figure 1

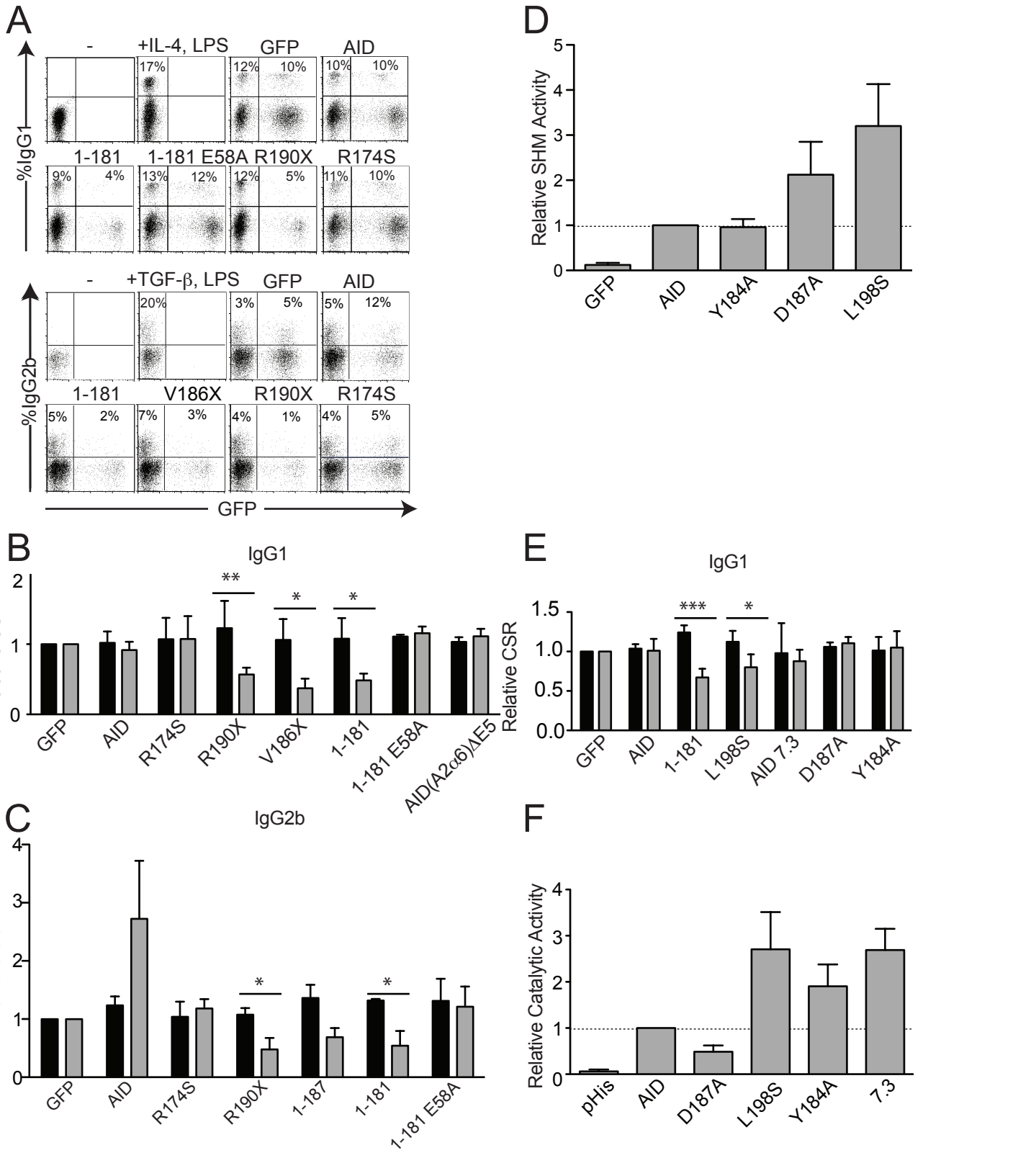

A
-
+IL-4, LPS
GFP
AID
%IgG1
17%
12%
10%
10%
10%
1-181
1-181 E58A R190X
R174S
9%
4%
13%
12%
12%
5%
11%
10%
-
+TGF-β, LPS
GFP
AID
%IgG2b
20%
3%
5%
5%
12%
1-181
V186X
R190X
R174S
5%
2%
7%
3%
4%
1%
4%
5%
GFP
D
Relative SHM Activity
GFP
AID
Y184A
D187A
L198S
B
IgG1
Relative CSR
GFP
AID
R174S
R190X
V186X
1-181
1-181 E58A
AID(A2α6)ΔE5
E
IgG1
Relative CSR
GFP
AID
1-181
L198S
AID 7.3
D187A
Y184A
C
IgG2b
Relative CSR
GFP
AID
R174S
R190X
1-187
1-181
1-181 E58A
F
Relative Catalytic Activity
pHis
AID
D187A
L198S
Y184A
7.3

Figure 2

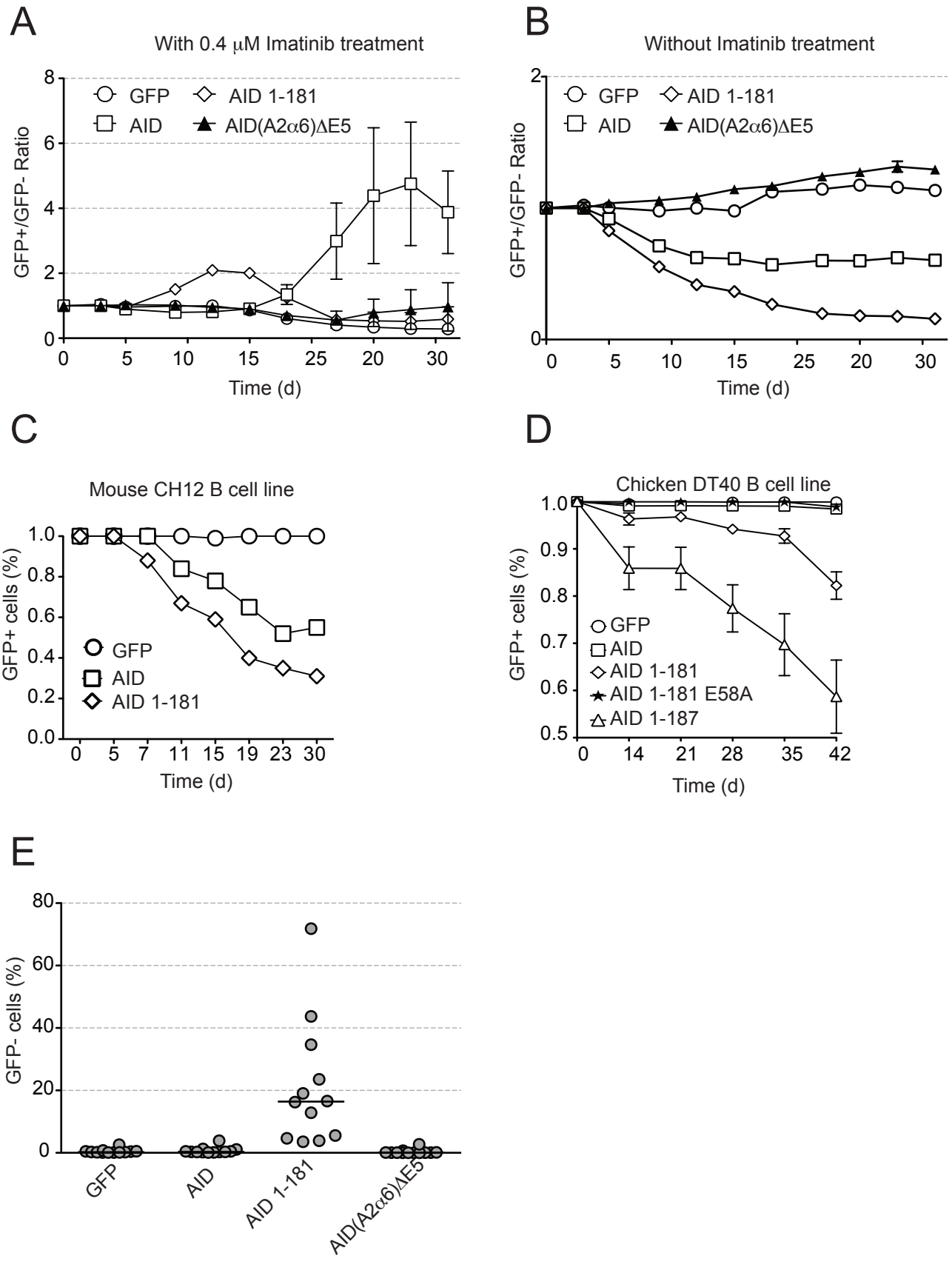

A
With 0.4 μM Imatinib treatment
GFP
AID
AID 1-181
AID(A2α6)ΔE5
GFP+/GFP- Ratio
Time (d)
B
Without Imatinib treatment
GFP
AID
AID 1-181
AID(A2α6)ΔE5
GFP+/GFP- Ratio
Time (d)
C
Mouse CH12 B cell line
GFP+ cells (%)
GFP
AID
AID 1-181
Time (d)
D
Chicken DT40 B cell line
GFP+ cells (%)
GFP
AID
AID 1-181
AID 1-181 E58A
AID 1-187
Time (d)
E
GFP- cells (%)
GFP
AID
AID 1-181
AID(A2α6)ΔE5

Figure 3

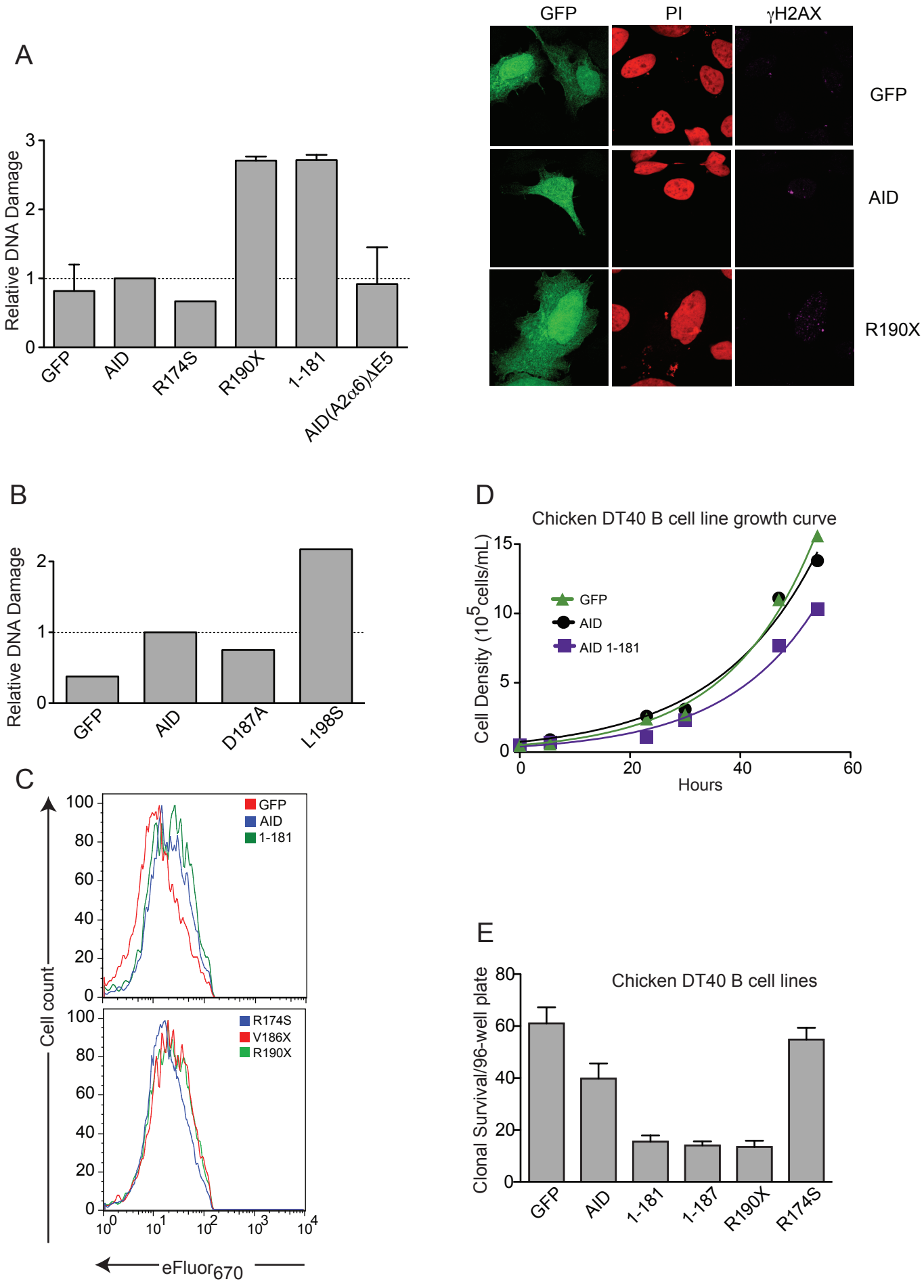

A
Relative DNA Damage
GFP
AID
R174S
R190X
1-181
AID(A2α6)ΔE5
GFP
PI
γH2AX
GFP
AID
R190X
B
Relative DNA Damage
GFP
AID
D187A
L198S
C
Cell count
GFP
AID
1-181
R174S
V186X
R190X
eFluor670
D
Chicken DT40 B cell line growth curve
Cell Density (10^5 cells/mL)
GFP
AID
AID 1-181
Hours
E
Chicken DT40 B cell lines
Clonal Survival/96-well plate
GFP
AID
1-181
1-187
R190X
R174S

Figure 4

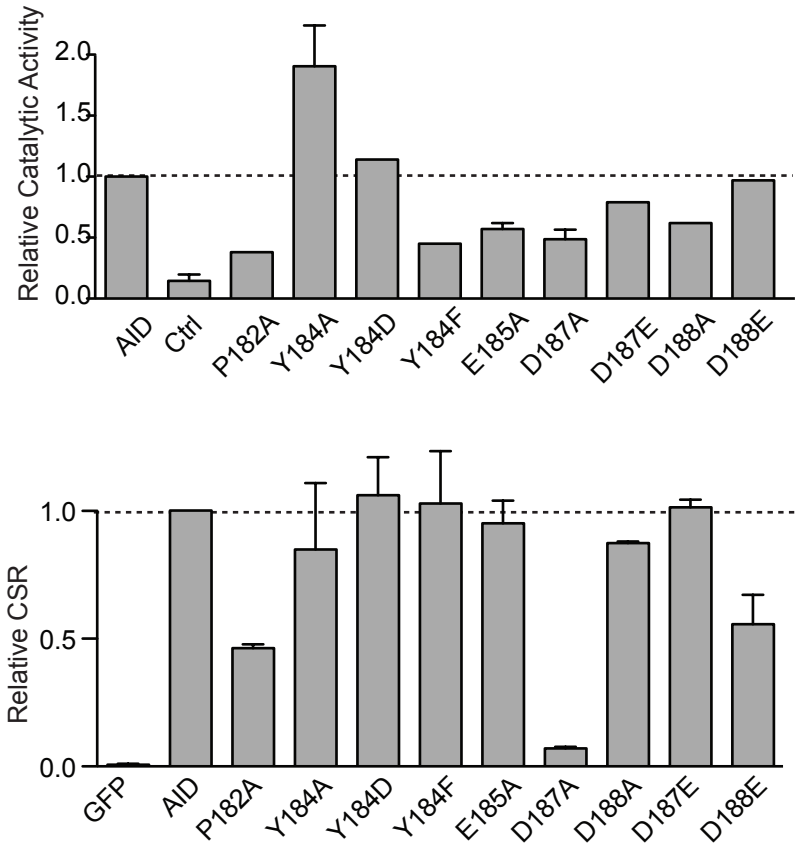

Relative Catalytic Activity
2.0
1.5
1.0
0.5
0.0
AID
Ctrl
P182A
Y184A
Y184D
Y184F
E185A
D187A
D187E
D188A
D188E
Relative CSR
1.0
0.5
0.0
GFP
AID
P182A
Y184A
Y184D
Y184F
E185A
D187A
D188A
D187E
D188E

Figure 5

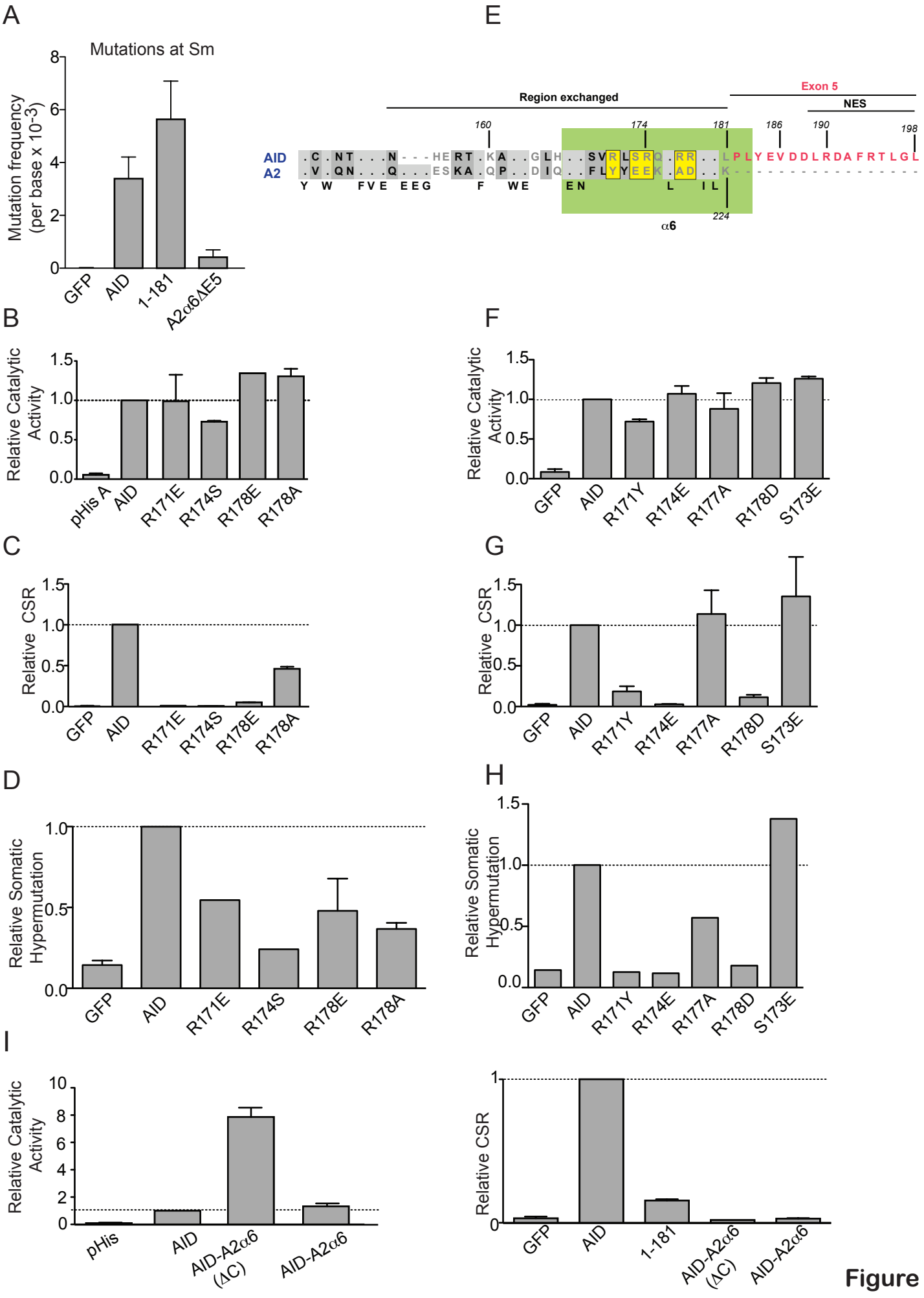

A
Mutations at Sm
Mutation frequency (per base x 10-3)
GFP
AID
1-181
A2α6ΔE5
E
Region exchanged
Exon 5
NES
160
174
181
186
190
198
AID
A2
α6
224
B
Relative Catalytic Activity
pHis A
AID
R171E
R174S
R178E
R178A
F
GFP
AID
R171Y
R174E
R177A
R178D
S173E
C
Relative CSR
G
D
Relative Somatic Hypermutation
H
I
pHis
AID-A2α6 (ΔC)
AID-A2α6

Figure 6

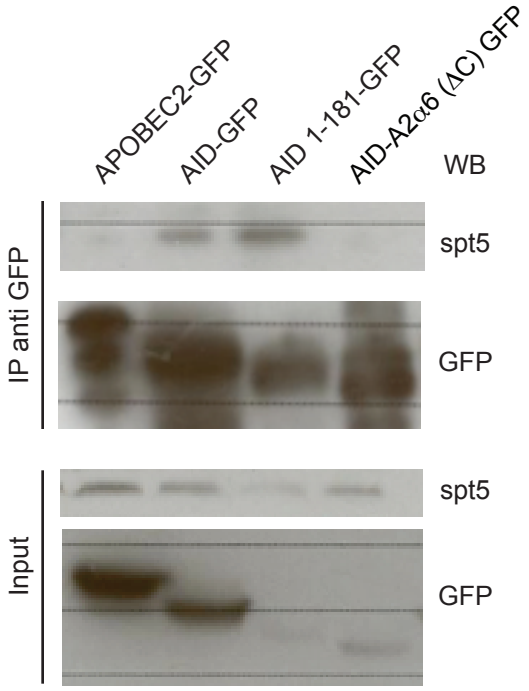
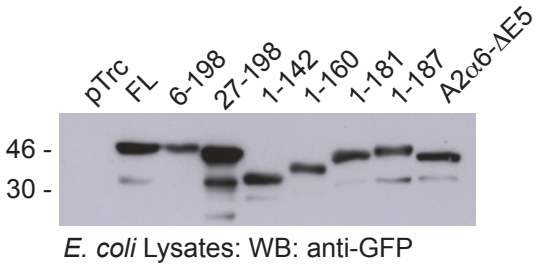

APOBEC2-GFP
AID-GFP
AID 1-181-GFP
AID-A2α6 (ΔC) GFP
WB
IP anti GFP
spt5
GFP
Input
spt5
GFP

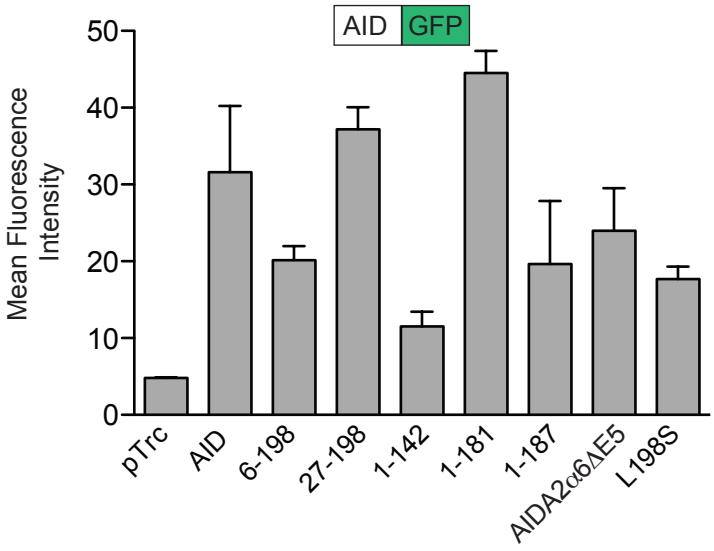
Figure 7

Supplementary 1A



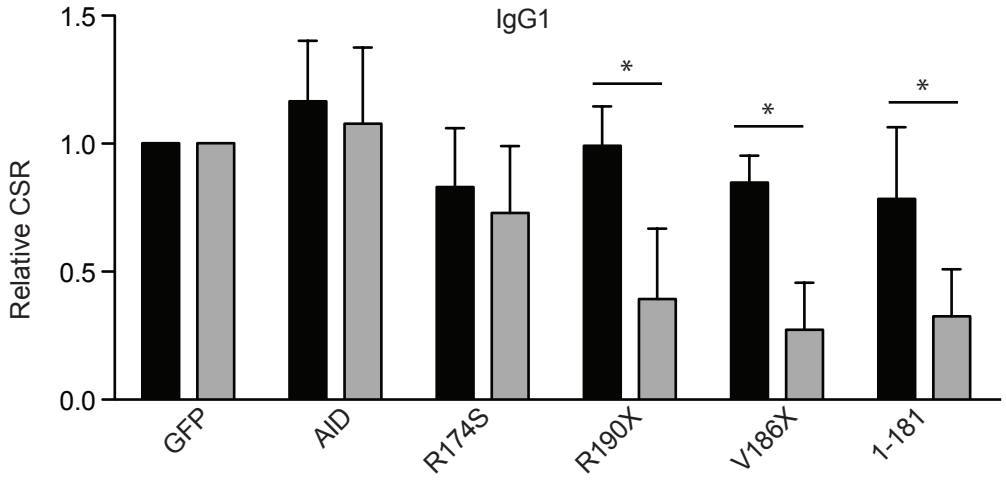
E. coli Lysates: WB: anti-GFP

Supplementary 1B

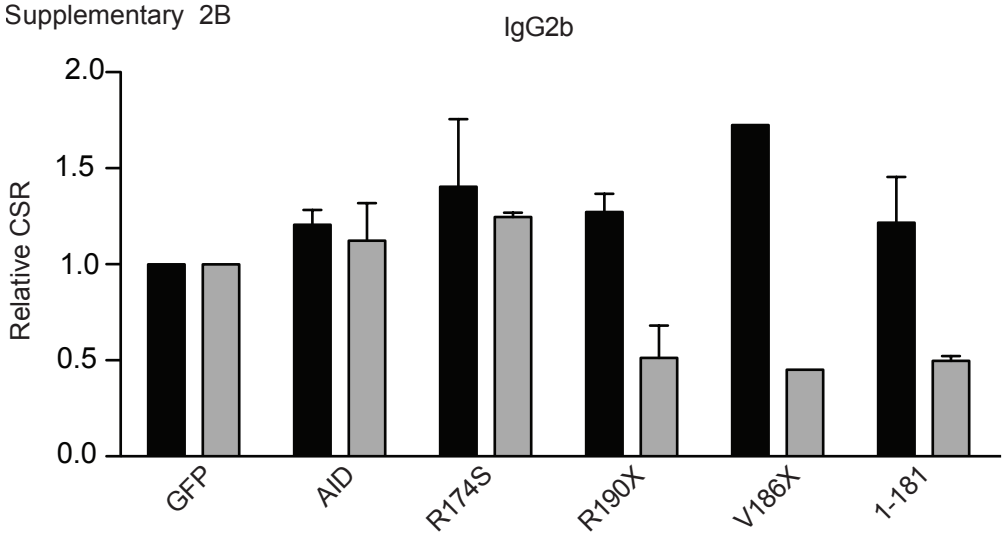


Supplementary Figure 1

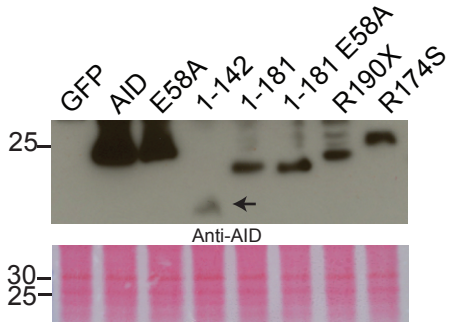
Supplementary 2A



Supplementary 2B

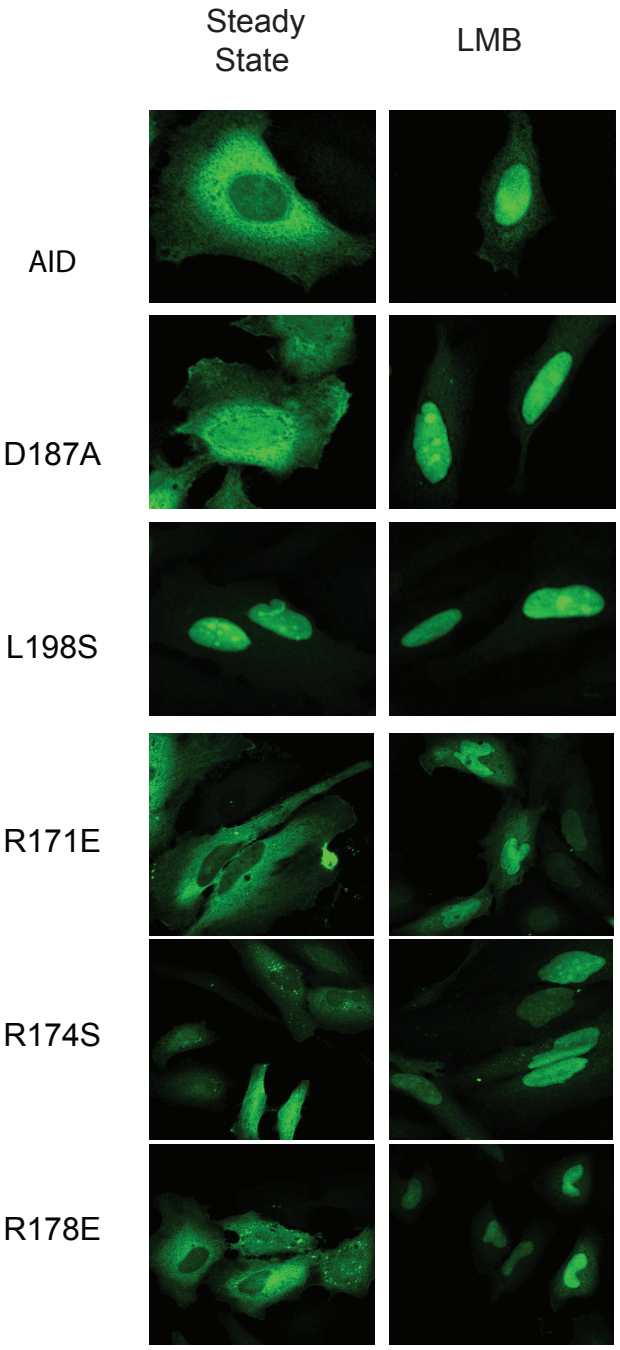


Supplementary 2C



Supplementary Figure 2

Supplementary 3



Supplementary Figure 3

patient	AID status	Ig levels (g/L)			CSR in vitro	SHM	Lymphadenopathies
		IgM	IgG	IgA			
NORMAL		0.5-1.1	5-12	0.3-1.3			
1-I-1*	R190X/+	4.05	1.28	0.1	defective	N	-
1-II-2*	R190X/+	2.85	0.5	<0.05	defective	N	-
1-II-3*	R190X/+	3.02	6.33	<0.05	defective	N	-
1-III-1*	R190X/+	1.22	7.58	<0.05	defective	N	-
2-I-1*	R190X/+	2.93	2.82	<0.05	defective	N	-
2-II-2*	R190X/+	2.34	0.21	<0.05	defective	defective	++
3-II-1*	R190X/+	2.34	0.38	<0.05	defective	N	-
4-I-1	R190X/+	2.03	1.30	<0.05	defective	ND**	-
5-I-1	V186X/+	2.02	5.41	1.19	defective	ND**	-
5-II-1	V186X/+	0.84	NE**	<0.05	ND**	defective	++
5-II-2	V186X/+	1.32	4.71	<0.05	defective	ND**	-
6-I-1	R174S/R174S	12.20	1.70	0.42	defective	ND**	++
7-I-1	R174S/R174S	9.0	0.7	<0.05	defective	defective	++
8-I-1	L181_P182ins31/ L181_P182ins31	12.0	1.0	<0.05	defective	N	-
8-I-2	L181_P182ins31/+	0.72	4.57	0 .19	ND**	ND**	-

Table 2

FIGURE LEGENDS

Figure 1: Different structural requirements for AID mutagenicity in *E coli* versus B cells

(A) A schematic representation of wild-type AID and mutants used in this study. Mutations causing hyper-IgM II syndrome that are inherited in a dominant manner are in red, while those that are transmitted in an autosomal recessive manner are in green. The region of AID that is replaced with residues from APOBEC2 in the chimeric protein AIDA2α6ΔE5 is shown in orange.

(B) *E. coli* mutation assay to measure the catalytic activity of C-terminal GFP-tagged wild-type AID and the indicated mutants. For each independent experiment, the mutational frequencies of AID mutants were normalized to wild-type AID control. Data are plotted as mean relative catalytic activity ± SD. The relative catalytic activities are representative of at least two independent experiments. (C) Same as (B), except that wild-type AID and mutants were cloned into pTrcHisA plasmid with an N-terminal 6XHistidine tag (D) Somatic hypermutation (SHM) was measured in IgM+ ψV^{-} $AID^{-/-}$ DT40 cell line infected with the retroviral vector pMx-IRES-GFP encoding wild-type AID and mutants with a consensus Kozak sequence. GFP+ IgM+ cells were seeded at a density of 50 cells/well twelve individual clones expanded for four weeks, at which time the surface IgM expression was measured by flow cytometry; representative of two independent experiments (E) Representative flow cytometry panels showing the loss of sIgM expression in IgM+ ψV^{-} $AID^{-/-}$ DT40 cells infected with full-length and indicated AID mutants four weeks post-sorting. (F) Lysates from 1×10^{6} infected and sorted for

GFP+ sIgM+ DT40 cells were used for anti-AID western blot. A non-specific band at about 30 KDa serves as an arbitrary loading control.

Figure 2: Dominant negative effect on CSR requIRES enzymatic activity

(A) Representative flow cytometry dot plots from one class-switching experiment with AID heterozygous mice. The numbers represent the percent switching of GFP- and GFP+ populations from each infection. Aggregate switching data from at least two independent experiments with mouse splenic B cells heterozygous for *Aicda* gene stimulated to switch to IgG1 with 50 ng/mL IL-4 and 25 μg/mL LPS (B) or to IgG2b with 1ng/mL TGFβ1 and 25 μg/mL LPS (C) only after retroviral infection. For each independent experiment, the switching of GFP- population (black bars) and GFP+ population (grey bars) of every sample is normalized to the GFP control. (* $P<0.05$,** $P<0.01$ Student's T test, Mann-Whitney, PRISM 5.0d). (D) Somatic hypermutation was measured as in 1D, the relative SHM activity was calculated by normalizing the median of %IgM negatives of the twelve clones to that of the full-length AID control, for each independent experiment, and presented as the mean SHM ± SD. (E) Same as (B) but with artificial point mutants of AID. (F) *E. coli* mutation assay to measure the catalytic activity of N-terminal 6XHistidine-tagged wild-type AID and the indicated mutants. For each independent experiment, the mutational frequencies of the variants were normalized to the wild-type AID control. Data are plotted as mean relative catalytic activity ± SD. The relative catalytic activities are representative of at least two independent experiments.

Figure 3: Increased off-target mutation by truncated C-terminally truncated AID requIRES α6

(A) BCR-ABL^{+} K562 cells were transduced with GFP control and pMXs-IRES-GFP vectors encoding AID and AID 1-181. Infected (GFP+) and uninfected (GFP-) populations were mixed to obtain a starting GFP+/GFP- 1:1 ratio and plated in triplicates. The cells were grown in the presence of 0.4 μM imatinib. The GFP+/GFP- ratio was periodically monitored for 30 days by flow cytometry. Data are plotted as the mean GFP+/GFP- ratio from triplicate populations ± SEM.

(B) Same experiment as in (A) but cells were grown in the absence of imatinib.

(C) Mouse CH12 cells infected with retroviruses encoding untagged AID variants were sorted for GFP expression and plated in 6-well plates at (5×10^5 cells/mL) and continuously maintained in logarithmic growth conditions. GFP expression was measured over time by flow cytometry.

(D) Chicken DT40 cells infected with pMx-IRES-GFP retroviruses encoding untagged wild-type AID and mutants were sorted for GFP expression and seeded at 50 cells/well. 12 clones were expanded for the indicated duration, periodically monitoring GFP expression by flow cytometry. Data are plotted as mean relative catalytic activity ± SD. One representative of two independent experiments is shown.

(E) Graph showing the levels of GFP expression of IgM+ ψV^{-} AID$^{-/-}$ DT40 cells infected with full-length AID and indicated mutants for twelve clones four weeks post-sorting for GFP.

Figure 4: CSR dominant negative AID variants are genotoxic and impair clonal expansion

(A) γH2AX foci in transduced HeLa Cells. HeLa cells were transfected with pMx-IRES-GFP retroviral vectors encoding untagged full-length AID and variants with a consensus Kozak sequence and were mounted onto glass slides and stained for γH2AX and counterstained with PI. The foci were counted in a double-blind manner, in which only the GFP+ cells with more than 3 foci per cell were considered as positive hits. The number of foci-positive cells expressing each of the mutants for each experiment was normalized to the foci-positive cells expressing the full-length control, and presented as mean number of foci-positive cells ± SEM. The data is representative of at least two independent experiments. Representative microscopic images showing the GFP (green), PI (for nuclear staining) and γH2AX channels. (B) Same as (A) with retroviruses expressing artificial point mutants. (C) Relative cell proliferation of infected primary mouse splenic B cells. Splenic B cells from AID-/- mice were purified, stained with e-Fluor-670 (eBioscience Catalog #65-0840) as per manufacturer's instructions and placed in culture with 25 μg/mL LPS. The cells were infected with pMx-IRES-GFP retroviral vectors encoding GFP, full-length AID and variants without a consensus Kozak sequence and activated with 10 ng/mL IL-4. Cell proliferation was assessed by flow cytometry on day 4 past infection and presented as histograms. (D) Relative *in vitro* clonal survival. DT40 cells transduced with pMx-IRES-GFP retroviral vectors encoding GFP, full-length AID or AID 1-181 with a consensus Kozak sequence were sorted for GFP expression and placed in culture at an initial density of 25×10^5cells/mL. The cell density was then measured by trypan blue exclusion twice a day for at least 60 hours.

The population growth is graphed on a semi-logarithmic graph. (E) Chicken DT40 cells infected with pMx-IRES-GFP retroviral vectors encoding GFP, full-length AID and AID 1-181 with a consensus Kozak sequence were positively sorted for GFP expression at one cell per well in 96-well plates, and the number of clones that appear after a 10-day incubation are scored, and the percent clonal survival calculated. The data are presented as mean ± SD.

Figure 5: The C-terminal region of AID modulates its catalytic activity

(A) *E. coli* mutation assay to measure the catalytic activity of N-terminal His-tagged wild-type AID and the indicated mutants. For each independent experiment, the mutational frequencies of the variants were normalized to the full-length AID control. Data are plotted as mean relative catalytic activity ± SD. The relative catalytic activities are representative of at least two independent experiments

(B) Class switch recombination assay to measure the switching activity to IgG1 isotype, in primary splenic B cells from AID-/- mice infected with pMx-IRES-GFP vectors encoding wild-type AID and the indicated mutants, and activated with 10 μg/mL LPS and 5 ng/mL IL-4. For each independent experiment, the class-switch recombination of the mutants were normalized to that of the full-length AID control. Data are plotted as mean relative CSR activity ± SD. The relative CSR activities are representative of at least two independent experiments.

Figure 6: AID α6 is functionally important

(A) Primary B cells from Ung-/- were purified (a lack of base excision repair means a higher accumulation of mutations), and infected with pMx-IRES-GFP vectors encoding wild-type AID and the indicated mutants, and activated with 10 μg/mL LPS and 5 ng/mL IL-4. Afterwards, the DNA was extracted to amplify the Sμ switch region by PCR, and the sequences determined and analyzed for mutational frequencies. Data are plotted as mean mutation frequency per base $\times 10^{-3}$ ± SD from several sequences analyzed. (B) *E. coli* mutation assay to measure the catalytic activity of N-terminal His-tagged wild-type AID and the indicated mutants. For each independent experiment, the mutational frequencies of the variants were normalized to the wild-type AID control. Data are plotted as mean relative catalytic activity ± SD. The relative catalytic activities are representative of at least two independent experiments (C) Class switch recombination assay to measure the switching activity to IgG1 isotype, in primary splenic B cells from AID-/- mice infected with pMx-IRES-GFP vectors encoding wild-type AID and the indicated mutants, and activated with 10 μg/mL LPS and 5 ng/mL IL-4. For each independent experiment, the class-switch recombination of the mutants were normalized to that of the full-length AID control. Data are plotted as mean relative CSR activity ± SD. The relative CSR activities are representative of at least two independent experiments (D) Somatic hypermutation (SHM) was measured in IgM+ ψV^{-} $AID^{-/-}$ DT40 cell line infected with the retroviral vector pMx-IRES-GFP encoding wild-type AID and mutants with a consensus Kozak sequence. GFP+ IgM+ cells were seeded at a density of 50 cells/well and twelve individual clones expanded for four weeks, at which time the surface IgM expression was measured by flow cytometry; representative of two independent experiments (E) Amino acid sequence alignment of AID and APOBEC2.

The region spanning α6 of AID (shaded in green) exchanged with that of Apo2 in the chimeric protein AID(A2α6)ΔE5 is labeled. The arginines replaced in AID with those from APOBEC2 in the analysis represented by figures 6F-H, are shaded in yellow. (F) Same as (B). (G) Same as C, except the primary B cells from AID-/- mice were infected with pmx-IRES-GFP vectors encoding wild-type AID and mutants with a consensus Kozak sequence. (H) Same as in (D). (I) *E. coli* mutation assay to measure the catalytic activity of N-terminal His-tagged wild-type AID, the chimeric protein AID(A2α6)ΔE5, and the fusion protein AIDA2α6. For each independent experiment, the mutational frequencies of the variants were normalized to the wild-type AID control. Data are plotted as mean relative catalytic activity ± SD. The relative catalytic activities are representative of at least two independent experiments

Figure 7: AID α6 is necessary to bind to Spt5

APOBEC2 (A2)-GFP, AID-GFP, AID 1-181 GFP, and AID(A2α6)ΔE5 were immunoprecipitated with an anti-GFP antibody from transiently expressing HEK293T cells. Immunoprecipitates (IP) were probed with anti-GFP and anti-spt5 in Western blots. The filters were then probed with anti-Spt5 antibodies, to test for immunoprecipitation of Spt5 with the mutant AID constructs.

Supplementary Figure 1

(A) Anti-GFP Western blot:

Lysates from overnight induced cultures of BW310 *E. coli* transformed with full-length AID and mutants expressed from pTrc23-24. Lysates were prepared as outlined in Methods. The samples were loaded into a 10% SDS-PAGE gel, transferred onto a nitrocellulose membrane and Western blotted against GFP.

(B) GFP fluorescence

Colonies of BW310 *E. coli* transformed with full-length AID and mutants expressed from pTrc23-24 were placed in 2XTY medium and induced with 1mM final concentration of IPTG. After an overnight induction, 100-500 µL of the sample was placed in a FACS tube, washed with 1X PBS, and the GFP fluorescence of the bacteria measured by flow cytometry. Data are presented as mean ± SD.

Supplementary Figure 2

Aggregate switching data from at least two independent experiments with splenic B cells from mice wild-type for the *Aicda* gene and stimulated to switch to IgG1 with 50 ng/mL IL-4 and 25 µg/mL LPS (A) or to IgG2b with 1ng/mL TGFβ1 and 25 µg/mL LPS (B) only after retroviral infection. For each independent experiment, the switching of the GFP- population (black bars) and GFP+ population (grey bars) of every sample is normalized to the GFP control. (* $P<0.05$,**$P<0.01$ Student's T test, Mann-Whitney, PRISM 5.0d).

(C) Whole-cell extracts from Plat-E virus producing cells were obtained 48 hours post-transfection and assessed for AID expression by Western blotting with a mouse anti-

human AID primary antibody and a goat anti-mouse-HRP secondary antibody. The Ponceau S staining of transferred proteins is shown as a loading control.

Supplementary Figure 3

Confocal images of HeLa cells transiently expressing C-terminally GFP-tagged wild-type AID and mutants. Cells were treated with leptomycin B where indicated according to (Patenaude et al., 2009).

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat. Struct. Mol. Biol. 16, 517–527.