Post-translational regulation of mammalian DNA cytidine deaminases

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Zachary Lee Demorest

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Reuben S. Harris

July 2011

Acknowledgements

With wholehearted gratitude, I would like to thank:

My advisor, Reuben Harris, for his dedication to the excitement of science and for providing a work atmosphere that is filled with motivated, driven and intelligent people.

My parents, Robert and Judy Demorest, who have always been supportive, nurtured my curiosities and encouraged me to do my best.

My wife, Shana Demorest, for being unconditionally understanding and for helping me in difficult times.

The Biochemistry, Molecular Biology and Biophysics department for travel awards to attend scientific conferences.

The Immunology Training Grant (NIH T32 AI0007313) for support during my final year of dissertation studies.

The Minnesota Medical Foundation for the Charles Carr/William Peterson award.

Harris lab members, especially past and current graduate students John Albin, Michael Burns, Guylaine Haché, Judd Hultquist, Lela Lackey, Rebecca LaRue, Steve Offer, Eric Refsland and Mark Stenglein; Dr. Bill Brown for expert technical advice and for keeping the lab in running order; and all the technician and undergraduates for making reagents and providing support.

Members of Friday lab meeting for their input on my projects and for providing yummy treats on Friday afternoons.

Dr. Lincoln Potter for helpful discussions regarding the detection and analysis of protein phosphorylation.

Dr. James Ervasti and Davin Henderson for assistance with baculovirus and insect cell culture.

My friends, who have endured this journey with me, helped me to be strong and have enjoyed many celebrations and happy hours with.

My undergraduate advisors, Walter Low and John Ohlfest, for guidance early in my career and for introducing me to the excitement of scientific research.

My Dog, Piton, who has never blamed me for missed walks or hunting trips due to the commitments of graduate school.

Dedication

This thesis is dedicated to my parents, Robert and Judy Demorest, and to my wife, Shana Demorest.

Abstract

Our immune system is faced with the challenge of neutralizing the daily bombardment of invading pathogens. This requires two major response mechanisms in order to achieve this task. The innate immune system acts quickly in response to non-specific bacterial, viral and nucleic acid based antigens to prevent the onset of disease. In contrast, the adaptive immune system functions slowly but has the added advantage of providing memory and long-term immunity to specific pathogens. Two members of the AID/APOBEC family of cytidine deaminases, APOBEC3G (A3G) and activation-induced deaminase (AID), are critical components of the innate and adaptive immune responses.

A3G is a potent inhibitor of Human Immunodeficiency Virus type-1 (HIV-1) as well as a number of other endogenous replicative elements that threaten our genomic integrity. This is accomplished by actively mutating cytosines to uracil in the viral genome during replication. These mutations render the virus non-functional and therefore incapable of spreading infection. AID is necessary for our bodies to generate a diverse antibody repertoire. This is achieved through the initiation of class-switch recombination and somatic hypermutation, processes that allow developing B-cells to generate antibodies of varying isotype with a high specificity for a given antigen. Both of these enzymes, AID and A3G, are DNA mutating enzymes with the potential to wreak havoc on our genome and contribute to cancer. The regulation of these enzymes is multi-faceted, involving differential transcription, miRNAs, subcellular localization, interactions with regulatory proteins, and post-translational modifications.

This thesis focuses specifically on understanding the regulatory mechanisms that govern the mutating capacity of AID and A3G. I have identified and characterized two novel AID interacting proteins, CIB1 and DND1, which I propose to have a role in the regulation of AID. Additionally, I discovered that AID, A3G and related DNA deaminases are negatively regulated by phosphorylation of a conserved threonine located adjacent to the active site of these enzymes. A phospho-mimetic T-to-E amino acid substitution impairs deaminase activity and blocks their function in cells [defective in class switch recombination (AID) and HIV-1 restriction (A3G)]. Due to near-universal conservation of this solvent exposed threonine, it is likely that this phospho-regulatory mechanism will be applicable to nearly all DNA deaminase family members spanning the entire vertebrate phylogeny. It is an extremely attractive adaptive/innate immunity mechanism because it provides a "real-time" switch to finely tune the deaminase activity of expressed proteins without the time delays intrinsic to other mechanisms. Finally, using chimeras and single amino acid substitutions, I identified cis-acting elements in human and mouse AID that are important for nucleocytoplasmic shuttling of these proteins.

Table of Contents

Page

Acknowledgements ..... **i**

Dedication ..... **ii**

Abstract ..... **iii**

Table of Contents ..... **v**

List of Figures ..... **vii**

Chapter 1: Introduction **1**

Mammalian DNA cytidine deaminases 2

Antibody gene diversification 3

Retroelement restriction 5

Mechanisms of regulation 7

Figures 12

Chapter 2: The Interaction Between AID and CIB1 is Nonessential for Antibody Gene Diversification by Gene Conversion or Class Switch Recombination **16**

Chapter summary 17

Introduction 18

Results 20

Discussion 25

Experimental procedures 27

Figures 34

Chapter 3: Phosphorylation Directly Regulates the DNA Cytidine Deaminase Activity of AID and APOBEC3G **40**

Chapter summary 41

Introduction 42

Results 44

Discussion 51

Experimental procedures 53

Figures 58

Chapter 4: Conclusions and Discussions **69**

Post-translational modifications 69
AID in germ cell reprogramming 70
The complexities of AID trafficking 71
Concluding remarks 73
Figures 74
References 75
Appendix I: AID interacts with DND1 88
Appendix summary 89
Introduction 90
Results 92
Discussion 95
Experimental procedures 98
Figures 100
Appendix II: Nucleocytoplasmic trafficking of AID 106
Appendix summary 107
Introduction 108
Results 110
Discussion 114
Experimental procedures 117
Figures 118
Appendix III: Phospho-peptide mapping of APOBEC3G 123
Appendix summary 124
Introduction 125
Results 127
Discussion 130
Experimental procedures 132
Figures 135

List of Figures

Page

Fig. 1-1. Phylogenetic tree illustrating the expansion of APOBEC members 12
Fig. 1-2. Mechanisms of antibody diversification 13
Fig. 1-3. Retroelement restriction by AID/APOBEC proteins 14
Fig. 1-4. Cellular distribution of human APOBEC3s 15
Fig. 2-1. AID interacts with the calcium and integrin binding protein 1 34
Fig. 2-2. CIB1 over-expression in DT40 35
Fig. 2-3. CIB1 is dispensable for immunoglobulin gene conversion in DT40 36
Fig. 2-4. AID localization in $CIB^{-/-}$ DT40 cells 38
Fig. 2-5. CIB1 is dispensable for CSR in mice 39
Fig. 3-1. An active site Thr/Ser is conserved in mammalian DNA cytidine deaminases.. 58
Fig. 3-2. Intrinsic DNA cytidine deaminase activity of AID and A3G constructs 60
Fig. 3-3. A3G-T218E has diminished deaminase activity but retains DNA binding ability 61
Fig. 3-4. Mutants of A3G and AID localize normally within living cells 63
Fig. 3-5. AID-T27E is defective for class switch recombination 64
Fig. 3-6. A3G-T218E fails to restrict Vif-deficient HIV-1 65
Fig. 3-S1. A3G-T218E has reduced deaminase activity in human cell lysates 66
Fig. 3-S2. Phospho-mimetic mutants retain DNA binding ability 67
Fig. 3-S3. General model for the activation of DNA deaminases during immune activation 68
Fig. 4-1. Post-translational modifications of A3G 74
Fig. I-1. Identification of DND1 100
Fig. I-2. AID interacts with DND1 101
Fig. I-3. AID and DND1 co-localize in living cells 102
Fig. I-4. DND1 is not required for class-switch recombination 103
Fig. I-5. AID and DND1 may be involved in testicular tumorigenesis 104
Fig. I-6. Model for functional implications of AID and DND1 interactions 105
Fig. II-1. Mouse AID-GFP is compromised for import activity 118

Fig. II-2. Differences in the putative NLS do not account for import defect 119
Fig. II-3. The C-terminus of mouse AID contains cytoplasmic retention activity 120
Fig. II-4. A single amino acid substitution disrupts import activity in human AID 121
Fig. II-5. A single amino acid substitution disrupts cytoplasmic retention in mouse AID 122
Fig. III-1. APOBEC3G-strep is phosphorylated *in vivo* 135
Fig. III-2. PKA phosphorylates APOBEC3G-strep *in vitro* 137
Fig. III-3. CaMKII phosphorylates APOBEC3G *in vitro* 138

Chapter 1: Introduction

Mammalian DNA cytidine deaminases

The information encoded by the genome of a cell is of vital importance. Cells go to great lengths to protect the integrity of this information. There are many exogenous sources of DNA mutation including UV damage, radiation and carcinogens. Another threat to genomic integrity is the hydrolytic deamination of cytosine by water, which results in the conversion to uracil[1]. The cell combats these mutations using enzymes called uracil DNA glycosylases, which recognize the aberrant uracil bases in DNA and remove them.

It is curious then, why would organisms possess enzymes that actually catalyze the deamination of cytidine? The AID/APOBEC family of cytidine deaminases is comprised of enzymes that actively facilitate the conversion of cytosine to uracil. Activation induced deaminase (AID) is thought to be the most ancestral of the APOBEC family, with its origin rooted in bony fish[2,3]. A duplication of the gene gave rise to APOBEC1 (A1), which is conserved down to chickens, lizards, and opossum (**Fig. 1-1**). Further duplication and evolution of this gene family resulted in the generation of the APOBEC3 (A3) family, with ancestral placental mammals believed to encode three copies (A3Z1, A3Z2 and A3Z3). Interestingly, mice, rats and pigs seem to have lost the A3Z1 domain. On the other hand, in humans and primates there has been an explosion of the APOBEC3 locus, predominantly of the A3Z2 subtype, totaling 7 separate domains present today (**Fig. 1-1**).

Overall, the human genome encodes 11 different AID/APOBEC family members: AID, APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H and APOBEC4. Several members of the

AID/APOBEC family of cytidine deaminases are capable of generating uracil bases, which are typically only found in RNA, in DNA polynucleotides molecules. These aberrant bases would seemingly be detrimental to the organism, but these enzymes have been maintained throughout evolution due to their beneficial specialized functions in the adaptive and innate immune response by facilitating antibody gene diversification and inhibition of exogenous and endogenous retroelements.

Antibody gene diversification

One of the mechanisms mammals use to defend themselves against invading foreign pathogens is through neutralization with highly specific antibodies. Great diversity is required in the antigen binding region in order to generate antibodies capable of recognizing invading pathogens. Antibodies, or immunoglobulins (Ig), are composed of heavy and light polypeptide chains. Each Ig is made up of two heavy and two light chains that are connected together by disulfide bonds. The N-terminal ends of the heavy and light chains form the antigen binding domain, called the variable region. The C-terminal end of the heavy chain composes the constant region, the region that determines the antibody's effector function.

Individual B lymphocytes diversify their specific immunoglobulin through several mechanisms, the first of which is V(D)J recombination (**Fig. 1-2A**). This process, catalyzed by the RAG1/2 heterodimer, randomly joins together a combination of variable (V), diversity (D) and joining (J) regions from tandem arrays encoded at the Ig locus[4]. This generates a unique functional immunoglobulin that the B lymphocyte can express on

its cell surface as it circulates looking for antigen.

In most mammals, two additional processes can further diversify the antibody repertoire: somatic hypermutation (SHM) and class switch recombination (CSR) (**Fig. 1-2B,C**). These processes, amazingly, were both found to be initiated by AID catalyzed deamination at the Ig locus[5-7]. Somatic hypermutation is a process that occurs in germinal center B cells undergoing affinity maturation in which AID introduces non-templated point mutations into the variable region of the immunoglobulin[8] (**Fig. 1-2B**). This randomly alters the specificity of the Ig for its cognate antigen. Mutations that result in increased affinity towards an antigen are selected for, while those that decrease affinity are selected against. This process results in clonal B lymphocytes that produce Igs with greater affinity for a specific antigen than V(D)J recombination can generate alone.

The second mechanism, class switch recombination, is also catalyzed by AID deamination at the Ig locus[5,9]. AID deaminates cytidine to uracil in the repetitive switch regions found immediately upstream of an array of constant regions at the Ig heavy chain locus (**Fig. 1-2C**). These lesions are processed by repair machinery to result in recombination that changes the default constant region Cμ (designated IgM) to a downstream constant region (e.g. Cα, Cε, Cγ; IgA, IgE, IgG). These constant regions (and therefore isotypes) differ in several ways including the length of the hinge region, the number and location of interchain disulfide bonds and the number of attached oligosaccharide moieties[10]. These differences determine the immunoglobulin's effector function and generate further diversity in the antibody response. For example, IgG1 and IgG3 antibodies are recognized by receptors present on the surface of phagocytic cells

such as macrophages and neutrophils, whereas IgE is bound by mast cells, basophils, and eosinophils[11,12]. A single rearranged variable region may be expressed in IgG, IgA, or IgE antibodies to provide a multifaceted immune response.

The absence of AID in humans manifests as an immunodeficiency called hyper-IgM syndrome type-2 (HIGM2)[13]. These patients are characterized by B lymphocytes that are unable to perform CSR and therefore only generate IgM antibodies. The case for SHM is more difficult to study in human patients, however evidence suggests that humans lacking AID have a decreased capacity to generate high affinity antibodies by affinity maturation[14]. In mouse models for AID deficiency, defects in CSR and SHM are both apparent[5].

Retroelement restriction

Of the 2.9 billion nucleotides in the human genome, only 1.5% actually codes for known proteins[15]. A large portion is made up of transposable elements such as DNA transposons, retrotransposons and retroviruses. DNA transposons, which move via a "cut and paste" mechanism that results in the transposon moving about the genome without increasing the total number of insertions, only comprise a small portion. These insertions, however, can be detrimental when they land within a gene and interrupt its normal expression. On the other hand, replicative retrotransposons including long terminal repeat (LTR) transposons, and non-LTR elements such as long interspersed nuclear elements (LINEs) and small interspersed nuclear elements (SINEs), make up a much larger portion of the genome. These replicative elements move via a "copy and paste" mechanism,

where they are transcribed to generate an RNA copy that is subsequently reverse-transcribed to DNA and reinserted somewhere else in the genome. This results in an increase in the number of insertions upon each replicative cycle. There are estimated to be more than 500,000 long interspersed nuclear element-1 (LINE-1) integrations in the human genome, comprising about 17%[15,16].

Exogenouses retroviruses, including human immunodeficiency virus-1 (HIV-1), are very similar to retrotransposons in that they reproduce via a "copy and paste" mechanism. The major difference is that they have an extracellular stage in their lifecycle. After transcription of the retroviral genome into RNA, it is packaged into viral particles along with other viral proteins and released from the cell by budding from the cell membrane. When the viral particle encounters a new cell, it fuses with the membrane and deposits its cargo into the cytoplasm. Here, the RNA is reverse-transcribed to DNA and integrated into the newly infected cell's genome.

Several members of the AID/APOBEC family have been shown to inhibit the replication of these mobile elements. The best documented is APOBEC3G's (A3G) ability to inhibit the replication of HIV-1[17-21]. A3G is co-packaged with the HIV-1 RNA genome into budding virions. Upon infection of a new cell, A3G deaminates cytidine to uracil in the first strand cDNA during reverse transcription (**Fig. 1-3, left cycle**). This results in a hypermutated viral genome that effectively terminates further replication of the virus. The ability of other members of the AID/APOBEC family to inhibit the replication of HIV-1 has been debated, however strong evidence suggests APOBEC3F is also fully capable of restriction[22-24].

As for endogenous retroelements, AID/APOBEC proteins have also been shown to inhibit their replication. The replication of LINE-1, the most common retroelement in the human genome, is strongly inhibited by AID, A3A, A3B and A3F[25-28]. Furthermore, APOBEC3 proteins have been reported to inhibit the replication of other common retroelements from several other species including MusD (mouse), IAP (mouse) and Ty1 (yeast)[29-31] (**Fig. 1-3, right cycle**). It is clear that AID/APOBEC proteins can play an important role in maintaining the integrity of the genome through the inhibition of mutagenic retroelements. However, these enzymes themselves are capable of generating genomic mutation and therefore must be tightly regulated.

Mechanisms of regulation

The described beneficial properties of cytidine deaminases in innate and adaptive immunity must outweigh the potential costs of compromised genomic integrity in order for these enzymes to have thrived in the evolutionary process. For this to be true, one would expect tight regulation of the AID/APOBEC family members to prevent off-target deamination of the genome. To date, multiple mechanisms have been described that contribute to minimizing these potentially damaging effects.

Transcription

AID/APOBEC proteins are not ubiquitously expressed in every tissue. AID was first identified as a B cell specific factor that was required for SHM and CSR[5,6]. However, more recent reports have also identified its expression in human and mouse testes[27,32]. APOBEC1, which edits an mRNA important in lipid metabolism, is found in highest

quantities in the small intestine, liver and spleen[33,34]. APOBEC2 is expressed primarily in muscle[35], however no known function has been described until very recently. APOBEC2 deficiency has now been reported to result in left-right axis defect during frog embryogenesis and altered muscle fiber types and myopathy in mice[36,37].

Due to high sequence similarity between the different APOBEC3s, it has been difficult to generate an accurate expression profile. Many of the early reports were flawed in their ability to actually distinguish between the 7 human A3s. Our laboratory devoted a significant amount of effort to generate a panel of quantitative PCR reagents capable of profiling the expression of the A3s separately[38]. The detailed results of this study are too exhaustive to describe here, however some general observations can be made. Overall, the highest A3 expression was found in lymphoid organs such as the thymus and spleen, where large populations of T and B lymphocytes are found. A3A expression was found to be highest in tissues containing macrophages, consistent with its role in degrading foreign DNA in phagocytic cells[39].

Inhibition by micro RNA

Another common mechanism for gene regulation is through translational inhibition by micro-RNAs (miRNA)[40-42]. miRNAs are endogenous small double-stranded RNA molecules that are processed into a ~22 nucleotide single-stranded RNA by the enzymes drosha and dicer. This small RNA molecule is then guided by an RNA interference silencing complex to bind to the 3'UTR of a target mRNA to inhibit translation. An endogenous miRNA (miR-155) has been reported to negatively regulate AID expression via this mechanism[43,44].

Ubiquitin mediated proteasomal degradation

In the cellular environment, proteins modified with poly-ubiquitin chains are targeted for degradation by the proteasome. Three members of the AID/APOBEC family have been demonstrated to be poly-ubiquitinated for degradation via this mechanism. In one example, HIV-1 counteracts APOBEC3 restriction by encoding a small protein called vif that binds to APOBEC3G and APOBEC3F and recruits members of a ubiquitin ligase complex including elongin B/C, cullin5, and RBX[22-24,45-49]. The ubiquitinated APOBEC3 proteins are degraded by the proteasome to effectively deplete the cell of these restriction factors and allow for viral replication.

In another case, proteasomal degradation of an AID/APOBEC member is thought to be beneficial for the cell. Nuclear AID has been shown to have a reduced lifespan due to proteasomal degradation[50]. This is presumed to help protect the genomic DNA by limiting the amount of AID that resides in the nucleus. However, as of yet, the specific ubiquitin ligase complex members have not been identified.

Subcellular localization

The AID/APOBEC family presents a complex distribution of subcellular localization. Each of the APOBEC3 proteins has differential cellular distributions (**Fig. 1-4**). A3A can be found throughout the cell, evenly distributed between nucleus and cytoplasm, while A3D, A3F and A3G are mainly restricted to the cytoplasmic compartment. A3C and A3H have a slight bias towards the nucleus, but can be found in the cytoplasm as well. A3B seems to be the only APOBEC3 that is predominantly nuclear.

The most ancestral member, AID, is a nucleocytoplasmic shuttling protein, with nuclear export via CRM1/exportin, nuclear import by the karyopherin α family and cytoplasmic retention all contributing to the dynamics of subcellular localization[51-53]. AID is found predominantly in the cytoplasmic compartment due to active export from the nucleus via the CRM1 export machinery. However, AID must enter the nucleus to reach its physiological target, the immunoglobulin locus. There is some controversy in the literature regarding how this is achieved and this thesis addresses these issues through the identification of cis-acting elements in human and mouse AID that contribute to the active import and cytoplasmic retention of these proteins (**Appendix II**).

Regulatory protein interactions

Numerous proteins have been proposed to interact with various members of the AID/APOBEC family. The functional implications of these interactions are more apparent for some than others. APOBEC1 complementation factor (ACF), an RNA-binding protein, is necessary for APOBEC1 to act on its known target, apolipoprotein B mRNA[54,55].

Several proteins have been proposed to interact with AID including CTNNBL1[56,57], GANP[58], HSP90[59], MDM2[60], protein kinase A (PKA)[61-63] and RPA[64]. The implications of these interactions vary, with some facilitating antibody diversification[56], affecting protein stability[59], enabling interaction with single-stranded DNA[64] or affecting cellular localization[57,58]. The implications of phosphorylation by cellular kinases will be discussed in detail below.

The work described in this thesis extends this field through the identification and

characterization of two novel AID interacting proteins: calcium and integrin binding protein-1 (CIB1) (**Chapter 2**) and dead-end (DND1) (**Appendix I**).

Phosphorylation

Post-translational modification by phosphorylation is a common mechanism for regulation of cellular processes. AID is subject to phosphorylation at numerous sites including S3, T27, S38, S41, S43, T140 and Y184[61,63,65-67]. However, only the modifications at S3, S38, and T140 have been characterized. Phosphorylation at serine-3 has been reported to negatively regulate AID's ability to facilitate CSR without altering its intrinsic deaminase activity[65]. In contrast, phosphorylation at threonine-140 seems to increase the rate of somatic hypermutation and class switch recombination[66]. Lastly, serine-38 phosphorylation enables interaction with the ssDNA binding protein RPA and facilitates SHM and CSR[61,64,67,68].

APOBEC3G appears also to be a substrate for post-translational modification by phosphorylation. A report indicates that it is phosphorylated at threonine-32 by protein kinase A (PKA)[69]. The authors suggest that this modification dampens vif's ability to target A3G for proteasomal degradation. This thesis describes a novel regulatory mechanism involving the phosphorylation of a threonine that is conserved among nearly all mammalian AID/APOBEC members (**Chapter 3**).

Figures

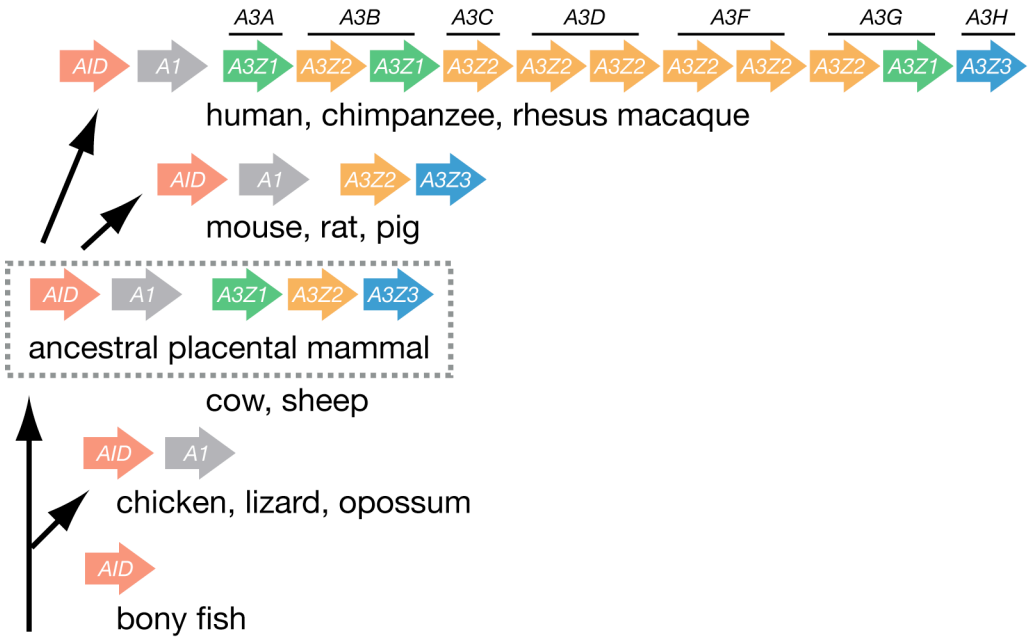


Fig. 1-1. Phylogenetic tree illustrating the expansion of APOBEC members. Schematic illustrating the expansion of the AID/APOBEC family in vertebrates. AID originates in bony fish. A1 is found in chicken, lizard and opossum, while the A3 family is rooted in placental mammals. Figure courtesy of Rebecca LaRue.

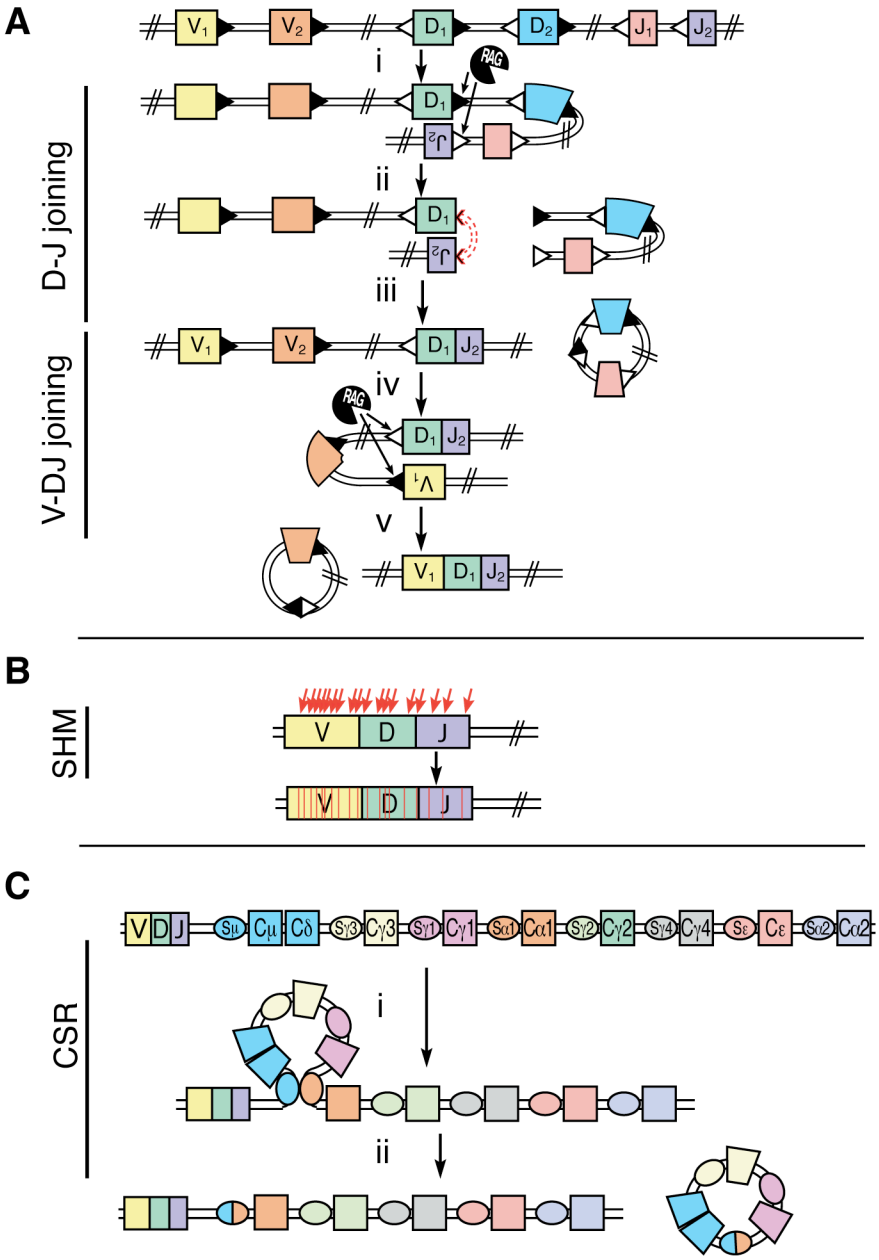


Fig. 1-2. Mechanisms of antibody diversification. *A*. The process of V(D)J recombination is facilitated by the RAG1/2 heterodimer. A single variable (V), diversity (D) and joining (J) region are recombined at random to form a unique combination for antibody recognition. *B*. Somatic hypermutation (SHM) is a process where non-templated point mutations are introduced into the antigen binding region after V(D)J recombination. *C*. Class switch recombination (CSR) occurs between repetitive switch regions (e.g., Sμ, Sγ3) to result in exchange of the default constant region (μ) for one of several found downstream (e.g., Cγ3, Cγ1, Cα1, Cγ2, Cε, Cα2). Figure adapted from[70].

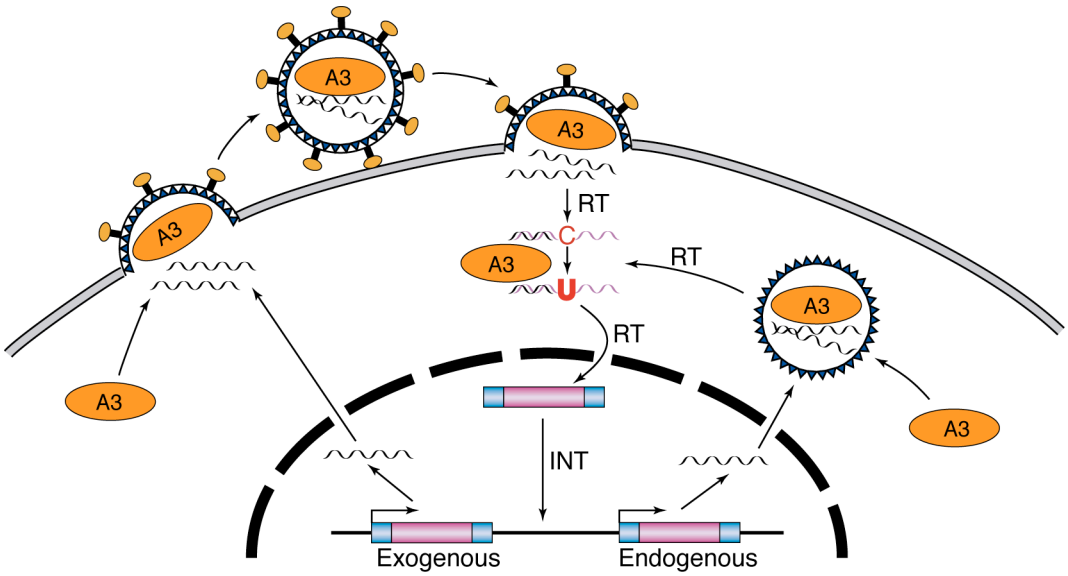


Fig. 1-3. Retroelement restriction by AID/APOBEC proteins. A model for APOBEC3 (A3) mediated restriction of exogenous retroelements (e.g., HIV, left cycle) and endogenous elements (e.g., LINE-1, right cycle). The A3 is packaged into budding virions and travels with the virus to a new cell. Upon fusion, A3 converts cytosine (C) to uracil (U) in the first strand cDNA during reverse transcription. In the case of endogenous elements, the entire lifecycle occurs within the cell. RT=reverse transcription, INT= integration. Modified from[71].

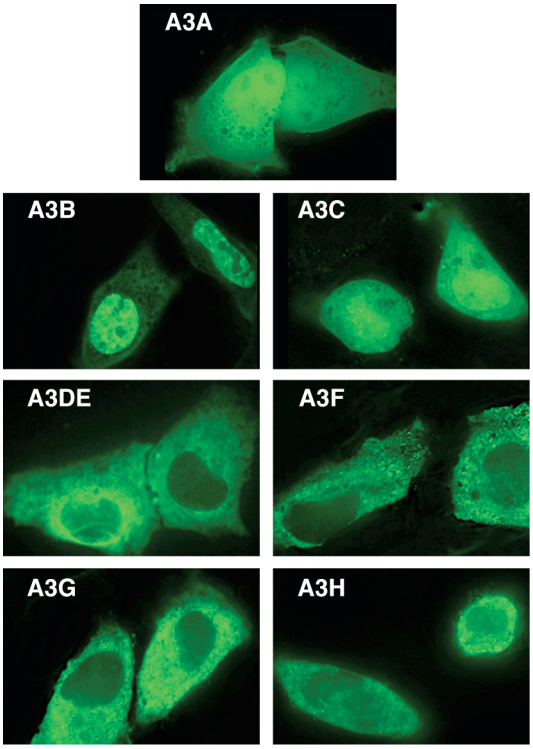


Fig. 1-4. Cellular distribution of human APOBEC3s. Images of human APOBEC3s expressed as N-terminal GFP fusion proteins transiently transfected into HeLa cells. Live cell images of steady-state localization were taken 48 hours after transfection. Images courtesy of Lela Lackey.

Chapter 2: The Interaction Between AID and CIB1 is Nonessential for Antibody Gene Diversification by Gene Conversion or Class Switch Recombination

Adapted, with permission, from: Demorest et al. (2010) PLoS One 5(7):e11660.
W.L. Brown conducted the experiments presented in Fig. 2-1. D.A. MacDuff generated the data for Fig. 2-2. Z.L. Demorest performed the experiments for Figs. 2-3, 2-4, and 2-5.

Chapter summary

Activation-induced deaminase (AID) initiates somatic hypermutation, gene conversion and class switch recombination by deaminating variable and switch region DNA cytidines to uridines. AID is predominantly cytoplasmic and must enter the nuclear compartment to initiate these distinct antibody gene diversification reactions. Nuclear AID is relatively short-lived, as it is efficiently exported by a CRM1-dependent mechanism and it is susceptible to proteasome-dependent degradation. To help shed light on mechanisms of post-translational regulation, a yeast-based screen was performed to identify AID-interacting proteins. The calcium and integrin binding protein CIB1 was identified by sequencing and the interaction was confirmed by immunoprecipitation experiments. The AID/CIB1 interaction resisted DNase and RNase treatment, and it is therefore unlikely to be mediated by nucleic acid. The requirement for CIB1 in AID-mediated antibody gene diversification reactions was assessed in CIB1-deficient DT40 cells and in knockout mice, but immunoglobulin gene conversion and class switch recombination appeared normal. The DT40 system was also used to show that CIB1 over-expression has no effect on gene conversion and that AID-EGFP subcellular localization is normal. These combined data demonstrate that CIB1 is not required for AID to mediate antibody gene diversification processes. It remains possible that CIB1 has an alternative, a redundant or a subtle non-limiting regulatory role in AID biology.

Introduction

Following V(D)J recombination, the expressed antibody repertoire in vertebrate B lymphocytes gains additional diversity through the processes of immunoglobulin (Ig) gene conversion (IGC), somatic hypermutation (SHM), and class-switch recombination (CSR) (reviewed by[8,9,72]). Most vertebrates use SHM to diversify variable (V) region sequences, but a few species, such as birds and rabbits, use a template-dependent IGC mechanism to generate antigen-binding diversity and improve antibody affinity. All vertebrates use CSR to irreversibly change the constant (C) region of the heavy chain Ig gene, which dictates the antibody tissue distribution and function.

All three of these antibody gene diversification reactions are initiated at the DNA level by activation-induced cytidine deaminase (AID)-catalyzed deoxy-cytidine to deoxy-uridine editing[7] (reviewed by[8,9,72]). Once these lesions occur in Ig gene DNA, ubiquitious DNA 'repair' enzymes catalyze further reactions that ultimately result in the distinct outcomes described above. For instance, both the uracil DNA glycosylase UNG2 and the mismatch recognition proteins MSH2/6 help process AID-catalyzed DNA uridines[73-75].

Several lines of evidence indicate that AID is subject to numerous levels of post-translational regulation. First, AID is predominantly cytoplasmic, and recent data suggest that this may occur through an active retention mechanism[53,76,77]. Second, AID is actively imported into the nuclear compartment, a step that is undoubtedly required for its essential role in antibody gene diversification[51,53]. Third, AID is exported from the nuclear compartment by virtue of a carboxy-terminal leucine-rich motif, which is bound by CRM1 and necessary for export[51-53,77,78]. Fourth, AID is susceptible to post-

translational regulation by ubiquitination and phosphorylation[50,61,63,66,68,79,80]. The responsible E3 ubiquitin ligation complex has yet to be identified, but protein kinase A (PKA) has been shown to phosphorylate AID[61,63,80]. Finally, phosphorylation enhances the interaction between AID and replication protein A, which may be an important DNA level co-factor[64]. These studies combine to indicate that multiple cellular proteins regulate AID, and several of these have yet to be identified.

To gain further insights into the proteins that regulate AID, we used high-throughput yeast two-hybrid screening to identify interacting proteins. We found that AID interacts with the calcium and integrin binding protein 1 (CIB1), a 22 kDa regulatory protein that is broadly expressed, comprised of four EF hand motifs (two of which bind calcium), and required for spermatogenesis in mice[81,82]. CIB1 is a provocative AID-interacting protein, because it could potentially link B cell receptor signaling, activation through calcium fluctuations, and the germinal center specific antibody gene diversification reactions. The requirement for CIB1 in AID-mediated processes was examined in two model systems for antibody gene diversification, the chicken B cell line DT40 and murine B lymphocytes.

Results

The AID-CIB1 interaction

A large scale, Gal4-based yeast two-hybrid screen was conducted as described[83]. A bait plasmid expressing human AID residues 56-198 was used to screen a tongue and tonsil cDNA library. DNA sequencing was used to identify candidate interactors, and false positives were eliminated by cross checking literature, public databases, and unpublished data. Only one positive hit resisted triage: residues 1-191 of the calcium and integrin binding protein 1 (CIB1). Other reported AID-interaction proteins, DNA-PKcs[84], MDM2[60], RPA[64], PKA[61,63], and CTNNBL1[56], were not recovered.

To confirm the yeast two-hybrid result, we co-expressed AID-EGFP and myc-CIB1 in HEK-293T cells and asked whether the two proteins interacted by co-immunoprecipitation (co-IP). These experiments showed that myc-CIB1 was able to co-IP AID-EGFP (**Fig. 2-1**A). In contrast, myc-CIB1 did not co-IP an EGFP control.

AID can bind single-stranded nucleic acids, including both DNA and RNA[85]. AID homologs such as APOBEC3G also bind RNA, and this interaction enables indirect associations with many different cellular proteins[86-90]. To ask whether the AID-CIB1 interaction occurs through a nucleic acid bridge, co-IP experiments were repeated in the presence of DNase or RNase. An α-GFP (AID) IP resulted in strong myc-CIB1 immunoblot signals, which were not diminished by incubating extracts with either DNase I or RNase A (**Fig. 2-1**B). These data strongly suggested that the interaction between AID and CIB1 occurs independently of a nucleic acid bridge.

To further substantiate the AID-CIB1 interaction, a pull-down experiment was performed with a tandem affinity purification construct AID-STZ (Experimental procedures). This construct, or an EGFP-STZ control, was transfected into HEK-293T cells, expressed for 48 hours, and ultimately pulled-down with IgG Sepharose, which binds the Z domain of Protein A. Immunoblotting with a CIB1-specific monoclonal antibody revealed that endogenous CIB1 associates with AID-STZ but not with EGFP-STZ, which was even more highly expressed (**Fig. 2-1**C).

CIB1 over-expression in DT40

To begin to test whether CIB1 regulates AID-dependent antibody gene diversification reactions, we over-expressed both human and chicken CIB1 proteins in the chicken B cell line DT40. This cell line constitutes a good model system for studying IGC, because the reversion of an Ig frameshift mutation is AID-dependent and can be monitored at the cellular level by flow cytometry of surface Ig positive cells[91,92]. CIB1 over-expressing lines were established and subjected to limiting dilution, and 8-12 of the resulting subclones were cultured continuously for 6 weeks. The median percentage of surface Ig positive cells among the control vector transfected subclones was 7.5%, and the medians of the CIB1 over-expressing subclones were not significantly different (**Fig. 2-2**). These data indicated that CIB1 protein levels are not limiting in DT40 for gene conversion at the antibody locus.

Immunoglobulin gene conversion in CIB1 knockout DT40

To determine whether CIB1 is required for IGC in DT40, serial gene targeting was used to delete exons 5 and 6 of the *CIB1* gene (**Fig. 2-3**A). A linearized targeting vector encoding the puromycin-resistance gene was transfected into a surface Ig-negative DT40 line and resistant colonies were isolated. Gene targeting events were identified by PCR (**Figs. 2-3**A & B). *CIB1* heterozygous lines were subjected to a second round of gene targeting with a blasticidin resistance construct. CIB1 null lines were identified by allele-specific PCR reactions and confirmed by reverse-transcription PCR and Southern blot analyses (**Figs. 2-3**B & C and data not shown). The CIB1 defective cells showed no obvious growth defects or cellular abnormalities demonstrating that this protein is not essential in DT40 cells.

To ask whether AID-dependent IGC requires CIB1, surface Ig-negative $CIB1^{+/+}$, $CIB1^{+/-}$, and $CIB1^{-/-}$ cell lines were subcloned by limiting dilution, maintained in culture for 6 weeks, labeled with α-Ig antibody, and quantified by flow cytometry. We found that the median frequency of surface Ig-positive subclones was nearly identical for $CIB1^{+/+}$, $CIB1^{+/-}$, and $CIB1^{-/-}$ cell lines, 4.5%, 3.9%, and 4.4%, respectively (**Fig. 2-3**D). These data provided an unambiguous demonstration that CIB1 is not required for IGC in DT40.

AID localization in CIB1 knockout DT40

Another property of AID is that it localizes predominantly to the cytoplasm of cells (*e.g.*,[27,53,76,77]). To ask whether CIB1 is required for this process, we stably transduced wildtype and CIB1-deficient DT40 cells with an AID-EGFP expression

construct and used fluorescent microscopy to image the resulting localization patterns (**Fig. 2-4**). The cytoplasmic localization of AID-EGFP was indistinguishable in wildtype and CIB1-deficient cells indicating that CIB1 is also dispensable for this property.

Isotype switching in CIB1 knockout mice

To evaluate the potential contribution of CIB1 to another AID-dependent process in another model organism, we examined the CSR phenotype of CIB1-deficient mice[82]. First, we measured the abundance of alternative antibody isotypes in sera from 3 wildtype and 6 CIB1-deficient mice by ELISA (**Fig. 2-5**A). Normal levels of IgM, IgG1, IgG2a, IgG2b, IgG3, and IgA were observed indicating that a CIB1 deficiency does not result in a gross CSR defect.

However, minor deficiencies in CSR can be readily eclipsed by antigen driven clonal expansions of alternative isotype expressing B lymphocytes. We therefore next examined the capacity of wildtype and CIB1-deficient primary B lymphocytes to undergo CSR *ex vivo* in response to lipopolysacharride (LPS) and interleukin-4 (IL-4) treatment. This stimulatory regimen drives CSR from IgM to IgG1, which is easily quantified by α-IgG1 staining and flow cytometry (*e.g.*,[73]). After 4 days in culture, 28% of the wildtype cells expressed surface IgG1 and near-identical results were obtained with CIB1-deficient B lymphocytes (**Fig. 2-5**B). These data combined to demonstrate that CIB1 is not required for CSR in mice or in murine B lymphocytes *ex vivo*.

Germinal center morphology in CIB1 knockout mice

AID-deficient mice have enlarged germinal centers and spleenomegaly[5]. To ask whether the CIB1 interaction with AID may be important for proper development *in vivo*, we compared the morphology of the spleens of wildtype and CIB1-deficient mice. Spleens were removed, sectioned, and stained with hematoxylin and eosin. CIB1-deficient mice were found to have a normal spleen and germinal center morphology (**Fig. 2-5**C).

Discussion

AID-mediated antibody diversification reactions are conserved processes in vertebrates. However, the precise mechanisms that regulate AID remain to be fully understood. We identified the calcium and integrin binding protein 1 (CIB1) as a strong AID-interacting protein. This protein is an attractive regulatory candidate because calcium signaling is an integral part of antigen-dependent adaptive immune responses. However, the complete ablation of CIB1 in DT40 or mice demonstrated that this protein is dispensable for both IGC and CSR in these organisms (*i.e.*, CIB1 is not an antibody diversification co-factor). Moreover, the lack of any detectable antibody diversification phenotype also indicated that CIB1 is unlikely to be a critical positive or negative regulator of AID. This point was further underscored by the lack of an Ig diversification phenotype in CIB1 over-expressing DT40.

CIB1 is however important for other functions in vertebrates. It is a nonenzymatic regulatory molecule that binds an array of target proteins and modulates their activity. For example, CIB1 activates the serine/threonine kinase p21-activated kinase (PAK1) and thereby modulates the migratory and proliferative capacities of cells[93]. It inhibits polo-like kinase 3 (PLK3), which may slow the G1 phase of the cell cycle[94], and it inhibits apoptosis signal-regulating kinase (ASK1), which negatively regulates the stress-activated MAP kinases JNK and p38[95]. The CIB1 knockout mouse develops near normally under non-stressed conditions except that the CIB1-/- male is sterile[82]. This is most likely due to a disrupted progression of the haploid phase of spermatogenesis. Significantly slowed proliferation of mouse embryo fibroblasts was also noted. While the

exact cause of the disrupted haploid phase is unclear, it may be related to elevated levels of the cell cycle regulator CDC2/CDK1.

Since the interaction between AID and CIB1 is robust, we were surprised that antibody gene diversification reactions are not obviously affected. This could be due to a functional overlap between CIB1 and the related calcium and integrin binding proteins, CIB2 and CIB3, which share over 60% amino acid similarity[81,96]. However, AID may have other roles in the biology of vertebrates, such as the provision of innate immunity to retrotransposition[27,31,97,98], the restriction of foreign DNA[39], or the epigenetic reprogramming of stem cells[99,100]. Future studies will be necessary to address the role of CIB1 in these potentially interesting alternatives.

Experimental procedures

Ethics Statement

All experimental procedures for working with mice, described in full below, were approved by the University of Minnesota Institutional Animal Care and Use Committee (IACUC Protocol Number: 0703A04446).

DNA constructs

The parental myc expression vector encoding 6 copies of an N-terminal myc tag was a generous gift from Dr. Shigeki Miyamoto (University of Wisconsin)[101]. Human *CIB1* cDNA was PCR amplified from a spleen cDNA library (Invitrogen) using primers 5'-GAA-TTC-TGA-TGG-GGG-GCT-CGG-GCA-GTC-GC and 5'-CTC-GAG-TCA-CAG-GAC-AAT-CTT-AAA-GGA-GCT-G, digested with EcoRI/XhoI and ligated into p6xmyc to generate p6xmyc-hCIB1 (pRH1023). Chicken *CIB1* cDNA was PCR amplified from DT40 cDNA using primers 5'-GAA-TTC-TCA-TGG-GGG-GCT-CCA-GCA-GTC-TGC and 5'-CTC-GAG-TCA-CAG-GAC-AAT-CTT-GAA-GG, digested with EcoRI/XhoI, and ligated into p6xmyc to generate p6xmyc-cCIB1 (pRH1511).

Human AID cDNA[60] was PCR amplified using primers 5'-TAA-TAC-GAC-TCA-CTA-TAG-GG and 5'-GTC-GAC-AAG-TCC-CAA-AGT-ACG-AAA-TGC, digested with HindIII/SalI, and ligated into similarly cut pEGFP-N3 (Clontech) (pRH984). The lentiviral transduction vector pCSII-AID-EGFP (pRH1892) was generated by directly subcloning AID-EGFP from pEGFP-N3-AID (pRH986) into pCSII-EF-MCS (pRH1892)[102] using the XhoI/NotI sites. The lentiviral packaging and

helper plasmids pMDG (pRH429) and ΔNRF (pRH1899) were provided by Dr. Nik Somia (University of Minnesota)[103,104].

The tandem affinity purification constructs, pAID-STZ and pEGFP-STZ (Strep, TEV, Z domain), were made as follows. First, a Tobacco Etch Virus (TEV) protease cleavage site (ENLYFQG) was cloned into the EcoRI/XmaI sites of pBluescript II KS+ (Stratagene). Second, two tandem copies of the IgG binding Z domain of protein A were amplified by PCR from pRAV-flag[105] using primers 5'-CCC-GGG-ATG-AGG-TTA-ACC-ATG-GCG-CAA and 5'-GAG-CTC-TCT-AGA-TTA-CGC-GTC-TAC-TTT-CGG-CGC-CTG and cloned into SacI/SmaI sites of pBluescript-TEV. Third, a XhoI-NotI-EcoRI linker was added by ligating annealed primers 5'-TCG-AGA-GCG-GCC-GCA-TG and 5'-AAT-TCA-TGC-GGC-CGC-TC into the XhoI/EcoRI sites. Fourth, a Streptavidin tag was added by ligating annealed oligonucleotides 5'-GGC-CGC-ATG-GCT-AGC-TGG-AGC-CAC-CCG-CAG-TTC-GAA-AAA-GGC-GCC and 5'-GGC-CGG-CGC-CTT-TTT-CGA-ACT-GCG-GGT-GGC-TCC-AGC-TAG-CCA-TGC into the NotI site. Fifth, Strep-TEV-ZZ was subcloned into the XhoI/XbaI sites of pcDNA3.1(Invitrogen). Finally human AID or EGFP were subcloned into KpnI/SalI sites of this construct to generate pAID-STZ (pRH1216) and pEGFP-STZ (pRH1233).

The chicken *CIB1* targeting vectors were generated by amplifying targeting arms from DT40 genomic DNA using primers 5'-GGC-CCG-TTT-TCG-TCC-CCC-GGA and 5'-ACG-GCA-CCT-CCG-TGC-GGG-AGC (1.2 kb left arm) and 5'-TCC-AGC-AGC-ATA-AGG-GGT-CCC and 5'-CTG-CAC-AGA-GCT-CGT-TCC-CCA (1.0 kb right arm). These PCR products were cloned into pCR-BluntII-TOPO (Invitrogen) and

subcloned into pBluescript II KS+ (Stratagene) using NotI/BamHI (left arm) and EcoRI/HindIII (right arm). The puromycin or blasticidin resistance cassettes were then subcloned from bML4 or bML5[91] into the BamHI site between the homology arms (pRH1479).

Immunoprecipitation experiments

HEK-293T cells were transfected with 5 μg of plasmid pEGFP-N3-AID or pEGFP-N3 and p6xmyc-CIB1 in a 10 cm dish (~60% confluent) using a ratio of 1 μg DNA to 3 μl FuGene6 (Roche). 48 hrs post-transfection, cells were treated with 50 μM MG132 for 5 hours prior to harvesting. MG132 was included to minimize the degradation of AID by the proteasome[50], although it was not necessary to detect the AID/CIB1 interaction (data not shown). The cells were washed in cold PBS and resuspended in lysis buffer [25 mM HEPES, pH 7.4; 150 mM NaCl; 1 mM EDTA; 1 mM $MgCl_2$; 1 mM $ZnCl_2$; 10% glycerol; 1% NP40; protease inhibitors (Complete, EDTA-free, Roche)], on a rotating wheel for 1 hour at 4ºC. The lysate was clarified by centrifugation at 13,000 *g* for 20 minutes and 300 μl was incubated with 2.5 μL (2.5 μg) α-myc mAb (LabVision) on a rotating wheel for 30 minutes at 4ºC before the addition of 25 μL of pre-washed Protein G Sepharose (GE Healthcare). Bound complexes were washed three times with lysis buffer and once with TBS/0.05% Tween 20. Proteins were eluted off the beads with 0.1M glycine (pH 2.5) and resolved by 10% SDS-PAGE. The proteins were transferred to PVDF membrane and blotted with a rabbit α-GFP polyclonal Ab (Invitrogen) to detect AID-GFP.

Reciprocal pulldowns were conducted by transfecting pEGFP-N3-AID or pEGFP-N3 and 6xmyc-hCIB1 as above. Lysates (300 μL) were incubated with 20 μg of RNase A, DNase I or BSA at 37 ºC for 20 minutes (Sigma). They were then incubated with 5 μL (5 μg) of α-GFP mAb (Clontech JL8) on a rotating wheel for 30 minutes at 4ºC before the addition of 25 μL of washed Protein G Sepharose beads. The beads were washed, eluted, and transferred as above then blotted with α-myc (LabVision 9E11 mAb).

For the endogenous CIB1 pull-down experiments, HEK-293T cells were transfected as above with pAID-STZ or pEGFP-STZ. These proteins were precipitated with IgG Sepharose (GE Healthcare), removed from beads by TEV protease cleavage, processed as above to a PVDF membrane, and probed first with an α-CIB1 monoclonal antibody[106] and second with an α-Strep monoclonal antibody (Novagen). Primary antibodies were detected with α-rabbit-HRP or α-mouse-HRP secondary reagents (BioRad).

DT40 experiments

Surface Ig-negative DT40 cells were maintained in RPMI (Hyclone) supplemented with 10% FBS (Gibco) and 50 μM 2-mercaptoethanol. Cells were subcloned by limiting dilution and grown continuously for 6 weeks. The IGC capacity was determined by measuring the percentage of cells that reverted to surface Ig-positive by staining with R-PE conjugated mouse α-chicken Ig antibody (Southern Biotechnology Associates, Inc.) and quantifying labeled cells by flow-cytometry (FACSCalibur, BD Biosciences). 8-12 independent subclones were examined for each condition and the

median percentage of surface Ig-positive cells was used to assess gene conversion capacity.

CIB1-deficient DT40 was constructed by serial gene targeting[60,92]. Targeting constructs were linearized with NotI and electroporated into 20 million cells (500 V, 25 μF; BioRad GenePulser II). Following the appropriate drug selection (0.25 μg/ml puromycin, 10μg/ml blasticidin), single clones were isolated and expanded for screening by PCR, RT-PCR, and Southern blotting. For PCR screening, a common reverse primer 5'-CTT-TGT-GCC-TCC-CGT-TAG-AG located 3' to the targeting arm was used in combination with forward primers designed specifically for exon 5 5'-GAT-GGC-ACC-ATC-AAC-CTC-TC, the blasticidin drug cassette 5'-GAA-GAC-CCC-AAG-GAC-TTT-CC, or the puromycin cassette 5'-CCC-CCT-GAA-CCT-GAA-ACA-TA. To confirm loss of CIB1 mRNA expression, cDNA was generated from total mRNA and subjected to RT-PCR using primers spanning intron 4 5'-GGG-ATG-ACA-GCA-TGT-CCT-TT and 5'-TGA-GCT-GCT-CCA-TCT-CAA-TG. Primers for AID expression were used to verify the presence of cDNA 5'-TGG-ACA-GCC-TCT-TGA-TG and 5'- GTC-CCA-GAG-TTT-TAA-AG. For Southern blotting, a 3' probe was generated by PCR from DT40 genomic DNA using primers 5'-GAT-CCC-TCC-CTC-CTT-GGG-AGA and 5'-CCG-GAG-GCG-TTG-GCT-GGT-GCC.

DT40 lines stably expressing chicken CIB1 or human CIB1 were generated by electroporating linearized (BglII) p6xmyc-hCIB1 or p6xmyc-cCIB1 (250 V, 950 μF, 2 pulses; BioRad GenePulser II) into surface Ig negative DT40, selecting with 1.5 mg/ml Geneticin (Invitrogen), and screening for expression by immunoblotting. 2 x 10^6 cells

were lysed in NP40 lysis buffer [25mM HEPES, pH 7.4; 150mM NaCl; 1mM EDTA; 1mM $MgCl_2$; 1mM $ZnCl_2$; 10% glycerol; 1% NP40; protease inhibitors (Complete, EDTA-free, Roche)] on ice for 1 hr and then clarified by centrifugation (16,100 g, 5 min). The supernatants were denatured in Laemmli loading buffer and separated by SDS-PAGE. Proteins were transferred to PVDF membrane and blotted with α-myc (LabVision 9E11 mAb) to detect transfected CIB1 or α-tubulin (Covance mAb).

AID localization was determined in DT40 by transducing cells with a lentivirus encoding AID-EGFP (above). Lentiviruses were produced by transfecting HEK-293T cells with pCSII-AID-EGFP, the VSV-G envelope vector pMDG, and the ΔNRF packaging construct encoding the HIV-1 gag, pol, rev, tat, and vpu genes as described previously[104]. DT40 cells were cultured in lentiviral supernatant for 12 hrs to enable viral transduction. After an additional 36 hrs, the cells were washed with fresh media and AID-EGFP localization was determined by fluorescence microscopy (40x objective; Deltavision).

Mouse experiments

Experimental procedures were conducted in accordance with the University of Minnesota Institutional Animal Care and Use Committee guidelines. The CIB1-deficent mice were reported previously[82]. Heterozygous mice were bred to produce experimental wildtype and CIB1-deficient littermates, which were housed approximately 7 weeks and then used for experiments.

To quantify serum Ig levels, 50-100 μl blood was collected from mice via facial vein bleeding and allowed to clot at room temperature. After removing cells by

centrifugation, serum antibody titers for isotypes IgM, IgG1, IgG2a, IgG2b, IgG3, and IgA were determined by ELISA using a Mouse Immunoglobulin Isotyping Kit (BD Pharmingen).

To determine the frequencies of *ex-vivo* CSR, resting B-cells were isolated from dissociated spleen through magnetic sorting by negative selection using an antibody cocktail against CD43 (Ly-48), CD4 (L3T4), and Ter-119 (Miltenyi Biotech). Isolated B-cells shown to be over 90% CD19 positive were then cultured for in RPMI media (Gibco) supplemented with 10% FBS plus 50 ng/ml IL-4 (R&D Systems) and 50 μg/ml LPS (Sigma). After 4 days, cells were analyzed for CSR by flow-cytometry using a α-IgG1-PE antibody (BD Biosciences).

For histology experiments, spleens were removed from euthanized animals, preserved in 10% phosphate-buffered formalin, sliced into 10 μm thick sections (Leica Instruments), stained with hematoxylin and eosin, and examined by light microscopy to identify germinal centers (Zeiss Axiovert).

Figures

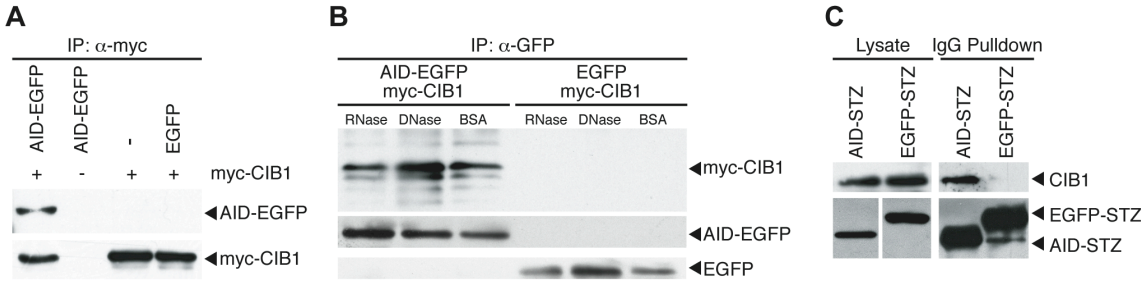


Figure 2-1. AID interacts with the calcium and integrin binding protein 1. A. myc-CIB1 co-IPs AID-EGFP but not EGFP only (left and right lanes, respectively). A myc-specific monoclonal antibody was used to precipitate complexes, and AID-GFP was detected with an α-GFP polyclonal antibody. **B**. AID-EGFP co-IPs myc-CIB1 in a DNase I- and RNase A-resistant manner. A α-GFP monoclonal antibody was used to precipitate AID-GFP, and myc-CIB1 was detected with a α-myc monoclonal antibody. **C**. AID-STZ pulls down endogenous CIB1 from HEK-293T cell extracts. IgG Sepharose was used to pull-down STZ complexes, and CIB1 was detected using a α-CIB1 polyclonal antibody. AID-STZ and GFP-STZ were detected with a α-strep antibody. A low level of non-specific background was observed in the vicinity of AID-STZ. For cell lysate (input) control blots, two panels are shown because EGFP-STZ is expressed over 100-fold better than AID-STZ. A quantification of the input versus pull-down signal indicated that 1% of cellular CIB1 can be pulled-down with AID complexes when IgG sepharose beads are limiting.

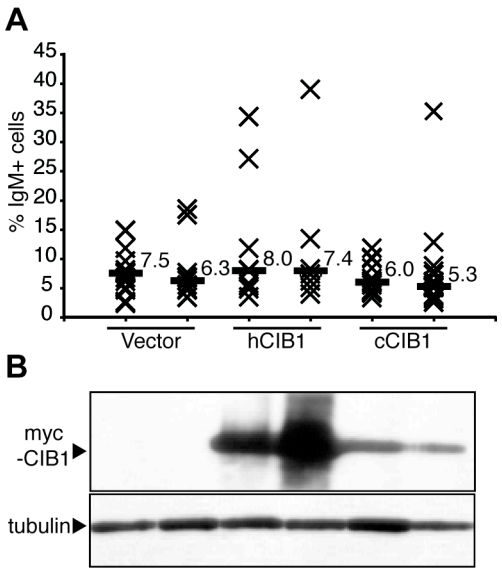


Figure 2-2. CIB1 over-expression in DT40. A. An IGC fluctuation analysis showing the percentage of surface Ig-positive cells in subclone cultures over-expressing human (h) or chicken (c) CIB1. Each X represents data from an individual subclone and the labeled horizontal bars report the medians for each data set. **B**. CIB1 over-expression confirmed by immunoblotting. Loading was controlled by stripping and re-probing the blot with an α-tubulin antibody.

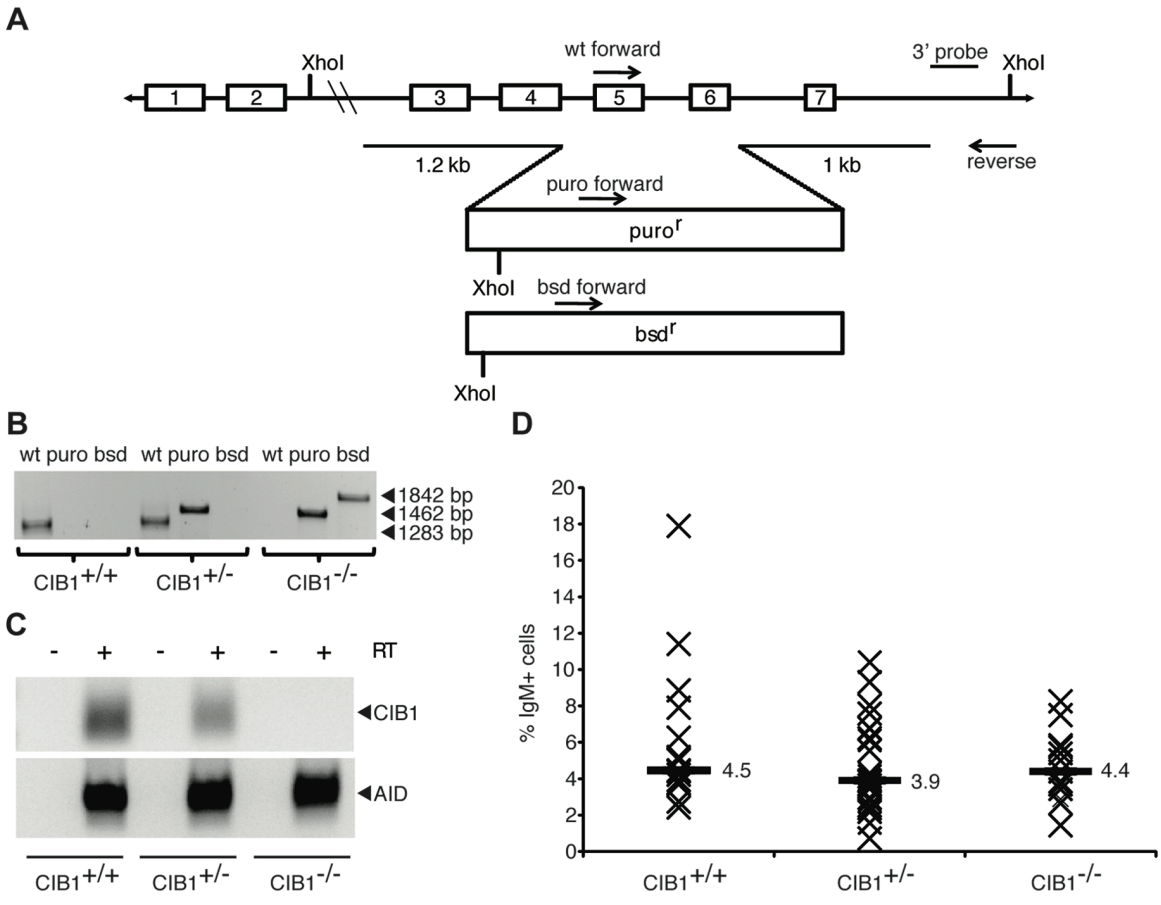

A
XhoI
wt forward
3' probe
XhoI
1
2
3
4
5
6
7
1.2 kb
1 kb
reverse
puro forward
puro^r
XhoI
bsd forward
bsd^r
XhoI
B
wt puro bsd wt puro bsd wt puro bsd
1842 bp
1462 bp
1283 bp
CIB1+/+
CIB1+/-
CIB1-/-
C
- + - + - + RT
CIB1
AID
CIB1+/+
CIB1+/-
CIB1-/-
D
% IgM+ cells
20
18
16
14
12
10
8
6
4
2
0
4.5
3.9
4.4
CIB1+/+
CIB1+/-
CIB1-/-

Figure 2-3. CIB1 is dispensable for immunoglobulin gene conversion in DT40. A. Schematic showing the constructs used to replace exons 5 and 6 of *CIB1* with the indicated drug resistance cassettes. The positions of the allele-specific PCR primers for genotyping and the XhoI sites and the position of the 3' external probe used for Southern blot analysis are shown. **B**. An agarose gel image showing the allele-specific PCR products from CIB1+/+, +/-, and -/- cell lines. **C**. An agarose gel image of CIB1-specific RT-PCR products from CIB1+/+, +/-, and -/- cell lines. AID-specific reactions were used to demonstrate the integrity of each cDNA preparation. **D**. An IGC fluctuation analysis showing the percentage of surface Ig-positive cells in subclone cultures of the indicated genotype. Each X represents data from an individual subclone and the labeled horizontal bars report the medians for each data set.

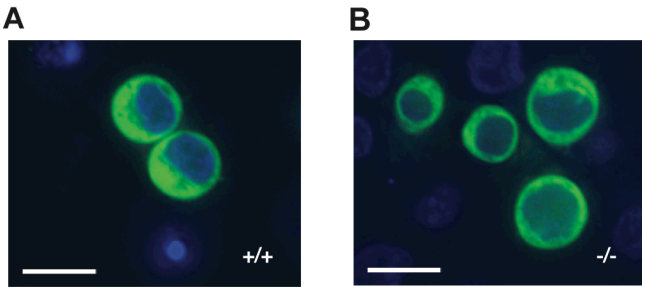


Figure 2-4. AID localization in CIB$^{-/-}$ DT40 cells. A. AID-EGFP localization in CIB1$^{+/+}$ DT40. **B**. AID-EGFP localization in CIB1$^{-/-}$ DT40. Images were taken using a 40X objective and the scale bars indicate 10 μm.

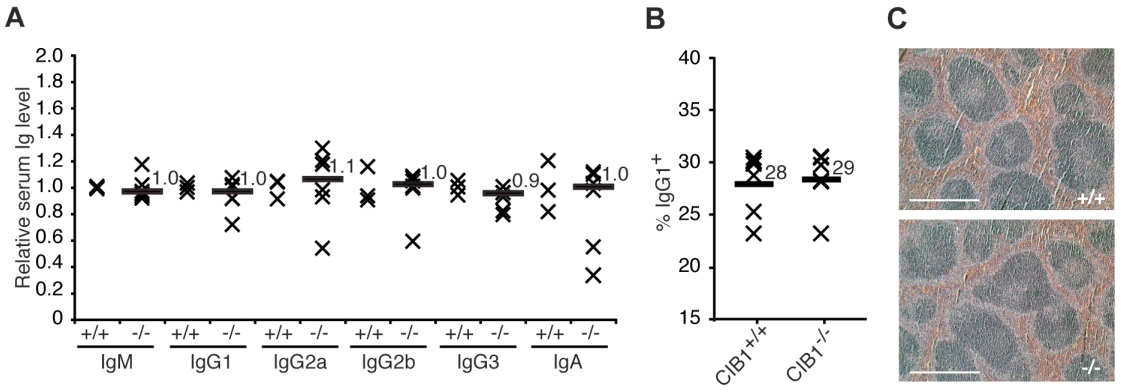


Figure 2-5. CIB1 is dispensable for CSR in mice. A. Relative levels of each antibody isotype in sera from CIB1$^{+/+}$ or CIB1$^{-/-}$ mice as measured by ELISA. The CIB1$^{-/-}$ data were normalized to the mean antibody levels in sera from wildtype (WT) littermates (arbitrarily set to 1 for comparison). Each X represents data from an independent animal and the horizontal bars and labels report the median values (n=3 for CIB$^{+/+}$ and n=6 for CIB$^{-/-}$). **B**. IgM to IgG1 CSR *ex vivo*. B-cells were isolated from the spleens of CIB1$^{+/+}$ or CIB1$^{-/-}$ mice, cultured for 4 days in the presence of LPS and IL-4, and analyzed by α-IgG1-PE labeling and flow cytometry. Each X represents data from an independent animal and the horizontal bars and labels report the median values. **C**. Images of hematoxylin and eosin stained sections of spleen isolated from CIB1$^{+/+}$ and CIB1$^{-/-}$ mice. Scale bars indicate 500 μm.

Chapter 3: Phosphorylation Directly Regulates the DNA Cytidine Deaminase Activity of AID and APOBEC3G

Adapted, with permission, from: Demorest et al. (2011) J Biol Chem Jun 9. [Epub ahead of print].
Z.L. Demorest performed the experiments for Figs. 3-1, 3-2, 3-3, 3-4, 3-5, 3-6, and 3-S2. M. Li generated the data for Fig. 3-S1.

Chapter 3 summary

The beneficial effects of DNA cytidine deamination by AID (antibody gene diversification) and APOBEC3G (retrovirus restriction) are tempered by probable contributions to carcinogenesis. Multiple regulatory mechanisms serve to minimize this detrimental outcome. Here, we show that phosphorylation of a conserved threonine attenuates the intrinsic activity of AID (T27) and APOBEC3G (T218). Phospho-null alanine mutants maintain intrinsic DNA deaminase activity, whereas phospho-mimetic glutamate mutants are inactive. The phospho-mimetic variants fail to mediate isotype-switching in activated mouse splenic B lymphocytes or suppress HIV-1 replication in human T cells. Our data combine to suggest a model in which this critical threonine acts as a phospho-switch that fine-tunes the adaptive and innate immune responses and helps protect mammalian genomic DNA from pro-carcinogenic lesions.

Introduction

Activation-induced deaminase (AID) and APOBEC3G (A3G) are the archetypal members of a larger family of polynucleotide cytidine-to-uridine (C-to-U) deaminases with critical functions in adaptive and innate immunity[8,107]. All mammals have AID, apolipoprotein B mRNA editing catalytic subunit 1 (APOBEC1), and APOBEC2 and variable numbers of APOBEC3s (A3) ranging from one in rodents to seven in most primates including humans (A3A/B/C/D/F/G/H)[3]. AID, A1, A3A, A3C, and A3H are single domain proteins with one zinc-coordinating active site, whereas several A3s including rodent A3 and human A3B, A3D, A3F, and A3G are double domain proteins with two zinc-coordinating motifs (both are conserved but typically only one is active). Atomic structures for the catalytic domain of human A3G[108-111] and APOBEC2[112] are available, and these enable structure-function studies and homology models.

Considerable effort has been devoted to understanding the multiple mechanisms that combine to regulate AID and A3G activity. First, the transcription of each of these genes is tissue specific, with AID being expressed predominantly in B lymphocytes and A3G in most cell types[6,38,113]. Second, AID and A3G transcription levels are upregulated by distinct signal transduction pathways (STAT/NFκB for AID and NFAT/IRF for A3G)[114,115]. Third, AID expression is regulated by at least one micro RNA miR-155[43,44]. Fourth, both proteins are predominantly cytoplasmic, with AID having additional nuclear import and export capabilities[51-53,116,117]. Finally, both proteins are subject to proteasome-dependent degradation, AID in the nuclear compartment[50] and A3G in the cytoplasmic compartment[118-120].

In addition, numerous proteins have been implicated in regulating AID and A3G function [MDM2[60], RPA[64], HSP90[59], GANP[58] CIB1[121], CTNNBL1[56]], with protein kinase A [(PKA[61-63,69]] being most relevant to the work described here. PKA phosphorylates AID at threonine-27 and serine-38, with serine-38 phosphorylation promoting interactions with replication protein A (RPA) and facilitating class-switch recombination (CSR) and somatic hypermutation (SHM)[61-63,67]. Phosphorylation of the homologous residue in A3G, threonine 32, also has functional consequences by rendering the protein less susceptible to HIV-1 Vif-induced ubiquitylation and degradation[69]. Here, we provide evidence in support of a new role for threonine/serine phosphorylation in directly suppressing the intrinsic DNA deaminase activity of AID and A3G. Extensive conservation of this particular residue suggests that phospho-regulation may extend to most other DNA deaminase superfamily members.

Results

AID-T27, A3G-T32, and A3G-T218 are homologous and located within a region of high sequence and structural conservation.

Prior studies demonstrated phosphorylation of AID-T27 *in vivo* and *in vitro* by mass spectrometry and radiolabeling[61,63,67] and A3G-T32 by immunoblotting[69]. We noted that these two threonines are homologous to A3G-T218, which high-resolution structures have shown to be located within a solvent-accessible loop[108-111] (**Fig. 3-1A**). This threonine anchors a conserved motif that matches a consensus PKA phosphorylation site (R-H/R-x-T)[122] (**Fig. 3-1B**). Notably, nearly all AID/A3 family members have homologous threonine or serine residues at this exact position (**Fig. 3-1C**). Rare exceptions are only apparent in specific mammalian lineages (carnivores and rodents) or in redundant or inactive domains (most alleles of human A3H are unstable)[123]. In the catalytic domain of A3G, the first arginine in this motif (R215) is located adjacent to the catalytic glutamate, and it has been implicated in binding substrate single-stranded DNA[109-111]. Taken together, these observations and particularly the high level of conservation and the structural positioning next to the active site suggest that phosphorylation and dephosphorylation may serve as a post-translational switch that helps control the DNA deaminase activity of these mutagenic enzymes.

PKA and CaMKII phosphorylate A3G-T218 in vitro.

AID-T27 and A3G-T32 can be phosphorylated by PKA[61,63,67,69]. To determine whether these observations extend to A3G-T218, we asked whether recombinant PKA could phosphorylate a peptide representing the soluble loop in which this residue resides,

VRGRHET$_{218}$YLCYE. We found that PKA could readily phosphorylate this peptide but not a T218A mutant derivative that is otherwise identical (**Fig. 3-1D**). Similarly, CaMKII, which also phosphorylates R-x-x-T motifs[122], was able to phosphorylate the A3G-T218 peptide but not the alanine mutant derivative. Both enzymes were also able to phosphorylate a serine in a control peptide (Kemptide). These data demonstrate that A3G-T218 is a suitable substrate for at least two kinases PKA and CaMKII.

Phospho-mimetic mutations inhibit DNA cytidine deaminase activity.

To address whether phosphorylation is capable of attenuating the DNA cytidine deaminase activity of AID and A3G, phospho-null and -mimetic derivatives of these proteins were tested in an *E. coli*-based activity assay. The rifampicin-resistance (Rif^R) mutation assay has been used extensively to assess intrinsic DNA cytidine deaminase activity[7,124]. Consistent with prior reports, AID and A3G triggered 3- and 4-fold increases in the median Rif^R mutation frequency compared to catalytically inactive controls, AID-E58Q and A3G-E259Q (**Fig. 3-2A, B**). In comparison, phospho-mimetic AID-T27E and A3G-T218E proteins also showed greatly reduced activity approaching background levels. Phospho-null alanine mutants showed slightly higher levels of mutator activity. Mutation of another predicted surface threonine in AID (T140) or the homologous threonine in the non-catalytic N-terminal domain of A3G (T32) had little effect. All proteins expressed similarly in *E. coli*, indicating that these data are not due to variable protein expression levels (lower panels in **Fig. 3-2A, B**).

To ask whether these observations extended to A3G purified from human cells, we used a DNA oligonucleotide deamination assay optimized to measure catalytic activity. As expected, wildtype A3G catalyzes dose-dependent C-to-U deamination of labeled deoxy-oligonucleotide substrates, which following uracil excision and NaOH-mediated phosphodiester backbone cleavage is detected as a shorter DNA fragment (**Fig. 3-3A, C, E**). As anticipated from the *E. coli* mutation experiments, the A3G phospho-mimetic variant T218E showed considerably lower levels of catalytic activity. Interestingly, the A3G phospho-null variant T218A showed significantly elevated levels of catalytic activity consistent with a proportion of the wildtype protein being already phosphorylated (and thereby inactivated) in HEK293T cells. Taken together, the *E. coli* and the purified protein activity data indicate that phosphorylation of the conserved threonine, AID-T27 or A3G-T218, may serve to attenuate the intrinsic DNA deaminase activity of these proteins (supported further by HEK293T cell extract data in **Fig. 3-S1**).

DNA binding is unaffected by phospho-mimetic substitutions.

To ask whether the diminished catalytic activity of the phospho-mimetic substitution mutants is due to defective ssDNA binding, we tested the ssDNA binding ability of AID and A3G in electrophoretic mobility shift assays. A3G-myc-His used in the deaminase reactions above was used for ssDNA binding experiments. Purified protein was diluted serially, incubated with a fluorescently labeled oligo, and fractionated on a native polyacrylamide gel. As expected, A3G and the catalytic mutant E259Q bound ssDNA in a dose dependent manner (**Fig. 3-3B, D, E**)[125]. Likewise, A3G-T218A and

A3G-T218E had nearly identical ssDNA binding abilities, which were indistinguishable from the wildtype enzyme (**Fig. 3-3B, D, E**).

Similar EMSA experiments were done with wildtype AID and mutant derivatives, except the sensitivity of the assay had to be increased by using a radiolabeled ssDNA oligo substrate. Again, the wildtype and the phospho-null and -mimetic variants produced near identical mobility shifts (**Fig. 3-3F, G, H**). As a control to demonstrate AID's specificity for ssDNA, an AID-R24E mutant was analyzed in parallel and shown to be defective in DNA binding. This arginine is conserved and homologous to A3G-R215, which NMR chemical shift perturbation and mutagenesis experiments have implicated strongly in DNA binding[109-111]. Additional EMSA data can be found in **Fig. 3-S2**. Overall, these EMSA results clearly show that phospho-mimetic substitutions in A3G and AID do not cause visible decreases in each protein's ability to bind ssDNA.

Mutants of AID and A3G localize normally within living cells.

The subcellular localization of AID/APOBEC family members has been well studied[51-53,116,117]. A3G is predominantly cytoplasmic. AID is also mostly cytoplasmic, but it is imported into the nuclear compartment by an importin-alpha pathway and exported back to the cytoplasm by a CRM1 pathway. To ask whether our phospho-null or phospho-mimetic mutants retain normal, steady-state subcellular distributions, we performed a series of AID/A3G-GFP localization studies in living HeLa cells. No detectable alteration in the steady-state cytoplasmic distribution of A3G-EGFP, AID-EGFP, or their mutant derivatives was detected (**Fig. 3-4A, B**). Moreover, experiments

done in the presence and absence of the CRM1 inhibitor leptomycin B indicated that the nuclear import and export activities were also intact for all AID-EGFP constructs. These data therefore indicate that A3G, AID, and their mutant derivatives are capable of interacting with the cellular factors responsible for localization and, furthermore, that AID is able to enter the nucleus where it will have the opportunity to access the immunoglobulin loci, its physiologic DNA deamination target.

AID-T27E is defective for class-switch recombination.

One of the physiological functions of AID is catalyzing C-to-U deamination events in immunoglobulin heavy chain gene switch region DNA, and thereby triggering additional DNA repair processes that ultimately manifest as antibody isotype switch recombination[5,8]. Therefore, as a functional test of AID activity, we assayed the phospho-null and phospho-mimetic mutants in an *ex vivo* B-cell CSR system[56,61-63]. Naïve splenic B-lymphocytes were isolated from AID-deficient mice, cultured in the presence of IL-4 and LPS to induce cell division and isotype switching from IgM to IgG1, transduced with retroviruses encoding AID-IRES-EGFP, mutants of AID, or EGFP alone, and four days later subjected to flow cytometry for IgG1-positve cells. Mock (not shown) or EGFP virus transduced cells remained AID-defective and showed no class switching to IgG1 (**Fig. 3-5**). Also, as expected, wildtype AID expression complemented the endogenous defect and enabled class switching to IgG1 in a significant proportion of cells (representative plots in **Fig. 3-5A** and average of 4 experiments in **Fig. 3-5B**). Conversely

and surprisingly, neither T27A nor T27E was capable of promoting the switch to IgG1 despite similar protein expression levels (**Fig. 3-5A, B, C**).

The T27E result was anticipated based on the lower level of catalytic activity, but not DNA binding or localization activities, elicited by this mutant. However, the aforementioned data on AID-T27A showing normal deaminase, ssDNA binding, and cellular localization/trafficking activities strongly suggested that this variant would be capable of normal or even elevated CSR levels, in stark contrast to the defect in CSR shown here. This result makes the CSR data set more difficult to interpret. One possibility, noted previously[61], is that phosphorylation of S38 may depend first upon phosphorylation of T27. An alternative may be that each of these residues has a distinct mechanistic contribution to CSR, with our studies favoring a role for T27 in regulating catalysis.

A3G-T218E lacks HIV-1 restriction activity.

A3G potently inhibits HIV-1 replication by blocking reverse transcription and deaminating viral cDNA cytosines to uracils[107]. This antiviral activity is most evident in HIV-1 lacking viral infectivity factor (Vif), a small basic protein that triggers A3G degradation. Thus, a rigorous test of A3G's functional activity is whether it suppresses the spreading infection of Vif-deficient HIV-1[126-128]. We therefore created a panel of CEM-SS T cell lines stably expressing wildtype A3G-EGFP, an E259Q catalytic mutant control, a phospho-null T218A construct, or a phospho-mimetic T218E protein. As anticipated from prior studies, wildtype A3G completely suppressed the replication of

Vif-deficient HIV-1 and its strong antiviral effect was largely dependent upon the integrity of the catalytic glutamate E259[29,129] (**Fig. 3-6A**). A3G-T218A showed wildtype levels of restriction consistent with full or elevated levels of enzymatic activity. In contrast, A3G-T218E failed to prevent the replication of Vif-deficient HIV-1. However, this mutant protein did cause reproducible delays in peak viral replication consistent with severely attenuated but not fully defective catalytic activity. As additional controls, N-terminal A3G-T32A or T32E substitutions had no discerable effect, and all cell lines supported similar levels of Vif-proficient HIV-1 spreading infection (**Fig. 3-6A** and data not shown). It is notable that, although we were able to confirm A3G-T32 phosphorylation by mass spectrometry, we found no differences in the subcellular localization, HIV restriction capacity, or Vif susceptibility in alanine or glutamate substituted derivatives (**Fig. 3-4** and data not shown).

An additional possibility is that A3G-T218E may not restrict HIV-1 because it is not efficiently packaged into budding virions. To test and eliminate such a possibility, we harvested virus produced from HEK293T cells expressing A3G and mutants thereof and blotted for the presence of A3G in these viral particles. We found no significant difference in the ability of any of the mutants to get into virions as compared to wildtype A3G (**Fig. 3-6C**).

Discussion

The AID/APOBEC family of cytidine deaminases is an important facet of the adaptive and innate immune responses in humans. However, their mutagenic activity must be tightly regulated in order to prevent potentially detrimental off-target effects. Regulation of these proteins has been described at multiple levels including transcription, micro-RNAs, cytoplasmic localization, proteasomal degradation, and phosphorylation (see Introduction). Here we describe a novel phosphorylation regulatory mechanism capable of attenuating the intrinsic deaminase activity of AID and A3G. In this study, we demonstrate that phospho-mimetic substitution of a highly conserved threonine renders these proteins inactive in several independent assays. We show that ssDNA binding ability and steady-state subcellular localization (and for AID also trafficking) are unaffected, indicating that these proteins are structurally intact. In functional assays, this modification prevents AID from facilitating CSR and A3G from restricting HIV-1ΔVif replication. It is intriguing that two neighboring phosphorylation sites can have such contrasting effects on AID's function, with S38 phosphorylation enabling interaction with RPA and allowing CSR and SHM, and T27 phosphorylation rendering the protein inactive. This begs the question of how PKA is regulated to distinguish between these neighboring residues. Further studies are warranted to better understand these post-translational regulatory events and investigate the possible involvement of other Ser/Thr kinases that can also recognize PKA consensus motifs such as CaMKII described here.

The obvious utility of post-translational regulation by phosphorylation is two-fold (illustrated by model in **Fig. 3-S3**). First, a threonine or serine phosphorylated DNA

deaminase would possess a low level of DNA deaminase activity and pose less of a threat to genomic DNA. Genomic DNA integrity is further ensured by the fact that AID, A3G and many other A3 proteins are predominantly cytoplasmic. Second, signal transduction pathways, which are critical for both adaptive and innate immune responses, could readily switch-on DNA deaminase activity by triggering the removal of the phosphate group (phosphatase or phospho-transferase activity). This would ensure an expedited immune response that could be further bolstered by up-regulating AID or A3 expression at the transcriptional and/or translational levels.

We propose that the post-translational modification of AID and the A3 proteins by phosphorylation provides a means of directly controlling the intrinsic DNA cytidine deaminase activity of these proteins (**Fig. 3-S3**). It is likely that this mechanism will be conserved in vertebrates because residues homologous to AID-T27, A3G-T32 or -T218 are apparent in almost all other known polynucleotide cytidine deaminase family members (**Fig. 3-1**). It is further possible that defects in these signal transduction pathways may manifest as immunodeficiency syndromes (over-phosphorylated protein), autoimmune diseases (under-phosphorylated protein), and/or carcinogenesis (under-phosphorylated protein), especially in combination with other regulatory defects.

Experimental procedures

DNA constructs- pEGFP-N3-A3G and pEGFP-N3-AID have been described[27,28]. Mutants of AID and A3G were made by Quickchange site directed mutagenesis (Stratagene). The retroviral vector pMX-EGFP was constructed by subcloning EGFP from plasmid pEGFP-N3 (Clontech) into pMX-PIE (a gift from V. Barreto) using EcoRI/NotI. pMX-AID-IRES-EGFP was generated by first amplifying untagged AID from pEGFP-N3-AID by PCR using primers (5'- GCT AGC GCC ACC ATG GAC AGC CT) and (5'- CCT GCA GGT CAA AGT CCC AAA GTA). The insert was cut with NheI/SbfI and ligated into pCSII-IRES-EGFP (a gift from N. Somia). The AID-IRES-EGFP insert was then PCR amplified using primers (5'- GAA TTC ATG GAC AGC CTC TTG ATG AAC) and (5'- CCA CAT AGC GTA AAA GGA GCA AC), cut using EcoRI/NotI, and ligated into pMX-PIE. The MLV amphotrophic envelope vector pRK5-10A1 and the HIV-1 accessory vector ΔNRF were generous gifts from N. Somia. The MLV accessory vector pMD-OGP was provided by F. Randow. The Vif-deficient HIV-1_{IIIB} provirus has been used previously[126,128]. The *E. coli* expression constructs pTrc99a-AID-GST and pTrc99a-A3G-GST were generated by subcloning AID from pEGFP-N3-AID or A3G from pEGFP-N3-A3G into pTrc99a-GST using NcoI/SalI. Untagged bacterial expression plasmids pTrc99a-hAID and pTrc99a-A3G have been described[27].

In vitro *peptide kinase assays* – CaMKII (New England Biolabs) or PKA [New England Biolabs or gift from L. Masterson and G. Veglia[130]] were incubated with 1 ug of either kemptide (CLRRASLG; American Peptide Company), a peptide containing A3G-T218 (VRGRHETYLCYE; New England Peptides), or an A3G-T218A mutant peptide

(VRGRHEAYLCYE; New England Peptides) in the manufacturers recommended buffer supplemented with ^{32}P-γ-ATP. Kinase reactions were incubated at 30°C for 1 hour before fractionating on a 16% Tricine/urea-acrylamide gel. The gel was dried and imaged by phosphorimaging (FLA-5000, Fuji).

E. coli *mutation assays* - The rifampicin resistance assay has been published[7,124]. BW310 *E. coli* cells were transformed with pTrc99a-AID or pTrc99a-A3G expression constructs and plated on ampicillin containing media. Four individual colonies were picked and seeded into media containing 1 mM IPTG and 100 μg/ml ampicillin. After shaking overnight at 37°C, the cultures were plated on ampicillin media to obtain viable cell counts and to 100 μg/ml rifampicin containing media for mutational frequency. AID and A3G expression levels were determined by western blot using antibodies against AID (EK25G9, Cell Signaling) or A3G (#10201 rabbit anti-A3G polyclonal serum provided by J. Lingappa through the AIDS Research and Reference Reagent Program).

Recombinant protein preparations- Protein samples for AID/A3G were prepared by growing 500 mL cultures of BL21 *E. coli* cells transformed with pTrc99a-GST, pTrc99a-AID-GST, pTrc99a-A3G-GST, or mutants of AID/A3G. The cultures were centrifuged and the cell pellet was resuspended in lysis buffer [50 mM Tris-Cl pH 7.9, 200 mM NaCl, 50 μM $ZnCl_2$, Complete protease inhibitors (Roche)]. After centrifugation (17,000 x g, 20 minutes), clarified lysates were then incubated overnight at 4°C with Glutathione Sepharose (GE Healthcare). The resin was washed 4 times with lysis buffer and purified protein was eluted by cleavage with TEV protease (Invitrogen).

For preparation of A3G from human cells, pcDNA3.1-A3G-myc-his or mutants thereof were transfected into HEK293T cells cultured in DMEM (Gibco) and 10% FBS (Hyclone) using TransIT-LT1 reagent (Mirus). After 48 hours, cells were lysed in buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$ 1 mM ZnCl, 10% glycerol) and bound to Ni-NTA agarose beads (Qiagen). The beads were washed and purified protein was eluted using imidazole (250 μM), as described[131].

DNA binding assays- Protein samples prepared from *E. coli* as above were used for AID DNA binding reactions. 6 pmol of purified protein was diluted serially 1:2 and mixed with 0.5 pmol ^{32}P labeled oligo (5'-ATT ATT ATT ATT CCA ATG GAT TTA TTT ATT TWR CTA TTT ATT T) in binding buffer (10 mM HEPES pH 7.6, 10% glycerol, 100 mM KCl, 10 mM $MgCl_2$, 100 μM EDTA, 500 μM DTT). Reactions were incubated for 30 minutes at 37°C before separation on a 7% TBE-acrylamide gel. The gel was dried and imaged by phosphoimager (Storm, Molecular Dynamics).

Protein samples purified from human cells from above were used for A3G DNA binding reactions. 25 pmol of purified protein was diluted serially 1:2 and each dilution was mixed with 1 pmol fluorescein labeled oligo (5'-ATT ATT ATT ATT CCA ATG GAT TTA TTT ATT TAT TTA TTT ATT T-fluorescein) in binding buffer. The reactions were incubated for 30 minutes at 37°C before separation on a 7% TBE-acrylamide gel. The free and bound oligos were then detected by fluorescence imaging (FLA-5000, Fuji).

Oligo based deaminase assays- Protein samples purified from human cells as above were used for A3G deaminase activity reactions. Starting with 1.2 pmol purified

protein, 2-fold serial dilutions were made and mixed with 1 pmol of substrate oligo (5'-ATT ATT ATT ATT CCA ATG GAT TTA TTT ATT TAT TTA TTT ATT T-fluorescein), 0.1 μg/μl of RNase A (Qiagen) and 0.001 U/μl of UDG (NEB). The reactions were incubated at 37°C for 2 hours and then NaOH was added to 100 μM before incubating at 90°C for another 30 min. The reactions were separated on a 16% Tris/Urea-acrylamide gel and visualized by fluorescence imaging (FLA-5000, Fuji).

Fluorescence microscopy studies- For steady state A3G and AID localization, pEGFP-N3-A3G, pEGFP-N3-AID or mutants thereof were transfected into HeLa cells grown in DMEM (Gibco) supplemented with 10% FBS (Hyclone). 48 hours later, the cellular localization was determined by fluorescence microscopy (Deltavision). For AID import activity, leptomycin-B (20 ng/mL) was added 2 hours prior to imaging as above.

HIV restriction assays- CEM-SS and CEM-GFP (courtesy of M. Malim) were maintained in RPMI 1640 (Gibco) supplemented with 10% FBS (Hyclone). Stable cell lines expressing pEGFP-N3, pEGFP-N3-A3G or mutants were generated in the permissive cell line CEM-SS by electroporating linearized DNA and selecting for stable integrants with 1 mg/ml G418 (Mediatech) as described[126]. Clones were confirmed to have similar expression levels by western blot using an antibody against A3G (rabbit polyclonal raised against a C-terminal peptide). Virus was produced by transfecting HIV-1 provirus using TransIT-LT1 (Mirus) HEK293T cells maintained in DMEM (Gibco) supplemented with 10% FBS (Hyclone). 48 hours after transfection, virus-containing supernatants were filtered with a 0.45 μm filter. Viruses were then titered using the CEM-GFP reporter cell line as described[126]. Spreading infections were initiated by adding

virus to CEM-SS stable cell lines at an MOI of 0.05. Supernatants from infected cultures were collected at 2-4 day intervals and added to CEM-GFP. After 48-hours, the cells were fixed in 4% paraformaldehyde and analyzed for GFP expression by flow-cytometry (Quanta SC MPL, Beckman Coulter). Procedures for the detection of A3G in producer cells and viral particles have been described[132].

Class-switch recombination assays- All experiments were conducted in accordance with the University of Minnesota Animal Care and Use Committee guidelines. The C57BL/6 $AID^{-/-}$ mice have been described[5]. *Ex vivo* CSR assays were conducted by purifying resting B-cells from spleen by magnetic sorting (130-090-862, Miltenyi Biotech). Isolated B-cells were then cultured in RPMI supplemented with 10% FBS, 50 ng/ml IL-4, and 50 μg/ml LPS. After 48 hours, the media was replaced with transducing viral supernatant supplemented with 20 mM HEPES and 16 μg/ml polybrene and centrifuged (600 x g, 2 hours, 30°C). The cells were then resuspended into fresh media containing IL-4 and LPS and cultured for 4 days. Efficiency of switching to IgG1 was determined by staining with anti-IgG1-PE (BD Biosciences) and analyzed by flow-cytometry (FACSCanto II, BD Biosciences).

Figures

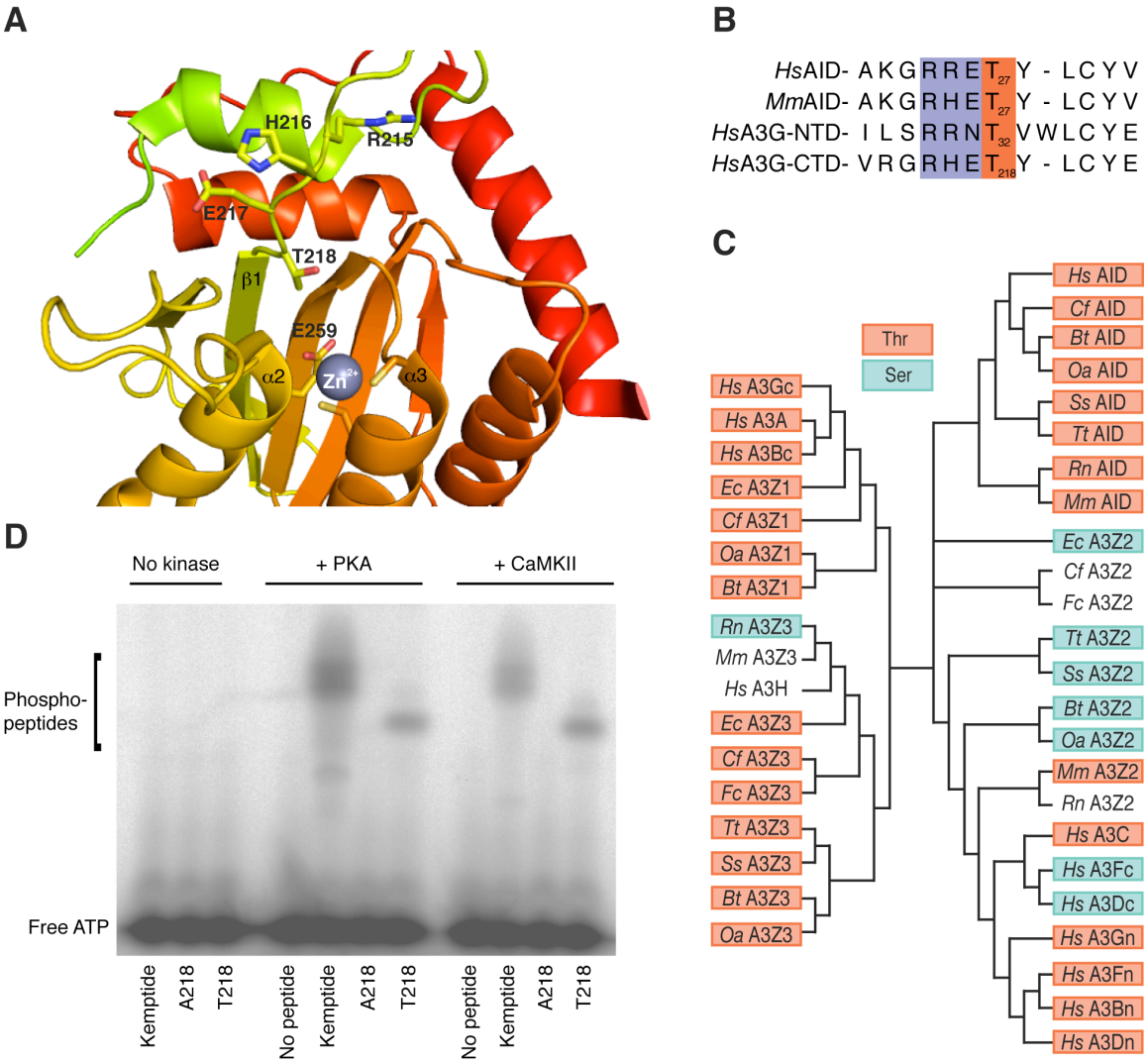

A
H216
R215
E217
T218
β1
E259
α2
Zn2+
α3
B
HsAID- A K G R R E T27 Y - L C Y V
MmAID- A K G R H E T27 Y - L C Y V
HsA3G-NTD- I L S R R N T32 V W L C Y E
HsA3G-CTD- V R G R H E T218 Y - L C Y E
C
Thr
Ser
Hs A3Gc
Hs A3A
Hs A3Bc
Ec A3Z1
Cf A3Z1
Oa A3Z1
Bt A3Z1
Rn A3Z3
Mm A3Z3
Hs A3H
Ec A3Z3
Cf A3Z3
Fc A3Z3
Tt A3Z3
Ss A3Z3
Bt A3Z3
Oa A3Z3
Hs AID
Cf AID
Bt AID
Oa AID
Ss AID
Tt AID
Rn AID
Mm AID
Ec A3Z2
Cf A3Z2
Fc A3Z2
Tt A3Z2
Ss A3Z2
Bt A3Z2
Oa A3Z2
Mm A3Z2
Rn A3Z2
Hs A3C
Hs A3Fc
Hs A3Dc
Hs A3Gn
Hs A3Fn
Hs A3Bn
Hs A3Dn
D
No kinase
+ PKA
+ CaMKII
Phospho-peptides
Free ATP
Kemptide
A218
T218
No peptide
Kemptide
A218
T218
No peptide
Kemptide
A218
T218

Figure 3-1. An active site Thr/Ser is conserved in mammalian DNA cytidine deaminases. *A*. T218 is positioned adjacent to the catalytic glutamate E259 in the crystal structure of A3G (3IR2)[108]. The residues comprising the kinase recognition sequence and the side chains of the two cysteines that coordinate zinc are also shown. *B*. An alignment of the PKA consensus sequence, R-x-x-T, found in human AID, mouse AID, human A3G C-terminal domain (CTD), and human A3G N-terminal domain (NTD). *C*. A phylogenetic analysis showing the conservation of active site Thr/Ser residues in the AID/A3 proteins of nearly all mammals. Domains highlighted in orange represent the presence of a threonine and domains highlighted in cyan contain a serine. *D*. PKA and CaMKII can phosphorylate A3G-T218 *in vitro*.Abbreviations: *Hs* = human, *Mm* = mouse, *Rn* = rat, *Bt* = cow, *Oa* = sheep, *Ss* = pig, *Tt* = peccary, *Ec* = horse, *Cf* = dog, *Fc* = cat.

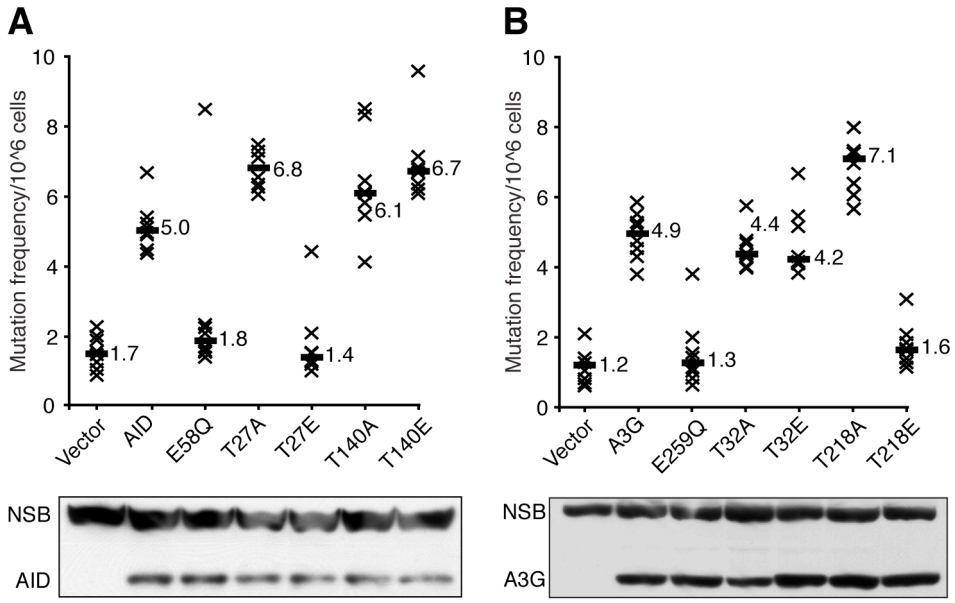


Figure 3-2. Intrinsic DNA cytidine deaminase activity of AID and A3G constructs. *A*,*B*. Results from *E. coli*-based Rif^R mutation assays with each X representing data from an independent culture. Median mutation frequencies are indicated by horizontal bars and numbers. Western blots of AID or A3G constructs from representative cultures with a non-specific band (NSB) as a loading control.

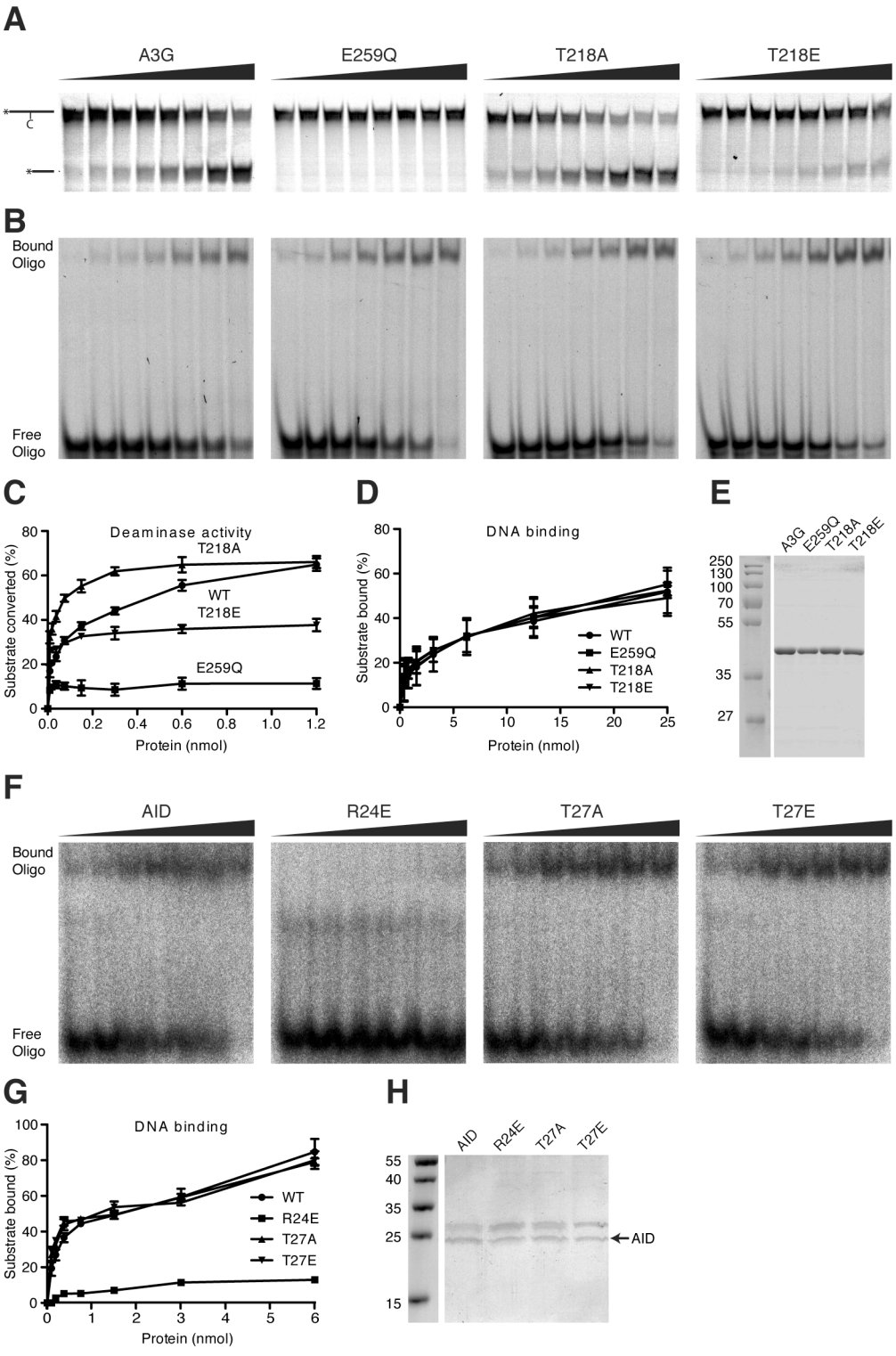

A
A3G
E259Q
T218A
T218E
C
B
Bound Oligo
Free Oligo
C
Deaminase activity
T218A
WT
T218E
E259Q
Substrate converted (%)
Protein (nmol)
D
DNA binding
WT
E259Q
T218A
T218E
Substrate bound (%)
Protein (nmol)
E
A3G
E259Q
T218A
T218E
250
130
100
70
55
35
27
F
AID
R24E
T27A
T27E
Bound Oligo
Free Oligo
G
DNA binding
WT
R24E
T27A
T27E
Substrate bound (%)
Protein (nmol)
H
AID
R24E
T27A
T27E
55
40
35
25
15
AID

Figure 3-3. A3G-T218E has diminished deaminase activity but retains DNA binding ability. *A*. Representative images from titrated A3G oligo deaminase assays. The upper band is the intact oligo and the lower band is the product of deamination, uracil excision, and strand cleavage. *B*. Representative images from A3G ssDNA binding assays. Free oligo and protein bound complexes are labeled. *C*. Quantification of A3G deaminase activity in (a) and replicas not shown. Data are plotted as the mean of 3 independent experiments +/- S.D. *D*. Quantification of A3G EMSA data in (b). Data are plotted as the mean of 3 independent experiments +/- S.D. *E*. Coomassie stained gel illustrating the purity of A3G enzymes used in these experiments. *F*. Representative images from AID ssDNA binding assays. *G*. Quantification of AID EMSA data in (f). Data are plotted as the mean of 3 independent experiments +/- S.D. *H*. Coomassie stained gel illustrating the purity of AID enzymes used in the ssDNA binding assays. The identity of the AID bands was confirmed by immunoblotting (not shown).

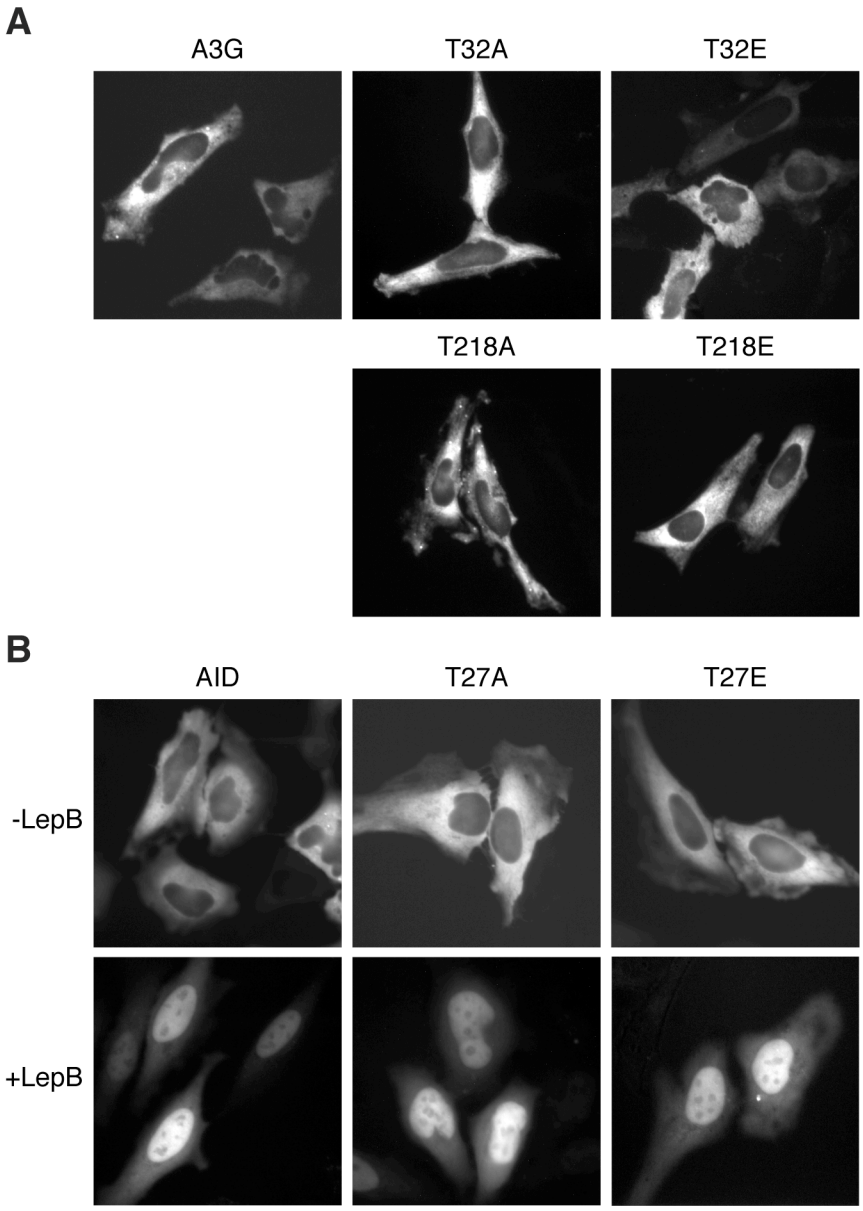


Figure 3-4. Mutants of A3G and AID localize normally within living cells. *A*. Representative fluorescent images of A3G-EGFP localization in HeLa cells. *B*. Representative images of AID-EGFP localization in HeLa cells in the presence or absence of leptomycin-B (LepB).

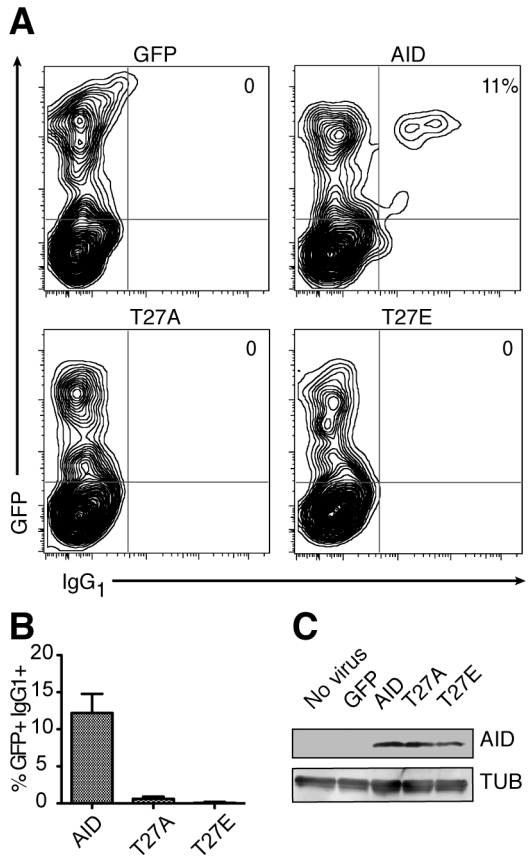


Figure 3-5. AID-T27E is defective for class switch recombination. *A*. Representative flow cytometry plots of stimulated B lymphocytes transduced with the indicated human AID-IRES-EGFP constructs. Transduction is indicated by GFP expression and CSR by IgG1 expression. *B*. A histogram summarizing the CSR activity from 4 independent experiments (mean and S.D. of the percentage of IgG1 cells within the GFP-positive transduced population). *C*. A representative immunoblot of AID expression with tubulin (TUB) as a loading control.

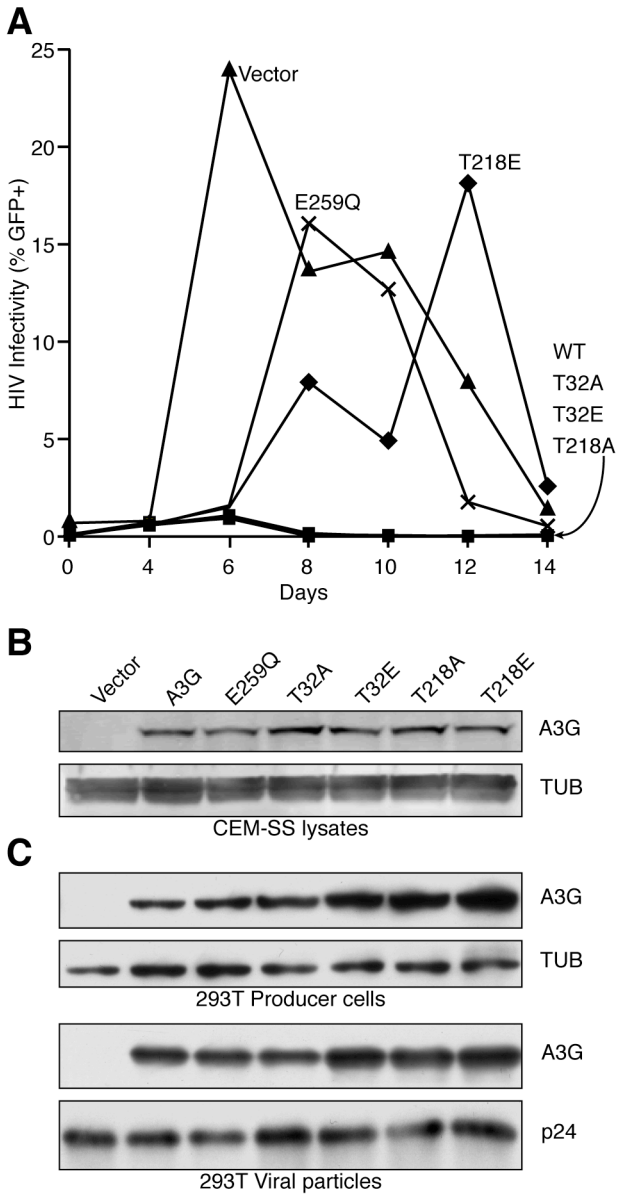


Figure 3-6. A3G-T218E fails to restrict Vif-deficient HIV-1. *A*. The kinetics of Vif-deficient HIV-1 replication in the indicated stable A3G expressing T cell lines. An MOI of 0.05 was used to initiate infection on day 0, and viral infectivity was measured on subsequent days by titering cell-free supernatants on CEM-GFP indicator cells. *B*. A representative immunoblot of A3G-EGFP (anti-GFP) expression with tubulin (TUB) as a loading control. *C*. Western blots indicating the presence of A3G in producer cells and viral particles. TUB and p24 are loading controls for cells and viral particles respectively.

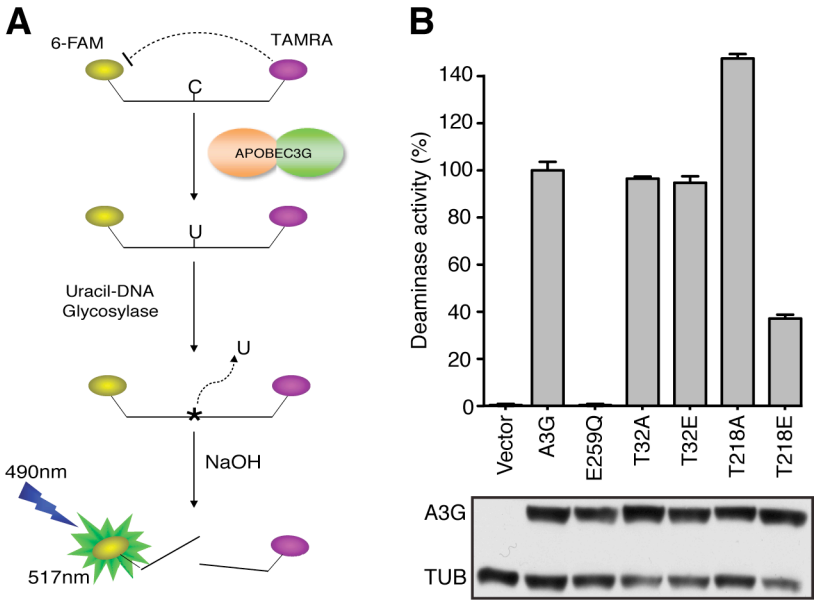


Fig. 3-S1. A3G-T218E has reduced deaminase activity in human cell lysates. *A*. Schematic representation of the FRET-based DNA deaminase assay. *B*. Results from a FRET-based DNA cytidine deaminase assay for the indicated A3G constructs expressed in HEK293T cell lysates. Percent activity is normalized to wildtype and plotted as mean activity from 4 independent experiments +/- S.D. An anti-GFP western blot indicates similar levels of A3G-EGFP protein expression with tubulin (TUB) serving as a loading control.

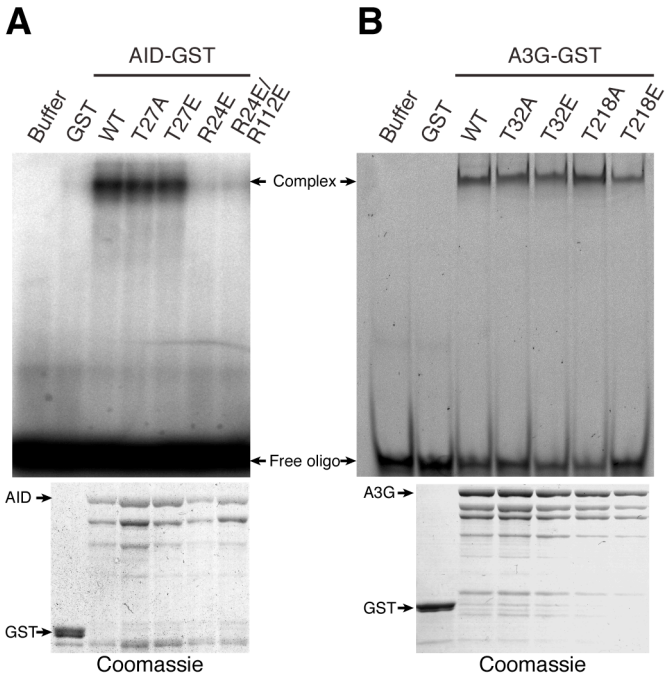


Fig. 3-S2. Phospho-mimetic mutants retain DNA binding ability. *A*. An autoradiogram of a ^{32}P-labeled deoxy-oligonucleotide supershifted by AID binding. A coomassie stained gel of the indicated protein preparations is shown below. The identity of the GST and AID-GST bands was confirmed by immunoblotting (not shown). *B*. An image of fluorescein-labeled deoxy-oligonucleotide supershifted by A3G binding. A coomassie stained gel of the indicated protein preparations is shown below. The identity of the GST and A3G-GST bands was confirmed by immunoblotting (not shown). The migration of the DNA complexes in panels (a) and (b) is coincident and not intended to reflect the underlying protein stoichiometries.

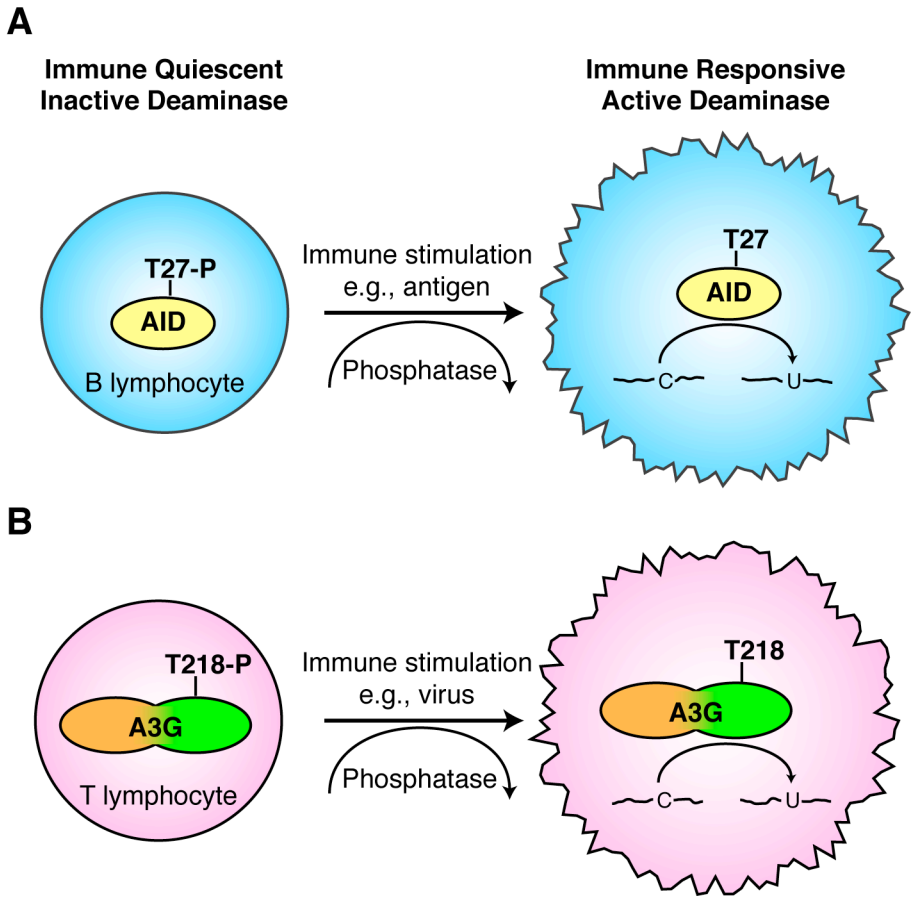


Fig. 3-S3. General model for the activation of DNA deaminases during immune activation. *A*. Schematic representation of a quiescent B lymphocyte with inactive AID phosphorylated at T27. Upon stimulation by antigen, the phosphate is removed to activate the deaminase. *B*. Schematic diagram of inactive A3G phosphorylated at T218 in a quiescent T lymphocyte. Immune stimulation by viral infection would facilitate the activation of A3G through phosphatase activity.

Chapter 4: Conclusions and Discussion

The overarching theme of this thesis has been mechanisms of post-translational regulation of the AID/APOBEC family of cytidine deaminases. The major findings are the identification and characterization of two novel AID interacting proteins that potentially have roles outside of the B lymphocyte compartment (**Chapter 2 and Appendix I**), and the discovery that phosphorylation of AID and A3G at a threonine conserved in the AID/APOBEC family negatively regulates deaminase activity (**Chapter 2**). **Appendix II** describes the identification of cis-acting sequences involved in the nucleocytoplasmic trafficking activities of AID. The following is a discussion of the implications, complexities and questions raised by these findings.

Post-translational modifications

Evidence suggests that APOBEC proteins may be differentially modified when expressed in different cell types. For example, A3G isolated from the human kidney cell line, HEK293, is fully active in our *in vitro* deaminase assays. However, when similar quantities are purified from *E. coli* or from human T lymphocytes (e.g., CEM-SS), the protein has little to no detectible activity. We hypothesize that these differences may be due to alternative modifications on A3G, or alternatively that certain cell lines may possess a specific inhibitory protein. Our lab and others have attempted to determine the causative modifications (or inhibitory proteins) responsible for this phenotype, but they remain elusive[133].

During the course of these studies, several modifications were identified by mass-spectrometry work conducted by Nevin Krogan and Jeff Johnson at UCSF. Analyses of APOBEC3G purified from Sf9 insect cells using the baculovirus system revealed a number of phosphorylation, formylation and ubiquitination sites (**Fig. 4-1**). In addition to confirming phosphorylation at threonine-32, which has been published[69], we detected phosphorylation at threonine-16 and threonine-231. We also detected formyl modifications in a clustering around residues 42-63 on the N-terminus, and at lysine-249 and 303 in the C-terminus. Very little is known about formyl modification in cellular proteins, though it seems to be common in histones and chromatin modifying proteins[134]. Interestingly, three of the residues where formyl modifications were detected, were also found to be subject to ubiquitination (K42, K52, K63). This could simply be an artifact resulting from the potentially reactive nature of these surface exposed lysines, or there could be a biological relationship between these two modifications. To date, we have not conducted functional studies to investigate the implications of these modifications.

It is clear that the AID/APOBEC enzymes are subject to post-translational modifications, many of which have not been identified or functionally characterized. In the future, I expect to see progression in this area highlighting the importance of these modifications in physiologically relevant systems.

AID in germ cell reprogramming

AID was originally identified for its contributions to antibody diversity in developing B lymphocytes. Being the oldest member of the AID/APOBEC family, it is conceivable that ancestral AID may have once provided the functions of all the APOBEC

proteins present in higher vertebrates today. More recent studies have demonstrated AID possesses activities that function outside of the immunoglobulin locus. The discovery that AID is capable of inhibiting the replication of LINE-1 provides support for this theory[27]. In addition, building evidence suggests that AID plays a role in epigenetics in germ cells[135]. While this is interesting, the fact remains that AID deficient mice are viable and do not develop abnormally. The only described defect is in their ability to diversify the immunoglobulin locus. The possibility exists that AID may not be required for normal development, however its unregulated expression may result in germ cell defects. AID expression in germ cells lacking a specific regulatory protein(s) could result in aberrant AID activity. In support of this, the two novel AID interacting proteins described in this thesis, CIB1 (**Chaper 1**) and DND1 (**Appendix I**), both have phenotypes in the male gonads. Male mice lacking CIB1 are deficient in spermatogenesis and DND1-null mice have germ cell loss and a propensity to develop testicular tumors[82,136]. It will be interesting to see the results of future studies in the lab utilizing the Aid-/-Dnd-/- mouse line.

The complexity of AID trafficking

AID is the only AID/APOBEC family member known to have genomic DNA as its physiologic target (the immunoglobulin locus). As described in **Appendix II**, AID is normally kept out of the nucleus through a combination of active nuclear export and cytoplasmic retention activities. However, the mechanisms of nuclear entry have been controversial. AID is theoretically small enough (24kD) to be able to passively diffuse through the nuclear pore. However, studies utilizing GFP tagged AID have indicated that

some form of cytoplasmic retention activity prevents this from occurring and that active nuclear import is required[53]. To complicate this, the position of a GFP tag seems to affect the protein's localization properties as well. Placing the tag on the N-terminus appears to block active import activity, while placing the tag on the C-terminus contributes cytoplasmic retention[53,76].

Based on published work and my own observations, I think two conclusions can be made. First, AID is kept out of the nucleus by the CRM1/exportin nuclear export machinery (this is well supported by the literature). Second, AID is actively imported into the nucleus. In my opinion, the best evidence for this is that AID accumulates in the nucleus within 5 minutes after inhibition of CRM1 export when expressed in HeLa cells. Several groups claim to have identified the sole necessary import factor responsible for this activity (e.g., importin α, GANP and CTNNBL1[53,57,58]). However, clearly they cannot all be correct. It seems more likely that several cellular factors are involved, maybe even redundant pathways, in the passage of AID into the nucleus. I believe that several observations in the literature are likely influenced by differences in experimental design, especially in regards to the cell type used and the method of detecting AID localization (GFP tag).

In the future, I believe this will all be sorted out. This will require a thorough investigation of endogenous AID trafficking in primary B lymphocytes and germ cells using immunocytochemistry or immunohistochemistry. However, at this time, commercially available antibodies to detect AID are not available.

Concluding remarks

The AID/APOBEC family of cytidine deaminases comprises an important facet of mammalian immunity, both innate and adaptive. However, this does not come without the potential cost of compromised genomic integrity. Multiple levels of regulation are employed to protect the cell's most valuable information against mistargeted deamination reactions. This thesis contributes to our knowledge of these mechanisms through the identification of a conserved phospho-regulatory mechanism that is capable of dampening the mutagenic activity of these enzymes. Further, this work characterizes two novel AID interacting proteins that are potentially involved in the prevention of aberrant mutagenesis in developing germ cells.

Figures

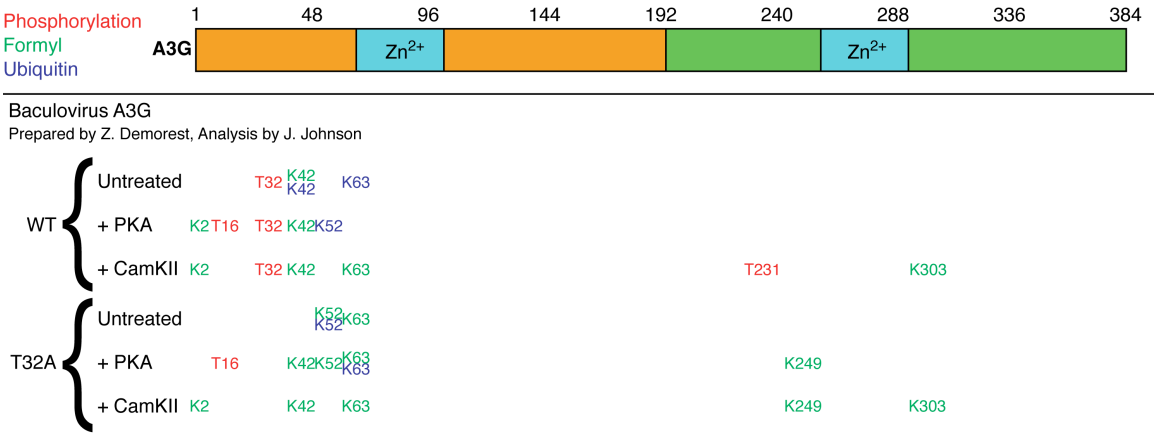


Fig. 4-1. Post-translational modifications of A3G. A3G purified from Sf9 insect cells was incubated *in vitro* with protein kinase A (PKA), or calcium calmodulin kinase II (CamKII) before analysis by mass-spectrometry (Jeff Johnson and Nevin Krogan, UCSF). A mutant of A3G with a T32A substitution was also subject to analysis.

References

1. Frederico, L. A., Kunkel, T. A., and Shaw, B. R. (1990).A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* **29**, 2532-2537
2. LaRue, R. S., Jonsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hotzel, I., Andresdottir, V., Smith, T. P., and Harris, R. S. (2008).The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol* **9**, 104
3. Conticello, S. G. (2008).The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229
4. Agrawal, A., and Schatz, D. G. (1997).RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* **89**, 43-53
5. 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
6. Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., and Honjo, T. (1999).Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* **274**, 18470-18476
7. 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
8. Di Noia, J. M., and Neuberger, M. S. (2007).Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22
9. Longerich, S., Basu, U., Alt, F., and Storb, U. (2006).AID in somatic hypermutation and class switch recombination. *Curr Opin Immunol* **18**, 164-174
10. Davies, D. R., and Metzger, H. (1983).Structural basis of antibody function. *Annu Rev Immunol* **1**, 87-117
11. Helm, B. A., Sayers, I., Higginbottom, A., Machado, D. C., Ling, Y., Ahmad, K., Padlan, E. A., and Wilson, A. P. (1996).Identification of the high affinity receptor binding region in human immunoglobulin E. *The Journal of biological chemistry* **271**, 7494-7500
12. Jefferis, R., Lund, J., and Goodall, M. (1995).Recognition sites on human IgG for Fc gamma receptors: the role of glycosylation. *Immunol Lett* **44**, 111-117
13. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A., and Durandy, A. (2000).Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565-575

14. Meffre, E., Catalan, N., Seltz, F., Fischer, A., Nussenzweig, M. C., and Durandy, A. (2001).Somatic hypermutation shapes the antibody repertoire of memory B cells in humans. *The Journal of experimental medicine* **194**, 375-378

15. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J.,

Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. J. (2001).Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921

16. Cordaux, R., and Batzer, M. A. (2009).The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703
17. Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002).Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650
18. Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S., and Malim, M. H. (2003).DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809
19. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003).Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103
20. Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A. J. (2003).Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300**, 1112
21. Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., and Gao, L. (2003).The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-98
22. Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., and Peterlin, B. M. (2004).Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *Journal of virology* **78**, 6073-6076
23. Wiegand, H. L., Doehle, B. P., Bogerd, H. P., and Cullen, B. R. (2004).A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *The EMBO journal* **23**, 2451-2458
24. Liddament, M. T., Brown, W. L., Schumacher, A. J., and Harris, R. S. (2004).APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Current biology : CB* **14**, 1385-1391
25. Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G. G., and Munk, C. (2006).APOBEC3 proteins inhibit human LINE-1 retrotransposition. *The Journal of biological chemistry* **281**, 22161-22172
26. Lovsin, N., and Peterlin, B. M. (2009).APOBEC3 proteins inhibit LINE-1 retrotransposition in the absence of ORF1p binding. *Ann N Y Acad Sci* **1178**, 268-275
27. MacDuff, D. A., Demorest, Z. L., and Harris, R. S. (2009).AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res* **37**, 1854-1867
28. Stenglein, M. D., and Harris, R. S. (2006).APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* **281**, 16837-16841
29. Schumacher, A. J., Haché, G., Macduff, D. A., Brown, W. L., and Harris, R. S. (2008).The DNA deaminase activity of human APOBEC3G is required for Ty1,

MusD, and human immunodeficiency virus type 1 restriction. *J Virol* **82**, 2652-2660
30. Schumacher, A. J., Nissley, D. V., and Harris, R. S. (2005).APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9854-9859
31. Esnault, C., Millet, J., Schwartz, O., and Heidmann, T. (2006).Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res* **34**, 1522-1531
32. Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., and Niedobitek, G. (2006).Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. *J Pathol* **210**, 26-31
33. Teng, B., Burant, C. F., and Davidson, N. O. (1993).Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816-1819
34. Nakamuta, M., Oka, K., Krushkal, J., Kobayashi, K., Yamamoto, M., Li, W. H., and Chan, L. (1995).Alternative mRNA splicing and differential promoter utilization determine tissue-specific expression of the apolipoprotein B mRNA-editing protein (Apobec1) gene in mice. Structure and evolution of Apobec1 and related nucleoside/nucleotide deaminases. *The Journal of biological chemistry* **270**, 13042-13056
35. Liao, W., Hong, S. H., Chan, B. H., Rudolph, F. B., Clark, S. C., and Chan, L. (1999).APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochemical and biophysical research communications* **260**, 398-404
36. Vonica, A., Rosa, A., Arduini, B. L., and Brivanlou, A. H. (2011).APOBEC2, a selective inhibitor of TGFbeta signaling, regulates left-right axis specification during early embryogenesis. *Dev Biol* **350**, 13-23
37. Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S., and Rada, C. (2010).Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *The Journal of biological chemistry* **285**, 7111-7118
38. Refsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L., and Harris, R. S. (2010).Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* **38**, 4274-4284
39. Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., and Harris, R. S. (2010).APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol* **17**, 222-229
40. Bartel, D. P. (2004).MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297
41. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008).Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**, 102-114

42. Meister, G., and Tuschl, T. (2004).Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-349
43. Dorsett, Y., McBride, K. M., Jankovic, M., Gazumyan, A., Thai, T. H., Robbiani, D. F., Di Virgilio, M., San-Martin, B. R., Heidkamp, G., Schwickert, T. A., Eisenreich, T., Rajewsky, K., and Nussenzweig, M. C. (2008).MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* **28**, 630-638
44. 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
45. Conticello, S. G., Harris, R. S., and Neuberger, M. S. (2003).The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Current biology : CB* **13**, 2009-2013
46. Marin, M., Rose, K. M., Kozak, S. L., and Kabat, D. (2003).HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature medicine* **9**, 1398-1403
47. Mehle, A., Strack, B., Ancuta, P., Zhang, C., McPike, M., and Gabuzda, D. (2004).Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *The Journal of biological chemistry* **279**, 7792-7798
48. Sheehy, A. M., Gaddis, N. C., and Malim, M. H. (2003).The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature medicine* **9**, 1404-1407
49. Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X. F. (2003).Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060
50. Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J. C., and Reynaud, C. A. (2008).Proteasomal degradation restricts the nuclear lifespan of AID. *J Exp Med* **205**, 1357-1368
51. 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
52. 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
53. 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
54. Lellek, H., Kirsten, R., Diehl, I., Apostel, F., Buck, F., and Greeve, J. (2000).Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. *J Biol Chem* **275**, 19848-19856

55. Mehta, A., Kinter, M. T., Sherman, N. E., and Driscoll, D. M. (2000).Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Molecular and cellular biology* **20**, 1846-1854
56. Conticello, S. G., Ganesh, K., Xue, K., Lu, M., Rada, C., and Neuberger, M. S. (2008).Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol Cell* **31**, 474-484
57. Ganesh, K., Adam, S., Taylor, B., Simpson, P., Rada, C., and Neuberger, M. (2011).CTNNBL1 Is a Novel Nuclear Localization Sequence-binding Protein That Recognizes RNA-splicing Factors CDC5L and Prp31. *The Journal of biological chemistry* **286**, 17091-17102
58. Maeda, K., Singh, S. K., Eda, K., Kitabatake, M., Pham, P., Goodman, M. F., and Sakaguchi, N. (2010).GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J Biol Chem* **285**, 23945-23953
59. Orthwein, A., Patenaude, A. M., Affar el, 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
60. MacDuff, D. A., Neuberger, M. S., and Harris, R. S. (2006).MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. *Mol Immunol* **43**, 1099-1108
61. Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J. P., Manis, J. P., and Alt, F. W. (2005).The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508-511
62. McBride, K. M., Gazumyan, A., Woo, E. M., Barreto, V. M., Robbiani, D. F., Chait, B. T., and Nussenzweig, M. C. (2006).Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc Natl Acad Sci U S A* **103**, 8798-8803
63. Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006).PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc Natl Acad Sci U S A* **103**, 395-400
64. Chaudhuri, J., Khuong, C., and Alt, F. W. (2004).Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992-998
65. 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. *Molecular and cellular biology* **31**, 442-449
66. 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. *The Journal of experimental medicine* **205**, 2585-2594
67. Pham, P., Smolka, M. B., Calabrese, P., Landolph, A., Zhang, K., Zhou, H., and Goodman, M. F. (2008).Impact of phosphorylation and phosphorylation-null

mutants on the activity and deamination specificity of activation-induced cytidine deaminase. *J Biol Chem* **283**, 17428-17439
68. Cheng, H. L., Vuong, B. Q., Basu, U., Franklin, A., Schwer, B., Astarita, J., Phan, R. T., Datta, A., Manis, J., Alt, F. W., and Chaudhuri, J. (2009).Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proc Natl Acad Sci U S A* **106**, 2717-2722
69. Shirakawa, K., Takaori-Kondo, A., Yokoyama, M., Izumi, T., Matsui, M., Io, K., Sato, T., Sato, H., and Uchiyama, T. (2008).Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat Struct Mol Biol* **15**, 1184-1191
70. MacDuff, D. A., Offer, S. M., Demorest, Z. L., and Harris, R. S. (2007) *Antibody Gene Diversification by Aid-Catalyzed DNA Editing*, John Wiley & Sons, Inc.
71. Macduff, D. A., and Harris, R. S. (2006).Directed DNA deamination by AID/APOBEC3 in immunity. *Current biology : CB* **16**, R186-189
72. Neuberger, M. S. (2008).Antibody diversification by somatic mutation: from Burnet onwards. *Immunology and cell biology* **86**, 124-132
73. Rada, C., Di Noia, J. M., and Neuberger, M. S. (2004).Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* **16**, 163-171
74. Di Noia, J., and Neuberger, M. S. (2002).Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43-48
75. Di Noia, J. M., Williams, G. T., Chan, D. T., Buerstedde, J. M., Baldwin, G. S., and Neuberger, M. S. (2007).Dependence of antibody gene diversification on uracil excision. *J Exp Med* **204**, 3209-3219
76. Rada, C., Jarvis, J. M., and Milstein, C. (2002).AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc Natl Acad Sci U S A* **99**, 7003-7008
77. 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. *Mol Cell* **12**, 501-508
78. Geisberger, R., Rada, C., and Neuberger, M. S. (2009).The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. *Proc Natl Acad Sci U S A* **106**, 6736-6741
79. Basu, U., Wang, Y., and Alt, F. W. (2008).Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase. *Mol Cell* **32**, 285-291
80. Chatterji, M., Unniraman, S., McBride, K. M., and Schatz, D. G. (2007).Role of activation-induced deaminase protein kinase A phosphorylation sites in Ig gene conversion and somatic hypermutation. *J Immunol* **179**, 5274-5280
81. Gentry, H. R., Singer, A. U., Betts, L., Yang, C., Ferrara, J. D., Sondek, J., and Parise, L. V. (2005).Structural and biochemical characterization of CIB1

delineates a new family of EF-hand-containing proteins. *J Biol Chem* **280**, 8407-8415

82. Yuan, W., Leisner, T. M., McFadden, A. W., Clark, S., Hiller, S., Maeda, N., O'Brien, D. A., and Parise, L. V. (2006).CIB1 is essential for mouse spermatogenesis. *Molecular and cellular biology* **26**, 8507-8514
83. Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Cote, M., Rich, R. L., Myszka, D. G., and Sundquist, W. I. (2001).Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**, 55-65
84. 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
85. 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
86. Chiu, Y. L., Soros, V. B., Kreisberg, J. F., Stopak, K., Yonemoto, W., and Greene, W. C. (2005).Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. *Nature* **435**, 108-114
87. Chiu, Y. L., Witkowska, H. E., Hall, S. C., Santiago, M., Soros, V. B., Esnault, C., Heidmann, T., and Greene, W. C. (2006).High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 15588-15593
88. Gallois-Montbrun, S., Holmes, R. K., Swanson, C. M., Fernandez-Ocana, M., Byers, H. L., Ward, M. A., and Malim, M. H. (2008).Comparison of cellular ribonucleoprotein complexes associated with the APOBEC3F and APOBEC3G antiviral proteins. *J Virol* **82**, 5636-5642
89. Gallois-Montbrun, S., Kramer, B., Swanson, C. M., Byers, H., Lynham, S., Ward, M., and Malim, M. H. (2007).Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. *J Virol* **81**, 2165-2178
90. Kozak, S. L., Marin, M., Rose, K. M., Bystrom, C., and Kabat, D. (2006).The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. *J Biol Chem* **281**, 29105-29119
91. Arakawa, H., Hauschild, J., and Buerstedde, J. M. (2002).Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **295**, 1301-1306
92. Harris, R. S., Sale, J. E., Petersen-Mahrt, S. K., and Neuberger, M. S. (2002).AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Current Biology* **12**, 435-438
93. Leisner, T. M., Liu, M., Jaffer, Z. M., Chernoff, J., and Parise, L. V. (2005).Essential role of CIB1 in regulating PAK1 activation and cell migration. *The Journal of cell biology* **170**, 465-476

94. Naik, M. U., Pham, N. T., Beebe, K., Dai, W., and Naik, U. P. (2010).Calcium-dependent inhibition of Polo-like kinase 3 activity by CIB1 in breast cancer cells. *International Journal of Cancer* **In press**
95. Yoon, K. W., Cho, J. H., Lee, J. K., Kang, Y. H., Chae, J. S., Kim, Y. M., Kim, J., Kim, E. K., Kim, S. E., Baik, J. H., Naik, U. P., Cho, S. G., and Choi, E. J. (2009).CIB1 functions as a Ca(2+)-sensitive modulator of stress-induced signaling by targeting ASK1. *Proc Natl Acad Sci U S A* **106**, 17389-17394
96. Denofrio, J. C., Yuan, W., Temple, B. R., Gentry, H. R., and Parise, L. V. (2008).Characterization of calcium- and integrin-binding protein 1 (CIB1) knockout platelets: potential compensation by CIB family members. *Thrombosis and haemostasis* **100**, 847-856
97. Rosenberg, B. R., and Papavasiliou, F. N. (2007).Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. *Advances in immunology* **94**, 215-244
98. Gourzi, P., Leonova, T., and Papavasiliou, F. N. (2007).Viral induction of AID is independent of the interferon and the Toll-like receptor signaling pathways but requires NF-kappaB. *J Exp Med* **204**, 259-265
99. Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., and Blau, H. M. (2010).Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042-1047
100. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*
101. Huang, T. T., Feinberg, S. L., Suryanarayanan, S., and Miyamoto, S. (2002).The zinc finger domain of NEMO is selectively required for NF-kappa B activation by UV radiation and topoisomerase inhibitors. *Molecular and cellular biology* **22**, 5813-5825
102. Bai, Y., Soda, Y., Izawa, K., Tanabe, T., Kang, X., Tojo, A., Hoshino, H., Miyoshi, H., Asano, S., and Tani, K. (2003).Effective transduction and stable transgene expression in human blood cells by a third-generation lentiviral vector. *Gene therapy* **10**, 1446-1457
103. Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M., and Trono, D. (1996).In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263-267
104. Xu, K., Ma, H., McCown, T. J., Verma, I. M., and Kafri, T. (2001).Generation of a stable cell line producing high-titer self-inactivating lentiviral vectors. *Mol Ther* **3**, 97-104
105. Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W., and Liu, X. (2003).Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. *Mol Cell Proteomics* **2**, 1225-1233
106. Naik, U. P., Patel, P. M., and Parise, L. V. (1997).Identification of a novel calcium-binding protein that interacts with the integrin alphaIIb cytoplasmic domain. *J Biol Chem* **272**, 4651-4654

107. Malim, M. H., and Emerman, M. (2008).HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388-398
108. Shandilya, S. M., Nalam, M. N., Nalivaika, E. A., Gross, P. J., Valesano, J. C., Shindo, K., Li, M., Munson, M., Royer, W. E., Harjes, E., Kono, T., Matsuo, H., Harris, R. S., Somasundaran, M., and Schiffer, C. A. (2010).Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **18**, 28-38
109. Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S., and Matsuo, H. (2008).Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116-119
110. Harjes, E., Gross, P. J., Chen, K. M., Lu, Y., Shindo, K., Nowarski, R., Gross, J. D., Kotler, M., Harris, R. S., and Matsuo, H. (2009).An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J Mol Biol* **389**, 819-832
111. Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., and Chen, X. S. (2008).Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121-124
112. 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
113. Koning, F. A., Newman, E. N., Kim, E. Y., Kunstman, K. J., Wolinsky, S. M., and Malim, M. H. (2009).Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**, 9474-9485
114. Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F. W., and Geha, R. S. (2004).Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol* **16**, 395-404
115. Farrow, M. A., Kim, E. Y., Wolinsky, S. M., and Sheehy, A. M. (2011).NFAT and IRF proteins regulate transcription of the anti-HIV protein, APOBEC3G. *J Biol Chem*
116. Stenglein, M. D., Matsuo, H., and Harris, R. S. (2008).Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *J Virol* **82**, 9591-9599
117. Bennett, R. P., Presnyak, V., Wedekind, J. E., and Smith, H. C. (2008).Nuclear Exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *J Biol Chem* **283**, 7320-7327
118. Sheehy, A. M., Gaddis, N. C., and Malim, M. H. (2003).The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **9**, 1404-1407
119. Stopak, K., de Noronha, C., Yonemoto, W., and Greene, W. C. (2003).HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* **12**, 591-601

120. Conticello, S. G., Harris, R. S., and Neuberger, M. S. (2003).The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* **13**, 2009-2013
121. Demorest, Z. L., MacDuff, D. A., Brown, W. L., Morham, S. G., Parise, L. V., and Harris, R. S. (2010).The interaction between AID and CIB1 is nonessential for antibody gene diversification by gene conversion or class switch recombination. *PLoS One* **5**, e11660
122. Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S. G., and Pandey, A. (2007).A curated compendium of phosphorylation motifs. *Nat Biotechnol* **25**, 285-286
123. OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S., and Emerman, M. (2008).Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **4**, 249-259
124. 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. *Mol Cell* **10**, 1247-1253
125. Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J. G. (2006).Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J Virol* **80**, 5992-6002
126. Albin, J. S., Haché, G., Hultquist, J. F., Brown, W. L., and Harris, R. S. (2010).Long-term restriction by APOBEC3F selects human immunodeficiency virus type 1 variants with restored Vif function. *J Virol* **84**, 10209-10219
127. Gervaix, A., West, D., Leoni, L. M., Richman, D. D., Wong-Staal, F., and Corbeil, J. (1997).A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proc Natl Acad Sci U S A* **94**, 4653-4658
128. Haché, G., Shindo, K., Albin, J. S., and Harris, R. S. (2008).Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Curr Biol* **18**, 819-824
129. Browne, E. P., Allers, C., and Landau, N. R. (2009).Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. *Virology* **387**, 313-321
130. Masterson, L. R., Shi, L., Metcalfe, E., Gao, J., Taylor, S. S., and Veglia, G. (2011).Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. *Proc Natl Acad Sci U S A*
131. Shlyakhtenko, L. S., Lushnikov, A. Y., Li, M., Lackey, L., Harris, R. S., and Lyubchenko, Y. L. (2011).Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J Biol Chem* **286**, 3387-3395
132. LaRue, R. S., Lengyel, J., Jonsson, S. R., Andresdottir, V., and Harris, R. S. (2010).Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. *J Virol* **84**, 8193-8201
133. Thielen, B. K., Klein, K. C., Walker, L. W., Rieck, M., Buckner, J. H., Tomblingson, G. W., and Lingappa, J. R. (2007).T cells contain an RNase-insensitive inhibitor of APOBEC3G deaminase activity. *PLoS Pathog* **3**, 1320-1334

134. Wisniewski, J. R., Zougman, A., and Mann, M. (2008).Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic acids research* **36**, 570-577
135. Morgan, H. D., Dean, W., Coker, H. A., Reik, W., and Petersen-Mahrt, S. K. (2004).Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of biological chemistry* **279**, 52353-52360
136. Youngren, K. K., Coveney, D., Peng, X., Bhattacharya, C., Schmidt, L. S., Nickerson, M. L., Lamb, B. T., Deng, J. M., Behringer, R. R., Capel, B., Rubin, E. M., Nadeau, J. H., and Matin, A. (2005).The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature* **435**, 360-364
137. Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., and Cairns, B. R. (2008).DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201-1212
138. Marr, S., Morales, H., Bottaro, A., Cooper, M., Flajnik, M., and Robert, J. (2007).Localization and differential expression of activation-induced cytidine deaminase in the amphibian Xenopus upon antigen stimulation and during early development. *Journal of immunology* **179**, 6783-6789
139. Bascove, M., and Frippiat, J. P. (2010).Molecular characterization of Pleurodeles waltl activation-induced cytidine deaminase. *Molecular immunology* **47**, 1640-1649
140. Chen, S. H., Habib, G., Yang, C. Y., Gu, Z. W., Lee, B. R., Weng, S. A., Silberman, S. R., Cai, S. J., Deslypere, J. P., Rosseneu, M., and et al. (1987).Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* **238**, 363-366
141. Navaratnam, N., Shah, R., Patel, D., Fay, V., and Scott, J. (1993).Apolipoprotein B mRNA editing is associated with UV crosslinking of proteins to the editing site. *Proc Natl Acad Sci U S A* **90**, 222-226
142. Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., and Scott, J. (1987).A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831-840
143. Blanc, V., Kennedy, S., and Davidson, N. O. (2003).A novel nuclear localization signal in the auxiliary domain of apobec-1 complementation factor regulates nucleocytoplasmic import and shuttling. *J Biol Chem* **278**, 41198-41204
144. Kedde, M., Strasser, M. J., Boldajipour, B., Oude Vrielink, J. A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P. M., van Duijse, J., Orom, U. A., Lund, A. H., Perrakis, A., Raz, E., and Agami, R. (2007).RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273-1286
145. Slanchev, K., Stebler, J., Goudarzi, M., Cojocaru, V., Weidinger, G., and Raz, E. (2009).Control of Dead end localization and activity--implications for the function of the protein in antagonizing miRNA function. *Mech Dev* **126**, 270-277
146. Weidinger, G., Stebler, J., Slanchev, K., Dumstrei, K., Wise, C., Lovell-Badge, R., Thisse, C., Thisse, B., and Raz, E. (2003).dead end, a novel vertebrate germ

plasm component, is required for zebrafish primordial germ cell migration and survival. *Current biology : CB* **13**, 1429-1434
147. Bhattacharya, C., Aggarwal, S., Kumar, M., Ali, A., and Matin, A. (2008).Mouse apolipoprotein B editing complex 3 (APOBEC3) is expressed in germ cells and interacts with dead-end (DND1). *PloS one* **3**, e2315
148. Watanabe, Y., and Maekawa, M. (2010).Methylation of DNA in cancer. *Adv Clin Chem* **52**, 145-167
149. Fritz, E. L., and Papavasiliou, F. N. (2010).Cytidine deaminases: AIDing DNA demethylation? *Genes Dev* **24**, 2107-2114
150. Mattaj, I. W., and Englmeier, L. (1998).Nucleocytoplasmic transport: the soluble phase. *Annual review of biochemistry* **67**, 265-306
151. Pemberton, L. F., Blobel, G., and Rosenblum, J. S. (1998).Transport routes through the nuclear pore complex. *Curr Opin Cell Biol* **10**, 392-399

Appendix I: AID interacts with DND1

D.M. Henderson conducted the experiments presented in Fig. I-1B. W.L. Brown generated the data for Fig. I-1C. Z.L. Demorest performed the experiments for Figs. I-2, I-3, and I-4.

Appendix I Summary

Activation induced deaminase (AID) functions to generate antibody diversity in developing B lymphocytes through the processes of somatic hypermutation and class-switch recombination. Recent evidence also suggests it has a role in germ cell epigenetics. APOBEC1, a related deaminase, requires an RNA-binding cofactor called APOBEC1 complementation factor (ACF) in order to deaminate its physiological target, *APOB* mRNA. We identified dead-end (DND1) as a putative AID interacting protein based on its sequence homology to ACF. We hypothesized that DND1 regulates AID deaminase activity in antibody diversification reactions, or in epigenetic reprogramming in germ cells. Through *in vitro* GST-pulldown and co-immunoprecipitation studies, we confirmed that AID and DND1 are bonafide interactors. Additionally, AID and DND1 were found to co-localize in distinct foci when co-expressed in living HeLa cells. To examine the functional significance of this interaction, we acquired DND1 knockout mice and first investigated their ability to generate a diverse antibody repertoire (through AID dependent mechanisms). We found that DND1 is not required for antibody diversification by class-switch recombination. Interestingly, *Dnd*1-/- mice are known for a significant germ cell defect. Male mice have testicular germ cell loss, and therefore are sterile, and are also predisposed to formation of testicular germ cell tumors. To ask whether the DND1 and AID interaction is relevant in germ cells, we are in the midst of generating *Aid*-/-*Dnd*1-/- mice to test the hypothesis that AID activity may be required for the generation of testicular tumors in *Dnd*-/- mice.

Introduction

Antibody diversity in developing B lymphocytes is facilitated by the processes of class switch recombination and somatic hypermutation. Remarkably, both of these processes are initiated by a single enzyme, activation induced deaminase (AID) [reviewed in[8]]. To achieve this, AID converts DNA cytosines to uracil at immunoglobulin switch regions (class switch recombination) and variable regions (somatic hypermutation). AID has also been implicated in functions outside of the B cell compartment beginning with the discovery that AID can deaminate ssDNA methyl-cytosine to thymine[135]. The resulting G:T base pairs are thought to be repaired by the cell's mismatch repair machinery to effectively facilitate germ cell reprogramming by replacing the methylated cytosine with a new demethylated cytosine. In support of this hypothesis, AID has been shown to be expressed in human and mouse testes[27,32] as well as very early in development in several lower vertebrates including zebrafish[137], frogs[138] and newts[139].

AID is the founding member of a large family of cytidine deaminases found in vertebrates[3]. The namesake of the APOBEC family originates from APOBEC1, which deaminates cytidine 6666 of the Apolipoprotein B (ApoB) mRNA[140-142]. Multiple forms of regulation are involved in the transcription, translation and post-translational control of AID and related APOBEC members[6,38,43,44,51-53,113-117]. APOBEC1 forms a complex with an accessory factor called APOBEC1 Complementation Factor (ACF) that is required for *APOB* mRNA editing[54,55]. ACF is an RNA-binding protein that is believed to provide the substrate specificity for APOBEC1 and, in addition, contains a nuclear localization sequence that facilitates the complex's entry into the nucleus[143].

Through comparative studies in collaboration with Angabin Matin at MD Anderson Cancer Center, we identified an RNA-binding protein called dead-end (DND1) which has a high level of sequence homology to ACF (**Fig. I-1A**). DND1 is expressed primarily in primordial germ cells, but has also been found in the heart[136]. *Dnd*1-/- mice are characterized by primordial germ cell loss on all backgrounds and a predisposition to develop testicular germ cell tumors on the 129 inbred strain[136] (**Fig. I-1B**). Recent evidence suggests it functions to counteract miRNA inhibition in germ cells by competitively binding to the target sequence in the 3'-UTR, thus displacing the miRNA and relieving inhibition[144].

Due to the common evolutionary history of AID and APOBEC1 and the homology between DND1 and ACF, we hypothesize that DND1 regulates AID targeting and/or activity in developing B lymphocytes or in germ cells. In the absence of DND1, we predict there would be a deficiency in AID dependent antibody diversification in B lymphocytes. In germ cells, unregulated AID activity may directly contribute to tumorigenesis by causing genomic mutations, or indirectly by altering the epigenetic code during development. As an alternative hypothesis, the AID and DND1 interaction may be significant in DND1's ability to regulate miRNA activity in developing germ cells[144].

In this study, we test these hypotheses by examining whether Dnd1-/- mice are defective for AID dependent antibody diversification. Further, we describe the generation of *Aid*-/-*Dnd*1-/- knockout mice to ask whether AID is required for the generation of testicular germ cell tumors seen in *Dnd*1-/- mice.

Results

AID interacts with DND1

To determine whether AID and DND1 are capable of interaction, we performed a series of GST-pulldown and co-immunoprecipitation experiments. GST-AID expressed in *E. coli* was bound to glutathione beads and incubated with lysates from HEK-293T cells expressing DND1-GFP. We found that GST-AID successfully pulled down DND1-GFP but not GFP alone (**Fig. I-2A**). Since DND1 and AID have the ability to bind RNA, we treated the reactions with RNaseA to rule out the possibility that this interaction was facilitated through a nucleic acid bridge. We found that the interaction between AID and DND1 persisted in the presence of RNaseA (**Fig. I-2A**).

To validate our findings in another interaction system, we performed a co-immunoprecipitation assay in human cell lines. AID-HA and DND1-myc were co-expressed in HEK-293T cells. The cells were lysed and an antibody against myc was used to precipitate DND1-myc. Using a detection antibody against HA, we found AID to be interacting with DND1 in this system as well (**Fig. I-2B**). These combined data provide strong support for an *in vivo* interaction between DND1 and AID that is not mediated by an RNA bridge.

AID and DND1 co-localize within living cells

After confirming that DND1 and AID do indeed interact, we were interested to see whether we would be able to see this interaction *in vivo*. DND1 has been shown to localize to perinuclear granules in germ cells, whereas it distributes cell-wide in somatic cell lines such as COS7, HEK293T or NIH3T3[145-147]. We transfected AID-GFP and

DND1-mCherry either alone or in combination into HeLa cells, which do not express AID endogenously, and examined their localization by fluorescence microscopy. As expected in this somatic cell line, AID localized predominantly to the cytoplasm, while DND1 was predominantly nuclear with some found in the cytoplasm (**Fig. I-3A**). Interestingly, when AID and DND1 were co-expressed, DND1 was no longer found in the nucleus, but instead had relocalized to distinct foci in the perinuclear area where AID was also found (**Fig I-3B**). These results are intriguing because first, they further support that DND1 and AID interact in living cells, and second, they implicate AID in having a role in DND1's perinuclear localization seen in germ cells.

DND1 is not required for class-switch recombination

To evaluate the functional significance of the interaction between AID and DND1, we assayed AID dependent antibody diversification in DND1 knockout mice. AID is required to initiate class-switch recombination, a process that creates diversity at the constant region of the immunoglobulin. This generates antibodies with different constant regions, thus changing their effector function. We collected serum from *Dnd1*+/+ and -/- mice and analyzed it for Ig isotypes by ELISA. We found no significant differences in the serum Ig isotype levels between *Dnd1*+/+ and *Dnd1*-/- mice (**Fig. I-4A)**. The levels for IgG3 and IgA appeared slightly lower in the knockout mice, however this slight difference was not significant.

Another technique commonly used to monitor AID dependent antibody diversification is the *in vitro* class-switch recombination assay. Naïve B cells from *Dnd1*+/+ and -/- mice were isolated from spleen by magnetic sorting. The cells were

cultured in the presence of IL-4 and LPS to induce switching to IgG1. After 4 days, the cells were assayed by flow-cytometry using an antibody specific for IgG1. We found a reproducible modest decrease in the percent of IgG1 positive cells in the DND1-null cells as compared to wildtype (**Fig. I-4B**). While these results may be interesting, B cells from *Dnd1*-/- mice retain ample ability to switch to IgG1 and are not significantly different from wildtype mice (**Fig. I-4C**). Further, we sectioned spleen isolated from *Dnd*1-/- and *Dnd1*+/+ mice, stained with hematoxylin and eosin and compared them for morphological differences. We determined that spleen from *Dnd*1-/- mice were normal with typical germinal centers visible(**Fig. I-4D**).

DND1 may have a role in regulating AID activity in germ cell reprogramming

To test the hypothesis that DND1 is regulating AID activity in germ cells *in vivo*, we are generating *Aid*-/-*Dnd1*-/- mice through the crossing of C57Bl/6-*Aid*-/- and 129-*Dnd1*-/- mouse strains we have in the laboratory. On the 129 background, *Dnd1*-/- mice have germ cell loss and testicular tumor formation, whereas on other backgrounds only germ cell loss is observed. The AID mice are on a C57Bl/6 background, so repeated backcrossing is necessary and still ongoing to generate a congenic mouse line needed for these studies (**Fig. I-5**). To date, we have performed three of the ten backcrosses required to generate a fully congenic mouse strain.

Discussion

In this study, we investigated the role of a novel AID interacting partner, DND1. DND1 was identified as a putative AID regulatory protein due to its high degree of homology to ACF, an accessory factor required for APOBEC1 to deaminate *APOB* mRNA. DND1 has previously been characterized as having a role in germ cell development, where its absence in mice results in germ cell loss and a predisposition to testicular tumors. We hypothesized that DND1 regulates AID activity in developing B lymphocytes, or alternatively, in germ cells. In B lymphocytes, we predicted the absence of DND1 would manifest as a deficiency in AID dependent antibody diversification. In germ cells lacking DND1, AID activity could promote tumorigenesis directly by causing genomic mutation (**Fig. I-6A**). Alternatively, AID may facilitate cancer through epigenetics by relieving suppression of methylated chromatin regions[reviewed in[148,149]] (**Fig. I-6A**). Indirectly, the AID/DND1 interaction may be significant in DND1's ability to inhibit miRNAs in germ cells (**Fig. I-6B**).

We showed that AID and DND1 are capable of interaction using two independent *in vitro* assays, GST-pulldown and co-immunoprecipitation experiments. Importantly, of concern because both AID and DND1 are RNA binding proteins, we also showed that this interaction is RNase resistant, suggesting that this is not simply occurring through an RNA bridge. Using fluorescence microscopy, we examined the cellular localization of DND1 in living HeLa cells in the presence and absence of AID. When DND1 was expressed alone in somatic cells, it localized predominantly in the nucleus with some apparent in the cytoplasm. Remarkably, the addition of AID to the cell resulted in a drastic change in DND1 localization, where AID and DND1 were found to co-localize in

distinct foci in the perinuclear area. This result is especially intriguing because prior reports have shown that DND1 localizes in perinuclear granules only in germ cells, not in somatic cells. Additional histological experiments examining endogenous AID and DND1 localization in germ cells are warranted to determine if AID co-localizes with DND1 in the perinuclear granules (once specific antibodies become available). Notably, our colleagues have reported similar findings for an interaction between mouse APOBEC3 and DND1, where co-localization is seen in perinuclear granules in COS7 cells[147]. However, we have not been able to reproduce the GST-pulldown, co-immunoprecipitation, or cell localization phenotypes reported (Brown, Demorest, not shown).

Functionally, we examined whether the absence of DND1 had any effect on AID's ability to facilitate antibody diversification by class-switch recombination. We collected blood from *Dnd1*+/+ and -/- mice and compared their serum Ig isotype levels by ELISA, but found no significant difference. Additionally, naïve B cells were harvested and subjected to an *ex vivo* class-switch recombination assay. In these experiments, we found no significant defect in AID's ability to promote switching to IgG1 in the absence of DND1. These combined data indicate that the AID/DND1 interaction does not have an obvious role in class-switch recombination.

Going forward, we are testing the hypothesis that AID and DND1 are important partners involved in germ cell development and reprogramming. The generation of a congenic mouse model (129-*Aid*-/-*Dnd1*-/-) requires ten consecutive backcrosses between the C57Bl/6-*Aid*-/- and 129-*Dnd1*-/- founders. To date, we have performed 3 of these

backcrosses and anticipate it will take an additional 1-2 years to complete this cross. Once the experimental animals are available, they will be assessed for the loss of testicular germ cells as well as the prevalence of testicular tumors by histological examination such as hematoxylin and eosin staining as in **Fig. I-1B**.

Experimental procedures

Plasmids

DND1-myc (pRH909)

DND1-GFP (pRH1053)

GST-AID (pRH2502)

AID-HA (pRH1275)

AID-GFP (pRH984)

DND1-mCherry (pRH1054)

GST-pulldowns and co-immunoprecipitation

GST-AID was transformed into BL21 *E. coli* cells and grown in LB media supplemented with ampicillin (100 μg/mL). Cells were lysed in GST lysis buffer (5 mM HEPES pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$ 1 mM ZnCl, 10% glycerol). After clarification by centrifugation, Glutathione Sepharose (GE Healthcare) beads were added to immobilize AID-GST. After extensive washing with lysis buffer, clarified cell lysates from untransfected HEK-293T cells or cells expressing DND1-GFP were added. The GST-AID complexes were eluted from the beads using glutathione (10mM), separated on a 10% SDS-PAGE gel and transferred to a PVDF membrane. Detection of GST-AID or DND1-GFP was done using an antibody to GST (GE Healthcare 27-4577-01) or GFP (Clontech JL8) respectively.

For co-immunoprecipitation experiments, AID-HA and DND1-myc were transfected into HEK-293T cells using TransIT-LT1 (Mirus). 48-hours after transfection, cells were lysed in NP-40 lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP40) and clarified by centrifugation. An antibody against myc (ThermoFisher 9E11) was added, followed by the addition of Protein G Sepharose 4 Fast Flow (GE Healthcare). Bound complexes were washed with lysis buffer x4 and eluted with 0.1M glycine. The complexes were separted as above and detected with antibodies against HA (12CA5 Roche) and myc (Sigma C3956).

Microscopy

AID-GFP and DND1-mCherry were transfected into HeLa cells using TransIT-LT1 (Mirus). 48-hours later, AID and DND1 localization was determined by fluorescence microscopy (Deltavision).

Mouse experiments

All experiments were conducted with the approval of the University of Minnesota Institutional Animal Care and Use Committee (IACUC). The 129-*Ter* DND1 knockout mice were a kind gift from Angabin Matin[136]. Genotyping was carried out by PCR amplification of tail snip DNA using primers 5'-gta gtt cag gaa ctc cac ttg tg (RSH1952) and 5'- gcc taa tga tga cct tca gtg g (RSH1953). The resulting amplicons were digested with DdeI and separated on a 7% acrylamide gel. The presence of 180, 154, and 97 bp bands indicates wildtype. Heterozygous animals are indicated by 180, 154, 123, and 97 bp bands. Homozygous *Dnd1*-/- animals are identified by 180, 123, and 97 bp bands. Analysis of serum Ig isotype levels by ELISA and the *in vitro* class-switch recombination assay have been described[121]. For histology, mice were euthanized by CO_2 inhalation and their testes dissected. The tissues were then fixed in 10% buffered formalin and imbedded in paraffin before cutting 10μm sections by microtome (Leica Instruments). The sections were then stained with hematoxylin and eosin and examined by light microscopy (Zeiss Axiovert).

Figures

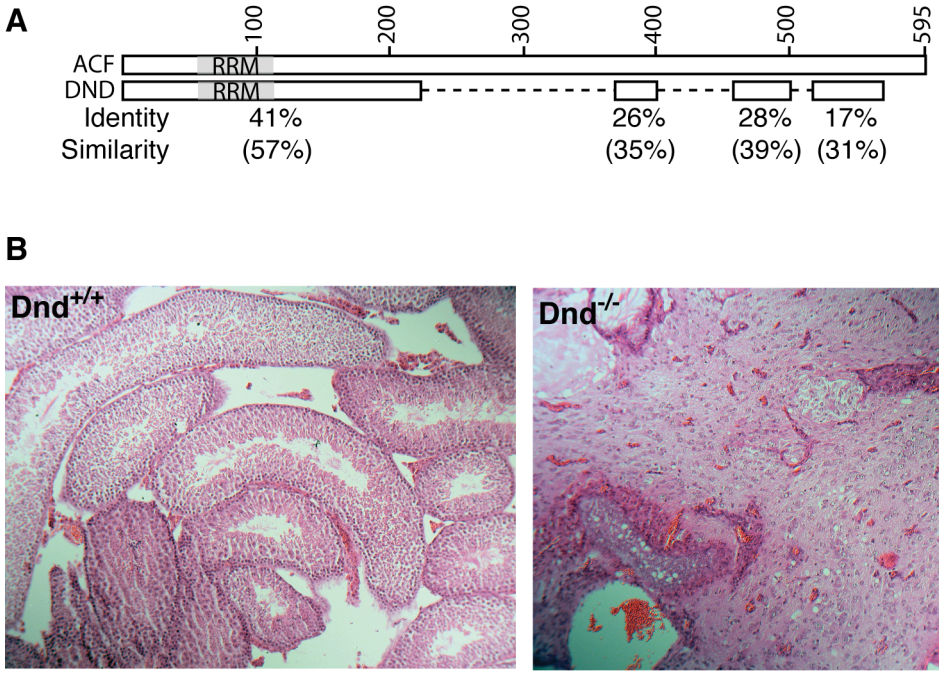


Fig. I-1. Identification of DND1. *A*. Alignment of ACF and DND1 proteins indicating percent conserved identity in each region. The conserved RNA Recognition Motif (RRM) is also indicated. Percent sequence identity is listed below each region and the percent similarity is shown in brackets. *B*. Hematoxylin and eosin stained sections of testes harvested from *Dnd1*+/+ and -/- mice. The highly organized seminiferous tubules are clearly seen in *Dnd1*+/+ tissue, while the *Dnd1*-/- tissue is disorganized and lacks defined seminiferous lumen.

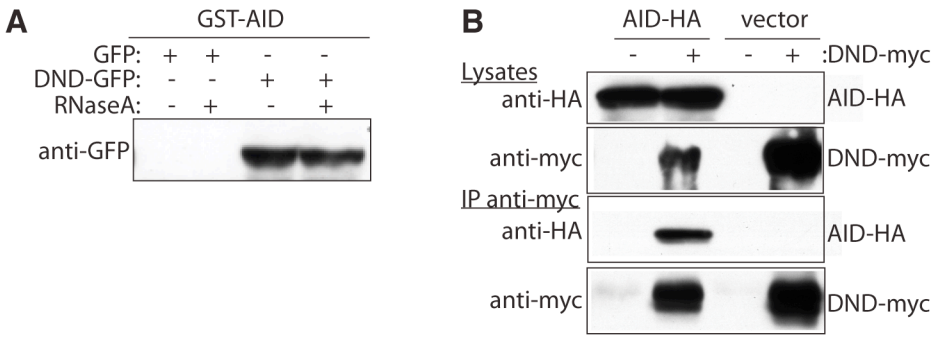


Fig. I-2. AID interacts with DND1. *A*. GST-tagged AID pulls down DND1-GFP from human cell extracts. *B*. DND1-myc co-immunoprecipitates AID-HA when co-expressed in human cells.

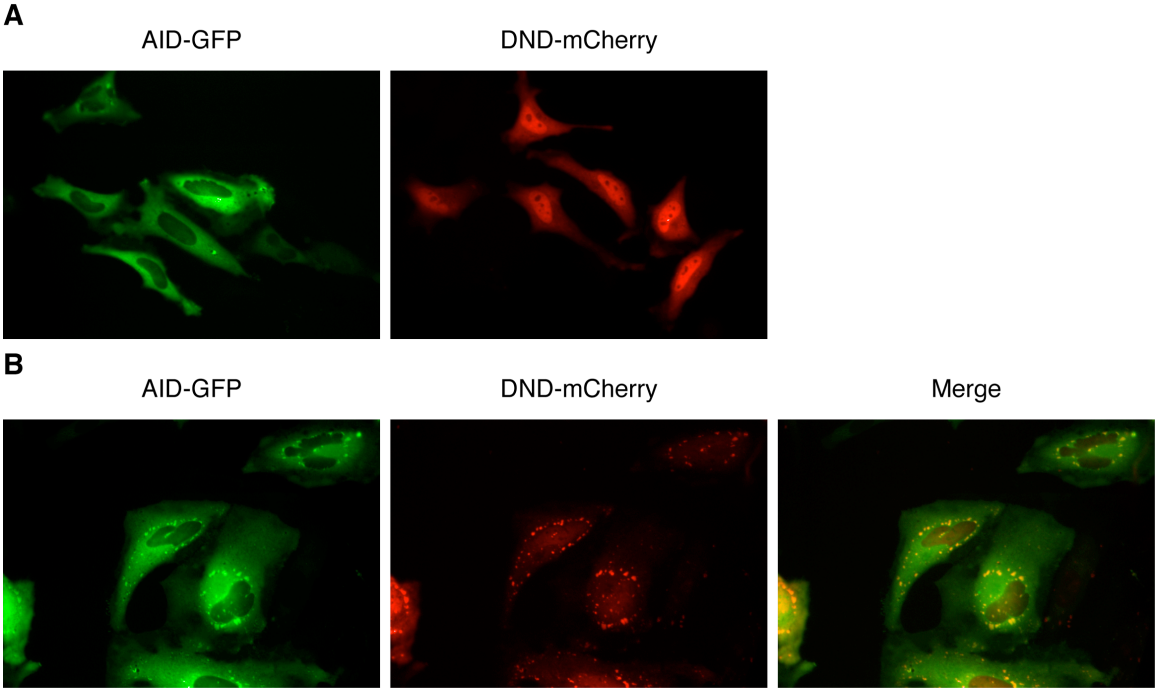


Fig. I-3. AID and DND1 co-localize within living cells. *A*. AID is predominantly cytoplasmic under steady state conditions, whereas DND1 is primarily nuclear. *B*. When co-expressed, DND1 re-localizes to distinct foci coincident with AID in the cytoplasm of living HeLa cells.

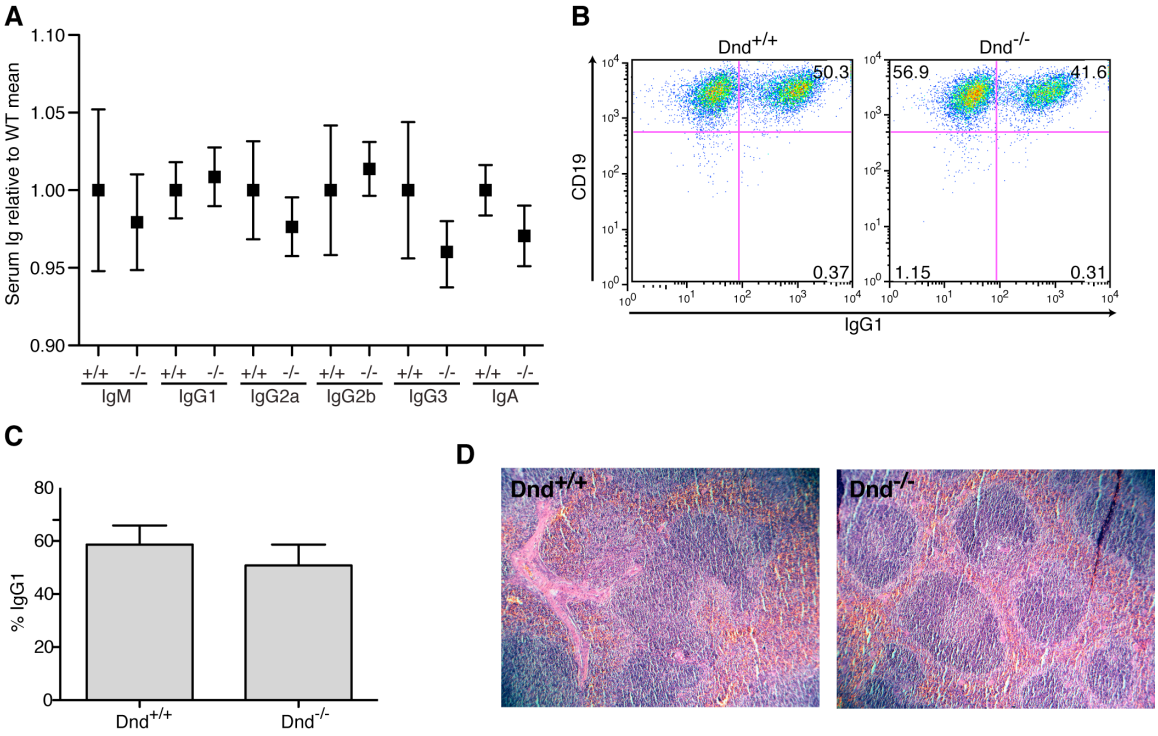


Fig. I-4. DND1 is not required for class switch recombination. *A*. Results from analysis of serum Ig isotype levels from *Dnd*-/- and +/+ mice as measured by ELISA. Data are represented as the normalized mean of 4 mice analyzed +/- S.D. *B*. Representative flow-cytometry plots from *in vitro* CSR assay indicating percent of CD19+ B cells that have undergone switching to IgG1. *C*. Bar graph indicating the mean switching activity +/- S.D. from 3 age matched littermates. *D*. Images of hematoxylin and eosin stained sections of spleen isolated from *Dnd1*+/+ and *Dnd1*-/- mice.

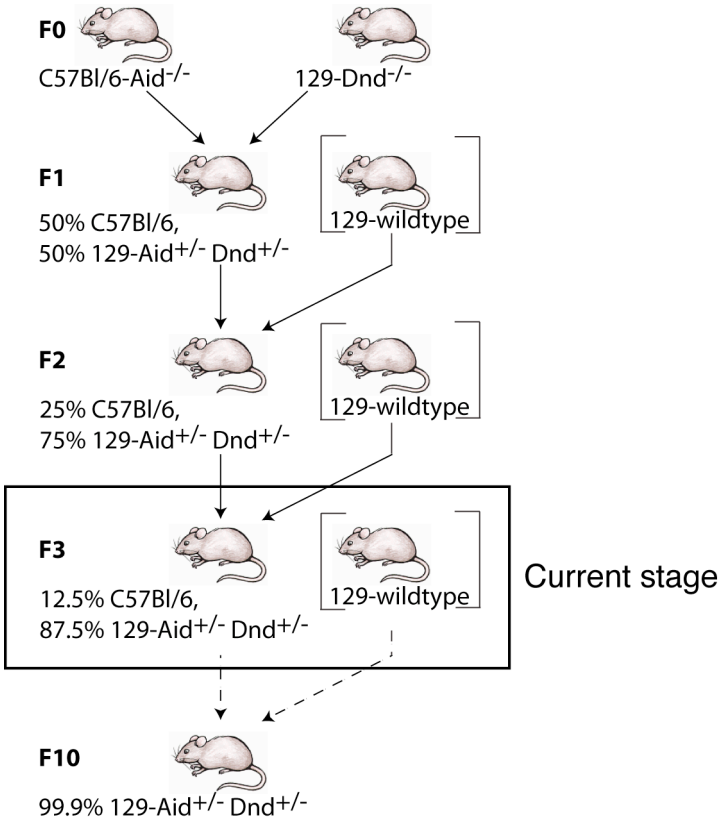


Fig. I-5. AID and DND1 may be involved in testicular tumorigenesis. *A*. Schematic illustrating the breeding scheme required to generate congenic 129-*Aid*-/-*Dnd1*-/- mice. Current breeding pairs are from the F3 generation.

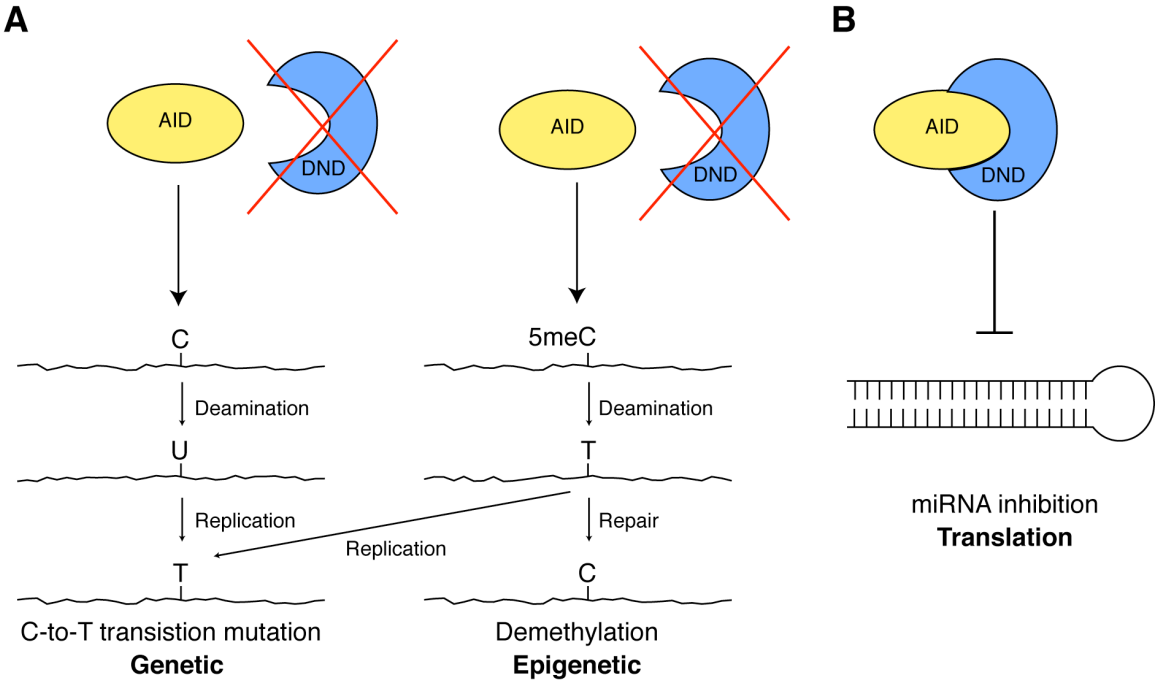


Fig. I-6. Model for functional implications of AID and DND1 interactions. *A*. In the absence of DND1, AID can promote tumorigenesis genetically by converting cytidine to uracil. Upon replication, these mutations become fixed as thymine, resulting in a C-to-T transition. AID can also promote tumorigenesis epigenetically by converting 5meC to T. Replication will result in a C-to-T transition, or these T:G mismatches can be fixed by the mismatch repair machinery to result in a replacement with a demethylated cytidine. *B*. The AID/DND1 interaction may be important for regulating miRNA inhibition.

Appendix II: Nucleocytoplasmic trafficking of AID

Z.L. Demorest performed all experiments described in this appendix

Appendix II summary

The processes of class switch recombination and somatic hypermutation are critically important in the generation of a diverse antibody repertoire in all vertebrates. Remarkably, one enzyme, activation induced deaminase (AID), initiates both of these processes through deamination of cytosine to uracil at the immunoglobulin locus. AID is tightly regulated to prevent aberrant mutations outside of its intended target. One mechanism believed to help protect genomic integrity is nuclear export of AID by CRM1, thus keeping it away from DNA in the nucleus. However, upon B cell activation, AID must enter the nucleus and target the immunoglobulin locus to facilitate antibody diversification. Here we show that conflicting reports in the literature regarding differential nuclear import and cytoplasmic retention activities in AID are due to interspecies differences in AID isolated from human and mouse. We identify a single amino acid difference at position 48 in the N-terminus of mouse AID that renders it import defective. Further, a single amino acid at position 195 in the cytoplasmic tail of mouse AID is responsible for additional cytoplasmic retention activity.

Introduction

Activation induced cytidine deaminase (AID) is a small 24kD protein that is required to initiate class-switch recombination and somatic hypermutation, two processes which diversify antibodies in vertebrates [reviewed in [8]]. AID is thought to initiate class-switch recombination and somatic hypermutation by converting cytidine to uracil at the immunoglobulin switch regions or the variable region, repectively. AID is a member of the APOBEC family of cytidine deaminases, which are well known for their involvement in innate immunity to retroelements, retrotransposons and foreign DNA[39,107]. Of the 11 APOBEC members in humans, AID is unique in its targeting of genomic DNA. The precise mechanism by which AID is able to specifically identify the immunoglobulin locus remains elusive, although multiple levels of regulation of AID have been described.

AID is normally found in the cytoplasmic compartment of cells, presumably to keep it away from genomic DNA and prevent aberrant mutations. This cytoplasmic localization is achieved through two mechanisms. First, AID is actively exported from the nucleus via the CRM1/exportin1 nuclear export machinery[51,52]. Second, recent reports suggest that AID may also be retained in the cytoplasm, similarly to APOBEC3G[53,116,117].

In order to reach its physiological target, AID must enter the nucleus. AID is theoretically small enough to passively diffuse into the nucleus[150,151] Alternatively, AID may be actively imported into the nucleus by the importin family of nuclear chaperone proteins. Indeed, a putative bipartite nuclear localization sequence (NLS) has been identified in the N-terminal region of AID[51]. In this study, AID expressed with a C-terminal GFP tag shuttled from cytoplasmic compartment to the nuclear compartment upon treatment with leptomycin B (LMB), a CRM1-exportin1 inhibitor. Further, they

show that the putative NLS from residues 8-25 is indeed functional using deletion and point mutants, which disrupt the import activity. In addition, fusion of the N-terminal 29 residues from AID to GFP resulted in a more nuclear distribution than GFP alone (found throughout the cell).

Unfortunately, there are conflicting reports in the literature regarding import activity. Two reports have concluded that the canonical NLS from 8-25 is non-functional [52,53]. In one case, the authors conclude that AID is actively imported, however the NLS is conformational and not dependent on the residues from 8-25[53]. In the other study, the authors observed that AID fused to GFP on the N- or C-terminus did not facilitate nuclear localization of the fusion protein upon disruption of CRM1 export by treatment with LMB or by mutation of the NES sequence in AID (F198A)[52]. Instead, the authors report that AID displays a cell-wide distribution, consistent with the ability to passively diffuse through the nuclear pore.

We hypothesized that the apparent discrepancies in these conflicting studies might be due to species differences. Both studies reporting a functional NLS (canonical or conformational) used human AID in their experiments[51,53], while the other concluding that AID enters the nucleus through passive diffusion was performed with mouse AID[52]. We hypothesize that human and mouse AID have differential activities for import, export or cytoplasmic retention. To test this, we performed a thorough localization comparison between human and mouse AID using chimeric proteins and amino acid substitution mutants.

Results

Mouse AID is compromised for import activity

To evaluate the import potential of human and mouse AID, hAID and mAID, respectively, expression constructs were C-terminally tagged with GFP (with identical linker regions), transfected into HeLa cells, and examined by live cell fluorescence microscopy. Consistent with prior reports, both human and mouse AID were found predominantly in the cytoplasmic compartment (**Fig. II-1A**). We then added LMB, incubated the cells for 2 hours, and again determined the cellular distribution of human and mouse AID. Interestingly, only human AID appeared to accumulate in the nucleus, whereas mouse AID remained predominantly cytoplasmic (**Fig. II-1B**). These data suggest that mouse AID is defective for import activity, or alternatively, is retained in the cytoplasm via a mechanism independent of CRM1 export.

Differences in the putative NLS do not account for mouse AID import defect

Human and mouse AID have several differences in amino acid sequence, including within the putative NLS region described[51] (**Fig II-2A**). To determine if these differences rendered the NLS in mouse AID non-functional, we generated reciprocal mutants of human and mouse AID exchanging human AID asparagine-7 and arginine-8 for lysine-7 and glutamine-8 (hAID-N7K/R8Q) and vice versa (mAID-K7N/Q8R). We chose not to mutate residue 9 because it was similar and of equal charge. We tagged these mutants with GFP, transfected them into HeLa cells, and determined their steady-state localization. Similar to the wildtype proteins, both the human and mouse AID mutants were found to be predominantly cytoplasmic (**Fig. II-2B**). After treatment with LMB, consistent with wildtype proteins, hAID-N7K/R8Q was found to have shuttled into the

nucleus while mAID-K7N/Q8R resided in the cytoplasm (**Fig. II-2C**). These results indicate that putative NLS residues are not responsible for nuclear import or for the differential import phenotype.

The C-terminus of mouse AID contains cytoplasmic retention activity

Altogether, there are only 15 differences in amino acid sequence between human and mouse AID. To gain further insight into the amino acid difference(s) responsible for the apparent disparities in nuclear accumulation, we took advantage of a common SexAI restriction endonuclease cleavage site in the cDNA (corresponding to phenylalanine-81) to generate chimeras between the human and mouse proteins. This resulted in chimeras designated hAID(1-81)mAID(82-198) and mAID(1-81)hAID(82-198) (**Fig. II-3A**). These mutants were tagged with GFP and transfected into HeLa cells as above and imaged for their cellular distribution. We found that both chimeras were predominantly cytoplasmic under steady-steady state conditions, similar to the wildtype proteins (**Fig. II-3B**). However, upon addition of LMB, two surprising phenotypes were observed. The hAID(1-81)mAID(82-198) chimera remained in the cytoplasmic compartment, consistent with wildtype mouse AID and indicating there may be cytoplasmic retention activity in the region from 82-198. The other chimera, mAID(1-81)hAID(82-198), also displayed an interesting localization. This chimera was found to localize throughout the cell, with no preference for the cytoplasmic or nuclear compartments. These data are consistent with the N-terminal portion of mouse AID containing no import activity and thus achieving this localization by passive diffusion.

Substitution of a single amino acid in the N-terminus of human AID disrupts import activity

The region from amino acids 1-81 used in the chimera above contains 8 differences between human and mouse AID (**Fig. II-3A**). We previously determined that 2 of these differences (N7K/R8Q) do not account for the apparent discrepancy in import activity (see **Fig. II-2**). To determine which of the remaining differing residues were responsible for the lack of import activity observed in mouse AID, we generated reciprocal single point mutants in human and mouse AID. These mutant proteins were expressed with a C-terminal GFP tag in HeLa cells for live cell imaging. Of the 5 mutants tested (hAID-R9K was never tested), only hAID-Y48H displayed a cellular localization phenotype different from the wildtype protein (**Fig. II-4A**). This mutant localized to the cytoplasmic compartment under steady state as expected, however, in the presence of LMB, it failed to accumulate in the nucleus in the majority of cells imaged. These data suggest that mAID is deficient for import activity due to a specific variance at amino acid residue 48. However, the reciprocal substitution in mouse AID (mAID-H48Y) does not result in an actively imported protein in the presence of LMB, consistent with cytoplasmic retention activity as described in **Fig. II-3B** or a conformational NLS that requires coordination between the N- and C-terminus.

Substitution of a single amino acid in the C-terminus disrupts cytoplasmic retention in mouse AID

AID contains a leucine rich nuclear export sequence at its C-terminus which is recognized by CRM1 to facilitate exit from the nucleus[51,52]. In order to characterize the residues responsible for the cytoplasmic retention activity seen in the C-terminus of

mouse AID, we generated mutants of human and mouse AID to disrupt CRM1 mediated export (hAID-L198A and mAID-F198A). These mutants were expressed in HeLa cells as GFP fusion proteins and their cellular localization was determined by fluorescence microscopy. We found that hAID-L198A was localized predominantly in the nuclear compartment, while mAID-F198A was found throughout the cell, consistent with prior reports[51,52] (**Fig. II-5A**). To determine which residue(s) were responsible for the contrasting localization phenotypes, we substituted single amino acid residues of mouse AID with the corresponding residue found in human AID. These mutants were generated in the context of mAID-F198A to exclude nuclear export by CRM1. We identified a single amino acid substitution in the C-terminus (M195T) that, when paired with F198A, facilitated nuclear accumulation of mouse AID (**Fig. II-5A**). One explanation for the apparent differences seen in nuclear accumulation between hAID and mAID could be that the cellular import factors are species specific. To address this, we expressed hAID-L198A, mAID-F198A and mAID-M195T/F198A in murine NIH-3T3 fibroblasts and examined their localization. We found no apparent difference between the localization phenotypes of these mutants regardless of the host cell species (**Fig. II-5B**). These data are consistent with mouse AID containing additional cytoplasmic retention activity that is coordinated by M195 and not present in human AID.

Discussion

AID contributes to adaptive immunity by initiating the processes of class switch recombination and somatic hypermutation in developing B lymphocytes. Because its intended target is endogenous genomic DNA, multiple levels of regulation are utilized to prevent aberrant mutation. Nuclear export via CRM1 has been shown to keep AID in the cytoplasmic compartment, a safe distance away from genomic DNA. However, AID must get into the nucleus to facilitate class switch recombination and somatic hypermutation at the immunoglobulin locus. Some controversy surrounds this mechanism, with reports indicating active import due to a canonical bipartite[51] or tertiary conformational NLS[53], and another proposing passive diffusion[52].

We noticed that these studies were conducted with AID isolated from different species (active import reported for human AID and passive diffusion for mouse). We hypothesized that interspecies differences were the source of the discrepancies reported. To address this, we performed a thorough analysis of import, export and retention activities in human and mouse AID through the use of chimeras and reciprocal point mutations. First, we determined that differing residues between human and mouse AID in the putative NLS are not responsible for the lack of mouse AID accumulation. We then discovered that replacement of the first 81 amino acids in human AID with mouse-derived sequence rendered this protein unable to accumulate in the nucleus upon inhibition of CRM1, suggesting that the N-terminus of mouse AID has a defect in active import. Further, through single amino acid substitution, we identify the causative amino acid difference at position 48 (tyrosine in human and histidine in mouse). This is

especially interesting in light of the report indicating the region in human AID responsible for binding to importin-α spans residues 46-52[53].

It is curious however, that making the reciprocal mutation in mouse AID (mAID-H48Y) does not result in a protein that accumulates in the nucleus upon LMB treatment. This implies that mouse AID is not only defective for import, but also contains activity that keeps the protein sequestered in the cytoplasmic compartment. In support of this, a chimera we tested, hAID(1-81)mAID(82-198), had a surprising phenotype. This protein contains human sequence in the N-terminal region and therefore should possess import activity. However, upon treatment with LMB, the protein remained predominantly cytoplasmic, consistent with cytoplasmic retention preventing it from entering the nucleus. To characterize this activity, we made single amino acid substitutions in mAID-F198A, which is not exported by CRM1, with the corresponding residue from human sequence. We showed that exchanging methionine-195 with a threonine (as found in human AID) now facilitated nuclear accumulation. This was unexpected, as mouse AID was found to have a defect in import activity in the first 81 amino acids. One possible explanation is that the residues required for import activity are not in a linear arrangement in the protein, but instead are part of a tertiary functional group comprised of residues found in both the N- and C-terminal regions of the protein. The methionine at position 195 in mouse AID may contribute cytoplasmic retention activity while simultaneously affecting active import. In support of this, the Di Noia lab reported that AID NLS activity is conformational, and also describes cytoplasmic retention activity in the C-terminal tail from residues 158-198[53].

These combined data support a model in which import, export, and cytoplasmic retention activity are all involved in contributing to the overall cellular localization of AID. The differences we identified between human and mouse AID trafficking are intriguing and raise the question of why AID from human and mouse would seemingly differ in their nucleocytoplasmic transport processes. One possibility may relate to the fact that humans have 6 more APOBEC3 members than mice. AID in mice may have retained functions necessary to contribute to innate immunity to retroelements, as the human APOBEC3s are known for [reviewed in[71]]. Further studies examining the cellular compartmentalization of AID in human and mouse primary B lymphocytes will be critical to resolve this issue.

Experimental procedures

Plasmids

pEGFP-N3-hAID (pRH984)

pEGFP-N3-mAID (pRH986)

pEGFP-N3-hAID(N7K-R8Q)

pEGFP-N3-mAID(K7N-Q8R)

pEGFP-N3-hAID(1-81)mAID(82-198) (pRH1561)

pEGFP-N3-mAID(1-81)hAID(82-198) (pRH1562)

pEGFP-N3-hAID(Q14H) (pRH1777)

pEGFP-N3-hAID(R25H) (pRH1778)

pEGFP-N3-hAID(F42C) (pRH1779)

pEGFP-N3-hAID(Y48H) (pRH1780)

pEGFP-N3-hAID(N53S) (pRH1781)

pEGFP-N3-mAID(H48Y) (pRH1785)

pEGFP-N3-hAID(L198A) (pRH1932)

pEGFP-N3-mAID(F198A) (pRH1933)

pEGFP-N3-mAID(M195T-F198A) (RH2326)

Microscopy

AID expression constructs in pEGFP-N3 were transiently transfected into HeLa cells grown in DMEM (Gibco) supplemented with 10% FBS (Hyclone) using TransIT-LT1 (Mirus). 48 hours later, the media was removed and replaced with DMEM without phenol red. The cellular distribution of AID-GFP fusion proteins was imaged by fluorescence microscopy using a 20X or 40X objective (Deltavision). For inhibition of CRM1 export, leptomycin B (20 ng/mL) was added to the media 2 hours prior to imaging.

Figures

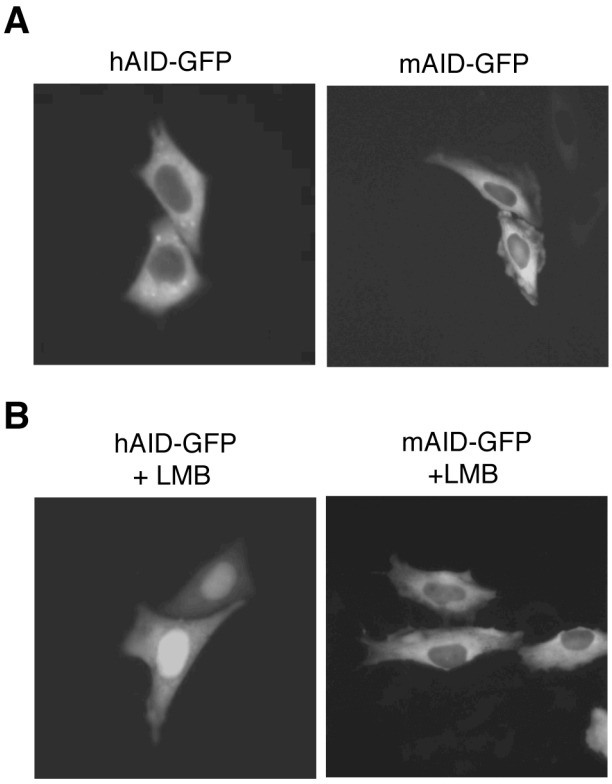


Fig. II-1. Mouse AID-GFP is compromised for import activity. *A*. Steady state localization of GFP tagged human and mouse AID in HeLa cells. *B*. Localization of human and mouse AID 2 hours after treatment with leptomycin B (LMB).

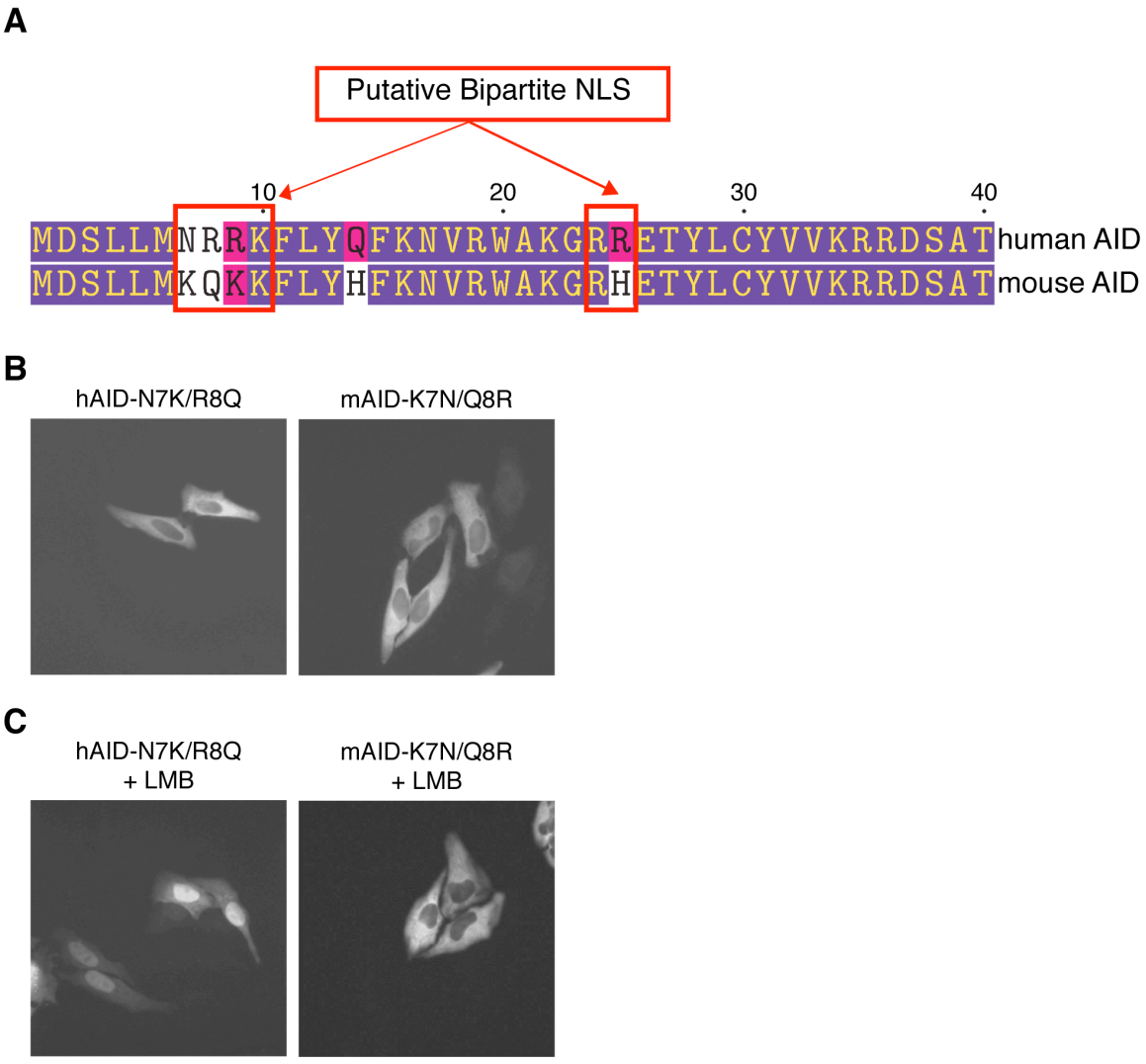


Fig. II-2. Differences in the putative NLS do not account for import defect. *A*. Alignment of human and mouse AID with the putative bipartite NLS highlighted[51]. *B*. Steady-state localization of GFP tagged mutant human and mouse AID in HeLa cells. *C*. Localization of human and mouse AID mutants 2 hours after leptomycin B (LMB) treatment.

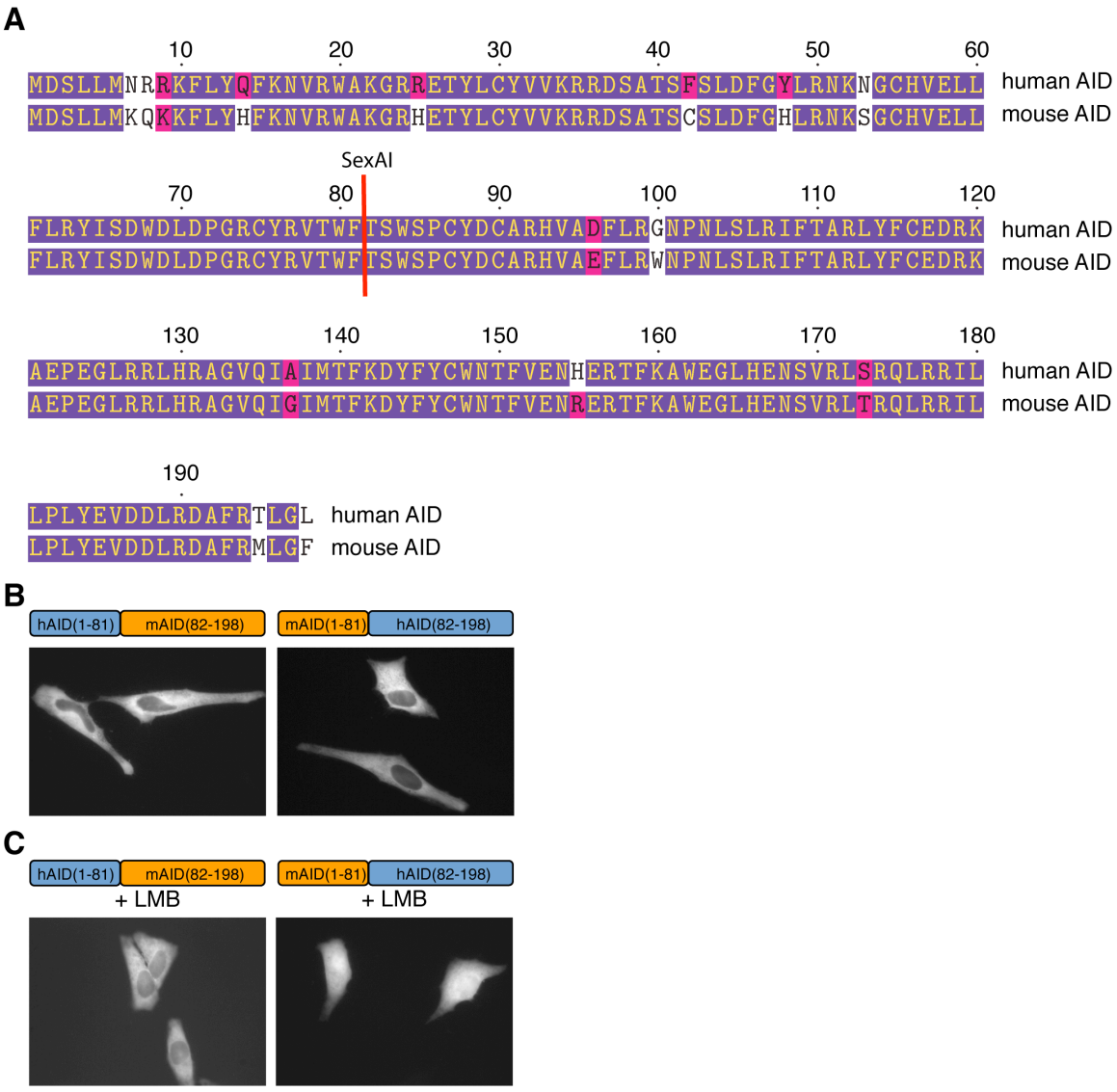


Fig. II-3. The C-terminus of mouse AID contains cytoplasmic retention activity. *A*. Schematic representation of chimeric human and mouse AID constructs tested. *B*. Steady-state localization of GFP tagged chimeric human and mouse AID in HeLa cells. *C*. Localization of human and mouse AID chimeras 2 hours after leptomycin B (LMB) treatment.

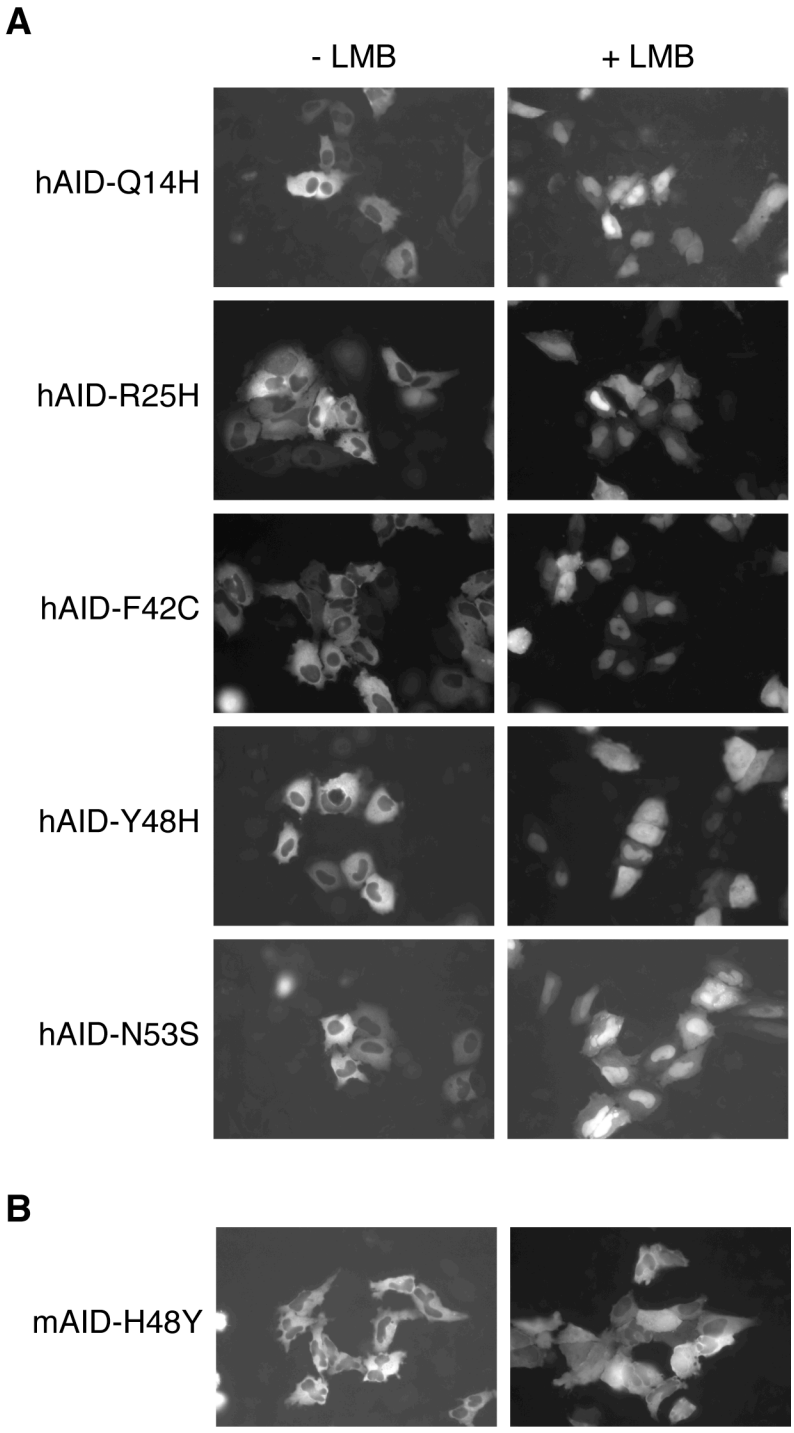


Fig. II-4. A single amino acid substitution disrupts import activity in human AID. *A*. Localization of human AID mutants tested in steady state and in the presence of leptomycin B (LMB). hAID-Y48H displays a cell-wide distribution in most cells. *B*. The reciprocal mutant, mAID-H48Y, does not gain the ability to accumulate in the nucleus.

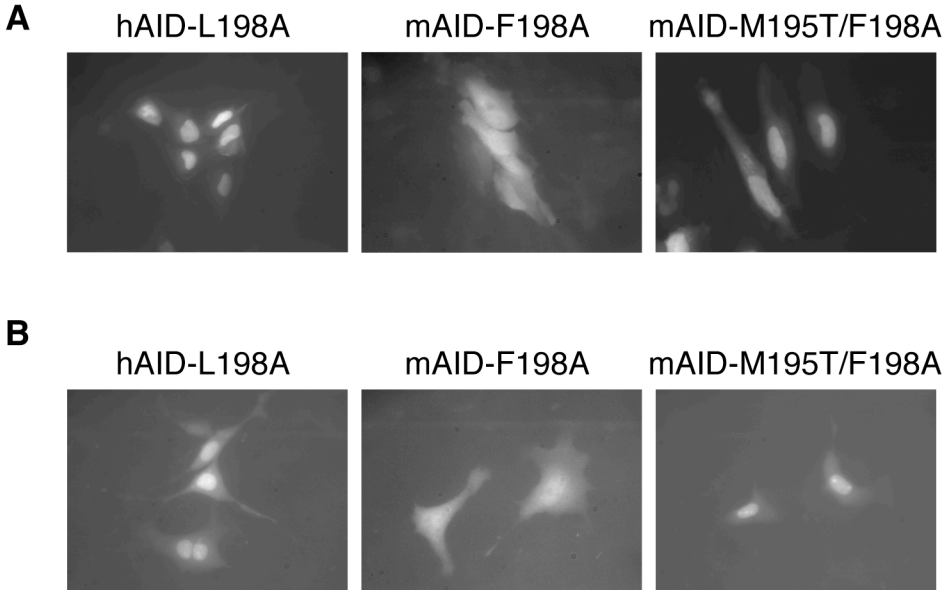


Fig. II-5. A single amino acid substitution disrupts cytoplasmic retention in mouse AID. *A*. Representative images of human AID-L198A, mouse AID-F198A, and mouse AID-M195T/F198A localization in HeLa cells. *B*. Representative images of human AID-L198A, mouse AID-F198A, and mouse AID-M195T/F198A localization in NIH-3T3 cells

Appendix III: Phospho-peptide mapping of APOBEC3G

Z.L. Demorest performed all experiments described in this appendix.

Appendix III Summary

Post-translational modification by phosphorylation is a common mechanism for regulation of protein function. In this study, we used a combination of *in vivo* and *in vitro* labeling techniques combined with phospho-amino acid and phospho-peptide analyses to characterize phosphorylation modifications on APOBEC3G (A3G). We found that A3G with a C-terminal strep tag is phosphorylated at multiple sites *in vivo* comprising at least one serine, one threonine and one tyrosine residue. *In vitro,* we detected at least one phosphorylated serine residue on A3G-strep when incubated in the presence of recombinant protein kinase A (PKA). In parallel studies utilizing mass-spectrometry, we identified a phosphorylated serine residue in the strep tag present on A3G. Further studies were performed using untagged A3G generated using the baculovirus expression system. *In vitro*, we found that A3G can be phosphorylated at a number of sites by calcium calmodulin dependent kinase II (CaMKII) and have evidence to support the identification of threonine-218 as a candidate substrate for phosphorylation.

Introduction

There are several methods for the detection of protein phosphorylation including mass-spectrometry, phospho-specific antibodies, *in vitro* kinase assays, and *in vivo* manipulation of cellular kinases and phosphatases using activators or inhibitors. Another powerful method is the use of the radioisotope ^{32}P. Radioactive phosphorus can be used either *in vitro* in the context of γ-^{32}P-ATP, or *in vivo* using inorganic orthophosphate.

For *in vitro* studies, the protein of interest is incubated with a candidate kinase in the presence of γ-^{32}P-ATP. The transfer of a radioactive phosphate molecule to the target protein can be detected by autoradiography. For *in vivo* studies, cells expressing a tagged version of the target protein are cultured for several hours in the presence of ^{32}P-orthophosphate. After purification of the target protein, phosphorylation by endogenous cellular kinases can be detected by autoradiography.

As an extension of the labeling techniques outlined above, labeled protein samples can be further processed to determine the type of amino acids phosphorylated (serine, threonine or tyrosine) by phospho-amino acid analysis. The labeled protein is hydrolyzed to individual amino acids with hydrochloric acid and then mixed with excess phospho-amino acids standards (phospho-serine, phospho-threonine, and phospho-tyrosine). The resulting mixture is applied to a thin-layer chromatography (TLC) plate and separated by 2-dimensional electrophoresis. The plate is sprayed with ninhydrin to visualize the location of the phospho-amino acid standards and imaged by autoradiography to detect phosphorylated residues from the protein sample. These two images are overlaid to determine the identity of phosphorylated residues.

Another powerful technique used in the analysis of protein phosphorylation is phospho-peptide mapping. In this analysis, phosphorylated protein labeled either *in vivo* or *in vitro* is digested with an enzyme (such as trypsin) to generate peptide fragments. These peptide fragments are applied to a TLC plate and separated by charge using electrophoresis in one direction and by hydrophobicity using chromatography in a second direction. The plates are imaged by autoradiography to identify specific phosphorylated peptides. For this analysis to be useful, mutant derivatives of the target protein with candidate phosphorylation sites mutated to alanine are examined in parallel. The absence of a specific phospho-spot in the mutant protein compared to the wildtype implies that that residue is a target for phosphorylation. In this study, we performed both *in vivo* metabolic labeling of APOBEC3G (A3G) and *in* vitro kinase assays utilizing protein kinase A (PKA) and calcium calmodulin dependent kinase II (CaMKII) and generated 2-dimentional maps for phospho-amino acid and phospho-peptide analysis.

Results

APOBEC3G-strep is phosphorylated in vivo

To evaluate whether A3G is phosphorylated *in vivo*, A3G-strep was purified from transiently transfected HEK293T cells that were metabolically labeled using ^{32}P-orthophosphate. Using phospho-amino acid analysis, we determined that A3G-strep is phosphorylated at a minimum of three sites *in vivo* (one serine, one threonine and one tyrosine) (**Fig. III-1A**).

To determine whether threonine-218 was one of the sites phosphorylated *in vivo,* we performed phospho-peptide mapping using metabolically labeled A3G-strep, A3G(191-384)-strep, A3G(191-384)T218A-strep and GFP-strep. We found a number of phosphorylated peptides present in A3G-strep, indicated by the presence of spots on the autoradiogram (**Fig. III-1B**, top left). Interestingly, most of the phospho-sites present in full-length A3G are also present in the A3G(191-384) construct (Fig. III-1B, top right). This suggests that the C-terminal domain of A3G is the predominant substrate for phosphorylation. When we compared the maps for A3G(191-384)-strep to A3G(191-384)T218A-strep, we could not reproducibly detect the disappearance of a specific phospho-peptide spot (**Fig. III-1B**, top right and bottom left). As a control, we also purified GFP-strep from metabolically labeled cells and performed phospho-peptide mapping. Unfortunately, we discovered that GFP-strep contained one prominent phosphorylation site that matched a major site in the A3G analyses (**Fig. III-1B**, bottom right). This is consistent with a phosphorylation site present in the tag that is shared between GFP-strep and A3G-strep. However, the A3G-strep maps contain numerous

spots that are not present in the GFP-strep map indicating that there are multiple other sites that are phosphorylated on A3G *in vivo*.

PKA phosphorylates APOBEC3G-strep

To ask whether PKA is capable of phosphorylating A3G *in vitro*, we first purified A3G-strep from HEK293T cells. The purified protein was incubated with PKA in the presence of γ-^{32}P-ATP and then the labeled protein was subjected to phospho-amino acid analysis. We found that PKA is capable of phosphorylating at least one serine residue in A3G-strep (**Fig. III-2A**). Many attempts were made to use phospho-peptide mapping to pinpoint the identity of the phosphorylated serine residue(s), however they were unsuccessful.

These analyses were also paired with mass-spectrometry studies to assist in the identification of novel A3G modification in collaboration with Nevan Krogan and Jeff Johnson at UCSF. Using mass-spectrometry, we confirmed the identity of one phosphorylation site in A3G-strep as a serine in the tag between strep and the TEV protease cleavage site (depicted in **Fig. III-2B** and data not shown).

CaMKII phosphorylates untagged APOBEC3G

To determine whether CaMKII is capable of phosphorylating A3G, we performed an *in vitro* kinase assay using purified CaMKII and γ-^{32}P-ATP. To eliminate experimental complications arising from the phosphorylation site identified in the C-terminal strep tag, we generated a baculovirus expression vector encoding A3G-TEV-ZZ for expression and purification of A3G from Sf9 insect cells. After cleavage with TEV protease, the only remaining non-A3G residues on the C-terminus are ENLYFQ, which are not viable

targets for serine/threonine kinase phosphorylation. After tryptic digest of the labeled proteins A3G, A3G(T32A), A3G(T218A) and A3G(T32A-T218A), we performed phospho-peptide mapping to identify phosphorylated peptides. We found at least 4 major phosphorylated peptides and several other minor spots indicated on the 2-dimensional map for A3G (**Fig. III-3**, top left). When we compared the maps of A3G and A3G(T32A), we could not reproducibly identify the disappearance of a specific phospho-peptide spot, indicating that T32 is not likely a substrate for phosphorylation by CaMKII (**Fig. III-3**, top panels). However, when we compared the maps for A3G(T218A) and A3G(T32A-T218A) with the wildtype protein, we noticed the disappearance of a specific phospho-peptide spot (**Fig. III-3**, black arrow). This provides evidence that A3G-T218 is a potential site for phosphorylation by CaMKII.

Discussion

Phosphorylation can act as a molecular switch to result in the activation or inactivation of cellular proteins. In this study, we used phospho-amino acid analysis and phospho-peptide mapping to identify candidate residues on A3G that can be phosphorylated by cellular kinases *in vivo,* or *in vitro* by PKA or CaMKII. We found that A3G with a C-terminal strep tag is phosphorylated at a minimum of three sites (one serine, one threonine and one tyrosine) *in vivo*. Unfortunately, a phosphorylation site in the tag complicated the interpretation of the phospho-peptide maps generated. However, we did observe a number of unique phospho-spots present in the maps for A3G-strep that were not present in the GFP-strep control. This indicates that there are several unique sites for phosphorylation of A3G *in vivo*. Further, the majority of these spots were seen in the C-terminal domain of A3G(191-384), indicating that this domain is more prone to phosphorylation.

We then performed an *in vitro* kinase assay and phospho-amino acid analysis of PKA phosphorylated A3G-strep and found that PKA phosphorylates A3G-strep at a minimum of one serine residue. Through parallel experiments utilizing mass-spectrometry, we found that PKA phosphorylated a serine in the tag on the C-terminus of A3G. To remove experimental complications arising from the phosphorylation site found in the C-terminal tag, we generated a baculovirus expression vector in which A3G can be purified using TEV protease to result in untagged purified protein. Using this untagged protein preparation, we were unable to detect phosphorylation of A3G by PKA (not shown). From these analyses, we do not have sufficient evidence to support PKA as a relevant A3G kinase *in vivo*.

We then asked whether another cellular serine/threonine kinase, CaMKII, was capable of phosphorylating A3G by performing an *in vitro* kinase assay. Examination of the phospho-peptide maps revealed that CaMKII phosphorylates A3G at a number of sites *in vitro*. In one experiment, the evidence suggests that CaMKII phosphorylates A3G at threonine-218.

In summary, we have shown that A3G is subject to post-translational modification by phosphorylation at multiple sites *in vivo*. Our evidence suggests that PKA is not a relevant A3G kinase. However, CaMKII phosphorylates A3G *in vitro* at multiple sites with threonine-218 as a potential candidate site.

Experimental procedures

Plasmids

A3G-strep-TEV-ZZ=pRH4403

A3G(T218A)-strep-TEV-ZZ=pRH2450

A3G(191-384)-strep-TEV-ZZ=pRH2172

A3G(191-384)T218A-strep-TEV-ZZ=pRH2169

GFP-strep-TEV-ZZ=pRH1233

pFastBac-Dual-A3G-TEV-ZZ=pRH4356

pFastBac-Dual-A3G(T32A)-TEV-ZZ=pRH4357

pFastBac-Dual-A3G(T218A)-TEV-ZZ=pRH4359

pFastBac-Dual-A3G(T32A-T218A)-TEV-ZZ=pRH4361

In vitro kinase assays

HEK293T cells transiently transfected with A3G-strep-TEV-ZZ or mutants thereof were lysed in buffer (25mM HEPES pH 7.4; 150 mM NaCl; 5 mM EDTA; 1% Triton X-100; 0.5% sodium deoxycholate; 0.1% SDS; 10% glycerol; 20 mM beta-glycerophosphate; 1 mM sodium vanadate; 10 mM sodium fluoride) and clarified by centrifugation. IgG sepharose beads were added and incubated overnight at 4°C on a rotating shaker. The beads were washed four times in lysis buffer, then three times in PKA buffer (NEB). 40 μL of PKA buffer was added, followed by 1 μL γ-^{32}P-ATP and 1 μL PKA. The reaction was incubated for one hour at 30°C, followed by three washes in TEV protease buffer (Invitrogen). 25 μL of TEV buffer and 1 μL of TEV protease were added and incubated overnight at 4°C on a rotating shaker. The eluted protein was collected, separated by SDS-PAGE, transferred to PVDF membrane, then dried and imaged by phosphorimager to detect labeled protein.

For CaMKII assays, 10 μg of untagged A3G purified from Sf9 insect cells was incubated in 20 μL CamKII buffer (NEB) supplemented with 2 μL γ-^{32}P-ATP and 2 μL CaMKII (NEB). The reaction was incubated at 30°C for 1 hour then separated by SDS-

PAGE, transferred to PVDF membrane, dried and imaged by phosphorimager to detect labeled protein.

In vivo metabolic labeling

HEK293T cells were transiently transfected with A3G-strep-TEV-ZZ or mutants thereof and incubated for 8 hours in DMEM without phosphate supplemented with 0.5 mCi/mL ^{32}P-orthophosphate. The labeled cells were lysed in 25mM HEPES pH 7.4; 150 mM NaCl; 5 mM EDTA; 1% Triton X-100; 0.5% sodium deoxycholate; 0.1% SDS; 10% glycerol; 20 mM beta-glycerophosphate; 1 mM sodium vanadate; 10 mM sodium fluoride. IgG Sepharose beads were added to clarified lysates followed by 4 washes in lysis buffer. The beads were washed 3 times in TEV protease buffer (Invitrogen) then A3G-strep was eluted by TEV cleavage. The eluted proteins were separated by SDS-PAGE, transferred to PVDF membrane, dried and imaged by phosphorimager to detect labeled protein.

Phospho-amino acid analysis

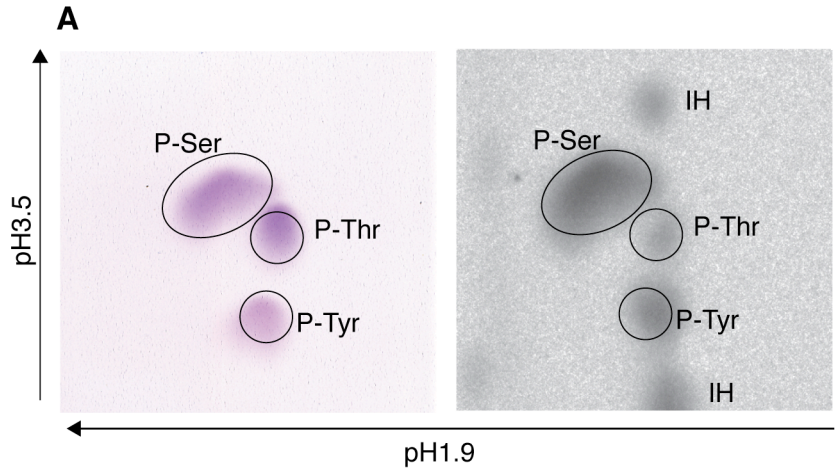
The labeled protein of interest was cut from a PVDF membrane and rehydrated in methanol. After several washes with water, 6 M HCl was added and incubated at 110°C for 60 minutes to hydrolyze the sample. The sample was then evaporated to dryness in a SpeedVac and resuspended in water. After mixing with phospho-amino acid standards (1 μg each of phospho-serine, phospho-threonine, and phosphotyrosine), the sample was spotted on a thin-layer chromatography (TLC) cellulose plate. Electrophoresis was performed in a HTLE 7000 in pH 1.9 buffer (0.58 M formic acid, 1.36 M acetic acid) for 20 minutes at 1.5 kV in the X-direction. The TLC plate was dried and then electrophoresis was performed in pH 3.5 buffer (0.87 M acetic acid, 0.5% pyridine, 0.5 mM EDTA) for 16 minutes at 1.3 kV in the Y-direction. The plate was dried, sprayed with 0.25% ninhydrin and incubated at 65°C for 15 minutes to visualize the phospho-amino acid standards. After phosphorimaging, the presence of specific radiolabeled phospho-amino acids is determined by aligning the two images.

Phospho-peptide mapping

The labeled protein of interest was cut from a PVDF membrane and rehydrated in methanol. The membrane was washed three times with water and then incubated for 30

minutes at 37°C in 0.5%PVP-360 in 100 mM acetic acid. After washing three times with water and two times in 50 mM ammonium bicarbonate pH 8.0, the membrane was incubated overnight at room temperature with 5 μg TPCK-trypsin in 200 μL of 50 mM ammonium bicarbonate. The liquid was transferred to a new tube and lyophilized in a SpeedVac. 1 mL of water was added and the lyophilization was repeated. The digest was resuspended in 10 μL of pH 1.9 electrophoresis buffer (0.58 M formic acid, 1.36 M acetic acid) and spotted onto a TLC plate. Electrophoresis was performed in pH 1.9 buffer for 25 minutes at 1.0 kV. The plate was dried and then chromatography was performed overnight using phospho-chromatography buffer (37.5% n-butanol, 25% pyridine, 7.5% acetic acid). The plate was then dried and imaged by phosphorimaging to identify phosphorylated peptides.

Figures


A
pH3.5
P-Ser
P-Thr
P-Tyr
IH
P-Ser
P-Thr
P-Tyr
IH
pH1.9

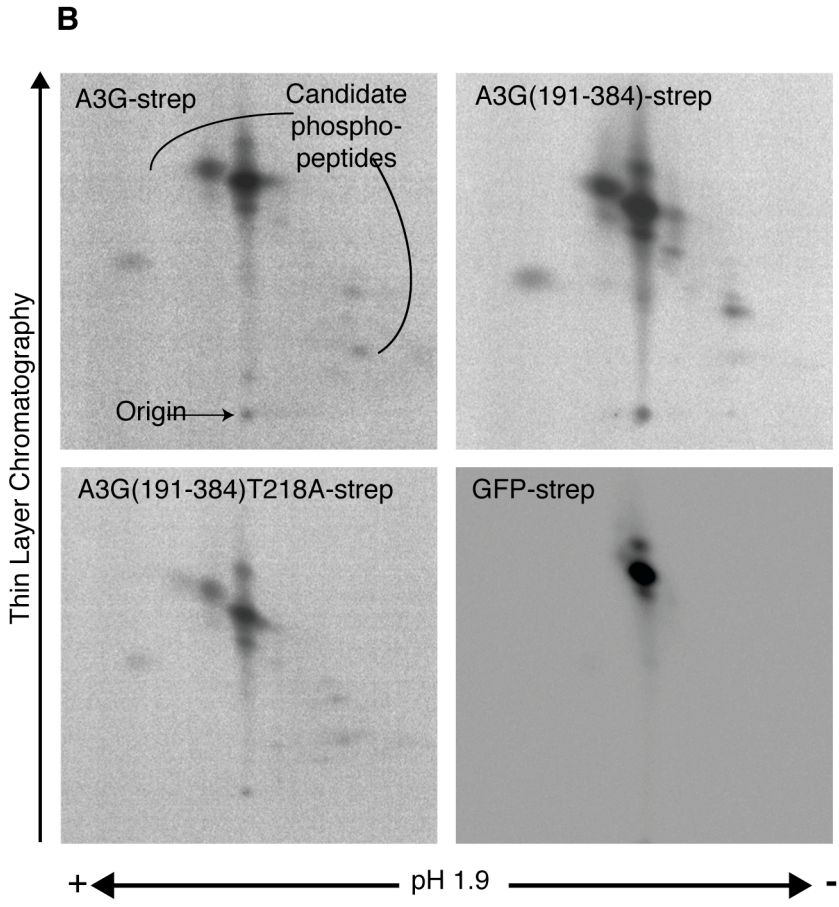

B
Thin Layer Chromatography
A3G-strep
Candidate phospho-peptides
Origin
A3G(191-384)-strep
A3G(191-384)T218A-strep
GFP-strep
+
pH 1.9
-

Fig. III-1. APOBEC3G-strep is phosphorylated *in vivo.* *A*. Phospho-amino acid analysis of *in vivo* labeled A3G-strep. Phospho-amino acid standards are visualized by ninhydrin staining after 2-dimensional electrophoresis (Left panel). The presence of phospho-serine, phospho-tyrosine and phospho-threonine is detected by autoradiography (Right panel). *B*. 2-dimensional phospho-peptide maps indicating the presence of multiple phosphorylation sites on A3G-strep *in vivo*. GFP-strep is also phosphorylated on one major site. IH:incomplete hydrolysis

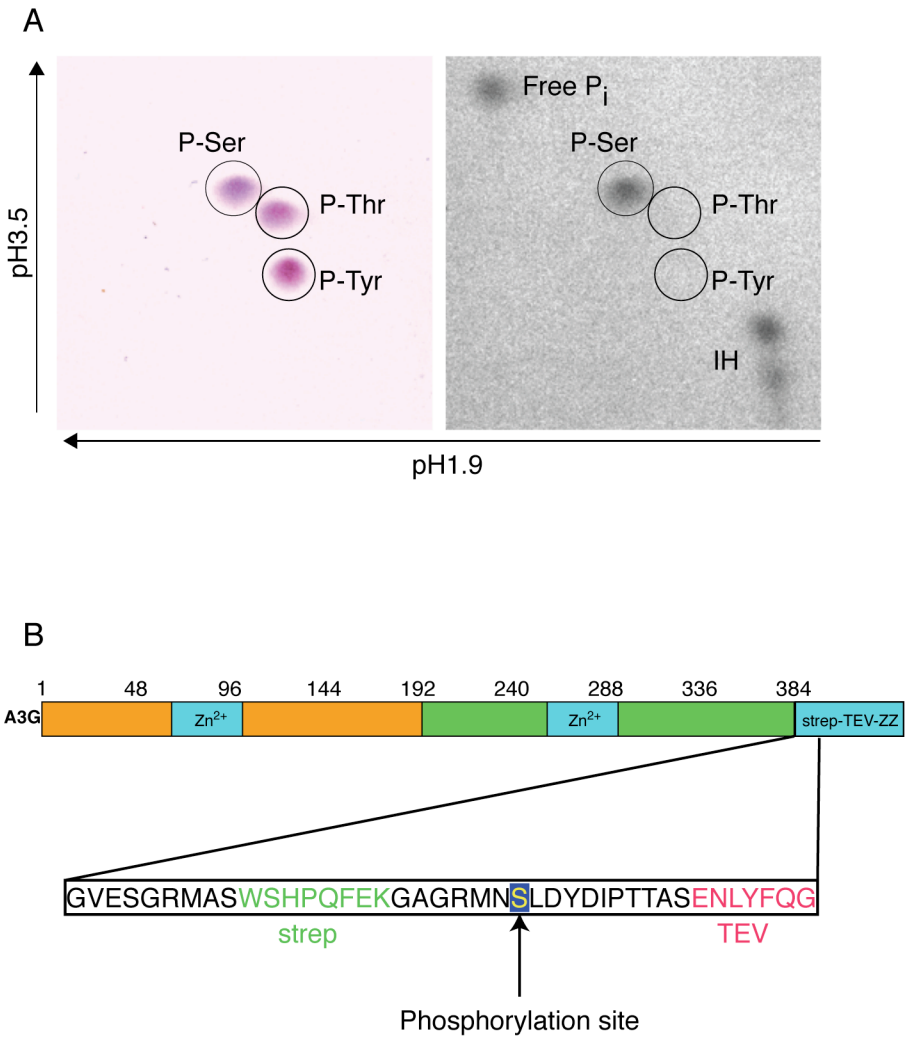


Fig. III-2. PKA phosphorylates APOBEC3G-strep *in vitro*. *A*. Phospho-amino acid analysis of *in vitro* labeled A3G-strep. Phospho-amino acid standards are visualized by ninhydrin staining after 2D electrophoresis (Left panel). Only the presence of phospho-serine is detected by autoradiography (Right panel). IH=incomplete hydrolysis. *B*. Schematic of A3G indicating the phosphorylation site detected by mass-spectrometry in the tag between strep and the TEV protease cleavage site.

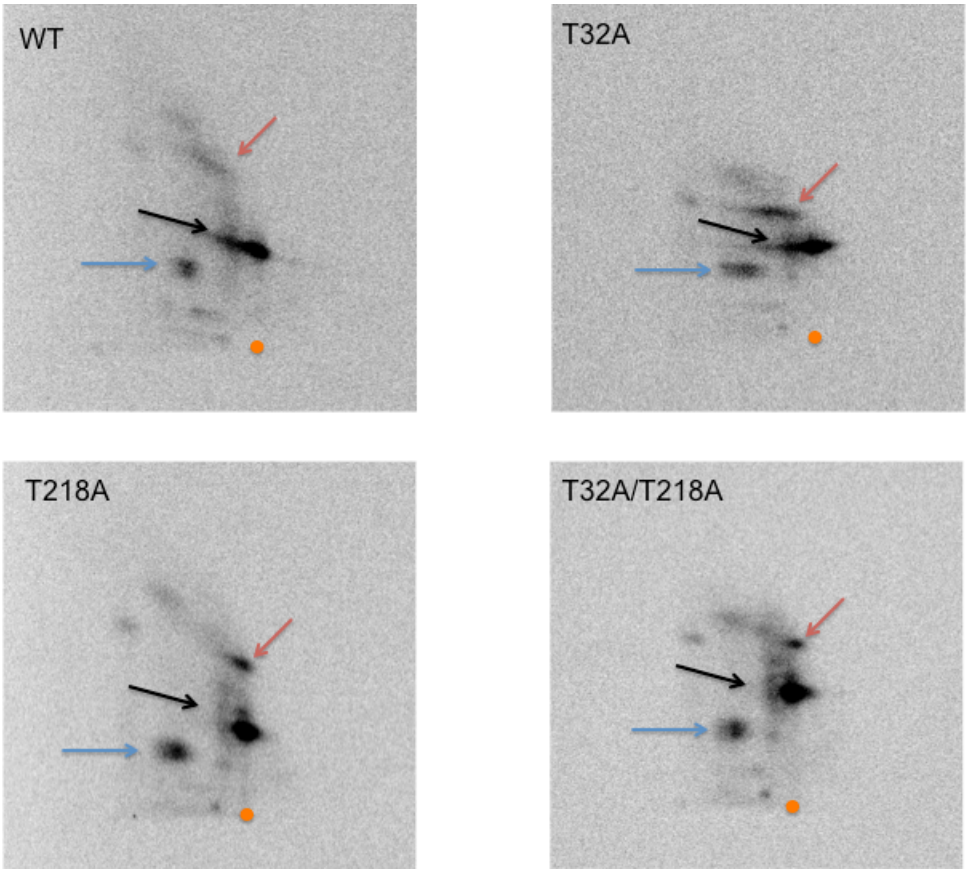


Fig. III-3. CaMKII phosphorylates APOBEC3G *in vitro*. Phospho-peptide maps generated from *in vitro* phosphorylation of A3G with CaMKII. Several major phospho-peptide spots can be seen in the WT enzyme. Orange spot indicates the origin. Colored arrows indicate matching phospho-peptides.