APOBEC3 Proteins Restrict Mobile and Foreign DNA

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

Mark Daniel Stenglein

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

Advisor, Reuben S. Harris, Ph.D.

25 August, 2009

Acknowledgements

I would like to thank my advisor, Reuben Harris, for setting an enthusiastic, ambitious, and inquisitive example, for working hard to create an environment were it is possible to be successful, and for fun lab get-togethers and curling parties.

I would like to thank my family especially my parents, John and Sharon, my siblings Jeremy and Rose, and my sister in law Qingqing for all of their love and support. Sunday dinner with my family has been a ritual that has provided bedrock and, often, great leftovers.

I would like to thank all of my good friends who have eased the passage with their good humor, good ideas, and camaraderie. I am lucky to have as good friends: Laura Breshears, Zach and Shana Demorest, Ron and Erin Eid, Julia Hatler, Davin Henderson, Medora Huseby, David Kast, Brendan Kavaney, Drew Kile, Tony Kujawa Andy Lane, Marjan Mashhadi, Ben Moore, Laura Ramsey, Eric Rahrman, and Brad Scutvick.

I would like to thank all of the Harris lab members. I would like to especially thank my fellow graduate students Guylaine Haché, Rebecca LaRue, and Zach Demorest. Also thank you to Bill Brown who does a great job keeping things under control. Thank you to the many lab workers and undergraduate assistants who have provided invaluable help. Thank you to those who have helped with experiments in this thesis including: John Valesano, Natalia Martemyanova, Ming Li, Brian Hoium, Mengfei Liu, Phil Gross, and Melissa Cornwell.

I would like to thank the Hendrickson, Bielinksy, Sobeck, and Livingston lab members for all of the good cheer and science at Friday lab meetings. Eric Hendrickson has been particularly supportive and has written letters for several award and job applications.

I would like to thank my committee members: Tim Griffin, Do-Hyung Kim, Hiroshi Matsuo, Nik Somia, and Meg Titus. Thank you especially to Meg Titus for providing helpful, honest, and constructive criticism. Meg has also written several letters for fellowships and job applications and has very generously shared her microscope.

I would like to thank the University of Minnesota's Institute for Molecular Virology and Center for Genome Engineering.

Thank you to John Hanten at 3M pharmaceuticals for providing intriguing preliminary data, reagents, and good discussion.

Thank you to Yoji Shimizu, Sandi Sherman, and Chris Penell for efficiently running the cancer biology training grant and related meetings and seminars.

For reagents, I would like to thank Roger Tsien, Scott McIvor, Perry Hackett, Christine Rada, Michael Neuberger, and Javier Di Noia.

I would like to acknowledge financial support from the University of Minnesota Cancer Biology Training Grant (CA009138) and the 3M Science and Technology Graduate Fellowship.

Dedication

This dissertation is dedicated to my parents, John and Sharon, and to Medora Huseby.

Abstract

A cell's DNA provides its operating instructions. Consequently, cells rigorously maintain their DNA. At the same time, cells must prevent foreign DNA from persisting, or risk having their operations hijacked. This thesis describes how APOBEC3 proteins help fulfill this double imperative, by limiting the replication of mobile genetic elements, and by triggering the destabilization of foreign intracellular DNA.

The APOBEC3s are a family of DNA modifying enzymes that convert deoxycytidines to deoxyuridines. Introducing uridines into DNA can alter the sequence of genes or regulatory elements, in essence removing information content from the molecule. Such an alteration also often leads to the destabilization and degradation of the DNA. Previously, APOBEC3s had been shown to act on a relatively limited subset of viral and retrotransposon DNA replication intermediates. The major contribution of this work is to expand the range of biologically relevant APOBEC3 substrates to include LINE-1 retrotransposons and foreign intracellular DNA in general.

LINE-1s are the only retrotransposons currently active in humans, and their mobilization can cause disease by a variety of mechanisms, the most straightforward of which is by inserting in or near a gene and disrupting its function. Expression of several APOBEC3 proteins reduces the rate of LINE-1 retrotransposition in human cells. Less retrotransposed LINE-1 DNA accumulates in these cells, suggesting that the block occurs prior to retrotransposon integration. The APOBEC3s therefore limit the mutagenic potential of LINE-1 elements.

Foreign intracellular DNA is inherently dangerous and is often associated with microbial infection. In recognition of this danger, sensors within cells detect foreign

DNA, but, previously, little was known about mechanisms that respond to DNA detection to mediate its clearance. This thesis demonstrates that APOBEC3 family members are induced following DNA detection and destabilize foreign DNA. This defense system is likely to have evolved to respond to intracellular microbial DNA, but it may also diminish the efficacy of processes such as genetic engineering and gene therapy.

Finally, this thesis explores the molecular basis of the differential sub-cellular localization of the APOBEC3 proteins. Sequestering APOBEC3 proteins in different cellular compartments may be a way to regulate the DNA substrates to which they have access. This thesis therefore offers several insights into a family of proteins that modify potentially harmful DNA in order to protect the well-being and proper function of cells.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Portions of this chapter are adapted with permission, from Stenglein, LaRue, Schumacher, and Harris. "Host Factors that Restrict Retrovirus Replication", a chapter of: Viral Genome Replication Cameron, Craig E.; Götte, Matthias; Raney, Kevin D. (Eds.) 2009. ISBN: 978-0-387-89425-6

DNA is the genetic material of cellular organisms. It is therefore critical that cells maintain the integrity of their genomes, and multiple systems detect and repair damaged DNA[139]. In addition to protecting their own DNA, cells must prevent foreign DNA from persisting or being expressed. This thesis is about a family of DNA modifying enzymes, the APOBEC3s, and how these proteins safeguard the cell by limiting the mobility of viruses and transposable elements and by counteracting foreign DNA.

The APOBEC3s form a family of seven DNA cytidine deaminases in humans. These enzymes modify DNA, converting deoxycytidine (for short, cytidine, or C) to deoxyuridine (uridine, or U) (**Fig. 1-1**). Uridines base pair like thymidines (T) and a C-to-U conversion creates a G:U mismatch in double-stranded (ds) DNA. Like any DNA lesion, such a mismatch has the potential to disrupt cis-acting regulatory elements and protein coding capacity. If left unrepaired, the C-to-U lesion can become fixed as a C-to-T mutation when the DNA is replicated. Consequently, DNA uridines are normally recognized and removed by DNA repair pathways[139] (

Fig. 1-2).

So in essence APOBEC3 proteins do the opposite of what would seem to be beneficial to the cell (

Fig. 1-2). However, the very fact that APOBEC3s exist suggests that they must perform some useful function. The beneficial function of APOBEC3s can be understood by answering two fundamental questions: what DNA molecules are APOBEC3 substrates? And, what is the purpose of introducing uridines into these molecules?

The major findings of this thesis help answer these questions. The findings are that: APOBEC3s limit the mobility of LINE-1 retrotransposons **(chapter 2),** and that APOBEC3s engage and destabilize intracellular foreign DNA **(chapter 3).** In other words, this thesis defines two new classes of APOBEC3 substrates and suggests biological purposes for deaminating them.

One obvious follow-up question is: How are APOBEC3 proteins specifically targeted to some DNA molecules and kept away from others (*e.g.,* genomic DNA)? One means of accomplishing this may be to sequester APOBEC3s in different subcellular compartments, and **chapter 4** of this thesis examines the molecular bases for differential APOBEC3 subcellular localization.

The Discovery of APOBEC3s

The APOBEC3 (A3s) proteins were discovered independently and virtually simultaneously three different ways[60,70,147]. One study used genomic DNA sequencing and a computational approach to identify a cluster of genes on chromosome 22 that were predicted to encode proteins that shared an amino acid motif with the RNA-editing enzyme APOBEC1 (**apo**lipoprotein **B**, mRNA-**e**diting **c**atalytic polypeptide-1)[70]. These newly identified genes were named APOBEC3A – APOBEC3H (A3A – A3H; **Fig. 1-4**). A second study answered a longstanding question in the human immunodeficiency virus (HIV) field, by identifying A3G as a cellular factor that prevented certain HIV strains from replicating in certain T-cell lines[147]. A third report demonstrated that, unlike their namesake APOBEC1, A3s were not *RNA* cytidine deaminases but instead were *DNA*-editing enzymes[60]. An explosion of studies followed these three foundational reports, and their key findings are discussed below.

The APOBEC3s are part of a larger family of related enzymes.

A3s were so named because their predicted amino acid sequences share a motif with the mRNA-editing enzyme APOBEC1 (H-x-E-x_{24}-C-P-x_{2-4}-C, where x is nearly any amino acid; **Fig. 1-3**) [27,59,114]. The conserved amino acids in this motif are catalytically important. The histidine and the two cysteines coordinate a zinc ion and the glutamate is thought to act as a proton donor in the deaminase reaction [114,179]. In addition to APOBEC1 and the A3s, the larger APOBEC family includes APOBEC2, APOBEC4, and activation induced deaminase (AID)[27]. Of these, APOBEC2 and AID appear to be the oldest evolutionarily, as they are the only ones found in all vertebrates[37,90,105]. APOBEC4 is specific to higher vertebrates[135]. APOBEC1 and the A3 proteins are found only in mammals[29,30,49,59,70,173]. A brief overview of these other family members follows, with a focus on what is known about their physiological nucleic acid substrates and functions.

APOBEC1: The namesake of the APOBEC family, APOBEC1's best-characterized function is to deaminate cytidine 6666 of the Apolipoprotein B (ApoB) mRNA[22,115,131]. Deaminating this nucleotide converts a glutamine codon (CAA) into a stop codon (UAA). APOBEC1 is specifically targeted to this cytidine by cis-acting elements in the ApoB mRNA and by a protein co-factor, APOBEC1-complementation factor[85,102,103]. APOBEC1 is predominantly expressed in the intestine and liver. Consequently, editing is restricted to those tissues, where it results in the expression of a shortened ApoB protein[22,62,131].

APOBEC1 was found to possess both RNA and DNA deaminase activity[60], but it's principle physiological function appears to be to edit the ApoB mRNA. The only discernible phenotypes of APOBEC1-deficient mice are a loss of ApoB mRNA editing

and subtle changes in blood lipoprotein composition[62,111]. Thus, APOBEC1's biologically relevant substrate is the ApoB mRNA and deamination serves to generate two protein isoforms from this mRNA.

APOBEC2: APOBEC2 was identified as a protein containing the characteristic APOBEC cytidine deaminase motif[70,79,86]. However, APOBEC2 has shown no DNA, RNA or free nucleotide cytidine deaminase activity in various assays, and its biological function remains unclear[60,105]. APOBEC2 deficient mice have no discernable phenotypic difference from their APOBEC2 proficient littermates [105].

APOBEC4: More recently, another gene predicted to encode a protein with sequence homology to the other APOBEC family members was identified and named APOBEC4[135]. Like APOBEC2, it remains unclear whether APOBEC4 possesses enzymatic activity or what its function *in vivo* might be.

Activation Induced Deaminase (AID): AID is expressed predominantly in activated B cells, where it deaminates genomic cytidines at the immunoglobulin (antibody) locus [110,130]. This deamination is required for somatic hypermutation and class switch recombination, two components of antibody gene diversification, an integral part of the adaptive immune response[1,37,61,90,93,95,109]. Mice lacking AID are defective in somatic hypermutation and do not undergo class switch recombination[109]. Similarly, humans with AID deficiency suffer from immunodeficiency because they are unable to undergo class-switch recombination[133]. Thus, AID's physiologically relevant substrate is genomic DNA (at a particular locus), which AID deaminates to trigger antibody gene diversification. It remains unclear how AID is targeted specifically to the antibody locus and not to other parts of the genome. In addition to this role in adaptive immunity, AID

may engage retroelement DNA substrates, and therefore may also function in the innate immune system in a manner similar to the A3s (described below)[92].

The APOBEC3s: Because of their similarity to AID (in sequence and in enzymatic activity) it has been proposed that duplications of an ancestral AID gene gave rise to the precursors of the present day APOBEC1 and A3 genes[27,59,83]. A3 gene number varies in different mammalian species (**Fig. 1-4**). For instance, the mouse has a single A3 gene, but the human family is composed of seven members (**Fig. 1-4**). Phylogenetic analyses suggest that these differences are due to lineage-specific gene expansions and deletions.

A3s can be classified into one of two categories, termed single or double domain. Single domain A3s (in humans A3A, A3C, and A3H) are about 200 amino acids long (~25 kDa). Their structure is predicted to be similar to that of the C-terminal half of A3G (**Fig. 1-3**). At around 400 amino acids (~50 kDa), double domain A3s (A3B, A3DE, A3F, and A3G in humans) are about twice as long as the single domain proteins. Double domain A3s contain two characteristic zinc-binding motifs, and likely arose from the duplication or fusion of two single-domain A3 genes [27,30,83].

A3 proteins exhibit two major biochemical activities *in vitro*: cytidine deamination and nucleic acid binding. Cytidine deamination appears to be restricted to single-stranded (ss) DNA substrates; A3s show no activity on RNA, RNA/DNA hybrids, or double-stranded (ds) DNA substrates[60,68,70,161,183]. Although A3s only deaminate ssDNA *in vitro*, they bind single- and double-stranded nucleic acids in a sequence non-specific manner[68].

Although they originated from an AID-like ancestor, A3s function in the innate immune system and not in adaptive immunity, like AID[105,109]. The innate immune response does not involve the direct production of antigen-specific T- and B-cell receptors. Instead the innate immune response targets invariant molecular features of pathogens[69]. The emerging picture of A3 function is that they employ their cytidine deaminase activity to neutralize potentially harmful DNA molecules, including viral and transposable element DNA.

A3s limit retroviral replication.

The most studied and best-known A3s function is to block retroviral replication. This antiviral activity was first demonstrated in 2002 when it was shown that A3G was a cellular factor antagonistic to HIV-1 replication[147]. The existence of such a factor was inferred before its identity was unveiled. This inference was based on the observation that HIV-1 strains lacking the virion infectivity factor (Vif) accessory gene were only able to replicate in a subset of human T cell lines. The Vif-proficient parent virus replicated normally in these cells [*e.g.*, [45,50,112,151,158,170]]. Hybrid T cells, created by fusing permissive and non-permissive cells, also failed to replicate Vif-defective viruses[94,150]. These fusion experiments argued that Vif was not required to compensate for a cellular factor lacking in non-permissive cells. Instead, the results were consistent with the existence of a dominant cellular factor in the non-permissive cells that prevented replication of Vif-deficient HIV-1.

To identify this presumed antiviral factor, Malim and coworkers made use of a related pair of T cell lines: the permissive CEM line and its non-permissive derivative,

CEM-SS [151]. Reasoning that the dominant factor would be more highly expressed in non-permissive cells, these investigators performed subtractive hybridization experiments to isolate mRNAs that were more abundant in CEM than in CEM-SS [147]. Among the mRNAs so identified was A3G. Expressing A3G in CEM-SS cells rendered them non-permissive for HIV-1ΔVif replication. A number of investigations following this study answered two key questions: (1) what is the molecular mechanism underlying the antiviral effect of A3G? And, (2) how does Vif permit HIV-1 to replicate in the presence of this potent anti-viral protein?

The fact that A3G is a DNA deaminase suggested that it might exert its antiviral effect by deaminating viral DNA[60]. It was hypothesized that A3G deaminated ssDNA intermediates during reverse transcription, the process by which retroviruses convert their genome from a ssRNA copy to a dsDNA copy (**Fig. 1-5**). In support of this hypothesis, proviral DNA (integrated dsDNA viral genomes) from viruses produced in A3G-expressing cells contained dramatic increases in C/G-to-T/A mutations[58,84,96,185]. Because they were biased to one strand, the mutations could be attributed to deamination of cytidines on the first DNA strand synthesized during reverse transcription (**Fig. 1-5**)[58,96,183,185].

Numerous studies have further elucidated the mechanism by which A3G blocks HIV-1 replication, and have shown that there are several major requirements for this antiviral activity. The first such requirement is for cytidine deaminase activity. There are reports that overexpression of A3G mutants lacking enzymatic activity partially inhibit retroviral replication[7,119,149]. But the inhibition disappears at lower expression

levels[15,106,144]. And, when expressed at levels comparable to endogenous protein in primary cells, A3G mutants lacking enzymatic activity do not limit the replication of HIV-1ΔVif[15,106,144]. The second major requirement is that A3G must be packaged into newly forming virus particles as they bud from an infected cell (**Fig. 1-5**)[116,149]. This is because reverse transcription in a newly infected cell takes place in the protected environment of the viral core. Therefore, A3G in the cell being infected has no access to the reverse transcription reaction, and does not block viral replication. The third major requirement is that A3G has to evade viral countermeasures, such as the Vif protein (this is obviously not the case in most HIV-1 infected individuals).

Indeed, understanding the mechanism of A3G's antiviral activity did not help explain how Vif allowed HIV-1 to replicate in the presence of A3G. However, it was not long before several groups showed that Vif triggered the proteasome-dependent degradation of A3G in infected cells[28,88,98,101,148,184]. This degradation reduces A3G protein levels below an effective threshold, allowing the virus to replicate unhindered. An epitope-tagged HIV-1 Vif protein was used to affinity purify an E3 ubiquitin ligase complex consisting of CUL5, ELONGIN B, ELONGINC and RBX2[184]. Cells depleted for these proteins or expressing dominant negative variants preserved A3G and resisted HIV-1 (Vif+) infection[101,184].

Several salient points have emerged from the studies on the antiviral activity of A3 proteins. First, other A3s in addition to A3G are capable of limiting retroviral replication. A3F, in particular, appears to be as good as A3G at blocking HIV-1 replication[87,176,186]. Other A3s have been shown to inhibit HIV-1 weakly or not at

all[8,153,182]. Furthermore, various A3s have been demonstrated to restrict the replication of other viruses, and not just HIV-1. The list of A3-susceptible viruses includes retroviruses such as human T-lymphotropic virus type I (HTLV-1), murine leukemia virus (MLV), simian immunodeficiency viruses (SIV), and equine infectious anemia virus (EIAV) [39,96]. The list also includes viruses similar to retroviruses, such as hepatitis B virus (HBV), and more distant DNA viruses such as adeno-associated virus (AAV), however the data on non-retroviruses is much less abundant[18,165]. There also appears to be species specificity to these antiviral activities. So, for instance, human A3G does not limit wild type HIV-1 replication (because of Vif), but human A3G restricts the simian immunodeficiency virus that infects African green monkeys (SIVagm). Conversely, the African green monkey A3G restricts HIV-1, but not SIVagm. Thus, an important function of A3 proteins may be to prevent cross-species viral transmission.

APOBEC3s limit the mobility of retrotransposons

As described above, the A3 gene family has expanded in the primate lineage, and there are seven human A3 genes (**Fig. 1-4**). An evolutionary analysis has estimated that this expansion occurred in the period between 100 and 25 million years ago, *i.e.*, long before HIV began infecting humans within the past hundred years[53,83]. Therefore, strictly speaking A3 genes did not evolve to block HIV replication. Of course, A3s may have evolved to block the replication of other, more ancient, retroviruses, and that is likely the case. But, some A3 family members appear to have very limited or no activity against retroviruses (at least against those viruses that have been tested). So, a major question remains: what are the other functions of these proteins? A number of studies, including

one described in **chapter 2** have demonstrated that one major additional A3 function is to limit the mobility of transposable retroelements.

Judging from their prevalence in many genomes, transposable elements are very successful genetic pathogens. For instance, approximately one half of the human genome is composed of DNA that is recognizably derived from these mobile DNA elements[81]. Such elements can be broadly divided into two classes: retroelements and DNA transposons (**Fig. 1-6**)[34,81]. Retroelements mobilize via a "copy and paste" mechanism: They transcribe their genome, and this transcript is reverse transcribed and integrated into a new genomic locus. DNA transposons, in contrast, move via a "cut and paste" mechanism. The transposon-encoded transposase enzyme excises the entire dsDNA transposon and subsequently integrates it into a new locus.

Retroelements can be further divided into four major classes (**Fig. 1-6**). The first class consists of retroviruses such as HIV. The second class is known as the LTR retrotransposons, because, like retroviruses, their genomes contain long terminal repeats. Because LTR retrotransposons share many characteristics of genome organization and life cycle with (exogenous) retroviruses, these are sometimes referred to as "endogenous retroviruses". The main difference between exogenous and endogenous retroviruses is that the life cycle of endogenous retroviruses is confined to a single cell, whereas the retroviruses lifecycle begins in one cell and ends in another. The third major class of retroelements is the non-LTR retrotransposons whose genomes, like their name suggests, lack LTRs. The genome organization and lifecycle of these elements differs substantially from the LTR retroelements (these differences are explained in greater detail in **chapter**

2). The final class of retroelement is the non-autonomous retrotransposon. These are relatively short DNA elements that do not encode any gene products. Instead, these retrotransposons rely on proteins encoded by non-LTR retrotransposons to mobilize.

Because of their similarities to exogenous retroviruses, and because of the broad anti-retroviral activity of A3 proteins, it was perhaps not too surprising to find that human A3s limit the mobility of endogenous retroviruses. This was shown for mouse, pig, and yeast retroelements[40,41,72,145]. Perhaps more surprising was the finding that A3s also decrease the transposition efficiency of non-LTR retrotransposons. **Chapter 2** describes one of the first studies arriving at this conclusion [10,11,18,26,78,108,120,153].

Are APOBEC3 activities relevant *in vivo*?

Although many of the above-mentioned studies showing that A3 proteins possess antiviral and anti-retrotransposon activity are based on cell culture experiments, there is ample evidence that A3 proteins perform these functions *in vivo*. First, A3-deficient mice are more susceptible to infection by murine leukemia virus and mouse mammary tumor virus[91,122,162]. Second, there are many reports of patient-derived HIV sequences containing high levels of C/G-to-T/A transition mutations (*e.g.*, reference[46]). In fact, this phenomenon, termed "hypermutation", was evident well before A3G was discovered. There are also reports of patient-derived hypermutated hepatitis B virus (HBV) and human papilloma virus (HPV) sequences[159,169]. Third, the existence of viral countermeasures argues that A3 proteins constitute a potent barrier to viral replication. HIV-1 only encodes 15 proteins, but a major function of one of them, Vif, is to cause A3 proteins to be degraded. Other retroviruses don't necessarily encode Vif proteins, but

many have ways of evading A3s. For example, the foamy viruses (a family of retroviruses) encode Bet, a protein that allows the virus to evade A3s[89,137]. Finally, bioinformatics analyses have shown that endogenized mouse and human retroelements bear scars of A3-like deamination events (*i.e.,* a higher level of C/G-to-T/A mutations than would be predicted to occur randomly)[2,42,71,77,129]. These studies testify to the *in vivo* relevance of A3s.

Figures

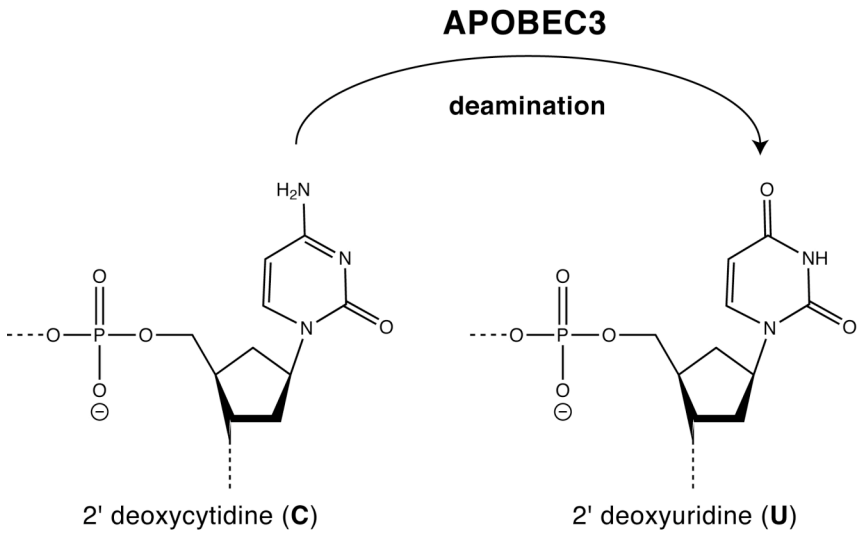

Fig. 1-1. APOBEC3s are DNA cytidine deaminases. A3s convert 2' deoxycytidine to 2' deoxyuridine within single-stranded DNA substrates.

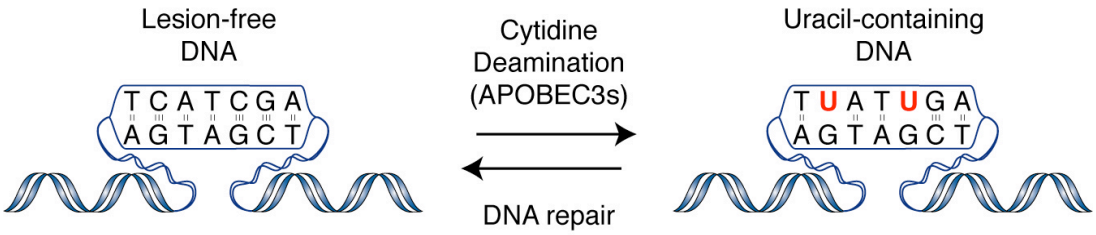

Fig. 1-2. APOBEC3s seemingly oppose DNA repair systems. By deaminating cytidines to uridines, A3 proteins introduce lesions into DNA. Such lesions are normally repaired by DNA repair systems.

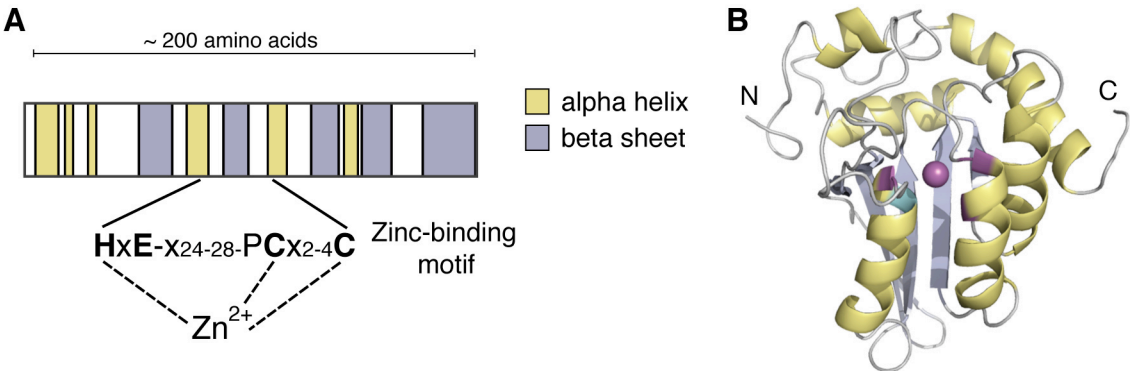

Fig. 1-3. A typical APOBEC3 domain. (A) A schematic of an A3 domain, typified by the C-terminal half of A3G. Secondary structure elements and catalytically important active site residues are indicated. (B) A ribbon diagram of the NMR structure of the C-terminal half of A3G (adapted from [20,57]; PDB 2KEM). Secondary structure elements are colored as in (A). The active site zinc ion is colored purple, as are zinc-binding residues. The catalytic glutamic acid is colored cyan. The N- and C- termini are indicated.

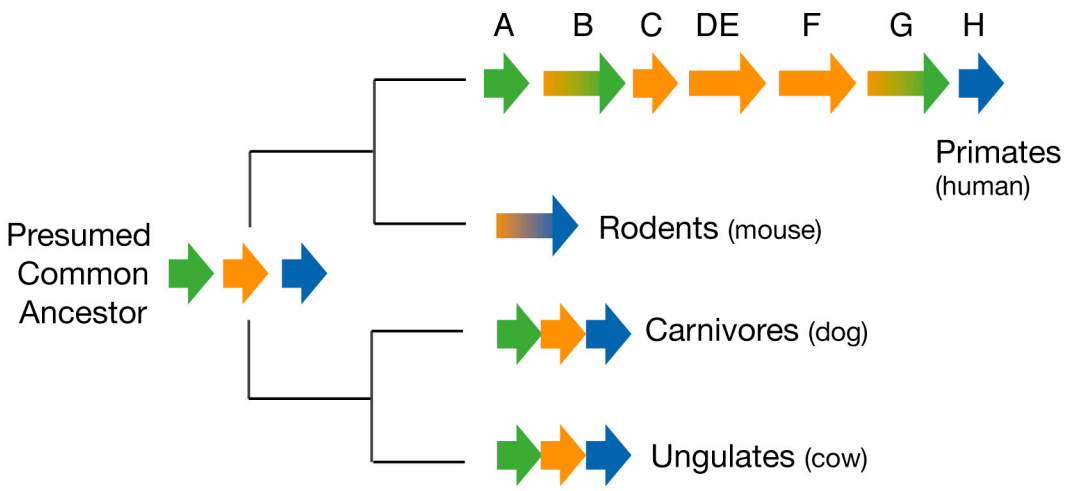

Fig. 1-4. The mammalian A3s. The A3 repertoire of several representative mammalian species is shown. Coloring reflects deduced evolutionary relationships. Double domain A3s are depicted as being twice as long as single domain A3s (adapted from[82]).

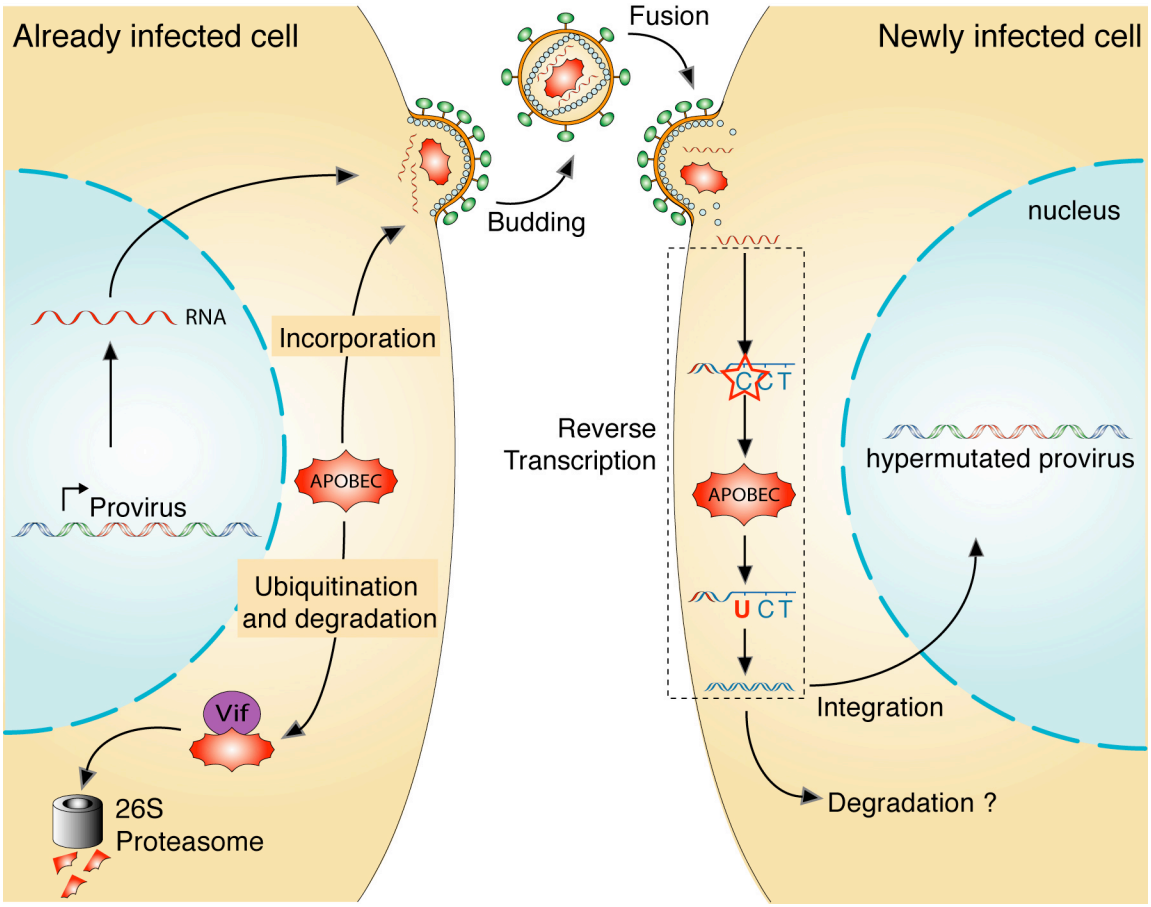

Fig. 1-5. A model of how A3G blocks retroviral replication. In the absence of Vif, A3G is packaged into retroviral particles as they bud from an infected cell (left). Once the virus infects a new cell (right), A3G acts, as the viral RNA genome is reverse transcribed into a dsDNA copy. A3G deaminates cytidines in nascent viral DNA, converting them to uridines. The result is either integration of a hypermutated, genetically inactivated, viral genome (provirus), or degradation of the viral DNA. In the presence of Vif, A3G is degraded in a proteasome-dependent manner and thus is unavailable to be packaged into a virus particle. See text for additional details.

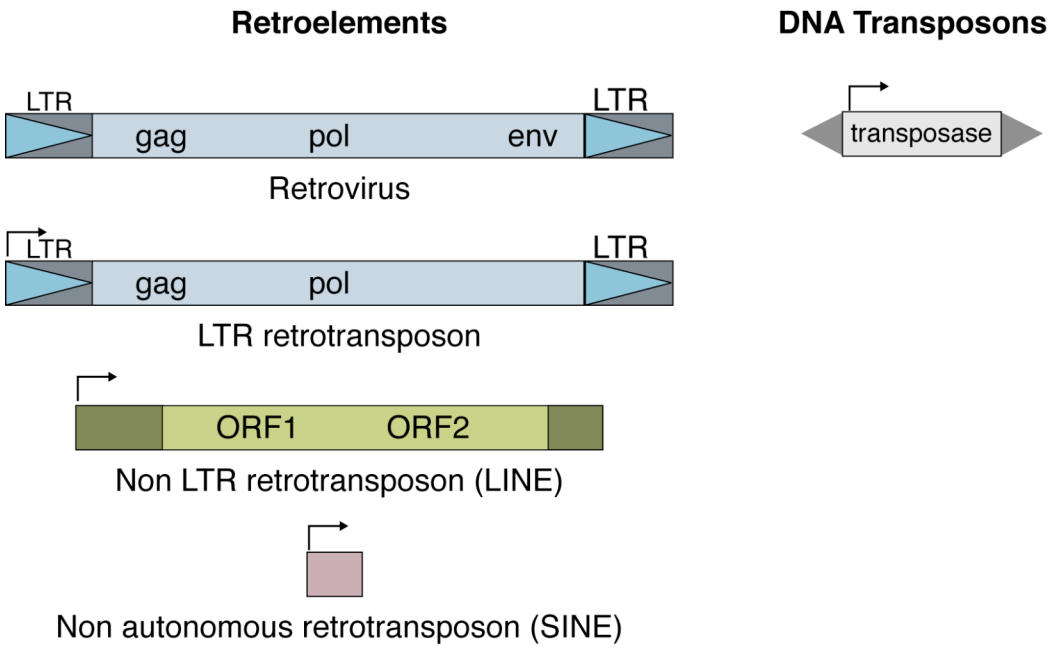

Fig. 1-6. The major classes of transposable elements. Cartoon schematics of the genomes of the major classes of transposable elements. Genomes are drawn approximately to scale. Dark shaded ends of elements indicate the long terminal repeats (LTR) of retroviruses and LTR retrotransposons, the untranslated regions of non-LTR retrotransposons, and the inverted repeats of DNA transposons. Element-encoded gene products are indicated. Figure adapted from: [34].

Chapter 2: APOBEC3B and APOBEC3F Inhibit L1 Retrotransposition by a DNA Deamination-Independent Mechanism

This chapter provides evidence that A3 proteins limit the mobility of LINE-1 retrotransposons. This is important because LINE-1 elements are the only active transposable element in humans, and because previously APOBECs were only known to inhibit LTR-containing retroelements such as retroviruses and related retrotransposons.

Adapted, with permission, from: Stenglein and Harris (2006) J Biol Chem 281(25) 16837-16841

Chapter Summary

The most common transposable genetic element in humans, the long interspersed nuclear element-1 (LINE-1, or L1), constitutes about 20% of the genome. The activity of L1 and related retrotransposons such as Alu elements causes disease and contributes to speciation. Little is known about the cellular mechanisms that control their spread posttranscriptionally. Here, we show that expression of human A3B or A3F decreased the rate of L1 retrotransposition by 5 to 10-fold. Expression of two related proteins, A3D or A3G, had little effect. Two lines of evidence indicated that these A3 proteins use a deamination-independent mechanism to inhibit L1. First, a catalytically inactive A3B mutant maintained L1 inhibition activity. Second, C-to-T hypermutations were not detected in L1 elements that had replicated in the presence of A3B or A3F. In addition, lower levels of retrotransposed L1 DNA accumulated in the presence of A3B and A3F. Together, these data suggest a model in which A3B or A3F provide a pre-integration barrier to L1 retrotransposition. An inverse correlation between L1 activity and A3 gene number suggest the relevance of this mechanism to mammals.

Introduction

Several human APOBEC3 (A3) proteins have been shown to be capable of inhibiting the replication of a diverse and growing number of retroelements (for a review see chapter 1). Retroelements replicate by reverse transcribing their genomic RNA into a double-strand DNA. Prior to this study, A3 anti-retroelement activity was thought to be limited to those whose genomes contained long terminal repeats (LTRs). For such LTR-containing retroelements, reverse transcription occurs in the cytoplasm of the host cell. It is in this subcellular compartment that the human A3 proteins are thought to engage the retroelement as it assembles its core components into a ribonucleoprotein complex. Much of the mechanistic understanding of this activity has derived from studying A3G (A3G). A3G accesses assembling virus particles through interactions with their Gag structural protein[41,147]. There, A3G is positioned to exert its antiviral effect by deaminating nascent cDNA cytidines during reverse transcription[58,96,185]. The uridines template the incorporation of second cDNA strand adenosines and, ultimately this mutagenic process results in a lethal level of retroelement C/G -> T/A transition mutations (hypermutations). Additional evidence indicates that several of the other human A3 proteins can also inhibit a variety of retrotransposons and retroviruses by a similar cDNA cytidine deamination mechanism [e.g., [8,10,15,35,58,87,96,106,137,144,145,159,176,182]]. However, partial inhibition of HIV-1[56,116,119,149], HBV [136,146], HTLV[141] and yeast Ty1[40] was still observed with A3G catalytic mutants, indicating that, although deaminase activity is the major mechanism, other deamination independent mechanisms might also contribute to inhibition.

LTR retrotransposons such as the human endogenous retrovirus make up about 8% of the human genome, and most (if not all) of these are thought to be inactive[81]. In

contrast, about 75% of the repetitive DNA in the human genome (34% of the total) is composed of retroelements that do not require LTRs for replication. The most prominent examples include long interspersed nuclear element-1 (LINE-1 or L1), which replicates autonomously, and Alu elements, whose replication requires L1 protein activities[34,124]. These elements make up 17 and 11% of the genome respectively[81]. The mechanism of L1 replication is fundamentally different from that of LTR-dependent retroelements. Most importantly, L1 elements reverse transcribe their RNA genome upon integration into the host cell genomic DNA by a process called target-primed reverse transcription[124]. Although most of the estimated 500,000 L1 elements in the human genome are incapable of mobilizing, about 100 are active and they can compromise genomic integrity by disrupting genes, interfering with transcriptional programs, mobilizing other elements and precipitating recombination[34,124]. One of every 10 to 250 human infants is estimated to contain a new L1 insertion[124]. L1 elements can thereby cause disease and, over many generations, contribute to speciation. Hypermethylation of L1 element genes has been shown to lead to their transcriptional silencing[143], but cellular mechanisms that limit L1 activity post-transcriptionally are poorly understood.

During A3 protein localization studies we discovered that A3B is predominantly nuclear, in contrast to A3G, which is mostly cytoplasmic (see chapter 4 and refs[96,98]). Because L1 elements undergo nuclear reverse transcription, and because A3s are thought to block retroelements at this stage, we hypothesized that A3B could inhibit L1 retrotransposition. Using a GFP-based L1 retrotransposition reporter assay, we found that expression of A3B or A3F strongly inhibited L1 retrotransposition. Expression of the related human proteins A3D or, as previously shown[166], A3G did not significantly

effect the rate of L1 retrotransposition. Interestingly, the mechanism of inhibition appeared to be deaminase-independent, as hypermutations were not apparent in retrotransposed L1 DNA and inhibition was still observed with a catalytically inactive A3B protein. Such a mechanism was further supported by PCR experiments showing that less integrated L1 DNA accumulated in the presence of A3B or A3F. These data combined to suggest a model in which these A3 proteins interfere with L1 transposition prior to retroelement integration.

Results

The cytoplasmic localization of the human A3G protein[96,98] correlates with its ability to inhibit the replication of LTR-dependent retroelements [reviewed in **chapter 1** and in ref[32]]. We confirmed prior reports of A3G's predominantly cytoplasmic localization by visualizing A3-GFP fusion proteins in live HeLa cells, and showed that A3D and A3F located similarly (**Fig. 2-1**A). In contrast, A3B-GFP localized predominantly to the nucleus, and GFP alone distributed throughout the cell (**Fig. 2-1**A). Similar results were obtained in human 293T cells with A3-GFP and A3-HA constructs, indicating that localization was not cell, tag or fixation-dependent (**Fig. 2-1**B and see chapter 4).

The nuclear compartmentalization of A3B and the fact that L1 elements reverse transcribe in the nucleus suggested that A3B might inhibit replication of L1. To test this hypothesis, it was necessary to establish human 293 cell clones expressing near-equivalent amounts of HA-tagged A3B, A3D, A3F and A3G (**Fig. 2-2**A). It is important to note that these A3 constructs were not massively over-expressed, as control immunoblots using a polyclonal antibody specific to A3G showed that the A3G-HA levels were comparable to the levels of endogenous A3G in two well-studied, HIV-infectable T cell lines [H9 and CEM; *e.g.*, [147,176]; data not shown]. Thus, to a reasonable approximation, the stable cell lines were expressing near physiological levels of A3F, A3G and likely also the other A3 proteins.

Successful replication of an L1 element requires that it be transcribed, transported to the cytosol, assembled into a ribonucleoprotein complex, returned to the nucleus, and

reverse-transcribed and integrated[124]. To monitor L1 retrotransposition, we made use of a plasmid-based L1 element harboring a reporter gene (GFP) in its 3' UTR (**Fig. 2-2**B). The GFP gene is transcribed from its own promoter and is in the opposite orientation as the L1 genome. The GFP gene is disrupted by an intron that is in the same orientation as the L1 element. Thus, functional GFP protein cannot be expressed from the plasmid. However if the L1 element is transcribed, spliced, reverse transcribed, and integrated, then functional GFP protein can be expressed from the newly integrated, now intronless GFP gene. This provides a system for monitoring retrotransposition kinetics [**Fig. 2-2**B; [107,125]].

To assess whether A3B was able to inhibit L1 retrotransposition, the GFP-based L1 reporter construct was transfected into several independent 293 clones stably expressing a control vector, A3B, A3D, A3F, or A3G (between 3 and 5 clones for each A3 were transfected). These analyses focused on these A3s because they are the double A3 proteins, and hence are overall similar to each other, and because one of them (A3G) was expected to function as a negative control[166]. The steady accumulation of GFP-positive cells (around 1% increase in GFP+ cells per day) indicated that similar retrotransposition rates occurred in 293 cells expressing the control vector, A3D and A3G (**Fig. 2-2**C). In contrast, expression of A3B decreased the rate of L1 retrotransposition by 5- to 10-fold. Additionally, contrary to expectations based on our localization studies, expression of A3F also provided a strong barrier to L1 retrotransposition.

As an additional control for A3 protein function, we monitored the infectivity of an HIV-based retrovirus produced in the presence of each of the A3-HA constructs or in the presence of an empty control vector (**Fig. 2-2**D). As previously reported by several

laboratories, the infectivity of Vif-deficient HIV was decreased strongly by A3G, and A3F and weakly by A3B [*e.g.*, [8,87,176,182]]. Expression of A3D had little effect. These controls are important because they show that the L1 inhibition data are not simply phenocopies of the HIV restriction activities of the human A3 proteins. Moreover, these data strongly suggest that other trivial explanations for the L1 phenotype (such as non-specific nucleic acid binding) are not likely. For instance, A3G potently restricts many LTR-dependent retroviruses (**Fig. 2-2**D and introductory references) and it binds single strand nucleic acid[68,116], but it does little to the rates of L1 retrotranspostion (**Fig. 2-2**C).

To begin to dissect the mechanism of L1 inhibition, we isolated a DNA cytidine deaminase-deficient A3B mutant, $A3B_{W228L,\ D316N}$. Several previous studies have demonstrated that expression of human A3 proteins in *E. coli* causes a mutator phenotype, thereby providing a measure of their DNA cytidine deaminase activity (*e.g.*, [56,60]). *E. coli* expressing A3B and A3G showed, respectively, 58- and 23-fold increases in the frequency of mutation to rifampicin-resistance (Rif^R; **Fig. 2-3**A). In comparison, control vector- and A3B $_{W228L,\ D316N}$ -transformed cells showed much lower and nearly identical levels of Rif^R mutation, which were attributable to spontaneous mutations that occur during *E. coli* growth. The *E. coli* studies therefore demonstrated that A3B $_{W228L,\ D316N}$ is not able to catalyze DNA cytidine deamination and that catalytically active A3B is a potent DNA mutator.

To address whether deaminase activity is required for L1 inhibition, retrotransposition assays were done using 293 clones stably expressing a vector control or A3B $_{W228L,\ D316N}$. Interestingly, the catalytically inactive A3B protein also caused 5- to 10-fold reductions in the rate of L1 retrotransposition (**Fig. 2-3**B). These data indicated

that the L1 inhibition activity of A3B is independent of deaminase activity. In further support of such a mechanism, retrotransposed L1 DNA showed no signs of the hallmark, strand-specific C/G to T/A transition mutations that are commonly found in A3-restricted LTR-type retroelements. Between 7 and 22 retrotransposed L1-GFP substrates were sequenced for each experimental condition, and only 3 base substitutions were found in 26,640 bp sequences. In contrast, HIV experiments with A3F and A3G yielded 8 and 21 C to T transitions per 1000 bp of viral DNA analyzed[56]. To further elucidate the mechanism of inhibition, we recovered genomic DNA from cells at the end of the retrotransposition experiments and used PCR to quantify integrated L1 DNA (see experimental procedures). These experiments repeatedly produced less intense or barely visible retrotransposition bands from the genomic DNA samples of cells expressing A3B or A3F (**Fig. 2-3**C). In many instances, an additional semi-nested PCR was required to visualize retrotransposed L1 DNA (data not shown). These data combined to strongly indicate that A3B and A3F interfere with L1 retrotransposition prior to integration by a process that does not require DNA cytidine deamination.

Discussion

Initially, we were surprised that the L1 inhibition activity of A3B and A3F failed to correlate with obvious subcellular localization. However, the similar inhibitory potentials of A3B and A3F suggested that this activity might derive from a common domain. Indeed, these proteins are 96% identical between residues 66-190 and 65-189, respectively, whereas the remainder of these polypeptides shares less than 57% identity. It is likely that this region of concentrated identity mediates L1 inhibition. This is particularly interesting because the corresponding region of A3G is less than 50% identical, and it is required for associating with HIV-1 Gag and thereby mediating its specific encapsidation[116]. By analogy, this region of A3B and A3F may mediate L1 inhibition by associating with the L1 ORF2 protein, which, like the nucleocapsid region of HIV Gag, binds both zinc and nucleic acid[124]. This ORF2 region is conserved amongst L1 and related elements. Thus, an A3-ORF2 association may also interfere with the mobilization of related autonomous retroelements and with some non-autonomous ones, such as Alu, because they require ORF2 for transposition.

The precise mechanism by which A3s limit LINE-1 retrotransposition remains unclear, and additional studies could help answer some key remaining questions. For instance, although mutated LINE-1 DNA was not detected in cells expressing A3B or A3F, it is possible that such DNA could be detected using a sensitive PCR technique that selectively amplifies hypermutated DNA[160]. It would also be possible to determine *where* in the cell inhibition occurs by using A3B and A3F proteins that are more definitely restricted to particular subcellular compartments. For instance, experiments using A3B fused to a nuclear export signal or a protein that is retained in the cytoplasm could

determine whether A3B normally acts in the nucleus. Experiments using transgenic mice harboring a marked human LINE-1 element could be very informative[123]. Measuring the retrotransposition rate in these mice (and in mice crossed with transgenic mice expressing human A3s) would help demonstrate the *in vivo* relevance of A3-mediated LINE-1 inhibition.

The primate A3 genes bear strong signatures of positive selection (defined as a high ratio non-synonymous to synonymous mutations)[142]. Such selection is often evident in genes involved in host-pathogen interactions. This positive selection and the A3 gene expansion in the primate lineage may indicate that primate A3s are able to engage a variety of retroelements, including L1, as the data presented here suggests. This affords a possible explanation for why L1 retrotransposition levels are much higher in germ cells of mice than in those of humans[34]. Indeed, sequencing of the mouse and human genome has revealed a profound difference in the mobility of L1 retrotransposons – and transposable elements in general – between mice and humans[81,172]. It is possible to estimate the age of retrotransposon insertions by comparing their sequence to a consensus sequence (newer insertions have accumulated fewer mutations). In the rodent lineage, transposable elements appear to be as active today as they were 100 million years ago[81,172]. In contrast, in the lineage leading to humans, transposable elements have been declining in mobility over the last 100 million years[81,172]. This decline corresponds to an increase in A3 gene number (rodent genomes only encode a single A3 gene; the human genome encodes 7). Obviously, many factors could account for this difference in addition to the anti-retrotransposon activities of A3 proteins. Nevertheless, there is a striking inverse correlation between transposable element mobility and A3 gene number.

Going forward, it will be particularly interesting to assess A3 function in the germ lineage, for it is only retrotransposon insertions in these cells that have a chance of being inherited and increasing in frequency in the population.

Experimental Procedures

DNA constructs. The L1 construct, pL1-ac002980-EGFP[14], and the functionally active A3B construct, pcDNA3.1-A3B-HA[182], have been described. The deaminase-deficient A3B construct has two amino acid substitutions, W228L and D316N, and it was isolated by PCR using oligonucleotides 5'-GAG CTC GGT ACC ACC ATG AAT CCA CAG ATC AGA AAT and 5'-GTC GAC CAT CCT TCC GTT TCC CTG ATT CTG GAG and IMAGE clones 3915193 and 4707934 as amplification templates. A3D, A3F and A3G were amplified from existing laboratory constructs[58,60,87] using oligonucleotides A3D: 5'-GAG CTC GGT ACC ACC ATG AAT CCA CAG ATC AGA AAT/5'-GTC GAC TCC CTG GAG AAT CTC CCG TAG C, A3F: 5'-GAG CTC AGG TAC CAC CAT GAA GCC TCA CTT CAG AAA C/5'- GTC GAC TCC CTC GAG AAT CTC CTG CAG CTT, and A3G: 5'-GAG CTC AGG TAC CAC CAT GAA GCC TCA CTT CAG AAA C/5'-GTC GAC TCC GTT TTC CTG ATT CTG GAG AAT. The A3B, A3D, A3F and A3G PCR products were digested with SacI and SalI and cloned into a similarly digested pEGFP-N3 (Clontech). They were also digested with KpnI and SalI and cloned into a modified version of pcDNA3.1(+) (Invitrogen) containing a carboxy-terminal 3x influenza hemagglutinin (HA) tag. This modified pcDNA3.1 vector was created by cloning a triple HA tag between the XhoI and XbaI sites in the polylinker. The *E. coli* expression plasmids, pTrc99A and pTrc99A-A3G were reported previously[56,60]. The A3B expressing derivatives of pTrc99A were constructed by PCR using oligonucleotides 5'-GGT ACC ACC ATG AAT CCA CAG ATC AGA AAT and 5'-GTC GAC CCC ATC CTT CAG TTT CCC TGA TTC TGG AG, KpnI and SalI restriction and ligation into similarly digested vector. All constructs were verified by

DNA sequencing. The cDNA sequences for A3F and A3G were identical to NCBI Reference Sequences. The cDNA sequence for A3D is also called A3D-E, and it is represented by GenBank cDNA BE888971. Active A3B differed in eight positions from GenBank NP_004891 (K62E, L80P, F107L, T146K, M193V, D205G, T337A and R372K). The limited number of A3B sequences in GenBank prevented us from determining whether these variations are naturally occurring.

Cell Lines. Human embryonic kidney (HEK) 293 cells and HeLa cells were maintained in DMEM supplemented with 10% FBS and 25 U/ml penicillin and 25 ug/ml streptomycin (all Gibco) at 37˚C and 5% CO_2. Transfections were performed using FuGene 6 according to the manufacturer's protocol (Roche). Stable, A3-expressing 293 cell lines were established by transfecting with linearized A3-HA expression constructs, followed by a selection with growth media containing 880 ug/ml G418 (Roche). Clones expressing HA-tagged A3 proteins were identified and quantified by anti-HA (Covance) immunoblotting. Maintaining all clones in growth media supplemented with 220 ug/ml G418 ensured stable A3 protein expression.

Microscopy. 15,000 HeLa or 50,000 293T cells were seeded on LabTek chambered coverglasses (Nunc). After 24 hrs of incubation, these cells were transfected with 250 ng of the pEGFP-N3-based DNA constructs. After an additional 24 hrs of incubation, images of the live cells were collected on a Zeiss Axiovert 200 microscope. 293T cells transfected with the A3-GFP constructs were fixed in 4% formaldehyde, permeablized in 0.2% TritonX-100, stained with 100 ng/ml DAPI and imaged as above. The A3B $_{W228L\ D316N}$ construct was used to acquire the images presented in Fig. 1. Additional

experiments using other A3B constructs have also shown a predominantly nuclear localization (see chapter 4).

L1 Retrotransposition Assays. L1 assays were performed as originally described with minor modifications[125]. For each condition, 3-5 independent A3 or control 293 clones were plated at a density of $3x10^5$ cells per well in 6 well plates. After 24 hrs of incubation, the cells were transfected with 500 ng of pL1-ac002980-EGFP. After an additional 24 hrs of incubation, L1 plasmid-containing cells were selected with 0.67 μg/ml puromycin (US Biological). Maintenance of the L1 substrate was ensured by the continual presence of puromycin (0.15 μg/ml) through the duration of the experiment. At 72 hr intervals, the cells were passaged and at least 25,000 cells were examined by flow cytometry. A GFP-positive region was established such that less than 0.01% of negative-control 293 cells profiled within this gate.

Immunoblotting. Total cellular protein extracts were prepared using a serial freeze/thaw procedure. The extracted proteins were quantified by Bradford assays (BioRad) and equal amounts were fractionated by SDS-PAGE, transferred to a PVDF membrane (Millipore) and probed with monoclonal anti-HA (Covance), monoclonal anti-alpha-tubulin (BioRad) or polyclonal anti-A3G [119]. Primary antibodies were detected by incubation with horseradish peroxidase-coupled anti-mouse IgG (BioRad) or anti-rabbit IgG (BioRad), followed by chemi-luminescence (Roche).

L1 DNA analyses. To ensure the exclusive amplification and detection of L1 retrotransposition events, we designed a PCR-based assay that would specifically amplify intronless GFP DNA. Most importantly, a splice-specific oligonucleotide, 5'-CAG CGT

GCA GCT GGC C was designed to span the intron by annealing to exonic sequences 10 bp immediately upstream and 6 bp downstream. The GFP gene-specific annealing of this oligonucleotide would only occur if the intron were spliced out, reverse transcribed and integrated. This primer was used in combination with 5'-CGA TCC CCT CAG AAG AAC TCG or 5'-CGA TCC CCT CAG AAG AAC TCG to generate a larger 360 bp or a smaller (semi-nested) 240 bp PCR product.

E. coli mutation assays. The intrinsic DNA cytidine deaminase activity of several A3 proteins including A3G has been assayed by quantifying the accumulation of rifampicin-resistance mutations in *E. coli* expressing these proteins [*e.g.*, [56,60]].

Figures

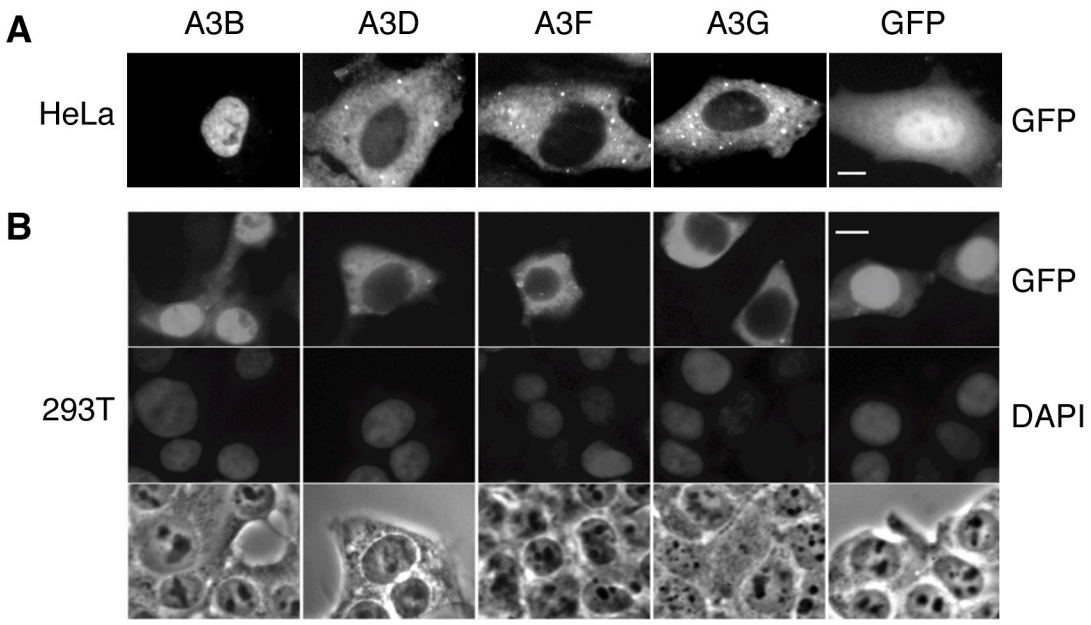

Fig. 2-1. Subcellular distribution of human A3B, A3D, A3F and A3G. (A) Representative live HeLa cells showing the subcellular localization of GFP-tagged human A3B, A3D, A3F and A3G. **(B)** Representative fixed 293T cells showing the subcellular localization of GFP-tagged human A3B, A3D, A3F and A3G (top panel). The middle and bottom panels show DAPI-stained and bright field images of the same fields, respectively. The scales indicate 10 μm. (See also **Fig. 4-1**).

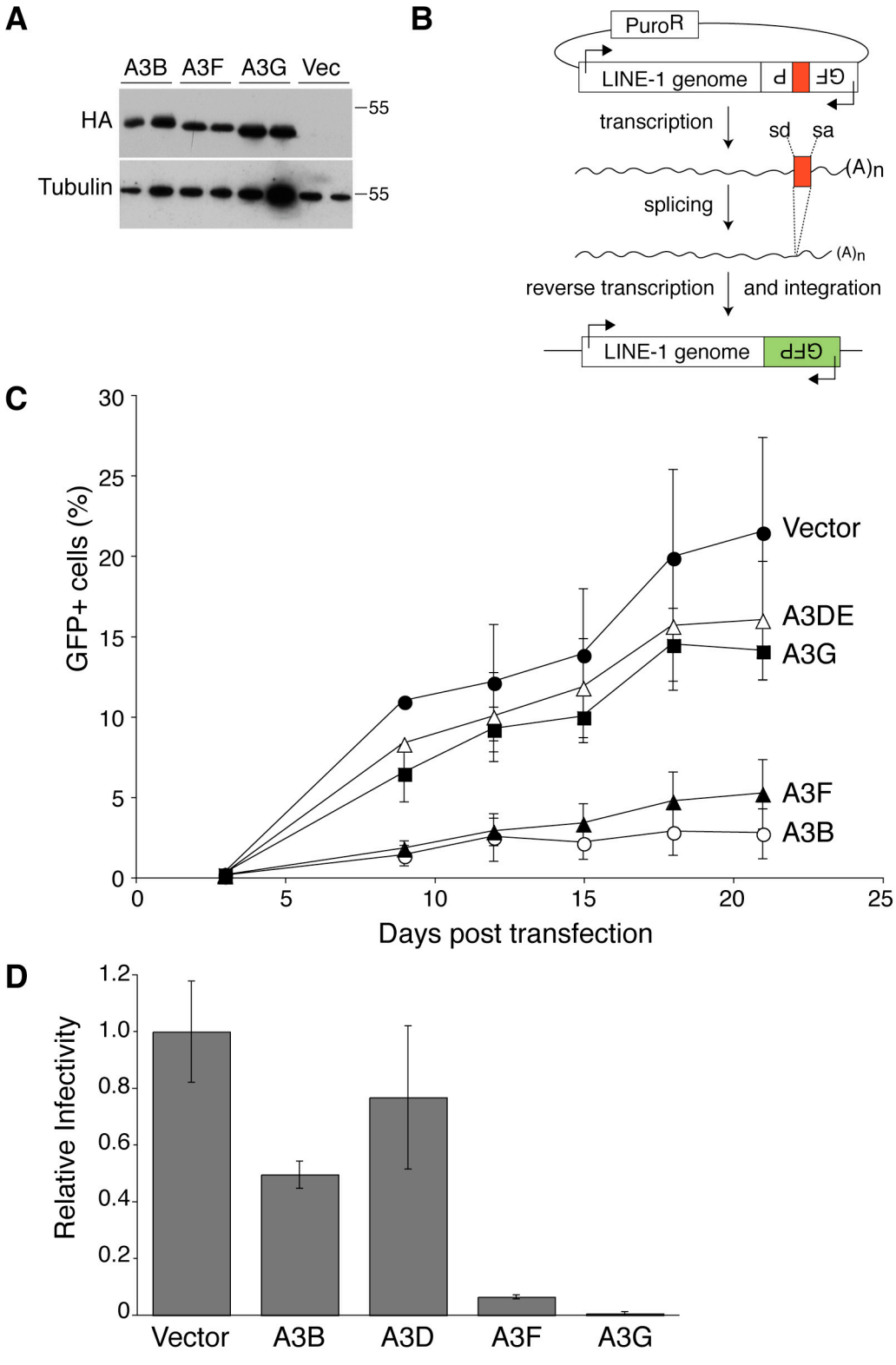
A
A3B
A3F
A3G
Vec
HA
Tubulin
55
55
B
PuroR
LINE-1 genome
transcription
sd
sa
(A)n
splicing
reverse transcription
and integration
GFP
C
GFP+ cells (%)
Days post transfection
Vector
A3DE
A3G
A3F
A3B
D
Relative Infectivity
Vector
A3B
A3D
A3F
A3G

Fig. 2-2. Expression of A3B or A3F inhibits L1 retrotransposition. (A) anti-HA and anti-alpha-tubulin immunoblots showing expression of HA-tagged A3 proteins in representative stable 293 cell lines. The A3B clones shown express the catalytically inactive A3B $_{W228L,\ D316N}$ construct. The active A3B construct contains a single HA epitope[182] and, as expected, its immunoblot intensity was approximately one-third that of the triple HA-tagged proteins (data not shown). **(B)** a schematic of the GFP-based L1 retrotransposition system[125]. A complete L1 genome contained in the plasmid harbors an antisense, intron-interrupted GFP cassette in its 3' UTR (intron colored red). L1 transcription, splicing, reverse transcription and integration are necessary for GFP expression (green color indicates non-interrupted GFP gene) (sd: splice donor; sa: splice acceptor). GFP fluorescence as monitored by flow cytometry provides a quantitative measure of L1 activity. **(C)** A3B and A3F inhibit L1 retrotransposition. The accumulation of GFP-positive cells over time is shown. For each dataset, the mean and SEM of 3-5 independently derived 293 clones, which stably express the control vector or a human A3 protein, are indicated. **(D)** Single cycle HIV infectivity assays. HIV-GFP viruses were produced in the presence of the control or the indicated A3-HA constructs and used to infect non-fluorescent 293T target cells, as described previously[56,87]. After 48 hours of incubation, target cell GFP fluorescence was monitored by flow cytometry, providing a quantitative measure of a single round of infectivity. Data are presented relative to control, and the mean of three parallel experiments is indicated.

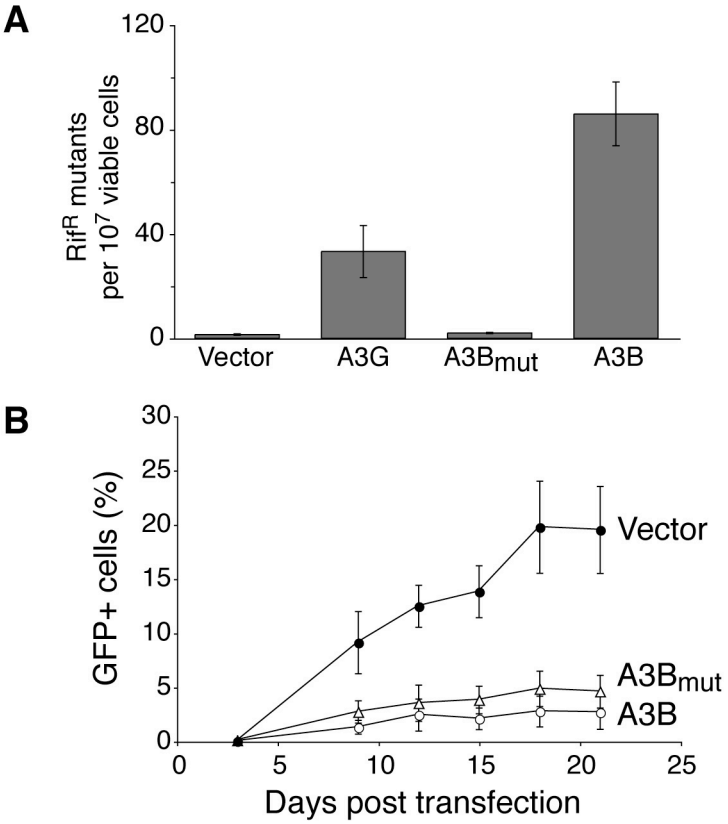
A
120
80
40
0
Rif^R mutants per 10^7 viable cells
Vector
A3G
A3B_mut
A3B
B
30
25
20
15
10
5
0
GFP+ cells (%)
Vector
A3B_mut
A3B
0
5
10
15
20
25
Days post transfection
C
Vector
A3B_mut
A3B
A3F
A3G
293
H2O
Control
Integrated L1

Fig. 2-3. The inhibition of L1 retrotransposition does not require DNA cytidine deaminase activity. (A) The DNA cytidine deaminase activity of A3B, $A3B_{W228L,\ D316N}$, and A3G was quantified using an *E. coli*-based mutation assay (see experimental procedures). Empty-vector transformed bacteria serve as a negative control and indicate the background mutation frequency. **(B)** L1 inhibition by catalytically active and inactive A3B proteins. $A3B_{mut}$ denotes the catalytically inactive A3B $_{W228L,\ D316N}$ construct. This experiment was performed as described in **Fig. 2-2** and, for comparison, the A3B data from **Fig. 2-2**B are shown (dotted line). **(C)** an ethidium bromide stained agarose gel showing less integrated L1 DNA in cells expressing A3B or A3F (a 360 bp specific band). A non-specific higher molecular weight band (ca. 950 bp) indicates that similar amounts of genomic DNA were used in each PCR reaction. See the experimental procedures for additional details.

Chapter 3: APOBEC3 Proteins Restrict Foreign DNA

The experiments described in this chapter show that APOBEC3 proteins restrict foreign intracellular DNA. This expands the range of biologically relevant APOBEC substrates and has implications for therapeutic processes involving DNA transfer such as gene therapy and for standard molecular biology procedures such as transfection.

This chapter is adapted from a manuscript submitted for publication in June 2009 that is awaiting peer review

Chapter Summary

Intracellular foreign DNA is inherently dangerous and is often associated with viral or bacterial infection. Innate immune sensors detect intracellular foreign DNA, but little is known about mechanisms that respond to clear it. Here, we show that cytidine deaminase A3A is a key effector in one such mechanism. In primary cells, A3A is strongly induced by interferon following DNA detection. A3A triggers degradation of transfected foreign DNA by editing cytidines to uridines. These noncanonical DNA nucleosides are converted into destabilizing abasic sites by uracil DNA glycosylase. This mechanism is evident in cell lines and primary monocytes, where up to 97% of transfected DNA cytidines show signs of editing. These findings expand the notion of what is a biologically relevant A3 substrate, offer insights into innate immune responses to infection, and may impact applied processes such as genetic engineering and gene therapy.

Introduction

Intracellular foreign DNA threatens the well-being and proper function of a cell. If the DNA is pathogenic in origin, it could encode cytotoxic gene products or it could be replicated to propagate an infection. Even non-pathogenic DNA, for instance DNA from necrotic cells internalized by a phagocyte, threatens to insert into the genome or disrupt the cell's gene expression equilibrium. Foreign DNA is sensed by Toll-like receptor (TLR)-dependent and TLR-independent mechanisms[100,167]. TLR9 senses DNA in endosomal compartments and signals to induce the production of type 1 interferons and pro-inflammatory cytokines and chemokines. The identity of the TLR-independent cytosolic DNA sensor is unknown, but this pathway also results in robust interferon (IFN) induction[65,121,156]. IFN in turn stimulates the transcription of many genes whose products orchestrate a wide variety of innate immune responses[138]. For instance, Tetherin blocks viral budding, PKR inhibits translation, RNaseL degrades intracellular RNA, and ADAR1 deaminates double-strand RNA adenosines to inosines[3,75,104,117,134,168]. However, little is known about cellular proteins that mediate the clearance of foreign intracellular DNA.

The enzymatic conversion of DNA cytidines to uridines has at least two fundamental roles in immunity. Activation-induced deaminase (AID) deaminates antibody gene DNA and triggers antibody gene diversification by somatic hypermutation and class switch recombination[37]. The APOBEC3s (A3s): A3A, A3B, A3C, A3DE, A3F, A3G and A3H, are DNA cytidine deaminases that have been shown to inhibit the replication of a diverse set of retroviruses and retrotransposons (see chapter 1 and reviews [6,24,32,54,63,128]). A3F and A3G, the best-studied members of this family, strongly

inhibit the replication of Vif-defective HIV-1 by deaminating viral cDNA during retrovirus reverse transcription. In contrast, A3A does not restrict HIV-1, but it has been shown to limit retrotransposition of L1 and Alu elements and replication of adeno-associated virus (AAV) [8,10,11,18,26,108,113]. Other DNA viruses such as hepatitis B virus (HBV) and human papillomavirus (HPV) may also be targets of A3 proteins[159,169].

The fact that human cells have innate mechanisms to sense foreign DNA and the fact that A3s are potent DNA deaminases with established roles in immunity led us to hypothesize that one or more of these proteins acts downstream of DNA-sensing molecules to mediate the clearance of foreign DNA. This hypothesis was also prompted by two observations. First, we had noted that transient expression of A3A seemed to decrease expression of co-transfected reporter genes. Second, we had preliminary data that suggested that A3A was upregulated in DNA-stimulated primary leukocytes.

Here, we show that A3A has all of the hallmarks of a foreign DNA restriction factor. We find that A3A is predominantly expressed in phagocytic cells, is strongly induced following DNA detection, and triggers the degradation of foreign DNA by a cytidine deamination and uracil excision mechanism.

Results

A3A is expressed in phagocytes and strongly induced by IFN and CpG DNA

We reasoned that if A3s were indeed DNA restriction factors, then they ought to be expressed in cells frequently exposed to foreign DNA and would likely be induced by foreign DNA detection. We focused on A3A, because of reports suggesting that A3A is expressed predominantly in monocytes, macrophages, and neutrophils[18,126]. These highly phagocytic cells ingest large quantities of foreign materials, including nucleic acids. To confirm and extend these studies, we obtained fresh human peripheral blood mononuclear cells (PBMCs), purified CD14-positive cells (monocytes and macrophages) by magnetic cell sorting, and used a panel of specific PCR assays to quantify A3 expression (**Fig. 3-1** & **Fig. 3-2**). Several A3s were expressed in PBMCs, with A3A, A3C and A3G consistently showing the highest basal mRNA levels (**Fig. 3-1**B). The relative A3A expression level increased 3-fold in the CD14-enriched cell population and decreased proportionately in the CD14-depleted fraction, whereas A3C and A3G mRNA levels were similar in the different populations (**Fig. 3-1**B). These observations indicated that A3A is expressed predominantly in CD14-positive phagocytic cells.

Although prior reports have indicated that several A3s can be induced by interferon (IFN)[12,19,126,127,140,157,163,171], a systematic and quantitative analysis of the entire repertoire had not been done. Therefore, fresh PBMCs were treated with IFN, and RNA was isolated at multiple time points for quantification by real-time PCR. A3A was strongly IFN-responsive with mRNA levels consistently peaking 100-fold above those in untreated cells (**Fig. 3-1**C). It is unclear whether this results from an increase in transcription or in mRNA stability. Other A3s showed modest 2- to 3-fold increases.

Moreover, we found that incubating PBMCs with the CpG-containing ssDNA oligonucleotide (known as CpG DNA, these are potent TLR9 ligands) caused a 200-fold increase in A3A, and a 2- to 4-fold increase in the expression of other A3s (**Fig. 3-1**D**,** E). The dramatic induction of A3A by CpG DNA, but not GpC control DNA, was also visible by immunoblotting using an A3A-specific polyclonal antibody (**Fig. 3-1**F). We also compared the degree of A3A induction to other well-characterized IFN-stimulated genes[138]. A3A mRNA was induced slightly less than ISG15 (one of the most IFN-stimulated genes[138]) but significantly higher than IFNα2, IFNβ, PKR, and ADAR1 (**Fig. 3-1**E). As a control, we measured the mRNA levels of RNaseL, which is activated post-transcriptionally in response to IFN signaling. Accordingly, RNaseL mRNA abundance did not increase upon stimulation. The relative mRNA levels of 3 "housekeeping" genes: TBP, RPL13A, and HPRT, varied little under each experimental condition and they were therefore used as internal normalization standards.

A3A inhibits stable gene transfer

The finding that A3A is expressed in cells that ingest pathogens and extracellular debris, and the fact that its expression is induced by DNA detection suggests that it might play a role in the clearance of foreign DNA. To directly test this hypothesis, we quantified the impact of A3A expression on several measures of foreign DNA integrity and stability. First, we measured the impact of A3A on the stable gene transfer efficiency (**Fig. 3-3**). HEK-293 cells, which lack detectable A3A mRNA, were co-transfected with a neomycin-resistance (Neo^R) reporter plasmid and an expression construct for A3A, A3A-E72A (a catalytically inactive mutant; (**Fig. 3-4**), and[18]), or a GFP control. Two days were allowed for integration of the reporter plasmid into chromosomal DNA and for

expression of the Neo^R gene product. The cells were then plated into neomycin (G418)-containing medium, and drug resistant colonies were allowed to grow. The resulting Neo^R colonies were fixed, stained, scanned, and counted. A3A expression resulted in a dose-dependent decrease in the frequency of Neo^R colonies in comparison to either control condition (**Fig. 3-3**B, C). This effect was potent: transfecting as little as 16 ng of A3A expression plasmid decreased gene transfer efficiency to 40% of controls (**Fig. 3-3**C). A3A expression did not decrease the number of colony-forming cells (assessed by plating in drug-free medium), nor was there any indication of increased apoptosis or decreased proliferation in A3A-expressing cells (**Fig. 3-5**). Essentially identical data were obtained with HeLa cells, and using a more efficient DNA transposon-mediated gene transfer system (**Fig. 3-6**).

A3A glutamate 72 is part of a conserved zinc-coordinating catalytic motif, H-x-E-x_{24}-C-P-x_{2-4}-C (where x is any amino acid)[27,83,114]. The fact that A3A-E72A-GFP expression did not decrease the efficiency of stable gene transfer suggested that this effect requires A3A's cytidine deaminase activity (addressed below).

A3A inhibits transient gene expression

As a second measure of foreign DNA stability, we tested whether transient gene expression would be affected by A3A. GFP-encoding plasmid DNA was delivered to HEK-293 cells and GFP fluorescence was monitored by flow cytometry at 2 and 5 days post-transfection (**Fig. 3-3**). After two days, the higher A3A expression levels caused a small decrease in GFP fluorescence. By five days, all A3A levels caused significant reductions in GFP fluorescence, and a clear dose response was apparent. The rate of GFP fluorescence decay was also accelerated in A3A-expressing cells compared to control

cells (~4% decline in GFP-positive cells per day versus 2% in the control cells). Thus, A3A expression caused a more rapid disappearance of transient reporter gene expression.

A3A destabilizes transfected plasmid DNA

Because reduced gene transfer efficiency and diminished reporter gene expression could both result from a reduction in the quantity or integrity of the transfected DNA itself, we used quantitative PCR to directly measure plasmid DNA levels in A3A expressing cells. These experiments revealed a 60% reduction in plasmid DNA levels in comparison to control cells two days after transfection (**Fig. 3-3**E). It is noteworthy that this is likely an underestimate of the potency of A3A, because transfected DNA may resist DNase digestion (done prior to cell lysis to remove free DNA) or become sequestered in A3A-impermeable compartments (*e.g.*, endosomes). Nevertheless, these data showed that A3A can directly compromise the integrity of foreign plasmid DNA, and that this effect requires the conserved catalytic glutamate.

A3A triggers the degradation of foreign DNA by a mechanism that requires uracil excision and yields C/G-to-T/A hypermutations

If A3A indeed triggers the clearance of foreign DNA by a mechanism involving DNA deamination, then, in addition to a requirement for catalytic activity, two major predictions follow. First, the immediate products of deamination, DNA uridines, might be substrates for excision by cellular uracil DNA glycosylases. The most obvious candidate is the major cellular uracil DNA glycosylase, UNG2, which functions normally in error-free excision repair of genomic uracils but also processes uracils during error-prone

antibody gene diversification events[37,74]. Second, since DNA uridines base-pair like thymidine and template the insertion of adenosines, they ought to amplify by PCR and result in C/G-to-T/A transition mutations. Furthermore, if A3A was causing these mutations, then the mutated cytidines should preferentially occur within 5'-TC dinucleotides, since A3A has been shown to exhibit this local preference *in vitro*[18].

To test these predictions, we used a uracil DNA glycosylase inhibitor (UGI) to inhibit UNG2[36], and a PCR-based technique called differential DNA denaturation PCR (3D-PCR)[159] that uses lower than normal PCR denaturation temperatures. We did this because if UNG2 processes A3A-generated uridines, then inhibiting it should cause uridine-containing intermediates to accumulate. 3D-PCR is based on the idea that DNA with relatively fewer inter-strand hydrogen bonds (*e.g.*, duplex DNA with G-U mismatches or A/T rich DNA) will denature at lower temperatures and thus will selectively amplify at lower PCR denaturation temperatures. Therefore, 3D-PCR combined with UNG2 inhibition allows for the detection of potentially short-lived uridine-containing DNA intermediates.

UGI-expressing HEK-293T cells were co-transfected with a reporter plasmid and A3A or A3A-E72A expression plasmids. Two days later, transfected DNA was recovered from the cells and analyzed by 3D-PCR. At the highest denaturation temperatures, plasmid DNA recovered from A3A and A3A-E72A-expressing cells amplified by PCR (**Fig. 3-7**A). However, at lower denaturation temperatures, only plasmid DNA recovered from A3A-expressing cells amplified. Cells not expressing UGI were similary transfected and amplification at lower denaturation temperatures was not

detected. These results suggested that A3A could be editing transfected plasmid DNA and that UNG2 might normally be processing the edited molecules.

To confirm that UNG2 was acting on uridine-containing DNA molecules generated by A3A, we transfected HEK-293 cells with a reporter plasmid, A3A or A3A-E72A expression plasmid, and increasing amounts of a UGI expression plasmid. In support of the prior experiments, higher doses of UGI (greater UNG2 inhibition) led to a greater abundance of lower denaturation temperature PCR products (**Fig. 3-7**B). Moreover, UNG2 inhibition alone was not sufficient to enable the recovery of low denaturation temperature PCR amplicons (expression of catalytically active A3A was also required).

To confirm that the low denaturation temperature amplification corresponded to A3A-edited molecules, we cloned and sequenced a number of PCR products (**Fig. 3-7**C). These analyses revealed high levels of C/G-to-T/A transition mutations in molecules recovered from A3A-expressing cells. In contrast, no mutations were found in molecules recovered from A3A-E72A-expressing cells. Mutations were even apparent in the majority of molecules amplified at a non-selective denaturation temperature (94.4˚C), indicating that selective amplification by 3D-PCR was not required to detect edited sequences. The edited cytidines were preferentially within 5'-T<u>C</u> dinucleotides (15/15 TC cytidines were edited in at least one sequence, but only 14/78 of cytidines in AC, CC, or GC contexts were edited in at least one sequence). 5'-TC dinucleotides have previously shown to be the preferred A3A deamination target *in vitro*[18]. Overall, these data demonstrated that A3A deaminates foreign DNA in human cells and that the resulting uracils are substrates for UNG2.

Extensive foreign DNA hypermutation occurs in primary human monocytes

We next wanted to assess the impact of foreign DNA editing in A3A's normal physiological context, *i.e.* in primary human phagocytes. We took advantage of the fact that a two log difference in A3A expression exists between CpG-treated and mock-treated PBMCs (**Fig. 3-1**D, E, F). These two cell populations were prepared and transfected with GFP-encoding plasmid DNA or mock-transfected with buffer alone. We found that the introduction of plasmid DNA into cells caused A3A to be induced, regardless of whether the cells had been pre-treated with CpG DNA (**Fig. 3-8**A). These data are consistent with prior studies showing the introduction of plasmid DNA into leukocytes induces a robust TLR-independent IFN-response[65,156].

Next, to determine whether endogenous A3A mutates transfected plasmid DNA, we recovered DNA from cells at 8 and 24 hours post-transfection and used 3D-PCR to amplify a region of the GFP gene. For a control template, we mixed GFP plasmid DNA with DNA recovered from mock-transfected cells. Using this control template, products amplified only at the highest denaturation temperatures (**Fig. 3-8**B, "pDNA"). In contrast, we were able to detect plasmid DNA amplification at lower denaturation temperatures using the DNA recovered from transfected monocytes. Cloning and sequencing these PCR products confirmed that they resulted from amplification of extensively edited plasmid DNA (**Fig. 3-8**C), with some molecules having as many as 72/93 cytidines edited in the 257 nucleotide amplicon. We also noted that amplification occurred at lower denaturation temperatures using DNA recovered from the CpG-pretreated cells than from the non-pretreated cells at 8 hours post-transfection. This suggested that the plasmid DNA had been more extensively mutated in the cells that were expressing higher levels

of A3A at the time of transfection. At 24 hours post-transfection, we detected weak bands at the lowest denaturation temperatures in the mock-pretreated cells, likely corresponding to mutations catalyzed by A3A that had been induced by plasmid DNA transfection itself (**Fig. 3-8**B).

In addition to cloning and sequencing individual molecules, we also directly sequenced populations of PCR products (**Fig. 3-9**). By analyzing these sequences, we found evidence for C-to-U conversion of 70/72 cytidines (97%) within a 200 nucleotide region of the GFP gene (*i.e.,* that these cytidines had been edited in at least some of the molecules amplified; **Fig. 3-9**). We were also able to confirm a strong bias towards deamination at 5'-TC dinucleotides, consistent with catalysis by A3A (**Fig. 3-8**D**, Fig. 3-9**). This approach also detected evidence for C/G-to-T/A mutation in the DNA populations amplified at the standard PCR denaturation temperature of 95˚C (**Fig. 3-9**). Thus, even without the selective amplification of 3D-PCR, a detectable fraction of the DNA transfected into monocytes was mutated by endogenous A3A (and/or by one of the other A3s expressed in monocytes). We also found that, in contrast to experiments in HEK-293 cells, we were able to detect hyper-edited DNA molecules recovered from primary monocytes without inhibiting UNG2. This suggested that UNG2 activity is greater in HEK-293 cells, that transfected A3A inactivates its own expression plasmid (attenuating its efficacy in cotransfection experiments), and/or that editing is more extensive in primary monocytes (consistent with the dramatic induction of A3A). Essentially identical results were obtained using PBMCs isolated from three independent donors and an additional set of PCR primers (**Fig. 3-10**).

These results demonstrated that endogenous A3A hypermutates foreign DNA in primary cells. The extensive levels of mutation suggested that foreign DNA becomes genetically inert prior to eventual degradation. Indeed, A3A-mediated foreign DNA clearance is possibly one of the causes underlying the extremely short-lived GFP expression evident in transfected monocytes, which decreased from ~5% GFP-positive at 8 hours to 0.5% by 24 hours in the CpG-pretreated transfected cells (**Fig. 3-8**E).

Discussion

The recognition and clearance of foreign intracellular DNA has been postulated to be a fundamental arm of the innate immune response[64,65,156]. Although TLR-dependent and -independent pathways sense foreign DNA and trigger IFN production, little is known about effector proteins downstream of IFN signalling, some of which presumably must function to mediate DNA clearance. Here, we show that A3A is such a foreign DNA restriction factor. A3A is expressed predominantly in CD14-positive phagocytic cells and it is strongly IFN-inducible, consistent with a front-line role in foreign DNA restriction. A3A is also induced by CpG single-strand DNA or double-strand DNA, presumably through the induction of IFN by DNA detection pathways. A3A reduces foreign DNA stability and integrity by three measures: stable gene transfer, transient gene expression, and DNA persistence. The extraordinary levels of foreign DNA C/G-to-T/A hypermutation, a requirement for the A3A active site glutamate, and the influence of the uracil DNA glycosylase inhibitor UGI combine to demonstrate that foreign DNA restriction is a deamination and uracil-excision dependent process.

Our data indicate a model in which A3A deaminates foreign DNA cytidines to uridines and that these lesions are subsequently converted by the cellular uracil DNA glycosylase UNG2 to nuclease-susceptible abasic sites (**Fig. 3-11**). It is likely that these abasic sites are processed by the endonuclease APEX, which cleaves the DNA phosphodiester backbone adjacent to abasic sites[74], but many other cellular nucleases could also be involved. We are additionally intrigued by the apparent strand specificity of the hypermutations, and we are investigating its basis.

Although the studies described in this chapter focus on A3A, humans have six other A3 proteins. All of the A3 proteins possess DNA deaminase activity, and the other A3s have broader expression profiles than that of A3A[18,60,70,108]. It may be that these other APOBE3 proteins mediate foreign DNA restriction in other cell types. In support of this, several of the other human A3 proteins decrease the efficiency of stable gene transfer in a cell culture assay, although none as efficiently as A3A (**Fig. 3-12**). The presence of A3s in mammals and related DNA cytidine deaminases in all vertebrates suggests that the DNA restriction mechanism described here may constitute a general innate immune response.

In addition to shedding light on a potential response to infection, the findings described here may also impact applied processes that rely on gene transfer. For example, many laboratory experiments rely on the introduction of DNA into cells. Similarly, gene therapy requires that DNA be stably incorporated into and expressed by a cell. And, in order to create induced pluripotent stem cells, multiple transgenes must be stably introduced into primary somatic cells. In all of the above cases, a major obstacle is that many cells - especially primary cells - express and incorporate foreign DNA very inefficiently. It is possible that endogenous A3 proteins are one of the barriers that make this process so inefficient. Further experiments will be necessary to validate this prediction and to assess the breadth of editing-based DNA restriction.

Several important unanswered questions remain. First, what DNA molecules are the targets of this activity *in vivo*? Candidate targets include bacterial, viral, and parastic DNA, as well as self DNA released from necrotic cells. Such DNA could enter the cell by phagocytosis, other endocytic pathways, or by infection. Studies that have detected

C/G-to-T/A hypermutated HBV and HPV DNA in primary tissues suggest that viral DNAs are subject to editing by A3s *in vivo*[159,169]. But this activity is not limited to viral DNA, as we have shown that transfected plasmid DNA (*i.e.*, bacterial DNA) is restricted.

Additional studies will be required to determine the range of foreign DNA that is subject to A3-mediated restriction. For example, to determine whether (non-transfected) bacterial DNA is subject to A3A-mediated restriction, primary monocytes could be infected with an intracellular bacterium such as *Listeria monocytogenes*. Following infection, DNA could be recovered from cells and 3D-PCR could be used to detect A3A-edited bacterial DNA. Similarly, cells could be infected with different DNA viruses to determine whether their genomes were being edited during infection.

Another unresolved question is *where* in the cell does editing occur? Epitope-tagged A3A distributes throughout the cell (see chapter 4), so it could act in the cytosol, following DNA escape from endocytic compartments, or in the nucleus. Experiments with chimeric A3 proteins that are restricted to the nucleus or the cytoplasm could help resolve this issue.

Finally, it is worth noting that strong parallels can be drawn between the mechanism of foreign DNA restriction described here and the mechanisms of antibody gene diversification by AID and HIV-1 restriction by A3F and A3G. In all three cases, a DNA deaminase is targeted to a particular DNA substrate: foreign DNA, antibody gene DNA, and retroviral cDNA, respectively. In all three cases, ubiquitous DNA 'repair' enzymes are diverted to process the resulting DNA uridines, resulting in foreign DNA restriction, antibody gene diversification, and retrovirus restriction. Thus, we conclude

that DNA deamination has at least three fundamental roles in the vertebrate innate and adaptive immune responses.

Experimental Procedures

Plasmids: The A3A coding region (Genbank accession NM_145699) was amplified by PCR using 5'-GAG CTC GGT ACC ACC ATG GAA GCC AGC CCA GC-3' and 5'-GTC GAC CAT CCT TCC GTT TCC CTG ATT CTG GAG-3', digested with SacI/SalI and ligated into similarly digested pEGFP-N3 (Clontech) and pcDNA3.1-HA vectors[154]. The A3A-GFP fusion reading frame was cut out of pEGFP-N3-A3A by digesting with NheI/NotI and ligated into an XbaI/NotI-digested pEF, a derivative of pEAK8 (Edge Biosystems)[36]. Amino acid substitution mutants were generated using site-directed mutagenesis (QuikChange, Stratagene; primer sequences available on request). All constructs were confirmed by restriction digestion and DNA sequencing. The uracil DNA glycosylase inhibitor expression plasmid, pcDNA3.1-UGI, was created by digesting pEF-UGI[36] with EcoRI/NotI and ligating the UGI-encoding fragment into a similarly digested pcDNA3.1 (Invitrogen). The Sleeping Beauty DNA transposon plasmids, pT2/SV-neo and pCMV-SB10 are as described[31,67]. pTre2-mCherry was created by amplifying the mCherry coding region from pRSETB-mCherry (a gift from Roger Tsien) using primers 5'-GCG GCC GCT TAC TTG TAC AGC TCG TCC ATG CCG-3' and 5' GGA TCC ATC GCC ACC ATG GTG AGC AAG GGC GAG G-3'. The PCR product was digested BamHI/NotI and ligated into a similarly digested pTre2pur (Clontech). This plasmid was digested with XhoI, and the larger fragment was purified on an agarose gel and self ligated.

Cell Culture: Human embryonic kidney (HEK) 293, 293T, and HeLa cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 25 units/ml penicillin and 25 μg/ml streptomycin at 37 ˚C and 5% CO_2.

Transfections of these cells used TransIT-LT1 (Mirus Bio Corporation) according to the manufacturer's protocol.

Cell Lines: HEK-293T cell stably expressing uracil DNA glycosylase were created by transfecting cells with pEF-UGI[36]. Stably transfected cells were selected in puromycin and clonal cell lines screened for inhibition of UDG activity as described[36].

Primary Cells: Primary human peripheral blood mononuclear cells (PBMCs) were isolated from peripheral blood obtained from healthy donors (Memorial Blood Centers, Minneapolis, MN). PBMCs were isolated by density gradient centrifugation using Ficoll Paque Plus according to the manufacturer's directions (GE Healthcare Life Sciences). In some experiments, monocytes were further enriched by negative selection of non-monocytes using magnetic separation (MACS separation, Miltenyi Biotec) or centrifugation (RosetteSep, Stem Cell Technologies), according to the manufacturer's protocols. Enrichment was monitored by flow cytometry and staining with fluorescein-labeled anti-CD14 antibody (Miltenyi Biotec).

PBMC stimulation: PBMCs or monocytes were treated with 2 ng/ml universal type 1 interferon (R&D Systems) or with 3 uM CpG DNA oligonucleotide ("ODN2216" 5'- ggG GGA CGA TCG TCg ggg gg -3', lower case letters designate nucleotides linked by phosphorothioate bonds) or with 3 uM control GpC DNA oligonucleotide (5'- tgc tgc ttt tgt gct ttt gtg ctt -3'; this GpC-containing oligo is not an efficient TLR9 ligand) for the times indicated in the text and figures.

Flow Cytometry: Cells were analyzed using a Cell Lab Quanta SC flow cytometer (Beckman Coulter). CD14+ cells were stained with CD14-FITC (Miltenyi Biotec). Cells were incubated with propidium iodide to assess cell viability.

Immunoblotting: Cells were harvested and total protein extracted in a buffer containing 25 mM HEPES (pH 7.4), 10% glycerol, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM $ZnCl_2$, and protease inhibitors. The extracts were clarified by centrifugation for 10 minutes at 20,800g at 4˚C. The extracted proteins (15 μg) were fractionated by SDS-PAGE, transferred to a polyvinylidene difluoride membrane (Millipore), and probed with an anti-A3A polyclonal antiserum (see below), an anti-GFP monoclonal antibody (Clontech), or anti-EF1a monoclonal antibody (Upstate). Primary antibodies were detected by incubation with fluorescently labeled secondary antibodies and imaging on an Odyssey imaging device (LI-COR Biosciences). In some experiments, the PVDF membrane was stained with Ponceau S to visualize total protein as a loading control.

A3A antiserum: A rabbit was immunized with a peptide corresponding to A3A residues 171-199 (CPFQPWDGLEEHSQALSGRLRAILQNQGN) mixed with TiterMax Gold adjuvant (Sigma). Serum was collected and the ability of the serum to detect A3A was confirmed (data not shown).

Quantitative reverse-transcription PCR assays: Total RNA was isolated from cells (RNeasy, Qiagen). cDNA was reverse transcribed from 1 μg of total RNA using random hexamer primers and AMV reverse transcriptase (Roche). cDNA was used as template in qPCR reactions performed in a Lightcycler 480 instrument (Roche) according to the manufacturer's protocol. Reactions were performed in duplicate and the mean of the duplicates reported. All gene expression data were normalized to the geometric mean of at least two of the following three reference genes, whose levels varied little in our experiments: TATA-box binding protein (TBP), Ribosomal protein 13A (RPL13A), or

hypoxanthine-guanine phosphoribosyltransferase (HPRT). The sequences of all primers and probes are listed in **Table S1**.

Gene transfer experiments: 250,000 293 or HeLa cells were plated into 6-well plates. After 24 hours of incubation, the cells were transfected with neo-encoding plasmid pcDNA3.1 (Invitrogen) and pEF-A3A-GFP or pEF-A3A-E72A-GFP. When A3A expression plasmid was titrated, additional empty pEF vector was added to transfections to keep the pDNA mass constant. Two days post transfection, 100,000 cells were plated into 10cm dishes containing media supplemented with 1 mg/ml G418 (Cellgro). At the same time, to confirm that equivalent numbers of colony-forming cells were plated, serial dilutions of cells were plated into drug-free media in 24 well plates. After 12-14 days of additional incubation, colonies were fixed, stained with crystal violet, and counted. Experiments were performed in triplicate and the mean and standard deviation is reported. Essentially identical results were obtained in 293 and HeLa cells. Sleeping Beauty DNA transposon-mediated gene transfer assays were performed similarly, essentially as described[67].

Cell viability assays: 293 or HeLa cells were transfected as above, and at the indicated time points post transfection, cells were harvested, labeled with propidium iodide, a fluorescent dye to which live cells are impermeable, and analyzed by flow cytometry.

Cell proliferation assays: 293 or HeLa cells were transfected as above, and at each day post transfection, cells were treated with the Cell Titer 96 aqueous reagent (Promega), according to the manufacturer's instructions.

Transient GFP-expression assays: 293 or HeLa cells were transfected as above, and at the indicated times post transfection, GFP-fluorescence was analyzed cells by flow cytometry.

Plasmid DNA qPCR assays: 293 or HeLa cells were transfected as above, and at the indicated times post transfection, cells were harvested and total DNA extracted (DNeasy, Qiagen). 50 ng of total DNA was used as input to qPCR reactions. Reactions were performed on an iCycler instrument using SYBR Green dye (BioRad) according to the manufacturer's protocol. Reactions were performed in duplicate and the mean of duplicates reported. Plasmid DNA levels were normalized to genomic DNA levels as measured by a qPCR assay specific for the beta actin locus. Primers are listed in **Table S1**.

Cytidine deaminase assays: 293T cells were transfected as above and two days post transfection, cells lysates were prepared as above for immunoblotting. The deaminase activity in the lysates was determined using a FRET-based assay essentially as described[164]. Briefly, serial dilutions of the lysate were incubated for two hours at 37˚C with a ssDNA oligonucleotide (5'-AAATTCTAATAGATAATGTGA) labeled on one the 5' end with FAM and the 3' end with TAMRA. These fluorophores are close enough that FRET occurs between them and FAM fluorescence is diminished. Cytosine deaminase activity in the lysates converts a cytidine in the oligo to a uridine. Recombinant UDG is added to excises the resulting uracil, leaving an abasic site. The reactions are heated to cleave the abasic site. Once cleaved, the FAM and TAMRA labels are physically separated, FRET stops, and FAM fluorescence increases. The FAM fluorescence is therefore proportional to the amount of cytosine deaminase activity.

Assays to detect hyper-editing in cell-lines: 293 cells, 293T cells, or 293T cells stably expressing UGI were transfected with reporter plasmids pTRE2-Δpuro-mCherry or pEGFP-N3, pcDNA3.1-UGI or pcDNA3.1, and pEF-A3A-GFP or pEF-A3A-E72A-GFP. When UGI expression was titrated empty pcDNA3.1 vector was added to keep pDNA mass constant. One day post transfection, cells were treated with DNaseI (Roche) to remove extracellular input plasmid. One day later, cells were harvested and total DNA extracted (DNeasy, Qiagen). 50 ng of total DNA was used as input to PCR using primers listed in **Table S2** and Taq DNA polymerase (Roche), which amplifies uracil-containing DNA (the uridine templates the incorporation of an adenosine), with other reagents according to the manufacturer's protocol. Reaction conditions were: 94˚C for 30 seconds, then 25 cycles of 94˚C for 15 seconds, 50˚C for 30 seconds, and 68˚C for 2 minutes, and a final extension at 68˚C for 7 minutes. 2.5 ul of this PCR was used as template in a second PCR using nested primers (**Table S2**). Nested PCR reactions used Phusion DNA polymerase (Finnzymes), with other reagents according to the manufacturer's protocol. Reaction conditions were: a gradient of denaturation temperatures (T_d) for 30 seconds, then 25 cycles of the T_d gradient for 15 seconds, 52˚C for 30 seconds, and 72˚C for 2 minutes, and a final extension at 72˚C for 7 minutes. T_d gradients are as indicated in figures. PCR products were separated on agarose gels and detected by ethidium-bromide staining.

DNA sequencing: PCR products were separated by agarose gel electrophoresis and purified (QIAquick, Qiagen). 20 ng of purified PCR product was sequenced directly (Genewiz) or PCR products were cloned (Clonejet, Fermentas) and sequenced.

DNA sequence analysis: Sequences of PCR products were analyzed using phred software version 0.071220.b (www.phrap.org; reference[44]). Phred calls bases but also calls "secondary bases" if there is a second peak above the background at the same position as the primary peak. At each position in the nucleotide sequence, if the primary or secondary base was not the expected base, the identity of the unexpected base and the fraction editing was calculated as follows: fraction editing = [area of the unexpected peak / (area of the unexpected peak + area of the expected peak)].

Primary cell transfection: Monocytes were treated for 20-24 hours with 3 μM CpG DNA oligonucleotide or mock treated. Monocytes were then transfected with endotoxin-free DNA preparations (Qiagen) using a Nucleofector 2 electroporation device according to the manufacturers instructions (Amaxa). Three hours post-transfection, cells were treated with DNaseI to remove non-transfected extracellular DNA. At various time points pre and post transfection, as indicated in the text, cells were harvested and RNA and protein was isolated. Gene expression was monitored using qRT-PCR, and recovered DNA was analyzed as above.

Figures

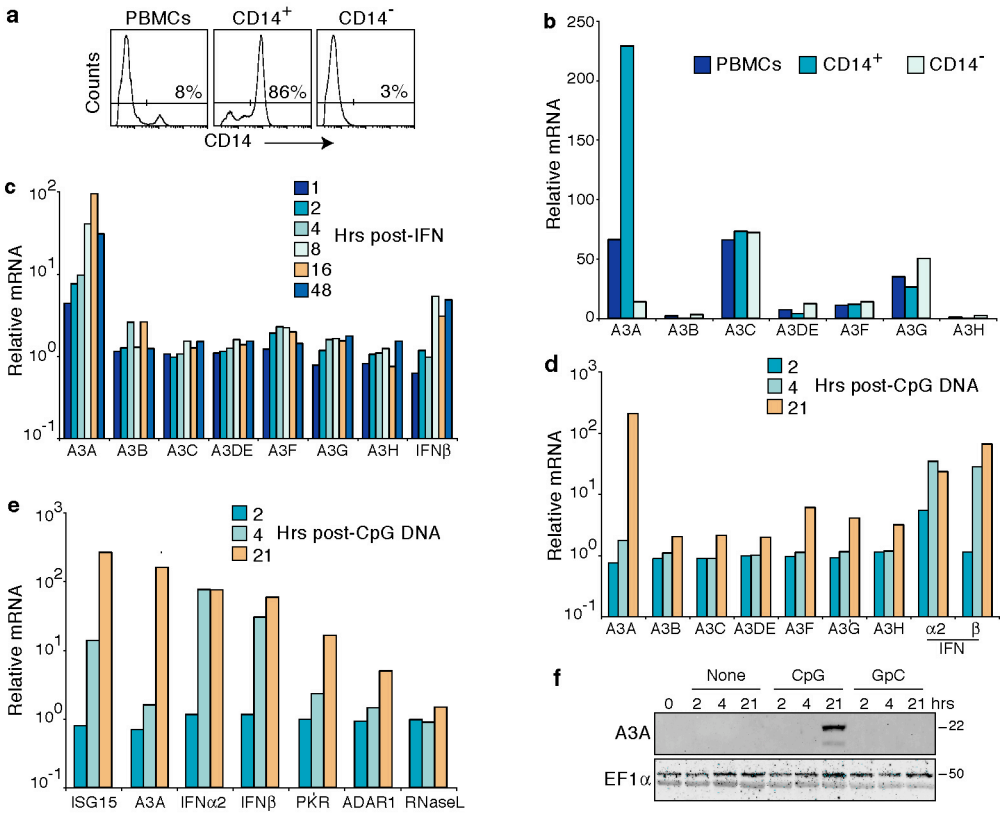

Fig. 3-1. A3A is expressed in monocytes and macrophages and is induced by interferon. a, flow cytometry histograms showing the efficiency of CD14-positive cell enrichment. **b,** basal relative A3 mRNA levels in PBMCs or the specified cell subpopulations, with a value of 1 assigned to the level of A3H mRNA in bulk PBMCs. **c & d,** A3 and IFN mRNA levels in PBMCs treated with recombinant IFN or CpG DNA oligonucleotide for the indicated number of hours. mRNA levels are relative to those measured in untreated cells at each time point. **e**, mRNA levels of the indicated IFN-responsive genes in PBMCs treated with CpG DNA for the indicated times. mRNA levels are relative to those measured in untreated cells at each time point. **f**, A3A protein levels in PBMCs treated with CpG or with a control GpC DNA oligonucleotide. The same membrane was probed with anti-eEF1α for a loading control. Samples from one representative experiment were analyzed in d, e, and f.

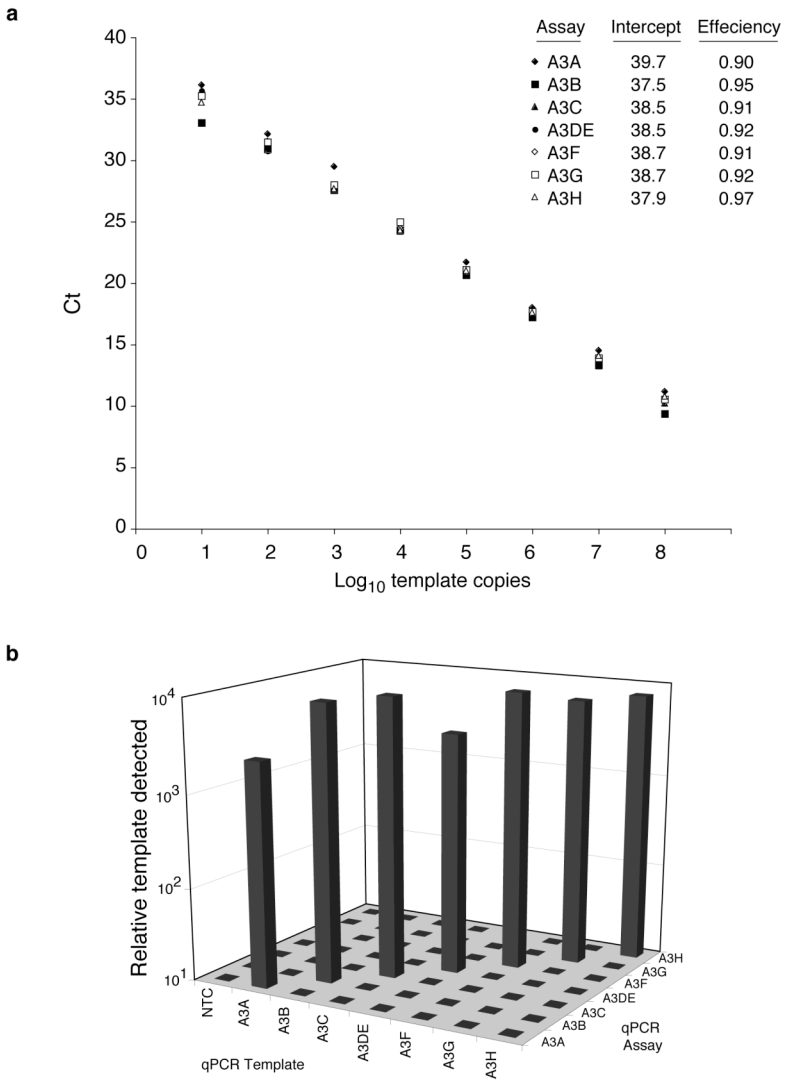

Fig. 3-2. Characterization of A3 qPCR Assays. a, assay efficiencies: a dilution series of templates containing the indicated A3 coding and partial UTR sequences were amplified by qPCR. The cycle at which PCR products accumulated above a threshold level (Ct) is indicated. The y-axis intercept and reaction efficiency for each assay is indicated. **b,** assay specificities: Templates containing the indicated A3 coding and partial UTR sequences were amplified by the indicated assays. The relative amount of template detected by each assay is shown.

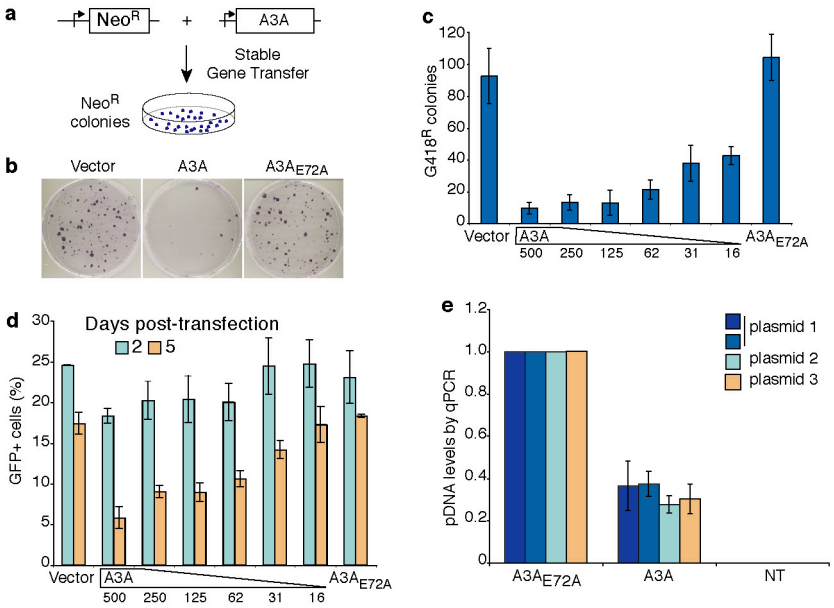

Fig. 3-3. Foreign DNA restriction by A3A. a, schematic of plasmid-based stable gene transfer studies. **b**, images of representative plates of Neo^R clones obtained in a stable gene transfer experiment initiated by co-transfecting HEK-293 cells with a Neo^R plasmid and expression construct for A3A, A3A-E72A or an empty control vector. **c**, the mean and standard deviation of the data in (b) and two additional experiments are shown. The amount of A3A-expression plasmid was decreased in a two-fold dilution series from 500 to 16ng as indicated. **d**, transient expression of a GFP reporter plasmid transfected into 293 cells in the presence of A3A, A3A-E72A or an empty expression vector. GFP fluorescence was quantified by flow cytometry at 2 and 5 days post-transfection, and the mean and standard deviation of three experiments is shown. **e**, Total DNA from HeLa cells transfected with A3A or A3A-E72A was recovered 2 days post-transfection. The amount of three co-transfected plasmids was monitored by quantitative PCR. The A3A-E72A data were normalized to one, and the mean and standard deviation from three experiments is shown. NT, non-transfected control.

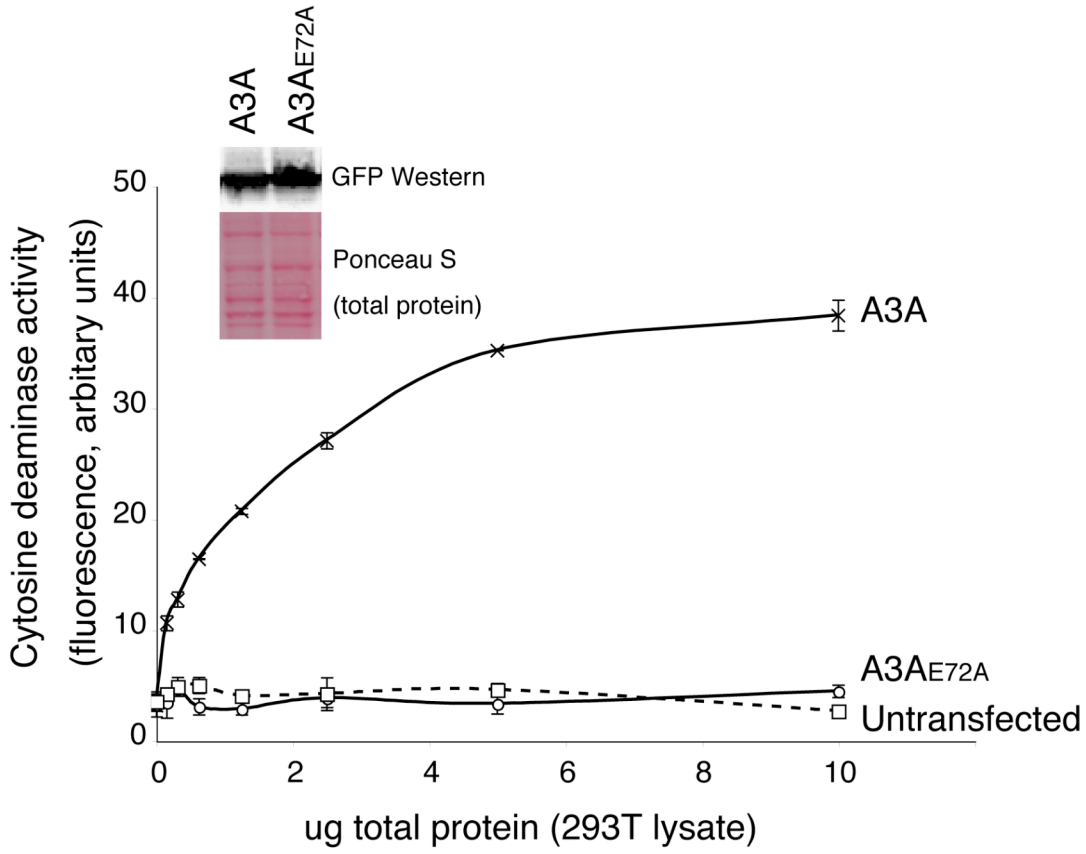

Fig. 3-4. A3A(E72A) lacks deaminase activity and is expressed at wild type levels. A fluorescence-based cytidine deaminase assay was used to compare the cytidine deaminase activity in the lysate of 293T cells transfected with wild type A3A-GFP or $A3A_{E72A}$-GFP (see experimental procedures for details). The mean and standard deviation of two experiments is shown. Inset: a western blot comparing A3A and $A3A_{E72A}$ expression levels in the lysates.

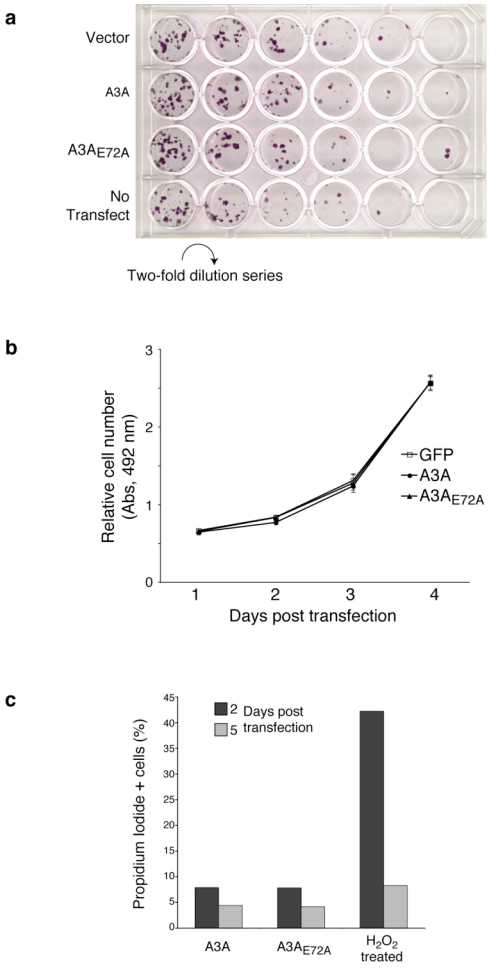

Fig. 3-5. A3A expression does not negatively impact several measures of cellular viability and proliferation. a, in addition to plating cells transfected with the indicated expression constructs into drug-containing media (as in **Fig. 3-3**), dilutions of the same transfected cell populations were plated into drug-free media to quantify the number of viable colony-forming cells. Twelve days post-plating, cells were fixed, stained, and imaged. **b,** HEK-293 cells were transfected with the indicated expression constructs. At the indicated number of days post-transfection, viable cells were quantified using the CellTiter reagent (Promega). The mean and standard deviation of three experiments are shown. **c,** Cells were transfected as above. At the indicated number of days post-transfection, cells were stained with propidium iodide and quantified by flow cytometry. Live cells with intact membranes are impermeable to propidium iodide. Hydrogen-peroxide (H_2O_2; 0.5 mM) served as a cytotoxic control treatment.

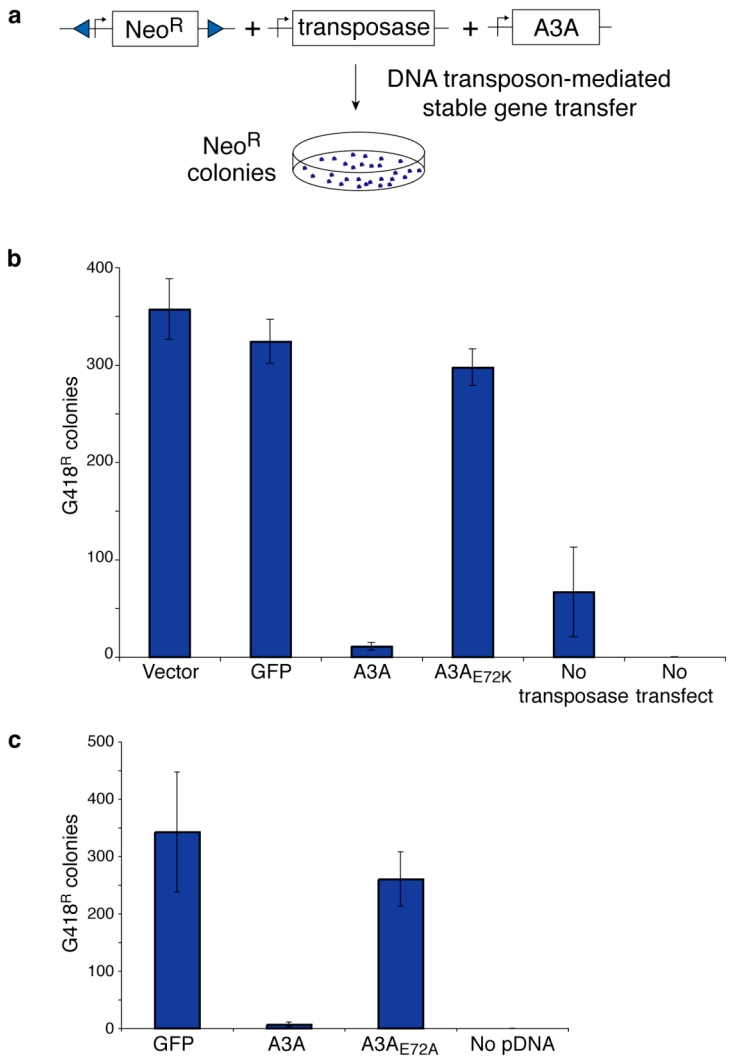

Fig. 3-6. A3A limits DNA-transposon-mediated gene transfer in multiple cell types a, HEK-293 cells were transfected with a transposase-expressing plasmid, a Neo^R reporter transposon plasmid, and A3A-expressing or control constructs as indicated. Two days post-transfection, cells were plated into G418-containing medium and colonies allowed to grow for twelve days, at which point the colonies were fixed, stained, and counted. No transposase controls were not transfected with transposase expression plasmid, so $G418^R$ colonies likely arise from random integration of the neo^R plasmid, as in **Fig. 3-3**C. **b,** A3A limits stable gene transfer in multiple cell types. HeLa cells were transfected as in **Fig. 3-3**C with a Neo^R reporter plasmid (pcDNA3.1) and gene transfer efficiency was measured as above.

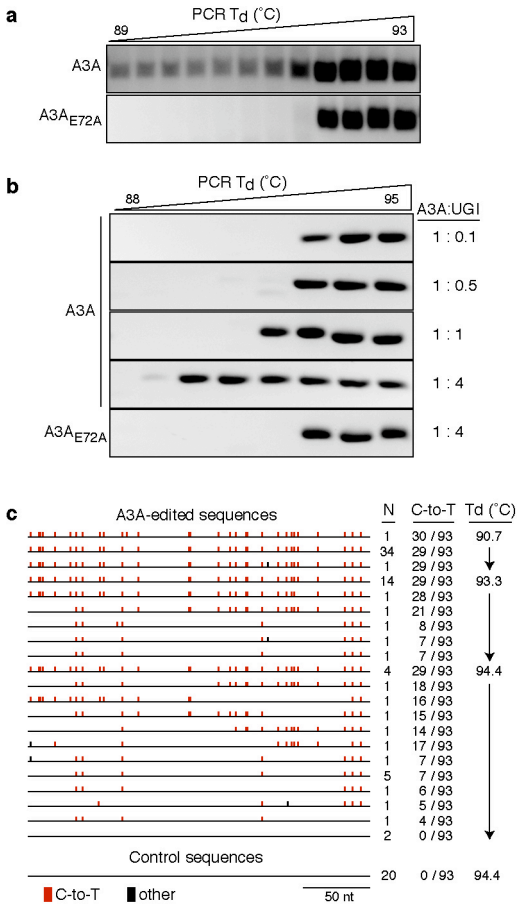

Fig. 3-7. A3A deaminates transfected plasmid DNA to generate lesions for uracil DNA glycosylase. a, agarose gel analysis of hyper-editing assay PCR products. UGI-expressing HEK-293T cells were transfected with plasmids encoding A3A or A3A-E72A and a Neo^R reporter construct. Total DNA was recovered 48 hours post-transfection and analyzed by 3D-PCR at the indicated denaturation temperatures (Td). **b,** agarose gel analysis of hyper-editing assay PCR products. HEK-293 cells were transfected with increasing amounts of UGI plasmid, a Neo^R reporter construct, and a plasmid encoding A3A or A3A-E72A. Total DNA was recovered 48 hours post-transfection and analyzed by 3D-PCR at the indicated Td. **c,** summary of the plasmid DNA sequences recovered from the experiment in (b). C/G-to-T/A hypermutations are indicated as red tics along the consensus sequence, and all other base substitutions as black tics. The control sequences were obtained from cells expressing A3A-E72A. The number of times each sequence was recovered, the number of C-to-T conversions in each sequence (out of 93 total cytidines),

and the PCR Td used to amplify the populations of molecules from which the sequences are derived are indicated.

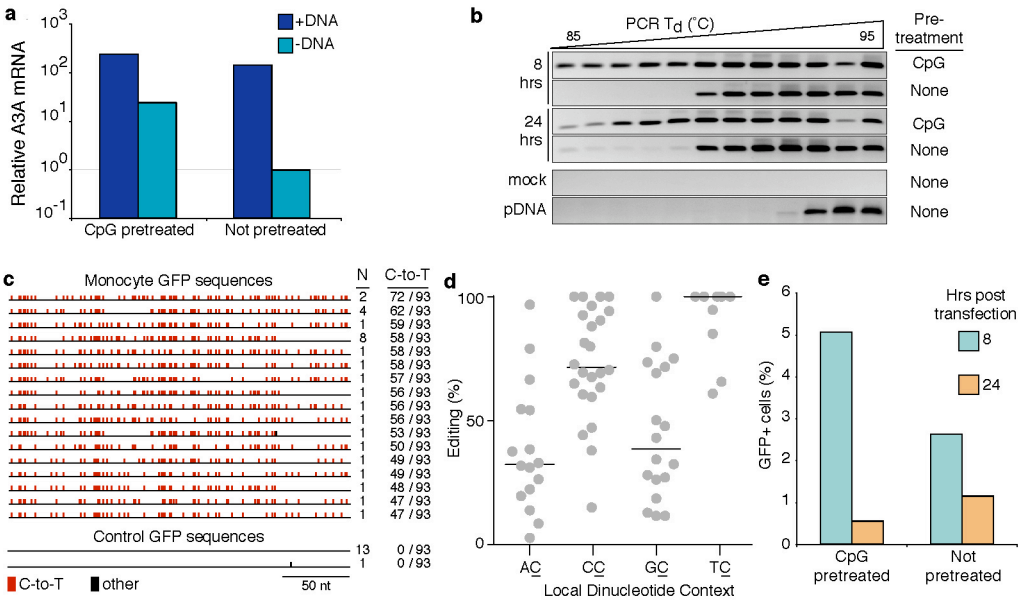

Fig. 3-8. A3A mutates foreign DNA in primary cells. **a**, A3A expression levels in CpG-pretreated or untreated monocytes transfected with plasmid DNA (+DNA) or with buffer alone (-DNA). RNA was harvested 8 hours post transfection, and A3A mRNA levels were determined by qRT-PCR and are shown relative to the level in non-pretreated, mock-transfected cells. **b**, agarose gel analysis of hyper-editing assay PCR products from the experiment described in panel (a). Total DNA was isolated and a portion of the GFP gene was amplified using a gradient of PCR denaturation temperatures (Td). "Mock": Total DNA extracted from mock-transfected cells was subjected to the same PCR scheme. "pDNA": DNA from the same preparation used to transfect the monocytes was mixed with total DNA from mock-transfected cells and subjected to the same PCR scheme. **c**, summary of plasmid DNA sequences recovered from the experiment described in panel (b). Plasmid DNA recovered from primary monocytes was subjected to PCR as in (a). Molecules from the population amplified at 87˚C Td were cloned and sequenced. C/G-to-T/A mutations are indicated as red tics along the consensus sequence, and all other base substitutions as black tics. **d**, DNA recovered from CpG-treated monocytes was amplified by PCR as in (c) and the percentage of molecules in the population that were edited at each cytidine was determined by analyzing the chromatograms (partially edited cytidines exhibited mixed C/T peaks, see also **Fig. 3-9**). The percentages were grouped according to dinucleotide context and the median percentage per group is indicated by a horizontal line. For

instance, 5 of the 9 cytidines within 5'-TC dinucleotides were edited in 100% of the molecules amplified (*i.e.*, only a T peak was evident in the chromatogram). **e**, GFP expression in primary monocytes was determined at the indicated time post transfection using flow cytometry.

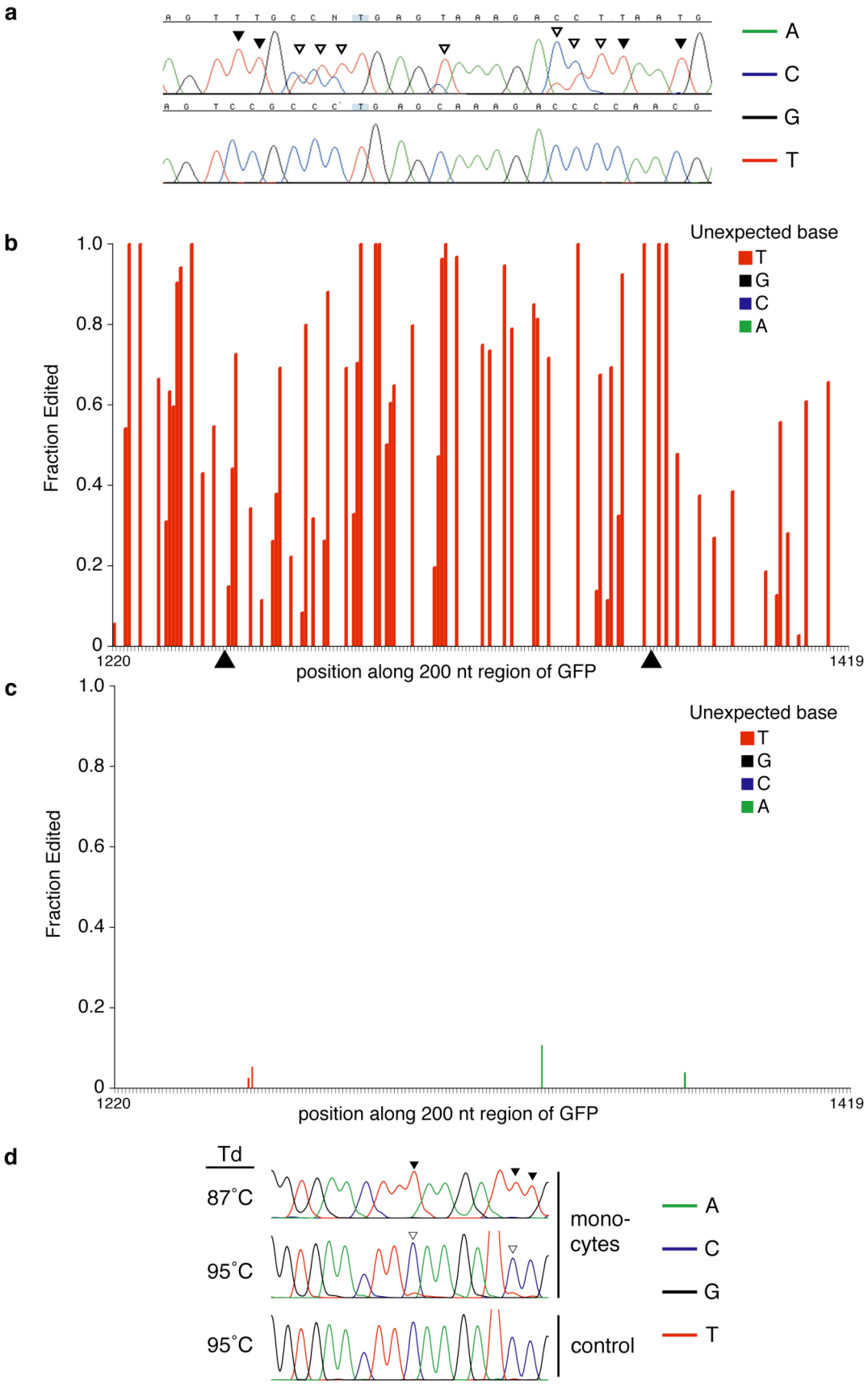
a
A
C
G
T
b
Unexpected base
T
G
C
A
Fraction Edited
1.0
0.8
0.6
0.4
0.2
0
1220
1419
position along 200 nt region of GFP
c
Unexpected base
T
G
C
A
Fraction Edited
1.0
0.8
0.6
0.4
0.2
0
1220
1419
position along 200 nt region of GFP
d
Td
87˚C
95˚C
95˚C
mono-
cytes
control
A
C
G
T

Fig. 3-9. Analysis of populations of hyperedited molecules. Plasmid DNA was recovered from transfected monocytes, amplified using an 87˚C PCR denaturation temperature, purified by agarose gel electrophoresis, and sequenced. These sequences originated from a population of molecules, so if a sufficiently large fraction differed from the expected nucleotide at each position, then a mixed chromatogram peak would be evident. **a,** A representative region of chromatograms of sequences of PCR products from plasmid DNA recovered from monocytes (top) or control PCR products (bottom). Multiple instances of cytidines that had been completely (filled arrowheads) or partially (open arrowheads) converted to uridines were evident. **b,** a 200 nucleotide region of the amplicon, for which there was chromatogram data of sufficiently high quality, was analyzed using phred software[44] (see Methods). 70/72 cytidines (97%) in this region displayed some evidence of C-to-T (C-to-U) conversion (*i.e.* had been edited in a fraction of the molecules in the population). The fraction edited for each of these cytidines is indicated. The position of the two cytidines for which there was no evidence of editing are marked by black arrowheads (there was no obvious sequence context that explained why those cytidines remained unedited). The position of the 200 nucleotide region in the plasmid pEGFP-N3 (Clontech) is indicated. **c,** the same analysis was performed on a control sample, which was generated by mixing GFP-encoding plasmid with total DNA isolated from mock-transfected cells and subjected to the same PCR amplification and sequencing scheme. PCR products from the lowest denaturation temperature at which product amplified (95 ˚C) were sequenced. Only minor secondary peaks are evident, likely arising from sequencing error or mutations introduced during PCR. **d,** representative regions of chromatograms are shown. Cytidines that have experienced 100% editing are indicated with filled arrowheads. Cytidines that have undergone partial editing (*i.e.*, a fraction of the molecules in the population have been edited at that position) are indicated with open arrowheads. The PCR denaturation temperatures used to amplify the populations of molecules and the source of the template DNA are indicated.

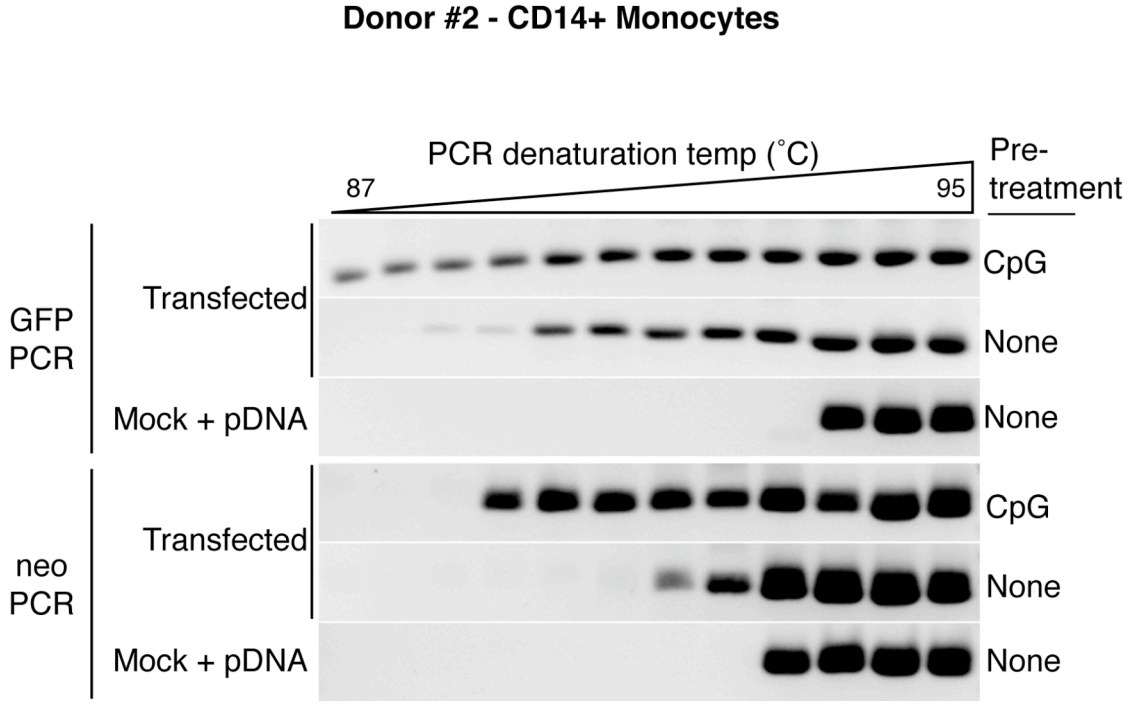

Fig. 3-10. Hyperedited plasmid DNA recovered from transfected monocytes is detectable using multiple sets of PCR conditions and multiple donors. a, primary monocytes were transfected with a GFP expression plasmid (pEGFP-N3). Twenty hours later, total DNA was recovered from the cells and PCR to detect hyperedited DNA were performed as in Fig. 4b. Regions of the plasmid containing portions of the GFP and the Neo^R coding sequences were amplified.

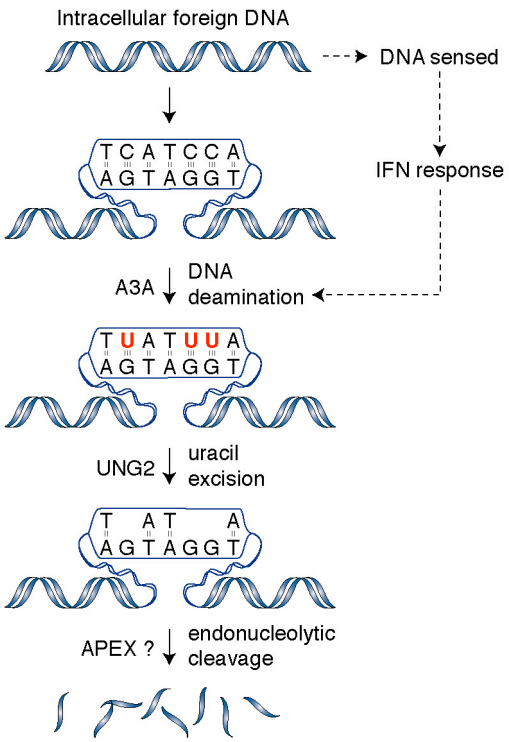

Fig. 3-11. **A model for foreign DNA restriction triggered by A3A**. Foreign DNA enters the cell by escaping from an endosomal or phagosomal compartment, by infection, or by other means. TLR9 senses DNA in endosomes or an unknown factor senses it in the cytosol. Both sensors initiate signaling cascades that result in production of type 1 IFNs. The IFN response induces A3A expression. A3A engages the foreign DNA, deaminating multiple cytidines in a molecule. The resulting uracils are excised by UNG2, creating nuclease-sensitive abasic sites. Cleavage of the backbone by APEX or other endonucleases results in fragmentation and degradation of the foreign DNA.

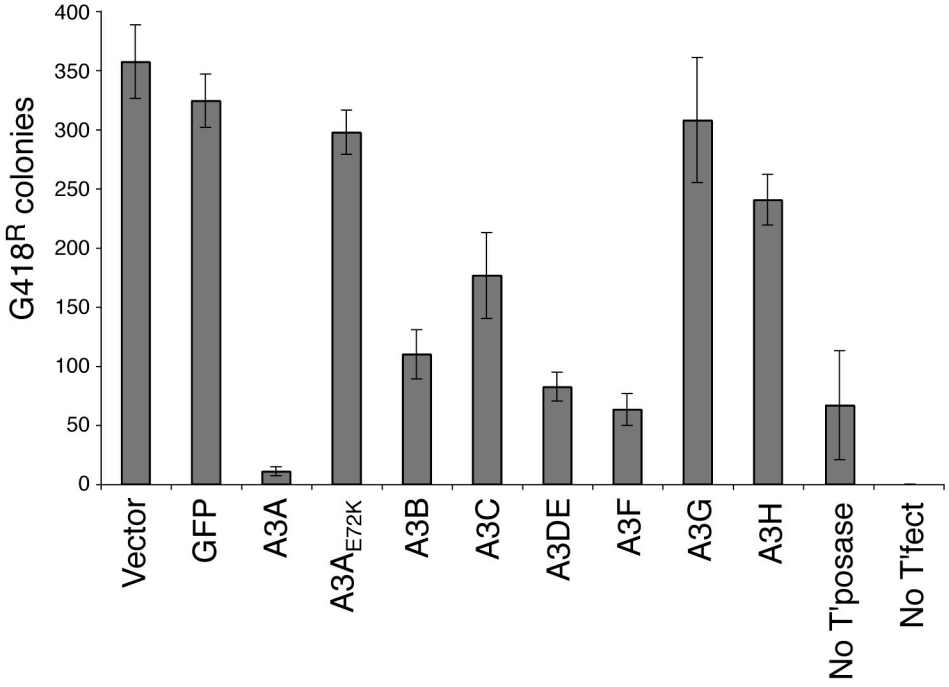

Fig. 3-12. **Multiple A3 proteins possess DNA restriction activity.** A DNA transposon-mediated stable gene transfer experiment was performed as in **Fig. 3-6**. No t'posase: transposase expression plasmid was omitted from transfections. No t'fect: untransfected cells.

Gene Symbol !	NCBI Accession	5' Primer		3' Primer		Probe	
		Name	Seq (5'-3')	Name	Seq (5'-3')	Name	Seq[b]
APOBEC3s							
APOBEC3A	NM_145699	RSH2742	gagaagggacaagcacatgg	RSH2743	tggatccatcaagtgtctgg	UPL26	ctgggctg
APOBEC3B	NM_004900	RSH3220	gacctttggtccttcgac	RSH3221	gcacagccccaggagaag	UPL1	cctggagc
APOBEC3C	NM_014508	RSH3085	agcgcttcagaaaagagtgg	RSH3086	aagtttcgttccgatcgttg	UPL155	ttgccttc
APOBEC3D/ APOBEC3DE	NM_152426/ ENST00000216099[a]	RSH2749	acccaaacgtcagtcgaatc	RSH2750	cacatttctgcgtggttctc	UPL51	ggcaggag
APOBEC3F	NM_145298	RSH2751	ccgtttggacgcaaagat	RSH2752	ccaggtgatctggaaacactt	UPL27	gctgcctg
APOBEC3G	NM_021822	RSH2753	ccgaggacccgaaggttac	RSH2754	tccaacagtgctgaaattccg	UPL79	ccaggagg
APOBEC3H	NM_181773	RSH2757	agctgtggccagaagcac	RSH2758	cggaatgtttcggctgtt	UPL21	tggctctg
Ref Genes							
TBP	NM_003194	RSH3231	cccatgactcccatgacc	RSH3232	tttacaaccaagattcactgtgg	UPL51	ggcaggag
RPL13A	NM_012423	RSH3227	ctggaccgtctcaaggtgtt	RSH3228	gccccagataggcaaactt	UPL74	ctgctgcc
HPRT1	NM_000194	RSH2959	tgaccttgatttatttttgcatacc	RSH2960	cgagcaagacgtttcagtcct	UPL73	gctgagga
ACTB[c]	NM_001101	RSH730	atcatgttgagaccttcaa	RSH731	agatgggcacagtgtgggt		
Other Genes							
IFNA2	NM_000605	RSH2951	tcctgcttgaaggacagaca	RSH2952	tttcagcctttggaactgg	UPL63	aggaggag
IFNB	NM_002176	RSH2953	ctttgctatttttcagacaagattca	RSH2954	gccaggaggtttctcaacaat	UPL20	ctggctgg
ISG15	NM_005101	RSH3540	gcgaactcatcttttgccagt	RSH3541	agcatcttcaccgtcaggtc	UPL76	tggctgtg
EIF2AK2 [PKR]	NM_002759	RSH3477	gacaaagcttccaaccagga	RSH3478	cggtatgtattaagttcctccatga	UPL39	ctccacct
ADAR	NM_001111	RSH3479	cgagaatcccaaacaaggaa	RSH3480	ctggattccacagggattgt	UPL3	ctgctggg
RNASEL	NM_021133	RSH3481	gggaacatgtgagggactgt	RSH3482	cacattccgaagcgtcctat	UPL62	acctgctg
pDNA assays[c]							
NeoR gene[d]	N/A	RSH951	gggtagccaacgctatgtcc	RSH2466	gcttgccgaatatcatggtg		
Sleeping Beauty IR/DR(L)[d]	N/A	RSH2608	aaaatcgttttcaactactccac	RSH2609	ggaaaaatgacttgttgtcatgc		
Sleeping Beauty Transposase[e]	N/A	RSH2606	cgttatgtttggaggaagaaagg	RSH2607	tgccatcttatttgtgaagtgc		
GFP[f]	N/A	RSH1960	agaacggcatcaagtgaac	RSH1961	tgctcaggtagtggttgtcg		

(a) Ensembl transcript accession number (ensembl.org).
(b) It is not known whether the UPL probes correspond to the coding or template DNA strands of their target sequences (Roche proprietary information).
(c) These assays used SYBR Green I dye (Qiagen) and thus did not require a specific probe.
(d) Amplicon in pT2-SV-neo
(e) Amplicon in pCMV-SB-10
(f) Amplicon in pEAK8-GFP-based constructs

Table 3-1. qPCR primers and probes.

Gene !		5' Primer		3' Primer	
		Name	Seq (5'-3')	Name	Seq (5'-3')
GFP[a]	First PCR	RSH1358	gccaccatggtgagcaagggcgagga	RSH2872	aatagggttccgcgca
GFP[a]	Nested PCR	RSH1960	agaacggcatcaaggtgaac	RSH550	gcaagtaaaacctctacaaatgtggt
Neo[a]	First PCR	RSH1337	cagttccgcccattctccg	RSH266	ttgtttgcaagcagcagatta
Neo[a]	Nested PCR	RSH2388	attgaacaagatggattgcacg	RSH3566	atactttctcggcaggagca
mCherry[b]	First PCR	RSH1052	cgctgttttgacctccatag	RSH3467	tggcatatgttgccaaactc
mCherry[b]	Nested PCR	RSH2842	agggcgagatcaagcagag	RSH3626	tcatgagggtccatggtg

(a) In plasmid pEGFP-N3
(b) In plasmid pTRE2-mCherry

Table 3-2. 3D-PCR primers and probes.

Chapter 4: Two Regions within the Amino-Terminal Half of APOBEC3G Cooperate To Determine Cytoplasmic Localization.

This chapter explores the molecular bases for the cytoplasmic localization of APOBEC3G. Sequestering some APOBEC3s in the cytoplasm may be a way to keep them away from genomic DNA.

Adapted with permission from: Stenglein, Matsuo, and Harris (2008) J Virology 81(18) pp. 9591-9599. Permission granted by The American Society for Biochemistry and Molecular Biology.

Chapter Summary

The human A3 proteins share certain characteristics. For instance, all possess nucleic acid binding and cytidine deaminase activities. But it is also possible to differentiate the A3s, with one of the most obvious differences being that these proteins occupy different subcellular compartments (**Fig. 4-1**). This differential compartmentalization may relate to differences in A3 function. A3G, for instance, localizes predominantly to the cytoplasm of cells. This localization pattern is consistent with a model wherein cytosolic A3G packages into virus particles as they assemble on the plasma membrane. Sequestering some A3s in the cytoplasm may also be a way to keep them away from genomic DNA. To better understand the basis for differential A3 subcellular localization, we compared A3G to A3B, which is predominantly nuclear. A3G/A3B chimeric proteins mapped a primary subcellular localization determinant to a region within the first 60 residues of each protein. A panel of 25 A3G mutants, each with a residue replaced by the corresponding A3B residue, revealed that several positions within this region were particularly important, with Y19D showing the largest effect. The mislocalization phenotype of these mutants was only apparent in the context of the amino-terminal half of A3G and not the full-length protein, suggesting the existence of an additional localization determinant. Indeed, a panel of 5 single amino acid substitutions within the 113-128 region had little effect by themselves, but, in combination with Y19D, two (F126S and W127A) caused full-length A3G to redistribute throughout the cell. The critical localization-determining residues were predicted to cluster on a solvent-exposed surface, suggesting a model in which these two regions of A3G combine to mediate an intermolecular interaction that controls subcellular localization.

Introduction

The human genome encodes seven apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 (APOBEC3 or A3) proteins [70]. These proteins are capable of defending cells against a variety of genetic pathogens such as endogenous retrotransposons and exogenous retroviruses [see introduction in chapter 1 and these reviews [6,24,32,54,63,128]]. The most-studied A3 target is human immunodeficiency virus-1 (HIV-1). A3 proteins, in particular APOBEC3F (A3F) and APOBEC3G (A3G), are able to limit the replication of Vif-deficient HIV-1. They do this by gaining access to viral particles as they bud from infected cells. The A3s transit within the virus particle to a cell that will be infected. There, the A3s employ their DNA cytidine deaminase enzymatic activity to convert cytidine bases in the viral genome to uracils during reverse transcription, thereby blocking infection.

Members of the A3 family possess DNA cytidine deaminase activity, but they can be differentiated on the basis of other activities [*e.g.*, [8,11,17,33,60,68,87,153,176,186]]. For instance, some A3s such as A3F and A3G are potent inhibitors of HIV-1 replication, while others such as APOBEC3B (A3B) have modest anti-HIV activity [*e.g.*, [8,38,153,182]]. The subcellular distribution of A3 proteins is also a distinguishing feature (**Fig. 4-1** and refs[4,5,11,12,16,18,51,52,55,78,96-98,108,120,153,174,175]). For instance, A3G is predominantly cytoplasmic, whereas A3B is primarily nuclear. These differences in subcellular localization presumably reflect differences in the sets of proteins and/or nucleic acids with which the A3s interact and they also suggest that the A3s may perform distinct functions within the cell.

The A3 proteins are evolutionarily related to the mRNA editor APOBEC1 and the antibody gene DNA deaminase AID. AID is present in all vertebrates but A3s are found only in mammals, suggesting that an ancestral *AID* gene duplicated and diverged to give rise to the present day *A3* genes [30,59,70,93,135,173]. It was therefore reasonable to hypothesize that the A3s and AID share conserved properties in addition to DNA cytidine deaminase activity. For instance, both AID and APOBEC1 are nucleo-cytoplasmic shuttling proteins that are exported from the nucleus by the CRM1-dependent nuclear export pathway [13,23,66,99,180,181]. A3G also has a putative leucine-rich CRM1 nuclear export sequence spanning residues 369-379 (**Fig. 4-2**). However, localization studies indicated that A3G is not subject to CRM1-dependent nuclear export[4]. Rather, its cytoplasmic localization was ascribed to a 16-residue peptide, dubbed a cytoplasmic retention signal (CRS), spanning residues 113-128[5].

In the studies described in this chapter, we inhibited CRM1 and deleted the leucine-rich region of A3G to independently confirm some of this prior work. However, our analyses of chimeric A3G-A3B proteins and A3G amino acid substitution mutants suggested a more complex molecular explanation for A3G's cytoplasmic localization, namely that residues within amino acids 1-60 and 113-128 cooperate to determine A3G's cytoplasmic localization. A structural model of the amino(N)-terminal half of A3G supported these data by indicating that critical residues within this region are likely to be solvent-exposed and therefore available for intermolecular interactions.

Results

A predicted nuclear export signal is not required for A3G's cytoplasmic localization.

A3G residues 369-379 QDLSGRLRAIL are similar to the canonical leucine-rich nuclear export motif (LXLX$_{2-3}$LX$_{2-3}$L) that enables proteins to be bound by CRM1 and shuttled to the cytoplasm [*e.g.*, the nuclear export signal (NES) of AID: LRDAFRTLGL [13,47,66,99,118,152]]. We analyzed the localization of A3G-GFP and AID-GFP in the presence and absence of the CRM1 inhibitor leptomycin B [LMB [178]]. LMB caused AID-GFP to accumulate in nuclei of transiently transfected HeLa cells, as shown previously [**Fig. 4-3**A; [4,13,66,99]]. In contrast, LMB treatment of cells expressing A3G-GFP had no effect [**Fig. 4-3**A and [4]]. To support this finding, we deleted the entire putative NES of A3G and observed that this construct, encoding A3G$_{1-369}$ fused to GFP, was still predominantly cytoplasmic (**Fig. 4-3**B). These data corroborated prior studies that used LMB and substitution mutants to show that A3G is not being shuttled by CRM1 out of the nucleus [4].

A putative nuclear localization signal in A3B is not required for its nuclear localization.

To contrast with A3G, we also tested the hypothesis that A3B, a nuclear resident, is actively imported into the nucleus. In support of this hypothesis, A3B has a putative nuclear localization signal (NLS) at amino acids 206-212 (PLVLRRR; **Fig. 4-2**), and it has been reported to be a nucleo-cytoplasmic shuttling protein[11]. This amino acid motif is similar to a functional NLS in the SV40 large T antigen [PKKKRKVE [48,73]]. We

performed site-directed mutagenesis, replacing the proline and arginines in this motif with alanines. All of these A3B mutants remained nuclear, including a quadruple mutant with the putative NLS motif changed at four positions to ALVLAAA (**Fig. 4-3**C and data not shown**)**. This result demonstrated that the putative NLS is not required for A3B's nuclear localization in HeLa cells, and that A3B may be transported to or retained in the nucleus by other means. We did note that there was a slight increase in cytoplasmic fluorescence with the quadruple mutant, but the bulk of the fluorescence clearly remained nuclear.

The N-terminal halves of A3G and A3B recapitulate the localization of the full-length proteins.

A3G and A3B are termed "double-domain" A3s because each has two zinc-binding domains. The genes of the double-domain proteins are likely to have arisen from the duplication of an ancestral "single-domain" *AID*-like *A3* gene[30,59,70,173]. We therefore examined the localization of the N- and C-terminal halves of A3B and A3G by themselves to map the determinants of the localization of the full-length protein. Previous structural studies on the C-terminal domain informed the selection of the dividing line between the domains, which was chosen so as to not disrupt secondary structure [**Fig. 4-2**, **Fig. 4-5**, and [20,21,57]].

The N-terminal domain of A3G ($A3G_{1-196}$-GFP, A3G-NTD) localized predominantly to the cytoplasm and the N-terminal domain of A3B ($A3B_{1-192}$-GFP, A3B-NTD) to the nucleus, much like their respective full-length parental proteins (**Fig. 4-4**A). In contrast, their C-terminal domains, $A3G_{197-384}$ (A3G-CTD) and $A3B_{193-382}$-GFP (A3G-

CTD), distributed throughout the cell in a manner indistinguishable from GFP alone. Similar data were obtained with HA-tagged proteins in fixed cells (**Fig. 4-4**B). Thus, the N-terminal halves of the proteins clearly contained their primary localization determinants. These data are largely consistent with prior deletion studies[5,11].

As an additional test that the N-terminal half of A3G and A3B contained these proteins' major subcellular localization determinants, we constructed full-length chimeric proteins combining one domain of A3B with one domain of A3G. We created a chimera combining the N-terminal domain of A3B with the C-terminal domain of A3G ($A3B_{1-190}G_{195-384}$) and one that fused the N-terminal domain of A3G with the C-terminal domain of A3B ($A3G_{1-194}B_{191-382}$). The chimera crossover point was selected to maintain overall structural integrity, occurring within a region of high similarity that is predicted to link the N- and C-terminal domains of these proteins (*i.e.*, after α5 in the A3G-NTD model structure; **Fig. 4-2** and **Fig. 4-5**). Similar crossover points were previously used to generate functional A3G/F chimeras, and an A3G-A3A chimera[7,55,56]. Like A3B, the chimera consisting of the NTD of A3B and the CTD of A3G localized predominantly to the nucleus (**Fig. 4-4**A and **Fig. 4-4**B). In contrast, like A3G, the chimera fusing the NTD of A3G and the CTD of A3B localized predominantly to the cytoplasm (although we noted that this chimera exhibited a higher degree of nuclear fluorescence than wild-type A3G). Virtually identical results were obtained with GFP-tagged proteins in live cells and HA-tagged proteins and immunofluorescence in fixed cells (**Fig. 4-4**A and **Fig. 4-4**B**)**. These data support the conclusion that the N-terminal halves of A3B and A3G provide the primary determinants of their subcellular localization.

The first 60 amino acids of A3G and A3B harbor key subcellular localization determinants.

We next wanted to more finely map the region within the N-terminal half of A3B and A3G that determines their subcellular distributions. Our strategy to do this was guided by A3B, A3F and A3G amino acid alignments (**Fig. 4-2**). We and others have previously shown that A3F, like A3G, is cytoplasmic[9,78,153,175]. We also noted that, although the full-length proteins distribute to opposite compartments of the cell, the amino terminal halves of A3B and A3F are largely identical (**Fig. 4-2**; residues 65-191 of A3B and 66-192 of A3F are 93% identical). In contrast, residues 1-65 of A3F and 1-64 of A3B share only 56% identity, whereas A3F and A3G are virtually identical over the same region [59 out of the first 60 residues; shaded blue in **Fig. 4-2** and [87]]. These correlations strongly implied that a key determinant of the subcellular localization of both A3G and A3B would reside within the first 60 amino acids.

We therefore constructed chimeras that replaced the first 60 amino acids of A3G with the corresponding residues or A3B and *vice versa* (**Fig. 4-2**). These chimeric proteins are also likely to be structurally sound, because the fusion junctions were predicted to lie in a flexible loop between the β2 strand and α1 helix [see **Fig. 4-2**, **Fig. 4-5**, and [20]]. To verify their structural integrity, we performed *E. coli*-based mutation assays. All of the chimeras exhibited intrinsic DNA deaminase activity equal to or greater than that of the proteins from which they were derived (data not shown). We next assessed the localization patterns of these chimeras. Like A3G, the $A3G_{1-59}B_{61-382}$ chimera exhibited a predominantly cytoplasmic localization in the majority of cells (**Fig. 4-6C)**. In contrast, the $A3B_{1-60}G_{60-384}$ chimera localized predominantly to the nucleus in

most cells. Similar results were again obtained using GFP- and HA-tagged proteins (**Fig. 4-6**C**)**. We noted that these results were less clear-cut than those for the chimeras crossing over at the midway point. Nevertheless, the first 59 residues of A3G conferred an A3G-like localization pattern to the rest of A3B. And replacing the first 59 residues of A3G with those of A3B caused A3G to adopt an A3B-like localization pattern. These results supported the conclusion that the first 60 amino acids of A3G harbor a major subcellular localization determinant.

Single amino acid substitutions can disrupt the subcellular distribution of A3G.

To more finely map the region of A3G responsible for its cytoplasmic localization, we constructed a panel of amino acid substitution mutants within this critical 60 amino acid N-terminal region. A3G and A3B differ at 27 residues within their first 63 amino acids (**Fig. 4-2**). At 25 of these differing positions, we generated a mutant with the A3G residue replaced by the corresponding A3B residue. For instance, we created an A3G mutant with lysine 2 replaced by asparagine 2 of A3B (A3G-K2N). We did not generate two possible mutants because they were deemed conservative (M9/V9, I53/V54).

We first examined the subcellular distribution of the single amino acid substitution mutants, initially as A3G-NTD-GFP derivatives. The localization of most of the mutants was predominantly cytoplasmic and indistinguishable from the wild-type protein (*e.g.*, A3G-NTD-S18Y-GFP in **Fig. 4-7**A). However several mutants, such as Y19D and Y22E, exhibited a clearly disrupted localization pattern (**Fig. 4-7**A). For instance, 86% of the Y19D-expressing cells and 64% of the Y22E-expressing cells had a cell-wide or predominantly nuclear fluorescence localization pattern (**Fig. 4-7**B). Several

other mutants such as R24E, S28Y, T32Y and S60F also mislocalized (**Fig. 4-7**B). As a control, we characterized the expression level of the A3G-NTD-GFP mutants by immunoblotting (**Fig. 4-7**C). Some of the mutations resulted in an apparent decrease in steady-state protein expression level or a decrease in protein solubility. Immunoblotting also revealed the presence of cross-reacting bands of higher mobility that could correspond to fragments of the A3G-NTD-GFP proteins. However, there was no apparent correlation between mislocalization and expression level. These data therefore defined single amino acids required for A3G-NTD localization and confirmed that the first 60 amino acids form a region critical for determining the protein's subcellular distribution.

Two regions of A3G cooperate to determine cytoplasmic localization.

Despite the fact that A3G-NTD-Y19D or -Y22E localized aberrantly, none of these substitutions by themselves significantly altered the cytoplasmic localization of the full-length protein (**Fig. 4-8**, **Fig. 4-9**, and data not shown). A previous study had suggested that A3G residues 113-128 were important for cytoplasmic localization [5]. We therefore hypothesized that residues within amino acids 1-60 and 113-128 might cooperatively determine A3G's subcellular distribution. To test this hypothesis, we combined Y19D with single amino acid substitutions within the F126 to D130 region, which is predicted to be solvent exposed (see below and **Fig. 4-10**). Specifically, we analyzed the subcellular distribution of A3G-F126S, -W127A, -D128K, -P129R and -D130K alone and in combination with Y19D. Like the single amino acid substitutions within the first 60 amino acids, none of these mutations by themselves significantly disrupted full-length A3G's cytoplasmic localization (**Fig. 4-8**A). However, when combined with Y19D, two of these substitutions, F126S and W127A, caused full-length

A3G to distribute cell-wide (**Fig. 4-8**B). These data supported the hypothesis that two distinct regions of A3G cooperate to determine cytoplasmic localization.

Residues critical for cytoplasmic localization are predicted to cluster on a common solvent exposed surface.

The Matsuo and Harris laboratories recently used NMR spectroscopy to obtain several high-resolution structures of the C-terminal domain of A3G [20,57]. This structural information and sequence homology between the N- and C-terminal domains were used to generate a three-dimensional model of the N-terminal half of A3G (**Fig. 4-10**, **Fig. 4-5**, and experimental procedures). Interestingly, the predicted A3G-NTD structure indicated that most of the critical localization-determining residues, including Y19, Y22, F126, and W127, cluster on a common solvent exposed surface. The proximity of the two critical regions on the protein's surface strongly suggests that they cooperate to mediate an intermolecular interaction that governs A3G's cytoplasmic localization.

Discussion

Many prior studies have noted that A3G localizes to the cytoplasm of cells, in a diffuse cytosolic manner and in some cells in punctate bodies thought to be mRNA-processing centers (P bodies) or other RNA-containing structures[4,5,11,12,16,18,25,26,52,55,78,80,96-98,108,120,153,174,175]. However, a clear molecular explanation for this property has been elusive. Here, we confirmed prior work showing that, unlike its homolog AID, A3G is not subject to CRM1-dependent nuclear export and therefore that it is unlikely to be a nucleo-cytoplasmic shuttling protein[4]. We additionally used A3G deletion, chimera and single amino acid substitution derivatives to show that the first 60 residues strongly influence the protein's cytoplasmic localization. Several N-terminal amino acids, including aromatic residues Y19 and Y22, appeared particularly important.

Several lines of evidence clearly showed that the first 60 amino acids of A3G contribute to cytoplasmic localization. First, N-terminal amino acid conservation between cytoplasmic A3G and A3F, but not nuclear A3B, implicated this region in such a role. Second, replacing the first 60 amino acids of A3G with the corresponding A3B residues caused the chimeric protein to redistribute to the nucleus. Third, replacing the first 60 residues of A3B with this portion of A3G caused the resulting chimera to become predominantly cytoplasmic. Fourth, single amino acid substitutions in this region of A3G caused the N-terminal half of the protein, which is normally cytoplasmic, to redistribute throughout the cell. Fifth, the critical amino acids so identified clustered on a predicted solvent-exposed surface with the potential to mediate an interaction required for

cytoplasmic localization. These data therefore combined to demonstrate that this N-terminal portion of A3G provides key determinants of cytoplasmic localization.

However, there was also evidence that this 60-residue region alone is not solely responsible. In particular, the single amino acid substitutions that disrupted the localization of the N-terminal half of the protein failed to mislocalize full-length A3G. A recent study from the Smith group indicated that A3G residues 113-128 may also be involved in cytoplasmic localization[5]. Our data showed that, like the Y19 region, this region alone is not sufficient for cytoplasmic localization of full-length A3G, because mutation of five of the predicted solvent-exposed residues within this region had no effect on the protein's cellular distribution. However, two double mutants that combined substitutions within the two critical regions, Y19D-F126S and Y19D-W127A, caused full-length A3G to distribute cell-wide. These data therefore indicated that these two regions cooperate to determine the cytoplasmic localization of full-length A3G. These regions are predicted to be juxtaposed on the same solvent-accessible surface (**Fig. 4-10**). Taken together with prior studies demonstrating the propensity for A3G to multimerize, to bind RNA and to form ribonucleoprotein complexes [25,26,52,68,70,80], it is likely that the surface defined here mediates at least one of these important macromolecular interactions and thereby controls A3G's subcellular distribution.

Experimental Procedures

DNA Constructs: The pEGFP-N3 (Clontech) based plasmids encoding carboxy(C)-terminally green fluorescent protein (GFP)-tagged full-length AID, A3B and A3G were described previously [AID: [132]; A3B and A3G: [153]].

PCR was used to create the truncated A3B and A3G/GFP fusion constructs. The primers were: $A3G_{1-369}$: 5'-NNN NGA GCT CAG GTA CCA CCA TGA AGC CTC ACT TCA GAA AC (RSH431) and 5'-NNN NGT CGA CTT GGC TGT GCT CAT CTA GTC C; A3B-NTD: 5'-NNN NGA GCT CGG TAC CAC CAT GAA TCC ACA GAT CAG AAA T (RSH567) and 5'-NNN NGT CGA CCA TCC TTC CCA GGT ATC TGA GAA TCT CCT TTA G; A3B-CTD: 5'-NNN NGT CGA CCA TCC TTC ACA GGT ATC TGA GAA TCT CCT TTA G and 5'-NNN NGT CGA CCA TCC TTC CGT TTC CCT GAT TCT GGA G (RSH580); A3G-NTD: RSH431 and 5'-NNN NGT CGA CCG AGT GTC TGA GAA TCT CCC CC; A3G-CTD: 5'-NNN GAG CTC AGG TAC CAC CAT GGA TCC ACC CAC ATT CAC TTT C and 5'-NNN GTC GAC TCC GTT TTC CTG ATT CTG GAG AAT (RSH498). PCR products were digested with SacI and SalI and ligated into similarly digested pEGFP-N3.

The A3B/G chimeric constructs were created using overlapping PCR. The primers used were: $A3B_{1-60}G_{60-384}$: RSH567, 5'-CTG GGT GGT ACT TAA GTT CGG AAT ACA CCT GGC CTC GAA AGA C, 5'-TCC GAA CTT AAG TAC CAC CCA G, and 5' RSH498; $A3G_{1-59}B_{61-382}$: RSH431, 5'-GCG TGG TAC TGA GGC TTG AAA TAC ACC TGG CCT CGA AAG AC, 5'-TTC AAG CCT CAG TAC CAC GC, and RSH580; $A3B_{1-190}G_{195-384}$: RSH567, 5'-GTG GGT GGA TCC ATC GAG TGT CTG AGA ATC TCC TTT AGC G, 5'-CAC TCG ATG GAT CCA CCC AC, and RSH498;

$A3G_{1-194}B_{191-382}$: RSH431, 5'-GTG TCT GGA TCC ATC AGG TAT CTG AGA ATC TCC CCC AGC A, 5'-TAC CTG ATG GAT CCA GAC AC, and RSH580. PCR products were digested with SacI and SalI and ligated into a similarly digested pEGFP-N3.

The GFP-tagged truncated and chimeric constructs were subcloned into C-terminal hemagglutinin (HA)-tagged mammalian and bacterial expression vectors. The A3 coding region was cut out of the pEGFP-N3-based plasmids using KpnI and SalI and ligated into KpnI/XhoI digested pcDNA3.1-HA and pTrc99-HA [153].

Amino acid substitution mutants were generated using site-directed mutagenesis (QuikChange, Stratagene; primer sequences available on request). A3G double mutants were created by site-directed mutagenesis using single mutants as PCR templates. pEGFP-N3-A3G-NTD mutants were subcloned as KpnI/RsrII fragments into similarly digested full-length A3G expression constructs pEGFP-N3-A3G and pcDNA3.1-A3G-HA, reported previously [153]. All constructs were confirmed by restriction digestion and DNA sequencing.

Cell Culture: Human embryonic kidney 293T and HeLa cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 25 units/ml penicillin and 25 μg/ml streptomycin at 37 ˚C and 5% CO_2. All transfections used FuGENE (Roche Applied Science) or TransIT-LT1 (Mirus Bio Corporation) according to the manufacturers' protocols.

Microscopy: HeLa or 293T cells were seeded on 8-well LabTek chambered cover glasses (Nunc). After 24h of incubation, the cells were transfected with the indicated pEGFP-N3-based constructs. After an additional 24h of incubation, images of the live cells were captured on a Zeiss Axiovert 200 microscope. Images were cropped and their

brightness and contrast adjusted linearly using ImageJ software (http://rsb.info.nih.gov/ij/). For leptomycin B experiments, cells were treated for two hours with 20 ng/ml leptomycin B (LC Laboratories) or mock treated with an equivalent dilution of ethanol (the vehicle).

For immunofluorescence experiments, HeLa cells were seeded onto sterilized coverglasses set in culture dishes. After 24h of incubation, the cells were transfected with the indicated pcDNA3.1-HA-based constructs. After an additional 24h of incubation, cells were fixed with 4% formaldehyde, permeabilized in 0.2% Triton X-100, incubated, first, with anti-HA antibody (Covance), and, second, with fluorescein-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch), and imaged as above.

***E. coli* Mutation Assays:** An *E. coli*-based rifampicin-resistance assay was used to quantify the intrinsic DNA cytidine deaminase activity of A3B, A3G and chimeric and mutant derivatives. This assay has been described previously [*e.g.*, [56,60]].

Immunoblotting: 293T cells were plated in 6-well plates. After 24h of incubation, the cells were transfected with APOBEC expression plasmid (or appropriate control plasmids). 36h later, the cells were harvested and total proteins extracted in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 1% Nonidet P-40, 0.1% sodium dodecyl sulfate, 0.5% sodium deoxycholate, and a protease inhibitor cocktail (Roche). The extracts were clarified by centrifugation for 10 minutes at 20,800g at 4˚C. The extracted proteins were fractionated by SDS-PAGE, transferred to a polyvinylidene difluoride membrane (Millipore), and probed with monoclonal anti-HA (Covance) or monoclonal anti-GFP (Clontech). Primary antibodies were detected by incubation with horseradish peroxidase-coupled anti-mouse IgG (Bio-Rad) or anti-rabbit IgG (Bio-Rad) followed by

chemiluminescence (Roche Applied Science). Ponceau S (Sigma) staining of the PVDF membrane following immunoblotting served as a protein loading control.

Amino acid alignments: The amino acids sequences of A3B, A3F and A3G were aligned using ClustalW and edited manually. The protein sequences correspond to GenBank accession numbers NP_004891, NP_660341 and NP_068594, respectively.

Structural modeling of A3Gntd: The primary amino acid sequences of A3G-NTD (residues 1-196) and A3G-CTD (residues 197-384) were aligned using the homology modeling module of the InsightII program (Accelrys). These amino acid sequences are 35% identical and 52% similar (according to the bl2seq program, NCBI). The secondary structure elements of A3G-NTD were predicted by projecting the actual secondary structure of A3G-CTD [PDB 2jyw; [20,57]] over the A3G-NTD/A3G-CTD primary amino acid sequence alignment. Because of amino acid conservation in the corresponding regions, the secondary structure elements of A3G-NTD were predicted to correspond directly with the actual elements in the A3G-CTD (**Fig. 4-5**). Finally, this information was used to construct a three-dimensional model of the A3G-NTD (InsightII program, Accelrys). Images of the model were created using MacPyMol (DeLano Scientific).

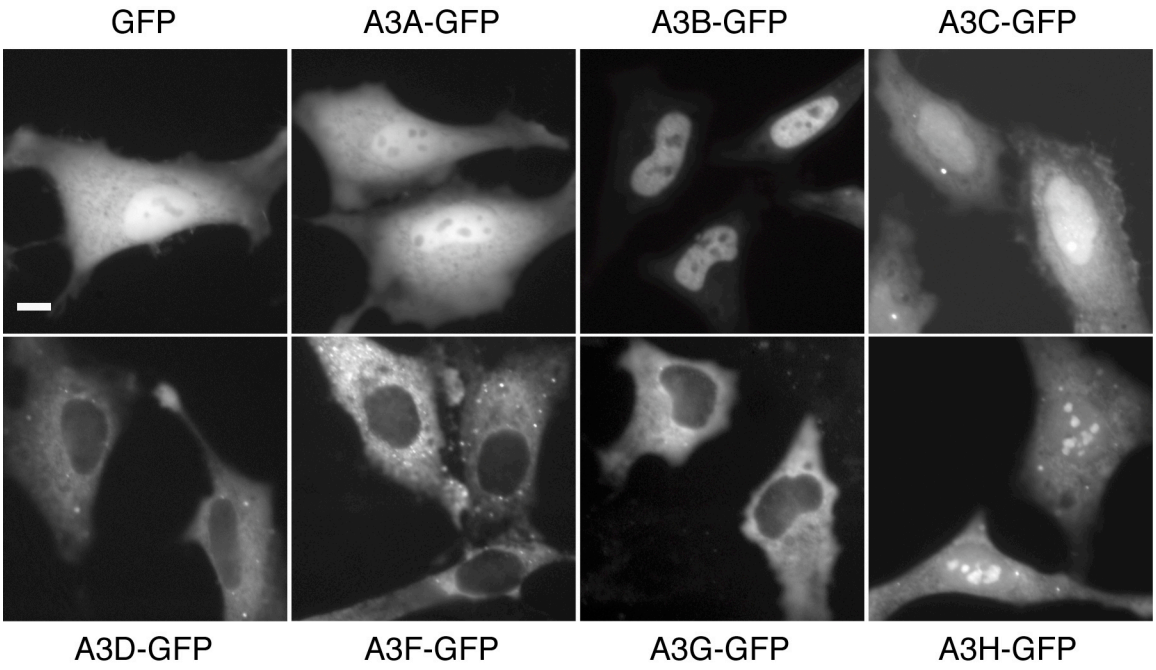

Fig. 4-1. A3 subcellular localization. Images of representative live HeLa cells transiently transfected with the indicated A3-GFP expression plasmids. The scale bar indicates 10 μm.

Fig. 4-2. An alignment of A3B, A3F and A3G amino acid sequences, highlighting regions of concentrated identity. Blue and green shading indicates regions of concentrated identity between A3G and A3F and A3B and A3F, respectively. Gray-shaded residues are identical between at least two of the proteins. The crossing arrows indicate the junctions of chimeric proteins. Dashed boxes outline the putative cytoplasmic retention signal (CRS) and nuclear export signal (NES) of A3G and the putative nuclear localization signal (NLS) of A3B. The histidine, glutamic acid and cysteine residues of the two zinc-coordinating motifs are indicated in bold. The amino acids deleted in $A3G_{1-369}$ are underlined. Additional details can be found in the text.

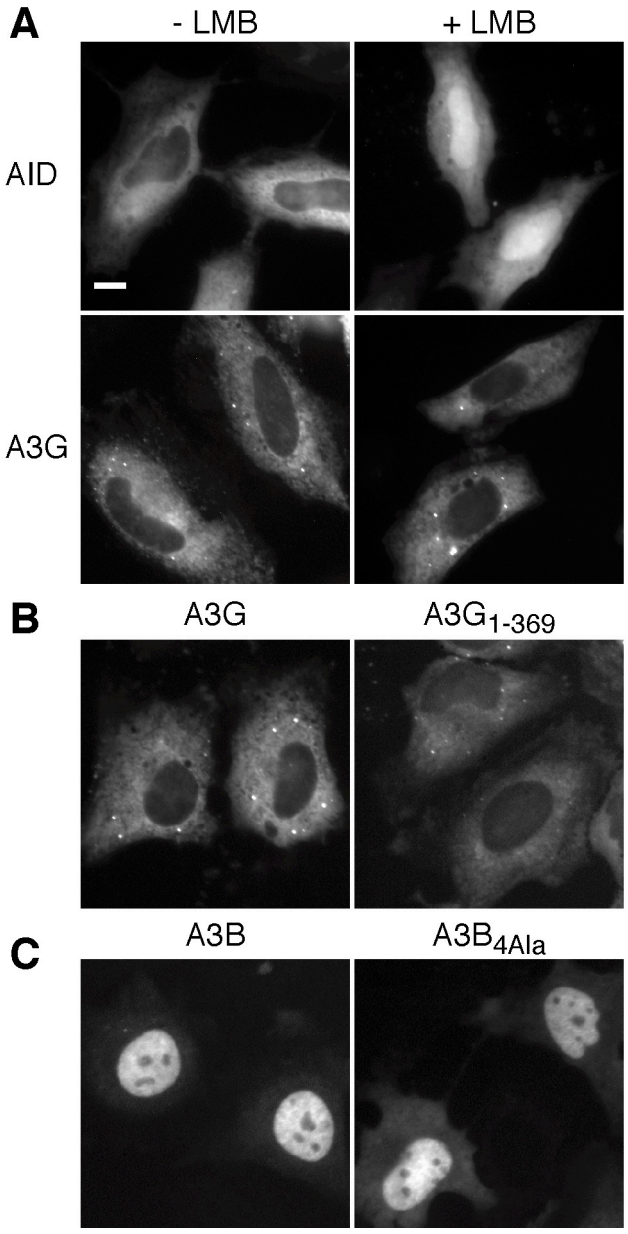

Fig. 4-3. Predicted nucleo-cytoplasmic shuttling signals in A3G and A3B are not required for their subcellular distributions. Images of representative live HeLa cells transiently transfected with the indicated A3-GFP expression plasmids. The scale bar indicates 10 μm. (A) LMB caused AID-GFP but not A3G-GFP to accumulate in the nucleus. Minus (-)LMB indicates mock-treated cells. (B) Localization of A3G-GFP and the $A3G_{1-369}$.GFP mutant lacking the putative C-terminal NES. (C) Localization of wild type A3B and a mutant A3B with four alanines replacing residues in the putative NLS ($A3B_{4Ala}$).

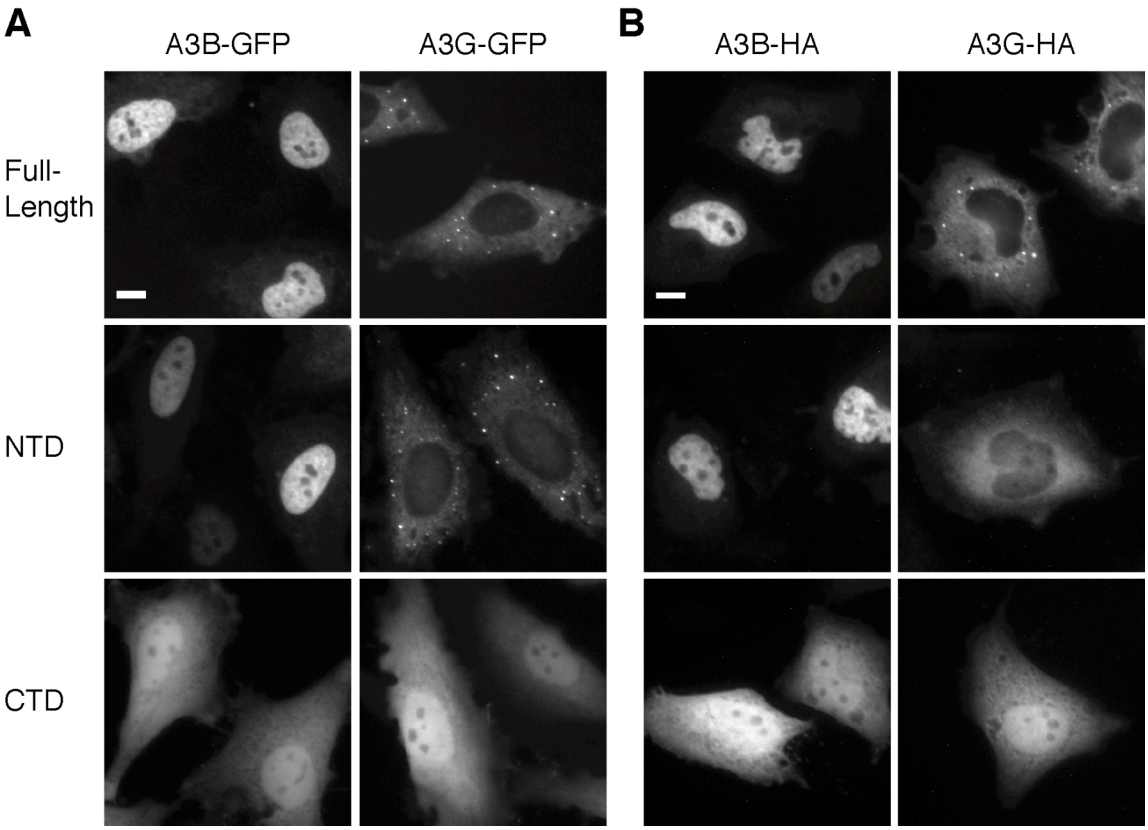

Fig. 4-4. The N-terminal halves of A3B and A3G recapitulate the localization patterns of the full-length proteins. Representative images of HeLa cells transfected with the indicated A3-GFP or A3-HA expression plasmids. The NTD of A3B and A3G consists of amino acids 1-192 and 1-196, respectively. The CTD of A3B and A3G consists of amino acids 193-382 and 197-384, respectively. (A) Live cells expressing the indicated GFP-fusion proteins. (B) Fixed cells expressing the indicated HA-fusion proteins. The scale bar indicates 10 μm.

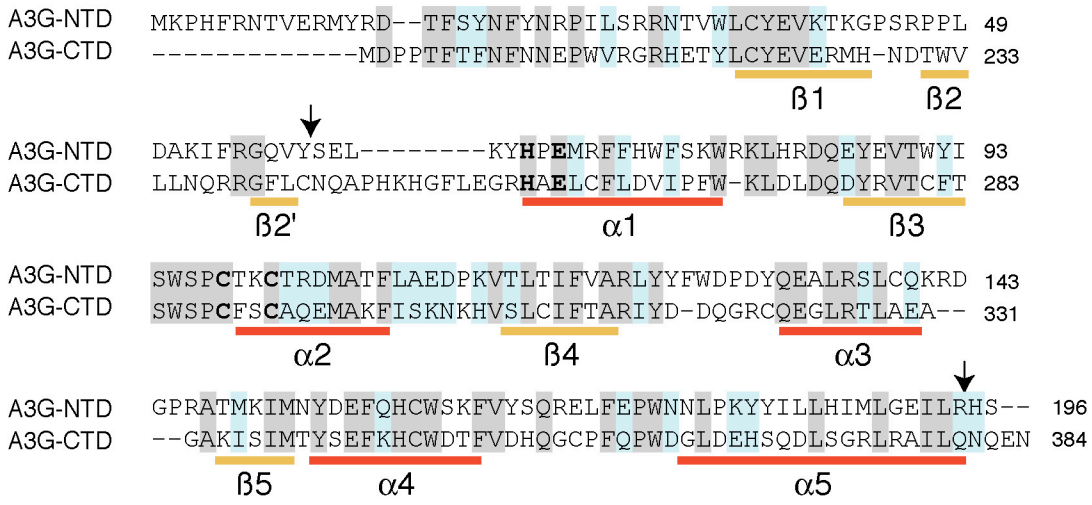

Fig. 4-5. Alignment of the amino acid sequences of A3G-NTD and A3G-CTD. The amino acid alignment between A3G-NTD (residues 1-196) and A3G-CTD (residues 197-384) used to create the A3G-NTD model structure is shown. Identical amino acids are shaded gray and similar amino acids blue. The beta strands and alpha helices of the actual A3G-CTD NMR structure are illustrated for comparison [20]. Arrows indicate chimera crossover points (described in text).

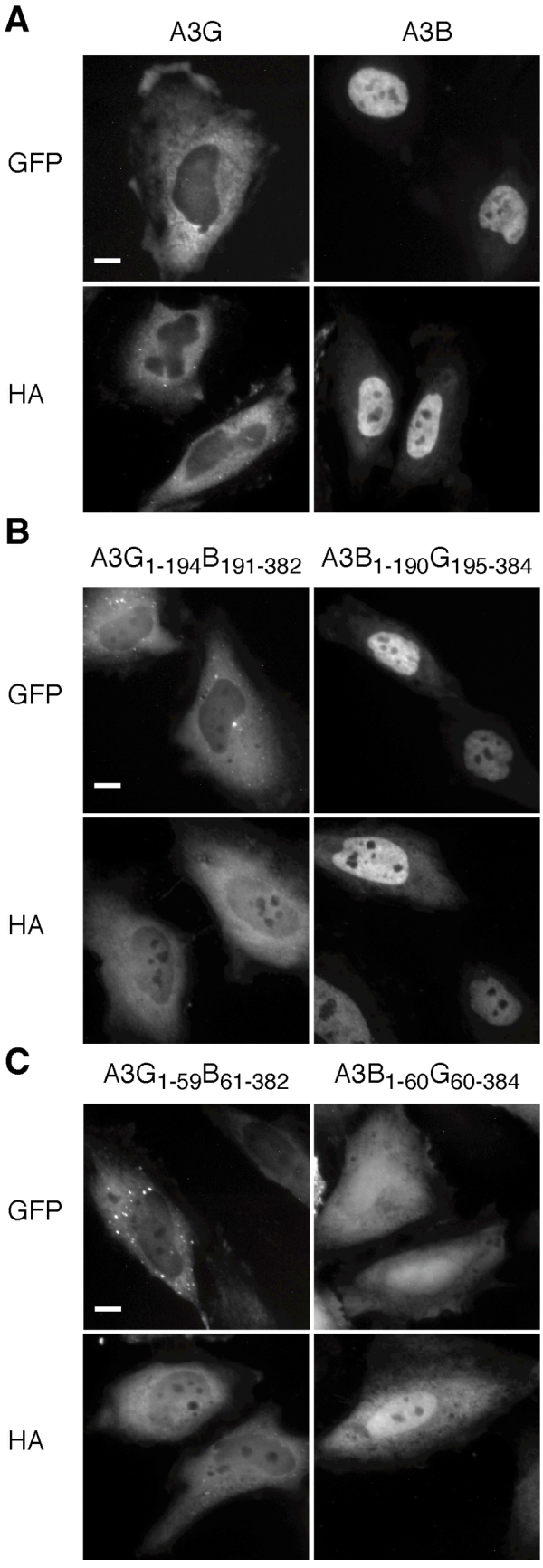
A
A3G
A3B
GFP
HA
B
A3G1-194B191-382
A3B1-190G195-384
GFP
HA
C
A3G1-59B61-382
A3B1-60G60-384
GFP
HA

Fig. 4-6. The first 60 amino acids of A3B and A3G possess essential subcellular localization determinants. Representative images of HeLa cells transfected with the indicated chimeric-A3 expression plasmids. The GFP-tagged constructs were visualized in live cells and the HA-tagged proteins were visualized in fixed cells. (A) Cells expressing A3B or A3G. (B) Cells showing the localization of chimeras that fuse the NTD of A3B with the CTD of A3G and *vice-versa*. (C) Cells showing the localization of chimeras that swap the first 60 amino acids of A3B for those of A3G and *vice-versa*. The scale bar indicates 10 μm.

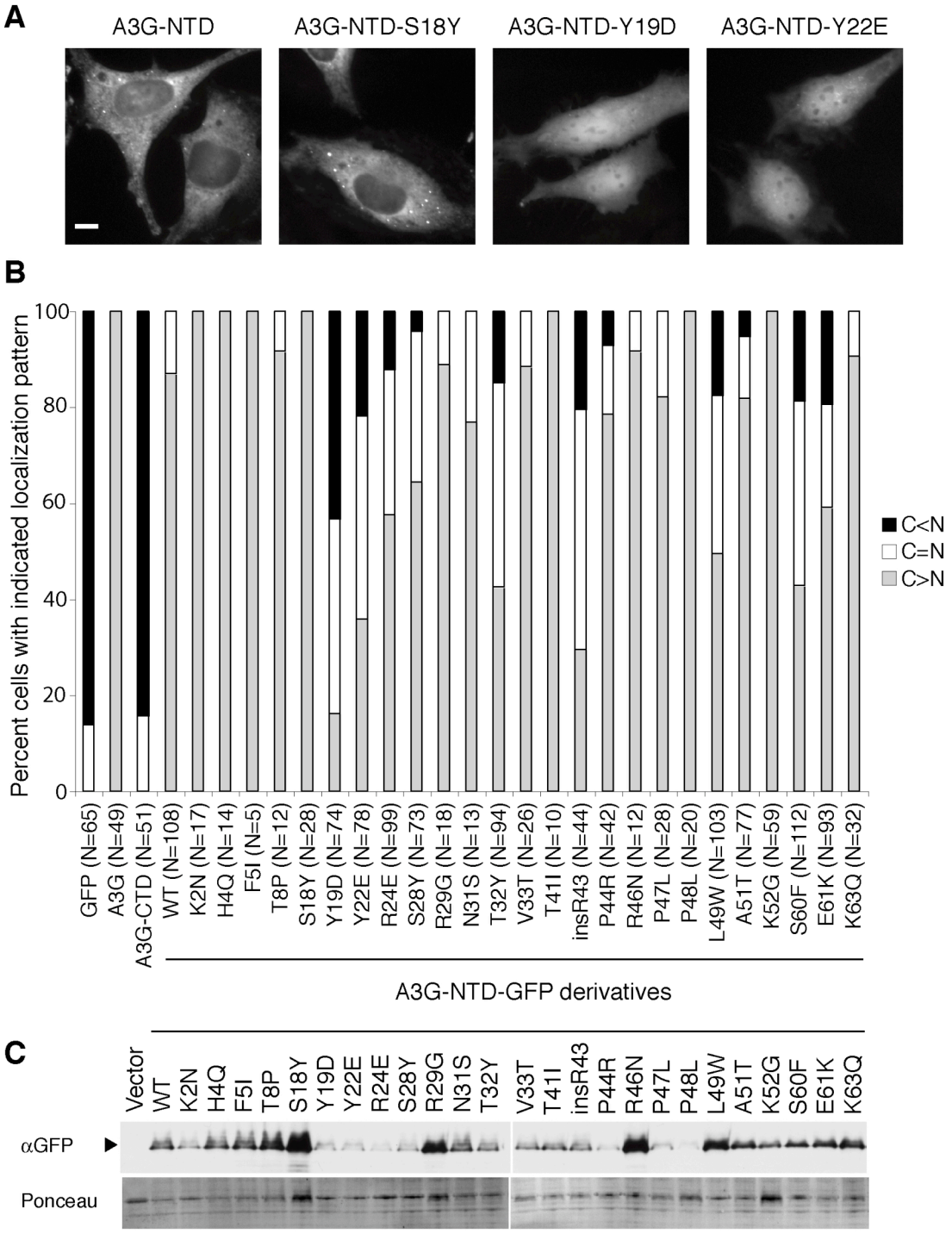
A
A3G-NTD
A3G-NTD-S18Y
A3G-NTD-Y19D
A3G-NTD-Y22E
B
Percent cells with indicated localization pattern
100
80
60
40
20
0
C<N
C=N
C>N
GFP (N=65)
A3G (N=49)
A3G-CTD (N=51)
WT (N=108)
K2N (N=17)
H4Q (N=14)
F5I (N=5)
T8P (N=12)
S18Y (N=28)
Y19D (N=74)
Y22E (N=78)
R24E (N=99)
S28Y (N=73)
R29G (N=18)
N31S (N=13)
T32Y (N=94)
V33T (N=26)
T41I (N=10)
insR43 (N=44)
P44R (N=42)
R46N (N=12)
P47L (N=28)
P48L (N=20)
L49W (N=103)
A51T (N=77)
K52G (N=59)
S60F (N=112)
E61K (N=93)
K63Q (N=32)
A3G-NTD-GFP derivatives
C
Vector
WT
K2N
H4Q
F5I
T8P
S18Y
Y19D
Y22E
R24E
S28Y
R29G
N31S
T32Y
V33T
T41I
insR43
P44R
R46N
P47L
P48L
L49W
A51T
K52G
S60F
E61K
K63Q
αGFP
Ponceau

Fig. 4-7. Single amino acid substitutions within the first 60 amino acids of A3G disrupt cytoplasmic localization. (A) Representative images of live HeLa cells transiently transfected with the indicated A3G-NTD-GFP or derivative expression plasmids. The scale bar indicates 10 μm. (B) A quantification of the extent of disrupted localization for the mutants. Individual cells expressing the indicated proteins were scored and grouped into three categories based on their overall pattern of cellular fluorescence: those with more apparent nuclear than cytoplasmic fluorescence (C < N), equivalent nuclear and cytoplasmic fluorescence (C = N), or more apparent cytoplasmic than nuclear fluorescence (C > N). The percentage of cells falling into each category is indicated for each mutant. Cells from two to four independent experiments were scored and the tallies were combined. The total number of cells (N) scored for each construct is indicated. (C) An anti-GFP immunoblot showing the expression of the indicated A3G-NTD-GFP proteins. An arrowhead indicates the band corresponding to A3G-NTD-GFP and mutant derivatives. Ponceau S staining of the membrane served as a protein loading control.

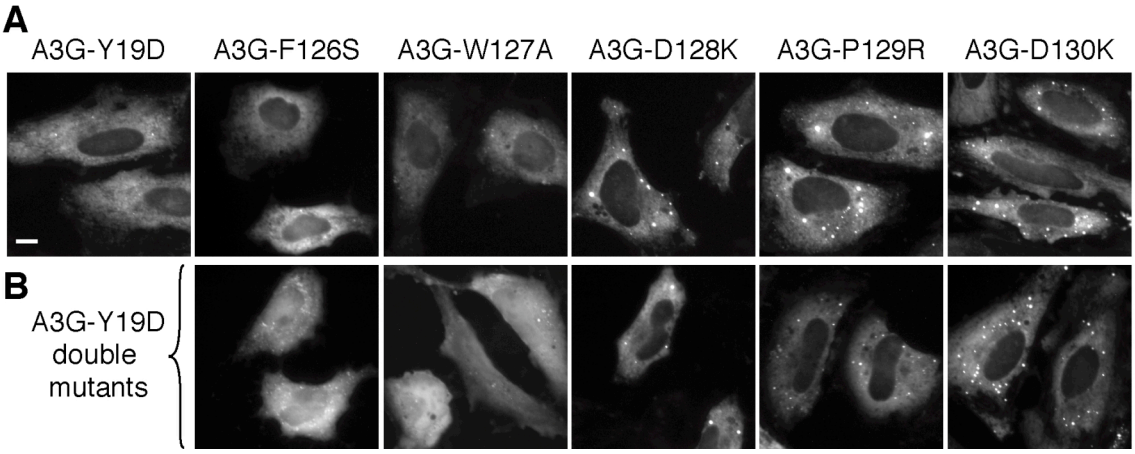

Fig. 4-8. Two regions within the N-terminal half of A3G cooperatively determine the cytoplasmic localization of the full-length protein. Representative images of live HeLa cells transfected with the indicated A3-GFP expression plasmids. The scale bar indicates 10 μm. (A) Representative cells showing the localization of the indicated (A) full-length A3G single mutants, or (B) full-length A3G double mutants.

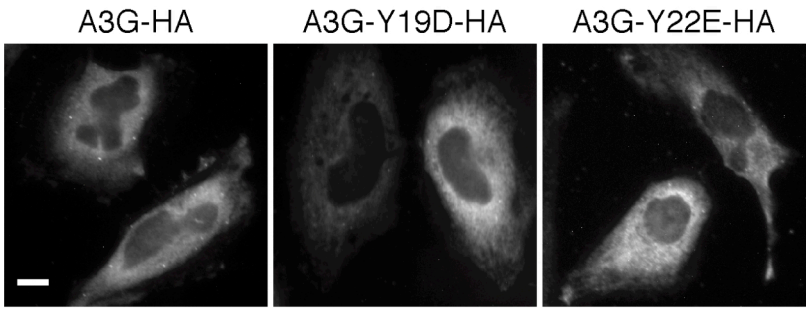

Fig. 4-9. Y19D and Y22E do not significantly alter the localization of the full-length A3G. Representative images of HeLa cells transfected with the indicated A3-HA expression plasmids. Protein localization was visualized by immunofluorescence in fixed cells. The scale bar indicates 10 μm.

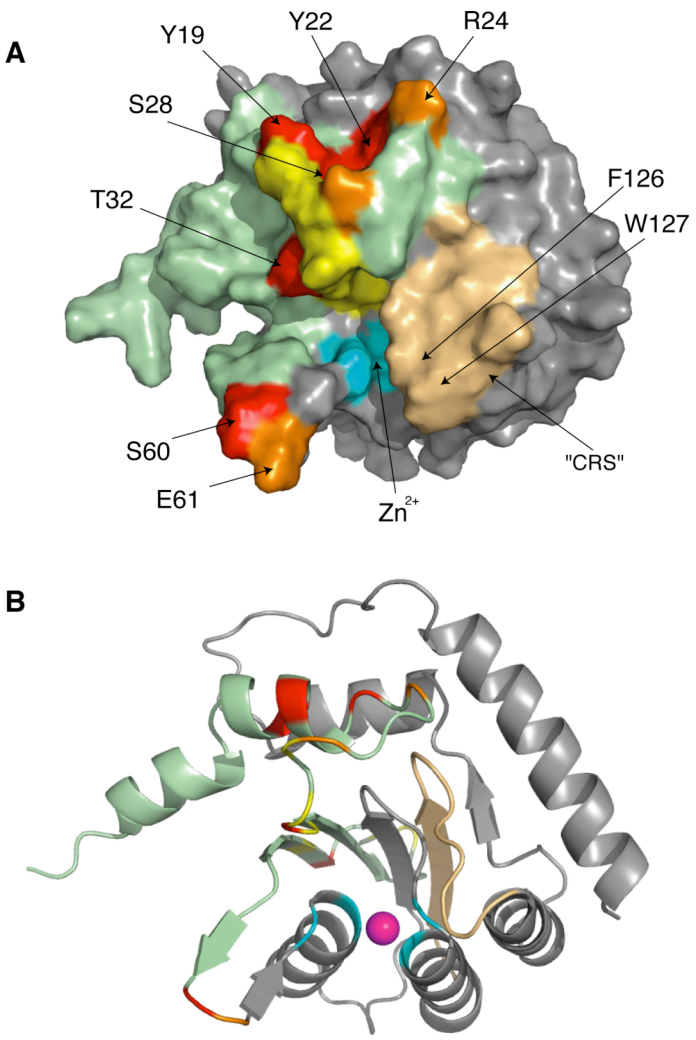

Fig. 4-10. A predicted structure of the N-terminal half of A3G highlighting residues that influence cytoplasmic localization. A predicted three-dimensional structure of A3G-NTD, based on NMR structures of A3G-CTD [20,57]. A3G residues 1-60 are colored light green. The conserved zinc-coordinating residues H65, E67, C97 and C100 are colored cyan and labeled (Zn^{2+}). Residues within the first 60 amino acids whose mutation results in a mislocalization of A3G-NTD-GFP are highlighted in red, orange and yellow according to the degree of mislocalization (>50%, 25-50% or 10-25% respectively in

Fig. 4-7B). The previously reported CRS [5] is colored tan and critical residues within this motif are indicated (F126 and W127). Surface (A) and ribbon (B) representations are shown. Zinc ion in (B) is colored magenta. Two residues with a significantly altered localization pattern [L49W and ins(R42)] were not predicted to be on the same surface, but it is probable that these mutations perturb the protein's structure.

Chapter 5: Conclusions and Discussion

The major findings of this thesis, that A3 proteins inhibit LINE-1 retrotransposition, and that A3 proteins restrict foreign DNA have been discussed in detail in chapters 2 and 3. The following is a discussion of several of the unifying themes of the thesis, a discussion of some of the most important unresolved questions raised by these studies, and future experimental directions that could be used to address these questions.

The relationship between A3 subcellular localization and function

The most striking (or at least the most visual) difference between the A3s is their different subcellular localization patterns. These different cellular distributions seem like they should reflect differences in function. And indeed, A3s perform different functions, or at least possess different activities. For instance, some A3s, like A3F and A3G, are particularly good at inhibiting HIV replication. Others, like A3A and A3B are potent limiters of LINE-1 retrotransposition. But, there does not seem to be a good correlation between subcellular localization and these other activities. One likely explanation for this is that the coarse localization patterns determined by visualizing overexpressed, epitope-tagged A3s do not accurately represent the full range of cellular compartments to which these proteins have access. Therefore, although it is tempting to base hypotheses about A3 function on coarse A3 localization patterns (as we did when originally undertaking the LINE-1 inhibition studies), this approach has the potential to be misleading.

One hypothesis that would seem to still be based on good rationale, however, is that A3 proteins that are sequestered in the cytoplasm pose less of a risk to genomic DNA than the nuclear A3s. This idea is supported by the observation that the only A3 with a predominant nuclear localization, A3B, is being lost from some human populations[76]. The worldwide allele frequency of the A3B deletion is 22%, and there are some populations (*e.g.*, Micronesians), that are essentially A3B null. This suggests that the risk posed by this nuclear DNA-mutating enzyme may outweigh its beneficial functions.

Deamination-dependent vs. deamination-independent A3 activities

Another potentially confusing characteristic of the A3s is that in some cases their activities seem to depend on their catalytic activity and in other cases they apparently function in a deamination-independent fashion. For instance, catalytically inactive A3B decreases the rate of LINE-1 retrotransposition to the same extent as catalytically competent A3B. Thus, LINE-1 inhibition does not depend on catalytic activity, which is one of the major conclusions of **chapter 2**. In contrast, the foreign-DNA restriction activity of A3A appears to depend on deaminase activity **(chapter 3)**.

There are several possible explanations for this apparent contradiction. The most straightforward is that LINE-1 inhibition and foreign DNA restriction are fundamentally different activities and one requires deaminase activity and one doesn't. Another possibility is that LINE-1 inhibition is indeed a deamination-dependent process and that the A3B mutant that was used maintains a very low level of deaminase activity that is undetectable in our deaminase assays but is sufficient to inhibit retrotransposition. The third possible explanation is that foreign DNA restriction does not require deaminase

activity, but a separation of function mutant that demonstrates this fact has not been discovered.

The experimental complications of DNA restriction

One implication of the findings of **chapter 3** is that transient transfection experiments employing A3s are fraught with potential complications. For instance, A3A mutates its own expression plasmid (**Fig. 3-3**), which likely attenuates its own expression, and dampens its own effects in such experiments. This may explain why DNA restriction in primary monocytes appears to be significantly more potent than that in 293 cells expressing A3A from a plasmid (compare **Fig. 3-7** & **Fig. 3-8**). Other A3s appear to have similar activity and likely mutate *any* transfected plasmids, not just their own expression construct.

Does knowing this nullify the conclusions of **chapter 2**, that A3s inhibit LINE-1 retrotransposition? This is a possibility because the LINE-1 assay is plasmid-based, so it is possible that A3s are simply destabilizing the LINE-1 plasmid, and not blocking retrotransposition *per se*.

There are two main arguments against this possibility. First, retrotransposition rates were measured in cells harboring plasmid DNA. This is true because the LINE-1 plasmid contains a puromycin^{R} gene and an autonomous replication origin, and the transfected cells were selected in puromycin. This was originally done to mitigate the possible effect of differing transfection efficiencies between the different clonal cell lines. But it also means that retrotransposition rates were being measured in drug resistant cells that were therefore harboring plasmid DNA. Of course it is possible to imagine scenarios

where the portion of the plasmid containing the puromycin^{R} gene was maintained intact in the cells but the LINE-1 portion of the plasmid was mutated or lost.

The second main argument that LINE-1 inhibition and DNA restriction are fundamentally different activities is that the former does not require A3 catalytic activity whereas the latter does. This different requirement suggests a major difference in mechanism.

Nevertheless, to conclusively demonstrate that A3s limit LINE-1 retrotransposition, a new system measuring retrotransposition from a genomic LINE-1 element would have to be developed. Alternatively, it would be possible to take advantage of the fact that some humans are A3B deficient (due to the deletion described above)[76]. In principle, it would be possible to assemble cohorts of A3B-proficient and –deficient people and determine the frequency of novel LINE-1 insertions in these different populations. This can be done using high-throughput DNA sequencing or whole genome tiling array technologies. In any case, this example illustrates the caution that must be taken when interpreting A3-related results based on transient transfection experiments.

Other DNA restriction systems

That A3 proteins recognize and eliminate foreign DNA is the most important finding of this thesis. This discovery can be put in the context of other DNA defense systems that operate in eukaryotic and prokaryotic cells.

In animal cells, one of the most potent DNA defenses is that provided by the DNase-II enzymes[43]. DNase-II homologs have been identified in many vertebrate and invertebrate species. These endonucleases cleave the DNA phosphodiester backbone,

introducing single-stranded nicks. DNase-II localizes to endosomal compartments and is thought to facilitate the degradation of DNA that enters the cell by endocytosis, the major way that most foreign material enters cells. One obvious difference between nucleases such as DNase-II and DNA deaminases such as A3s is that nucleases actually cleave the DNA molecule; A3s introduce a relatively subtle modification. Another major difference is their localization: different A3 proteins occupy essentially every cellular compartment (**Fig. 4-1**), and therefore have access to many more DNA substrates than endosomal DNase-II proteins. Therefore, A3s have the potential to engage nuclear and cytosolic retroelement replication intermediates, but DNase-II enzymes are probably not properly situated to do so. Another difference is that DNase-II enzymes are ubiquitously and constitutively expressed, whereas A3 expression is restricted to certain tissues and is inducible by cytokines such as IFN.

Another IFN-inducible DNA-defense system is provided by the exonuclease TRex1[155]. Unlike A3A, TRex1 is expressed ubiquitously, but like A3A it is induced by IFN following foreign DNA detection. Similar to some A3 family members, TRex1 localizes to the cytosol of cells. TRex1-deficient mice die prematurely, apparently of a toxic autoimmune response to due the accumulation of intracellular DNA[155]. Also, like some A3s, it has been suggested that TRex1 operates on DNA generated by retroelement reverse transcription[155]. Thus, in addition to A3 family members, TRex1 may be an important DNA-modifying enzyme that facilitates the clearance of intracellular foreign and retroelement DNA.

The A3s, TRex1, and DNaseII operate in eukaryotic cells. In bacterial cells, anti-DNA defense systems are provided by restriction endonucleases[177]. These sequence-

specific endonucleases are typically co-expressed with cognate DNA methyl transferases, which mark the cell's own DNA by methylating it. The endonucleases cleave DNA lacking the protective methylation, thereby neutralizing non-self intracellular DNA. This system is analogous to the A3-mediated anti-foreign-DNA mechanism described here. Thus, the recognition and clearance of foreign DNA is a characteristic of prokaryotic and eukaryotic life forms.

Going forward

The studies described in this thesis create as many questions as they answer. Some of the most interesting of these unresolved questions relate to the finding that A3s restrict foreign DNA.

One such question is whether cytidine deamination activity is absolutely required to restrict foreign DNA. This question arises because in other contexts, such as LINE-1 inhibition, deaminase activity is not required (nucleic acid binding may suffice). Our data suggest that deamination is required, but several experimental approaches could fortify this conclusion. One approach could use a panel of hypomorphic A3A mutants that we have generated. These mutants exhibit a range of intrinsic deaminase activity ranging from dead to fully active. Demonstrating a strong correlation between deaminase activity and DNA restriction would support the conclusion that restriction requires deamination. (Showing that these mutants bound DNA with equal affinities would be a necessary control). A second approach would rely on inhibitors of A3 deaminase activity. Currently, our laboratory is conducting a screen of a large library of chemical compounds for small molecules that inhibit deaminase activity. Such compounds would be useful if they inhibited deamination but left nucleic acid binding activity unaltered. It is likely

that these studies will conclusively demonstrate that A3 deamination activity is indeed required for foreign DNA restriction.

Another question relates to the generality of A3-mediated DNA restriction. Although this activity is pronounced in primary monocytes, it is possible that A3s act against foreign DNA in a variety of human cell types. As a first step towards testing this, it will be necessary to determine in which cells A3s are expressed. This could be done by applying the panel of q-RT-PCR assays described in **chapter 3** to RNA samples from a panel of tissues. It would also be straightforward to transfect plasmid DNA into a variety of primary human cell types, and then recover the DNA and subject it to 3D-PCR to detect editing. It would be particularly interesting to determine whether this occurs in dermal fibroblasts, the cells most commonly used to create induced pluripotent stem cells.

A related question dealing with the breadth of DNA-editing mediated DNA restriction is whether it extends beyond humans and beyond A3s. All mammalian species examined to date contain A3 proteins, and orthologs of A3A, the human family member with the most potent DNA restriction activity, exist in other mammals (**Fig. 1-4**). It is likely that these A3 proteins restrict foreign DNA in these other species. But editing-mediated DNA restriction need not be limited to A3s. All vertebrates contain related cytidine deaminases (such as AID) and it is possible that these enzymes restrict foreign DNA in non-mammalian species by a similar mechanism.

Perhaps the most important unanswered A3-related question is: how are these enzymes specifically targeted to foreign and retroelement DNA and not (presumably) to genomic DNA? Three simple models can be proposed. One is that foreign DNA may be

recognizably foreign, for instance by different methylation patterns. In this model (the "sentinel" model), A3A or a targeting cofactor recognizes the foreign characteristic. A cofactor could be an enzyme such as a DNA helicase that unwinds the foreign dsDNA, rendering it susceptible to editing by A3s. In the second model (the "failsafe" model), self DNA is protected by associated factors or epigenetic markings. DNA lacking this marking is attacked by A3s. This latter mechanism would be analogous to prokaryotic restriction/modification systems, in which self DNA is protected from restriction by a nucleic acid modification (usually methylation). The third model (the "location" model) is that any DNA in certain cellular compartments (for example the cytosol) is subject to A3-mediated editing. Experiments using synthetic or chromatinized DNA substrates and localization-restricted A3 chimeras could help differentiate between these models.

Concluding Remarks

The A3 paradox is that they are DNA mutating enzymes that protect the cell. This paradox is explainable by the observation that A3 deamination seems to be restricted to DNA substrates that threaten the cell's well-being. Here, I have significantly expanded the notion of what constitutes an A3 target, by showing that A3 proteins act against LINE-1 retrotransposons and foreign DNA in general. The latter finding is particularly important because A3-mediated DNA restriction may reduce the efficacy of therapies reliant on DNA transfer. Such therapies include gene therapy and the generation of induced pluripotent stem cells. Because of their potential to cure previously intractable diseases, these therapies rank among the most promising on medicine's horizon. Transient administration of A3-inhibiting drugs may facilitate the delivery of beneficial DNA and could help convert therapeutic promise into reality.

References

1. Arakawa, H., J. Hauschild, and J.M. Buerstedde, *Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion.* Science, 2002. **295**(5558): p. 1301-6.
2. Armitage, A.E., A. Katzourakis, T. de Oliveira, J.J. Welch, R. Belshaw, K.N. Bishop, B. Kramer, A.J. McMichael, A. Rambaut, and A.K. Iversen, *Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences.* J Virol, 2008. **82**(17): p. 8743-61.
3. Bass, B.L. and H. Weintraub, *An unwinding activity that covalently modifies its double-stranded RNA substrate.* Cell, 1988. **55**(6): p. 1089-98.
4. Bennett, R.P., E. Diner, M.P. Sowden, J.A. Lees, J.E. Wedekind, and H.C. Smith, *APOBEC-1 and AID are nucleo-cytoplasmic trafficking proteins but APOBEC3G cannot traffic.* Biochem Biophys Res Commun, 2006. **350**(1): p. 214-9.
5. Bennett, R.P., V. Presnyak, J.E. Wedekind, and H.C. Smith, *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, 2008. **283**(12): p. 7320-7.
6. Bieniasz, P.D., *Intrinsic immunity: a front-line defense against viral attack.* Nat Immunol, 2004. **5**(11): p. 1109-15.
7. Bishop, K.N., R.K. Holmes, and M.H. Malim, *Antiviral potency of APOBEC proteins does not correlate with cytidine deamination.* J Virol, 2006. **80**(17): p. 8450-8.
8. Bishop, K.N., R.K. Holmes, A.M. Sheehy, N.O. Davidson, S.J. Cho, and M.H. Malim, *Cytidine deamination of retroviral DNA by diverse APOBEC proteins.* Curr Biol, 2004. **14**(15): p. 1392-6.
9. Bogerd, H.P., B.P. Doehle, H.L. Wiegand, and B.R. Cullen, *A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor.* Proc Natl Acad Sci U S A, 2004. **101**(11): p. 3770-4.
10. Bogerd, H.P., H.L. Wiegand, B.P. Doehle, K.K. Lueders, and B.R. Cullen, *APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.* Nucleic Acids Res, 2006. **34**(1): p. 89-95.
11. Bogerd, H.P., H.L. Wiegand, A.E. Hulme, J.L. Garcia-Perez, K.S. O'Shea, J.V. Moran, and B.R. Cullen, *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8780-5.
12. Bonvin, M., F. Achermann, I. Greeve, D. Stroka, A. Keogh, D. Inderbitzin, D. Candinas, P. Sommer, S. Wain-Hobson, J.P. Vartanian, and J. Greeve, *Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication.* Hepatology, 2006. **43**(6): p. 1364-74.
13. Brar, S.S., M. Watson, and M. Diaz, *Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks.* J Biol Chem, 2004. **279**(25): p. 26395-401.
14. Brouha, B., J. Schustak, R.M. Badge, S. Lutz-Prigge, A.H. Farley, J.V. Moran, and H.H. Kazazian, Jr., *Hot L1s account for the bulk of retrotransposition in the human population.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5280-5.
15. Browne, E.P., C. Allers, and N.R. Landau, *Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent.* Virology, 2009. **387**(2): p. 313-21.
16. Burnett, A. and P. Spearman, *APOBEC3G Multimers Are Recruited to the Plasma Membrane for Packaging into Human Immunodeficiency Virus Type 1 Virus-Like Particles in an RNA-Dependent Process Requiring the NC Basic Linker.* J Virol, 2007. **81**(10): p. 5000-13.
17. Chelico, L., P. Pham, P. Calabrese, and M.F. Goodman, *APOBEC3G DNA deaminase acts processively 3′ --> 5′ on single-stranded DNA.* Nat Struct Mol Biol, 2006. **13**(5): p. 392-9.
18. Chen, H., C.E. Lilley, Q. Yu, D.V. Lee, J. Chou, I. Narvaiza, N.R. Landau, and M.D. Weitzman, *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.* Curr Biol, 2006. **16**(5): p. 480-5.
19. Chen, K., J. Huang, C. Zhang, S. Huang, G. Nunnari, F.X. Wang, X. Tong, L. Gao, K. Nikisher, and H. Zhang, *Alpha interferon potently enhances the anti-human immunodeficiency virus type 1 activity of APOBEC3G in resting primary CD4 T cells.* J Virol, 2006. **80**(15): p. 7645-57.

20. Chen, K.M., E. Harjes, P.J. Gross, A. Fahmy, Y. Lu, K. Shindo, R.S. Harris, and H. Matsuo, *Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G.* Nature, 2008. **452**(7183): p. 116-9.
21. Chen, K.M., N. Martemyanova, Y. Lu, K. Shindo, H. Matsuo, and R.S. Harris, *Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G.* FEBS Lett, 2007. **581**(24): p. 4761-6.
22. Chen, S.H., G. Habib, C.Y. Yang, Z.W. Gu, B.R. Lee, S.A. Weng, S.R. Silberman, S.J. Cai, J.P. Deslypere, M. Rosseneu, and et al., *Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon.* Science, 1987. **238**(4825): p. 363-6.
23. Chester, A., A. Somasekaram, M. Tzimina, A. Jarmuz, J. Gisbourne, R. O'Keefe, J. Scott, and N. Navaratnam, *The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay.* Embo J, 2003. **22**(15): p. 3971-82.
24. Chiu, Y.L. and W.C. Greene, *The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements.* Annu Rev Immunol, 2008. **26**: p. 317-353.
25. Chiu, Y.L., V.B. Soros, J.F. Kreisberg, K. Stopak, W. Yonemoto, and W.C. Greene, *Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells.* Nature, 2005. **435**(7038): p. 108-14.
26. Chiu, Y.L., H.E. Witkowska, S.C. Hall, M. Santiago, V.B. Soros, C. Esnault, T. Heidmann, and W.C. Greene, *High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15588-93.
27. Conticello, S.G., *The AID/APOBEC family of nucleic acid mutators.* Genome Biol, 2008. **9**(6): p. 229.
28. Conticello, S.G., R.S. Harris, and M.S. Neuberger, *The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G.* Curr Biol, 2003. **13**(22): p. 2009-13.
29. Conticello, S.G., M.A. Langlois, and M.S. Neuberger, *Insights into DNA deaminases.* Nat Struct Mol Biol, 2007. **14**(1): p. 7-9.
30. Conticello, S.G., C.J. Thomas, S.K. Petersen-Mahrt, and M.S. Neuberger, *Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases.* Mol Biol Evol, 2005. **22**(2): p. 367-77.
31. Cui, Z., A.M. Geurts, G. Liu, C.D. Kaufman, and P.B. Hackett, *Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon.* J Mol Biol, 2002. **318**(5): p. 1221-35.
32. Cullen, B.R., *Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors.* J Virol, 2006. **80**(3): p. 1067-76.
33. Dang, Y., X. Wang, W.J. Esselman, and Y.H. Zheng, *Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family.* J Virol, 2006. **80**(21): p. 10522-33.
34. Deininger, P.L., J.V. Moran, M.A. Batzer, and H.H. Kazazian, Jr., *Mobile elements and mammalian genome evolution.* Curr Opin Genet Dev, 2003. **13**(6): p. 651-8.
35. Delebecque, F., R. Suspene, S. Calattini, N. Casartelli, A. Saib, A. Froment, S. Wain-Hobson, A. Gessain, J.P. Vartanian, and O. Schwartz, *Restriction of foamy viruses by APOBEC cytidine deaminases.* J Virol, 2006. **80**(2): p. 605-14.
36. Di Noia, J. and M.S. Neuberger, *Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase.* Nature, 2002. **419**(6902): p. 43-8.
37. Di Noia, J.M. and M.S. Neuberger, *Molecular mechanisms of antibody somatic hypermutation.* Annu Rev Biochem, 2007. **76**: p. 1-22.
38. Doehle, B.P., A. Schafer, and B.R. Cullen, *Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif.* Virology, 2005. **339**(2): p. 281-8.
39. Doehle, B.P., A. Schafer, H.L. Wiegand, H.P. Bogerd, and B.R. Cullen, *Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion.* J Virol, 2005. **79**(13): p. 8201-7.
40. Dutko, J.A., A. Schafer, A.E. Kenny, B.R. Cullen, and M.J. Curcio, *Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases.* Curr Biol, 2005. **15**(7): p. 661-6.
41. Esnault, C., O. Heidmann, F. Delebecque, M. Dewannieux, D. Ribet, A.J. Hance, T. Heidmann, and O. Schwartz, *APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses.* Nature, 2005. **433**(7024): p. 430-3.

42. Esnault, C., S. Priet, D. Ribet, O. Heidmann, and T. Heidmann, *Restriction by APOBEC3 proteins of endogenous retroviruses with an extracellular life cycle: ex vivo effects and in vivo "traces" on the murine IAPE and human HERV-K elements*. Retrovirology, 2008. **5**: p. 75.
43. Evans, C.a.A., R, *DNase II: genes, enzymes and function.* Gene, 2003. **322**: p. 1-15.
44. Ewing, B., L. Hillier, M.C. Wendl, and P. Green, *Base-calling of automated sequencer traces using phred. I. Accuracy assessment*. Genome Res, 1998. **8**(3): p. 175-85.
45. Fisher, A.G., B. Ensoli, L. Ivanoff, M. Chamberlain, S. Petteway, L. Ratner, R.C. Gallo, and F. Wong-Staal, *The sor gene of HIV-1 is required for efficient virus transmission in vitro*. Science, 1987. **237**(4817): p. 888-93.
46. Fitzgibbon, J.E., S. Mazar, and D.T. Dubin, *A new type of G-->A hypermutation affecting human immunodeficiency virus.* AIDS Res Hum Retroviruses, 1993. **9**(9): p. 833-8.
47. Fornerod, M., M. Ohno, M. Yoshida, and I.W. Mattaj, *CRM1 is an export receptor for leucine-rich nuclear export signals*. Cell, 1997. **90**(6): p. 1051-60.
48. Fried, H. and U. Kutay, *Nucleocytoplasmic transport: taking an inventory.* Cell Mol Life Sci, 2003. **60**(8): p. 1659-88.
49. Fujino, T., N. Navaratnam, and J. Scott, *Human apolipoprotein B RNA editing deaminase gene (APOBEC1)*. Genomics, 1998. **47**(2): p. 266-75.
50. Gabuzda, D.H., K. Lawrence, E. Langhoff, E. Terwilliger, T. Dorfman, W.A. Haseltine, and J. Sodroski, *Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes.* J Virol, 1992. **66**(11): p. 6489-95.
51. Gallois-Montbrun, S., R.K. Holmes, C.M. Swanson, M. Fernandez-Ocana, H.L. Byers, M.A. Ward, and M.H. Malim, *Comparison of cellular ribonucleoprotein complexes associated with the APOBEC3F and APOBEC3G antiviral proteins.* J Virol, 2008. **82**(11): p. 5636-42.
52. Gallois-Montbrun, S., B. Kramer, C.M. Swanson, H. Byers, S. Lynham, M. Ward, and M.H. Malim, *Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules.* J Virol, 2007. **81**(5): p. 2165-78.
53. Gao, F., E. Bailes, D.L. Robertson, Y. Chen, C.M. Rodenburg, S.F. Michael, L.B. Cummins, L.O. Arthur, M. Peeters, G.M. Shaw, P.M. Sharp, and B.H. Hahn, *Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes.* Nature, 1999. **397**(6718): p. 436-41.
54. Goff, S.P., *Genetic control of retrovirus susceptibility in mammalian cells.* Annu Rev Genet, 2004. **38**: p. 61-85.
55. Goila-Gaur, R., M.A. Khan, E. Miyagi, S. Kao, and K. Strebel, *Targeting APOBEC3A to the viral nucleoprotein complex confers antiviral activity.* Retrovirology, 2007. **4**: p. 61.
56. Hache, G., M.T. Liddament, and R.S. Harris, *The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain*. J Biol Chem, 2005. **280**(12): p. 10920-4.
57. Harjes, E., P.J. Gross, K.M. Chen, Y. Lu, K. Shindo, R. Nowarski, J.D. Gross, M. Kotler, R.S. Harris, and H. Matsuo, *An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model.* J Mol Biol, 2009. **389**(5): p. 819-32.
58. Harris, R.S., K.N. Bishop, A.M. Sheehy, H.M. Craig, S.K. Petersen-Mahrt, I.N. Watt, M.S. Neuberger, and M.H. Malim, *DNA deamination mediates innate immunity to retroviral infection.* Cell, 2003. **113**(6): p. 803-9.
59. Harris, R.S. and M.T. Liddament, *Retroviral restriction by APOBEC proteins*. Nat Rev Immunol, 2004. **4**(11): p. 868-77.
60. Harris, R.S., S.K. Petersen-Mahrt, and M.S. Neuberger, *RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators*. Mol Cell, 2002. **10**(5): p. 1247-53.
61. Harris, R.S., J.E. Sale, S.K. Petersen-Mahrt, and M.S. Neuberger, *AID is essential for immunoglobulin V gene conversion in a cultured B cell line.* Curr Biol, 2002. **12**(5): p. 435-8.
62. Hirano, K., S.G. Young, R.V. Farese, Jr., J. Ng, E. Sande, C. Warburton, L.M. Powell-Braxton, and N.O. Davidson, *Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48*. J Biol Chem, 1996. **271**(17): p. 9887-90.
63. Holmes, R.K., M.H. Malim, and K.N. Bishop, *APOBEC-mediated viral restriction: not simply editing?* Trends Biochem Sci, 2007. **32**(3): p. 118-28.
64. Ishii, K.J. and S. Akira, *Innate immune recognition of, and regulation by, DNA*. Trends Immunol, 2006. **27**(11): p. 525-32.

65. Ishii, K.J., C. Coban, H. Kato, K. Takahashi, Y. Torii, F. Takeshita, H. Ludwig, G. Sutter, K. Suzuki, H. Hemmi, S. Sato, M. Yamamoto, S. Uematsu, T. Kawai, O. Takeuchi, and S. Akira, *A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA.* Nat Immunol, 2006. **7**(1): p. 40-8.
66. Ito, S., H. Nagaoka, R. Shinkura, N. Begum, M. Muramatsu, M. Nakata, and T. Honjo, *Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 1975-80.
67. Ivics, Z., P.B. Hackett, R.H. Plasterk, and Z. Izsvak, *Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.* Cell, 1997. **91**(4): p. 501-10.
68. Iwatani, Y., H. Takeuchi, K. Strebel, and J.G. Levin, *Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect.* J Virol, 2006. **80**(12): p. 5992-6002.
69. Janeway, C.A., Jr. and R. Medzhitov, *Innate immune recognition.* Annu Rev Immunol, 2002. **20**: p. 197-216.
70. Jarmuz, A., A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott, and N. Navaratnam, *An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.* Genomics, 2002. **79**(3): p. 285-96.
71. Jern, P., J.P. Stoye, and J.M. Coffin, *Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses.* PLoS Genet, 2007. **3**(10): p. 2014-22.
72. Jonsson, S.R., R.S. LaRue, M.D. Stenglein, S.C. Fahrenkrug, V. Andresdottir, and R.S. Harris, *The restriction of zoonotic PERV transmission by human APOBEC3G.* PLoS ONE, 2007. **2**(9): p. e893.
73. Kalderon, D., W.D. Richardson, A.F. Markham, and A.E. Smith, *Sequence requirements for nuclear location of simian virus 40 large-T antigen.* Nature, 1984. **311**(5981): p. 33-8.
74. Kavli, B., M. Otterlei, G. Slupphaug, and H.E. Krokan, *Uracil in DNA--general mutagen, but normal intermediate in acquired immunity.* DNA Repair (Amst), 2007. **6**(4): p. 505-16.
75. Kerr, I.M., R.E. Brown, and A.G. Hovanessian, *Nature of inhibitor of cell-free protein synthesis formed in response to interferon and double-stranded RNA.* Nature, 1977. **268**(5620): p. 540-2.
76. Kidd, J.M., T.L. Newman, E. Tuzun, R. Kaul, and E.E. Eichler, *Population Stratification of a Common APOBEC Gene Deletion Polymorphism.* PLoS Genet, 2007. **3**(4): p. e63.
77. Kieffer, T.L., P. Kwon, R.E. Nettles, Y. Han, S.C. Ray, and R.F. Siliciano, *G-->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo.* J Virol, 2005. **79**(3): p. 1975-80.
78. Kinomoto, M., T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima, T. Kurata, T. Sata, and K. Tokunaga, *All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.* Nucleic Acids Res, 2007. **35**(9): p. 2955-64.
79. Kostic, C. and P.H. Shaw, *Isolation and characterization of sixteen novel p53 response genes.* Oncogene, 2000. **19**(35): p. 3978-87.
80. Kozak, S.L., M. Marin, K.M. Rose, C. Bystrom, and D. Kabat, *The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules.* J Biol Chem, 2006. **281**(39): p. 29105-19.
81. Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E.J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson,

G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, M.V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G.A. Evans, M. Athanasiou, R. Schultz, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M. Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.P. Yang, R.F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, A. Patrinos, M.J. Morgan, P. de Jong, J.J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y.J. Chen, *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.

82. LaRue, R.S., V. Andresdottir, Y. Blanchard, S.G. Conticello, D. Derse, M. Emerman, W.C. Greene, S.R. Jonsson, N.R. Landau, M. Lochelt, H.S. Malik, M.H. Malim, C. Munk, S.J. O'Brien, V.K. Pathak, K. Strebel, S. Wain-Hobson, X.F. Yu, N. Yuhki, and R.S. Harris, *Guidelines for naming nonprimate APOBEC3 genes and proteins.* J Virol, 2009. **83**(2): p. 494-7.
83. LaRue, R.S., S.R. Jonsson, K.A. Silverstein, M. Lajoie, D. Bertrand, N. El-Mabrouk, I. Hotzel, V. Andresdottir, T.P. Smith, and R.S. Harris, *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, 2008. **9**: p. 104.
84. Lecossier, D., F. Bouchonnet, F. Clavel, and A.J. Hance, *Hypermutation of HIV-1 DNA in the absence of the Vif protein.* Science, 2003. **300**(5622): p. 1112.
85. Lellek, H., R. Kirsten, I. Diehl, F. Apostel, F. Buck, and J. Greeve, *Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex.* J Biol Chem, 2000. **275**(26): p. 19848-56.
86. Liao, W., S.H. Hong, B.H. Chan, F.B. Rudolph, S.C. Clark, and L. Chan, *APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family.* Biochem Biophys Res Commun, 1999. **260**(2): p. 398-404.
87. Liddament, M.T., W.L. Brown, A.J. Schumacher, and R.S. Harris, *APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo.* Curr Biol, 2004. **14**(15): p. 1385-91.
88. Liu, B., X. Yu, K. Luo, Y. Yu, and X.F. Yu, *Influence of primate lentiviral Vif and proteasome inhibitors on human immunodeficiency virus type 1 virion packaging of APOBEC3G.* J Virol, 2004. **78**(4): p. 2072-81.
89. Lochelt, M., F. Romen, P. Bastone, H. Muckenfuss, N. Kirchner, Y.B. Kim, U. Truyen, U. Rosler, M. Battenberg, A. Saib, E. Flory, K. Cichutek, and C. Munk, *The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein.* Proc Natl Acad Sci U S A, 2005. **102**(22): p. 7982-7.
90. Longerich, S., U. Basu, F. Alt, and U. Storb, *AID in somatic hypermutation and class switch recombination.* Curr Opin Immunol, 2006. **18**(2): p. 164-74.
91. Low, A., C.M. Okeoma, N. Lovsin, M. de las Heras, T.H. Taylor, B.M. Peterlin, S.R. Ross, and H. Fan, *Enhanced replication and pathogenesis of Moloney murine leukemia virus in mice defective in the murine APOBEC3 gene.* Virology, 2009. **385**(2): p. 455-63.
92. MacDuff, D.A., Z.L. Demorest, and R.S. Harris, *AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity*. Nucleic Acids Res, 2009. **37**(6): p. 1854-67.
93. MacDuff, D.A. and R.S. Harris, *Directed DNA deamination by AID/APOBEC3 in immunity*. Curr Biol, 2006. **16**(6): p. R186-9.

94. Madani, N. and D. Kabat, *An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein.* J Virol, 1998. **72**(12): p. 10251-5.
95. Maizels, N., *Immunoglobulin gene diversification.* Annu Rev Genet, 2005. **39**: p. 23-46.
96. Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono, *Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.* Nature, 2003. **424**(6944): p. 99-103.
97. Marin, M., S. Golem, K.M. Rose, S.L. Kozak, and D. Kabat, *Human immunodeficiency virus type 1 Vif functionally interacts with diverse APOBEC3 cytidine deaminases and moves with them between cytoplasmic sites of mRNA metabolism.* J Virol, 2008. **82**(2): p. 987-98.
98. Marin, M., K.M. Rose, S.L. Kozak, and D. Kabat, *HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation.* Nat Med, 2003. **9**(11): p. 1398-403.
99. McBride, K.M., V. Barreto, A.R. Ramiro, P. Stavropoulos, and M.C. Nussenzweig, *Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase.* J Exp Med, 2004. **199**(9): p. 1235-44.
100. Medzhitov, R., *Recognition of microorganisms and activation of the immune response.* Nature, 2007. **449**(7164): p. 819-26.
101. Mehle, A., B. Strack, P. Ancuta, C. Zhang, M. McPike, and D. Gabuzda, *Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway.* J Biol Chem, 2004. **279**(9): p. 7792-8.
102. Mehta, A., S. Banerjee, and D.M. Driscoll, *Apobec-1 interacts with a 65-kDa complementing protein to edit apolipoprotein-B mRNA in vitro.* J Biol Chem, 1996. **271**(45): p. 28294-9.
103. Mehta, A., M.T. Kinter, N.E. Sherman, and D.M. Driscoll, *Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA.* Mol Cell Biol, 2000. **20**(5): p. 1846-54.
104. Meurs, E., K. Chong, J. Galabru, N.S. Thomas, I.M. Kerr, B.R. Williams, and A.G. Hovanessian, *Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.* Cell, 1990. **62**(2): p. 379-90.
105. Mikl, M.C., I.N. Watt, M. Lu, W. Reik, S.L. Davies, M.S. Neuberger, and C. Rada, *Mice deficient in APOBEC2 and APOBEC3.* Mol Cell Biol, 2005. **25**(16): p. 7270-7.
106. Miyagi, E., S. Opi, H. Takeuchi, M. Khan, R. Goila-Gaur, S. Kao, and K. Strebel, *Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1.* J Virol, 2007. **81**(24): p. 13346-53.
107. Moran, J.V., S.E. Holmes, T.P. Naas, R.J. DeBerardinis, J.D. Boeke, and H.H. Kazazian, Jr., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917-27.
108. Muckenfuss, H., M. Hamdorf, U. Held, M. Perkovic, J. Lower, K. Cichutek, E. Flory, G.G. Schumann, and C. Munk, *APOBEC3 proteins inhibit human LINE-1 retrotransposition.* J Biol Chem, 2006. **281**(31): p. 22161-72.
109. Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo, *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.* Cell, 2000. **102**(5): p. 553-63.
110. Muramatsu, M., V.S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N.O. Davidson, and T. Honjo, *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, 1999. **274**(26): p. 18470-6.
111. Nakamuta, M., B.H. Chang, E. Zsigmond, K. Kobayashi, H. Lei, B.Y. Ishida, K. Oka, E. Li, and L. Chan, *Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1.* J Biol Chem, 1996. **271**(42): p. 25981-8.
112. Nara, P.L. and P.J. Fischinger, *Quantitative infectivity assay for HIV-1 and-2.* Nature, 1988. **332**(6163): p. 469-70.
113. Narvaiza, I., D.C. Linfesty, B.N. Greener, Y. Hakata, D.J. Pintel, E. Logue, N.R. Landau, and M.D. Weitzman, *Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase.* PLoS Pathog, 2009. **5**(5): p. e1000439.

114. Navaratnam, N., S. Bhattacharya, T. Fujino, D. Patel, A.L. Jarmuz, and J. Scott, *Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site.* Cell, 1995. **81**(2): p. 187-95.
115. Navaratnam, N., R. Shah, D. Patel, V. Fay, and J. Scott, *Apolipoprotein B mRNA editing is associated with UV crosslinking of proteins to the editing site.* Proc Natl Acad Sci U S A, 1993. **90**(1): p. 222-6.
116. Navarro, F., B. Bollman, H. Chen, R. Konig, Q. Yu, K. Chiles, and N.R. Landau, *Complementary function of the two catalytic domains of APOBEC3G.* Virology, 2005. **333**(2): p. 374-86.
117. Neil, S.J., T. Zang, and P.D. Bieniasz, *Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu.* Nature, 2008. **451**(7177): p. 425-30.
118. Neville, M., F. Stutz, L. Lee, L.I. Davis, and M. Rosbash, *The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export.* Curr Biol, 1997. **7**(10): p. 767-75.
119. Newman, E.N., R.K. Holmes, H.M. Craig, K.C. Klein, J.R. Lingappa, M.H. Malim, and A.M. Sheehy, *Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity.* Curr Biol, 2005. **15**(2): p. 166-70.
120. Niewiadomska, A.M., C. Tian, L. Tan, T. Wang, P.T. Sarkis, and X.F. Yu, *Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association.* J Virol, 2007. **81**(17): p. 9577-83.
121. Okabe, Y., K. Kawane, S. Akira, T. Taniguchi, and S. Nagata, *Toll-like receptor-independent gene induction program activated by mammalian DNA escaped from apoptotic DNA degradation.* J Exp Med, 2005. **202**(10): p. 1333-9.
122. Okeoma, C.M., N. Lovsin, B.M. Peterlin, and S.R. Ross, *APOBEC3 inhibits mouse mammary tumour virus replication in vivo.* Nature, 2007. **445**(7130): p. 927-30.
123. Ostertag, E.M., R.J. DeBerardinis, J.L. Goodier, Y. Zhang, N. Yang, G.L. Gerton, and H.H. Kazazian, Jr., *A mouse model of human L1 retrotransposition.* Nat Genet, 2002. **32**(4): p. 655-60.
124. Ostertag, E.M. and H.H. Kazazian, Jr., *Biology of mammalian L1 retrotransposons.* Annu Rev Genet, 2001. **35**: p. 501-38.
125. Ostertag, E.M., E.T. Prak, R.J. DeBerardinis, J.V. Moran, and H.H. Kazazian, Jr., *Determination of L1 retrotransposition kinetics in cultured cells.* Nucleic Acids Res, 2000. **28**(6): p. 1418-23.
126. Peng, G., T. Greenwell-Wild, S. Nares, W. Jin, K.J. Lei, Z.G. Rangel, P.J. Munson, and S.M. Wahl, *Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression.* Blood, 2007.
127. Peng, G., K.J. Lei, W. Jin, T. Greenwell-Wild, and S.M. Wahl, *Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity.* J Exp Med, 2006. **203**(1): p. 41-6.
128. Perez, O. and T.J. Hope, *Cellular restriction factors affecting the early stages of HIV replication.* Curr HIV/AIDS Rep, 2006. **3**(1): p. 20-5.
129. Perez-Caballero, D., S.J. Soll, and P.D. Bieniasz, *Evidence for restriction of ancient primate gammaretroviruses by APOBEC3 but not TRIM5alpha proteins.* PLoS Pathog, 2008. **4**(10): p. e1000181.
130. Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger, *AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification.* Nature, 2002. **418**(6893): p. 99-103.
131. Powell, L.M., S.C. Wallis, R.J. Pease, Y.H. Edwards, T.J. Knott, and J. Scott, *A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine.* Cell, 1987. **50**(6): p. 831-40.
132. Rada, C., J.M. Jarvis, and C. Milstein, *AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 7003-8.
133. Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A.G. Ugazio, N. Brousse, M. Muramatsu, L.D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, and A. Durandy, *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2).* Cell, 2000. **102**(5): p. 565-75.

134. Roberts, W.K., A. Hovanessian, R.E. Brown, M.J. Clemens, and I.M. Kerr, *Interferon-mediated protein kinase and low-molecular-weight inhibitor of protein synthesis.* Nature, 1976. **264**(5585): p. 477-80.
135. Rogozin, I.B., M.K. Basu, I.K. Jordan, Y.I. Pavlov, and E.V. Koonin, *APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis.* Cell Cycle, 2005. **4**(9): p. 1281-5.
136. Rosler, C., J. Kock, M. Kann, M.H. Malim, H.E. Blum, T.F. Baumert, and F. von Weizsacker, *APOBEC-mediated interference with hepadnavirus production*. Hepatology, 2005. **42**(2): p. 301-9.
137. Russell, R.A., H.L. Wiegand, M.D. Moore, A. Schafer, M.O. McClure, and B.R. Cullen, *Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors.* J Virol, 2005. **79**(14): p. 8724-31.
138. Sadler, A.J. and B.R. Williams, *Interferon-inducible antiviral effectors*. Nat Rev Immunol, 2008. **8**(7): p. 559-68.
139. Sancar, A., L.A. Lindsey-Boltz, K. Unsal-Kacmaz, and S. Linn, *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.* Annu Rev Biochem, 2004. **73**: p. 39-85.
140. Sarkis, P.T., S. Ying, R. Xu, and X.F. Yu, *STAT1-independent cell type-specific regulation of antiviral APOBEC3G by IFN-alpha.* J Immunol, 2006. **177**(7): p. 4530-40.
141. Sasada, A., A. Takaori-Kondo, K. Shirakawa, M. Kobayashi, A. Abudu, M. Hishizawa, K. Imada, Y. Tanaka, and T. Uchiyama, *APOBEC3G targets human T-cell leukemia virus type 1*. Retrovirology, 2005. **2**(1): p. 32.
142. Sawyer, S.L., M. Emerman, and H.S. Malik, *Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G.* PLoS Biol, 2004. **2**(9): p. E275.
143. Schulz, W.A., C. Steinhoff, and A.R. Florl, *Methylation of endogenous human retroelements in health and disease.* Curr Top Microbiol Immunol, 2006. **310**: p. 211-50.
144. Schumacher, A.J., G. Hache, D.A. Macduff, W.L. Brown, and R.S. Harris, *The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction.* J Virol, 2008. **82**(6): p. 2652-60.
145. Schumacher, A.J., D.V. Nissley, and R.S. Harris, *APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast.* Proc Natl Acad Sci U S A, 2005. **102**(28): p. 9854-9.
146. Seppen, J., *Unedited inhibition of HBV replication by APOBEC3G*. J Hepatol, 2004. **41**(6): p. 1068-9.
147. Sheehy, A.M., N.C. Gaddis, J.D. Choi, and M.H. Malim, *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.* Nature, 2002. **418**(6898): p. 646-50.
148. Sheehy, A.M., N.C. Gaddis, and M.H. Malim, *The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif.* Nat Med, 2003. **9**(11): p. 1404-7.
149. Shindo, K., A. Takaori-Kondo, M. Kobayashi, A. Abudu, K. Fukunaga, and T. Uchiyama, *The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity.* J Biol Chem, 2003. **278**(45): p. 44412-6.
150. Simon, J.H., N.C. Gaddis, R.A. Fouchier, and M.H. Malim, *Evidence for a newly discovered cellular anti-HIV-1 phenotype.* Nat Med, 1998. **4**(12): p. 1397-400.
151. Simon, J.H. and M.H. Malim, *The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes.* J Virol, 1996. **70**(8): p. 5297-305.
152. Stade, K., C.S. Ford, C. Guthrie, and K. Weis, *Exportin 1 (Crm1p) is an essential nuclear export factor.* Cell, 1997. **90**(6): p. 1041-50.
153. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism*. J Biol Chem, 2006. **281**(25): p. 16837-41.
154. Stenglein, M.D., H. Matsuo, and R.S. Harris, *Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization.* J Virol, 2008. **82**(19): p. 9591-9.
155. Stetson, D.B., J.S. Ko, T. Heidmann, and R. Medzhitov, *Trex1 prevents cell-intrinsic initiation of autoimmunity*, in *Cell*. 2008. p. 587-98.
156. Stetson, D.B. and R. Medzhitov, *Recognition of cytosolic DNA activates an IRF3-dependent innate immune response.* Immunity, 2006. **24**(1): p. 93-103.

157. Stopak, K.S., Y.L. Chiu, J. Kropp, R.M. Grant, and W.C. Greene, *Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells*. J Biol Chem, 2007. **282**(6): p. 3539-46.
158. Strebel, K., D. Daugherty, K. Clouse, D. Cohen, T. Folks, and M.A. Martin, *The HIV 'A' (sor) gene product is essential for virus infectivity.* Nature, 1987. **328**(6132): p. 728-30.
159. Suspene, R., D. Guetard, M. Henry, P. Sommer, S. Wain-Hobson, and J.P. Vartanian, *Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo.* Proc Natl Acad Sci U S A, 2005. **102**(23): p. 8321-6.
160. Suspene, R., M. Henry, S. Guillot, S. Wain-Hobson, and J.P. Vartanian, *Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR.* J Gen Virol, 2005. **86**(Pt 1): p. 125-9.
161. Suspene, R., P. Sommer, M. Henry, S. Ferris, D. Guetard, S. Pochet, A. Chester, N. Navaratnam, S. Wain-Hobson, and J.P. Vartanian, *APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase.* Nucleic Acids Res, 2004. **32**(8): p. 2421-9.
162. Takeda, E., S. Tsuji-Kawahara, M. Sakamoto, M.A. Langlois, M.S. Neuberger, C. Rada, and M. Miyazawa, *Mouse APOBEC3 restricts friend leukemia virus infection and pathogenesis in vivo*. J Virol, 2008. **82**(22): p. 10998-1008.
163. Tanaka, Y., H. Marusawa, H. Seno, Y. Matsumoto, Y. Ueda, Y. Kodama, Y. Endo, J. Yamauchi, T. Matsumoto, A. Takaori-Kondo, I. Ikai, and T. Chiba, *Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes.* Biochem Biophys Res Commun, 2006.
164. Thielen, B.K., K.C. Klein, L.W. Walker, M. Rieck, J.H. Buckner, G.W. Tomblingson, and J.R. Lingappa, *T Cells Contain an RNase-Insensitive Inhibitor of APOBEC3G Deaminase Activity.* PLoS Pathog, 2007. **3**(9): p. e135.
165. Turelli, P., B. Mangeat, S. Jost, S. Vianin, and D. Trono, *Inhibition of hepatitis B virus replication by APOBEC3G.* Science, 2004. **303**(5665): p. 1829.
166. Turelli, P., S. Vianin, and D. Trono, *The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay.* J Biol Chem, 2004. **279**(42): p. 43371-3.
167. Uematsu, S. and S. Akira, *Toll-like receptors and Type I interferons*. J Biol Chem, 2007. **282**(21): p. 15319-23.
168. Van Damme, N., D. Goff, C. Katsura, R.L. Jorgenson, R. Mitchell, M.C. Johnson, E.B. Stephens, and J. Guatelli, *The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein.* Cell Host Microbe, 2008. **3**(4): p. 245-52.
169. Vartanian, J.P., D. Guetard, M. Henry, and S. Wain-Hobson, *Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions.* Science, 2008. **320**(5873): p. 230-3.
170. von Schwedler, U., J. Song, C. Aiken, and D. Trono, *Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells.* J Virol, 1993. **67**(8): p. 4945-55.
171. Wang, F.X., J. Huang, H. Zhang, X. Ma, and H. Zhang, *APOBEC3G upregulation by alpha interferon restricts human immunodeficiency virus type 1 infection in human peripheral plasmacytoid dendritic cells.* J Gen Virol, 2008. **89**(Pt 3): p. 722-30.
172. Waterston, R.H., K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S.E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M.R. Brent, D.G. Brown, S.D. Brown, C. Bult, J. Burton, J. Butler, R.D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A.T. Chinwalla, D.M. Church, M. Clamp, C. Clee, F.S. Collins, L.L. Cook, R.R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K.D. Delehaunty, J. Deri, E.T. Dermitzakis, C. Dewey, N.J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D.M. Dunn, S.R. Eddy, L. Elnitski, R.D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G.A. Fewell, P. Flicek, K. Foley, W.N. Frankel, L.A. Fulton, R.S. Fulton, T.S. Furey, D. Gage, R.A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T.A. Graves, E.D. Green, S. Gregory, R. Guigo, M. Guyer, R.C. Hardison, D. Haussler, Y. Hayashizaki, L.W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D.B. Jaffe, L.S. Johnson, M. Jones, T.A. Jones, A. Joy, M. Kamal, E.K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W.J. Kent, A. Kirby, D.L. Kolbe, I. Korf, R.S. Kucherlapati, E.J.

Kulbokas, D. Kulp, T. Landers, J.P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D.R. Maglott, E.R. Mardis, L. Matthews, E. Mauceli, J.H. Mayer, M. McCarthy, W.R. McCombie, S. McLaren, K. McLay, J.D. McPherson, J. Meldrim, B. Meredith, J.P. Mesirov, W. Miller, T.L. Miner, E. Mongin, K.T. Montgomery, M. Morgan, R. Mott, J.C. Mullikin, D.M. Muzny, W.E. Nash, J.O. Nelson, M.N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M.J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K.H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C.S. Pohl, A. Poliakov, T.C. Ponce, C.P. Ponting, S. Potter, M. Quail, A. Reymond, B.A. Roe, K.M. Roskin, E.M. Rubin, A.G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M.S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J.B. Singer, G. Slater, A. Smit, D.R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J.P. Vinson, A.C. Von Niederhausern, C.M. Wade, M. Wall, R.J. Weber, R.B. Weiss, M.C. Wendl, A.P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R.K. Wilson, E. Winter, K.C. Worley, D. Wyman, S. Yang, S.P. Yang, E.M. Zdobnov, M.C. Zody and E.S. Lander, *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

173. Wedekind, J.E., G.S. Dance, M.P. Sowden, and H.C. Smith, *Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business.* Trends Genet, 2003. **19**(4): p. 207-16.

174. Wichroski, M.J., K. Ichiyama, and T.M. Rana, *Analysis of HIV-1 viral infectivity factor-mediated proteasome-dependent depletion of APOBEC3G: correlating function and subcellular localization.* J Biol Chem, 2005. **280**(9): p. 8387-96.

175. Wichroski, M.J., G.B. Robb, and T.M. Rana, *Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies.* PLoS Pathog, 2006. **2**(5): p. e41.

176. Wiegand, H.L., B.P. Doehle, H.P. Bogerd, and B.R. Cullen, *A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins.* Embo J, 2004. **23**(12): p. 2451-8.

177. Wilson, G.G. and N.E. Murray, *Restriction and modification systems*, in *Annu. Rev. Genet.* 1991. p. 585-627.

178. Wolff, B., J.J. Sanglier, and Y. Wang, *Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA.* Chem Biol, 1997. **4**(2): p. 139-47.

179. Yamanaka, S., K.S. Poksay, M.E. Balestra, G.Q. Zeng, and T.L. Innerarity, *Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc motif is essential for catalytic activity, and noncatalytic auxiliary factor(s) of the editing complex are widely distributed.* J Biol Chem, 1994. **269**(34): p. 21725-34.

180. Yang, Y., M.P. Sowden, Y. Yang, and H.C. Smith, *Intracellular trafficking determinants in APOBEC-1, the catalytic subunit for cytidine to uridine editing of apolipoprotein B mRNA.* Exp Cell Res, 2001. **267**(2): p. 153-64.

181. Yang, Y., Y. Yang, and H.C. Smith, *Multiple protein domains determine the cell type-specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13075-80.

182. Yu, Q., D. Chen, R. Konig, R. Mariani, D. Unutmaz, and N.R. Landau, *APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication.* J Biol Chem, 2004. **279**(51): p. 53379-86.

183. Yu, Q., R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J.M. Coffin, and N.R. Landau, *Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome.* Nat Struct Mol Biol, 2004. **11**(5): p. 435-42.

184. Yu, X., Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, and X.F. Yu, *Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.* Science, 2003. **302**(5647): p. 1056-60.

185. Zhang, H., B. Yang, R.J. Pomerantz, C. Zhang, S.C. Arunachalam, and L. Gao, *The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.* Nature, 2003. **424**(6944): p. 94-8.

186. Zheng, Y.H., D. Irwin, T. Kurosu, K. Tokunaga, T. Sata, and B.M. Peterlin, *Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication*. J Virol, 2004. **78**(11): p. 6073-6.