Role of the single deaminase domain APOBEC3A in virus restriction, retrotransposition, DNA damage and cancer

Yaqiong Wang,1 Kimberly Schmitt,1† Kejun Guo,2 Mario L. Santiago2 and Edward B. Stephens1

Correspondence
Edward B. Stephens
estephen@kumc.edu

1Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, 3901 Rainbow Blvd, Kansas City, KS 66160, USA
2Departments of Medicine, Microbiology and Immunology, University of Colorado Denver Medical School, Aurora, CO 80045, USA

The apolipoprotein mRNA editing enzyme catalytic polypeptide-like 3 (APOBEC3; A3) proteins are a family of seven cytidine deaminases (A3A, A3B, A3C, A3D, A3F, A3G and A3H) that restrict certain viral infections. These innate defence factors are best known for their ability to restrict the replication of human immunodeficiency virus type 1 (HIV-1) lacking a functional Vif protein (HIV-1Δvif) through the deamination of cytidine residues to uridines during reverse transcription, ultimately leading to lethal G→A changes in the viral genome. The best studied of the A3 proteins has been APOBEC3G because of its potent activity against HIV-1Δvif. However, one member of this family, A3A, has biological properties that make it unique among the A3 proteins. In this review, we will focus on the structural and phylogenetic features of the human and non-human primate A3A proteins, their role in the restriction of retroviruses and other viruses, and current findings on other biological properties affected by this protein.

Introduction

Several innate host restriction factors have been described that inhibit retroviruses such as human immunodeficiency virus type 1 (HIV-1). These innate restriction factors include the apolipoprotein B mRNA editing enzyme catalytic-polypeptide-like proteins 3 (APOBEC3; A3), bone marrow stromal antigen 2 (BST-2, also known as CD317, HM1.24 or tetherin) and tripartite motif (TRIM) proteins (Sheehy et al., 2002; Stremlau et al., 2004; Neil et al., 2008; Van Damme et al., 2008). The discovery of human APOBEC3G (hA3G) occurred during the search for a protein that restricted the replication of HIV lacking a functional Vif protein (HIV-1Δvif) (Sheehy et al., 2002) and has been the most extensively studied of the three restriction factors mentioned above. It is now known that A3G is one of seven members of the A3 family in the human genome (hA3A, hA3B, hA3C, hA3D, hA3F, hA3G and hA3H) (Jarmuz et al., 2002; OhAinle et al., 2006). The A3 proteins, along with activation-induced deaminase (AID), APOBEC1 (A1) and APOBEC2 (A2) are cytidine deaminases. An 11th member of this family has been identified in humans, APOBEC4 (A4), but has no reported deaminase activity (Rogozin et al., 2005). This review aims to provide a summary of the studies on A3A including the restriction of lentiviruses from non-human primates and other retroviruses, restriction of select DNA viruses, inhibition of retrotransposition, editing of mRNA transcripts, its role in degradation of foreign DNA and its potential role in the genesis of cancer.

A3 proteins and the structure of hA3A

The seven genes for the A3 proteins are arranged tandemly on chromosomes 22 and 10 of the human and rhesus macaque genomes, respectively (Jarmuz et al., 2002; Schmitt et al., 2011). The A3 proteins are characterized by one or two zinc-coordinating domains (Z domains) with the sequence H-x-E-x23–28-PC-x2–4-C. hA3A, hA3C and hA3H have one Z domain, while A3B, A3D, A3F and A3G have two Z domains (Fig. 1). Phylogenetic analysis revealed a clustering among A3 zinc-finger motifs, further classified as A3Z1, A3Z2 or A3Z3 (Fig. 1) (Conticello et al., 2003; LaRue et al., 2009). This diversification probably preceded the split between placental mammals and marsupials (125–150 million years ago) (Bininda-Emonds et al., 2007; Münk et al., 2012). The structure of hA3A (aa 10–194 of 199 aa) was solved by nuclear magnetic resonance (NMR) spectroscopy (Byeon et al., 2013). Similar to other APOBEC (A2 and some A3 members) proteins for which a structure has been solved, the hA3A structure consists of six helices surrounding a central β-sheet of five strands. In another study, A3G was found to form high-order multimers as a function of protein
structure of the C-terminal domain of hA3G (Holden et al., 2008; Kitamura et al., 2015). A salient finding of this study was that A3A monomers would not act on the target cytidine but rather were involved in substrate binding. They further stated that the evolution of two domain enzymes may have allowed the separation of binding and enzymatic activities, which resulted in less active proteins that were more specific to their targets. Comparison of structures of hA3A with other available APOBEC structures revealed that the A3A NMR structure is most similar to the X-ray crystallographic structure of the C-terminal domain of hA3G (Holden et al., 2008; Kitamura et al., 2012). An interesting feature of hA3A that is not found in the other hA3 proteins is the presence of 4 aa between the two cysteine residues of the canonical deaminase domain. Byeon et al. (2013) found that hA3A has a slightly higher binding affinity for ssRNA than ssDNA and displayed a similar catalytic activity for TTCA and CCAA substrates.

**A3A sequence diversity among primates and tissue expression**

Lentiviruses are endemic in multiple mammalian species, but the closest ancestors of HIV-1 and HIV-2 are the simian immunodeficiency viruses (SIVs) that infect west-central African chimpanzees (Keele et al., 2006) and sooty mangabeys (Santiago et al., 2005), respectively. Over 30 Old World monkey species also harbour endemic SIV infections (Hahn et al., 2000), but these infections are limited to sub-Saharan Africa. For example, Asian rhesus macaques are considered Old World monkeys but do not have endemic SIV infections. New World monkeys have also not been found to harbour endemic lentivirus infections. Notably, infection of rhesus macaques with SIV from sooty mangabeys (SIVsm) resulted in an AIDS phenotype (Letvin et al., 1985). To date, the SIV/rhesus macaque model remains the gold standard for understanding lentivirus pathogenesis in vivo.

Full-length genomes of non-human primates are limited, thereby precluding an in-depth evolutionary analysis of the A3 locus during primate evolution. In fact, the draft genome of the rhesus macaque had significant gaps in the A3 locus, leading to an early hypothesis that rhesus macaques do not encode A3A (Virgen & Hatzioannou, 2007). However, by piecing together fragments from unassembled sequence contigs, our group cloned the rhesus macaque A3A (rhA3A) gene (Schmitt et al., 2011). The sequence of the A3A proteins from 20 primate species from the hominids (eight sequences), Old World monkeys (seven sequences) and New World monkeys (five sequences) are compared in Fig. 2. The primate A3A proteins vary in length from 199 to 206 aa with amino acid identity at 88 positions (~45 %) and conserved amino acids at an additional 23 positions. The most variable region of the primate A3A proteins is the N-terminal region (Fig. 2). Notably, all hominid A3A proteins (with the exception of siamang) have a 3 aa deletion (positions 27–29) that is not observed in A3A proteins from Old or New World monkeys (Fig. 2). This indel will be discussed in greater detail later. Additionally, the New World monkey A3A proteins have a proline insertion at position 45 and a 2 aa deletion at positions 108–109 (Fig. 2).

The interaction between host proteins and pathogens may leave a genetic mark on the host protein known as a positive-selection signature. This manifests as an increase in DNA sequence mutations that change the protein sequence (non-synonymous mutations, Ks) compared with mutations that preserve the protein sequence (synonymous mutations, Ks). A3G, A3C, A3D, A3F and A3H were noted concentration (Li et al., 2014). In contrast, A3A was only found as monomers at all concentrations tested. These investigators found a correlation with those A3 proteins with a propensity to multimerize (A3B, A3D, A3F, A3G and A3H) and restriction activities of various A3 proteins. Recently, however, the crystal structure of A3A E72A–C171A was solved at 2.85 Å resolution, which showed that hA3A exists as a homodimer (Logue et al., 2012). An interesting feature of hA3A proteins is a proline insertion at position 27–29 that is not observed in A3A proteins from Old or New World monkeys (Fig. 2). This indel will be discussed in greater detail later. Additionally, the New World monkey A3A proteins have a proline insertion at position 45 and a 2 aa deletion at positions 108–109 (Fig. 2).

The interaction between host proteins and pathogens may leave a genetic mark on the host protein known as a positive-selection signature. This manifests as an increase in DNA sequence mutations that change the protein sequence (non-synonymous mutations, Ks) compared with mutations that preserve the protein sequence (synonymous mutations, Ks). A3G, A3C, A3D, A3F and A3H were noted in DNA sequence mutations that change the protein sequence (non-synonymous mutations, Ks) compared with mutations that preserve the protein sequence (synonymous mutations, Ks). A3G, A3C, A3D, A3F and A3H were noted
The biology of APOBEC3A
Fig. 2. Alignment of the amino acid sequence of hA3A with other hominid (top group) and Old (middle group) and New World (bottom group) monkey A3A proteins. Residues in red indicate invariant residues found in hominids and Old and New World monkeys with the exception of *Symphalangus syndactylus* (siamang) whose sequence appears to be truncated. A dot represents identity to the hA3A sequence and – represents gaps introduced into the sequence for purposes of alignment. The amino acids in red are identical in all species analysed. The species analysed were: *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Pan paniscus* (bonobo), *Gorilla gorilla* (gorilla), *Pongo pygmaeus* (Bornean orangutan), *Nomascus leucogenys* (northern white-cheeked gibbon), *Hylobates lar* (common gibbon), *Symphalangus syndactylus* (siamang), *Macaca mulatta* (rhesus macaque), *Macaca fascicularis* (cynomolgus monkey), *Papio anubis* (baboon), *Chlorocebus aethiops* (African green monkey), *Colobus guereza* (black-and-white colobus), *Cercopithecus neglectus* (De Brazza’s monkey), *Mandrillus sphinx* (mandrill), *Aotus trivirgatus* (northern owl monkey), *Saimiri boliviensis boliviensis* (black-capped squirrel), *Saimiri boliviensis* (black-capped squirrel), *Saimiri sciureus* (squirrel monkey), and *Callithrix jacchus* (tufted capuchin).
to have undergone positive selection, with $K_a/K_s$ values > 1 (Sawyer et al., 2004; Zhang & Webb, 2004; OhAinle et al., 2006; Ortiz et al., 2009; Cagliani et al., 2011; Duggal et al., 2011). Lentivirus restriction factors such as BST-2, TRIM5α and SAMHD1 also exhibited a positive-selection signature. In contrast, A3A was not under positive selection (Sawyer et al., 2004), consistent with the notion that it may not have played a significant role in retrovirus restriction. We revisited this question using primate A3A sequences from five Old World monkeys and five hominids (Fig. 3a). Indeed, 25 pairwise comparisons of $K_a/K_s$ ratios between hominid versus monkey A3A revealed that A3A was under neutral selection (Fig. 3b). However, when we looked at specific regions of A3A that were close to the predicted polynucleotide-binding groove (Bulliard et al., 2011), particularly the AC-loop1 (minus the indels), AC-loop3, AC-loop5 and $\alpha$-helix4, there was a positive-selection signature. These data revealed that, while A3A does not exhibit signals of positive selection as do the other A3 genes, the regions surrounding the nucleic acid-binding groove are hotspots of genetic conflict. Another report found that A3A exhibited positive-selection signatures only on select branches of the primate A3A phylogenetic tree (Henry et al., 2012). Overall, the data suggest that A3A proteins from non-human primates may have intriguing antiretroviral properties. The antiretroviral properties of rhesus macaque as well as other Old and New World primate A3A proteins will be discussed later in this review.

---

**Fig. 3.** Selection pressure in A3A following the evolutionary split between hominids and Old World monkeys. (a) A3A phylogenetic tree. A neighbour-joining (unrooted) tree for nucleotide sequences of five Old World monkey and five hominid species was reconstructed with gaps excluded and correcting for multiple substitutions using CLUSTAL_X. Support (>80 %) for specific clades is shown following 1000 subreplicates. (b) Identification of a positive-selection ‘hotspot’ in A3A. Pairwise ratios of $K_a/K_s$ substitutions between hominid and Old World monkey A3A sequences were computed using the K-estimator program. Full-length A3A was essentially neutral. However, when sequences encoding only the AC-loop1, AC-loop3, AC-loop5 and $\alpha$-helix4 regions were analysed, $K_a/K_s$ ratios that were significantly greater than 1 were obtained. These regions are referred to as ‘hotspots’ of positive selection in A3A. The 3 aa indel was excluded from this analysis, and thus the level of positive selection may be higher. In contrast, residues outside the ‘hotspots’ had $K_a/K_s$ values indicative of low-level purifying selection. Each dot corresponds to a pairwise comparison. Differences were evaluated using a two-tailed Student's t-test (***P<0.0001).
Expression profile of A3A in tissues and cells

It is generally accepted that hA3G is expressed in the cytoplasm of cells, due to encoded cytoplasmic retention signals (Bennett et al., 2008; Stenglein et al., 2008), while hA3A has been shown to be expressed in both the cytoplasm and nucleus of cells (Muckenfuss et al., 2006). hA3A is known to be expressed in differentiated cell populations including monocyte/macrophages, dendritic cells and keratinocytes (Peng et al., 2007; Vartanian et al., 2008). To probe the localization of endogenous hA3A in differentiated macrophages, a rabbit polyclonal antibody was generated against the peptide CPFQPWWG1EH5QALSG-RLRAILQNQGN corresponding to the C terminus of A3A (Carpenter et al., 2012). Using this antibody, it was found that endogenous hA3A was expressed only in the cytoplasm of cells (Land et al., 2013). However, the C terminus of hA3A shows significant amino acid identity to hA3B and hA3G. The hA3A-derived peptide used to generate an antibody was 100% identical to hA3B and only differed at three amino acid positions from A3G, which has the sequence CPFQPWWG1EH5QALSG-RLRAILQNQGN (differences in bold). Thus, the detecting antibody for endogenous hA3A was probably not specific to hA3A. Convincing proof that hA3A has exclusive cytoplasmic localization may also require the identification of a cytoplasmic retention signal similar to that of hA3G.

A number of studies have comprehensively examined the presence of A3 mRNA species in human cells and tissues. In the first study, investigators found that hA3A mRNA was detected at low levels in T-cells and at higher levels in monocytes (Koning et al., 2009). They showed that IFN-α treatment of isolated CD4+T-cell populations resulted in the induction of A3A mRNA, which was followed by the appearance of protein detectable in Western blots. In a second study, hA3A mRNA was not detected in CD4+T-cells, and the addition of IFN-α did not induce hA3A (Refsland et al., 2010). In a third study, these same investigators found that neither hA3A nor rhA3A expression was increased in stimulated CD4+T-cells from the respective hosts. We found that stimulated rhesus macaque CD4+T-cells isolated from PBMCs expressed rhA3A at robust levels comparable to those of rhA3G (Schmitt et al., 2011). The exact reasons for these discrepancies are unclear. However, it should be noted that in the two studies on rhesus macaque CD4+T-cells, the cells were stimulated with either phytohaemagglutinin (Hultquist et al., 2011) or Staphylococcus enterotoxin B (Schmitt et al., 2011). Thus, the method of stimulation may be critical to the induction of rhA3A.

Mechanisms of Vif-mediated degradation of A3 proteins

As the mechanism of Vif-mediated degradation of hA3G has been the most extensively studied, any discussion should begin with this protein. The HIV-1 Vif protein acts as an adaptor protein that results in the polyubiquitination and proteasomal degradation of hA3G. Vif accomplishes this by binding to hA3G and CBF-β, an allosteric regulator of Vif. Vif then recruits the E3 ubiquitin ligase composed of Cullin 5, Elongin B, Elongin C and ring box subunit 2 (RBX2), resulting in the polyubiquitination of hA3G and targeting the complex to the proteasome for destruction (Conticello et al., 2003; Marin et al., 2003; Sheehy et al., 2003; Stopak et al., 2003; Dussart et al., 2004; Liu et al., 2004; Mehle et al., 2004b; Yu et al., 2004; Kobayashi et al., 2005; Zhang et al., 2011; Jäger et al., 2012). In terms of virus maturation, the end result of Vif-mediated ubiquitination and degradation of hA3G is that hA3G is not incorporated (or not to any meaningful extent) into budding virions. Several deletion and site-directed mutagenesis studies have identified regions on HIV-1 Vif that are important for the interactions with hA3G. One study showed that deletion of aa 43–59 abolished interactions with A3G, while another study using overlapping peptides found that a region from aa 33 to 88 formed a non-linear binding site for A3G (Wichroski et al., 2005; Mehle et al., 2007). Aa 14–17 and 40–44 of the HIV-1 Vif were specifically found to affect the interactions with human A3F and A3G, respectively (Russell & Pathak, 2007). Other investigators showed that domains 25SLVx4Yx9Y38, 69YXXL72, 81LgxxXxW89 and 173EDRW174 were also involved in neutralizing hA3G and hA3F (Chen et al., 2009; Pery et al., 2009; Dang et al., 2010). Highly conserved sites for the interaction of Vif with the proteasome machinery have also been identified. The viral BC box (SLQ(Y/F)LA) and the Zn2+-binding HCC1 domains are necessary for the interactions with Elongin C and Cullin 5, respectively (Stopak et al., 2003; Yu et al., 2003, 2004; Mehle et al., 2004a, b, 2006; Luo et al., 2005). Finally, several studies have shown the importance of the amino acid at position 128 in hA3G. Chimaeric proteins between the hA3G and African green monkey A3G (agmA3G) and subsequently site-directed mutants revealed that the aspartic acid at position 128 of hA3G (a lysine in agmA3A) was necessary to specify interactions with HIV-1 Vif (Bogerd et al., 2004; Schrödelbauer et al., 2004).

Comparison of deamination of the HIV-1 genome in the presence of hA3G and hA3A

Upon transfection of HIV-1Δvif and hA3G into producer cells that do not express hA3G (e.g. 293 cells or SupT1 cells), hA3G is not degraded and is incorporated into the viral core of nascent virions released from producer cells. In the next round of replication, hA3G causes deamination of cytidine residues to uracils in the minus strand of ssDNA during reverse transcription (Yu et al., 2004). The mechanism of deamination involves the glutamic acid protonation of a water molecule, which acts as the catalytic nucleophile in the oxidation of a cytidine amino group, resulting in conversion to uridine and release of NH3 (Jarmuz et al., 2002). The nucleotide context in which deamination occurs is also important, with 5’-CC being the preferred context.
for hA3G, and deamination by hA3G is processive in nature (Chelico et al., 2006). This results in G→A hypermutation of the viral genome, generation of premature termination codons in the viral sequence, or inhibition of replication or degradation of the ssDNA by apurinic apyrimidinic endonuclease (Sheehy et al., 2002; Lecossier et al., 2003; Mangeat et al., 2003; Zhang et al., 2003; Yu et al., 2004).

For hA3G, it is known that the Z3 domain has the catalytic activity hA3G (Fig. 1). Mutation of Glu67 in the active site of the Z1 domain of hA3G was found to be dispensable for antiviral activity, while mutation of Glu259 in the Z3 domain abolished antiviral activity (Navarro et al., 2005). In contrast, another group showed that Ghu259Gln, lacking deaminase activity, still retained antiviral activity (Newman et al., 2005). This study has been challenged by other studies showing that the mutation of the active site of the Z3 domain had little or no antiviral activity (Miyagi et al., 2007; Schumacher et al., 2008; Browne et al., 2009). This indicates that deaminase activity is important for hA3G antiviral activity.

While hA3A is incorporated into nascent HIV-1 or HIV-1Δvif virions, studies have shown that hA3A does not restrict HIV-1 or HIV-1Δvif (Bishop et al., 2004; Goila-Gaur et al., 2007; Aguiar et al., 2008; Schmitt et al., 2011, 2013). These studies were all performed in 293 or HeLa cells. The studies indicated that little or no hA3A is incorporated into the nucleoprotein complex of virions compared with hA3G (Goila-Gaur et al., 2007). Two groups later showed that fusion of the hA3A gene to either the N-terminal domain hA3G or to Vpr resulted in proteins that were incorporated into the nucleoprotein complex and restricted virus replication (Goila-Gaur et al., 2007; Aguiar et al., 2008). Deamination of cytosine residues does occur with hA3A, and this process prefers cytosines in the nucleotide context of 5′-TC (Love et al., 2012; Logue et al., 2014). However, hA3A has been shown to induce fewer mutations in the HIV-1 genome and, unlike hA3G, is not processive in nature (Love et al., 2012). A recent study showed that the recognition loop 1(RL1; previously designated AC-loop1), which is truncated in hominid species, is a determinant of target specificity of hA3A (i.e. 5′-TC). These investigators showed that a chimaeric hA3A/A3G protein having the RL1 region from hA3G was more flexible with respect to the nucleotide 5′ to the deaminated C, deaminating AC, CC and GC with increased frequency. These investigators suggested that the presence of glycine and isoleucine residues at amino acid positions 25 and 26 in the RL1 of hA3A specified the 5′ nucleotide (Logue et al., 2014). This will be important in the discussion of the differences in virus restriction activity by different hominid and Old and New World monkey A3A proteins.

**Virus restriction by A3 proteins by deaminase-independent mechanisms**

In addition to deaminase-dependent activity of hA3G, other studies have shown that hA3G can restrict virus replication through deaminase-independent mechanisms. Different investigators have reported that hA3G can disrupt reverse transcription. Reports of inhibition of primer binding, strand transfer, transcript accumulation and integration have been described (Bishop et al., 2006; Iwatani et al., 2007; Li et al., 2007; Luo et al., 2007; Wang et al., 2012; Adolph et al., 2013; Belanger et al., 2013). A major question that remains to be resolved using animal model systems is whether A3-mediated virus restriction of HIV-1 [or SIV/simian-human immunodeficiency virus (SHIV)] occurs through deaminase-dependent or -independent mechanisms.

**Restriction of retroviruses by hA3A**

As discussed above, it appears that A3A has little effect on the replication of HIV-1Δvif when expressed in 293 or HeLa cells. However, hA3A is expressed at high levels in macrophages and monocytes in response to IFN-α and was shown to inhibit incoming virus without incorporation (Peng et al., 2007; Berger et al., 2011; Koning et al., 2011). These studies suggest that hA3A could restrict replication in the producer cell and without being incorporated into the virion. A recent biochemical study showed that the pH can affect both the specificity and the deaminase activity of hA3A (Pham et al., 2013). These investigators found that purified A3A had optimal cytidine deaminase activity at a pH of 5.8–6.1 and a strict YYCR (where Y is a pyrimidine; R is a purine). In contrast, deaminase activity was reduced by between 13- and 30-fold and had a relaxed cytidine deamination specificity at a pH of 7.4–7.8 (Pham et al., 2013). This is potentially relevant to HIV-1-infected macrocyte-derived macrophages (MDMs) because induction of autophagy has been reported to be necessary for HIV-1 replication in MDMs (Espert et al., 2009; Kyei et al., 2009). MDMs, cultured in the presence of 3-methyladenine, which inhibits class III phosphatidylinositol 3-kinase (PI3K) and autophagosome formation, exhibited substantial reductions in HIV-1 production of both R5 and X4 viruses (Espert et al., 2009).

Retroviruses other than lentiviruses can be restricted by A3A. One study reported that Rous sarcoma virus was moderately susceptible to hA3A (Wiegand & Cullen, 2007). A3A, A3B and A3H haplotype II proteins were shown to restrict human T-cell lymphotropic virus type 1 (HTLV-1) (Ooms et al., 2012). In this study, A3A was incorporated into virions, and a catalytic-site mutant abolished HTLV-1 restriction, suggesting that restriction may be dependent on deaminase activity (Ooms et al., 2012). These investigators also showed that A3A could mutate viral genomes using a technique known as differential DNA denaturation (3D)-PCR, biased to yield AT-rich, low-denaturation amplicons. These authors concluded that multiple independently mutated HTLV-1 proviral genomes could be detected in HTLV-1 cell lines. However, the 3D-PCR findings are not quantitative in nature, and thus the true editing frequency is unknown and may not...
significantly impact HTLV-I replication. Human endogenous retroviruses (HERVs) are a class of transposable elements that comprise over 8% of the human genome (Lander et al., 2001; Bannert & Kurth, 2006; Jern & Coffin, 2008). HERVs most likely originated from germline infections by exogenous retroviruses during primate evolution. Most groups of HERVs are also present in Old World monkeys and apes, which suggests that their introduction into the germline probably occurred more than 30 million years ago (Barbulescu et al., 1999; Bock & Stoye, 2000). HERVs have accumulated numerous point mutations, deletions and insertions that have rendered them non-infectious. One such HERV [HERV-K(MHL-2)] was reconstituted to an infectious virus (Lee & Bieniasz, 2007). This resurrected virus was shown to be restricted by several A3 proteins including hA3A (Lee et al., 2008). This suggests that, during primate evolution, A3 proteins including A3A may have evolved to help combat the onslaught of retroviruses that eventually accumulated in the human germline. More recently, a study analysed the restriction of the retrovirus murine leukemia virus in hA3A and hA3G transgenic mice in a mouse A3-knockout background (Stravou et al., 2014). These investigators showed that hA3G and hA3A restricted murine retroviruses by different mechanisms. They observed that hA3G was packaged into virions and caused extensive deamination of the viral genome, while hA3A was not packaged into virions and restricted infection when expressed in target cells (Stravou et al., 2014). This study also found that hA3A can restrict murine retroviruses by a deamination-independent mechanism.

**Virus restriction by non-human primate A3A proteins**

In contrast to hA3A, a study in rhesus macaques showed that rhA3A was capable of restricting the replication of Vif-deficient SHIVΔvif at similar levels as rhA3G (Schmitt et al., 2011). Another study observed that rhA3A did not restrict replication (Hultquist et al., 2011). However, the reason for this discrepancy is probably due to where the haemagglutinin tag was fused to the rhA3A (Schmitt et al., 2013). Furthermore, hA3A was capable of significantly restricting SHIVΔvif although not to the same extent as rhA3G (Schmitt et al., 2011). These results indicate that hA3A is capable of restricting lentiviruses other than HIV-1Δvif. Similar to conclusions observed above for hA3A (Love et al., 2012), we observed that rhA3A is capable of cytidine deamination but not at the same frequency as rhA3G (Schmitt et al., 2011).

Our results with rhA3A raised the question of whether the A3A proteins from other non-human primates restrict HIV-1/SIV/SHIV and the importance of the 3 aa deletion in hominid A3A proteins. Recently, we examined the restriction properties of A3A proteins from an additional hominid (Lars gibbon; gibA3A), four additional Old World monkeys [African green monkey (agmA3A) and black and white colobus monkey (colA3A)] and two New World monkeys [squirrel monkey (sqmA3A) and northern owl monkey (nomA3A)] (Schmitt et al., 2013). Similar to hA3A, gibA3A did not restrict HIV-1Δvif or SHIVΔvif, indicating that the lack of HIV-1Δvif restriction is probably conserved among hominids. Similar to rhA3A, we found that agmA3A and colA3A restricted the replication of SHIVΔvif and that all three Old World monkey A3A proteins (agmA3A, colA3A, and rhA3A) restricted HIV-1Δvif to some extent, with mndA3A being the most effective. Among the New World monkeys, sqmA3A restricted SHIVΔvif but not HIV-1Δvif, whereas nomA3A restricted neither SHIVΔvif nor HIV-1Δvif (Schmitt et al., 2013). As the A3A proteins from select Old World monkeys could restrict HIV-1Δvif, we investigated the molecular determinants that define restriction. Using the distinct phenotypes of rhA3A (virus restriction) and hA3A (no virus restriction) in epithelial cells, a series of hA3A/rhA3A chimaeric proteins were constructed and their restriction properties on HIV-1Δvif were analysed (Schmitt et al., 2013). Our results showed that a chimaeric A3 protein, designated rh25–33hA3A, which had the 3 aa deletion of hA3A replaced by the amino acids of rhA3A (27SVR29) plus three additional amino acid substitutions in the AC-loop1, restored the restriction properties of hA3A to a similar level as rhA3A restriction of SHIVΔvif (Schmitt et al., 2013). Interestingly, rhA3A aa 25–33 map to the AC-loop1, and molecular modelling suggests that the AC-loop1 of rhA3A has a greater molecular surface compared with hA3A. Due to its predicted solvent accessibility, this additional molecular surface may be an important region for protein–protein interactions that do not occur with hA3A.

The finding that sqmA3A restricted SHIVΔvif is an interesting result, as the New and Old World monkeys branched approximately 30–35 million years ago with the separation of South America from Africa. To date, no lentiviruses have been isolated from New World monkeys. This brings up the question of whether some New World monkey A3A proteins may have evolved to restrict other viruses. One potential candidate is the foamy viruses (FVs), which are exogenous, persistent and non-pathogenic retroviruses in the subfamily Spumaretrovirinae (Switzer et al., 2005). FVs have been isolated from a broad range of mammals including non-human primates (both New and Old World monkeys), horses, cows and cats (Meiering & Linial, 2001; Hussain et al., 2003; Rethwilm, 2010). Previous studies have shown that the FVs have co-speciated with their mammalian hosts. For example, FVs express the Bet protein, which antagonizes A3 restriction and may play a role in both particle release and virus persistence (Saib et al., 1995; Alke et al., 2001; Löchelt et al., 2005; Russell et al., 2005; Chareza et al., 2012). A remarkable feature of the FVs is that they have the capacity to cross species barriers and pose a significant risk of interspecies transmission to humans (Heneine et al., 2003; Switzer et al., 2005). While no disease has
been associated with the transmission of FVs to humans, the ease of primate-to-primate transmission implies the co-evolution of host restriction factors, as viral disease mechanisms are not sufficient to prevent cross-species transmission (Leendertz et al., 2008; Khan, 2009). As the FVs are promising candidates for the development of viral vectors for gene delivery and vaccination, it will be of interest to determine whether New and Old World monkey A3A proteins are capable of restricting the replication of spumaviruses (Schwantes et al., 2003; Trobridge, 2009; Rethwilm, 2010).

**A novel mechanism of virus restriction of HIV-1 by mndA3A, colA3A and debA3A**

A significant finding from our analysis of non-human primate A3A proteins was that colA3A, mndA3A and debA3A restricted the replication of both HIV-1 and HIV-1Avif (Schmitt et al., 2013; Katuwal et al., 2014). Thus, deciphering the mechanism of restriction could yield insight into potential novel antiviral strategies. Attempts to recover HIV-1 released from cells expressing colA3A and the HIV-1 genome resulted in very little Gag p24 (~100-fold reduction) and infectious virus released from the producer cells. This was not observed with the other A3A or A3G proteins examined (Schmitt et al., 2013). We also analysed whether transfection of the viral genome into cells 24 h prior to transfection of the vector expressing colA3A would restrict the replication of HIV-1. Under these conditions, colA3A still restricted replication of HIV-1, suggesting that the early steps of virus replication (such as entry and reverse transcription) were not affected by the presence of colA3A (Schmitt et al., 2013). A recent study showed that both reverse transcription and integration were not affected by the presence of colA3A (Katuwal et al., 2014). Thus, these observations indicate that the mechanism of restriction occurs in the absence of being incorporated into virions and suggests that restriction may be at the level of transcription or translation.

As the substitution of the human A3A AC-loop 1 with that of rhesus A3A restored restriction activities of hA3A, we have also used a chimaeric protein approach to interrogate the structural basis for the restriction of HIV-1 by colA3A. A chimaeric protein with the N-terminal 100 aa fused to the C-terminal half of hA3A (col1-100hA3A) restricted HIV-1 replication, while the reverse chimaera (h1-100colA3A) had no effect on HIV-1 replication (Katuwal et al., 2014). Additional hA3A/colA3A chimaeric proteins revealed that aa 25–33 were critical to colA3A restriction of HIV-1. Comparison of the AC-loop 1 sequence from the A3A proteins of several Old World monkeys revealed that the sequences for three monkey species (rhesus macaque, African green monkey and baboon) were 25DLS(V/I) RGR(H/R)Q33, which is quite different from the AC-loop 1 sequence of the colobus monkey (25KPWV- SGGQRE33). We identified two additional Old World monkey A3A proteins with a similar or exact sequence. These were from *Cercopithecus neglectus* (De Brazza's monkey; debA3A; 25KPVVSGQRE33) and *Mandrillus sphinx* (mandrill; mndA3A; 25KLWVSQHQ33) (Fig. 2). Similar to colA3A, we observed that debA3A and mndA3A restricted HIV-1, with mndA3A consistently being more restrictive than colA3A and debA3A. Additionally, restriction also occurred in the producer cell (Katuwal et al., 2014). While speculative at this point, it will be of interest to determine whether the leucine at position 26 of mndA3A (both colA3A and debA3A have a proline at this position) is responsible for the slightly better restriction activity of mndA3A. As the mndA3A restricted HIV-1 to higher levels than colA3A or debA3A, we constructed a chimaeric hA3A having the AC-loop 1 region of mndA3A (mnd25–33hA3A) and determined whether it could restrict HIV-1. Our results indicated that this chimaeric A3A did indeed restrict HIV-1, confirming that this region is necessary and sufficient for hA3A to restrict HIV-1 in producer cells (Katuwal et al., 2014). The Old World monkeys (family Cercopithecidae) are composed of two subfamilies (Cercopithecinae and Colobinae) and four tribes (Papionini, Cercopithecini, Presbytini and Colobini) (Schrago & Russo, 2003; Schrago, 2007). Our findings indicate that restriction of WT HIV-1 by A3A is represented by species in three tribes (Colobini: colA3A; Cercopithecini: debA3A; and Papionini: mndA3A). These results also reinforce our previous findings that the sequence of the AC-loop 1 region of A3A is critical for restriction of WT HIV-1 by select Old World monkey species. It will be of interest to determine whether other closely related species also express A3A proteins that restrict HIV-1 and whether the hA3A protein can potently inhibit HIV-1 with a few amino acid substitutions in the AC-loop 1 region.

As the data above suggested that the N-terminal region and more specifically the AC-loop 1 region is very important for HIV-1 restriction in producer cells, we determined whether we could truncate colA3A or mndA3A to the N-terminal 100 aa and retain restriction activity. Our results indicated that truncation of these two proteins still restricted HIV-1 replication (Katuwal et al., 2014). As the truncation disrupted the Z domain of this protein, we can conclude that deamination is not required for colA3A or mndA3A restriction of HIV-1 (Katuwal et al., 2014). To the best of our knowledge, this is the first demonstration that a single deaminase domain A3 protein could be significantly truncated and retain virus restriction activity. Together with our data on mnd25–33hA3A, it will be of interest to determine whether a truncated mnd25–33hA3A is also active against HIV-1.

**APOBEC3A and mRNA editing**

*Wilms Tumour 1* (WT1) mutations and variants are implicated in several diseases, including Wilms tumour and acute myeloid leukaemia. WT1 is a regulatory protein with dual tumour suppressor/oncogene activity depending on the isoforms expressed, including the Lys-Thr-Ser (KTS) variant. Recently, two G→A changes were detected
(G1303A and G1586A) in cDNA clones from non-progenitor blood mononuclear cells (Nivaran et al., 2015). These investigators showed that hA3A was expressed at high levels in non-progenitor cells compared with progenitor cells. Using small interfering RNA, they showed that knockdown of A3A expression, but not of A3B, A3D or A3F, led to near complete reversal of WT1 c.1303G→A. Furthermore, overexpression of A3A in a Fujioka cell line resulted in a significant increase in the WT1 c.1303G→A change. This represents the first report detailing a specific G→A mRNA editing of transcripts by hA3A. This was followed by a second study that showed that A3A, which is expressed at high levels in monocytes and macrophages, induced RNA editing during the polarization of M1 macrophages and in monocytes under hypoxic conditions or in response to IFN (Sharma et al., 2015). It will be of interest to determine whether hA3A can edit viral mRNA from retroviruses, especially within infected macrophages where hA3A is expressed at high levels.

### A3A restriction of DNA viruses

hA3A has been reported to restrict certain DNA viruses (Chen et al., 2006; Vartanian et al., 2008; Henry et al., 2009; Narvai et al., 2009; Suspène et al., 2011b; Ahasan et al., 2015; Warren et al., 2015). The parovirus adeno-associated virus 2 (AAV-2) was shown to be restricted by hA3A (Chen et al., 2006). These investigators found that mutation of the cysteine at position 106 of the canonical deaminase motif (H-x-E-x_{23-28}-PC-x_{2-4}-C) to a serine did not affect its nuclear localization but did not restrict AAV-2, suggesting a role of the enzymatic site in restriction. In a subsequent study by the same group of investigators using comprehensive mutagenesis of A3A, they found that regions outside the catalytic site also contributed to A3A antiviral activities. They also found that deaminase-defective A3A mutants blocked the replication of both AAV-2 and the autonomous parovirus minute virus of mice (Narvai et al., 2009). In a study on herpes simplex virus type 1 (HSV-1), investigators found that A3A could edit a small fraction of HSV-1 genomes without seriously impacting viral titres (Suspène et al., 2011b). This suggests that A3A probably does not play a significant role in restricting the herpesviruses tested. Another study analysed the ability of the seven A3 proteins to edit viral mRNA from retroviruses, especially within infected macrophages where hA3A is expressed at high levels.

### A3A inhibition of retrotransposition

The retroelements, which includes the long interspersed nuclear elements (LINE-1), small interspersed nuclear elements, Alu elements and LTR retroelements (including human endogenous retroviruses and retrotransposons), comprise a large family of transposable genetic elements that make up approximately 42% of the human genome and have shaped the genomes of humans and other mammals over millions of years. Some retroelements can cause genetic diseases through retrotransposition events that occur not only in germ cells but also in somatic cells, posing a threat to genomic stability throughout all cellular populations. In an effort to combat retroelements, mammals have developed an innate mechanism that provides resistance against the deleterious effects of retrotransposition. Within this, seven members of the A3 family of cytidine deaminases serve as highly active, intrinsic, antiretroviral host factors. Based on their preferential expression in germ cells, in which retrotransposons may be active, it is likely that A3 proteins were acquired through mammalian evolution primarily to inhibit retrotransposition and thereby maintain genomic stability in these cells. The A3A protein, like the other six A3 proteins, has been shown to inhibit retrotransposition. In vitro experiments have demonstrated that ectopically expressed hA3A effectively inhibits the retrotransposition of LINE-1, Alu, intracisternal A particle (IAP) and MusD (Boger et al., 2006a, b; Chen et al., 2006; Muckenfuss et al., 2006; Kinomoto et al., 2006).
et al., 2007; Niewiadomska et al., 2007; Koito & Ikeda, 2011) through a deaminase-independent mechanism.

A3A and degradation of foreign DNA

hA3A has been reported to degrade foreign DNA introduced into cells (Stenglein et al., 2010). These investigators used an assay that was based on the number of fluorescent cells following the transfection of plasmids expressing GFP and hA3A. They found a decrease in the level of GFP-positive cells at 5 days but not at 2 days post-transfection, and proposed a model in which the foreign DNA is ‘degraded’, which leads to an IFN response and expression of A3A. This results in DNA deamination and excision of uracils by UNG2, and degradation. However, the mechanism for how foreign DNA is recognized is unknown. They concluded that related proteins exist in all vertebrates and that it may be a conserved innate immune mechanism. It could be argued that DNA viruses or viruses with DNA intermediates most likely constitute the single most common introduction of foreign DNA. Furthermore, many DNA viruses (e.g. poxviruses, herpesviruses, adenoviruses and some parvoviruses) will have entered, replicated, assembled and been released from cells within 48 h of infection, which may be concomitant with death of the cell. Similarly, most retroviruses will have entered, reverse transcribed their genomes and been integrated within 48 h. This study raises the question of why, if this is such an important innate immune mechanism, does it not occur early after transfection? Additionally, if this is evolutionarily conserved, it should also occur in other hominids and other non-human primates. The ability of A3A proteins from rhesus macaques and black-and-white colobus monkey to degrade foreign DNA has also been examined (Schmitt et al., 2011, 2013). In our assays, we introduced a vector expressing A3A protein 24 h prior to transfection of a vector expressing GFP. In these assays, we obtained more than 80 % of cells expressing GFP. While we observed a decrease in the level of fluorescent cells transfected with hA3A, neither rhA3A nor colA3A caused a decrease in the percentage of fluorescent cells (Schmitt et al., 2011, 2013).

DNA damage and A3A

Of the seven A3 proteins, three (A3A, A3C and A3H) have a nucleocytoplasmic localization (Muckenfuss et al., 2006). Thus, the potential exists for these proteins to deaminate cytidine residues of chromosomal DNA. The overexpression of A3A in cultured cells has been shown by some investigators to result in mutation of both nuclear and mitochondrial DNA (Landry et al., 2011; Suspène et al., 2011a; Aynaud et al., 2012; Lackey et al., 2013; Mussil et al., 2013). Moreover, ectopic expression of A3A in HEK293 cells is generally toxic to cells.

Histone H2AX is a variant histone that represents approximately 10 % of the total H2A histone protein in normal human fibroblasts (Rokagou et al., 1998, 1999). H2AX is required for checkpoint-mediated cell-cycle arrest and DNA repair following dsDNA breaks (Yuan et al., 2010). DNA damage, caused by ionizing radiation, UV light or radiomimetic agents, results in rapid phosphorylation of H2AX at Ser139 by PI3K-like kinases, including ATM, ATR and DNA-PK (Rokagou et al., 1998; Burma et al., 2001). Within minutes following DNA damage, H2AX is phosphorylated at Ser139 at sites of DNA damage and this variant is known as γH2AX (Rokagou et al., 1999).

Human A3A was shown to induce a DNA damage response and phosphorylation of H2AX (Landry et al., 2011). More recently, using an antibody raised against an A3A-specific peptide, other investigators found that A3A expression in terminally differentiated macrophages was predominantly cytoplasmic and was not genotoxic (Land et al., 2013). However, as mentioned earlier, the peptide to which this antiserum was raised is identical in hA3B and 90 % identical to hA3G. However, as A3G is expressed in macrophages and is localized to the cytoplasm, this could lead to predominant cytoplasmic staining. Finally, a recent study showed that, like hA3A, A3A proteins from different mammalian species are also capable of hypermutating genomic DNA, suggesting that this activity has been conserved over approximately 148 million years of evolution (Caval et al., 2014a).

Possible roles of A3A in cancer

As the A3 proteins have a specificity for ssDNA, which is available during DNA replication and transcription of the genome, these proteins have the potential to be mutagens that could in turn lead to cancer. One study examined the editing of the HPV genome and found that hA3A, hA3C and hA3H were expressed in keratinocytes (Vartanian et al., 2008). They found that HPV-1a genomes from one of six plantar warts had consistent G→A and C→T edits, suggesting that both DNA strands were edited. Similar results were obtained in vitro when 293T cells were co-transfected with the HPV-1a genome and either hA3A, hA3C or hA3H. HPVs are known to cause cervical cancer with subtypes 16 and 18 most commonly associated with this cancer (Guan et al., 2012). These investigators found that HPV-16 sequences were edited in pre-cancerous cervical biopsies and concluded that stochastic or transient overexpression of these A3 proteins could expose the viral genome to mutations that could influence the development of tumours. Other studies have associated the A3A/A3B deletion and breast cancer of women of both Chinese and European ancestry, which is one of the strongest common genetic risks identified to date (Kidd et al., 2007; Long et al., 2013; Xuan et al., 2013). These results appear to be discordant from those published recently suggesting that hA3B overexpression might contribute to breast cancer (Burns et al., 2013). The deletion removes genomic sequences from exon 5 of hA3A transcripts to exon 8 of hA3B, resulting in a 29.5 kb deletion and fusion of A3A to the last exon of A3B. While the deletion...
did not affect the coding sequences of A3A, this transcript would contain the hA3A 5’ upstream regulatory sequences and the 3’ UTR from hA3B (A3A-UTR3B) (Long et al., 2013). Recently, it was shown that A3A protein expression from the A3A-UTR3B transcripts was 20 times more active than A3A-UTR3A or A3A-ΔUTR (Caval et al., 2014b). Deletion of the entire A3B gene correlates with a higher risk of developing some cancers and a higher overall mutation burden per cancer genome (Nik-Zainal et al., 2014). These authors have suggested that because A3A expression correlates with inflammatory environments (such as type 1 and 2 IFNs) and chronic inflammation is usually associated with the genesis of cancer, A3A-induced damage may contribute to somatic mutation selection in different cancers. Genome-wide association studies have identified small nucleotide polymorphisms in a non-genic region approximately 25 kb centromeric of hA3A that were associated with bladder cancer (Rothman et al., 2010; Golka et al., 2011). Finally, a study revealed a homozygous deletion in the A3A and A3B genes in pancreatic cancer (Liang et al., 2014). Thus, whether hA3A overexpression leads to an increased incidence of cancer is still highly controversial.

Conclusions and future studies

While the A3A protein shares many structural properties with the other members of the A3 family of proteins, it has unique properties that could be either beneficial or detrimental to the host. The exact function of A3A in the host is still unclear. Similar to several other A3 proteins, A3A has probably contributed to the restriction and inactivation of HERVs, which comprise 8% of the human genome (Jern & Coffin, 2008; Magiorkinis et al., 2013). Unlike the better-studied A3G, which readily restricts HIV-1Δvif by incorporation of hA3G into virus from the producer cell and restriction by cytidine deamination of the viral genome in target cells, hA3A does not significantly restrict the replication of HIV-1Δvif and its cytidine deamination does not appear to be processive in nature (Goila-Gaur et al., 2007). However, the findings that hA3A can restrict HIV-1 in macrophages and that the A3A proteins from other Old World monkey species such as the rhesus macaque (rhA3A) can restrict SHIVΔvif provides evidence that these proteins can restrict lentiviruses (Peng et al., 2007; Berger et al., 2011; Schmitt et al., 2011). Structure–function studies suggest that the AC-loop1 region (also referred to as RL1) may contain determinants for the restriction of HIV-1Δvif. Interestingly, the A3A proteins from hominid species have a 3 aa indel in this region (see Fig. 2). Thus, from an evolutionary standpoint, it will be of interest to learn why the hominids have this 3 aa deletion in their A3A proteins. hA3A can also restrict other retroviruses such as HTLV-1, although the mechanism is probably different from A3A restriction of SHIV/SIV. This protein can also restrict other viruses such as paroviruses (both AAV-2 and spontaneously replicating paroviruses) and papillomaviruses. The available evidence indicates that this probably occurs through a deaminase-independent mechanism (Narvaiza et al., 2009; Warren et al., 2015). Currently, the mechanism used by hA3A to restrict these viruses is unknown. However, it was shown that HPV upregulation of A3A in keratinocytes required the expression of the E7 oncoprotein (Warren et al., 2015). This raises questions of whether hA3A can interact with one or more viral proteins from these viruses to restrict replication and whether an hA3A protein containing the AC-loop1 region of an Old World monkey can still restrict these viruses. Recent studies have shown that hA3A can edit at the RNA level. If these results are confirmed, it will be of interest to decipher whether A3A can contribute to sequence variation in the genomes of RNA viruses. Overexpression of hA3A has been shown by several investigators to be toxic to cells, and others have suggested that, because of its expression in the nucleus, it may cause cytidine deamination of the host cell chromosome. Whether the upregulation of hA3A during times of immune activation leads to genomic mutations and cancer is currently unknown. However, the finding that certain cancers are associated with a 25 kb deletion in the A3 locus that results in the deletion of hA3B and an upregulation in hA3A expression requires further study.

Acknowledgements

This work was supported by NIH grant R21 AI108391 to E. B. S. and NIH R01 AI090795 to M. L. S.

References


Riefsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L. & Harris, R. S. (2010). Quantitative profiling of the full APOBEC3


