Loop 1 of APOBEC3C regulates its antiviral activity against HIV-1

Ananda Ayyappan Jaguva Vasudevan[1,*,#,§], Kannan Balakrishnan[1,2,#], Christoph G. W. Gertzen[3,4,5], Fanni Borvető[6], Zeli Zhang[1,§§], Anucha Sangwiman[1], Ulrike Held[7], Caroline Küstermann[7], Sharmistha Banerjee[2], Gerald G. Schumann[7], Dieter Häussinger[1], Ignacio G. Bravo[6], Holger Gohlke[3,4], Carsten Münk[1,*]

[1]Clinic for Gastroenterology, Hepatology, and Infectiology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [2]Department of Biochemistry, School of Life Sciences, University of Hyderabad, Gachibowli, Hyderabad, India. [3]Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [4]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre & Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, Jülich, Germany. [5]Center for Structural Studies (CSS), Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [6]Centre National de la Recherche Scientifique, Laboratory MIVEGEC (CNRS, IRD, Uni Montpellier), Montpellier, France. [7]Division of Medical Biotechnology, Paul-Ehrlich-Institute, Langen, Germany.

[#] A.A.J.V and K.B contributed equally to this article.

Present address:

[§]Structural Cell Biology Group, Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences (NIEHS), NIH, Research Triangle Park, NC 27709, USA.

[§§]La Jolla Institute for Immunology, La Jolla, CA 92037, USA.

[*]Address correspondence to Ananda Ayyappan Jaguva Vasudevan, Structural Cell Biology Group, Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences (NIEHS), NIH, Research Triangle Park, NC 27709, USA. anand.jaguvavasudevan@nih.gov. Carsten Münk, Clinic for Gastroenterology, Hepatology, and Infectiology, Medical Faculty, Heinrich Heine University Düsseldorf, Building 23.12.U1.82, Moorenstr. 5, 40225 Düsseldorf, Germany, Tel: +49 (0) 211-81-10887, carsten.muenk@med.uni-duesseldorf.de.

Short title: Evolution and role of loop 1 of A3C

ABSTRACT

APOBEC3 deaminases (A3s) provide mammals with an anti-retroviral barrier by catalyzing dC-to-dU deamination on viral ssDNA. Within primates, A3s have undergone a complex evolution *via* gene duplications, fusions, arms race and selection. Human APOBEC3C (hA3C) efficiently restricts the replication of viral infectivity factor (*vif*)-deficient *Simian immunodeficiency virus* (SIVΔ*vif*), but for unknown reasons, it inhibits HIV-1Δ*vif* only weakly. In catarrhines (Old World monkeys and apes), the A3C loop 1 displays the conserved amino acid pair WE, while the corresponding consensus sequence in A3F and A3D is the largely divergent pair RK, which is also the inferred ancestral sequence for the last common ancestor of A3C and of the C-terminal domains of A3D and A3F in primates. Here, we report that modifying the WE residues in hA3C loop 1 to RK leads to stronger interactions with substrate ssDNA, facilitating catalytic function, which results in a drastic increase in both deamination activity and in the ability to restrict HIV-1 and LINE-1 replication. Conversely, the modification hA3F_WE resulted only in a marginal decrease in HIV-1Δ*vif* inhibition. We propose that the two series of ancestral gene duplications that generated A3C, A3D-CTD and A3F-CTD allowed neo/subfunctionalization: A3F-CTD maintained the ancestral RK residues in loop 1, while diversifying selection resulted in the RK→WE modification in Old World anthropoids' A3C, possibly allowing for novel substrate specificity and function.

HIGHLIGHTS

- Loop 1 residues in primate APOBEC3C and closely related A3s are under positive selection
- A3C in monkeys and apes replaced the highly active residues RK to WE during A3C gene evolution
- Replacing WE residues in loop 1 of A3C with RK turns A3C into a superior restriction factor.
- Loop 1 residues RK caused enhancement of A3C catalytic activity by improved DNA interaction.
- The cytosine deamination hotspots and cell-wide distribution of A3C were not affected by loop 1 residues RK.

Keywords: APOBEC3C_A3F_cytidine deaminase, Sooty Mangabey monkey, human immunodeficiency virus (HIV), LINE-1, evolution

INTRODUCTION

The APOBEC3 (A3) family of single-stranded (ss) DNA cytidine deaminases builds an intrinsic immune defense against retroviruses, retrotransposons, and other viral pathogens [1; 2; 3; 4]. There are seven human A3 proteins that possess either one (A3A, A3C, and A3H) or two (A3B, A3D, A3F, and A3G) zinc (Z)-coordinating DNA cytosine deaminase motifs, HXE[X_{23-28}]PC[X_{2-4}]C (where X indicates a non-conserved position) [5; 6; 7]. A3C is the only single-domain A3Z2 protein in humans. During primate evolution, the ancestor of the *A3C* gene duplicated several times and formed double-domain *A3Z2-A3Z2* genes, which are *A3D* and *A3F* [6]. Initially, A3G was characterized as the factor capable of restricting infection of HIV-1 lacking Vif (viral infectivity factor) protein in non-permissive T cell lines and its biochemical properties and biological functions have been extensively studied [3; 8; 9; 10; 11].

The encapsidation of A3s into the viral particles is crucial for virus inhibition [12; 13; 14; 15; 16; 17]. During reverse transcription, viral core-associated A3 enzymes can deaminate cytidines (dC) on the retroviral ssDNA into uridines (dU). These base modifications in the minus-strand DNA cause coding changes and premature stop codons in the plus-strand viral genome (dG→dA hypermutation), which impair or suppress viral infectivity [2; 9; 18; 19; 20; 21]. In addition to the mutagenic activity of the virus-incorporated A3s, deaminase-independent mechanisms of restriction have been identified such as impeding reverse transcription or inhibiting DNA integration [22; 23; 24; 25; 26; 27]. To counteract A3 mediated inhibition, lentiviruses evolved the Vif protein, which physically interacts with A3s, targeting them for polyubiquitination and proteasomal degradation [28; 29; 30]. These A3-Vif interactions are often species-specific and an important factor reducing virus cross-species transmission [31; 32; 33; 34; 35; 36].

In addition to A3G, A3D, A3F, and A3H were shown to restrict HIV-1 lacking *vif* (HIV-1Δ*vif*) [2; 35; 37; 38; 39; 40]. Recently, mutation signatures resulting from the catalytic activity of nuclearly localized A3s (especially A3A, A3B, and likely A3H) were reported in several cancer types [41; 42; 43; 44; 45; 46; 47; 48]. The knowledge about A3C is rather sparse, it is distributed in both cytoplasm and nucleus [49] and does not seem to be a causative agent of chromosomal DNA mutations. In addition, human A3C is known to act as a potent inhibitor of *Simian immunodeficiency virus* from African green monkey (SIVagm) and from rhesus macaque (SIVmac), limits the infectivity of herpes simplex virus, certain human papillomaviruses, murine leukemia virus, *Bet*-deficient foamy virus, and hepatitis B virus, and represses the replication of LINE-1 (L1) endogenous retrotransposons [49; 50; 51; 52; 53; 54; 55; 56; 57; 58; 59]. In contrast, the restrictive role of A3C on HIV-1 is marginal and there are several contradictory findings regarding its viral packaging and cytidine deamination activity [40; 50; 60; 61; 62]. Notably, A3C is ubiquitously expressed in lymphoid cells [5; 50; 63; 64], mRNA expression levels of A3C are higher in HIV-infected $CD4^+$ T lymphocytes [40; 50], and significantly elevated in elite controllers compared to ART-suppressed individuals [65]. A3C was found to moderately deaminate HIV-1 DNA if expressed in target cells of the virus with the effect of increasing viral diversity rather than causing restriction [63].

The crystal structure of A3C and its HIV-1 Vif-binding interface have been solved [66]. The study revealed several key residues in the hydrophobic V-shaped groove formed by the α2 and α3 helices of A3C that facilitate Vif binding resulting in proteasome-mediated degradation of A3C [66]. We have extended this finding and identified additional Vif interaction sites in the α4 helix of A3C [67]. Apart from a previous study that predicted putative DNA substrate binding pockets

[55], biochemical and structural aspects of A3C enzymatic activity and their relevance for antiviral activity remain hitherto not well investigated [3; 4].

Recently, we have shown that increasing the catalytic activity of A3C by an S61P substitution in loop 3 is not sufficient to restrict HIV-1Δ*vif* [68]. It is unknown why A3C can potently restrict SIVΔ*vif* while HIV-1Δ*vif* is largely resistant, despite the fact that wild-type (WT) human A3C possesses reasonable catalytic activity and is encapsidated efficiently into retroviral particles [68]. Here we set out to further explore the determinants of A3C's restrictive capacity of HIV-1. We generated a synthetic open reading frame derived from sooty mangabey monkey genome *(smm, Cercocebus atys (torquatus) lunulatus)* coding for an A3C-like protein (hereafter called smmA3C-like protein) capable of restricting HIV-1 to similar or higher extents than human A3G. This A3C-like protein was reported to be resistant to HIV-1 Vif-mediated depletion [67]. Using this smmA3C-like protein as a tool, we dissected a novel structure-function relationship of hA3C and discovered the importance of loop 1 for A3C to achieve strong inhibition of HIV-1.

RESULTS

Identification of an A3Z2 protein with enhanced antiviral activity

To determine whether A3C from non-human primates can potently restrict HIV-1Δ*vif* propagation, we produced HIV-1Δ*vif* luciferase reporter virus particles with A3C (an A3Z2 protein) from human, rhesus macaque, chimpanzee (cpz), African green monkey (agm), and with human A3G (an A3Z2-A3Z1 double domain protein), or with a synthetic smmA3C-like

protein and tested the infectivity of the respective viral particles. Viral particles were pseudotyped with the glycoprotein of *Vesicular stomatitis virus* (VSV-G) and normalized by reverse transcriptase (RT) activity before infection. The firefly luciferase enzyme activity of infected cells was quantified two days post infection. Figure 1A shows the level of relative infectivity of HIV-1Δ*vif* in the presence of the tested A3 proteins. Human, rhesus, chimpanzee, and African green monkey A3C proteins reduced the relative infectivity of HIV-1Δ*vif* similarly by approximately 60 to 70%. Conversely, smmA3C-like protein inhibited HIV-1Δ*vif* replication by more than one order of magnitude (Fig. 1A). Human A3G served as a positive control for major anti-HIV-1 activity. Viral vector-producing cells showed that expression levels of smmA3C-like protein and agmA3C were lower than those of A3Cs from human, rhesus, and cpz (Fig. 1B). Efficiency of viral incorporation of the smmA3C-like protein was similar to that of hA3G, but much lower compared to hA3C (Suppl. Fig. S1A).

The smmA3C-like construct was originally described to express A3C of sooty mangabey monkey [67]. However, using alignments of primate A3Z2 and related A3 proteins, we found that the open reading frame consists of exons from both smm*A3C* and smm*A3F* genes. We fused these exons during the PCR amplification step, which occurred because of the high sequence similarity and poor annotation of the smm genome (see discussion section). In the smmA3C-like construct, first (coding for amino acids ^{1}MNPQIR^{6}) and last "exon" (amino acids ^{153}FKYC to EILE^{190}) were derived from smm*A3C* (i.e, coding regions of exon 1 and exon 4 of the smm*A3C* gene) while second (amino acids ^{7}NPMK to FRNQ^{58}) and third "exon" (amino acids ^{59}VDPE to GYED^{152}) in smmA3C-like were of smm*A3F* origin (smmA3F C-terminal domain, CTD, exon 5 and exon 6 of smm*A3F* gene) (Suppl. Fig. S1B). To compare the deamination activity of smmA3C-like to the WT proteins, we cloned the genuine smmA3C and smmA3F-CTD.

Immunoblot analysis of cell lysates confirmed that cellular expression of smmA3C-like and smmA3C (WT) were comparable, but the smmA3F-CTD construct failed to yield detectable levels of protein in transfected cells (Fig. 1C). In contrast to our expectations, only the smmA3C-like protein and not smmA3C showed enhanced cytidine deaminase activity (Fig. 1D). Not surprisingly, like hA3C[68], smmA3C-like protein formed intracellular RNAse resistant oligomers or high molecular mass (HMM) complexes and did not self-associate in the cytosol (data not shown).

Because restriction of HIV-1Δ*vif* by smmA3C-like protein was similar to or slightly stronger than restriction by hA3G (Fig. 1A), we analyzed the DNA-editing capacity of these A3s during infection by "3D-PCR" [68; 69]. DNA sequences in which cytosines are deaminated by A3 activity contain fewer GC base pairs than non-edited DNA, resulting in a lower melting temperature than the original, non-edited DNA. Therefore, successful PCR amplification at lower denaturation temperatures (T_d) (83.5 - 87.6°C) by 3D-PCR indicates the presence of A3-edited sequences. 3D-PCR amplification of viral genomic cDNA with samples of cells infected with HIV-1Δ*vif* viruses encapsidating hA3C, rhA3C, cpzA3C, or agmA3C yielded amplicons at T_d ≥86.3°C, whereas the activity of smmA3C-like protein allowed to produce amplicons at T_d <84.2°C. In control reactions using virions produced in the presence of hA3G, PCR amplification of viral DNA was detectable at lower T_d (85.2°C and weakly at 84.2°C) (Fig. 1E). Importantly, using the vector control sample (no A3), PCR amplicons could be amplified only at higher T_d (87.6°C). To study the effect of smmA3C-like protein in HIV-1Δ*vif*, PCR products generated on smmA3C-like protein-edited samples formed at 84.2°C were cloned and independent clones were sequenced. The novel smmA3C-like protein caused hypermutation in HIV-1Δ*vif* with a rate of 17.16% and predominantly favored the expected GA dinucleotide

context (Suppl. Fig. S2A). Thus, smmA3C-like protein caused a higher G→A mutation rate in HIV-1Δ*vif* than our previously described enhanced activity mutant A3C.S61P (see Figs. 2A and 2B for sequence and structure), A3G and A3F [68]. In addition, we applied qualitative *in vitro* cytidine deamination assays using A3 proteins isolated from HIV-1Δ*vif* and SIVagmΔ*vif* viral particles [70; 71]. This PCR-based assay depends on the sequence change caused by A3s converting a dC→dU in an 80-nucleotide (nt) ssDNA substrate harboring the A3C-specific TTCA motif. Catalytic deamination of dC→dU by A3C is then followed by a PCR that replaces dU by dT generating an MseI restriction site. The efficiency of MseI digestion was monitored by using a similar 80-nt substrate-containing dU instead of dC in the recognition site. As expected, encapsidation of hA3C and hA3C.S61P into the HIV-1Δ*vif* particles, did not yield a substantial product resulting from ssDNA cytidine deamination [68], whereas smmA3C-like protein generated high amounts of deamination products (Fig. 1F). Using smmA3C-like protein, the deamination products were observed even after transfection of 10-fold smaller amounts of expression plasmid during virus production. In contrast, A3C and A3C.S61P proteins isolated from SIVagmΔ*vif* particles produced the expected deamination products, whereas smmA3C-like protein exhibited the strongest catalytic activity, regardless of whether encapsidated in SIVagmΔ*vif* or HIV-1Δ*vif* particles (Fig. 1F). Taken together, we conclude that smmA3C-like protein inhibits HIV-1 by cytidine deamination causing hypermutation of the viral DNA.

Identification of the regulatory domain of smmA3C-like protein that mediates HIV-1 restriction

Amino acid sequence identity and similarity between hA3C and smmA3C-like protein reach 77.9% and 90%, respectively (Fig. 2A). To facilitate the identification of distinct determinants of smmA3C-like protein that confer HIV-1 inhibition, ten different hA3C/smmA3C-like chimeras were constructed [67] (Fig. 2C). Next, viral particles containing different chimeric proteins were produced and their infectivity was tested. As shown in Fig. 2D, chimeras C2, C4, and C8 strongly reduced the infectivity of HIV-1Δ*vif*. Especially, chimera C2 (hA3C harboring a swap of 36 residues of the smmA3C-like protein at the N-terminal end) inhibited HIV-1Δ*vif* replication by about two orders of magnitude. In comparison, chimeras C6 and C9 reduced viral infectivity by only 72% relative to vector control (Fig. 2D).

Next, we determined the intracellular expression and virion incorporation efficiency of the chimeras by immunoblot analysis. Chimeras C2, C3, C5, C7, and C9, which contain residues 37 to 76 of hA3C (Fig. 2C), were more highly expressed than C1, C4, C6, and C10 (Suppl. Fig. S2B). Specifically, chimera C2 displayed higher protein levels than hA3C while C10 protein was below the detection threshold. Chimeras, C2, C4, C6, C7, and C9 were found to be encapsidated in HIV-1Δ*vif* (Suppl. Fig. S2B, viral lysate). In particular, C3 and C5 were less efficiently packaged into viral particles although they were present at higher intracellular expression levels. Conversely, C6 produced less protein but its viral incorporation was higher than that of C3 or C5. In addition, we analyzed the *in vitro* cytidine deaminase activity of these chimeras as described above (Suppl. Fig. S2C). Here we used lysates of transfected HEK293T cells to readily evaluate the catalytic activity of the chimeric A3Cs. Only chimers C2 and C4 showed the level of deamination similar to those produced by smmA3C-like protein (Suppl. Fig. S2C). Taken together, chimeras C2 and C4 have the strongest HIV-1Δ*vif*-restricting effect

among all tested chimeras and display corresponding *in vitro* deamination activity. Due to its superior antiviral activity, we mainly focused on chimera C2 in our following experiments.

Synergistic effects of residues in the RKYG motif of chimera C2 and smmA3C-like protein control their potent antiviral activity

To identify the specific residues in chimera C2 that are essential for its anti-HIV-1 activity, we targeted two N-terminal motifs of C2, namely ^{13}DPHIFYFH^{20} (shortly "DHIH") and ^{25}RKAYG^{29} (named "RKYG") as presented in the sequence alignments of Fig. 2A, and generated variants of C2 by swapping one, two, or four amino acids with the analogous residues of hA3C as presented in Fig. 2E. First, we cloned the C2 variants C2.DH-YG (YGTQ motif of helix α1) and C2.RKYG-WEND (WEND motif of loop 1, Figs. 2A and 2B) and tested their anti-HIV-1 and deamination activity. This pilot experiment revealed that loop 1 motif RKYG but not α1 helix motif DHIH in C2 is essential for its activity (Fig. 2F and Suppl. Fig. S3A). Hence, we constructed the mutants C2.R25W, C2.K26E, C2.Y28N, and C2.G29D (Fig. 2E) and tested them for catalytic and antiviral activity. The results of the deamination assay further demonstrated that the DH motif in C2 is not relevant for its potent catalytic activity, as the C2.DH-YG acted similar to C2 (Suppl. Fig. S3A), but mutation of the RKYG motif in the RKYG-WEND variant resulted in a loss of deamination activity (Suppl. Fig. S3A). Interestingly, none of the single amino acid changes in RKYG (R25W, K26E, Y28N, and G29D) resulted in the loss-of-function of C2, albeit the catalytic activities of R25W and K26E were partially reduced (Suppl. Fig. S3A). Consistent with the data obtained from the *in vitro* assay, the chimeric C2.RKYG-WEND variant failed to restrict the infectivity of HIV-1Δ*vif* (Fig. 2F). Immunoblot analysis of cell and viral lysates confirmed

that cellular expression and viral encapsidation of these variants were comparable (Suppl. Fig. S3B). Finally, to test the *in vivo* DNA-editing capacity, we performed 3D-PCR analysis using C2, C2.DH-YG, and C2.RKYG-WEND variants. As presented in the 3D-PCR experiment of Suppl. Fig. S3C, only HIV-1Δ*vif* particles produced in the presence of A3C chimera C2 and its mutant C2.DH-YG generated amplicons that were detected at low-denaturation temperature, and C2.RKYG-WEND behaved similar to the vector control. Likewise, replacing RKYG with WEND in the smmA3C-like protein (Fig. 2E) inhibited its antiviral activity (Figs. 3A and 3B), DNA-editing capacity of HIV-1 genomes (Fig. 3C), and catalytic activity *in vitro* (Fig. 3D) as did the active site mutant E68A.

The WE-RK mutation in loop 1 of hA3C determines its strong deaminase-dependent antiviral function

Mutational changes of the RKYG motif to WEND residues in loop 1 of C2 and smmA3C-like protein resulted in complete loss of enzymatic functions and anti-HIV-1 activities (Figs. 2F, 3A, 3C, 3D, and Suppl. Figs. S3A and S3C). To identify the residues in hA3C that are critically required for the deaminase-dependent antiviral activity against HIV-1Δ*vif*, we mutated the loop 1 of hA3C with ^{25}WE^{26} to ^{25}RK^{26} and ^{28}ND^{29} to ^{28}YG^{29} residues and compared their antiviral capacity (see A3C alignment and ribbon diagram Figs. 2A and 2B)**.** As controls, we included additional mutants such as a catalytically inactive non-Zn^{2+}-coordinating C97 mutant, A3C.C97S [55], and the variants A3C.S61P [68] and A3C.S188I [72] exhibiting enhanced deaminase activity. Compared to WT hA3C, WE-RK greatly enhanced inhibition of HIV-1Δ*vif* and the ND-YG variant behaved like WT A3C, while S61P and S188I demonstrated only marginally

increased HIV-1Δ*vif* restriction (Fig. 4A). Importantly, active site mutant A3C.C97S did not inhibit HIV-1Δ*vif* (Fig. 4A). Enhancement of the antiviral activity of hA3C.WE-RK compared to WT hA3C neither appear to result from higher protein expression in the virus producer cells nor from differences in encapsidation, as demonstrated in a titration experiment that directly compared these features for both proteins (Suppl. Fig. S4A).

Next, we asked if the antiviral activity of A3C.WE-RK is deamination-dependent. To achieve this, we introduced the C97S mutation in A3C.WE-RK. Additionally, we compared the ancillary effect of mutants such as S61P [68] and S188I [72] by introducing these mutations in the WE-RK variant of A3C. As expected, the inhibitory activities of A3C.WE-RK, A3C.WE-RK.S61P, and A3C.WE-RK.S61P.S188I against HIV-1Δ*vif* were abolished by the active site ablating mutation C97S, indicating the importance of the enzymatic activity of A3C (Fig. 4B). In comparison, introducing either the single mutation S61P or the double mutation S61P.S188I did not considerably change the activity of A3C.WE-RK (Fig. 4B). Immunoblot analysis of cell and viral lysates demonstrated that hA3C and all mutants (except A3C.WE-RK.S61P.S188I.C97S mutant) were expressed at comparable levels (Fig. 4C). However, viral incorporation of A3C.C97S, A3C.WE-RK.C97S, A3C.WE-RK.S61P.C97S, and WE-RK.S61P.S188I.C97S was slightly decreased relative to that of WT and mutant proteins that do not contain the C97S mutation (Fig. 4C). Moreover, we confirmed the effects of all mutants on HIV-1Δ*vif* propagation by 3D-PCR (Fig. 4D) and deamination assays *in vitro* (Fig. 4E). In both assays, we found that the C97S mutation destroyed the function of all A3C variants. Thus, we conclude that the loop 1-mediated enhanced activity of hA3C.WE-RK is dependent on catalytic deamination.

To address if the cellular localization of A3C is affected by the WE-RK mutations, we used confocal microscopy. HeLa cells were transfected with the HA-tagged hA3C or hA3C.WE-RK and the proteins were visualized by applying an anti-HA antibody. Both proteins, hA3C and hA3C.WE-RK were localized in cytoplasm and nucleus (Figs. 5A and 5B). This distribution was found in 65.5% and 75% of cells expressing hA3C or that of hA3C.WE-RK, respectively. Only 20% or 10% of the cells expressing hA3C or hA3C.WE-RK displayed these proteins solely in the nucleus, respectively (Fig. 5C). Together, we infer that hA3C and hA3C.WE-RK had a similar distribution in HeLa cells.

The RK-WE mutation in loop 1 moderately reduces the antiviral activity of hA3F

hA3C and hA3F-CTD display 77% sequence similarity, reflecting a common evolutionary origin [6]. Interestingly, the antiviral activity of hA3F is mediated by its CTD [73; 74]. Various loops within the A3F-CTD were recently investigated for their role in substrate binding and enzyme function [75] but it was not possible to unravel the antiviral activity of a protein consisting only of the A3F-CTD, mainly due to earlier reported difficulties in expressing this domain alone in human cells [68; 76]. The residues ^{25}RK^{26} in loop 1 of smmA3C-like protein are derived from exon 5 of the smm*A3F* gene, located in the CTD of A3F (Suppl. Fig. S1B) and are conserved in primate A3F proteins (see section evolution, below). To test the impact of RK residues in CTD loop 1 of the hA3F, we compared the antiviral activity of hA3F with A3F.RK-WE against HIV-1Δ*vif*. hA3F and hA3F.RK-WE yielded similar amounts of protein and were equally efficiently encapsidated in HIV-1 particles (Fig. 6A). However, the HIV-1Δ*vif* inhibiting effect of A3F.RK-WE was about 2-fold lower than WT A3F (Fig. 6B). Consequently, A3F.RK-WE showed decreased mutation

efficiency compared with WT A3F (Figs. 6C and 6D), which is consistent with data presented in a recent report [75]. Thus, we conclude that loop 1 with its residues RK in CTD of A3F is important for the enzymatic function of hA3F.

Inhibition of human LINE-1 retrotransposition by A3C variants

Since A3C and A3F restrict endogenous human LINE-1 (L1) retrotransposition activity by 40-75% and 66-85%, respectively [49; 59; 77; 78], we set out to elucidate how the WE and the RK residues in loop 1 of both hA3C and hA3F affect the L1 inhibiting activity. To this end, we quantified the L1-inhibiting effect of human WT A3A, A3C, and A3F proteins and their mutants hA3C.WE-RK, hA3C.WE-RK.S61P, and hA3F.RK-WE by a dual-luciferase retrotransposition reporter assay [79]. In this cell culture-based assay, the firefly luciferase gene is used as the reporter for L1 retrotransposition and the Renilla luciferase gene is encoded on the same plasmid for transfection normalization (Fig. 7A). Consistent with previous reports, overexpression of hA3A, hA3C, and hA3F inhibited L1 reporter retrotransposition by approximately 94%, 68%, and 56%, respectively (Fig. 7B). The mutant hA3C.WE-RK restricted L1 more strongly (from 56% to ~96%), but the introduction of the additional S61P mutation in hA3C.WE-RK.S61P did not further increase the ability of the enzyme to restrict L1 mobilization (Fig. 7B). Notably, hA3F and the mutant hA3F.RK-WE exhibited a comparable level of L1 restriction, indicating that regions other than loop 1 of A3F-CTD and, probably, the NTD (N-terminal domain) of hA3F are involved in L1 restriction (Fig. 7B). Immunoblot analysis of cell lysates of co-transfected HeLa-HA cells demonstrated comparable expression of the L1 reporter and HA-tagged A3- and A3 mutant proteins (Suppl. Fig. S4B). Furthermore, compared

to the inhibition of L1 retrotransposition by hA3C and chimpanzee A3C (~60%), hA3C.S61P inhibited L1 reporter retrotransposition by 75% (Suppl. Fig. S4C and S4D). These findings indicate that the WE-RK mutation in hA3C enhances its L1-inhibiting activity. Based on the observed antiviral activity and the L1-restricting effect of hA3C.WE-RK on L1, we hypothesize that the introduction of these positively charged residues in hA3C significantly fosters its interaction with nucleic acids, which was recently reported to mediate its L1 inhibiting activity [59].

The positively charged residues R25 and K26 in A3C form salt-bridges with the backbone of the ssDNA

To understand how the positively charged residues in loop 1 of A3C.WE-RK mediate the enhanced cytidine deamination activity, a structural model of hA3C variant hA3C-RKYG binding to ssDNA, based on the ssDNA-bound crystal structure of A3A was generated that shows a cytidine residue in the active center of hA3C.RKYG (Suppl. Fig. S5A). However, the ssDNA fragment (which was co-crystallized with hA3A) in this conformation is too short to interact with amino acids 25, 26, 28, and 29, which differ between hA3C WT and the hA3C.RKYG variant. Hence, this static binding mode model cannot explain why hA3C.RKYG has a higher cytidine deaminase activity than hA3C WT. To probe the impact of structural dynamics on residue-ssDNA interactions in order to explain the differences in A3C.WE-RK properties, this model was later subjected to molecular dynamics (MD) simulations.

To assess the binding to a longer ssDNA fragment, we generated a second complex model of ssDNA bound to the NTD of rhesus macaque A3G (rhA3G) [80], similar to the ssDNA-bound A3F-

CTD model built previously [81], and aligned the crystal structure of hA3C WT and the model of hA3C.RKYG to this complex (Figs. 8A, 8B, and 8C. Note that the A3G structure was used only for placing the DNA but not for modeling the protein part). This new model revealed that the positively charged residues R25 and K26 in hA3C.RKYG form salt-bridges with the backbone of the ssDNA (Fig. 8C) in contrast to hA3C WT (Fig. 8B). Thus, these two residues can form stronger interactions with ssDNA in hA3C.RKYG than their counterparts in hA3C, which may explain the enhanced cytidine deaminase activity of hA3C.WE-RK compared to hA3C (Fig. 4E). However, as the binding of ssDNA to NTDs, such as in the structure of rhA3G, differs from that in CTDs, we did not subject the former model to MD simulations.

We next performed five replicas of MD simulations of 2 µs length each for hA3C, hA3C.RKYG, and hA3C.S61P.S188I to assess the structural impact of the substitutions on the protein. For this, we used a hA3C crystal structure as starting structures and variants thereof generated by substituting respective residues. In all MD simulations, the cytidine remains bound to the Zn^{2+} ion in the active site. The root mean square fluctuations (RMSF), which describe atomic mobilities during the MD simulations, show distinct differences between the variants in the putative DNA-binding regions of the proteins: the RMSF of hA3C.RKYG and hA3C.S61P.S188I are up to 2 Å larger compared to hA3C WT in the regions carrying the substitutions (residues 21-32 for hA3C.RKYG and residues 55-67 for hA3C.S61P.S188I) (Suppl. Fig. S5B). This effect is specifically related to the respective substitutions, as no change in RMSF occurs for a variant in any region where it is identical to A3C WT. The increased movement of ssDNA-binding residues might improve the sliding of hA3C.RKYG and hA3C.S61P.S188I along the ssDNA, owing to more transient interactions with the ssDNA backbone. Conversely, the RMSF of loop

7 is up to 1 Å lower in both the hA3C.RKYG and hA3C.S61P.S188I variants compared to the hA3C WT (Suppl. Fig. S5B).

These results encouraged us to investigate possible interaction patterns between DNA and each of the three A3C variants that could be a result of the shift in loop 1 dynamics. For this purpose, we used the initial DNA-bound model of hA3C.RKYG with cytidine in the active center, modeled from the experimental A3A structure as described above, to generate DNA-bound complexes for hA3C WT and hA3C.S61P.S188I. While our MD simulations showed similar changes in the conformational dynamics of the loops as before (Suppl. Fig. S5B), we detected an interesting change in interactions between loop 1 residue R30 and the DNA. R30, which is present in all three variants and points away from the DNA in the A3C crystal structure, interacts more frequently with the DNA in both hA3C.S61P.S188I (16.4 ± 2.6% of the simulation time applying stringent criteria for H-bond formation (mean ± SEM for 10 trajectories)) and hA3C.RKYG (44.7 ± 2.7%) than in hA3C WT (0.1 ± 0.0%). In hA3C.RKYG, K26 similarly forms H-bonds with the DNA over 10.3 ± 2.8% of the MD trajectories, but, expectedly, E26 in hA3C WT and hA3C.S61P.S188I forms almost no H-bonds.

In addition, to rule out the possibility that the loop 7 residues might be influencing the loop 1 residues from binding DNA, we have analyzed the interaction between them. The average distance between the two loops in the absence of DNA is very similar for hA3C (12.1 ± 1.75 Å; SD, n=5000), hA3C.RKYG (12.7 ± 1.78 Å; SD, n=5000), and hA3C.S61P.S188I (12.12 ± 1.78 Å; SD, n=5000). Given the average distance of 12 Å it is not surprising that with the exception of N23 and A121, which are the only residues in spatial proximity and thus commonly interact, residues in loop 1 form H-bonds to those in loop 7 in less than 1% of the simulation time for

all variants. The average distance of any atom in residue 25 to residues in loop 7 is larger than 4.4 Å, suggesting that sustained interactions are unlikely.

Next, we used more lenient distance criteria suitable to evaluate the formation of interactions and evaluated, whether only the N-terminus (W25 in hA3C and hA3C.S61P.S188I and R25 in hA3C.RKYG) or only the C-terminus (R30 in all three variants) of loop 1, or both residues at the same time, interact with the DNA. In hA3C, only W25 interacts with the DNA in ~20% of the conformations (Suppl. Fig. S6A). In hA3C.S61P.S188I, interactions between W25 or R30 occur in ~20% of the conformations, thus showing an increase of a factor of 5 for R30 (Suppl. Fig. S6B). In hA3C.RKYG, both R25 and R30 simultaneously interact with DNA in ~29% of all investigated conformations besides the interactions of R30 with DNA alone in ~42% of the conformations (Suppl. Fig. S6C). Hence, these results suggest that W25 and R30 act additively in hA3C.S61P.S188I, whereas they act cooperatively in hA3C.RKYG. This correlates with the differences in activities, with hA3C.RKYG showing the highest activity against HIV-1Δ*vif*.

WE-RK mutation in the loop 1 of A3C enhances the interaction with ssDNA

To validate our structural modeling analysis and to address if the interaction of hA3C and hA3C.WE-RK with the substrate ssDNA was differentially affected, we performed electrophoretic mobility shift assays (EMSA) using hA3C-GST (A3C fused to glutathione S-transferase, GST) and hA3C.WE-RK-GST purified from HEK293T cells (Fig. 9A). We first confirmed that the purified GST fusion proteins are catalytically active (Fig. 9B). As expected hA3C.WE-RK-GST displayed a stronger enzymatic activity than the WT equivalent and no activity with GST was detected (Fig. 9B). For EMSA, as a probe, we used a biotin-labeled ssDNA

oligonucleotide that harbors a TTCA motif in its central region [68; 82]. Because hA3C-GST is known to form a stable DNA-protein complex when the protein concentration reaches ≥ 20 nM [68], we decreased the amount of A3C and its mutant protein to specifically test their inherent DNA binding capacity. In a titration experiment with concentrations ranging from 2 to 8 nM in steps of 2 nM of hA3C-GST and hA3C.WE-RK-GST purified protein, we detected a clear trend in the formation of DNA–protein complexes for hA3C-GST and hA3C.WE-RK-GST (Fig. 9C). Intriguingly, DNA-protein complexes of hA3C.WE-RK-GST started appearing at the lowest protein concentration used (2 nM), while hA3C-GST-DNA complexes were detected at protein concentrations ≥ 6 nM. To confirm the specificity of the DNA–protein complexes, we competed for the reaction with unlabeled DNA carrying the same nucleotide sequence as the used probe in 200-fold excess relative to that probe. The addition of the competitor DNA to the sample containing the maximum (8 nM) amount of A3C protein, efficiently disrupted the protein-DNA complex formation (Fig. 9C and Suppl. Fig. S7). Together, data from structural modeling and EMSA experiments allowed us to conclude that the two amino acid-change in loop 1 of A3C boosts the ssDNA binding capacity of A3C. Importantly, the GST moiety did not affect the binding (Suppl. Fig. S7 and [68]).

Evolution of A3Z2 loop 1 regions in primates

Because of the strong evolutionary relationship between A3C, the CTD of A3F, and related A3Z2 proteins [6], we performed a phylogenetic reconstruction for the A3Z2 domains in primates, using the A3Z2 sequences in the northern tree shrew as outgroup. Our analyses were performed at the A3Z2 domain level, separating the two Z2 domains of the double-

domain A3D and A3F proteins, thus generating five evolutionary units: the A3D-NTD, A3F-NTD, A3C, A3D-CTD and A3F-CTD (Suppl. Fig. S8). Remarkably, the results show that the A3Z2 domains underwent independent duplication in the two sister taxa, tree shrews and primates, as the three A3Z2 tree shrew sequences constitute a clear outgroup to all primate A3Z2 sequences. We identified a sharp clustering of the A3D-NTD and A3F-NTD on the one hand and of A3C, A3D-CTD, and A3F-CTD on the other hand. As to New World monkeys (Platyrrhini), we could only confidently retrieve A3C sequences from the white-faced sapajou *Cebus capucinus* and from the Ma's night monkey *Aotus nancymaae.* These sequences from A3C New World monkeys were basal to all Catarrhini (Old World monkeys and apes) A3C, A3D-CTD and A3F-CTD sequences, suggesting that the two gene duplications leading to the extant organization of A3C, A3D, and A3F occurred after the Platyrrhini/Catarrhini split 43.2 Mya (41.0 - 45.7 Mya) and before the Cercopithecoidea/Hominoidea (Old World monkeys/apes) split 29.44 Mya (27.95 - 31.35 Mya). The results show a tangled distribution within the A3D-NTD and A3F-NTD clade, and within the A3D-CTD and A3F-CTD clade. These confusing relationships are more obvious when comparing the phylogenetic reconstruction of the Z2 domains without imposing any topological constraint (Suppl. Fig. S8) with a tree in which monophyly of each of the large six clades identified was enforced (Suppl. Fig. S8). The tanglegram linking both, highlights those sequences whose phylogenetic position does not match the expected cluster, after the current annotation. Conversely, Catarrhini A3C sequences form a well-supported monophyletic taxon, and this A3C gene tree essentially adheres to the corresponding species tree (Fig. 10). Focusing exclusively on the nodes that we could identify with confidence, we performed ancestral phylogenetic inference of the most likely amino acid sequence for the A3 loop 1 (Suppl. Fig. S9) and, in parallel, performed a

consensus analysis of the extant sequences (Fig. 10 and Suppl. Fig. S9). Our results recover the well-conserved aromatic stacking stretch F[FY]FXF characteristic of all A3s. In the A3C, A3D-CTD, and A3F-CTD clade, we identified a small motif displaying striking divergent evolution flanked by conserved small hydrophobic amino acids. The most likely ancestral form is the amino acid motif LRKA, which is also the form present in extant New World monkeys A3C and the most common in extant A3F-CTD, while in the extant A3D-CTD the Arg residue is less conserved L[RLQ][KT]A (Fig. 10). Strikingly, in the ancestor of Catarrhini A3C at around 29.4 Mya (27.6-31.3 Mya), this motif had already evolved to LWEA (Suppl. Fig. S9), and this is the common extant form in Old World monkeys and apes (Fig. 10). Only subsequently, and exclusively in the Chlorocebus lineage (African green monkeys), this change was partly reverted to LREA by a TGG>CGG transition. This reversion should have occurred after the divergence within Cercopithecinae, around 13.7 Mya (10.7 - 16.6 Mya) and before the speciation within Chlorocebus at 3.42 Ma (2.05-4.15 Mya) (Fig. 10).

DISCUSSION

Compared to the many studies conducted over the past decade on the HIV-1 restriction factors A3G and A3F, investigations on A3C are very limited. A small number of studies have addressed the catalytic activity and substrate binding capacity of A3C [59; 68; 72; 83]. While the previously characterized hA3C mutants S61P and S188I boost the catalytic activity of the enzyme to a certain degree, none of these mutations is powerful enough to reduce the HIV-1Δ*vif* infectivity to the level accomplished by A3G and they do not directly partake in catalytic activity [68; 72; 83]. Because our repeated attempts to express A3F-CTD in human cells were not

successful ([68] and Fig. 1C), we assayed A3C proteins from different Old World primate species. Due to the high level of nucleotide sequence identity between the A3C (A3Z2) paralogs (see discussion below) in the sooty mangabey monkey genome, we generated by missannotation a smmA3C-like protein with superior anti-HIV-1 and enzymatic activity. We have identified the key role of two positively-charged residues in loop 1 of this smmA3C-like protein (and of the hA3F-CTD), namely R25 and K26 in the RKYG motif. Replacing RKYG of smmA3C-like by the WEND form of this motif in hA3C abolished its anti-HIV-1 and catalytic activity. Importantly, the converse strategy of introducing the substitution WE-RK in the loop 1 of hA3C generated the potent, deaminase-dependent anti-HIV-1 enzyme hA3C.WE-RK. Consistent with these observations, our EMSA data demonstrate that residues in the loop 1 of A3C regulate protein-DNA interaction. Thus, we postulate that this more intense DNA-protein interaction is causative for the enhanced deamination activity and enhanced anti-HIV and anti-L1 activity. Similarly, Solomon and coworkers discussed that loop 1 residues of hA3G-CTD strongly interact with substrate ssDNA and that this interaction distinguishes catalytic binding from non-catalytic binding [84]. However, the loop 1 of A3 proteins likely has multiple functions, as loop 1 of A3A was found to be important for substrate specificity but not for substrate binding affinity [85], and loop 1 of A3H, especially its residue R26, plays a triple role for RNA binding, DNA substrate recognition, and catalytic activity likely by positioning the DNA substrate in the active site for effective catalysis [86]. In accordance, our study indicates that ^{25}RK^{26} substitution in loop 1 of A3C provides the microenvironment that drives the flexibility in substrate binding and enzymatic activity.

The binding model developed here rationalizes how hA3C.RKYG can interact with the negatively charged backbone of ssDNA via the positively charged loop 1 side chains of R25 and

K26 (Fig. 8C). Like our modeling strategy, Fang *et al*. [81] used their binding mode model of A3F-CTD with ssDNA to identify residues in the A3G-CTD important for ssDNA binding. Furthermore, the increased mobility of DNA binding regions carrying the substitutions in hA3C.RKYG and hA3C.S61P.S188I, respectively, compared to hA3C (Suppl. Fig. S5B) suggests that hA3C.RKYG and hA3C.S61P.S188I can better slide along the ssDNA than hA3C. The higher mobility of the residues may allow them to adapt more quickly to the passing ssDNA, which, together with likely stronger interactions with the ssDNA backbone, may explain the increased deaminase activity. This idea is corroborated by the MD simulations, in which the complexes including DNA loop 1 residues show more frequent interactions with the DNA in the case of hA3C.RKYG than in any of the other two variants, suggesting a stronger binding of the DNA; by contrast, in hA3C.S61P.S188I in 39.2% of the time either W25 or R30 interact with the DNA such that the DNA could be passed on from one residue to the other, assisting in the sliding-down mechanism while possibly also increasing binding affinity. In addition, loop 7 exhibits a decreased mobility in both hA3C.RKYG and hA3C.S61P.S188I compared to hA3C (Suppl. Fig. S5B). Decreased mobility of loop 7 has been shown to predict higher deaminase activity, DNA binding, and substrate specificity of A3G and A3F, and has been reported to be also relevant for antiviral activity of A3B and A3D [74; 87; 88; 89]. These structural findings can explain the differences in deaminase activity among the three variants.

Unexpectedly, our experiments also demonstrated that LINE-1 restriction by A3C, which was reported earlier to be deaminase-independent [59], is enhanced after expression of the A3C.WE-RK variant. These data suggest that the reported RNA-dependent physical interaction between L1 ORF1p and A3C dimers might be mediated by A3C loop 1, is partly dependent on the two amino acids W25 and E26 and is enhanced by the R25 and K26 substitutions. However, L1

inhibition by A3F was not significantly altered by the A3F.RK-WE mutations, clearly indicating that other regions (and NTD) in A3F are likely to be relevant for L1 restriction.

Because selection likely had to balance between anti-viral/anti-L1 activity and genotoxicity of A3 proteins, we wanted to characterize loop 1 residues during the evolution of the closely related A3Z2 proteins A3C, A3D CTD and A3F CTD in primates, all of them descendant of an ancestral Z2 domain that had undergone two duplication rounds [6]. In the most recent common ancestor of these enzymes that existed before the split Old World and New World primates (Catarrhini-Platyrrhini) around 43 Mya, we infer the ancestral form of the sequence of this motif in loop 1 to be LRKAYG. In New World monkeys, the A3C genes were not duplicated and are basal to the three sister clades of Catarrhini A3C, A3D-CTD, and A3F-CTD. In extant A3C sequences in New World monkeys, the loop 1 motif has notably remained unchanged and reads LRKAYG. In Catarrhini, on the contrary, the ancestral A3C sequence underwent two rapid rounds of duplication that occurred after the split with the ancestor of Platyrrhini, and before the split between the ancestors of Old World monkeys (Cercopithecoidea) and apes (Hominoidea), some 29 Mya [6]. A3F has since then been involved in an Red Queen arms race with retroviral genes [90]. In extant A3F-CTD sequences, the consensus form of the loop 1 remains LRKAYG, albeit with a certain variability of the R residue, which is exchanged with other positively charged amino acids. In extant A3D-CTD enzymes, this motif has undergone erosion, is more variable and reads L[RLQ][KT]A[YC]G. Interestingly, loop 1 in A3C experienced rapid and swift selective pressure to exchange the positively charged RK amino acids by the largely divergent chemistry of WE, yielding LWEAYG. This selective sweep occurred very rapidly, as this is the fixed form in all Catarrhini. Notoriously, and exclusively in the Chlorocebus lineage (African Green monkeys), this amino acid substitution was partly reverted

to LREAYG, which is the conserved sequence in the four Chlorocebus A3C entries available (Fig. 10).

Overall, our results suggest that the two duplication events that generated the extant A3C, A3D-CTD, and A3F-CTD sequences in Catarrhines released the selective pressure on two of the daughter enzymes allowing them to explore the sequence space and to evolve via sub/neofunctionalization, as proposed for Ohno's in-paralogs [91]. Thus, the A3F-CTD form of the loop 1 diverged little from the ancestral chemistry and possibly maintained the ancestral function, while the release in conservation pressure on A3D-CTD allowed the enzyme loop 1 to accumulate mutations and diverge from the ancestral state. In turn, A3C was rapidly engaged into a distinct evolutionary pathway, which is unique due to the highly divergent chemistry of loop 1 but also because A3C is the only A3Z2 monodomain enzyme of the A3 family. It must also be noted that among the descendants of the ancestral A3C in Catarrhines, only extant A3C forms a well-supported monophyletic clade (Suppl. Figs. S8 and S9). Instead, in several instances and for different species, sequences annotated in the databases as A3D-CTD clustered together with sequences annotated as A3F-CTD, and vice versa, and the same is true for the corresponding N-terminal domains (see tanglegram Suppl. Fig. S8), overall resulting in a lack of support for common ancestry for the individual moieties of A3C and A3F, and preventing us from inferring the ancestral forms of the loop 1 in A3D-CTD and A3F-CTD. This lack of monophyly could simply reflect the lack of power of phylogenetic reconstruction or the potential for database misannotations when applied to genes undergoing complex evolution, including a full panel of duplications, deletions, adaptive radiation, differential selection among paralogs and Red Queen dynamics [3; 6; 90; 92; 93]. In this respect, the field is wanting for a systematisation of protocols and procedures for identifying selection signatures

in genes with complex evolutionary histories [94]. This lack of resolution could also reflect a biological basis of read-through of unmatured mRNAs resulting in differentially edited or in naturally chimeric mRNAs [95; 96; 97], which can hamper phylogenetic inference. Finally, the genetic architecture of the A3 locus, with the different gene copies located in tandem may favour non-homologous recombination between recently diverged, closely related sequences, and may also facilitate gene conversion between non-homologous alleles, overall leading to genetic information flow between gene copies and decoupling the true evolutionary history from our gene name and annotation-based phylogenetic reconstructions. The combined result of these novelty-generating mechanisms could be an enhanced inter-species or even inter-individual diversity in the A3 locus at either the genetic or the transcriptomic levels [97; 98]. The functional impact of such gene and mRNA diversity deserves further investigation, especially in the context of personalised medicine.

In conclusion, we postulate that the loop 1 region of A3s might have a conserved role in anchoring its ssDNA substrate for efficient catalysis and that weak deamination and anti-HIV-1 activity of hA3C might have been the result of losing DNA interactions in loop 1 during its evolution. It is thus possible that genes encoding A3C proteins with loop 1 residues with a higher ssDNA affinity were too genotoxic to benefit their hosts by superior anti-viral and anti-L1 activity. Tao *et al*. [99] noted that the level of A3C preferentially increased upon treatment with artesunate (Art) and suggested that upregulated A3C is involved in the Art-induced DNA damage response[99]. Conceptually, we cannot rule out the possibility that the residues characterized here in loop 1 of hA3C might have an impact on recognition of unknown substrates or targets.

MATERIALS AND METHODS

Cell culture. HEK293T cells (ATCC CRL-3216) were maintained in Dulbecco's high-glucose modified Eagle's medium (DMEM) (Biochrom, Berlin, Germany), supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 50 units/ml penicillin, and 50 µg/ml streptomycin at 37°C in a humidified atmosphere of 5% CO_2. Similarly, HeLa-HA cells [100] were cultured in DMEM with 10% FCS (Biowest, Nuaillé, France), 2mM L-glutamine and 20 U/ml penicillin/streptomycin (Gibco, Schwerte, Germany).

Plasmids. The HIV-1 packaging plasmid pMDLg/pRRE encodes *gag-pol*, and the pRSV-Rev for the HIV-1 *rev* [101]. The HIV-1 vector pSIN.PPT.CMV.Luc.IRES.GFP expresses the firefly luciferase and GFP [102]. HIV-1 based viral vectors were pseudotyped using the pMD.G plasmid that encodes the glycoprotein of VSV (VSV-G). SIVagm luciferase vector system was described before [31]. All A3 constructs described here were cloned in pcDNA3.1 (+) with a C-terminal hemagglutinin (HA) tag. The smmA3C-like expression plasmid was generated by exon assembly from the genomic DNA of a white-crowned mangabey (*Cercocebus torquatus lunulatus*), and the cloning strategy for smmA3C-like and the chimeras of hA3C/smmA3C-like plasmid construction was recently described [67]. The expression vector for A3G-HA was generously provided by Nathaniel R. Landau. Expression constructs hA3C, rhA3C, cpzA3C, agmA3C and A3C point mutant A3C.C97S were described before [55; 58; 68].

Various point mutants hA3C.WE-RK, hA3C.ND-YG, hA3C.WE-RK.C97S, hA3C.WE-RK.S61P, hA3C.WE-RK.S61P.C97S, hA3C.WE-RK.S61P.S188I, hA3C.WE-RK.S61P.S188I.C97S, hA3F.RK-

WE, smmA3C-like.E68A were generated by using site-directed mutagenesis. Similarly, single or multiple amino acid changes were made in expression vectors to produce chimera 2 mutants (C2.DH-YG, C2.RKYG-WEND, C2.R25W, C2.K26E, C2.Y28N, and C2.G29D) and smmA3C-like.RKYG-WEND. To clone C-terminal GST-tagged hA3C, hA3C.WE-RK, the ORFs were inserted between the restriction sites HindIII and XbaI in the mammalian expression construct pK-GST mammalian expression vector [103]. Individual exons of authentic smmA3C and smmA3F and smmA3F-like genes exons were amplified and cloned in pcDNA3.1. All the primer sequences are listed in Suppl. Table 1.

Virus production and isolation. HEK293T cells were transiently transfected using Lipofectamine LTX and Plus reagent (Invitrogen, Karlsruhe, Germany) with an appropriate combination of HIV-1 viral vectors (600 ng pMDLg/pRRE, 600 ng pSIN.PPT.CMV.Luc.IRES.GFP, 250 ng pRSV-Rev, 150 ng pMD.G with 600 ng A3 plasmid or replaced by pcDNA3.1, unless otherwise mentioned) or SIVagm vectors (1400 ng pSIVTan-LucΔ*vif*, 150 ng pMD.G with 600 ng A3 plasmid) in 6 well plate. 48 h post-transfection, virion containing supernatants were collected and for isolation of virions, concentrated by layering on 20% sucrose cushion and centrifuged for 4 h at 14,800 rpm. Viral particles were re-suspended in mild lysis buffer (50 mM Tris (pH 8), 1 mM PMSF, 10% glycerol, 0.8% NP-40, 150 mM NaCl and 1X complete protease inhibitor).

Luciferase-based infectivity assay. HIV-1 luciferase reporter viruses were used to transduce HEK293T cells. Prior infection, the amount of reverse transcriptase (RT) in the viral particles

was determined by RT assay using Cavidi HS kit Lenti RT (Cavidi Tech, Uppsala, Sweden). Normalized RT amount equivalent viral supernatants were transduced. 48 h later, luciferase activity was measured using SteadyliteHTS luciferase reagent substrate (Perkin Elmer, Rodgau, Germany) on a Berthold MicroLumat Plus luminometer (Berthold Detection Systems, Pforzheim, Germany). Transductions were done in triplicates and at least three independent experiments were performed.

Immunofluorescence microscopy. $1x10^5$ HeLa cells grown on polyethylene coverslips (Thermo Fisher Scientific) were co-transfected with plasmids for hemagglutinin (HA) tagged hA3C (0.25 µg) WT or hA3C.WE-RK (0.25 µg) using FuGENE transfection reagent (Promega, Wisconsin, USA). At day 2 post transfection, cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 mins, permeabilized 0.1% Triton X-100 for 10 min, incubated with blocking solution (10% FBS in PBS) for 1 h, and then cells were stained with mouse anti-HA antibody (Covance, Münster, Germany) 1:1,000 dilution in blocking solution for 1 h. Donkey anti-mouse Alexa Fluor 488 (Covance) was used as a secondary antibody, 1:300 dilution in blocking solution for 1 h. Finally, DAPI was used to stain nuclei for 2 mins. The images were captured by using a 63x objective on Zeiss LSM 510 Meta laser scanning confocal microscopy (Carl Zeiss, Cologne, Germany). For the quantification of cellular localization of A3Cs, 40 randomly chosen transfected cells with A3C or A3C.WE-RK were categorized and quantified.

Immunoblot analyses. Transfected HEK293T cells were washed with PBS and lysed in radioimmunoprecipitation assay buffer (RIPA, 25 mM Tris (pH 8.0), 137 mM NaCl, 1% glycerol, 0.1% SDS, 0.5% sodium deoxycholate, 1% Nonidet P-40, 2 mM EDTA, and protease inhibitor cocktail set III [Calbiochem, Darmstadt, Germany]) 20 min on ice. Lysates were clarified by centrifugation (20 min, 14800 rpm, 4°C). Samples (cell/viral lysate) were boiled at 95^{o}C for 5 min with Roti load reducing loading buffer (Carl Roth, Karlsruhe, Germany) and subjected to SDS-PAGE followed by transfer (Semi-Dry Transfer Cell, Biorad, Munich, Germany) to a PVDF membrane (Merck Millipore, Schwalbach, Germany). Membranes were blocked with skimmed milk solution and probed with appropriate primary antibody, mouse anti-hemagglutinin (anti-HA) antibody (1:7,500 dilution, MMS-101P, Covance); goat anti-GAPDH (C-terminus, 1:15,000 dilution, Everest Biotech, Oxfordshire, UK); mouse anti-α-tubulin antibody (1:4,000 dilution, clone B5-1-2; Sigma-Aldrich, Taufkirchen, Germany), mouse anti-capsid p24/p27 MAb AG3.0 [104] (1:250 dilution, NIH AIDS Reagents); rabbit anti S6 ribosomal protein (5G10; 1:10^3 dilution in 5% BSA, Cell Signaling Technology, Leiden, The Netherlands). Secondary Abs.: anti-mouse (NA931V), anti-rabbit (NA934V) horseradish peroxidase (1:10^4 dilution, GE Healthcare) and anti-goat IgG-HRP (1:10^4 dilution, sc-2768, Santa Cruz Biotechnology, Heidelberg, Germany). Signals were visualized using ECL chemiluminescent reagent (GE Healthcare). To characterize the effect of the expression of A3 proteins and their mutants on LINE-1 (L1) reporter expression, HeLa-HA cells were lysed 48 h post-transfection using triple lysis buffer (20 mM Tris/HCl, pH 7.5; 150 mM NaCl; 10 mM EDTA; 0.1% SDS; 1% Triton X-100; 1% deoxycholate; 1x complete protease inhibitor cocktail [Roche]), clarified and 20 µg total protein were used for SDS-PAGE followed by electroblotting. HA-tagged A3 proteins and L1 ORF1p were detected using an anti-HA antibody (MMS-101P; Covance) in a 1:5,000 dilution and the polyclonal

rabbit-anti-L1 ORF1p antibody #984 [105] in a 1:2,000 dilution, respectively, in 1xPBS-T containing 5% milk powder. ß-actin expression (AC-74, 1:30,000 dilution, Sigma-Aldrich Chemie GmbH) served as a loading control.

Differential DNA denaturation (3D) PCR. HEK293T cells were cultured in 6-well plates and infected with DNAse I (Thermo Fisher Scientific) treated viruses for 12 h. Cells were harvested and washed in PBS, the total DNA was isolated using DNeasy DNA isolation kit (Qiagen, Hilden, Germany). A 714-bp fragment of the luciferase gene was amplified using the primers 5'-GATATGTGGATTTCGAGTCGTC-3' and 5'-GTCATCGTCTTTCCGTGCTC-3'. For selective amplification of the hypermutated products, the PCR denaturation temperature was lowered stepwise from 87.6°C to 83.5°C (83.5°C, 84.2°C, 85.2°C, 86.3°C, 87.6°C) using a gradient thermocycler. The PCR parameters were as follows: (i) 95°C for 5 min; (ii) 40 cycles, with 1 cycle consisting of 83.5°C to 87.6°C for 30 s, 55°C for 30 s, 72°C for 1 min; (iii) 10 min at 72°C. PCRs were performed with Dream Taq DNA polymerase (Thermo Fisher Scientific). PCR products were stained with ethidium bromide. PCR product (smmA3C-like sample only) from the lowest denaturation temperature was cloned using CloneJET PCR Cloning Kit (Thermo Fisher Scientific) and sequenced. smmA3C-like protein-induced hypermutations of eleven independent clones were analysed with the Hypermut online tool (https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html) [106]. Mutated sequences (clones) carrying similar base changes were omitted and only the unique clones were presented for clarity.

***In vitro* DNA cytidine deamination assay.** A3 proteins expressed in transfected HEK293T cells, virion-incorporated A3s, or purified GST fusion proteins were used as input. Cell lysates were prepared with mild lysis buffer 48 h post plasmid transfection. Deamination reactions were performed as described [70; 107] in a 10 µL reaction volume containing 25 mM Tris pH 7.0, 2 µl of cell lysate and 100 fmol single-stranded DNA substrate (TTCA: 5'-GGATTGGTTGGTTATTTGTATAAGGAAGGTGGATTGAAGGTTCAAGAAGGTGATGGAAGTTATGTTTGGTAGATTGATGG). Samples were treated with 50 µg/ml RNAse A (Thermo Fisher Scientific). Reactions were incubated for 1 h at 37˚C and the reaction was terminated by boiling at 95˚C for 5 min. One fmol of the reaction mixture was used for PCR amplification Dream Taq polymerase (Thermo Fisher Scientific) 95˚C for 3 min, followed by 30 cycles of 61˚C for 30 s and 94˚C for 30 s using primers forward 5'-GGATTGGTTGGTTATTTGTATAAGGA and reverse 5'-CCATCAATCTACCAAACATAACTTCCA. PCR products were digested with MseI (NEB, Frankfurt/Main, Germany), and resolved on 15% PAGE, stained with ethidium bromide (7.5 µg/ml). As a positive control, substrate oligonucleotides with TTUA instead of TTCA were used to control the restriction enzyme digestion [68].

L1 retrotransposition reporter assay. Relative L1 retrotransposition activity was determined by applying a rapid dual-luciferase reporter based assay described previously [79]. Briefly, $2x10^5$ HeLa-HA cells were seeded per well of a six-well plate and transfected using Fugene-HD transfection reagent (Promega) according to the manufacturer's protocol. Each well was cotransfected with 0.5 µg of the L1 retrotransposition reporter plasmid pYX017 or pYX015 [79] and 0.5 µg of pcDNA3.1 or WT or mutant A3 expression construct resuspended in 3 µl Fugene-

HD transfection reagent and 100 μl GlutaMAX-I-supplemented Opti-MEM I reduced-serum medium (Thermo Fisher Scientific). Three days after transfection cultivation, the medium was replaced by complete DMEM containing 2.5 μg/ml puromycin, to select for the presence of the L1 reporter plasmid harboring a puroR-expression cassette. Next day, the medium was replaced once more by puromycin containing DMEM medium and 48 hours later, transfected cells were lysed to quantify dual-luciferase luminescence. Dual-luciferase luminescence measurement: Luminescence was measured using the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's instructions. For assays in 6-well plates, 200 μl Passive Lysis Buffer was used to lyse cells in each well; for all assays, 20 μl lysate was transferred to a solid white 96-well plate, mixed with 50 μl Luciferase Assay Reagent II and firefly luciferase (Fluc) activity was quantified using the microplate luminometer Infinite 200PRO (Tecan, Männedorf, Switzerland). Renilla luciferase (Rluc) activity was subsequently read after mixing 50 μl Stop & Glo Reagent into the cell lysate containing Luciferase Assay Reagent II. Data were normalized as described in the results section. The retrotransposition-defective L1RP/JM111 (located on pYX015) was used as the reference Fluc vector and the normalized luminescence ratio (NLR) resulting from cotransfection of pYX015 and pcDNA3.1(+) was set as 1.

Protein sequence alignment and visualization. Sequence alignment of hA3C and smmA3C-like protein was done by using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The alignment file was then submitted to ESPript 3.0 [108] (espript.ibcp.fr) to calculate the similarity and identity of residues between both proteins and to represent the pairwise

sequence alignment. Cartoon model of the crystal structure of A3C (PDB 3VOW) was constructed using PyMOL (PyMOL Molecular Graphics System version 1.5.0.4; Schrödinger, Portland, OR).

Structural model building of protein-DNA complexes. The structural models of hA3C or hA3C.RKYG binding to ssDNA were generated by first aligning the X-ray crystal structure of rhA3G-NTD (PDB ID 5K82 [80]) onto the X-ray crystal structure of hA3F-CTD (PDB ID 5W2M [81]), the latter of which was co-crystallized with ssDNA. Subsequently, the hA3C X-ray crystal structure (PDB ID 3VOW [66]) was aligned onto the NTD of rhA3G, which is structurally similar to hA3C. The ssDNA and the interface region of hA3C were subsequently relaxed in the presence of each other using Maestro [109]. The same program was used to mutate hA3C to obtain the hA3C.RKYG and hA3C.S61P.S188I variants, which were again relaxed in the presence of the ssDNA. Similarly, we obtained hA3C, hA3C.RKYG, and hA3C.S61P.S188I ssDNA binding models based on the ssDNA-binding X-ray crystal structure of hA3A (PDB ID 5SWW [110]). These three DNA complex structures were later used for MD simulations as they include a cytidine residue in the active center.

Molecular dynamics simulations. hA3C, hA3C.RKYG, and hA3C.S61P.S188I were subjected to MD simulations. For this, the above-mentioned structures without the DNA were N- and C-terminally capped with ACE and NME, respectively. The three variants were protonated with PROPKA [111] according to pH 7.4, neutralized by adding counter ions, and solvated in an octahedral box of TIP3P water [112] with a minimal water shell of 12 Å around the solute. The

Amber package of molecular simulation software [113] and the ff14SB force field [114] was used to perform the MD simulations. For the Zn^{2+}-ions the Li-Merz parameters for two-fold positively charged metal ions [115] were used. To cope with long-range interactions, the "Particle Mesh Ewald" method [116] was used; the SHAKE algorithm [117] was applied to bonds involving hydrogen atoms. As hydrogen mass repartitioning [118] was utilized, the time step for all MD simulations was 4 fs with a direct-space, non-bonded cut-off of 8 Å.

In the beginning, 17500 steps of steepest descent and conjugate gradient minimization were performed; during 2500, 10000, and 5000 steps positional harmonic restraints with force constants of 25 kcal mol^{-1} $Å^{-2}$, 5 kcal mol^{-1} $Å^{-2}$, and zero, respectively, were applied to the solute atoms. Thereafter, 50 ps of NVT (constant number of particles, volume, and temperature) MD simulations were conducted to heat up the system to 100 K, followed by 300 ps of NPT (constant number of particles, pressure, and temperature) MD simulations to adjust the density of the simulation box to a pressure of 1 atm and to heat the system to 300 K. During these steps, a harmonic potential with a force constant of 10 kcal mol^{-1} $Å^{-2}$ was applied to the solute atoms. As the final step in thermalization, 300 ps of NVT-MD simulations were performed while gradually reducing the restraint forces on the solute atoms to zero within the first 100 ps of this step. Afterwards, five independent production runs of NVT-MD simulations with 2 μs length each were performed. For this, the starting temperatures of the MD simulations at the beginning of the thermalization were varied by a fraction of a Kelvin. MD simulation of those three variants in complex with ssDNA were performed similarly, treating the DNA with the OL15 force field [119] and performing ten independent production runs of NVT-MD simulations with 2 μs length each. To evaluate the interactions between loop

1 (residues 25-30) of the three variants and the ssDNA present in the complexes, we employed two different measures using CPPTRAJ [120]. First, we used the h-bond command to detect hydrogen bonds between residues in loop 1 and the ssDNA. Second, we measured the minimal distance of the side chain atoms, not including C_β of the respective residues, and the DNA for each snapshot of the MD simulations and correlated both (Suppl. Fig. S6), considering a larger distance cut-off of 4 Å to detect interactions between the side chains and DNA. The minimal distance over time for residue 30 is shown in Suppl. Fig. S10A-C and the root mean square deviation (RMSD) over time in Suppl. Fig. S10D-F. The latter figure indicates that the systems structurally stabilized after ~250 ns.

Expression and purification of recombinant GST-tagged hA3C and hA3C.WE-RK from HEK293T cells. Recombinant C-terminal GST-tagged hA3C and hA3C.WE-RK were expressed in HEK293T cells and purified by affinity chromatography using Glutathione Sepharose 4B beads (GE Healthcare) as described previously [68]. Cells were lysed 48 h later with mild lysis buffer [50 mM Tris (pH 8), 1 mM PMSF, 10% glycerol, 0.8% NP-40, 150 mM NaCl, and 1X complete protease inhibitor and incubated with GST beads. After 2 h incubation at 4°C in end-over-end rotation, GST beads were washed twice with wash buffer containing 50 mM Tris (pH 8.0), 5 mM 2-ME, 10% glycerol and 500 mM NaCl. The bound GST hA3C and hA3C.WE-RK proteins were eluted with wash buffer containing 20 mM reduced glutathione. The proteins were 90-95% pure as checked on 15% SDS-PAGE followed by Coomassie blue staining. Protein concentrations were estimated by Bradford's method.

Electrophoretic mobility shift assay (EMSA) with hA3C-GST and hA3C.WE-RK-GST. EMSA was performed as described previously [68; 82; 121]. We mixed 20 fmol of 3' biotinylated DNA (30-TTC-Bio-TEG purchased from Eurofins Genomics, Ebersberg Germany) with 10 mM Tris (pH − 7.5), 100 mM KCl, 10 mM MgCl2, 1 mM DTT, 2% glycerol, and the respective amount of recombinant proteins in a 15 µl reaction mixture, and incubated at room temperature for 30 min. The reaction mixture containing the protein–DNA complex were resolved on a 5% native PAGE gel on ice and transferred to a nylon membrane (Amersham Hybond-XL, GE healthcare) using 0.5 X TBE. After the transfer, the membrane containing protein–DNA complex were cross-linked by UV radiation with 312-nm bulb for 15 min. Chemiluminescent detection of biotinylated DNA was carried out according to the manufacturer's instruction (Thermo Scientific, LightShift Chemiluminescence EMSA Kit).

Phylogenetic inference. In order to study the evolution of the A3-Z2 domains a representative set of 61 primate A3C, A3D, and A3F gene sequences were collected from GenBank (https://www.ncbi.nlm.nih.gov/genbank), as follows: 26 A3C sequences, 12 A3D sequences, and 21 A3F sequences (full list available in Suppl. Table 2). The phylogenetic relationships and divergence times among the species used were retrieved from http://www.timetree.org (Suppl. Fig. S8). A3 sequences from the northern tree shrew *Tupaia belangeri* were included as an outgroup to the primate ones. As A3D and A3F sequences contain each two Z2 domains, they were split into the corresponding N- and C-termini. The alignments were performed at the amino acid level using MAFFTv7.380 (http://mafft.cbrc.jp/alignment/software/) [122]. Phylogenetic inference was performed using RAxMLv8 [123], at either the nucleotide level under

the GTR+Γ model or at the amino acid level under the LG+Γ model. Node support was evaluated applying 5,000 bootstrap cycles. Additionally, phylogenies at the nucleotide level were also calculated after introducing constraints in the tree, forcing monophyly of each clade A3D_N and C-termini, A3F_N and C-termini, New World monkeys A3C, and catarrhine A3C. Differences in maximum likelihood between alternative topologies for the same alignment were evaluated by the Shimodaira-Hasegawa test. Ancestral state reconstruction of amino acids in the loop A3-Z2 loop 1 was performed only for the supported clades using RAxMLv8. A tanglegram with the two phylogenies was drawn with Dendroscope v3.6.3 [124]. Final layouts were done with Inkscape 0.92.4.

Statistical analysis. Data were represented as the mean with SD in all bar diagrams. Statistically significant differences between two groups were analyzed using the unpaired Student's t-test with GraphPad Prism version 5 (GraphPad Software, San Diego, CA, USA). A minimum *p*-value of 0.05 was considered as statistically significant.

ACKNOWLEDGMENTS

We thank Wioletta Hörschken for excellent technical assistance, Boris Görg for microscopy support, and Wolfgang A. Schulz for kindly proofreading the manuscript. We thank Alejandro Moisés Barbero Amezaga (University Francisco de Vitoria / Faculty of Experimental Sciences, Madrid, Spain) for his excellent technical support, and the University Francisco de Vitoria /Faculty of Experimental Sciences for financial support of A.M.B.A. We thank Michael

Emerman, Jens-Ove Heckel, Henning Hofmann, Yasumasa Iwatani, Nathanial R. Landau, Neeltje Kootstra, Bryan Cullen, Jonathan Stoye, Harald Wodrich, and Jörg Zielonka for reagents. The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: a monoclonal antibody to HIV-1 p24 (AG3.0) from Jonathan Allan. FB and IGB acknowledge the IRD itrop HPC (South Green Platform) at IRD Montpellier for providing computing resources. HG is grateful for computational support and infrastructure provided by the "Zentrum für Informations- und Medientechnologie" (ZIM) at the Heinrich-Heine-University Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) to HG on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) (user ID: HKF7). Graphical abstract was designed with BioRender.com.

FUNDING

This work was supported by a grant from the research commission of the medical faculty of the Heinrich-Heine-University Düsseldorf (grant #2019-13 to CM and HG). KB is supported by the German Academic Exchange Service (DAAD). ZZ was supported by China Scholarship Council (CSC). CK and GGS are supported by the German Ministry of Health (grant # G115F020001). CM is supported by the Heinz-Ansmann foundation for AIDS research. The Center for Structural Studies is funded by the Deutsche Forschungsgemeinschaft (DFG Grant number 417919780 and INST 208/761-1 FUGG).

Competing interests: The authors have declared that no competing interests exist.

AUTHOR CONTRIBUTIONS

Conceptualization: AAJV and CM, **Data curation**: AAJV, KB, CGWG, FB, UH, CK, SB, GGS, IGB, DH, and HG, **Formal analysis**: AAJV, KB, CGWG, FB, ZZ, AS, UH, SB, CK, GGS, IGB, DH, HG, and CM, **Funding acquisition**: AAJV, CM, and HG, **Investigation**: AAJV, KB, CGWG, AS, FB, UH, CK, HG, and IGB, **Methodology**: AAJV, KB, CGWG, SB, HG, and CM, **Project administration**: CM, **Resources**: GGS, IGB, HG, and CM, **Supervision**: CM, DH, HG and AAJV, **Validation**: AAJV and CM, **Visualization**: AAJV, KB, CGWG, FB, CK, GGS, IGB, HG, and CM, **Writing – original draft**: AAJV, and **Writing – review and editing**: AAJV, KB, CGWG, FB, ZZ, AS, UH, CK, SB, GGS, DH, IGB, HG, and CM.

REFERENCES

1. Goila-Gaur, R. & Strebel, K. (2008). HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* **5**, 51.
2. Harris, R. S. & Dudley, J. P. (2015). APOBECs and virus restriction. *Virology* **479-480**, 131-45.
3. Salter, J. D., Bennett, R. P. & Smith, H. C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* **41**, 578-94.
4. Silvas, T. V. & Schiffer, C. A. (2019). APOBEC3s: DNA-editing human cytidine deaminases. *Protein Sci* **28**, 1552-1566.
5. Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J. & Navaratnam, N. (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285-96.
6. Münk, C., Willemsen, A. & Bravo, I. G. (2012). An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol Biol* **12**, 71.
7. LaRue, R. S., Jonsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hotzel, I., Andresdottir, V., Smith, T. P. & Harris, R. S. (2008). The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol* **9**, 104.
8. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-50.
9. Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C. & Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-8.
10. Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J. & Malim, M. H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**, 1392-6.
11. Vasudevan, A. A., Smits, S. H., Hoppner, A., Häussinger, D., Koenig, B. W. & Münk, C. (2013). Structural features of antiviral DNA cytidine deaminases. *Biol Chem* **394**, 1357-70.
12. Zennou, V., Perez-Caballero, D., Gottlinger, H. & Bieniasz, P. D. (2004). APOBEC3G incorporation into human immunodeficiency virus type 1 particles. *J Virol* **78**, 12058-61.
13. Luo, K., Liu, B., Xiao, Z., Yu, Y., Yu, X., Gorelick, R. & Yu, X. F. (2004). Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J Virol* **78**, 11841-52.
14. Svarovskaia, E. S., Xu, H., Mbisa, J. L., Barr, R., Gorelick, R. J., Ono, A., Freed, E. O., Hu, W. S. & Pathak, V. K. (2004). Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like

3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *J Biol Chem* **279**, 35822-8.

15. Huthoff, H. & Malim, M. H. (2007). Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J Virol* **81**, 3807-15.
16. Schafer, A., Bogerd, H. P. & Cullen, B. R. (2004). Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. *Virology* **328**, 163-8.
17. Burnett, A. & Spearman, P. (2007). 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* **81**, 5000-13.
18. Browne, E. P., Allers, C. & Landau, N. R. (2009). Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. *Virology* **387**, 313-21.
19. Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S. & Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-9.
20. Yu, Q., König, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J. M. & Landau, N. R. (2004). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* **11**, 435-42.
21. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. & Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103.
22. Iwatani, Y., Chan, D. S., Wang, F., Maynard, K. S., Sugiura, W., Gronenborn, A. M., Rouzina, I., Williams, M. C., Musier-Forsyth, K. & Levin, J. G. (2007). Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res* **35**, 7096-108.
23. Holmes, R. K., Koning, F. A., Bishop, K. N. & Malim, M. H. (2007). APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J Biol Chem* **282**, 2587-95.
24. Münk, C., Jensen, B. E., Zielonka, J., Häussinger, D. & Kamp, C. (2012). Running Loose or Getting Lost: How HIV-1 Counters and Capitalizes on APOBEC3-Induced Mutagenesis through Its Vif Protein. *Viruses.* **4**, 3132-3161.
25. Bishop, K. N., Holmes, R. K. & Malim, M. H. (2006). Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J Virol* **80**, 8450-8.
26. Mbisa, J. L., Bu, W. & Pathak, V. K. (2010). APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. *J Virol* **84**, 5250-9.
27. Strebel, K. (2005). APOBEC3G & HTLV-1: inhibition without deamination. *Retrovirology* **2**, 37.
28. Mehle, A., Strack, B., Ancuta, P., Zhang, C., McPike, M. & Gabuzda, D. (2004). Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J Biol Chem* **279**, 7792-8.
29. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **9**, 1404-7.
30. Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P. & Yu, X. F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-60.
31. Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., König, R., Bollman, B., Münk, C., Nymark-McMahon, H. & Landau, N. R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**, 21-31.
32. Bogerd, H. P., Doehle, B. P., Wiegand, H. L. & Cullen, B. R. (2004). 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* **101**, 3770-4.

33. Mangeat, B., Turelli, P., Liao, S. & Trono, D. (2004). A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. *J Biol Chem* **279**, 14481-3.
34. Zhang, W., Huang, M., Wang, T., Tan, L., Tian, C., Yu, X., Kong, W. & Yu, X. F. (2008). Conserved and non-conserved features of HIV-1 and SIVagm Vif mediated suppression of APOBEC3 cytidine deaminases. *Cell Microbiol* **10**, 1662-75.
35. Smith, J. L. & Pathak, V. K. (2010). Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. *J Virol* **84**, 12599-608.
36. Zhang, Z., Gu, Q., de Manuel Montero, M., Bravo, I. G., Marques-Bonet, T., Häussinger, D. & Münk, C. (2017). Stably expressed APOBEC3H forms a barrier for cross-species transmission of simian immunodeficiency virus of chimpanzee to humans. *PLoS Pathog* **13**, e1006746.
37. Dang, Y., Wang, X., Esselman, W. J. & Zheng, Y. H. (2006). Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J Virol* **80**, 10522-33.
38. Wiegand, H. L., Doehle, B. P., Bogerd, H. P. & Cullen, B. R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J* **23**, 2451-8.
39. Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T. & Peterlin, B. M. (2004). Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J Virol* **78**, 6073-6.
40. Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L. & Harris, R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-34.
41. Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., McDougle, R. M., Brown, W. L., Law, E. K. & Harris, R. S. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-70.
42. Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G., Harris, S., Shah, R. R., Resnick, M. A., Getz, G. & Gordenin, D. A. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **45**, 970-6.
43. Buisson, R., Langenbucher, A., Bowen, D., Kwan, E. E., Benes, C. H., Zou, L. & Lawrence, M. S. (2019). Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **364**.
44. Cortez, L. M., Brown, A. L., Dennis, M. A., Collins, C. D., Brown, A. J., Mitchell, D., Mertz, T. M. & Roberts, S. A. (2019). APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS Genet* **15**, e1008545.
45. Henderson, S. & Fenton, T. (2015). APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol Med* **21**, 274-84.
46. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **5**, 704-12.
47. Green, A. M. & Weitzman, M. D. (2019). The spectrum of APOBEC3 activity: From anti-viral agents to anti-cancer opportunities. *DNA Repair (Amst)* **83**, 102700.
48. Olson, M. E., Harris, R. S. & Harki, D. A. (2018). APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem Biol* **25**, 36-49.
49. Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G. G. & Münk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* **281**, 22161-72.
50. Yu, Q., Chen, D., König, R., Mariani, R., Unutmaz, D. & Landau, N. R. (2004). APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. *J Biol Chem* **279**, 53379-86.

51. Langlois, M. A., Beale, R. C., Conticello, S. G. & Neuberger, M. S. (2005). Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* **33**, 1913-23.
52. Suspene, R., Guetard, D., Henry, M., Sommer, P., Wain-Hobson, S. & Vartanian, J. P. (2005). 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* **102**, 8321-6.
53. Baumert, T. F., Rosler, C., Malim, M. H. & von Weizsacker, F. (2007). Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. *Hepatology* **46**, 682-9.
54. Vartanian, J. P., Guetard, D., Henry, M. & Wain-Hobson, S. (2008). Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **320**, 230-3.
55. Stauch, B., Hofmann, H., Perkovic, M., Weisel, M., Kopietz, F., Cichutek, K., Münk, C. & Schneider, G. (2009). Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. *Proc Natl Acad Sci U S A* **106**, 12079-84.
56. Ahasan, M. M., Wakae, K., Wang, Z., Kitamura, K., Liu, G., Koura, M., Imayasu, M., Sakamoto, N., Hanaoka, K., Nakamura, M., Kyo, S., Kondo, S., Fujiwara, H., Yoshizaki, T., Mori, S., Kukimoto, I. & Muramatsu, M. (2015). APOBEC3A and 3C decrease human papillomavirus 16 pseudovirion infectivity. *Biochem Biophys Res Commun* **457**, 295-9.
57. Suspene, R., Aynaud, M. M., Koch, S., Pasdeloup, D., Labetoulle, M., Gaertner, B., Vartanian, J. P., Meyerhans, A. & Wain-Hobson, S. (2011). Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. *J Virol* **85**, 7594-602.
58. Perkovic, M., Schmidt, S., Marino, D., Russell, R. A., Stauch, B., Hofmann, H., Kopietz, F., Kloke, B. P., Zielonka, J., Strover, H., Hermle, J., Lindemann, D., Pathak, V. K., Schneider, G., Lochelt, M., Cichutek, K. & Munk, C. (2009). Species-specific inhibition of APOBEC3C by the prototype foamy virus protein bet. *J Biol Chem* **284**, 5819-26.
59. Horn, A. V., Klawitter, S., Held, U., Berger, A., Vasudevan, A. A., Bock, A., Hofmann, H., Hanschmann, K. M., Trosemeier, J. H., Flory, E., Jabulowsky, R. A., Han, J. S., Lower, J., Lower, R., Münk, C. & Schumann, G. G. (2014). Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. *Nucleic Acids Res* **42**, 396-416.
60. Hultquist, J. F., Binka, M., LaRue, R. S., Simon, V. & Harris, R. S. (2012). Vif proteins of human and simian immunodeficiency viruses require cellular CBFbeta to degrade APOBEC3 restriction factors. *J Virol* **86**, 2874-7.
61. Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., Candinas, D., Sommer, P., Wain-Hobson, S., Vartanian, J. P. & Greeve, J. (2006). Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**, 1364-74.
62. Refsland, E. W., Hultquist, J. F. & Harris, R. S. (2012). Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n. *PLoS Pathog* **8**, e1002800.
63. Bourara, K., Liegler, T. J. & Grant, R. M. (2007). Target cell APOBEC3C can induce limited G-to-A mutation in HIV-1. *PLoS Pathog* **3**, 1477-85.
64. Refsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L. & Harris, R. S. (2010). Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* **38**, 4274-84.

65. Abdel-Mohsen, M., Raposo, R. A., Deng, X., Li, M., Liegler, T., Sinclair, E., Salama, M. S., Ghanem Hel, D., Hoh, R., Wong, J. K., David, M., Nixon, D. F., Deeks, S. G. & Pillai, S. K. (2013). Expression profile of host restriction factors in HIV-1 elite controllers. *Retrovirology* **10**, 106.
66. Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A., Sugiura, W. & Iwatani, Y. (2012). The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat Struct Mol Biol* **19**, 1005-10.
67. Zhang, Z., Gu, Q., Jaguva Vasudevan, A. A., Jeyaraj, M., Schmidt, S., Zielonka, J., Perkovic, M., Heckel, J. O., Cichutek, K., Häussinger, D., Smits, S. H. & Münk, C. (2016). Vif Proteins from Diverse Human Immunodeficiency Virus/Simian Immunodeficiency Virus Lineages Have Distinct Binding Sites in A3C. *J Virol* **90**, 10193-10208.
68. Jaguva Vasudevan, A. A., Hofmann, H., Willbold, D., Häussinger, D., Koenig, B. W. & Münk, C. (2017). Enhancing the Catalytic Deamination Activity of APOBEC3C Is Insufficient to Inhibit Vif-Deficient HIV-1. *J Mol Biol* **429**, 1171-1191.
69. Suspene, R., Henry, M., Guillot, S., Wain-Hobson, S. & Vartanian, J. P. (2005). Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J Gen Virol* **86**, 125-9.
70. Nowarski, R., Britan-Rosich, E., Shiloach, T. & Kotler, M. (2008). Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. *Nat Struct Mol Biol* **15**, 1059-66.
71. Jaguva Vasudevan, A. A., Kreimer, U., Schulz, W. A., Krikoni, A., Schumann, G. G., Häussinger, D., Münk, C. & Goering, W. (2018). APOBEC3B Activity Is Prevalent in Urothelial Carcinoma Cells and Only Slightly Affected by LINE-1 Expression. *Front Microbiol* **9**, 2088.
72. Wittkopp, C. J., Adolph, M. B., Wu, L. I., Chelico, L. & Emerman, M. (2016). A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses. *PLoS Pathog* **12**, e1005865.
73. Hache, G., Liddament, M. T. & Harris, R. S. (2005). The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J Biol Chem* **280**, 10920-4.
74. Chen, Q., Xiao, X., Wolfe, A. & Chen, X. S. (2016). The in vitro Biochemical Characterization of an HIV-1 Restriction Factor APOBEC3F: Importance of Loop 7 on Both CD1 and CD2 for DNA Binding and Deamination. *J Mol Biol* **428**, 2661-70.
75. Wan, L., Nagata, T. & Katahira, M. (2018). Influence of the DNA sequence/length and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F. *Phys Chem Chem Phys* **20**, 3109-3117.
76. Nakashima, M., Ode, H., Kawamura, T., Kitamura, S., Naganawa, Y., Awazu, H., Tsuzuki, S., Matsuoka, K., Nemoto, M., Hachiya, A., Sugiura, W., Yokomaku, Y., Watanabe, N. & Iwatani, Y. (2016). Structural Insights into HIV-1 Vif-APOBEC3F Interaction. *J Virol* **90**, 1034-47.
77. Schumann, G. G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem Soc Trans* **35**, 637-42.
78. Schumann, G. G., Gogvadze, E. V., Osanai-Futahashi, M., Kuroki, A., Münk, C., Fujiwara, H., Ivics, Z. & Buzdin, A. A. (2010). Unique functions of repetitive transcriptomes. *Int Rev Cell Mol Biol* **285**, 115-88.
79. Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D. & An, W. (2011). Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res* **39**, e16.
80. Xiao, X., Li, S. X., Yang, H. & Chen, X. S. (2016). Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat Commun* **7**, 12193.
81. Fang, Y., Xiao, X., Li, S. X., Wolfe, A. & Chen, X. S. (2018). Molecular Interactions of a DNA Modifying Enzyme APOBEC3F Catalytic Domain with a Single-Stranded DNA. *J Mol Biol* **430**, 87-101.

82. Marino, D., Perkovic, M., Hain, A., Jaguva Vasudevan, A. A., Hofmann, H., Hanschmann, K. M., Muhlebach, M. D., Schumann, G. G., König, R., Cichutek, K., Häussinger, D. & Münk, C. (2016). APOBEC4 Enhances the Replication of HIV-1. *PLoS One* **11**, e0155422.
83. Adolph, M. B., Ara, A., Feng, Y., Wittkopp, C. J., Emerman, M., Fraser, J. S. & Chelico, L. (2017). Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. *Nucleic Acids Res* **45**, 3378-3394.
84. Solomon, W. C., Myint, W., Hou, S., Kanai, T., Tripathi, R., Kurt Yilmaz, N., Schiffer, C. A. & Matsuo, H. (2019). Mechanism for APOBEC3G catalytic exclusion of RNA and non-substrate DNA. *Nucleic Acids Res* **47**, 7676-7689.
85. Ziegler, S. J., Hu, Y., Devarkar, S. C. & Xiong, Y. (2019). APOBEC3A loop 1 is a determinant for ssDNA binding and deamination. *Biochemistry*.
86. Bohn, J. A., DaSilva, J., Kharytonchyk, S., Mercedes, M., Vosters, J., Telesnitsky, A., Hatziioannou, T. & Smith, J. L. (2019). Flexibility in nucleic acid binding is central to APOBEC3H antiviral activity. *J Virol*.
87. Rathore, A., Carpenter, M. A., Demir, O., Ikeda, T., Li, M., Shaban, N. M., Law, E. K., Anokhin, D., Brown, W. L., Amaro, R. E. & Harris, R. S. (2013). The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. *J Mol Biol* **425**, 4442-54.
88. Siu, K. K., Sultana, A., Azimi, F. C. & Lee, J. E. (2013). Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat Commun* **4**, 2593.
89. Dang, Y., Abudu, A., Son, S., Harjes, E., Spearman, P., Matsuo, H. & Zheng, Y. H. (2011). Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. *J Virol* **85**, 5691-5.
90. Murrell, B., Vollbrecht, T., Guatelli, J. & Wertheim, J. O. (2016). The Evolutionary Histories of Antiretroviral Proteins SERINC3 and SERINC5 Do Not Support an Evolutionary Arms Race in Primates. *J Virol* **90**, 8085-9.
91. Ohno, S. (1970). Evolution by gene duplication Springer-Verlag Heidelberg. Germany.
92. Nakano, Y., Aso, H., Soper, A., Yamada, E., Moriwaki, M., Juarez-Fernandez, G., Koyanagi, Y. & Sato, K. (2017). A conflict of interest: the evolutionary arms race between mammalian APOBEC3 and lentiviral Vif. *Retrovirology* **14**, 31.
93. Sawyer, S. L., Emerman, M. & Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* **2**, E275.
94. Picard, L., Ganivet, Q., Allatif, O., Cimarelli, A., Guéguen, L. & Etienne, L. (2020). DGINN, an automated and highly-flexible pipeline for the Detection of Genetic INNovations on protein-coding genes. *bioRxiv*, 2020.02.25.964155.
95. Münk, C., Beck, T., Zielonka, J., Hotz-Wagenblatt, A., Chareza, S., Battenberg, M., Thielebein, J., Cichutek, K., Bravo, I. G., O'Brien, S. J., Lochelt, M. & Yuhki, N. (2008). Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. *Genome Biol* **9**, R48.
96. Hassan, M. A., Butty, V., Jensen, K. D. & Saeij, J. P. (2014). The genetic basis for individual differences in mRNA splicing and APOBEC1 editing activity in murine macrophages. *Genome Res* **24**, 377-89.
97. Gu, T., Gatti, D. M., Srivastava, A., Snyder, E. M., Raghupathy, N., Simecek, P., Svenson, K. L., Dotu, I., Chuang, J. H., Keller, M. P., Attie, A. D., Braun, R. E. & Churchill, G. A. (2016). Genetic Architectures of Quantitative Variation in RNA Editing Pathways. *Genetics* **202**, 787-98.
98. Shen, F. & Kidd, J. M. (2020). Rapid, Paralog-Sensitive CNV Analysis of 2457 Human Genomes Using QuicK-mer2. *Genes (Basel)* **11**.
99. Tao, L., Jiang, Z., Xu, M., Xu, T. & Liu, Y. (2019). Induction of APOBEC3C Facilitates the Genotoxic Stress-Mediated Cytotoxicity of Artesunate. *Chem Res Toxicol* **32**, 2526-2537.

100. Athanassiou, M., Hu, Y., Jing, L., Houle, B., Zarbl, H. & Mikheev, A. M. (1999). Stabilization and reactivation of the p53 tumor suppressor protein in nontumorigenic revertants of HeLa cervical cancer cells. *Cell Growth Differ* **10**, 729-37.
101. Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D. & Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *J Virol* **72**, 8463-71.
102. Bähr, A., Singer, A., Hain, A., Vasudevan, A. A., Schilling, M., Reh, J., Riess, M., Panitz, S., Serrano, V., Schweizer, M., König, R., Chanda, S., Häussinger, D., Kochs, G., Lindemann, D. & Münk, C. (2016). Interferon but not MxB inhibits foamy retroviruses. *Virology* **488**, 51-60.
103. Russell, R. A., Wiegand, H. L., Moore, M. D., Schafer, A., McClure, M. O. & Cullen, B. R. (2005). Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol* **79**, 8724-31.
104. Simm, M., Shahabuddin, M., Chao, W., Allan, J. S. & Volsky, D. J. (1995). Aberrant Gag protein composition of a human immunodeficiency virus type 1 vif mutant produced in primary lymphocytes. *J Virol* **69**, 4582-6.
105. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Lower, J., Stratling, W. H., Lower, R. & Schumann, G. G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* **40**, 1666-83.
106. Rose, P. P. & Korber, B. T. (2000). Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics* **16**, 400-1.
107. Jaguva Vasudevan, A. A., Perkovic, M., Bulliard, Y., Cichutek, K., Trono, D., Häussinger, D. & Münk, C. (2013). Prototype foamy virus Bet impairs the dimerization and cytosolic solubility of human APOBEC3G. *J Virol* **87**, 9030-40.
108. Robert, X. & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-4.
109. Release, S. (2016). 2: Maestro, Schrödinger, LLC, New York, NY, 2017. *Received: February* **21**, 2018.
110. Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., McCann, J. L., Starrett, G. J., Duffy, J. V., Demir, O., Amaro, R. E., Harki, D. A., Harris, R. S. & Aihara, H. (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol* **24**, 131-139.
111. Bas, D. C., Rogers, D. M. & Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* **73**, 765-83.
112. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935.
113. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman (2014). AMBER 14, University of California, San Francisco.
114. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-713.
115. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M., Jr. (2013). Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J Chem Theory Comput* **9**, 2733-2748.
116. Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089-10092.

117. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* **23**, 327-341.
118. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. (2015). Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **11**, 1864-74.
119. Zgarbova, M., Sponer, J., Otyepka, M., Cheatham, T. E., 3rd, Galindo-Murillo, R. & Jurecka, P. (2015). Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J Chem Theory Comput* **11**, 5723-36.
120. Roe, D. R. & Cheatham, T. E., 3rd. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084-95.
121. Iwatani, Y., Takeuchi, H., Strebel, K. & Levin, J. G. (2006). Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J Virol* **80**, 5992-6002.
122. Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-80.
123. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-3.
124. Huson, D. H. & Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* **61**, 1061-7.

FIGURE CAPTIONS

Figure 1. A3C-like protein from sooty mangabey inhibits HIV1Δ*vif* by more than 10-fold. (A) HIV-1Δ*vif* particles were produced with A3C from human, rhesus macaque, chimpanzees (cpz), African green monkey (agm), and A3C-like protein from sooty mangabey monkey (smm), hA3G or vector only. Infectivity of equal amounts of viruses (RT-activity normalized), relative to the virus lacking any A3, was determined by quantification of luciferase activity in HEK293T cells. Presented values represent means ± standard deviations (error bars) for three independent experiments. Asterisks indicate statistically significant differences relative to the effect of the empty vector on infectivity: ***, $p < 0.0001$. (B) Immunoblot analysis of HA-tagged A3 and HIV-1 capsid expression in cell lysates using anti-HA and anti p24 antibodies, respectively. GAPDH served as a loading control. "α" represents anti. (C) Expression and (D) deamination activity of smmA3C and smmA3F-CTD: Immunoblot analysis of HA-tagged smmA3C-like protein, and (WT) smmA3C, and (WT) smmA3F-CTD expression in cell lysates using anti-HA antibody. Tubulin served as a loading control. *In vitro* deamination activity of smmA3C-like

protein, smmA3C, and smmA3F-CTD using lysates of cells that were previously transfected with the respective expression plasmids. Samples were treated with RNAse A; oligonucleotide-containing uracil (U) instead of cytosine served as a marker to denote the migration of deaminated product after restriction enzyme cleavage. S-substrate, P-product. (E) 3D-PCR: HIV-1Δ*vif* produced together with A3C orthologues, hA3G or vector controls were used to transduce HEK293T cells. Total DNA was extracted and a 714-bp fragment of reporter viral DNA was selectively amplified using 3D-PCR. T_d = denaturation temperature. Extensive viral DNA editing profile of smmA3C-like protein and its relative positions of G→A transition mutations are presented in Suppl. Fig. S2A. (F) *In vitro* deamination activity of A3Cs encapsidated in HIV-1Δ*vif*, and SIVagmΔ*vif* particles. Virions were concentrated and lysed in mild lysis buffer and equal amounts of lysate were used for the assay. Numbers 1 and 10 indicate 60 ng and 600 ng of A3 expression vector used for transfection, respectively.

Figure 2. Design and activity of hA3C/smmA3C-like protein chimeras. (A) Sequence alignment of hA3C and smmA3C-like protein, motif 1 (YGTQ) and motif 2 (WEND) are marked with red boxes, red lollipops indicate active site amino acids H66, E68, C97 and C100, while S61 and S188 are colored in purple. (B) Ribbon model of the crystal structure of A3C (PDB 3VOW) depicting the spatial arrangements of helix α1 (YGTQ motif) and loop 1 (WEND motif). Residues of both motifs are presented in purple. Key residues S61, S188, and zinc-coordinating active site residues are denoted as ball and sticks. Sphere represents Zn2+ ion. (C) Structures of the chimeras generated between A3C and smmA3C-like protein. Grey and white boxes indicate fractions of A3C and the smmA3C-like protein, respectively. Regions of hA3C protein derived from exons E1 (amino acids 1-5), E2 (6-58), E3 (59-151), and E4 (152-190) and residues at the borders are marked on top of the hA3C box. Each chimera ("C") encompasses 190 amino acids. Amino acid position (number) at the breakpoints of each chimera is indicated. (D) HIV-1Δ*vif* particles were produced with A3C from human, smm (A3C-like), and h/smm chimeras or vector only. Infectivity of equal amounts of viruses (RT-activity normalized), relative to the virus lacking any A3, was determined by quantification of luciferase activity in HEK293T cells. (E) Illustration of chimera 2 (C2) and variants of C2 or smmA3C-like protein having amino acid

exchanges in the DHIH (circle) or RKYG (square) motif. The red triangle denotes catalytic residue E68A mutation. Amino acid position (number) at the breakpoint of chimera C2 is indicated. (F) HIV-1Δ*vif* particles were produced with C2 and its variants or vector only. Infectivity of equal amounts of viruses (RT-activity normalized), relative to the virus lacking any A3, was determined by quantification of luciferase activity in HEK293T cells. Values are means ± standard deviations (error bars) for three independent experiments. Presented values represent means ± standard deviations (error bars) for three independent experiments. Asterisks indicate statistically significant differences relative to the effect of the empty vector on infectivity: ***, $p < 0.0001$; ns, not significant.

Figure 3. RKYG-WEND exchange in smmA3C-like protein abrogates its antiviral activity. (A) HIV-1Δ*vif* particles were produced with smmA3C-like protein, its mutants E68A (catalytically inactive), RKYG-WEND or vector only. Infectivity of equal amounts of viruses (RT-activity normalized), relative to the virus lacking any A3, was determined by quantification of luciferase activity in HEK293T cells. (B) Immunoblot analyses were performed to quantify HA-tagged A3 proteins and viral p24 proteins in cellular and viral lysates using anti-HA and anti-p24 antibodies, respectively. Tubulin served as a loading control. "α" represents anti. (C) Quantification of hypermutation in viral DNA by 3D-PCR. HIV-1Δ*vif* particles produced in the presence of overexpressed smmA3C-like protein, its variants or vector control were used to transduce HEK293T cells. Total DNA was extracted and a 714-bp fragment of reporter viral DNA was selectively amplified using 3D-PCR. T_d = denaturation temperature. (D) *In vitro* deamination activity of smmA3C-like protein and its variants using lysates of cells that were previously transfected with the respective expression plasmids. Samples were treated with RNAse A; oligonucleotide-containing uracil (U) instead of cytosine served as a marker to denote the migration of deaminated product after restriction enzyme cleavage. S-substrate, P-product.

Figure 4. A3C gains deaminase-dependent anti-HIV-1 activity by a WE-RK change in loop 1. (A) HIV-1Δ*vif* particles were produced with hA3C, its mutants (C97S, S61P, S188I, WE-RK, ND-

YG) or vector only. Infectivity of equal amounts of viruses (RT-activity normalized), relative to the virus lacking any A3C, was determined by quantification of luciferase activity in HEK293T cells. (B) HIV-1Δ*vif* particles were produced with hA3C, its variants such as C97S, WE-RK, WE-RK.C97S, WE-RK.S61P, WE-RK.S61P.C97S, WE-RK.S61P.S188I, WE-RK.S61P.S188I.C97S or vector only. Infectivity of equal amounts of viruses (RT-activity normalized), relative to the virus lacking any A3C, was determined by quantification of luciferase activity in HEK293T cells. (C) Quantification of HA-tagged WT and mutant A3C proteins in both cellular and viral lysates by immunoblot analysis. A3s and HIV-1 capsids were stained with anti-HA and anti-p24 antibodies, respectively. Tubulin served as a loading control. "α" represents anti. (D) 3D-PCR: HIV-1Δ*vif* produced together with hA3C, its variants (as in Fig. 4B), or vector controls were used to transduce HEK293T cells. Total DNA was extracted and a 714-bp fragment of reporter viral DNA was selectively amplified using 3D-PCR. T_d = denaturation. (E) *In vitro* deamination assays to examine the catalytic activity of A3C and its variants using lysates of cells that were previously transfected with respective expression plasmids (as in Fig. 4B). RNAse A-treatment was included; oligonucleotide containing uracil (U) instead of cytosine served as a marker to denote the migration of deaminated product after restriction enzyme cleavage. S-substrate, P-product. The two lower panels represent immunoblot analyses of expression levels of HA-tagged A3C and mutant proteins (α HA (A3C)) and tubulin (α tubulin) which was used as a loading control.

Figure 5. Subcellular localization of human A3C in transfected HeLa cells. Immunofluorescence confocal laser scanning microscopy images of HeLa cells transfected with HA-tagged A3C or A3C.WE-RK. Representative pictures are shown which illustrate nuclear and cytoplasmic localization of the A3Cs (A, B) x-y optical sections. To detect A3Cs (green) immunofluorescence, cells were stained with an anti-HA antibody. Nuclei (blue) were visualized by DAPI staining. (C) 40 randomly chosen transfected cells with A3C or A3C.WE-RK were categorized and cellular localization of A3Cs were quantified.

Figure 6. Mutations in loop 1 of A3F-CTD moderately affect the antiviral activity of A3F. (A) Immunoblot analyses were performed to quantify the amounts of HA-tagged WT hA3C and hA3F proteins and their loop 1 mutants in cell lysates and viral particles. HA-tagged A3s and HIV-1 capsid proteins were stained with anti-HA and anti-p24 antibodies, respectively. Tubulin served as a loading control. "α" represents anti. (B) Infectivity of equal amounts of HIV-1Δ*vif* viruses (RT-activity normalized) encapsidating hA3C, hA3F, or their loop 1 mutants relative to the virus lacking any A3 protein was determined by quantification of luciferase activity in transduced HEK293T cells. (C) 3D-PCR: HIV-1Δ*vif* produced together with hA3C, hA3F, and their loop 1 mutants or vector control were used to transduce HEK293T cells. Total DNA was extracted and a 714-bp fragment of reporter viral DNA was selectively amplified using 3D-PCR. T_d = denaturation temperature. (D) *In vitro* deamination assay to examine the catalytic activity of hA3C, hA3F, and their loop variants was performed using lysates of cells that were transfected with the respective A3 expression plasmids. RNAse A-treatment was included; oligonucleotide containing uracil (U) instead of cytosine served as a marker to denote the migration of the deaminated products after restriction enzyme cleavage. S-substrate, P-product. The two lower panels represent immunoblot analyses of expression levels of HA-tagged A3C, A3F and mutant proteins (α HA (A3s)) and tubulin (α tubulin), which was used as a loading control.

Figure 7. Expression of the hA3C.WE-RK variant enhances A3C-mediated L1 restriction significantly. Dual-luciferase reporter assay to evaluate the effect of WT and mutant A3 proteins on L1 retrotransposition activity. (A) Schematic of the L1 retrotransposition reporter construct pYX017 [79]. The $L1_{Rp}$ reporter element is under transcriptional control of the CAG promoter and a polyadenylation signal (A1) at its 3'end. The firefly luciferase (Fluc) cassette has its own promoter (P2) and polyadenylation signal (A2), is expressed from the antisense strand relative to the CAG promoter, and interrupted by an intron (with splice donor [SD] and splice acceptor [SA]) in the transcriptional orientation of the L1 reporter element. (B) Effect of WT and mutant A3 proteins on L1 retrotransposition activity indicated by normalized luminescence ratio (NLR). NLR indicating retrotransposition activity observed after

cotransfection of pYX015 and empty pcDNA3.1 (+) expression plasmid was set as 1. Error bars indicate standard deviation (N=4).

Figure 8. ssDNA-protein interaction model of hA3C and hA3C.RKYG. (A) Binding mode model of ssDNA (orange) to hA3C WT based on hA3F-CTD and rhA3G-NTD. Magnifications of the active center (green box) are shown at the bottom for hA3C WT (B) and hA3C.RKYG (C). The side chains of residues in the active center that differ between hA3C WT and the hA3C.RKYG variant are shown in cyan and dark blue, respectively. The Zn^{2+} ion in the active center is shown as a sphere. Ongoing from hA3C WT to the hA3C.RKYG variant, the interface changes from being negatively to being positively charged. The flexible arginine and lysine side chains in the hA3C.RKYG variant can interact with the negatively charged backbone of ssDNA (panel C), stabilizing this interaction.

Figure 9. Recombinant hA3C.WE-RK efficiently catalyzes and displays improved interaction with ssDNA. (A) The purity of the recombinantly produced and affinity-purified proteins GST, A3C-GST, and A3C.WE-RK-GST was demonstrated by SDS-PAGE and subsequent Coomassie blue staining of the gel. The prestained protein ladder (M) indicates molecular mass. (B) *In vitro* deamination assay to examine the catalytic activity of purified GST and GST fusion proteins A3C-GST and A3C.WE-RK-GST was performed. RNAse A-treatment was included; oligonucleotide containing uracil (U) instead of cytosine served as a marker to denote the migration of the deaminated products after restriction enzyme cleavage. S-substrate, P-product. (C) EMSA with GST-tagged hA3C.WE-RK-GST and A3C-GST produced in HEK293T cells was performed with 30-nt ssDNA target DNA labelled with 3'-labelled biotin. Indicated amounts of protein (at the bottom of the blot, in nM) were titrated with 20 fmol of DNA. Presence of competitor DNA (unlabeled 80-nt DNA used in deamination assay, 200-fold molar excess added) used to demonstrate the specific binding of the protein to DNA being causative for the shift.

Figure 10. Species tree and one-letter amino acid sequence consensus of the loop 1 in A3C, A3D-CTD and A3F-CTD. The size of the amino acid symbol is proportional to its conservation among the sequences used. The orange dots in the species tree indicate the nodes used for consensus inference and correspond to the different rows in the table. The median values for the most recent common ancestor and the 95% confidence interval (obtained from http://www.timetree.org/) are indicated close to these reference nodes.