Loop 1 of APOBEC3C Regulates its Antiviral Activity against HIV-1

Ananda Ayyappan Jaguva Vasudevan1*,†, Kannan Balakrishnan1,2†, Christoph G. W. Gertzen3,4,5, Fanni Borvet6, Zeli Zhang1†, Anucha Sangwiman1, Ulrike Held7, Caroline Küstermann7, Sharmistha Banerjee2, Gerald G. Schumann7, Dieter Häussinger1, Ignacio G. Bravo6, Holger Gohlke3,4 and Carsten Münk1*

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

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

https://doi.org/10.1016/j.jmb.2020.10.014
Edited by Eric O. Freed

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 (SIVΔvif), but for unknown reasons, it inhibits HIV-1Δvif 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-1Δvif 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.

© 2020 Elsevier Ltd. All rights reserved.
Introduction

The APOBEC3 (A3) family of single-stranded (ss) DNA cytidine deaminases builds an intrinsic immune defense against retroviruses, retrotransposons, and other viral pathogens. 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. Z motifs can be classified into three groups (Z1, Z2, Z3), but share the consensus signature HXE[X25-28]PC[X2-4]C (where X indicates a non-conserved position). A3C is the only single-domain A3Z2 protein in humans. The knowledge about A3C is rather sparse. A3C is distributed in both cytoplasm and nucleus and represses the replication of LINE-1 (L1) endogenous retrotransposons.

The crystal structure of A3C and its HIV-1 Vif-binding interface has been solved. 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. We have extended this finding and identified additional Vif interaction sites in the α4 helix of A3C. Apart from a previous study that predicted putative DNA substrate binding pockets, biochemical and structural aspects of A3C enzymatic activity and their relevance for antiviral activity remain hitherto not well investigated.

Recently, we have shown that increasing the catalytic activity of A3C by an S61P substitution in loop 3 is not sufficient to restrict HIV-1Δvif. It is unknown why A3C can potently restrict SIVΔvif while HIV-1Δvif is largely resistant, despite the fact that wild-type (WT) human A3C possesses reasonable catalytic activity and is encapsidated efficiently into retroviral particles. 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 (smmA3C-like 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 1(a) shows the level of relative infectivity of HIV-1Δvif in the presence of the tested A3 proteins. Human, rhesus, chimpanzee, and African green monkey A3C proteins reduced the relative infectivity of HIV-1Δvif similarly by approximately 60–70%. Conversely, smmA3C-like protein inhibited HIV-1Δvif replication by more than one order of magnitude (Figure 1(a)). 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 (Figure 1(b)). Efficiency of viral incorporation of the smmA3C-like protein was similar to that of hA3G, but much lower compared to hA3C (Suppl. Figure S1(a)).

The smmA3C-like construct was originally described to express A3C of sooty mangabey monkey.69 However, using alignments of primate A3Z2 and related A3 proteins, we found that the open reading frame consists of exons from both smmA3C and smmA3F 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 1MNPKIR6) and third “exon” (amino acids 153FKYC to EILE190) were derived from smmA3C (i.e., coding regions of exon 1 and exon 4 of the smmA3C gene) while second (amino acids 195DVE to GYED198) in smmA3F-origin (smmA3F C-terminal domain, CTD, exon 5 and exon 6 of smmA3F gene) (Suppl. Figure S1(b)). 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 (Figure 1(c)). In contrast to our expectations, only the smmA3C-like protein and not smmA3C showed enhanced cytidine deaminase activity (Figure 1(d)). Not surprisingly, like hA3C70 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-1Δvif by smmA3C-like protein was similar to or slightly stronger than restriction by hA3G (Figure 1(a)), we analyzed the DNA-editing capacity of these A3s during infection by “3D-PCR”.70,71 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 (Tₐ) (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-1Δvif viruses encapsidating hA3C, rhA3C, cpzA3C, or agmA3C yielded amplicons at Tₐ ≥ 86.3 °C, whereas the activity of smmA3C-like protein allowed to produce amplicons at Tₐ < 84.2 °C. In control reactions using virions produced in the presence of hA3G, PCR amplification of viral DNA was detectable at lower Tₐ (85.2 °C and weakly at 84.2 °C) (Figure 1(e)). Importantly, using the vector control sample (no A3), PCR amplicons could be amplified only at higher Tₐ (87.6 °C). To study the effect of smmA3C-like protein in HIV-1Δvif, 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-1Δvif with a rate of 17.16% and predominantly favored the expected GA dinucleotide context (Suppl. Figure S2(a)). Thus, smmA3C-like protein caused a higher G → A mutation rate in HIV-1Δvif than our previously described enhanced activity mutant A3C.S61P (see Figure 2(a) and (b) for sequence and structure), A3G and A3F.70 In addition, we applied qualitative in vitro cytidine deamination assays using A3 proteins isolated from HIV-1Δvif and SIVagmΔvif viral particles.72,73 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 Mse1 restriction site. The efficiency of Mse1 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-1Δvif particles, did not yield a substantial product resulting from ssDNA cytidine deamination,70 whereas smmA3C-like protein generated high amounts of deamination products (Figure 1(f)). 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 SIVagmΔvif particles produced the expected deamination products, whereas smmA3C-like protein exhibited the strongest catalytic activity, regardless of whether encapsidated in SIVagmΔvif or HIV-1Δvif particles (Figure 1(f)). 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 (Figure 2(a)). To facilitate the identification of distinct determinants of smmA3C-like protein that confer HIV-1 inhibition, ten different hA3C/smmA3C-like chimeras were constructed (Figure 2(c)). Next, viral particles containing different chimeric proteins were produced and their infectivity was tested. As shown in Figure 2(d), chimeras C2, C4, and C8 strongly reduced the infectivity of HIV-1Δvif. Especially, chimera C2 (hA3C harboring a swap of 36 residues of the smmA3C-like protein at the N-terminal end) inhibited HIV-1Δvif replication by about two orders of magnitude. In comparison, chimeras C6 and C9 reduced viral infectivity by only 72% relative to vector control (Figure 2(d)).

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 (Figure 2(c)), were more highly expressed than C1, C4, C6, and C10 (Suppl. Figure S2(b)). 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-1Δvif (Suppl. Figure S2(b), 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. Figure S2(c)). Here we used lysates of transfected HEK293T cells to readily evaluate the catalytic activity of the chimeric A3Cs. Only chimeras C2 and C4 showed the level of deamination similar to those produced by smmA3C-like protein (Suppl. Figure S2(c)). Taken together, chimeras C2 and C4 have the strongest HIV-1Δvif-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 15DPHIFYFH20 (shortly “DHIH”) and 25RKAYG29 (named “RKYG”) as presented in the sequence alignments of Figure 2(a), and generated variants of C2 by swapping one, two, or four amino acids with the analogous residues of hA3C as presented in Figure 2(e). First, we cloned the C2 variants C2.DH-YG (YGTQ motif of helix 1) and C2.RKYG-WEND (WEND motif of loop 1, Figure 2(f) and Suppl. Figure S3(a)). Hence, we constructed the mutants C2.R25W, C2.K26E, C2.Y28N, and C2.G29D (Figure 2(e)) 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

Figure 1. A3C-like protein from sooty mangabey inhibits HIV1Δvif by more than 10-fold. (a) HIV-1Δvif 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. “x” 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 A3C-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-1Δvif 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. Td = denaturation temperature. Extensive viral DNA editing profile of smmA3C-like protein and its relative positions of G → A transition mutations are presented in Suppl. Figure S2(a). (f) In vitro deamination activity of A3Cs encapsidated in HIV-1Δvif, and SIVagmΔvif 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.
Evolution, role of loop 1 of A3C
similar to C2 (Suppl. Figure S3(a)), but mutation of the RKYG motif in the RKYG-WEND variant resulted in a loss of deamination activity (Suppl. Figure S3(a)). 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. Figure S3(a)). Consistent with the data obtained from the in vitro assay, the chimeric C2.RKYG-WEND variant failed to restrict the infectivity of HIV-1Δvif (Figure 2(f)). Immunoblot analysis of cell and viral lysates confirmed that cellular expression and viral encapsidation of these variants were comparable (Suppl. Figure S3(b)). 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. Figure S3(c), only HIV-1Δvif 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 (Figure 2(e)) inhibited its antiviral activity (Figure 3(a) and (b)). DNA-editing capacity of HIV-1 genomes (Figure 3(c)), and catalytic activity in vitro (Figure 3(d)) 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 (Figures 2(f), 3(a), (c), (d), and Suppl. Figs. S3(a) and (c)). To identify the residues in hA3C that are critically required for the deamination-dependent antiviral activity against HIV-1Δvif, we mutated the loop 1 of hA3C with 25WE to 25RK and 28ND to 28YG residues and compared their antiviral capacity (see A3C alignment and ribbon diagram Figure 2(a) and (b)). As controls, we included additional mutants such as a catalytically inactive non-Zn2+-coordinating C97 mutant, A3C.C97S, and the variants A3C.S61P and A3C.S188I exhibiting enhanced deaminase activity. Compared to WT hA3C, WE-RK greatly enhanced inhibition of HIV-1Δvif and the ND-YG variant behaved like WT A3C, while S61P and S188I demonstrated only marginally increased HIV-1Δvif restriction (Figure 4(a)). Importantly, active site mutant A3C.C97S did not inhibit HIV-1Δvif (Figure 4(a)). Enhancement of the antiviral activity of hA3C.WE-RK compared to WT hA3C appear to result neither 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. Figure S4(a)).

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 and S188I 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-1Δvif were abolished by the active site ablating mutation C97S, indicating the importance of the enzymatic activity of A3C (Figure 4(b)). In comparison, introducing either the
single mutation S61P or the double mutation S61P.S188I did not considerably change the activity of A3C.WE-RK (Figure 4(b)). Immunoblot analysis of cell and viral lysates demonstrated that hA3C and all mutants (except A3C.WE-RK.C97S, A3C.WE-RK.S61P.C97S, and WE-RK.S61P.S188I.C97S) were expressed at comparable levels (Figure 4(c)). 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 (Figure 4(c)). Moreover, we confirmed the effects of all mutants on HIV-1Δvif propagation by 3D-PCR (Figure 4(d)) and deamination assays in vitro (Figure 4(e)). 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 (Figure 5(a) and (b)). This distribution was found in 85.5% 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, respectively, displayed these proteins solely in the nucleus (Figure 5(c)). Together, we infer that hA3C and hA3C.WE-RK had a similar distribution in HeLa cells.
Evolution, role of loop 1 of A3C
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. Interestingly, the antiviral activity of hA3F is mediated by its CTD. Various loops within the A3F-CTD were recently investigated for their role in substrate binding and enzyme function 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. The residues in loop 1 of smmA3C-like protein are derived from exon 5 of the smmA3F gene, located in the CTD of A3F (Suppl. Figure S1(b)) 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-1Δvif. hA3F and hA3F.RK-WE yielded similar amounts of protein and were equally efficiently encapsidated in HIV-1 particles (Figure 6(a)). However, the HIV-1Δvif inhibiting effect of A3F.RK-WE was about 2-fold lower than WT A3F (Figure 6(b)). Consequently, A3F.RK-WE showed decreased mutation efficiency compared with WT A3F (Figure 6(c) and (d)), which is consistent with data presented in a recent report. 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, 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 applying a dual-luciferase retrotransposition reporter assay. 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 (Figure 7(a)). Consistent with previous reports, overexpression of hA3A, hA3C, and hA3F inhibited L1 reporter retrotransposition by approximately 94%, 68%, and 56%, respectively (Figure 7(b)). 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 (Figure 7(b)). 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 (Figure 7(b)). 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. Figure S4(b)). Furthermore, compared to the inhibition of L1 retrotransposition by hA3C and chimpanzee A3C (~60%), hA3C.S61P inhibited L1 reporter retrotransposition by 75% (Suppl. Figure S4(c) and (d)). 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.

Figure 4. A3C gains deaminase-dependent anti-HIV-1 activity by a WE-RK change in loop 1. (a) HIV-1Δvif particles were produced with hA3C, its mutants (C97S, S61P, S188I, WE-RK, ND-YQ), 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-1Δvif 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-1Δvif produced together with hA3C, its variants (as in (b)), 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. Td = 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 (b)). RNAs were pooled, treated with uracilase (U) instead of cytosine served as a marker to denote the migration of the 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.
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. Figure S5(a)). 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), similar to the ssDNA-bound A3F-CTD model built previously, and aligned the crystal structure of hA3C WT and the model of hA3C.RKYG to this complex (Figure 8(a)–(c)). 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 (Figure 8(c)) in contrast to hA3C WT (Figure 8(b)). 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 (Figure 4(e)). 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 purpose, 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. Figure S5(b)). 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. Figure S5(b)).

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. Figure S5(b)), 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 between 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. Figure S6(a)). 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. Figure S6(b)). 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. Figure S6(c)). 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-1Δvif.

**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 (Figure 9(a)). We first
confirmed that the purified GST fusion proteins are catalytically active (Figure 9(b)). As expected hA3C, WE-RK-GST displayed a stronger enzymatic activity than the WT equivalent and no activity with GST was detected (Figure 9(b)). For EMSA, as a probe, we used a biotin-labeled ssDNA oligonucleotide that harbors a TTCA motif in its central region.70,84 Because hA3C-GST is known to form a stable DNA-protein complex when the protein concentration reaches \( \geq 20 \text{ nM} \);70 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 (Figure 9(c)). 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 (Figure 9(c) and Suppl. Figure 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. Figure S7).70

**Evolution of A3Z2 loop 1 regions in primates**

Because of the strong evolutionary relationship between A3C, the CTD of A3F, and related A3Z2
proteins, 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. Figure 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 Platyrhini/Catarrhini split 43.2 Mya (41.0–45.7 Mya) and before the Cercopithecoidea/Homoidea (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. Figure S8) with a tree in which monophyly of each of the large six clades identified was enforced (Suppl. Figure 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 (Figure 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. Figure S9) and, in parallel, performed a consensus analysis of the extant sequences (Figure 10 and Suppl. Figure 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 (Figure 10). Strikingly, in the ancestor of Catarrhini A3C at around 29.4 Ma (27.6–31.3 Ma), this motif had already evolved to LWEA (Suppl. Figure S9), and this is the common extant form in Old World monkeys and apes (Figure 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 Ma (10.7–16.6 Ma) and before the speciation within Chlorocebus at 3.42 Ma (2.05–4.15 Ma) (Figure 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. 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-1 \( \text{A} \text{vif} \) infectivity to the level accomplished by A3G and they do not directly participate in catalytic activity. Because our repeated attempts to express A3F-CTD in human cells were not successful (Figure 1(c)), 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 RYKG motif. Replacing RYKG 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 deaminase 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. 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 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. In accordance, our study indicates that 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.RYKG can interact with the negatively charged backbone of ssDNA via the positively charged loop 1 side chains of R25 and K26 (Figure 8(c)). Like our modeling strategy, Fang et al. 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. Figure S5(b)) 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. Figure S5(b)). 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. 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, 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 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. 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 (Cercopithecoidae) and apes (Hominoidea), some 29 Mya. A3F has since then been involved in an Red Queen arms race with retroviral genes. 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][KTJ][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 (Figure 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. 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. Figure 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. In this respect, the field is wanting for a systematisation of protocols and procedures for identifying selection signatures in genes with complex evolutionary histories. 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, 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. 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. 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.\textsuperscript{100} 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\textsubscript{2}. Similarly, HeLa-HA cells\textsuperscript{101} were cultured in DMEM with 10% FCS (BioWest, Nuaillé, France), 2 mM 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.\textsuperscript{102} The HIV-1 vector pSIN.PPT.CMV.Luc.IRES.GFP expresses the firefly luciferase and GFP.\textsuperscript{103} 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.\textsuperscript{33} 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 (Cercocetus torquatus lunulatus), and the cloning strategy for smmA3C-like and the chimeras of hA3C/smmA3C-like plasmids was described recently.\textsuperscript{98} 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.\textsuperscript{57,60,70} 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, hA3C.FK-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.\textsuperscript{104} Individual exons of authentic smmA3C and smmA3F and smmA3F-like genes exons were amplified and cloned in pcDNA3.1. All 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-LucΔvif, 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 luminoimeter (Berthold Detection Systems, Pforzheim, Germany). Transductions were done in triplicates and at least three independent experiments were performed.

**Immunofluorescence microscopy**

$1 \times 10^5$ HeLa cells grown on polyethylene coverslips (Thermo Fisher Scientific) were cotransfected 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, Wuerzburg, Germany) 1:1000 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 minutes. The images were captured by using a 63× 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% glycercol, 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, 14,800 rpm, 4 °C). Samples (cell/viral lysate) were boiled at 95 °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 (1:7500 dilution, MMS-101P, Covance); goat anti-GAPDH (C termminus, 1:15,000 dilution, Everest Biotech, Oxfordshire, UK); mouse anti-β-tubulin antibody (1:4000 dilution, clone B5-1-2; Sigma-Aldrich, Tauflkirchen, Germany); mouse anti-capsid p24/p27 MAb AG3.0 (1:250 dilution, NIH AIDS Reagents); rabbit anti S6 ribosomal protein (SG10; 1:105 dilution in 5% BSA, Cell Signaling Technology, Leiden, The Netherlands). Secondary Abs.: anti-mouse (NA931V), anti-rabbit (NA934V) horseradish peroxidase (1:105 dilution, GE Healthcare) and anti-goat IgG-HRP (1:105 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 genes, reagent 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 1:5000 dilution and the polyclonal rabbit-anti-L1 ORF1p antibody #984 (1:2000 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'GATAAGGTTGGTTATTTGTATAAGGA and reverse 5'GGATTGGTTGGTTATTTGTATAAGG. Samples were treated with 50 μg/ml RNAse A (Thermo Fisher Scientific). Samples were boiled at 95 °C for 5 min and 100 fmol single-stranded DNA substrate (TTCA: 5’-GGATTGGTTGGTTATTTGTATAAGG; GAAGGTGGATTGAAGGTTCAAGAAGGTGATGG). Samples 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). 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 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’-GGATTGGTTGGTTATTTGTATAAGG; GAAGGTGGATTGAAGGTTCAAGAAGGTGATGG). 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’-GGATTGGTTGGTTATTTGTATAAGG and reverse 5’-CCATCCTAAAAAATATATTTCC. 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.70

L1 retrotransposition reporter assay

Relative L1 retrotransposition activity was determined by applying a rapid dual-luciferase reporter-based assay described previously.81 Briefly, 2 × 10^5 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 pYX01581 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, 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 by puromycin containing DMEM medium and 48 h 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 M200PRO (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. L1 retrotransposition activities were expressed as normalized luminescence ratios (NLR) relative to the background signal obtained after cotransfection of pcDNA3.1(+) and pYX015 coding for the retrotransposition-defective L1RP/JM111 element. The 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 3.0109 (escript.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; Schr dinger, 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 hA3G-NDT (PDB ID 5K8282) onto the X-ray crystal structure of hA3F-CTD (PDB ID 5W2M80), the latter of which was co-crystallized with ssDNA. Subsequently, the hA3C X-ray crystal structure (PDB ID 3VOW86) was aligned onto the NTD of hA3G, 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.110 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 5SWW111), a much relevant model similar to 6BUX.112 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 PROPKA113 according to pH 7.4, neutralized by adding counter ions, and solvated in an octahedral box of TIP3P water114 with a minimal water shell of 12 Å around the solute. The Amber package of molecular simulation software115 and the ff14SB force field116 were used to perform the MD simulations. For the Zn^{2+}-ions the Li-Merz parameters for two-fold positively charged metal ions117 were used. To cope with long-range interactions, the “Particle Mesh Ewald” method118 was used; the SHAKE algorithm119 was applied to bonds involving hydrogen atoms. As hydrogen mass repartitioning120 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, 17,500 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 purpose, the starting temperatures of the MD simulations at the beginning of the thermalization were varied by a fraction of one Kelvin. MD simulation of those three variants in complex with ssDNA were performed similarly, treating the DNA with the OL15 force field 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. First, we used the h-bond command to detect hydrogen bonds between residues in the side chains and DNA. The minimal distance over the side chain atoms, not including Cᵦ of the respective residues, and the DNA for each snapshot of the MD simulations and correlated both (Suppl. Figure 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 can be seen in Suppl. Figure S10 (a)–(c) and the root mean square deviation (RMSD) over time is shown in Suppl. Figure S10(d)–(l). 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.70 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.70,84,123 We mixed 1.33 nM (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 complexes were resolved on a 5% native PAGE gel on ice and transferred to a nylon membrane (American Hybond-XL, GE healthcare) using 0.5 X TBE. After the transfer, the membrane containing protein-DNA complexes 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. Figure 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/).124 Phylogenetic inference was performed using RAxMLv8;125 at either the nucleotide level under the GTR+I model or at the amino acid level under the LG+Γ model. Node support was evaluated applying 5000 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 catarhine 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 v3.6.3.126 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.

CRediT authorship contribution statement

Ananda Ayyappan Jaguva Vasudevan: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. Kannan Balakrishnan: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - review & editing. Christoph G.W. Gertzen: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - review & editing. Fanni Borvet: Data curation, Formal analysis, Investigation, Visualization, Writing - review & editing. Zeli Zhang: Formal analysis, Writing - review & editing. Anucha Sangwiman: Formal analysis, Investigation, Writing - review & editing. Ulrike Held: Data curation, Formal analysis, Investigation, Writing - review & editing. Caroline Küstermann: Data curation, Formal analysis, Investigation, Visualization, Writing - review & editing. Sharmistha Banerjee: Data curation, Formal analysis, Methodology, Writing - review & editing. Gerald G. Schumann: Data curation, Formal analysis, Resources, Visualization, Writing - review & editing. Dieter Häussinger: Data curation, Formal analysis, Supervision, Writing - review & editing. Ignacio G. Bravo: Data curation, Formal analysis, Investigation, Resources, Visualization, Writing - review & editing. Holger Gohlke: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing - review & editing. Carsten Münk: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - review & editing.

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 Amézaga (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).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2020.10.014.

Received 31 May 2020; Accepted 9 October 2020; Available online 15 October 2020

Keywords: APOBEC3C_A3F_cytidine deaminase; sooty mangabey monkey; human immunodeficiency virus (HIV); LINE-1; evolution

† A.A.J.V and K.B contributed equally to this article.
‡ La Jolla Institute for Immunology, La Jolla, CA 92037, USA.
References


