Super Restriction Factors against HIV-1 created from APOBEC3

Mollie M. McDonnell

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington
2020

Reading committee:
Michael Emerman, Chair
Adam Geballe
Jaisiri Lingappa

Program authorized to offer degree:
Molecular and Cellular Biology

University of Washington

Abstract

Super Restriction Factors against HIV-1 created from APOBEC3

Mollie M. McDonnell

Chair of the Supervisory Committee:

Michael Emerman

Departments of Human Biology and Basic Sciences,

Fred Hutchinson Cancer Research Center

The innate immune system provides the first defense against novel pathogens. As a part of the innate immune system, humans encode proteins, called restriction factors, that inhibit replication of viruses like HIV-1. These restriction factors would pose a potent block to HIV-1 viral infection; however, HIV-1 encodes accessory proteins to evade or antagonize the host restriction factors. In this dissertation, I created “super restriction factors,” defined as evolution-guided variants of a natural antiviral protein with improved antiviral activity and resistance to viral antagonism. These super restriction factors provide useful insights about the evolution of host restriction factors and the complex interactions with their viral antagonists as well as a prospective approach to understand cross-species transmission.

Table of Contents

List of Figures

List of Supplemental Figures

Chapter 1. Introduction

Viruses are obligate intracellular parasites that depend on a host to coopt resources to replicate. They are the most abundant species on this planet with an estimated 10 nonillion (10 to the 31st power) individual viruses that infiltrate every niche of this world, including the ocean, soil, and all living species (1). Hundreds of different types of viruses, both those that are pathogenic and those that cause no disease, can infect humans. However, we remain largely protected from this virus-filled world because the human immune system has evolved diverse mechanisms to combat viral infections. As a prelude to my thesis work to develop novel variants of antiviral proteins with improved properties, this thesis introduction will describe some of the mechanisms of cell autonomous or intrinsic immunity that are an ancient form of host defense against viral pathogens.

HIV-1 origins and viral lifecycle section

While some viruses, such as Hepatitis B Virus and human papillomaviruses, share a long evolutionary history with humans, many other viruses such as SARS-CoV 2 and Ebola exist in animals and cross species to infect humans (2–6). These events, called zoonoses or zoonotic transmissions, sometimes require viral adaptations to overcome the barriers imposed from a new species and their immune system, in this case humans. Similarly, the pandemic caused by the Human Immunodeficiency Virus 1 (HIV-1), originated from a cross-species transmission from a similar virus circulating in chimpanzees, the Simian Immunodeficiency Virus of chimpanzees (SIVcpz).

More than 35 years after the discovery of HIV, this pandemic still has a global burden on human health. According to the World Health Organization, as of 2018 there are more than 37.9 million people infected with HIV in the world, with 1.7 million new infections occurring annually. While cases have been reported in all countries, sub-Saharan Africa has the largest disease burden. To date, despite the extensive research, there is no cure nor is there a vaccine to prevent HIV infection. However, extensive improvements have been made to make more effective and more tolerable antiretroviral therapies to improve the lifespan of HIV+ patients and prophylactic treatments (PrEP) for those people in higher risk groups (7).

HIV and SIV belong to the lentivirus genus of the virus family, Retroviridae. A broad range of vertebrates including cats, horses, sheep, goats, and primates harbor lentiviruses which cause chronic and deadly diseases in their hosts (8). All known lentiviruses infect immune cells, such as monocytes/macrophages and lymphocytes (7). Untreated HIV replication leads to progressive CD4+ T cell loss and a wide range of immunological abnormalities, leading to increased risk of infectious and oncological complications. Combination antiretroviral therapy drugs are highly effective at inhibiting viral replication, and for individuals that adhere to these drugs, the therapy leads to suppression of viral replication. Viral suppression may prevent or delay the risk of developing chronic HIV infection. If the infection is left untreated, chronic HIV infection can cause widespread immune cell death and progressive failure immune defense called acquired immune deficiency syndrome (AIDS).

HIV was first discovered in 1983 at the Pasteur Institute in Paris, France and categorized as a lentivirus because of its reverse transcriptase activity and its structural

similarity to other lentiviruses as seen via electron microscopy (9, 10). While the first cases were being described in the US, cases were reported in African populations soon after. The genetic diversity of the viruses isolated from African populations was much higher than viruses isolated from other parts of the world, suggesting that the origins of this viral pandemic lie in Africa. Furthermore, the presence of an AIDS-like disease in rhesus macaques eventually led scientists to look for similar viruses in wild populations of primates, linking the origins of HIV (11). The additional discovery and sequencing of viruses from different African primate species, including chimpanzees and gorillas, validated the hypothesis that HIV is a consequence of a cross-species transmission from primates (12, 13).

There are two types of HIV that circulate, HIV-1 (the virus that is responsible for the current AIDS pandemic) and HIV-2. There are four groups of HIV-1: M, N, O, and P, each which arose due to a separate cross-species transmission event directly from African apes into humans. SIVcpz strains from chimpanzees were transmitted on at least two occasions to humans forming the groups M (the precursor that spawned the current HIV-1 pandemic) and N. HIV-1 from groups O and P, however, originated from an SIV infecting gorillas. Furthermore, SIVcpz can be traced back to a recombination event from two SIVs that infect Old World Monkeys, SIVrcm (an SIV that infects red capped mangabeys) and SIVmus/mon/gsn (an SIV that infects the genus of *Cercopithecus* monkeys). On the other hand, HIV-2 infections are much less common and more highly concentrated to West Africa. HIV-2 is less pathogenic and progresses slower than HIV-1. HIV-2 can be categorized into seven distinct groups A-H, each group

coming from a distinct cross-species transmission event from an SIV that infects sooty mangabeys (SIVsmm).

HIV-1 has a genome size of 9.7 kb and encodes for nine genes (*gag*, *pol*, *env*, *vif*, *vpu*, *vpr*, *nef*, *tat*, and *rev*). When HIV-1 infects a susceptible cell, such as T cells, it binds to its receptors, CD4 and co-receptor CXCR4 or CCR5, using the envelope proteins. HIV fuses to the host cell membrane and releases its viral contents into the host cell that is shielded from the immune system by the viral capsid (CA) protein. Inside of capsid contains two copies of the RNA viral genome and other enzymes needed for the subsequent steps of the viral lifecycle. After entry the reverse transcriptase (RT) enzyme begins converting the RNA genome into a double stranded DNA product using the host lys-tRNA as a primer, and host dNTPs. Once the double stranded DNA product is completed, it is integrated into the host genome using a viral enzyme, integrase, and host proteins. Once HIV-1 DNA has integrated into host chromosomes, it can exist in either an active replication phase or a dormant phase (called latency), and the cell remains infected for the duration of its lifetime.

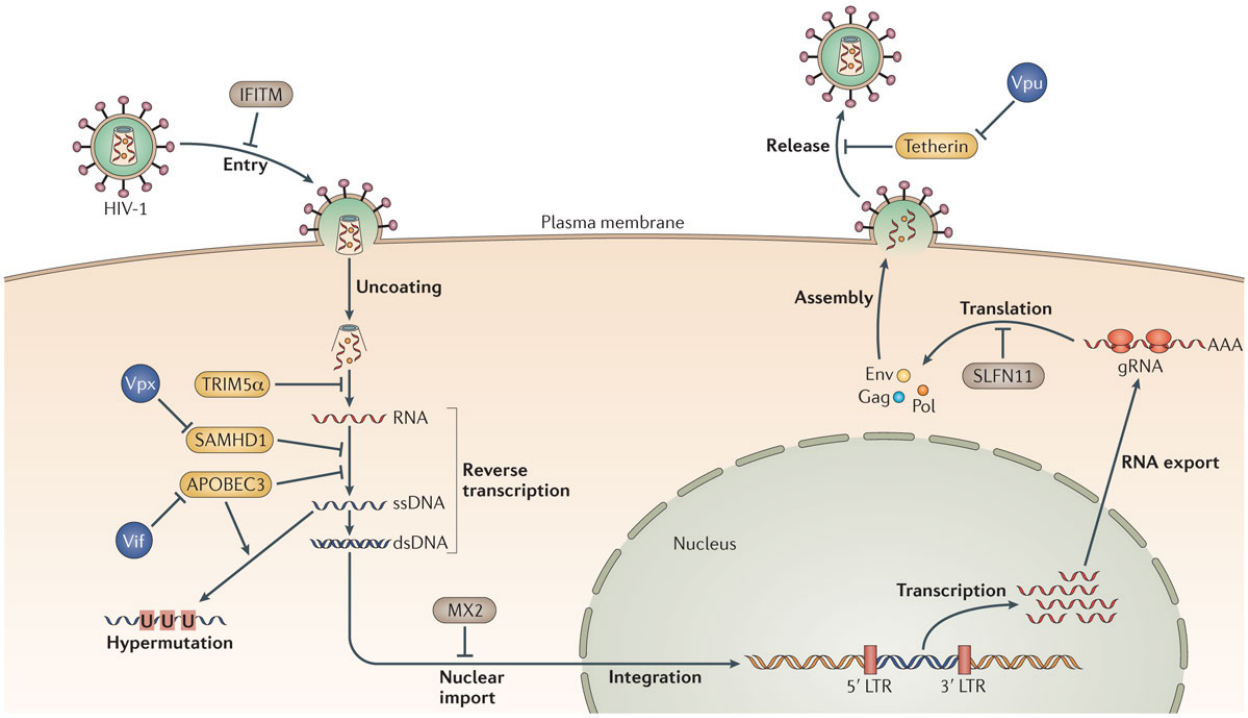
Post-integration, viral gene transcription may begin, but it is initially inefficiently elongated until sufficient Tat protein is produced. Tat enhances transcription elongation from the viral LTR by recruiting a host factor, pTEFb, to the viral RNA at the 5’ end in an RNA element called TAR (tat-response regions), producing completely spliced, multiple spliced, and unspliced full-length RNA. To export incompletely spliced viral mRNAs from the nucleus, the HIV-1 Rev protein binds to the incompletely spliced transcripts (*env*, *vpr*, *vpu*, and *vif*) and unspliced transcripts (*gag-pol* and the viral genome) using the RRE (Rev Response Element) to export these mRNA out of the nucleus via the host

CRM1 pathway. After translation of the viral genes occurs, virions assemble at the host cell membrane. The structural subunits are translated from *gag* to create matrix, capsid, and nucleocapsid as well as two spacer domains. As Gag multimers assemble to form an immature virion, other components, such as two copies of the viral genome, integrase, protease, and reverse transcriptase are packaged. Env proteins are also inserted into the viral membrane to produce the immature virion. Some host proteins are also incorporated into the virions such as the APOBEC3 antiviral proteins described below. The virion buds from the host membrane and is released through co-opting the endosomal sorting complex required for transport (ESCRT) machinery. To become an infectious particle, HIV virions must undergo maturation through cleavage of polyproteins by the viral protease forming the conical capsid shape. This infectious particle has completed the viral lifecycle and can now infect a new cell spreading the infection.

Innate immune system and evolutionary change

The innate immune system is the first response induced to protect the body from infections using germline encoded proteins to act cell autonomously to broadly and quickly recognize pathogens. When cells sense a novel viral pathogen, innate cytokines such as the type I and type III interferon molecules, are released and induce a signaling cascade to produce interferon-stimulated genes (ISGs). These ISGs take on numerous roles, but collectively are highly effective at resisting and controlling pathogens. Some antiviral genes are interferon stimulated, while others are constitutively expressed with tightly regulated activation (14).

Antiviral proteins, also called restriction factors, can act at different stages of the viral lifecycle and often times leverage the obligate nature of viruses to target essential components of viral replication, giving them broad activity (Figure 1). Many restriction factors are induced by interferon. For example, IFITMs are a family of interferon induced transmembrane proteins that alter membrane fluidity to broadly inhibit many viruses, including HIV-1, through targeting viral fusion (15–17). Another interferon induced restriction factor, Tetherin (BST-2), inhibits a wide range of enveloped viruses and works at the stage of viral release from the cell by tethering virions (such as HIV-1 virions) to the cell membrane (18). In contrast to the generalist strategies used by some antiviral proteins, other restriction factors target specific viral proteins. For example, TRIM5α (tripartite motif-containing protein 5α) is an interferon-induced restriction factor that intercepts HIV-1 replication by recognizing and binding to the incoming viral capsid and causes premature uncoating.



Nature Reviews | Microbiology

Figure 1. HIV-1 restriction and resistance factors

In the absence of virally encoded antagonists (or viral escape), host cell proteins called HIV-1 restriction factors (yellow) inhibit various stages of the replication cycle. The tripartite motif-containing protein 5α (TRIM5α) promotes accelerated fragmentation of viral cores, preventing cDNA synthesis. SAM and HD domain-containing protein 1 (SAMHD1) depletes the cellular levels of 2′-deoxynucleoside 5′-triphosphates (dNTPs), which are required for efficient cDNA synthesis. APOBEC3 (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3) proteins interfere with the processivity of HIV-1 reverse transcriptase and induce hypermutation of viral cDNA by cytidine deamination. Tetherin prevents the release of budded virions from the infected cell. Several viral proteins (blue) antagonize these cellular restriction factors. Viral infectivity factor (Vif) antagonizes APOBEC3 proteins, viral protein unique (Vpu) antagonizes tetherin, and the HIV-2 viral protein X (Vpx) antagonizes SAMHD1. HIV-1 resistance factors (brown) inhibit other stages of viral replication and are not counteracted by the virus. Myxovirus resistance 2 (MX2) prevents the nuclear import and integration of viral cDNA. Schlafen 11 (SLFN11) suppresses the translation of viral proteins. Interferon-induced transmembrane proteins (IFITMs) inhibit viral entry by interfering with membrane fusion. dsDNA, double-stranded DNA; gRNA, viral genomic RNA; LTR, long terminal repeat; ssDNA, single-stranded DNA. Reprinted with permission from (19) under license number 4891450268949.

Because restriction factors are germline-encoded, they are subject to slow evolutionary pressures that increase fitness of the host species. That is, polymorphisms in restriction factors within a host species that increase resistance to a pathogen, will eventually rise in frequency since they give a fitness advantage to individuals with that polymorphism. Eventually, these polymorphisms become fixed within the species, rendering the species less susceptible to this particular pathogen, but still sensitive to new variants of the pathogen that have adapted to the new polymorphism, or to other pathogens. This pattern of host resistance and virus adaptation leads to what has been called the host-virus arms race, or genetic conflict.

Signatures of these genetic conflicts can be measured by amino acid substitutions. Comparing the accumulation of nonsynonymous mutations (dN) to synonymous mutations (dS) over time can be analyzed using maximum likelihood statistical methods. Many non-coding regions in the genome have the same ratio of dN

to dS mutations. These types of genes, such as pseudogenes, are said to be evolving under neutral selection in that changes in allele frequency are due to genetic drift rather than fitness effects. On the other hand, most host protein-coding genes are evolving under negative (purifying) selection to purge the non-synonymous mutation from the population to maintain the function of the protein. Negative selection is reflected in a dN to dS ratio of less than one. By contrast, the interaction between restriction factors and viral proteins lead to evolution under positive selection. The genetic conflict these proteins results in an accumulation of dN over dS, with a ratio greater than 1. Ultimately, these positive selection analyses provide insight about restriction factor function, cross-species transmission events, and ancient viral infections.

Many types of genetic conflict between virus and host are driven by proteins encoded by the virus to directly counter-act restriction factors. We call these factors viral antagonists. However, viral escape can be also achieved through selecting for escape mutations that effect binding affinity of restriction factors. Together, these evasion techniques overcome the host pressure from the restriction factors and allow for successful infection of HIV-1. In turn, this exerts evolutionary pressure on the host to select for alleles that encode for antiviral proteins that are resistant to or escape viral antagonism by uncoupling the protein-protein interaction. However, the virus quickly adapts and evolves new mutations at sites that directly interact with restriction factors to restore the interaction. This back-and-forth evolution between host restriction factors and viral proteins is termed the host evolutionary arms' race since the adaptation and counter-adaptation is a continual feedback cycle of genetic conflict (Figure 2).

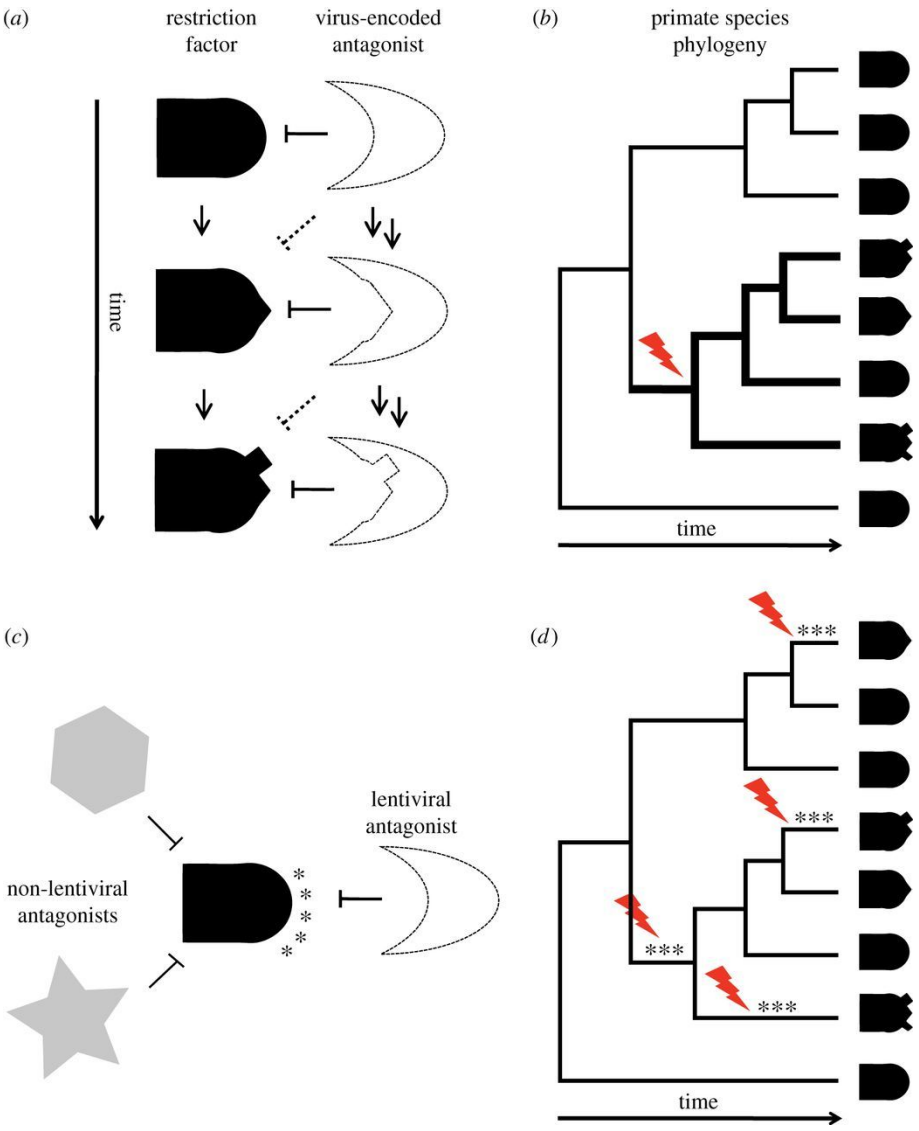


Figure 2. Virus-driven evolution fuels a genetic conflict.

(*a*) Depicted is a cycle of host–virus coevolution initiated when an antagonist encoded by a pathogenic virus counteracts a host restriction factor. Host factor variants are drawn in solid black, and viral antagonists are drawn in white. Selection of host factor variants that successfully evade antagonism is followed by counter-selection of viral variants that re-establish the host–virus interaction, applying selective pressure on the host once again, resulting in a cycle of coevolution. Note that the 'direction' of the arms race can be reversed, such that host antiviral effectors 'chase' their viral targets. (*b*) The age of host genetic adaptations, and consequently of the causal pathogenic viruses, can be estimated when superimposed onto the primate species phylogeny. For instance, the origins of positive selection on primate SAMHD1 appear to coincide with the birth of SAMHD1-degrading activity specific to some lentiviruses. (*c*) In cases where a particular restriction factor can be generally antagonized by a variety of different antagonists, information about binding sites can nevertheless delineate those amino acid changes that occurred on account of lentiviruses rather than due to antagonists encoded by other viruses (grey shading). (*d*) By tracing these lentiviral-specific changes on the primate phylogeny, one can infer whether one or several branches in primate phylogeny underwent

selection on account of pathogenic lentiviruses. For instance, despite being antagonized by several viruses, changes in APOBEC3G required to evade lentiviral Vif can be used to calibrate Vif-escaping adaptive episodes in primate APOBEC3G. Reprinted with permission from (20).

Classically defined restriction factors: rapidly evolving and have viral targets

Tetherin

For the purposes of this review, I am defining “classic” restriction factors against HIV as those that are both under positive selection and also have viral targets. There are several examples of restriction factors that act against HIV and have viral antagonists. For example, HIV-1 encodes a viral protein called Vpu to antagonize Tetherin. Tetherin uses its N-terminal transmembrane domain and a C-terminal glycosylphosphatidylinositol anchor to attach to the plasma membrane and simultaneously incorporate in HIV-1 particle membranes to physically tether virions to the plasma membrane (18, 21–24). HIV-1 Vpu prevents Tetherin incorporation into virions, downregulating the surface level expression of Tetherin, targeting it to the early endosome, and ultimately promoting its degradation (21, 22, 25). Although HIV-1, SIVcpz, SIVgor, SIVmus, encode a *vpu* gene, only HIV-1 Vpu can degrade Tetherin. SIVcpz, SIVgor, and SIVmus instead use Nef to antagonize Tetherin. Interestingly, the majority of primate lentiviruses, such as SIVmac and SIVsm, do not have a *vpu* gene (26, 27). In the absence of a *vpu* gene, SIV from Old World monkeys use Nef to counteract Tetherin (26, 27). HIV-2, which also does not encode Vpu, instead uses Env to antagonize Tetherin (28, 29). While SIVcpz Nef and SIVgor Nef can potently antagonize chimpanzee and gorilla Tetherin, these proteins cannot antagonize human Tetherin (27, 30, 31). Conversely, Vpu from SIVcpz and SIVgor are unable to antagonize human Tetherin, while this function has been gained by HIV-1 (27, 30, 31).

Indeed, HIV-1 Nef is unable to restrict human Tetherin (32). This species-specific activity maps to five amino acid sequence that is in the cytoplasmic domain of Tetherin and these sites are under strong positive selection in hominoids and Old World monkeys. Interestingly, this sequence is deleted in the human Tetherin, suggesting a viral pathogen drove human Tetherin to select for proteins with a deletion in this five amino acid sequence to avoid a viral antagonist. Furthermore, a cross-species transmission of HIV-1 to humans necessitated a gain of function by Vpu through adaptations in two regions within the N-terminal transmembrane domain of Vpu to replace the function of Nef.

SAMHD1

Other restriction factors have lentiviral antagonists, but interestingly not one encoded by HIV-1. For example, Human sterile alpha motif and HD-domain containing protein 1, SAMHD1, is a restriction factor that inhibits the process of reverse transcription of viral RNA through depleting the concentration of dNTPs. SAMHD1 is primarily active in non-dividing CD4+ T cells, dendritic cells, and macrophages (33). The lentiviral antagonist of SAMHD1 is either Vpx or the paralog Vpr (34). Interestingly, only two major lineages of lentiviruses encode *vpx*, whereas all encode *vpr*. In lineages that encode both *vpx* and *vpr* genes, the Vpx protein is used for SAMHD1 antagonism, whereas in lentiviral lineages that do not encode vpx, the Vpr protein sometimes functions to degrade SAMHD1 (35). HIV-1 does not have encode an antagonist of SAMHD1 (33, 35). Neither HIV-1 nor SIVcpz encodes a *vpx* gene and their respective Vpr proteins do not degrade SAMHD1 due to a deletion that occurred during the

generation of SIVcpz (36). The sites of positive selection in SAMHD1 are found in both the N- and C-terminal domain. These sites directly correlate with the phylogenetic separation of Vpx, with the SIVsmm/SIVmac/HIV-2 Vpx proteins interacting with the C-terminus of SAMHD1, while the SIVmnd2/SIVrcm Vpx proteins interacting with the N-terminus of SAMHD1 (33, 37–39). To determine if the sites under positive selection are responsible for the species-specificity of SAMHD1 antagonism by Vpx, amino acids in sites that are under positive selection can be swapped and tested to explain lentiviral evolution. For example, SIVmnd2 Vpx was challenged against an AGM SAMHD1 that was altered in the sites of positive selection to the amino acids found in mandrill SAMHD1. SIVmnd2 Vpx now had increased susceptibility to degrade SIVmnd2 Vpx (35). Thus, changes at the positively selected residues in SAMHD1 influences species-specificity. Together, the evolution between Vpx/Vpr and SAMHD1 is dynamic, with virus and host sides evolving to counteract each other through an evolvable interface.

TRIMs

TRIM5α (tripartite motif-containing protein 5 alpha) is a restriction factor against lentiviruses (and some other retroviruses) in which the positive selection directly correlates with binding of viral proteins, but without viral antagonism. TRIM5α is an interferon induced restriction factor (40, 41) that intercepts HIV-1 replication by recognizes the incoming viral capsid and causing premature uncoating (42). In this case, the arms race is set up by the escape of lentiviruses from TRIM5α by mutations in CA, rather than the direct antagonism of TRIM5. Human TRIM5α can block some retroviruses, but weakly inhibits HIV-1 (43, 44). However, TRIM5α from rhesus

macaque and other primates can potently inhibit HIV-1 (43, 45–48). Indeed, an *in vivo* study revealed that rhesus TRIM5α can potently inhibit SIV from sooty mangabey with 100-fold to 1,000-fold viral attenuation and suppress cross-species transmission (49). Trim5α is under positive selection and species specificity of TRIM5α for retroviruses can be altered with amino acid changes in the coil-coiled/ B30.2 protein domain that corresponds to amino acids that are under positive selection (50–52).

Cyclophilin A, referred to as CypA, interacts with Gag, is incorporated into virions, and was once thought to function as a foldase or protein chaperone during viral uncoating (53). However, more recent data suggests that its role is to protect the capsid core from TRIM5-mediated rapid uncoating. Additionally, the role of CypA itself in the virus life cycle is not completely resolved. Interestingly, only viruses from the HIV-1/SIVcpz lineage require CypA for replication (53). Nonetheless, primate genomes have also evolved a recurrent gene fusion between TRIM5 and CypA, called *TRIMcyp,* as an additional mechanism to counter new viral challenges. At least four times in primate evolutionary history additional innovation in the *TRIM5* locus has been sampled with TRIM-cyclophilin A chimeric proteins, which arose via the capture of a new exon (54–60). In this novel protein, cyclophilin A has replaced the B30.2 protein domain that was the antiviral specificity determinant. Through reconstructing evolutionary history and phylogenetic analysis, an ancient *TRIMCypA3* gene was discovered that has potent antiviral activity (61). This ancient gene could have arisen in response to a lentiviral pathogen encountered by primate ancestors but since the divergence of all simian primates, TRIMCypA3 has lost specificity and decayed in many extant sequences. Similar to TRIM5α, TRIM-Cyp does not have a lentiviral antagonist, but rather viral

escape from TRIM-CypA can occur because not all lentiviruses bind CypA. Thus, selection for mutations that disrupt the TRIM-Cyp interaction are another form of viral escape (62, 63).

TRIM5α belongs to a large family of proteins, with over 70 members in the Human genome, with a variety of functions. TRIM34, located adjacent to TRIM5α on the human gene locus, was recently characterized as a restriction factor to inhibit HIV-1 and SIV capsids (64). Interestingly, there is no evidence of positive selection in TRIM34 (65, 66). However, the restriction of TRIM34 on HIV-1 has been shown to be dependent on TRIM5α, which is under strong positive selection (51, 64). Since TRIM34 and TRIM5α interact and colocalize, perhaps TRIM34 is not under positive selection because its binding partner, TRIM5α, is the protein that is the target of host and viral selection (64, 67).

APOBEC3s

The apolipoprotein B mRNA editing enzyme catalytic-polypeptide like 3, APOBEC3 (shortened to A3 here) are a family of cytidine deaminases that hypermutate ssDNA. The *APOBEC3 (A3)* gene locus in primates encodes seven cytidine deaminase proteins (A3A-A3H), all with a conserved zinc coordinating motif His-X-Glu-X_{23-28}-Pro-Cys-X_{2-4}-Cys, that inhibit endogenous retroelements, like LINE-1, and retroviruses, such as HIV-1, among many other viruses. The antiviral activity of the A3 proteins is antagonized by the lentiviral accessory protein Vif.

A3s were first discovered because of a phenotype observed such that cell lines that were lacking a *vif* gene, termed nonpermissive cells, were not infectable by HIV

(68, 69). However, other cell lines, called permissive cells, could sustain full viral infection (70, 71). Through a series of fusion experiments, a dominant factor was identified as APOBEC3G (A3G) that was controlling the viral replication in the nonpermissive cells (72). Since then, A3s have been extensively studied as a potent restriction factor of HIV-1.

In order to inhibit HIV, A3s must get incorporated into budding HIV-1 virions in the producer cells. Once the virion with A3 packaged fuses to a target cell, A3s can act to inhibit HIV-1 by hypermutating ssDNA during the reverse transcription process. A3s convert cytidines to uridines on the minus strand, which in turn results in a G-to-A hypermutation on the positive sense strand. A3G is unique since it preferentially deaminate the 3' cytidine in a 5'-CC<u>C</u>-3' or 5'-C<u>C</u>-3' sequence on the newly synthesized minus strand. However, other antiviral A3s, such as A3C, A3D, A3F, and A3H, deaminate cytidines in a 5'-TT<u>C</u>-3' or 5'T<u>C</u>-3' motif. Furthermore, A3s tend to deaminate more heavily in the 5' region of the viral genome, perhaps because this single stranded region is exposed longer than the 3' end and/or because of the intrinsic biochemical properties of these A3s scanning ssDNA (73, 74). A3G scans processively ssDNA by a combination of primarily sliding and occasional jumping motions to reach the target cytidines (74). A3F, however, primarily uses jumping movements to scan ssDNA, which in turn leads to less efficient and less extensive deamination sites during reverse transcription (75). Together, these mutations can ultimately render the provirus inactive. Additionally, these A3s can inhibit HIV-1 in a cytidine deaminase independent mechanism, although this inhibition only accounts for less than 10% of the antiviral activity (76–79). Recent work has shown that A3G can restrict HIV-1 in a cytidine

deaminase independent mechanism through binding to reverse transcriptase and sterically hindering reverse transcription (76).

HIV-1 has overcome this barrier by devoting an accessory protein, Vif, to counteract the APOBEC3s. Vif employs multiple strategies to counteract A3 proteins, but the most well-established mechanism involves hijacking a cellular Cullin-RING ubiquitin ligase, resulting in ubiquitination and subsequent targeting of A3 for proteasomal degradation (72, 80–84). HIV-1 Vif is able to antagonize multiple A3 proteins including A3C, A3D, A3F, A3G, and A3H. To induce A3 degradation, Vif binds to CBF-β and recruits CUL5-EloB/C-E3 ubiquitin ligase complex to promote polyubiquitination and the subsequent proteasomal degradation of the A3s (85). Vif uses three distinct interfaces to bind to A3s: one for A3G, another for A3C/A3D/A3F, and another for A3H (86). Evolving an antagonist like Vif to target A3G must have been essential for lentiviruses, since almost all lentiviruses encode a Vif protein except equine infectious anaemia virus (8, 87). Furthermore, the interaction between Vif and A3 proteins is often species specific and the key determinant of interspecies transmission events such as the zoonotic transmission that resulted in the SIVcpz and HIV-1 (20).

The best evidence that a lentiviral antagonist is responsible for some of the positive selection seen in a restriction factor comes from the study of the evolution of A3G and Vif in different Old World monkeys and their endemic SIVs. As a result of antagonism-evolution, A3G has been adaptively diversifying within AGMs (88). Single amino acid substitutions that allow evasion of Vif-mediated degradation independently emerged on several occasions during primate evolution, the oldest of which appeared at

least 5 Mya in the common ancestor of the guenon species. An even older Vif-blocking adaptation in A3G, a multi-residue insertion event, arose at least 12 Mya in a subset of OWMs (*Colobinae*) (89). These examples show interspecies divergences, presumably fixed mutations that distinguish the antiviral gene of one species from another. However, selection can also occur as allele variation, known as polymorphism within a species. These polymorphisms provide powerful tools to reveal ongoing or more recent selective pressures.

The natural SIV infection that occurs in African green monkeys (AGMs) provides a unique opportunity to assess how host antiviral genes impact the evolution of lentiviruses, and vice versa, on a relatively recent timescale since AGMs have been diverging for approximately four million years (20). AGMs are the most abundant nonhuman primates in Africa that are comprised of four different subspecies, commonly known as grivet, sabaeus, vervet, and tantalus monkey (90). Each subspecies is infected with a distinct subspecies of SIV, named after the specific host it infects: SIVgri, SIVsab, SIVver, and SIVtan (91). In two of those subspecies, single amino acid changes in A3G emerged independently to allow evasion of the lentiviral antagonist, Vif. Specifically, these adaptive mutations prevent degradation by Vif from heterologous SIVagm isolates, suggesting that they are selected as a result of Vif-mediated selective pressure (88).

To better understand prospectively the dynamics of viral adaptation, plasma virema from infections of sabaeus monkeys with an SIVver strain was characterized. Vif readily counteracted A3G in individuals homozygous for the ancestral A3G allele. In animals that were homozygous for the A3G (D130H) allele that was resistant to Vif from

the initial virus inoculum, the *vif* gene evolved within the animal in order to counter this A3G polymorphism. However, in a sabaeus monkey that was heterozygous, encoding both an ancestral A3G allele and a D130H allele, SIVver Vif failed to evolve the capacity to antagonize both variants *in vivo* (88). Interestingly, this gain of antagonism to the A3G D130H mutation came at a cost to Vif, because as a result, Vif lost activity to the ancestral A3G (88). These studies demonstrate that A3G-driven evolution of Vif is a natural occurring phenomenon because the specificity of Vif proteins from the SIVagm subtypes reflects adaptation to subspecies-specific A3G variants in AGM populations exposed to SIV in the wild.

Furthermore, structural analysis of different species of Vif protein can provide insight to lentiviral evolution and cross-species transmissions. SIVrcm Vif, the precursor to SIVcpz/HIV-1 lineage, poorly antagonizes chimpanzee A3D and A3G proteins (36, 92). Adaptations to SIVrcm Vif were essential for the cross-species transmission of SIV from OWM into chimpanzees. Through comparing the HIV-1 Vif structure to the SIVrcm Vif structure, key similarities and differences were identified that contribute to the Vif substrate specificity(93). For example, a single amino acid change, Y86H, in SIVrcm Vif conferred specificity and adaptation towards cpzA3G and huA3G (93). This adaptation at position Y86 allowed for SIVrcm Vif to gain the ability to antagonize hominid primate A3Gs and to at least partially accommodate a change in A3G at position 128 (93).

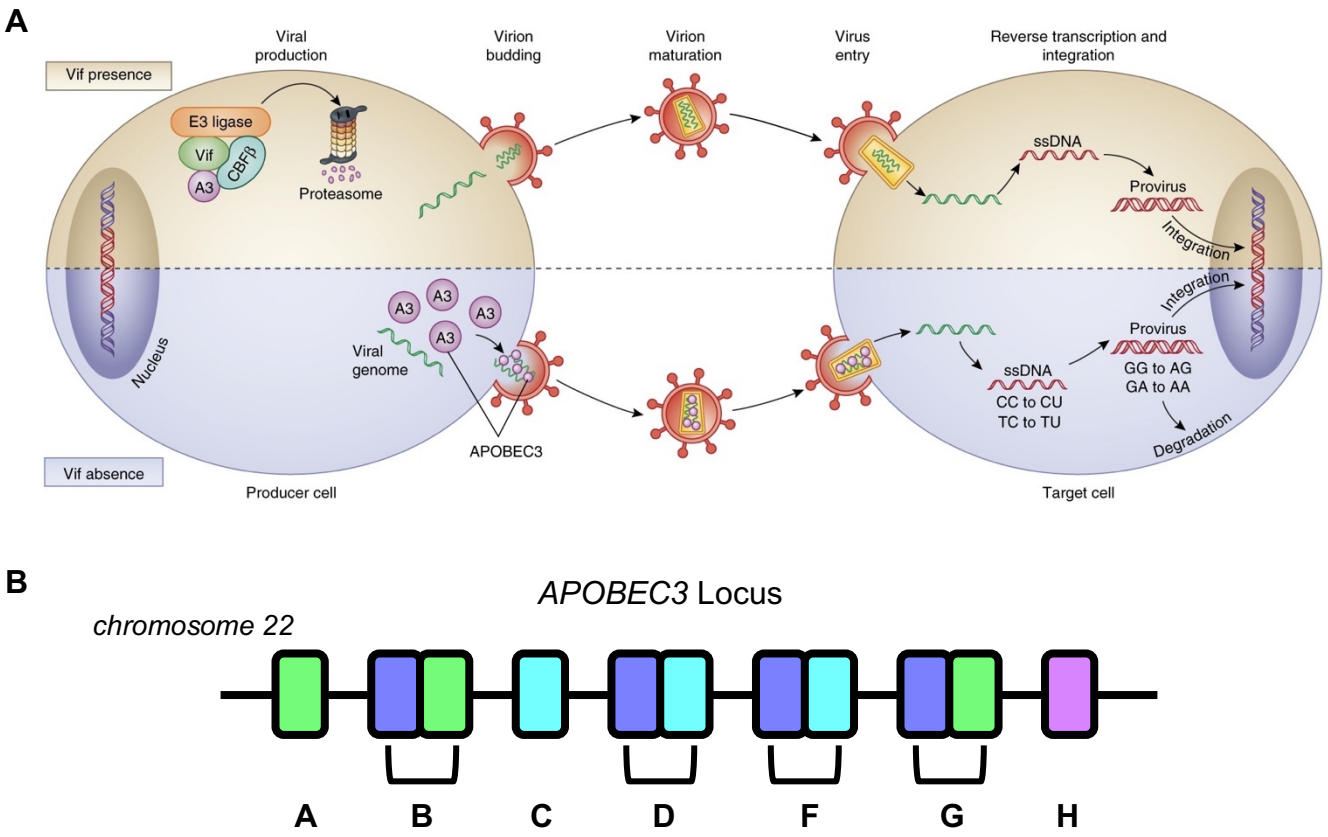


Figure 3. Overview of APOBEC3 mode of restriction against HIV and gene locus

(**A**) APOBEC3 proteins are degraded in the producer cell in the presence of Vif, but in the absence of Vif they are packaged into the budding viral particle. In the next cycle of infection, the APOBEC3 proteins mutagenize the viral genome during reverse transcription by deaminating cytosines to uracils in the minus-strand DNA. Mutated viral DNA may be degraded by DNA repair enzymes or integrated into the host cell genome. Reprinted with permission from (94) under license number 4917781351844. (**B**) Organization of the *A3* locus on chromosome 22. Colors indicate the evolutionary history of the corresponding zinc domain according to (95).

OTHER APOBEC3s

In addition to A3G, there are six other A3 proteins in the human genome that arose due to recombination and duplication to give rise to three single deaminase domain A3s (A3A, A3C, and A3H) and four double deaminase domain A3s (A3B, A3D, A3F, and A3G) (95). The duplication events of the *A3* locus occurred before the expansion and diversification of placental mammals, which is roughly between 95-120

Mya (95). Furthermore, these A3s can be categorized by the presence of a A3Z1, A3Z2, or A3Z3 domain based on sequence conservation (96). For example, an example of a A3Z1 is A3A, A3Z2 is A3C, and A3Z3 is A3H. The A3Z3 domain is the most evolutionary diverged A3 sequence. Not only do these A3s vary in the number of domains they have but they also vary in their ability to inhibit HIV-1.

Five of the human A3s have antiviral activity against HIV-1: A3C, A3D, A3F, A3G, and A3H with A3G being the most potent (97). In order to inhibit HIV, each of these A3s need to be expressed in CD4+ T lymphocytes and get packaged into budding HIV virions. A3C, A3D, A3F, A3G, and A3H are expressed in CD4+ T lymphocytes, while A3A and A3B are not, explaining why A3A and A3B do not inhibit HIV-1. Furthermore, A3D, A3F, A3G, A3H, and less so A3C, get packaged into virions, while A3A and A3B do not, again supporting the conclusion that only some A3s are antiviral (97, 98). A3D is interesting because human A3D is not as antiviral as its homolog in chimpanzees which corresponds to additional packaging of the chimpanzee version of A3D compared to human A3D (99).

A3s can also have one of three different subcellular localizations: nuclear, cytoplasmic, or cell wide. A3A, which is not packaged, has a cell-wide localization. However, when A3A is linked to Vpr, a viral protein that is efficiently packaged into virions, Vpr-A3A is packaged, suggesting that cellular localization could be an important determinate for packaging (100). A3B has a nuclear localization (97). A3C has a cell-wide localization and is not packaged very abundantly (97, 98). A3D, A3F, and A3G all have primarily cytoplasmic localizations (97). A3H has different haplotypes circulating in human populations that have different localization patterns. For example, A3H

haplotype I is more nuclear, less stable, packaged less, and therefore, less antiviral against HIV-1 (101–103). On the other hand, A3H haplotype II is more cytoplasmic, more stable, packaged more, and thus, more antiviral than A3H haplotype I (101–103). While it is known that antiviral A3s get packaged into virions, the exact mechanism for A3 packaging remains unknown. For A3G, it has been shown that A3G exists in cells in high molecular mass complexes that is bound to RNA and other proteins in stress granule and P-bodies. This form of A3G remains enzymatically inactive. However, the newly synthesized A3G binds to RNA that is in the same complex as HIV-1 gag and therefore, gets packaged into virions (104, 105). A3s are constitutively expressed in cells and have rather nonspecific substrate, ssDNA. Therefore, A3s that have nuclear localizations, such as A3B and A3H haplotype I, have also been associated with many cancers (106, 107). These oncological findings also suggest that there is a balance to A3 activity and too much activity could be detrimental to the host.

Like A3G, many of the other A3 genes have also been driven by positive selection in primates including A3A (108), A3C (98), A3D (99), and A3H (103). In the case of A3C and A3H, there is some evidence that the amino acids under positive selection also correspond to sites of interaction of Vif (98, 109), while for A3D the sites of positive selection overlap with amino acids that led to increase antiviral activity of the restriction factor in chimpanzees (99). However, there is considerably more positive selection in all the A3 genes than can be explained by interactions with Vif alone, which suggests that each of these A3 proteins has also been engaged in arms-races with other pathogens over evolutionary time in primates (110).

In addition to the multiple *A3* genes in humans, there is additional diversity in the *A3* locus in the form of human polymorphisms that affect the antiviral activity against HIV-1. For instance, there are at least twelve different haplotypes of A3H that vary in their ability to restrict HIV-1 (111, 112). The reconstructed sequence of the A3H from the common ancestor of humans and chimps encodes a stable and nuclear protein (101). However, the four major human A3H haplotypes (hap I-IV) differ in protein stability, which affects steady-state expression levels and subcellular localization, but none are both nuclear and stable (101, 102). Most humans possess the unstable A3H haplotype I, with frequencies ranging from 49%-96% depending on the population (101). However, a fraction of the population expresses a protein encoded by the stable, cytoplasmic A3H haplotype II (101, 102). In addition, the most common form of human A3C, which encodes for a serine at position 188, has little, if any, activity against HIV-1. However, about 10% of African individuals carry a polymorphism in A3C that encodes an isoleucine at position 188, instead of a serine. This one amino acid change is correlated with increased antiviral activity against HIV-1 resulting from an increased ability of A3C to form dimers and increased cytidine deaminase activity *in vitro* (98). Furthermore, chimpanzee and gorilla A3Cs have different amino acids at position 115 than humans and introducing these amino acids into human A3C also increases dimerization and cytidine deaminase activity (113). Nonetheless, even with these mutations, A3C is less active against HIV-1 compared to many of the other A3 proteins (114).

Restriction factors under positive selection with no viral target

MxB

Although in the examples above, some restriction factors against lentiviruses have been evolutionarily driven by ancient interactions with lentiviruses, this is not the case for all restriction factors against HIV-1. For example, MxB is an interferon induced restriction factor that has been shown to suppress infection of a wide range of HIV-1 strains (115–117). In an HIV-1 CRISPR screen designed to reveal the restriction factors that inhibit a lab-adapted strain to HIV-1, MxB was identified as a dominant mediator of IFN induced inhibition of HIV-1 infection (118). MxB inhibits HIV-1 infection by binding to HIV-1 capsids and inhibiting their entry into the nucleus.

While MxB has evolved under positive selection in primates, there is no overlap between the residues identified in the positive selection analysis and the residues identified for lentivirus restriction (119). Moreover, in contrast to what one would expect by a restriction factor driven by positive selection by lentiviruses, MxB from many different species, human, rhesus macaque, and African green monkey all potently inhibit HIV-1 unlike the case seen with TRIM5 or APOBEC3G (120). Thus, despite the potent block MxB imposes, there is no viral antagonist to counteract MxB and no evidence that escape of lentiviral capsids from MxB has driven positive selection (in contrast, for example, to TRIM5). One possible explanation is that perhaps HIV-1 cannot escape the pressures of MxB because many mutations would be necessary to alter the binding interface and these would ultimately disrupting the structural integrity of capsid (121). Alternatively, because MxB only can restrict HIV replication 5-10-fold, possibly it is not important enough selective pressure to target by HIV to make a

difference for viral replication (115–117). The sites under positive selection in MxB do not correspond to species specific lentiviral escape, suggesting that MxB evolution is not driven by lentiviruses. However, in an experiment designed to better understand HIV-1 infection in pigtail macaques, monkeys were infected with a simian tropic HIV-1 and viral sequences were analyzed for adaptation. Interestingly, adaptation in the viral sequence were found in CA, which was surprising given that these pig tail macaques do not encode a TRIM5 and CA has extreme genetic fragility (59, 121). Further experiments showed that there was partial resistant to MxB, suggesting that MxB could have driven the acquisition of these changes. However, sequences isolated with full resistant to MxB had a replication fitness cost, providing some insights as to why lentiviruses might not have evolved a viral target to counteract MxB (122).

ZAP/Schlafen11

Schlafen11, a interferon-induced restriction factor, inhibits viral replication by recognizing the differential codon usage between host and retroviral transcripts and blocking translation (123). Interestingly, this restriction factor is another example where HIV-1 has not evolved an antagonist against Schlafen11, possibly because human Schlafen11 only weakly inhibits HIV-1 (124). Similar to MxB, Schlafen11 is under positive selection, however the resides determined undergoing rapid evolution have nothing to do with anti-lentivirus activity, suggesting that another pathogen is driving this host-arms race (124). Perhaps, also similar to MxB, Schlafen11 does not pose an important enough block, and therefore lentiviruses have not bothered with evolving a counteract.

Zinc finger antiviral protein, ZAP, is another restriction factor that has potent activity against alphaviruses and filoviruses as well as moderate activity against retroviruses (125–130). ZAP blocks viral replication by degrading viral RNA with high CG dinucleotide content (128, 131, 132). ZAP alone does not have any enzymatic activity, but instead recruits other cellular proteins to inhibit viral replication, such as TRIM25 and KHNYN (128, 133). There are two human ZAP isoforms, ZAP-L and ZAP-S. Both isoforms contain a N-terminal RNA binding domain containing four CCCH0type zinc finger motifs, but ZAP-L also contains a catalytically inactive C-terminal poly(ADP-ribose) polymerase (PARP)-like domain and a functional CaaX prenylation motif (127, 134). ZAP-S functions as a negative feedback regulator of the interferon response through binding to and mediating the degradation of several host interferon messenger RNA (135). In contrast, the long isoform is the primary antiviral factor. The sites of positive selection are only found in the catalytically inactive C-terminal PARP domain of ZAP-L (127). If a viral protein were to prevent ZAP-L function by binding the N-terminal domain, it would also disrupt the activity of ZAP-S and increase the IFN response(135). This example of isoform-specific positive selection could be explained by ZAP-S acting as an evolutionary shield to protect the remainder of the protein from antagonism. Rather than dedicating an accessory protein to antagonize ZAP, retroviruses have evolved to evade ZAP by modulating the CG dinucleotide content of transcripts (136). It has been observed that HIV-1 and SIVcpz have fewer CpG dinucleotides than the precursor virus, SIVmon/gsn/mus, which could suggest than ZAP is the driving force for lowering the CG content of HIV-1 transcripts (136). However, HIV-2 has higher CG

content than its precursor SIVsmm, suggesting that most likely ZAP does not drive lentivirus evolution (136).

Restriction factors with no positive selection

IFITMs and SERINC3/5 and HUSH

There are restriction factors that have viral targets, but nonetheless, show no signs of genetic conflict with positive selection. IFITMs, specifically IFITM1, IFITM2, and IFITM3, are a family of interferon-induced transmembrane proteins that broadly inhibit many viruses (15, 17). In the context of lentiviruses, IFITM proteins have been shown to be incorporated into virions and inhibit viral entry (16, 137). In addition, IFITM proteins also impact late stages of the HIV-1 life cycle by reducing lentivirus envelope protein abundance (138). IFITMs has been shown to be under purifying selection, suggesting that the function of the IFITMs is conserved and perhaps related to the fact that IFITMs have roles in other cellular processes like embryonic development and cancer (17, 139). However, rather than containing the classical rapid evolution hallmarks of antiviral proteins, perhaps IFITMs have gotten around the viral block by diversification through gene duplication (IFITM1, IFITM2, and IFITM3) and cellular localization (140). This strategy which has also been shown in other antiviral proteins, such as IFIT, could lead to a boost in antiviral activity against broad viral targets (140, 141). On the other hand, viruses have evolved mechanisms to evade the entry blocks that IFITMs pose by mutations in envelope as well as having multiple routes of cellular entry (138).

The transmembrane proteins SERINC3/5 are restriction factors that are incorporated into virions to prevent HIV-1 viral entry (142, 143). HIV-1 Nef

downregulates the expression of SERINC3/5 from the cell surface to prevent incorporation into HIV-1 virions (142, 144). Interestingly, SERINC3/5 are not under positive selection in primates, perhaps because similar to IFITMs the cellular function of SERINCs is limiting the ability of these proteins to evolve (145). On the other hand, escape from SERINC3/5 can also occur through mutations in HIV-1 env (146), so it is possible that there is little pressure on SERINC5 because escape is too easy. Curiously though, other retroviruses, such as MLV also encode an antagonist of SERINC3/5, the glycoGag protein (147), yet again, there is no evidence that this antagonist has driven selection of SERINC3/5 in their natural host.

Vpx has another restriction factor target, Human silencing hub (HUSH) complex (148, 149). The HUSH complex functions to help impose and maintain HIV transcriptional silencing. Interestingly, despite HUSH complex being counteracted by Vpx in a species-specific manner, it is not under strong positive selection. While there have been no detectable strong signatures of evolutionary conflicts in the HUSH complex, proteins within the HUSH complex, such as TASOR, SETDB1, and periphilin, have signatures of other genetic innovations such as transcriptional variations and multiple isoforms (149). HUSH is involved in many other transcription programs, however, such as silencing endogenous retroviruses and retroelements (150, 151), so it is possible that the host would pay too high of a fitness cost to protect its genome if the specificity of these factors were to evolve.

Super Restriction Factors

Despite hosts encoding a cascade of restriction factors to inhibit lentiviral replication, lentiviruses encode means to counteract these blocks through antagonists and sequence escape. Therefore, my thesis work sought to create super restriction factors to improve the naturally found proteins. Super restriction factors are defined as evolution-guided variants of natural restriction factors that are more active and more resistant to viral antagonists (152, 153). Previous work on MxA, another broadly acting restriction factor that works to counteract both positive sense and negative sense RNA viruses like orthomyxoviruses, bunyaviruses, paramyxoviruses, picornaviruses, rhabdoviruses and togaviruses, has revealed that the antiviral activity of MxA can be improved through altering the five amino acids that are under positive selection (154). However, this work showed that broadly acting antiviral proteins like MxA skirt breadth-versus-specificity trade-offs that constrain their adaptations against multiple viruses (152). Here, I wanted to make an HIV restriction factor with increased potency, but also resistance to viral antagonism. My graduate dissertation focuses on the antiviral protein APOBEC3 and the viral antagonist HIV-1 Vif. Using an evolution-guided approach, the A3 super restriction factors are useful tools to prospectively study evolution for a protein that has not been sampled in nature. Additionally, these proteins have provided insights about A3 mechanisms to restrict HIV-1 and ultimately the evolutionary compromises for paths that have yet to be sampled in nature.

Chapter 2. Materials and Methods

APOBEC3 Tandem Deaminase Domain Sequences

The A3C-A3C synthetic tandem domain was designed after the most closely related A3, A3F. The N-terminal subunit consists of amino acid residues 1-187 of A3C (residue numbers 188-190 were deleted) and was codon optimized to distinguish the N-terminus from the C-terminus. The codon-optimized N-terminus was linked to the single domain human A3C at the C-terminus using a short amino acid sequence that is found in A3F and A3D, Arg-Asn-Pro (RNP), and naturally found in A3C single domain. The C-terminal domain begins at amino acid 6 (15 nucleotide deletion) to include this RNP sequence and extends to include the remainder of A3C followed by a flexible linker and a triple FLAG tag. The A3C-A3C tandem domain was constructed using overlapping extension PCR, as described in (98) and cloned into a pcDNA4/TO vector (ThermoFisher, V102020) using BamHI/XbaI restriction sites. All point mutations were made using Site Directed Mutagenesis QuikChange II Kit (Agilent; 200524).

Similar to A3C-A3C, A3C/A3H double domains were designed using A3C and A3H haplotype I, connected with an Arg-Asn-Pro (RNP) linker sequence, found in A3F and generated via gene synthesis (IDT) for both A3C-A3H hap I and A3H hap I-A3C. The constructs were cloned into the pCDNA4/TO backbone with a C-terminal 3XFLAG tag using restriction sites at EcoRI/XhoI. To create the A3C/A3H hap II variants, Site Directed Mutagenesis (Agilent; 200524) PCR introduced G105R, K121D, and E178D to convert A3H haplotype I to haplotype II.

Cell Culture and Transfections

HEK293T cells (ATCC, CRL-3216) were cultured in DMEM (Gibco, #11965092) with 10% HyClone Fetal Bovine Serum (GE Healthcare, #SH30910.03) and 1% penicillin-streptomycin (Gibco; #15140122) at 37°C in a humidified CO_2 incubator. SUPT1 cells, acquired from ATCC (CRL-1942), were maintained similarly but in RPMI medium (Gibco, #11875093), 10% HyClone Fetal Bovine Serum (GE Healthcare, #SH30910.03), 1% penicillin-streptomycin (Gibco, #15140122), and 10mM HEPES. HEK293T cells were plated in 6-well dishes for transfections at a density of $1.5x10^5$ cells per 1mL. Transfections were performed with the TransIT-LT1 transfection reagent (Mirus, #MIR2304) at a reagent:plasmid DNA ratio of 3:1.

Intracellular Protein Expression and Packaging Experiments

Intracellular expression of APOBEC3 proteins during virion production was determined by lysis of the virion-producing 293T cells with NP-40 buffer with protease inhibitor (200mM NaCl, 50mM Tris pH 7.4, 0.5% NP-40 Alternative, 1mM DTT, and Roche Complete Mini, EDTA-free tablets; 11836170001). Lysate samples were resolved on an 4-12% SDS-PAGE gel using MES buffer and transferred to a nitrocellulose membrane for western blot analysis. Anti-FLAG (Sigma; F3164), anti-tubulin (Sigma; T6199), and anti-p24gag (NIH-ARP; 3537) (155, 156) antibodies were used for western blots at a dilution of 1:5,000. StarBright Blue 700 Goat Anti-Mouse IgG (BIO-RAD, 12005866) were used to detect primary antibodies at a dilution of 1:5,000. The chemiluminescent signals from all western blots were imaged using a

ChemiDocMPImaging System (Bio-Rad) and images were processed using Fiji/ImageJ software to quantify the densitometry for each antibody detected band.

The amount of A3 packaged into virions was evaluated by co-transfecting either 100ng or 600ng pcDNA4/TO.A3.3XFLAG and 1000ng HIV-1ΔVifΔEnvLuc2 (unless denoted) in a 6-well plate. Three days post-transfection, cell lysates were harvested as described above and 1.5mL of the supernatant was collected, filtered through a 0.2 micron filter, and spun down using a tabletop microcentrifuge for 1 hour at max speed at 4°C to pellet the virions. The supernatant was aspirated off and 25uL of NuPAGE 4X loading dye (Invitrogen, #NP0007) was added to each sample. Samples were boiled for 10min at 95°C and loaded on an SDS-PAGE gel.

Single-cycle Infectivity Assay

Single-cycle infectivity assays using HIV-1 were previously described (109). 293T cells were seeded at a density of $1.5x10^5$ cells/mL in a 6-well plate. The following day, cells were transfected with 100ng or 600ng provirus (HIV-1ΔVifΔEnvLuc2, unless otherwise noted), 100ng L-VSV-G, and 400ng pcDNA4/TO.A3.3XFLAG or pcDNA4/TOPO empty vector unless otherwise indicated. 72 hours later, virus was harvested and normalized for virion production using an RT-qPCR assay, as described (157, 158). A volume of virus equivalent to 2000mU/mL of RT were used for infection of SUPT1 cells. For infectivity assays, SUPT1 cells were seeded at $2x10^4$ cells per well in a 96-well plate in media containing 20 ug/mL DEAE-dextran. 72 hours later, infected cells were lysed in luciferase lysis reagent (Bright-Glo, Promega #E2610) and luciferase expression was measured on a luminometer (LUMISTAR Omega, BMG Labtech).

Infectivity of each virus was normalized to 100% based on a No A3 control. All HIV-1 constructs are based on the LAI strain. A clade B patient derived Vif (PID:1203) that is able to antagonize A3H haplotype II was previously described (109).

Deep Sequencing of A3 Mediated Mutations

To analyze A3 mediated mutations, 293T cells were seeded at a density of $1.5x10^5$ cells/mL in a 6-well plate. The following day, cells were transfected with 600ng provirus (HIV-1ΔVifΔEnvLuc2), 100ng L-VSV-G, and 400ng pcDNA4/TO.A3.3XFLAG or pcDNA4/TO empty vector. Three days later, virions were harvested and quantified similar to single-cycle infectivity assay mentioned above. $2x10^6$ SUPT1 cells were infected with 500uL of each Benzonase-treated virus with and without a 10mM nevirapine control and spinoculated at 1,100 x g for 30min at 30°C. Twelve hours later, unintegrated viral cDNA was isolated using Qiagen Miniprep Kit (QIAprep Spin Miniprep Kit, 27106).

To determine A3 mediated mutations, we used a barcoded Illumina deep sequencing approach previously described (159) and explained in more detail at https://jbloomlab.github.io/dms_tools2/bcsubamp.html. This approach attaches unique molecular identifiers (8*N) to each DNA molecule, which enables increased correction of PCR and sequencing errors by determining molecule-specific consensus sequences. A region of *pol* (92, 160) was amplified using KOD Hot Start Master Mix (Millipore, 71842) in a first round of PCR with forward primer: CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGACAAGGAACTGTATCCTTTAACTT and reverse primer:

GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTGGTACAGTTTCAATAGGA CTAAT. The first round PCR was carried out using the following parameters: 95°C for 2min and 20 cycles of 95°C for 20sec, 70°C for 1sec, 50°C for 10sec, 70°C for 2 min, followed by a 95°C for 1min and a hold at 4°C. PCR products were cleaned up using AMPure beads (Beckman Coulter, A63880) and quantified with Qubit dsDNA HS Assay Kit (ThermoFisher, Q32854). Next, round 1 PCR products were bottlenecked such that each uniquely barcoded ssDNA molecules would be read ~2.9 times during sequencing. A second round of PCR performed with KOD Hot Start Master Mix (Millipore, 71842) supplemented with 1mM $MgCl_2$ under the following parameters: 95°C for 2min and 23 cycles of 95°C for 20sec, 70°C for 1sec, 60°C for 10sec, 70°C for 10sec, followed by a hold at 4°C. Samples were again cleaned up, quantified, pooled, purified via gel electrophoresis, and sequenced on an Illumina MiSeq, using 2x250 paired-end reads. Deep sequencing was carried out in technical replicate for each experimental condition, starting with the initial PCR reactions.

We used the software package dms_tools2 to align sequencing reads and build consensus sequences for each uniquely tagged DNA molecule (161). Error-corrected reads were compared to the target sequence to determine the number, identity, and surrounding nucleotides of all substitutions in each read. Reads with high numbers of substitutions (>10% of non-G nucleotides) at the junction of the two paired-end reads were removed from the analysis as these substitutions were most often found to be alignment artifacts. Since A3s are known to cause G-to-A substitutions, we initially subsampled our data to look only at G-to-A substitutions. We plotted the frequency of reads in each sample with a given number of G-to-A mutations (0, 1, 2, etc. up to 9, and

then 10+). Frequencies were calculated as the average frequency of technical replicates and read counts are shown as the sum of reads for both replicates.

To investigate if substitutions occurred more frequently in certain nucleotide contexts, we computationally extracted all three-nucleotide sequences centered on a substitution from each sample. This gave us a count for each sample of how many times each nucleotide was immediately upstream (site -1) of a substitution, underwent a substitution (site 0), or immediately downstream of a substitution (site 1). We then calculated the frequency for each nucleotide at each of these three sites in each sample. Data for technical replicates were combined by averaging the frequencies of each nucleotide at each site across the two replicates for each sample. Next, we calculated background-corrected frequencies for each nucleotide at each site using the frequencies in plasmid control sample as our background. We based this frequency correction on the 'type 2' logos from Gorodkin, et al. (162).

If $q_{ik}$is the frequency of base k at site i in the sample and $p_{ik}$is that frequency in the background, then the background corrected frequency is:

$$c_{ik} = \frac{q_{ik}/p_{ik}}{\sum_i q_{ik}/p_{ik}}$$

We then calculated the information content of each site as:

$$I_i = \log_2 N - \left(-\sum_{k \in A} c_{ik} \log_2 c_{ik} \right)$$

Where A is the set of nucleotides [A, C, G, T] and N is the number of elements in A (162, 163).

Finally, we calculated the letter height at each site as:

$$d_{ik} = c_{ik} I_i$$

The sequencing reads were uploaded to the NCBI SRA with the BioProject accession number PRJNA605864 and run accession numbers SRR11059567 to SRR11059589. The computational pipeline to analyze the sequencing data and generate Figure4 is available on Github (https://github.com/jbloomlab/SuperRestrictionFactor_Hypermutation).

qPCR Assay for HIV Late Reverse Transcription Products

For qPCR analysis of HIV late RT products, $2x10^6$ SUPT1 cells were infected with 500uL of each Benzonase-treated virus with and without a 10mM nevirapine control and spinoculated at 1,100 x g for 1hr at 30°C. 18 hours post infection, unintegrated viral cDNA was isolated using Qiagen Miniprep Kit (QIAprep Spin Miniprep Kit, 27106). HIV cDNA was amplified with TaqMan Gene Expression Master Mix (AppliedBiosystems #4369016), J1 FWD (Late RT F) – ACAAGCTAGTACCAGTTGAGCCAGATAAG, J2 REV (Late RT R) GCCGTGCGCGCTTCAGCAAGC, and LRT-P (late RT Probe) – FAM-CAGTGGCGCCCGAACAGGGA-TAMRA (164, 165). Data was acquired on an ABI QuantStudio5 Real Time (qPCR) machine and analyzed on Prism software.

Velocity Sedimentation

For analysis of A3 migration on a sucrose gradient, 293T were harvested 72 hours post-transfection (1000ng or 100ng of A3 per well in a 6-well plate) using a low salt and EDTA buffer (0.2M HEPES pH7.9, 0.1M NaCl, 0.01 $MgCl_2$, 0.002 EDTA pH8.0, 3.5% triton-X 100) (166) with protease inhibitor cocktail (Roche, 11836170001). For

each sample, 50uL of lysate was layered on a step gradient (10%, 15%, 40%, 50%, 60%, 70%, and 80% sucrose in lysis buffer) and subjected to velocity sedimentation in a Beckman MLS50 rotor at 45,000rpm (163,000 x g) for 37min at 4°C, as described previously (102). Gradients were fractioned (400μL) from top to bottom and analyzed via western blotting.

Chapter 3. APOBEC3C tandem domain proteins create super restriction factors against HIV-1

Abstract

Humans encode proteins, called restriction factors, that inhibit replication of viruses like HIV-1. One family of antiviral proteins, *apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (*APOBEC3, shortened to A3) acts by deaminating cytidines to uridines during the reverse transcription reaction of HIV-1. The *A3* locus encodes seven genes, named *A3A-A3H*. These genes either have one or two cytidine deaminase domains and several of these A3s potently restrict HIV-1. A3C, which has only a single cytidine deaminase domain, however, inhibits HIV-1 only very weakly. We tested novel double domain protein combinations by genetically linking two *A3C* genes to make a synthetic tandem domain protein. This protein functioned as a "super restriction factor" that had more potent antiviral activity than the native A3C protein, which correlated with increased packaging into virions. Furthermore, disabling one of the active sites of the synthetic tandem domain protein results in an even greater increase in the antiviral activity—recapitulating a similar evolution seen in A3F and A3G (double domain A3s that only use a single catalytically active deaminase domain). These A3C tandem domain proteins do not have an increase in mutational activity, but instead inhibit formation of reverse transcription products which correlates with their ability to form large higher order complexes in cells. Finally, the A3C-A3C super restriction factor largely escaped antagonism by the HIV-1 viral protein, Vif.

Importance

As a part of the innate immune system, humans encode proteins that inhibit viruses like HIV-1. These broadly acting antiviral proteins do not protect humans from viral infections because viruses encode proteins that antagonize the host antiviral proteins to evade the innate immune system. One such example of a host antiviral protein is APOBEC3C (A3C), which weakly inhibits HIV-1. Here, we show that we can improve the antiviral activity of A3C by duplicating the DNA sequence to create a synthetic tandem domain protein, and furthermore, these new proteins are relatively resistant to the viral antagonist, Vif. Together, these data give insights about how nature has evolved a defense against viral pathogens like HIV.

Introduction

The *APOBEC3 (A3)* gene locus in primates encodes cytidine deaminase proteins that inhibit endogenous retroelements, like LINE-1, and retroviruses, such as HIV-1, among other viruses including hepatitis B virus, human papillomavirus, and some herpes viruses (86, 167–169). This *A3* gene locus has expanded in primates to give rise to the seven members of the *A3* family, named *A3A* to *A3H* (95). For A3 proteins to inhibit HIV-1 replication, they must be packaged into budding virions and brought to the target cell, where they extensively deaminate cytidines on the negative strand of ssDNA to uridines during reverse transcription (86, 167). The resultant G-to-A hypermutation on the positive strand renders the provirus non-functional. The most potent naturally found A3, A3G, can mutate up to 10% of the guanines on a single provirus (170). In addition to extensive hypermutation, some A3s can also inhibit reverse transcription via deamination-

independent mechanisms as demonstrated by the presence of truncated cDNA products in the presence of A3s (77–79, 167, 171). Recent studies of A3G show that this non-enzymatic mechanism functions by binding and sterically hindering the enzymatic activity of reverse transcriptase (76).

There are four human A3s that have endogenous antiviral activity against HIV-1 in T cells: A3D, A3F, A3G, and A3H with A3G being the most potent (97). In order to replicate in the presence of these A3s, HIV-1 and other lentiviruses encode a protein, Vif, that targets A3s for proteasomal degradation. Vif has evolved three separate interfaces in order to degrade A3s: one binding A3G, another one A3H, and a third able to recruit A3C/A3D/A3F (86). Moreover, Vif also binds to the host factor CBF-β to help recruit an E3 ubiquitin ligase complex composed of CUL5, ELOB, ELOC, RBX2, and ARIH2 proteins that mediate polyubiquitination and rapid degradation of the A3s through the proteasome (83, 172).

In addition to the multiple *A3* genes in humans, there is additional diversity in the *A3* locus in the form of human polymorphisms that affect the antiviral activity against HIV-1. For instance, there are at least twelve haplotypes of A3H that vary in their ability to restrict HIV-1 (111, 112). In addition, the most common form of human A3C, which encodes for a serine at position 188, has little, if any, activity against HIV-1. However, about 10% of African individuals carry a polymorphism in A3C that encodes an isoleucine at position 188. This one amino acid change from a serine to an isoleucine is correlated with increased antiviral activity against HIV-1 resulting from an increased ability of A3C to form dimers and increased cytidine deaminase activity *in vitro* (98). Furthermore, chimpanzee and gorilla A3Cs have different amino acids at position 115 than humans and

introducing these amino acids into human A3C also increase dimerization and cytidine deaminase activity (113). Nonetheless, even with these mutations, A3C is less active against HIV-1 compared to many of the other A3 proteins (114).

Each *A3* gene can be classified according to the presence of a zinc-coordinating domain motif: A3Z1, A3Z2, or A3Z3 (96). The evolutionary history of the *A3* locus is characterized by duplication, recombination, and deletion events; in the last 50 million years, three duplication events have occurred in the A3Z1 and A3Z2 sub-families, but not in the A3Z3 (95). These fusion and recombination events gave rise to the seven *A3*s found in primates that include the three single deaminase domain *A3s* (*A3A*, *A3C*, and *A3H*) and four double deaminase domain *A3s* (*A3B*, *A3D*, *A3F*, and *A3G*).

Super restriction factors are defined as evolution-guided variants of natural restriction factors with improved properties based on previous work done on the restriction factor MXA (152). Super restriction factors provide a unique tool to study how restriction factors work and the evolutionary compromises for paths that have yet to be sampled in nature. Because the most active A3s are double domain proteins, we examined the hypothesis that super restrictor factors could be made by duplicating the poorly active single domain A3C protein into a synthetic tandem domain protein. We did this for both the common form of human A3C ($A3C_{S188}$) and the more active variant of A3C ($A3C_{I188}$). Remarkably, we found that all A3C tandem domain variants have greater antiviral activity relative to their single domain counterparts. We found that the tandem deaminase A3C variants are packaged into virions at higher levels than their single domain counterparts, likely explaining the majority of the increase in antiviral activity. In the natural double domain APOBEC3 proteins, only the C-terminal cytidine deaminase

active sites are used for hypermutation (173, 174). Surprisingly, we found that mutation of the cytidine deaminase active site in the C-terminal domain of $A3C_{S188}$-$A3C_{S188}$ results in even greater antiviral activity than the same protein with two active sites. This increase in antiviral activity is correlated with elevated packaging into virions and is largely independent of increased mutational load. Instead, there were lower amounts of reverse transcriptase products in cells infected with virus produced in the presence of these A3C tandem domain proteins. This inhibition of late reverse transcription products is also correlated with the formation of large higher-order complexes in the A3C-A3C variants compared to their single domain counterparts. Finally, we show that the A3C tandem domain proteins are largely resistant to antagonism by HIV-1 Vif. These A3C-A3C super restriction factors provide a unique tool to understand evolutionary paths of antiviral genes and their antagonism by viral proteins.

Results

Synthetic tandem domains of A3C have increased antiviral activity and are better packaged into virions

Of the four antiviral A3s, three of these, A3D, A3F, and A3G, all have two cytidine deaminase domains, whereas A3C has one cytidine deaminase domain and only weakly inhibits HIV-1 (98, 101, 103). These observations led us to hypothesize that synthetic double domain variants of the weakly active A3C would be more active than the single domain counterparts. To determine the antiviral effects of single versus double domain A3s, we created synthetic tandem domain A3Cs, called A3C-A3C here, using both $A3C_{S188,}$ the variant that is most common in the human population, and

$A3C_{I188}$, the variant that has increased antiviral activity and is present at a low frequency in humans (98).

We modeled the synthetic tandem deaminase protein after the most closely sequence-related double deaminase *A3* genes, *A3D* and *A3F*. A3C-A3C constructs were created by fusing two A3C sequences connected with a linker sequence (Arg-Asn-Pro) that is naturally found between the N- and C- terminal domains of A3D and A3F, as well as the natural deletion in the N-terminus of the second domain of A3C such that the C-terminal domain of A3C-A3C begins at the fifth amino acid of A3C, a methionine (Supplemental Figure 1).

We first examined the ability of each epitope-tagged variant to be expressed in cells and packaged into virions in 293T cells. All A3C variants were expressed at similar levels in cells (Figure 4A top), albeit with somewhat lower expression for the $A3C_{I188}$-$A3C_{I188}$ protein. However, the tandem domain A3C proteins were better packaged into virions compared to the natural single domain A3 proteins. In virions, we find that there is about a 1.6-fold increase in packaging for $A3C_{S188}$-$A3C_{S188}$ compared to $A3C_{S188}$ and about a 2.9-fold increase in packaging of $A3C_{I188}$-$A3C_{I188}$ relative to $A3C_{S188}$ (quantified in Figure 4A bottom, based on three independent experiments).

To evaluate the antiviral activity of these tandem domain proteins, we tested each variant at increasing concentrations in a single-cycle HIV-1ΔVif infectivity assay. As previously shown (98), $A3C_{S188}$ inhibits HIV-1ΔVif only weakly, and $A3C_{I188}$ has increased antiviral activity relative to $A3C_{S188}$ (Figure 4B: approximately 2-fold greater activity at the highest plasmid dosage). We found that $A3C_{S188}$-$A3C_{S188}$ has approximately 2-fold increased antiviral activity at every concentration relative to the

$A3C_{S188}$ single domain protein (black in Figure 4B). These levels of enhanced antiviral activity are similar to the previously characterized $A3C_{I188}$ (98). $A3C_{I188}$-$A3C_{I188}$, however, gained 3-fold greater antiviral activity when compared to its single domain counterpart $A3C_{I188}$, restricting HIV-1ΔVif to 5.8% infectivity at its highest concentration (blue in Figure 4B). Thus, these data show that A3C tandem domain proteins have increased antiviral activity relative to their single domain counterparts. This increased antiviral activity of the tandem domain variants relative to the parent single domains (Figure 4B) largely correlates with the increased packaging of the tandem domain A3Cs related to the single domain A3Cs (Figure 4A).

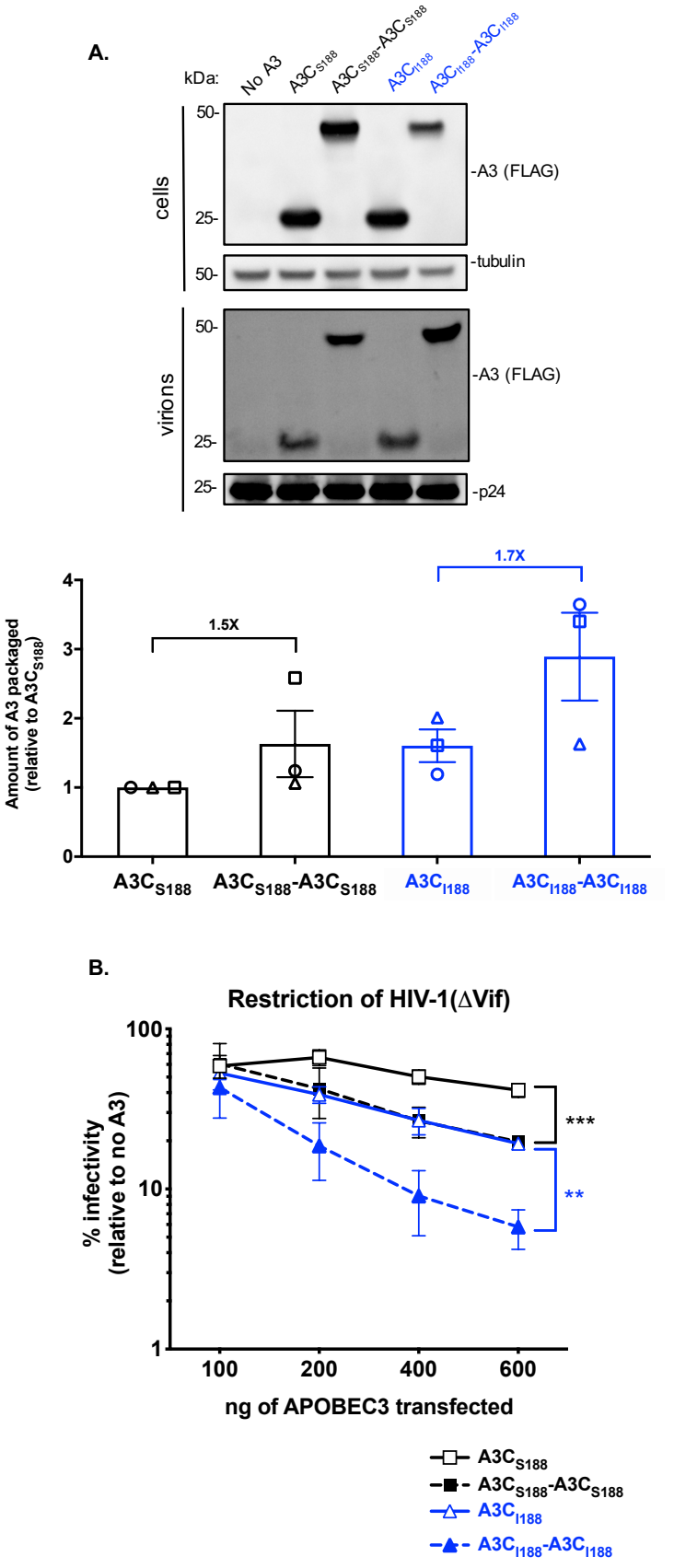

A.
kDa:
No A3
A3C$_{S188}$
A3C$_{S188}$-A3C$_{S188}$
A3C$_{I188}$
A3C$_{I188}$-A3C$_{I188}$
cells
50-
25-
-A3 (FLAG)
50-
-tubulin
virions
50-
25-
-A3 (FLAG)
25-
-p24
1.7X
1.5X
Amount of A3 packaged (relative to A3C$_{S188}$)
4
3
2
1
0
A3C$_{S188}$
A3C$_{S188}$-A3C$_{S188}$
A3C$_{I188}$
A3C$_{I188}$-A3C$_{I188}$
B.
Restriction of HIV-1(ΔVif)
100
10
1
% infectivity (relative to no A3)
***
**
100
200
400
600
ng of APOBEC3 transfected
A3C$_{S188}$
A3C$_{S188}$-A3C$_{S188}$
A3C$_{I188}$
A3C$_{I188}$-A3C$_{I188}$

Figure 4. A3C tandem domain proteins have increased antiviral activity and packaged more than their single domain counterparts

(A) Intracellular expression and packaging of A3C variants into virions. HIV-1ΔEnvΔVif provirus was co-transfected into 293T cells with 400ng of each A3C variant. Top: western blot of cellular lysates probed with anti-FLAG antibody showing intracellular expression levels for A3s and tubulin as a loading control. Middle: western blot of proteins in the pelleted virions and probed with anti-FLAG antibody for A3 levels and anti-p24gag for normalization. An empty vector condition was used a negative control and labeled no A3. Western blot shown is representative of 3 biological replicates. Bottom: Bar graph showing quantification of A3 protein packaged into virions from 3 biological replicates. The relative abundance of A3 in cell lysates and virions was quantified with Image Lab. Relative A3 packaged was calculated by dividing the relative abundance of A3 in the virions by the normalized levels of each A3 expressed in the cells. (B) The % infectivity of HIV-1ΔEnvΔVif pseudotyped with VSV-g and increasing doses of A3C variants is plotted, normalized to a control with no A3. The amount of each A3 plasmid transfected in ng is shown on the X-axis. $A3C_{S188}$ (black open squares, solid line), $A3C_{S188}$-$A3C_{S188}$ (black closed squares, solid line), $A3C_{I188}$ (blue open triangles, dashed blue line), and $A3C_{I188}$-$A3C_{I188}$ (blue close triangles, dashed blue line) are compared. Data points are an average of 3 biological replicates, with each biological replicate consisting of 3 triplicate infections. Error bars show the standard deviation of the mean (+/- SEM). Statistical differences were determined by unpaired *t* tests between $A3C_{S188}$ and $A3C_{S188}$-$A3C_{S188}$ (black bracket) and $A3C_{I188}$ and $A3C_{I188}$-$A3C_{I188}$ (blue bracket): ** P≤0.01, *** P≤0.001.

Mutation of one active site in $A3C_{S188}$-$A3C_{S188}$ increases antiviral activity even further

$A3C_{S188}$-$A3C_{S188}$ contains two cytidine deaminase motifs within the A3 conserved amino acid sequence His-X-Glu-X_{23-28}-Cys-Pro-X_{2-4}-Cys (Supplemental Figure 1). In A3F and A3G, the C-terminal catalytic domain exerts the key enzymatic activity, while the N-terminal catalytic domain mediates packaging into virions (79, 173, 174). Therefore, we next asked if the increase in antiviral activity of $A3C_{S188}$-$A3C_{S188}$ is due to two active cytidine deaminase sites and whether or not the enzymatic functions of A3C are important for restriction. We created active site point mutations by changing the essential glutamic acid to an alanine in each domain, E68A and E254A, respectively, and found that these changes did not affect protein expression levels in 293T cells (Figure 5A). As expected, when the glutamic acid was mutated to an alanine in the single domain $A3C_{S188}$ (called $A3C_{S188}$ E68A), the protein completely lost its

relatively weak antiviral activity (Figure 5B). In contrast and unexpectedly, when the C-terminal active site of A3C$_{S188}$-A3C$_{S188}$ was mutated, A3C$_{S188}$-A3C$_{S188}$ E254A, the antiviral activity instead *increased* by 5.8-fold (Figure 5B). On the other hand, when the N-terminal active site is inactivated in A3C$_{S188}$-A3C$_{S188}$ (called A3C$_{S188}$-A3C$_{S188}$ E68A), the antiviral activity did not significantly change (Figure 5B), suggesting that this site is not necessary for the antiviral activity, but that the active site at position 254 inhibits antiviral activity.

We further tested the abilities of A3C$_{S188}$-A3C$_{S188}$ E254A and A3C$_{S188}$-A3C$_{S188}$ to inhibit HIV-1ΔVif in a dose-response experiment by increasing the amount of *A3* plasmid. At the highest concentrations of plasmid transfected, A3C$_{S188}$-A3C$_{S188}$ E254A was able to reduce the infectivity to 3%, while the wild-type A3C$_{S188}$-A3C$_{S188}$ only reduced infectivity to 18% (Figure 5C). This potent antiviral activity of the A3C$_{S188}$-A3C$_{S188}$ E254A mutant suggests that having only one functional active site increases its antiviral effect.

In order to test if the increase in antiviral activity of the A3C$_{S188}$-A3C$_{S188}$ E254A mutant is also linked to increased packaging into virions, we examined the amount of packaged A3 via western blot. When compared to the intracellular expression, A3C$_{S188}$-A3C$_{S188}$ and A3C$_{S188}$-A3C$_{S188}$ E254A are expressed at relatively equal amounts (Figure 5D). However, when we evaluate the A3 packaged into virions, we see a 2.3-fold increase in A3C$_{S188}$-A3C$_{S188}$ E254A packaging compared to A3C$_{S188}$ (Figure 5D). This finding suggests that having only one, rather than two, functional catalytic sites increases the packaging and antiviral activity of A3C$_{S188}$-A3C$_{S188}$.

Both cytidine deaminase-dependent and cytidine deaminase-independent mechanisms of A3 inhibition of HIV-1 have been described (76–79, 167, 171). The antiviral activity of $A3C_{S188}$ requires an intact deaminase motif, as demonstrated by the total loss of restriction in $A3C_{S188}$ E68A (Figure 5B). However, when both active sites are mutated to make a catalytically inactive $A3C_{S188}$-$A3C_{S188}$, we still see 3.8-fold restriction of HIV-1ΔVif – a level of activity indistinguishable from the wild-type $A3C_{S188}$-$A3C_{S188}$ (Figure 5B). This suggests that the synthetic tandem domain $A3C_{S188}$-$A3C_{S188}$ functions as an antiviral protein in a cytidine deaminase-independent manner.

We also asked if $A3C_{I188}$-$A3C_{I188}$ increased antiviral activity with only one active deaminase domain. Therefore, we mutated each of the essential glutamic acids to an alanine in $A3C_{I188}$-$A3C_{I188}$ and tested for their ability to restrict HIV-1ΔVif. All of these $A3C_{I188}$-$A3C_{I188}$ active site mutants are expressed to similar levels in 293T cells (Supplemental Figure 2A). When the C-terminal active site was mutated, $A3C_{I188}$-$A3C_{I188}$ E254A lost antiviral activity (2.2-fold restriction) compared to the wild-type $A3C_{I188}$-$A3C_{I188}$. Similar to $A3C_{S188}$-$A3C_{S188}$, the N-terminal active site mutant had antiviral activity that was not statistically significant from the wild-type $A3C_{I188}$-$A3C_{I188}$ (Supplemental Figure 2B). However, in contrast to $A3C_{S188}$-$A3C_{S188}$, the C-terminally inactive mutant $A3C_{I188}$-$A3C_{I188}$ E254A and the double inactive mutant $A3C_{I188}$-$A3C_{I188}$ E68A E254A were indistinguishable in their antiviral activity, suggesting that $A3C_{I188}$-$A3C_{I188}$ only uses one catalytic site, like A3F and A3G. Thus, we speculate that $A3C_{I188}$-$A3C_{I188}$ is already optimized for its most potent restriction.

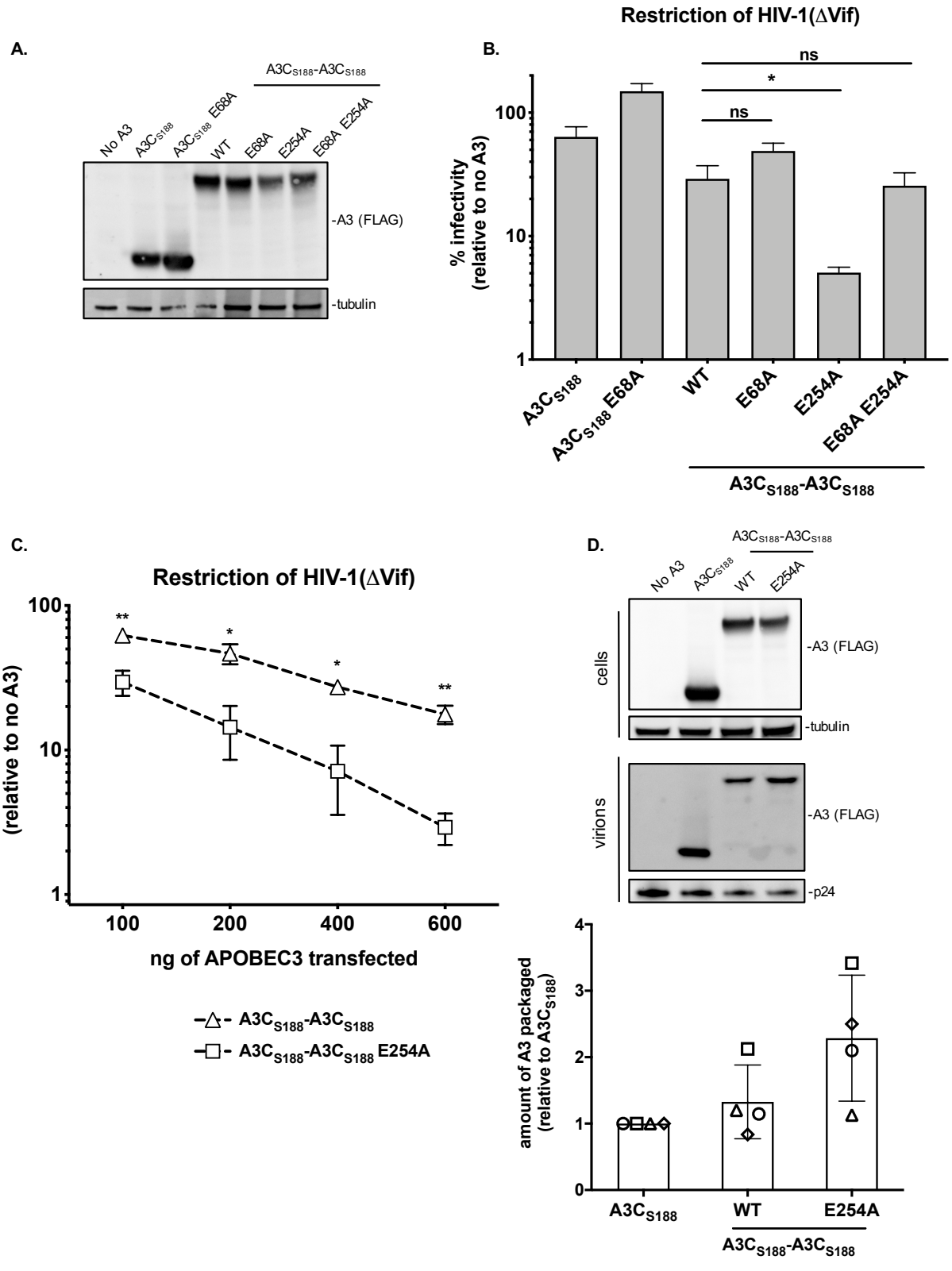

A.
A3C$_{S188}$-A3C$_{S188}$
No A3
A3C$_{S188}$
A3C$_{S188}$ E68A
WT
E68A
E254A
E68A E254A
-A3 (FLAG)
-tubulin
B.
Restriction of HIV-1(ΔVif)
ns
*
ns
% infectivity (relative to no A3)
100
10
1
A3C$_{S188}$
A3C$_{S188}$ E68A
WT
E68A
E254A
E68A E254A
A3C$_{S188}$-A3C$_{S188}$
C.
Restriction of HIV-1(ΔVif)
% infectivity (relative to no A3)
100
10
1
**
*
*
**
100
200
400
600
ng of APOBEC3 transfected
A3C$_{S188}$-A3C$_{S188}$
A3C$_{S188}$-A3C$_{S188}$ E254A
D.
A3C$_{S188}$-A3C$_{S188}$
No A3
A3C$_{S188}$
WT
E254A
cells
-A3 (FLAG)
-tubulin
virions
-A3 (FLAG)
-p24
amount of A3 packaged (relative to A3C$_{S188}$)
4
3
2
1
0
A3C$_{S188}$
WT
E254A
A3C$_{S188}$-A3C$_{S188}$

Figure 5. One functional deaminase domain in a tandem domain A3C protein is optimal for antiviral activity and packaging into virions

(A) Western bot analysis of intracellular expression levels of active site point mutation in $A3C_{S188}$-A3C $_{S188}$. WT denotes wild-type $A3C_{S188}$-A3C $_{S188}$; E68A refers to a mutation in the catalytic deaminase site in either the single domain A3C or N-terminus of the synthetic tandem domain $A3C_{S188}$-$A3C_{S188}$; E254A refers to a mutation in the C-terminus of the synthetic tandem domain $A3C_{S188}$-$A3C_{S188}$; E68A E254A refers to a double catalytic deaminase site mutant in $A3C_{S188}$-$A3C_{S188}$. Antibodies to FLAG were used to detect A3s and tubulin was used as a loading control. (B) Single-cycle infectivity assay measuring the percent infectivity of each A3C variant and active site mutant against HIV-1ΔEnvΔVif. Nomenclature of A3C variants is the same as in panel A. Cells were transfected with 400ng of A3 and 600ng of HIV-1ΔEnvΔVif pseudotyped with 100ng of VSV-g. Virus production was normalized and equal amounts of virus was used infect SUPT1 cells. Results from each experiment were normalized to a no APOBEC3 control. Bar graph shows an average of 3 biological replicates, each with triplicate infections (+/- SEM). Statistical differences were determined by unpaired *t* tests: * P≤0.05, ns= not significant. (C) Dose response analysis showing the infectivity of HIV-1ΔEnvΔVif pseudotyped with VSV-g, comparing $A3C_{S188}$-$A3C_{S188}$ (open triangles, solid line) to $A3C_{S188}$-$A3C_{S188}$ E254A (open square, dotted line). Infection was normalized to a no A3 control. Amount of A3 transfected in ng is on the x-axis. Data points are an average of 3 biological replicates, with triplicate infections. Error bars represent SEM. Statistical differences were determined by unpaired *t* tests: * P≤0.05, ** P≤0.01. (D) Representative western blot of the packaging of $A3C_{S188}$-$A3C_{S188}$ E254A into virions. HIV-1ΔEnvΔVif provirus was co-transfected into 293T cells with A3C variants. Left: intracellular expression levels western blot probed with anti-FLAG antibody for A3 levels and anti-tubulin as a loading control. Right: proteins in the pelleted virions shown in a western blot and probed with anti-FLAG antibody for A3 levels and anti-p24gag for normalization. No A3 was used a transfection control. Bar graph below shows quantification of A3 protein packaged into virions from 4 biological replicates. Relative abundance of A3 in cell lysates and virions was quantified with Image Lab. Relative A3 packaged was calculated by dividing the relative abundance of A3 in the virions by the A3 in the cells.

The A3C synthetic tandem domain antiviral activity does not correlate with hypermutation activity

In order to directly test the ability of the A3C-A3C variants to hypermutate HIV-1 ssDNA, we developed an assay to assess G-to-A mutations by sequencing large numbers of unintegrated viral DNA products 12 hours post infection. This assay utilizes primers with unique barcodes designed to distinguish between A3 mutations and Illumina sequencing error (175). A plasmid-only control was used to identify sequencing error rates and the frequency of mutations acquired during PCR. A “No A3” control was used to distinguish mutations made by reverse transcriptase from those introduced by

the A3 variants and any residual A3 activity in 293T cells. Each sequencing read spans the same length of HIV-1 *pol*. The number of unique sequencing reads for each condition ranged from 90,833 to 905,916. Only 0.2% of the reads from the plasmid-only control have any G-to-A mutations (**Error! Reference source not found.**A). 7% of the reads from the infection without any added A3 have a G-to-A mutation (**Error! Reference source not found.**B), presumably the result of mutations from reverse transcriptase or residual A3 activity from 293T cells. In contrast, A3G shows the highest rate of G-to-A hypermutation with 88% of reads that have at least one mutation and with over 30% of the reads having 10 or more mutations. A3G also induces a distribution of templates having 2 to 9 of such mutations (**Error! Reference source not found.**C). In contrast, $A3C_{S188}$ causes fewer G-to-A mutations than A3G with 27% of reads containing at least one mutation and 9% having two or more mutations, as shown with the shift towards the left in the bar graphs relative to A3G (comparing **Error! Reference source not found.**C to **Error! Reference source not found.**D). As expected from the single-cycle infectivity assay (Figure 5B), the mutational burden induced by the $A3C_{S188}$ E68A active site mutant is not above background levels – only 5% of reads have any G-to-A mutations and 2% of reads have two or more of such mutations (**Error! Reference source not found.**E).

Consistent with the result that much of the antiviral activity of $A3C_{S188}$-$A3C_{S188}$ does not depend on an active cytidine deaminase (Figure 5B), we find that $A3C_{S188}$-$A3C_{S188}$ does not increase G-to-A mutations when compared to $A3C_{S188}$ (compare **Error! Reference source not found.**F and **Error! Reference source not found.**D). For the $A3C_{S188}$-$A3C_{S188}$ and $A3C_{S188}$ samples, 26% and 27% of reads have any G-to-A

mutations, respectively, and the distribution of the G-to-A mutation plots are nearly identical. Interestingly, 24% of reads from the $A3C_{S188}$-$A3C_{S188}$ E254A sample have a G-to-A mutation, resulting in a similar to distribution as the bar graph for $A3C_{S188}$-$A3C_{S188}$ (**Error! Reference source not found.**G and **Error! Reference source not found.**F, respectively). This finding complements the increase in packaging results seen in Figure 5D. As an additional control, $A3C_{S188}$-$A3C_{S188}$ E68A E254A has a G-to-A substitution rate below background levels (**Error! Reference source not found.**H).

Previous studies have shown that the increase in antiviral activity of $A3C_{I188}$ is due to an increase in enzymatic activity (98). This finding is replicated here with a shift towards the right in the $A3C_{I188}$ bar graph when compared to $A3C_{S188}$ (comparing **Error! Reference source not found.**I and **Error! Reference source not found.**D) such that 49% of reads have at least one G-to-A mutation. $A3C_{I188}$-$A3C_{I188}$, despite having more antiviral activity than $A3C_{I188}$ (Figure 4A), does not have a higher percentage of mutations—49% of reads have at least one mutation in $A3C_{I188}$ and 45% in $A3C_{I188}$-$A3C_{I188}$ (**Error! Reference source not found.**I and **Error! Reference source not found.**J). In sum, the hypermutation activities for the double domain A3Cs appear to have nearly identical distributions as the corresponding single domain A3Cs. Taken together, these deep sequencing data of A3 hypermutation support our earlier conclusion from the double active site mutations (Figure 5B) that the increase in antiviral activity is not due to an increase in enzymatic activity.

In addition, we determined if the nucleotide preferences for G-to-A mutations were changed by adding a second cytidine deaminase domain to A3C. As expected (73), A3G has a strong preference of 5'-G<u>G</u>-3' on the positive sense strand with the -1

site having equal preference among the nucleotides present (**Error! Reference source not found.**K). However, both $A3C_{S188}$ and $A3C_{S188}$-$A3C_{S188}$ show a similar lack of preference for the -1 and +1 position (**Error! Reference source not found.**M and **Error! Reference source not found.**N), further supporting that an increase in mutation frequency does not explain the different antiviral activities of $A3C_{S188}$ and $A3C_{S188}$-$A3C_{S188}$.

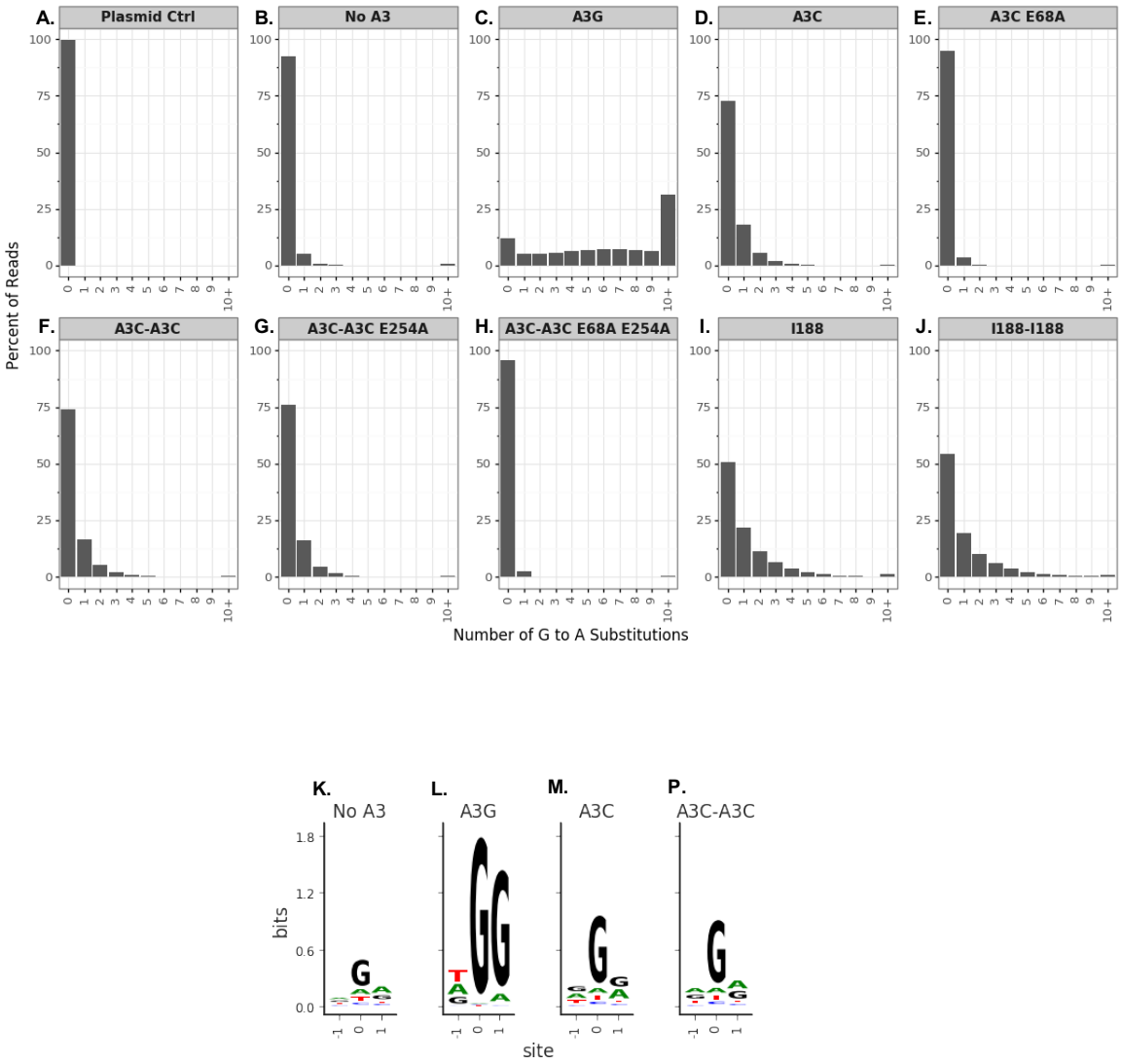


Figure 6. Frequency of G-to-A mutations does not increase in A3C tandem domain proteins

(A-J) Paired-end sequencing reads were analyzed for G-to-A mutations. Data is shown as frequency distribution bar graphs of the percent of reads by the number of G-to-A substitutions in each read for each A3 tested: A3G (C), A3C$_{S188}$ (D), A3C$_{S188}$ E68A (E), A3C$_{S188}$-A3C$_{S188}$ (F), A3C$_{S188}$-A3C$_{S188}$ E254A (G), A3C$_{S188}$-A3C$_{S188}$ E68A E254A (H), A3C$_{I188}$ (I), A3C$_{I188}$-A3C$_{I188}$ (J). Plasmid control (A) is used as a sequencing control and a no APOBEC3 (B) sample was used to distinguish background mutations, including RT mutations. Frequencies are calculated as the average frequency of technical replicates and read counts are shown as the sum of the reads for both replicates. (K-N) Extracted sequence information for all substitutions (and neighboring bases) in the No A3 (K), A3G (L), A3C$_{S188}$ (M), or A3C$_{S188}$-A3C$_{S188}$ (N) samples on viral cDNA. The constructed three-nucleotide logo plots are centered on the site of substitution, such that site 0 is fixed as the site of substitution and the -1 and +1 sites indicate the nucleotides immediately 5' and 3' of any substitution in these samples. The letter height is directly proportional to the frequency of that nucleotide at that site in the sample compared to the plasmid control.

A3C$_{S188}$-A3C$_{S188}$ tandem domain variants reduce the accumulation of reverse transcription products

Since we did not see an increase in G-to-A mutations in the A3C-A3C variants relative to A3C alone, we suspected that, similar to studies with an A3G active site mutant (76, 77, 171), the A3C-A3C tandem domain proteins might decrease reverse transcriptase products independent of hypermutation. To test this hypothesis, we used qPCR to quantify late reverse transcription (RT) products from unintegrated viral DNA. Virus produced in the presence of A3G showed a significant decrease in relative late RT products compared to the No A3 control, while A3C$_{S188}$ had equivalent levels of late RT products as the No A3 control (Figure 7). However, the amount of RT products produced from virus made in the presence of A3C$_{S188}$-A3C$_{S188}$ was significantly reduced (2.5-fold less) relative to that made in the presence of A3C$_{S188}$ (Figure 7). A3C$_{S188}$-A3C$_{S188}$ E254A had equivalent late RT products as A3C$_{S188}$-A3C$_{S188}$, consistent with the hypothesis that the increase in antiviral activity comes from the increase in packaging into virions (Figure 7, Figure 5C, and Figure 5D). Lastly, there was no significant difference in the late RT products of A3C$_{S188}$-A3C$_{S188}$ and A3C$_{S188}$-A3C$_{S188}$ E68A E254A, suggesting that inhibition of RT is likely the mechanism by which these variants

act (Figure 7 and Figure 5B). These findings further support the hypothesis that A3C super restriction factors function through a distinct mechanism to restrict HIV-1 compared to A3C single domain proteins.

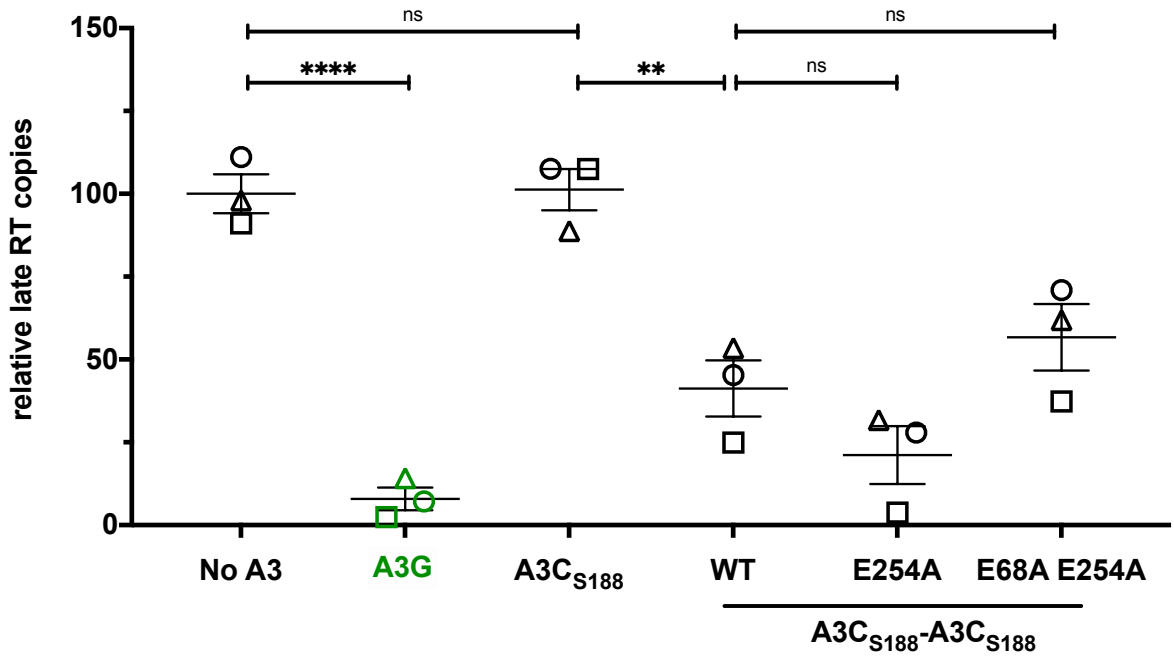


Figure 7. A3C tandem domain proteins operate in a deaminase-independent mechanism to inhibit reverse transcription products

(A) To evaluate the relative copies of late reverse transcription products, SUPT1 cells were infected with HIV-1ΔEnvΔvif and either no APOBEC3 or A3 to test for inhibition of HIV-1 reverse transcription. 18 hours later, viral cDNA was harvested and the levels of HIV-1 late reverse transcription products were assayed by qPCR. Each circle represents a normalized value for the respective biological replicate, with qPCR technical duplicates. Each sample has been adjusted for equal viral infection and a nevirapine control. Error bars indicate the SEM. Statistical differences were determined by unpaired *t* tests: ** P≤0.01, **** P≤0.0001, ns= not significant.

Tandem domain variants of A3C form larger higher-order complexes in cells relative to their native single domains

The ability of A3G to oligomerize in cells has been correlated with its antiviral activity because this oligomerization leads to increased packaging into virions (176, 177). The A3G oligomerization state also affects its binding to ssDNA and its catalytic activity (178). More specifically, A3G in lower molecular mass complexes has the ability to deaminate ssDNA (104, 178), while A3G residing in high molecular mass complexes is inactive, hinders rapid deamination of ssDNA, and is hypothesized to instead form a

roadblock to inhibit reverse transcription (104, 178). Because we observed that the A3C-A3C variants inhibited late RT products (Figure 7) rather than inducing hypermutation (**Error! Reference source not found.**), we examined the ability of each of the A3C variants to form high molecular weight complexes in a velocity sedimentation. As expected from previous reports (176, 179), A3G forms both lower and higher order complexes as demonstrated by its presence in the top and middle fractions of the a sucrose gradient (Figure5A). In contrast, both $A3C_{S188}$ and $A3C_{I188}$ were found in the top fractions of the gradient, overlapping with the GAPDH soluble control (Figure 8B and Figure 8D, respectively). This suggests the A3C single domain protein does not form higher-order complexes, unlike A3G. In contrast, we found that both $A3C_{S188}$-$A3C_{S188}$ and $A3C_{I188}$-$A3C_{I188}$ formed complexes that migrated farther down the sucrose gradient, similar to A3G (Figure 8C and Figure 8E, respectively). However, $A3C_{S188}$-$A3C_{S188}$ migrated even farther down the sucrose gradient than the pattern defined for A3G. Furthermore, both $A3C_{S188}$-$A3C_{S188}$ and $A3C_{I188}$-$A3C_{I188}$ have less protein in the top fractions than both A3G and their single domain counterparts (Figure 8F). Together, these findings suggest that the A3C tandem domain variants reside primarily in larger higher-order complexes.

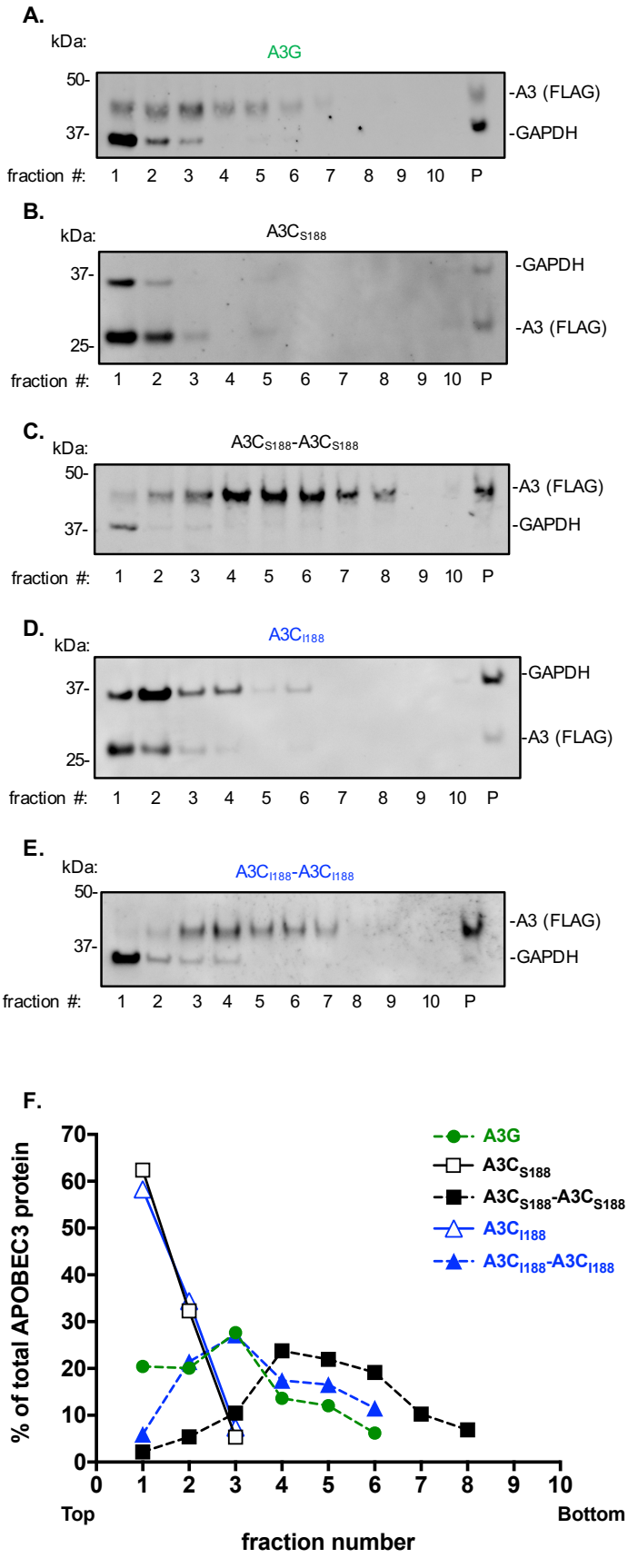

A.
kDa:
A3G
50-
37-
-A3 (FLAG)
-GAPDH
fraction #: 1 2 3 4 5 6 7 8 9 10 P
B.
kDa:
A3C$_{S188}$
37-
25-
-GAPDH
-A3 (FLAG)
fraction #: 1 2 3 4 5 6 7 8 9 10 P
C.
kDa:
A3C$_{S188}$-A3C$_{S188}$
50-
37-
-A3 (FLAG)
-GAPDH
fraction #: 1 2 3 4 5 6 7 8 9 10 P
D.
kDa:
A3C$_{I188}$
37-
25-
-GAPDH
-A3 (FLAG)
fraction #: 1 2 3 4 5 6 7 8 9 10 P
E.
kDa:
A3C$_{I188}$-A3C$_{I188}$
50-
37-
-A3 (FLAG)
-GAPDH
fraction #: 1 2 3 4 5 6 7 8 9 10 P
F.
% of total APOBEC3 protein
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8 9 10
Top
Bottom
fraction number
A3G
A3C$_{S188}$
A3C$_{S188}$-A3C$_{S188}$
A3C$_{I188}$
A3C$_{I188}$-A3C$_{I188}$

Figure 8. A3C double deaminase domain proteins form larger complexes than the single domains

(A-E) Velocity sedimentation and sucrose fractionation of A3C variants. Cell lysates transfected with 1000ng of A3 were subjected to velocity sedimentation fractionation (lanes 1-10) and 2.5% of each fraction was analyzed on a western blot: (A) A3G (B) $A3C_{S188}$ (C) $A3C_{S188\text{-}S188}$ (D) $A3C_{I188}$ (E) $A3C_{I188}$-$A3C_{I188}$. Each western blot is probed with anti-FLAG to analyze A3s levels and anti-GAPDH to mark the soluble fraction. (F) Quantification of percentage of the total A3 protein in each fraction. Relative abundance of A3 in each fraction from the sucrose gradient western blots was quantified with Image Lab and calculated as a percentage of total APOBEC3 protein found in all the fractions combined. A3G is shown in green circles, $A3C_{S188}$ in open black squares, $A3C_{S188}$-$A3C_{S188}$ in closed back squares, $A3C_{I188}$ in open blue triangles, $A3C_{I188}$-$A3C_{I188}$ in closed blue triangles.

A3C tandem domain proteins are largely resistant to viral antagonism by Vif

The most effective super restriction factors would need to both increase in antiviral activity and overcome viral antagonism. Therefore, we next examined the ability of the A3C-A3C super restriction factors to escape viral antagonism. To test if HIV-1 Vif from two different strains could degrade novel A3C tandem domain protein targets, we used a single-cycle infectivity assay using a HIV provirus containing a *Vif* gene from either a lab-adapted strain (LAI) or a primary-isolate *Vif* (patient ID 1203) that has been previously shown to degrade A3H and A3C variants (109). In the absence of the viral antagonist Vif, A3G potently inhibits HIV-1ΔVif, (

Figure 9A), and, as expected, the presence of either HIV-1 LAI Vif (light purple bars) or an HIV-1 primary-isolate of Vif (dark purple bars) leads to full antagonism of A3G (

Figure 9A). While, A3C$_{S188}$ and A3C$_{I188}$ inhibit HIV-1ΔVif, the presence of Vif fully antagonizes this antiviral activity (

Figure 9A). In contrast, each of the A3C tandem variants is partially resistant to Vif degradation as infectivity is not completely restored (

Figure 9A). Even in the presence of Vif, $A3C_{S188}$-$A3C_{S188}$ still restricted HIV-1 from LAI Vif to 20% infectivity and a primary-isolate Vif to 24%. Furthermore, $A3C_{S188}$-$A3C_{S188}$ E254A restricted HIV-1 to 7% and 12% infectivity (LAI Vif and primary-isolate Vif, respectively), and $A3C_{I188}$-$A3C_{I188}$ inhibits infection to 10% with both HIV-1 Vifs (

Figure 9A).

As expected from the inability of HIV-1 Vif to antagonize the A3C-A3C tandem domain proteins (

Figure 9A), we found that HIV-1 Vif did not decrease the amount of A3C-A3C packaged into virions (

Figure 9B). We used a western blot to evaluate levels of A3 proteins packaged into virions in the presence and absence of Vif (

Figure 9B). As a control, we used A3G, A3C$_{S188}$, and A3C$_{I188}$ as naturally found A3s that are degraded by HIV-1 Vif. Using densitometry, we quantified and compared the fraction of A3 probed in the presence of an HIV-1 provirus containing a deletion in the *Vif* gene to a LAI HIV-1 provirus retaining the *Vif* gene. As expected, A3G has 0.2 fraction of A3 remaining in the presence of Vif, A3C$_{S188}$ has 0.1, and A3C$_{I188}$ has 0.4, since HIV-1 Vif is able to degrade the naturally found A3s (

Figure 9B). Furthermore, these data support rescue of infection observed in the single-cycle infectivity assay (

Figure 9A). We found that $A3C_{S188}$-$A3C_{S188}$, $A3C_{S188}$-$A3C_{S188}$ E254A, and $A3C_{I188}$-$A3C_{I188}$ are packaged into budding virions in the presence of HIV-1 Vif with a fraction of A3 remaining in the presence of Vif to be 0.8, 0.7, and 1.7, respectively (

Figure 9B). This fraction of A3 remaining is much higher than the single domain counterparts and supports the inhibition of HIV-1 infection in the presence of Vif seen in the single-cycle infectivity assay (

Figure 9A and

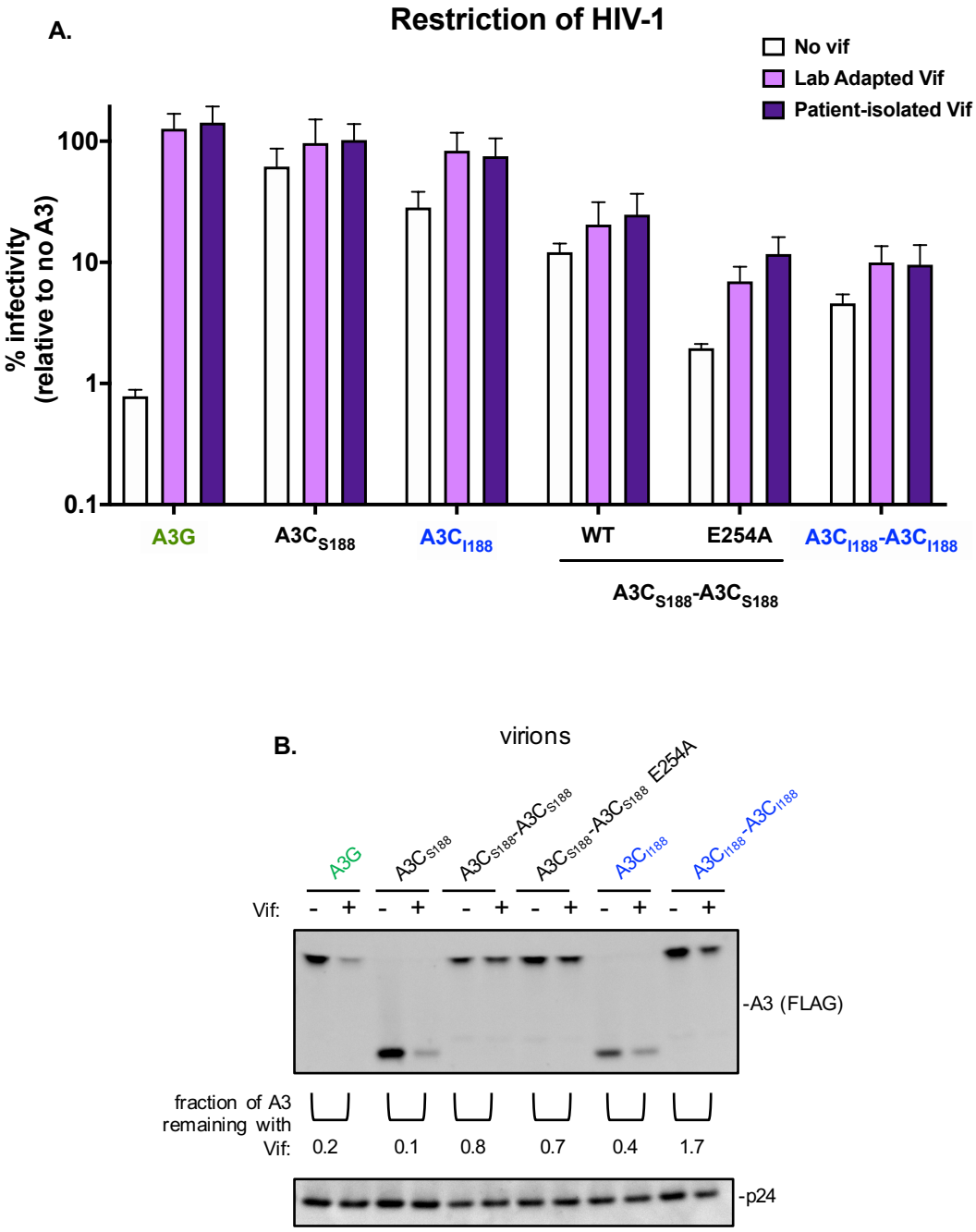


Figure 9B). Since HIV-1 Vif from two divergent strains is unable to fully antagonize these A3C-A3C variants, our data show that a synthetic tandem domain version of an A3 protein can both increase its potency and allow it to resist antagonism by HIV-1 Vif.

Figure 9. A3C tandem domain variants are resistant to viral antagonism

(A) Single-cycle infectivity assay performed in the presence of HIV-1ΔEnvΔvif (white), HIV-1ΔEnv +LAI Vif (light purple), and HIV-1ΔEnv+1203 Vif (dark purple). Cells were transfected with 400ng of A3 and 600ng of noted HIV-1ΔEnv provirus pseudotyped with 100ng of VSV-g. Virus production was normalized and equal amounts of virus was used infect SUPT1 cells. Results from each experiment were normalized to a no APOBEC3 control. Bar graph shows an average of 3 biological replicates and triplicate infections (+/- SEM). (B) Western blot of the packaging of A3C variants into virions. Either HIV-1ΔEnvΔVif or HIV-1ΔEnv+LAI Vif provirus was co-transfected into 293T cells with A3C variants. Proteins in the pelleted virions are shown in a western blot probed with anti-FLAG antibody for A3 levels and anti-p24gag as a loading control. Densitometry calculations were performed to determine the fraction of A3 remaining in the presence of Vif. Western blot shown is a representative of 3 biological replicates.

Discussion

Although human A3C can weakly inhibit HIV-1 replication, here we show that a super restriction factor can be created by linking two A3C sequences. These tandem domain proteins both increase anti-HIV-1 activity and yield a restriction factor that is partially resistant to antagonism by HIV-1 Vif. We have shown that this increase in antiviral activity is mostly explained by increased packaging of tandem domain A3s into budding virions. Moreover, we found that $A3C_{S188}$-$A3C_{S188}$ appears to use both a cytidine deaminase-dependent and independent mechanism for HIV-1 restriction. However, we could further increase the antiviral activity of $A3C_{S188}$-$A3C_{S188}$ by mutating the C-terminal active site ($A3C_{S188}$-$A3C_{S188}$ E254A). We found that A3C-A3C tandem domain proteins do not have an increase in mutation frequency when compared to their single domain counterparts, but rather the dominant mechanism of restriction is through inhibition of reverse transcription. Additionally, all the A3C tandem domain variants are resistant to HIV-1 Vif degradation. Together, these data point towards a selective advantage of double versus single domain A3 proteins against a viral target and support the idea that additional potency and escape from viral antagonism can be derived from combinations of APOBEC3 domains that have not been sampled in nature.

Two domain A3C proteins are packaged at higher levels than single domain A3C proteins

In order for the A3 proteins to be antiviral, they must be packaged into virions to function during the reverse transcription process that occurs in the target cell. We found that increase in antiviral activity of the A3C tandem domain proteins increase relative to their single domain counterparts closely parallels both the increase of A3 packaged into

budding virions (Figure 4 and Figure 5). However, the variability of the packaging assay does not allow us to conclude that packaging alone is responsible for the increased antiviral activity. Nonetheless, each of the A3C-A3C variants formed larger higher-order complexes when compared to their single domains (Figure 8) and it has been shown that oligomerization is important for A3G to be packaged into virions (177). A3H is an interesting exception to the idea that double cytidine deaminase A3 proteins have better oligomerization characteristics relative to single cytidine deaminase A3 proteins because it is the most distantly related A3 with only one deaminase domain that never duplicated or recombined to form a double deaminase protein (95). However, similar to A3G and the A3C-A3C variants, A3H can form large complexes in sucrose gradients, yet these complexes are RNA dependent (102, 180). Also, A3H is polymorphic in human populations like A3C, yet, A3H haplotype II confers strong antiviral activity (101). Recent structural work has shown that A3H binds to RNA to make functional dimers (180–182). This suggests that perhaps A3H evolved an independent mechanism to form higher-order structures that is dependent on RNA, but that A3C is unable to do this unless the second cytidine deaminase domain is artificially engineered. Furthermore, we can speculate that nature has selected for both single and double domain A3s because of the selective advantage of both monomer and dimer populations for deaminase-dependent activity as well as a larger order complex population for deaminase-independent activity.

Deaminase-dependent and deaminase-independent mechanisms of super restriction factor antiviral activity

Naturally found A3 proteins primarily restrict HIV-1 through hypermutation of ssDNA intermediates during reverse transcription. Studies have reported that A3G and A3F can also function through a cytidine deaminase-independent mechanism to restrict HIV-1 (77–79, 171). However, hypermutation rather than steric inhibition are the primary modes of restriction (76, 78, 79). Furthermore, A3G forms both low and high order complexes in cells, but the A3G residing in the higher-order complexes has been shown to have hindered deaminase activity (178). While A3G residing in the smaller complexes appears to be deaminating ssDNA targets, these large A3G complexes have been hypothesized to form a roadblock to inhibit reverse transcription (178). Interestingly, the $A3C_{S188}$-$A3C_{S188}$ double inactive site mutant ($A3C_{S188}$-$A3C_{S188}$ E68A E254A) still retained antiviral activity that, in fact, is indistinguishable from the wild type $A3C_{S188}$-$A3C_{S188}$ (Figure 5B). The A3 mediated hypermutation assay data added to these findings by showing that the tandem domain A3 proteins do not have more enzymatic activity than their single domain counterparts (**Error! Reference source not found.**). Rather, both $A3C_{S188}$-$A3C_{S188}$ and $A3C_{S188}$-$A3C_{S188}$ E254A inhibited the total late RT products more than their A3C single domain counterparts (Figure 7). Each of these results is also consistent with a lack of total A3 in the top fraction of the A3C tandem domains (Figure 8). This finding supports the hypothesis that these large higher-order complex contribute to hindering reverse transcription in a deaminase-independent mechanism. Thus, we found that these A3C-A3C variants gained a new mechanism to restrict reverse transcription: the ability to inhibit HIV-1 reverse transcription, most likely through forming large complexes of oligomers.

Only one cytidine deaminase domain is active in tandem domain APOBEC3 proteins

A3G and A3F have two deaminase domains, yet these naturally found A3s primarily rely on one domain, the C-terminal domain, for their catalytic activity (173, 174). Our finding that only one active domain is more antiviral than two active domains (Figure 5) is similar to how the A3F and A3G domains have evolved. However, A3G and A3F do not have enhanced antiviral activity when active site mutations are created (79, 174). Additionally, the preferred enzymatic domain for $A3C_{S188}$-$A3C_{S188}$ is the N-terminal domain, unlike with A3F and A3G. Nevertheless, the C-terminal active site mutant of $A3C_{S188}$-$A3C_{S188}$ can parallel the evolutionary process that the naturally found A3s have undergone. Since $A3C_{S188}$-$A3C_{S188}$ is less potent as a restriction factor than $A3C_{S188}$-$A3C_{S188}$ E254A, this suggests that there is a disadvantage to having two fully active deaminase sites. We speculate that the specialization of the A3 domains could be important for optimal efficiency as an enzyme, such that one domain is primarily used for packaging into virions and oligomerization while the other is important for scanning ssDNA for its deamination activity. Since A3F and A3G have evolved to only use one domain for hypermutation, mutating one catalytic site will not increase antiviral activity. However, $A3C_{S188}$-$A3C_{S188}$ can be further improve upon by mutating the protein such that only one active site is used, $A3C_{S188}$-$A3C_{S188}$ E254A, to recapitulate natural selection seen in A3F and A3G.

A3C-A3C super restriction factors are mostly resistant to Vif antagonism

Previous studies have shown that HIV-1 Vif binds and degrades $A3C_{S188}$ and $A3C_{I188}$ (98, 183). One viral protein, Vif, must counteract the antiviral activity of multiple

A3s to achieve maximal infectivity for the virus (86). HIV-1 Vif has evolved three separate interfaces in order to degrade A3s: one for binding A3G, another for A3H, and a third interface that is able to interact with A3C/A3D/A3F (86). In contrast to the human A3 proteins, HIV-1 has never evolved to antagonize an A3C-A3C tandem domain protein. The synthetic nature of these tandem domain A3s may explain why HIV-1 Vif is not able to completely antagonize any A3C-A3C variants (

Figure 9). The two different strains of HIV-1 Vif are approximately 85% identical and cannot fully antagonize A3C-A3C tandem domain proteins, suggesting a potential mechanism that the interface Vif previously used to bind to the A3C is now occluded by these tandem domain proteins. However, another possibility is that A3C-A3C is packaged so efficiently that Vif is unable to target all the active A3 prior to packaging

into virions, although only a ~1.5-fold increase in packaging of the super restriction factors makes this less likely. In either case, the increased activity and resistance to Vif antagonism of these super restriction factors provide useful insights about the initial constraints of Vif recognition to novel A3 variants.

Chapter 4. Synthetic double domain APOBEC3 chimeras create antiviral factors against HIV-1 that are more potent than APOBEC3G

Introduction

Positive selection in antiviral genes is a result of the host-virus "arms-race" due to repeated cycles of host resistance and virus adaptation. However, in addition to the increased evolutionary rate single amino acid substitutions that characterize host antiviral genes under positive selection, additional gene innovation in these antiviral genes has also been sampled through gene duplication and recombination to create antiviral gene families. Indeed, neo- or sub-functionalization upon gene duplication is an attractive evolutionary strategy to expand host anti-pathogen response through evasion of a viral gene or evolution of a new inhibitory mechanism. Such examples include the *Mx, IFIT* and *TRIM* genes. Most mammals, including humans, encode two paralogs of Mx proteins, MxA and MxB. Human MxA has broad and potent activity against a diverse range of RNA and DNA viruses, while the antiviral scope of human MxB is more restricted to lentiviruses and herpesviruses (115, 116, 154, 184). The number and identify of gene families can vary substantially between species. For example, humans have five *IFIT* genes (*IFIT1*, *1B*, *2*, *3*, and *5*), rats have four (*IFIT1*, *1b*, *2*, and *3*) and mice have six (*IFIT1*, *1b*, *1c*, *2*, *3*, and *3b*) (141). Conversely, *TRIM5*, a potent restriction factor against lentiviruses, is present in only a single copy in most primates, whereas it is present in up to six copies in rodents (185).

Another such example of antiviral gene family expansion is seen within the *apolipoprotein B mRNA editing enzyme catalytic-polypeptide like 3*, *APOBEC3* (shortened to *A3* here) locus. A3s are a family of cytidine deaminases that hypermutate

retroviruses such as HIV-1 and other endogenous retroelements. The *APOBEC3* (*A3*) locus, which is unique to placental mammals, as undergone dramatic expansion in many mammalian lineages including primates. The *A3* loci is flanked by *CBX6* and *CBX7* genes, suggesting that the amplification of A3 genes has mainly occurred via tandem gene duplication. The human genome encodes seven *A3* paralogs (named *A3A, A3B, A3C, A3D, A3F,* and *A3H*). In addition to this gene duplication, most of the A3 proteins have also evolved under positive selection in primates.

The *A3* gene family encodes a characteristic zinc-coordinating catalytic motif (His-X-Glu-X_{23-28}-Pro-Cys-X_{2-4}-Cys) and the A3 proteins can be grouped into 3 classes (A3Z1, A3Z2, and A3Z3) on the bases of their conserved Z domain sequences. Of the seven A3 paralogs in humans *A3A*, *A3C*, and *A3H* encode single domain proteins (*A3Z1*, *A3Z2*, and *A3Z3*, respectively), whereas the four remaining A3s are double Z domains. A3B and A3G are categorized as A3Z2-A3Z2, while A3D and A3F are A3Z2-A3Z1. In addition to characterizing A3s by their Z domains, these proteins vary in their ability to restrict HIV-1. A3A and A3B do not have antiviral activity against HIV-1, while A3G is the most potent naturally found A3. Despite the potent antiviral restriction these A3 pose, HIV-1 encodes an antagonist, Vif, to ameliorate this host block. To induce A3 degradation, Vif binds to CBF-β and recruits CUL5-EloB/C-E3 ubiquitin ligase complex to promote polyubiquitination and the subsequent proteasomal degradation of the A3s (72, 80–85). Vif is able to antagonize multiple A3 proteins including A3C, A3D, A3F, A3G, and A3H. Vif uses three distinct interfaces to bind to A3s: one for A3G, another for A3C/A3D/A3F, and another for A3H (86).

The human *A3* locus has sampled additional diversity through polymorphisms that encode proteins with different antiviral activities. For example, A3C also has a common form that encodes a serine at position 188 that weakly inhibits HIV-1, but a polymorphism found at 10% of African populations encodes for an isoleucine at position 188 of A3C has greater antiviral activity than the common form. A3H has four major haplotypes circulating in the human population with varying ability to restrict HIV-1. The viral Vif protein also has polymorphisms which affect its ability to degrade different A3H variants.

Despite the variation in A3 genes in their potency, domain composition, and ability to be antagonized by Vif, there is additional unsampled evolutionary variants of human A3 proteins that may prove to be more effective inhibitors of HIV-1. For example, not all of double Z domain combinations have been sampled by nature, and many combinations of A3 polymorphisms are unsampled. Our previous study showed that by duplicating the single A3Z2 domain protein A3C to make an A3C-A3C tandem domains protein, we could create a protein with increased antiviral activity relative to A3C single domain counterpart that was also largely resistant to degradation by HIV-1 Vif. These evolutionary-based variants of natural antiviral proteins with improved potency and/or escape from antagonism are called “super restriction factors”. Nonetheless, A3C-A3C is still not as potent as A3G, which is the most potent natural A3 protein against, HIV-1

In this study, we created novel A3 proteins by combining the single domain A3C with the single domain A3H protein in different orientations and with different natural polymorphisms. We found that these A3C/A3H double domains are potent inhibitors of HIV-1 in the absence of Vif, even more so than the most antiviral naturally found A3,

A3G. A3C/A3H hap II doubles are packaged into virions approximately 6-7x as much as A3C, likely explaining the increase in antiviral activity. Interestingly, A3C/A3H double domains do not have an increase in hypermutation activity relative to their single domain counterparts. HIV-1 Vif cannot fully degrade A3C/A3H double domains, similar to A3C-A3C. Together, this data suggests that novel A3Z2/A3Z3 domains are potent HIV-1 inhibitors.

Results

A3H/A3C chimeras are as potent or more potent than A3G

Each of the human A3s is comprised of either one or two of these conserved zinc-coordinating domains. A3H is unique because it is the only A3 with a Z3 domain. Furthermore, this Z3 domain has never been duplicated and recombined to make a double Z3 domain A3 in mammalian linages (95, 186, 187). Therefore, in order to explore the evolutionary potential of novel A3 combinations, we created synthetic tandem domain proteins consisting of one Z2 and one Z3 domain together in a single protein. These synthetic Z2-Z3 and Z3-Z2 proteins consist of A3H and the common form of $A3C_{S188}$ domains (Figure 10A). We also used two variants of A3H, hap I, the less stable and less antiviral A3H protein, and A3H hap II, the more stable and more antiviral A3H protein, and two different A3C variants, which, by themselves, are weak restrictors of HIV-1 (Figure 10A). We modeled these synthetic tandem domains after naturally found double Z2 domain A3s, A3D and A3F. We designed Z2/Z3 and Z3/Z2 double domains based on similar alignments as A3D and A3F, incorporating the naturally found short linker sequence between both domains Arg-Asn-Pro in A3D and A3F (

Figure 10A). These designed Z2/Z3 and Z3/Z2 double domains are combinations nature has yet to sample.

In order to test the antiviral activity of these proteins, expression vectors encoding these synthetic genes were transfected into 293T cells and an HIV-1 provirus expressing a reporter gene, and we performed a single-cycle infectivity assay to compare their antiviral activities. A3G was used as a positive control, as it is the most potent naturally found A3. A3G can restrict HIV-1 infection by over two orders of magnitude infection (

Figure 10B top). As previously reported, A3H hap I weakly inhibits HIV-1 and A3H hap II more potently inhibits A3H hap I but not as strongly as A3G, even at similar expression levels (

Figure 10B). Similar to A3H hap I, A3C also weakly inhibits HIV-1. As we previously reported, the antiviral activity of A3C can be increased by at least 2-fold, by creating a synthetic tandem domain, A3C-A3C, but A3C-A3C is nonetheless, still less antiviral than A3G at similar expression levels.

However, we found that A3C-A3H hap I and A3H hap I-A3C synthetic tandem domain proteins were as potent of antiviral proteins than the most active natural A3 protein, A3G (

Figure 10B top). That is, a synthetic tandem domain protein created from two A3 single domain proteins that on their own have little to no antiviral activity, produced an antiviral protein with the ability to decrease HIV-1 by two orders of magnitude. Moreover, combining A3C with the more active A3H haplotype, A3H hap II to create A3C-A3H hap II and A3H hap II-A3C, we could make antiviral proteins that are 9-fold and 11-fold, respectively, more active against HIV-1 than A3G (

Figure 10B top). This increase in antiviral activity could not be explained by increased expression levels alone since A3C and A3H hap II single domain proteins are expressed to similar levels as A3C-A3H hap II and A3H hap II-A3C (

Figure 10B bottom). Additionally, A3C-A3H hap II and A3H hap II-A3C are expressed to the same level as A3G (

Figure 10B bottom). In summary, A3 single domain proteins with weak antiviral activity can be genetically linked them together to create a new double domain A3 that are as good or better than the most potent naturally found A3, A3G.

In order to determine the dose-response of the antiviral potential from A3C-A3H hap II and A3H hap II-A3C at various expression levels, we transfected different amounts of plasmids encoding these synthetic tandem domain proteins along with a range of different levels of A3G. A3G could restrict HIV-1 to approximately 29% infectivity even at the lowest level of DNA transfected (10ng). However, both A3C-A3H hap II and A3H hapII-A3C were able to restrict HIV-1ΔVif more potently than A3G at every dose tested, which again is not explained by differences in expression levels (

Figure 10C). In summary, by creating novel double Z domains from single domains that are not very antiviral on their own, we can create a super restriction factor that is even more potent than A3G over a range of protein expression levels.

A3C/A3H chimeras are packaged better than the single domains

In order for A3s to be antiviral, A3s must get packaged into budding virions, with a direct correlation of increase in packaging to more potent A3. Therefore, we tested the ability of A3C/A3H hap II double domains to get packaged into virions via western blot. The intracellular expression levels of the naturally found A3s, A3G, A3H hap II, and A3C, are all similar to A3C-A3H hap II and A3H hap II-A3C (Figure 11 top). Both A3C and A3H hap II single domain proteins are incorporated into virions approximately in

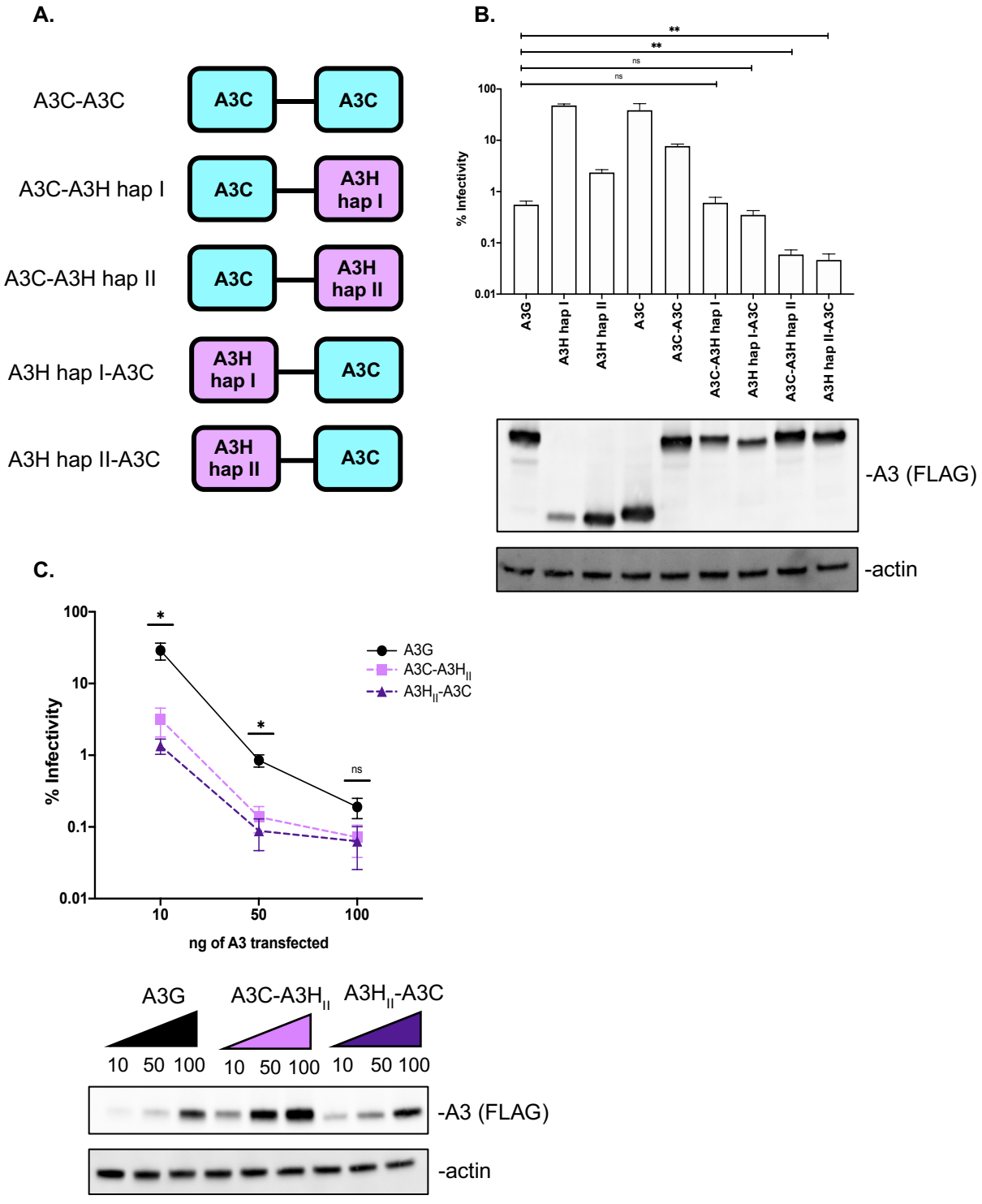

A.
A3C-A3C
A3C
A3C
A3C-A3H hap I
A3C
A3H hap I
A3C-A3H hap II
A3C
A3H hap II
A3H hap I-A3C
A3H hap I
A3C
A3H hap II-A3C
A3H hap II
A3C
B.
% Infectivity
A3G
A3H hap I
A3H hap II
A3C
A3C-A3C
A3C-A3H hap I
A3H hap I-A3C
A3C-A3H hap II
A3H hap II-A3C
-A3 (FLAG)
-actin
C.
% Infectivity
ng of A3 transfected
A3G
A3C-A3H$_{II}$
A3H$_{II}$-A3C
-A3 (FLAG)
-actin

Figure 10. A3C/A3H double domains are more potent restriction factors than A3G

(A) Cartoon schematic of A3C/A3H double domains synthesized. A3C is depicted in cyan and A3H is in purple. All double domains used in these experiments have a C-terminal 3Xflag tag for western blotting. (B) Top: Single-cycle infectivity assay measuring the percent infectivity of each A3 variant against HIV-1ΔEnvΔVif. Cells are transfected with 100ng of A3 and 600ng of HIV-1ΔEnvΔVif pseudotyped with 100ng of VSV-g. Virus production was normalized and equal amounts of virus was used infect SUPT1 cells. Results from each experiment were normalized to a no APOBEC3 control. Bar graph shows an average of 3 biological replicates, each with triplicate infections (+/- SEM). Statistical differences were determined by unpaired *t* tests: ** P≤0.01, ns= not significant. Bottom: Representative western blot of the intracellular levels of A3 in 293Ts. Antibodies to FLAG were used to detect A3s and actin was used as a loading control. (C) Top: The % infectivity of HIV-1ΔEnvΔVif pseudotyped with VSV-g and increasing doses of A3G (black, solid line), A3C-A3H hap II (light purple, dashed line), or A3H hap II-A3C (dark purple, dashed line) are plotted, normalized to a control with no A3. The amount of each A3 plasmid transfected in ng is shown on the X-axis. Data points are an average of 3 biological replicates, with each biological replicate consisting of 3 triplicate infections. Error bars show the standard deviation of the mean (+/- SEM). Statistical differences were determined by unpaired *t* tests between A3G and A3C-A3H hap II and A3G and A3H hap II-A3C: * P≤0.05, ns= not significant. Bottom: Representative western blot showing the intracellular expression levels of A3G, A3C-A3H hap II, and A3H hap II-A3C probed with anti-FLAG antibody showing intracellular expression levels for A3s and actin as a loading control. The ng of A3 transfected are denoted on top of the western blot.

similar quantities to each other, but more poorly than any of the double domain proteins (Figure 11 bottom). The synthetic tandem double domain protein A3C-A3C is packaged 4.9-fold more than A3C, which is similar to A3G (Figure 11 bottom). Similarly, A3C-A3H hap II is packaged 6.0-fold more than A3C and A3H hap II-A3C is packaged 7.6-fold more than A3C (Figure 11 bottom). However, despite this increase in packaging, it still does not explain the 650-fold increase in antiviral activity between A3C and A3C/A3H hap II double domains, nor the increased activity of A3C-A3H hap II and A3H hap II-A3C relative to A3G.

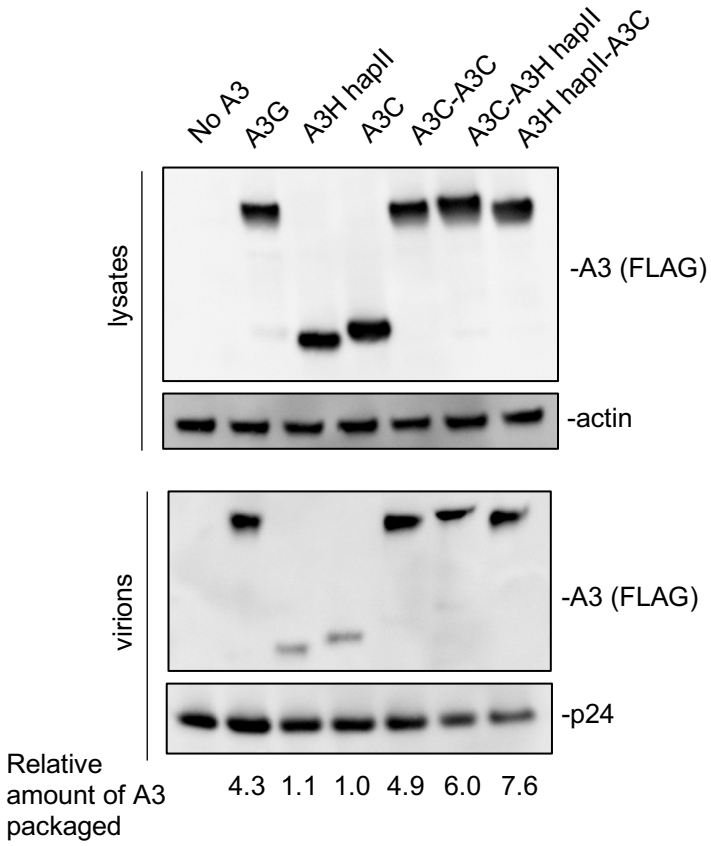


Figure 11. A3C/A3H double domains are packaged more than their single domain counterparts.

Intracellular expression and packaging of A3 into virions. HIV-1ΔEnvΔVif provirus was co-transfected into 293T cells with 100ng of each A3. Top: western blot of cellular lysates probed with anti-FLAG antibody showing intracellular expression levels for A3s and actin as a loading control. Bottom: western blot of proteins in the pelleted virions and probed with anti-FLAG antibody for A3 levels and anti-p24gag for normalization. An empty vector condition was used a negative control and labeled no A3. Western blot shown is representative of 3 biological replicates. The relative abundance of A3 in cell lysates and virions was quantified with Image Lab. Relative A3 packaged was calculated by dividing the relative abundance of A3 in the virions by the normalized levels of each A3 expressed in the cells and written below**.**

A3H/A3C chimeras use deaminase -dependent and -independent mechanisms to inhibit HIV

Naturally found A3s primarily use deaminase-dependent mechanisms to inhibit HIV-1. During the reverse transcription step of the HIV-1 life cycle, A3s act on the

opportunistic ssDNA intermediates to convert cytidines to uracils on the minus strand. As a result of the A3, this leads to G-to-A signatures in the dsDNA. A3G, the most potent A3G, has been documented to hypermutate up to 10% of guanosine in the HIV-1 genome (170). A3G primarily uses deaminase dependent methods of hypermutation to inhibit HIV-1 replication (76, 78). Previously, we found that A3C-A3C double domain proteins do not have an increase in deaminase-dependent restriction of HIV-1 compared to A3C, despite the 2-fold increase in antiviral activity (153). Instead, we found that A3C-A3C tandem domain proteins increased their antiviral activity through a gain in a cytidine deaminase independent inhibition of reverse transcription (153).

To test whether or not the large increase in antiviral activity of A3C/A3H double domains can be explained by an increase in hypermutation, we performed an A3 mediated hypermutation assay with each A3C/A3H double domain using a previously developed A3-mediated hypermutation assay to deep sequence all G-to-A mutations induced by a given A3 (153). Each sequencing read spanned the same region of HIV-1 *pol*. A “plasmid control” was used to identify PCR induced errors (Figure 12A). The “No A3” condition controlled for mutations that are induced from reverse transcriptase rather than those introduced from A3s (Figure 12B). A3G serves as a positive control since over 96% of the reads have more than 1 mutation and over 43% of the reads having more than 10 mutations. A3G also caused a distribution of reads with 2 to 9 G-to-A mutations (Figure 12C). In contrast, A3C had much fewer samples with G-to-A mutations, as demonstrated with a shift to the left in the frequency bar graph and only 12% of the reads have 2 or more mutations (Figure 12D). As expected, A3H hap II has more deaminase activity than A3H hap I (data not shown). More than 66% of the A3H

hap II reads have at least one G-to-A mutation and approximately 10% of the reads have 10 or more mutations (Figure 12E). Despite the extensive deaminase activity of A3H hap II, A3G still had more reads with higher frequency of mutations. Indeed, A3C-A3C has similar frequencies of G-to-A mutations as A3C (Figure 12F).

A3C-A3H hap II has 55% of the reads with at least one G-to-A mutation, with 38% of all reads having 2 or more mutations (Figure 12G). Interestingly, A3H hap II-A3C has approximately 30% of reads with at least one mutation (Figure 12H). However, despite the 500-fold increase in antiviral activity of A3C-A3H hap II and A3H hapII-A3C compared to A3H hap II, there is not 500-fold more hypermutation in either A3C-A3H hap II or A3H hapII-A3C. Intriguingly, the distribution of the reads with G-to-A mutations in A3C-A3H hap II and A3H hapII-A3C is shifted to the left, with more reads having fewer mutations, when compared to A3H hap II (compare Figure 12E, Figure 12G, and Figure 12H). These data suggest that the increase in antiviral activity cannot be explained by an increased rate of hypermutation.

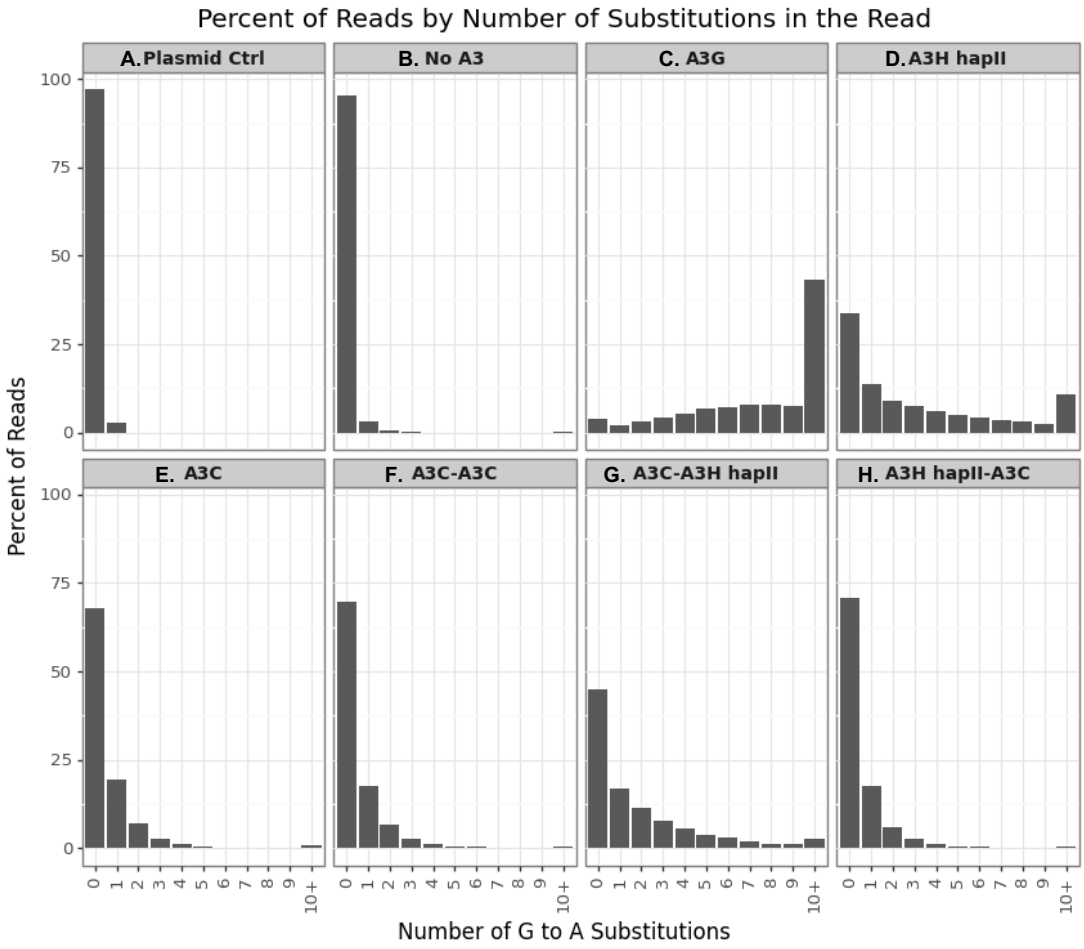


Figure 12. Frequency of G-to-A mutations does not increase in A3C/A3H hap II double domains

(A-H) Paired-end sequencing reads were analyzed for G-to-A mutations. Data is shown as frequency distribution bar graphs of the percent of reads by the number of G-to-A substitutions in each read for each A3 tested: A3G (C), A3H hap II (D), A3C (E), A3C-A3C (F), A3C-A3H hap II (G), A3H hap II-A3C (H). Plasmid control (A) is used as a sequencing control and a no APOBEC3 (B) sample was used to distinguish background mutations, including RT mutations. Frequencies are calculated as the average frequency of technical replicates and read counts are shown as the sum of the reads for both replicates.

Because the hypermutation data did not reveal an increase in deaminase activity, we next explored if A3C/A3H double domains inhibit reverse transcription in a deaminase independent manner. Previously, we found that A3C-A3C had an increase in deaminase independent antiviral activity that decreased the amount of reverse

transcriptase products after infection (153). To test this hypothesis, we transfected 293T cells using 100ng of A3 and 600ng of provirus to make A3 packaged virus and infected SUPT1 cells. 18hrs post-infection, we harvested unintegrated viral cDNA and quantified the amount late reverse transcription (RT) products normalized to the amount of virus used for each infection. Each infection was also done in biological triplicate. As expected, virus produced in the presence of A3G showed a significant decrease in relative late RT products compared to the no A3 control, while A3H hap II and A3C had more similar levels of late RT products as the no A3 control (Figure 13). At this lower concentration of A3C-A3C, we found that A3C-A3C had similar copies of late RT products as A3C (Figure 13). This difference in this finding compared to our previous results with A3C-A3C (153) finding most likely stems from the different concentration of A3 transfected during viral production since these transfections were done at levels of A3C-A3C that are barely antiviral (data not shown). Strikingly, virus made in the presence of A3C-A3H hap II or A3H hap II-A3C has even fewer accumulation of late RT products when compared to A3G (approximately 1.5 logs difference) (Figure 13). The difference in RT products between A3G and A3C/A3H hap II double domains infections closely correlates the difference in antiviral activity (Figure 10B). These results suggest that inhibition of RT is likely the major mechanism by which A3C/A3H double domains act and account for their greater antiviral activity relative to A3G. These findings support the hypothesis that A3C/A3H super restriction factors function in a novel deaminase independent mechanism compared to their single domain counterparts.

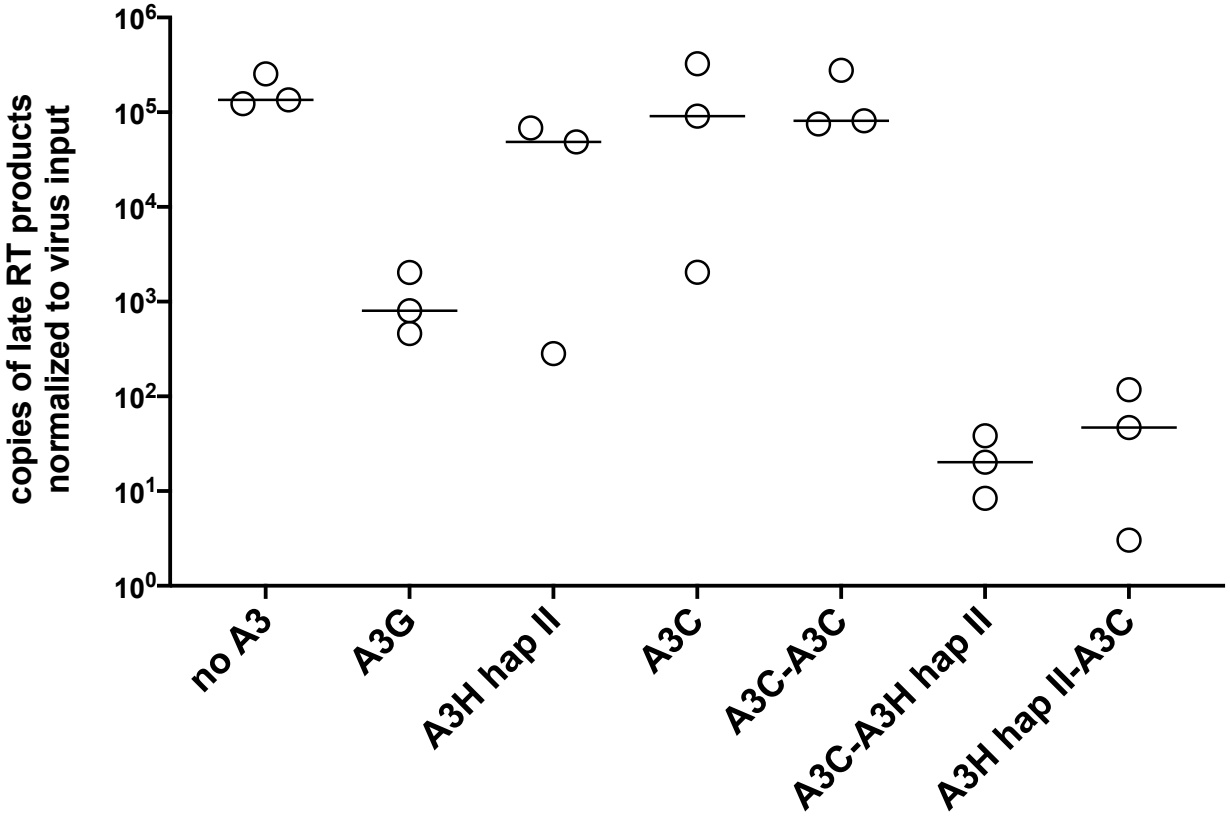


Figure 13. A3C/A3H double domains use deaminase independent mechanisms to restrict HIV-1

To evaluate the relative copies of late reverse transcription products, SUPT1 cells were infected with HIV-1ΔEnvΔvif and either no APOBEC3 or 100ng of A3 to test for inhibition of HIV-1 reverse transcription. 18 hours later, viral cDNA was harvested and the levels of HIV-1 late reverse transcription products were assayed by qPCR. Each circle represents a normalized value for the respective biological replicate, with qPCR technical duplicates. Each sample has been adjusted for equal viral infection and a nevirapine control. Bars represent the mean across 3 biological replicates.

A3C-A3H hap II forms both low molecular weight and high molecular weight complexes

The A3G oligomerization state affects its binding to ssDNA and its catalytic activity (178). A3G in lower molecular mass complexes (LMM) has the ability to deaminate ssDNA (104, 178), while A3G residing in high molecular mass (HMM) complexes is inactive, hinders rapid deamination of ssDNA, and is hypothesized to

instead form a block to inhibit reverse transcription (104, 178). Furthermore, the oligomerization state of A3G has been found to be RNA dependent (176). Based on the hypermutation data, we would expect that A3C-A3H hap II has the ability to form both LMM and HMM complexes since it retains some enzymatic activity, however, presumably largely functions in a deaminase independent mechanism. In order to test this hypothesis, we subjected cell lysates transfected with A3C-A3H hap II to velocity sedimentation, using A3G and A3C-A3C as controls.

As expected, A3G forms both LMM complexes and HMM complexes as prominent bands are present in the top and middle fractions (first 5 lanes of the western blot) (Figure 14A top). When cell lysates transfected with A3G are treated with RNase A, there is a shift in the migration pattern of A3G towards the top fractions, suggesting that the HMM complex are RNase dependent (Figure 14A bottom). We previously reported that A3C-A3C primarily forms HMM, with only a small detectable amount of protein found in the first lane and the prominent bands in lanes 2-4 (Figure 14B top). Here, we also tested if A3C-A3C oligomerization state is RNA dependent. We found that upon RNase treatment, A3C-A3C is spread across the first seven lanes as well as more A3C-A3C is found in the pellet, perhaps demonstrating that RNA is maintaining A3C-A3C in distinct complexes (Figure 14B bottom). Similar to A3G, A3C-A3H hap II forms both LMMs and HMMs, however, the distribution of the migration pattern is much larger, as A3C-A3H hap II is found in the top 7 fractions (Figure 14C top). Upon RNase treatment, A3C-A3H hap II can be found in all 11 fractions including the pellet, suggesting that similar to A3C-A3C, RNA is holding A3C-A3H hap II in distinct complexes (Figure 14C bottom). Similar to A3G, since A3C-A3H hap II is found in LMMs,

these data support the hypermutation data that there is some deaminase activity occurring, however it does not explain the entirety of the viral restriction. Furthermore, since A3C-A3H hap II forms even larger HMM complexes than A3G, which supports the hypothesis that A3C-A3H hap II acts primarily in a deaminase independent mechanism to restrict HIV-1 (Figure 12 and Figure 13).

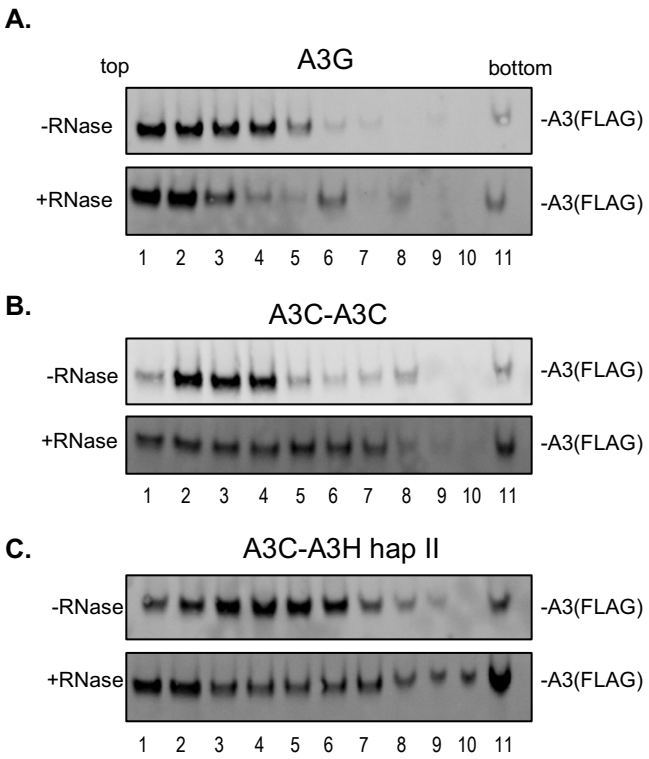


Figure 14. A3C/A3H double domains form both low molecular mass complexes and high molecular mass complexes.

(A-C) Velocity sedimentation and sucrose fractionation of APOBEC3s. Cell lysates transfected with A3 were subjected to velocity sedimentation fractionation +/- RNaseA treatment and 2.5% of each fraction was analyzed on a western blot: (A) A3G (B) A3C-A3H hap II (C) A3H hap II-A3C. Each western blot is probed with anti-FLAG to analyze A3s levels.

A3C-A3H hap II is partially resistant to HIV-1 Vif degradation

HIV-1 Vif binds to human A3C, A3D, A3F, A3G, and A3H and targets them for proteasomal degradation, thus alleviating the antiviral pressure that these A3 pose on

HIV-1 infection (167). Vif uses three interfaces to bind to antiviral A3s: one for A3G, another for A3H, and a third for A3C/A3D/A3F (86). However, we previously found that an A3C-A3C synthetic tandem domain protein was relatively resistant to Vif antagonism. To determine if Vif could target an A3 combination that it had never encountered naturally, we performed a single cycle infectivity assay with A3C-A3H hap II in the absence of HIV-1 Vif and in the presence of two HIV-1 Vifs, a patient-derived strain (patient identifier: 1203) and a lab adapted strain (LAI), that are approximately 85% identical. Vif 1203 was chosen for this experiment because previous work had shown that this Vif protein could degrade A3H hap II (109). A3G was used a positive control, as it potently restricts HIV-1ΔVif, but in the presence of a lab adapted Vif or a patient isolate Vif, Vif rescues infection to the levels of infection in the absence of A3 proteins (

Figure 15A). As expected from our previous work, the A3C-A3C combination is resistant to viral antagonism by Vif (

Figure 15A) (153). We find that the sensitivity of A3C-A3H hap II to Vif is partial. That is, in the absence of Vif, A3C-A3H hap II can restrict HIV-1 infection to 0.1% infectivity, while in the presence of either a lab adapted Vif or a patient isolate Vif, A3C-A3H hap II can still restrict HIV-1 infection to 2% (

Figure 15A). Therefore, HIV-1 Vif can degrade some of A3C-A3H hap II, but not enough to fully antagonize and remove the host defense.

The functional infectivity data of sensitivity of these A3 combinations to Vif antagonism is also supported by Western blot of the harvested cells transfected with HIV-1 provirus and A3s. A3G was used as a positive control, as it is expressed in the absence of HIV-1 Vif, but in the presence of either a patient isolate or a lab-adapted Vif, A3G is extensively degraded with only 10% of the A3G remaining (

Figure 15B). On the other hand, neither a patient-isolated or a lab-adapted HIV-1 Vif is able to fully degrade A3C-A3C, as shown with 50% of the protein band present in the western blot and quantification of the relative amount of A3 (

Figure 15B). Interestingly, A3C-A3H hap II is also able to evade Vif antagonism from both Vifs tested with 50% of the protein remaining in the presence of Vif, similar to A3C-A3C (

Figure 15B). Together, these data demonstrate that A3C-A3H hap II is a super restriction factor that evades Vif antagonism.

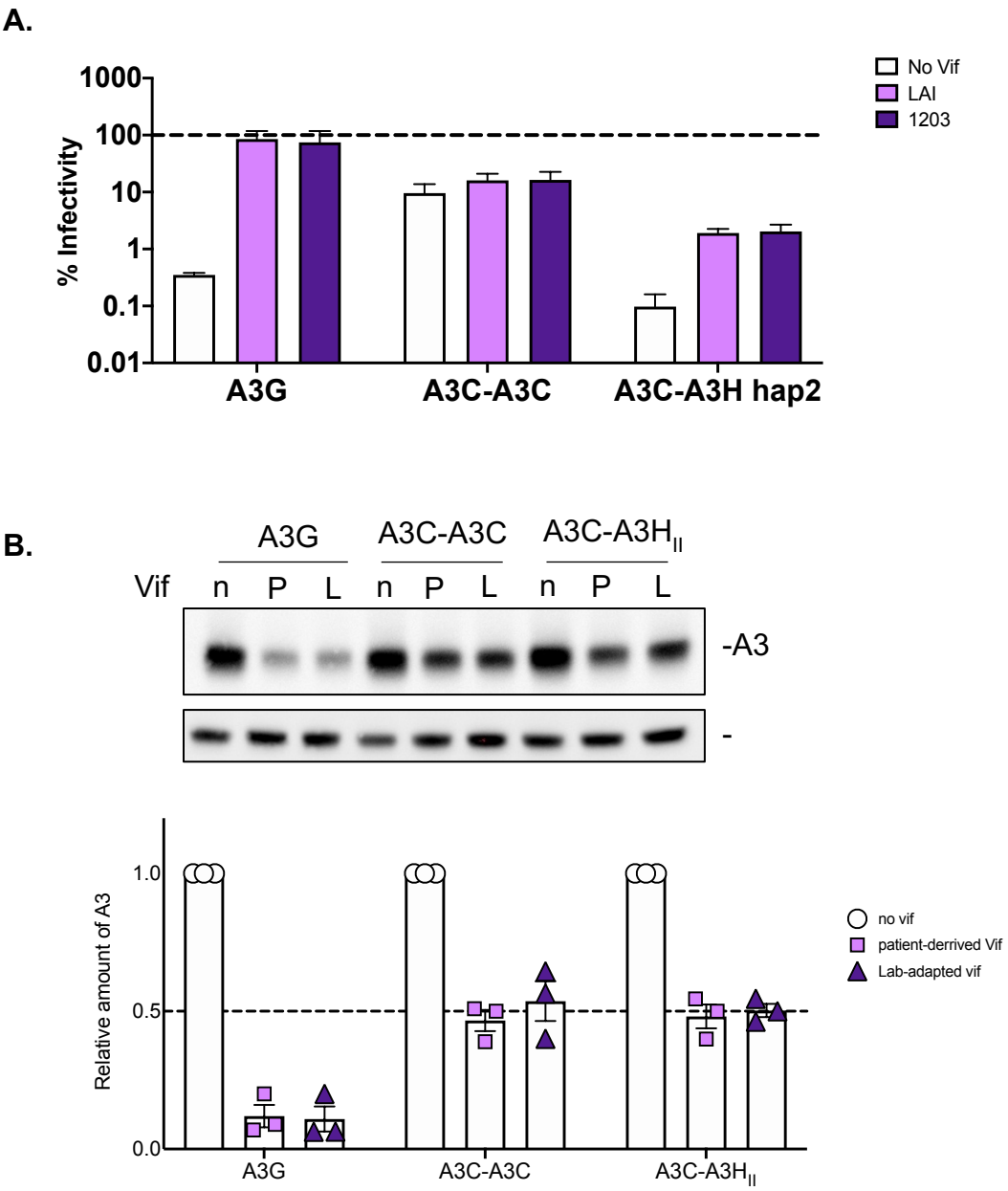


Figure 15. HIV-1 Vif cannot fully degrade A3C/A3H double domains

(A) Single-cycle infectivity assay performed in the presence of HIV-1ΔEnvΔvif (white), HIV-1ΔEnv +LAI Vif (light purple), and HIV-1ΔEnv+1203 Vif (dark purple). Cells were transfected with 100ng of A3 and 600ng of noted HIV-1ΔEnv provirus pseudotyped with 100ng of VSV-g. Virus production was normalized and equal amounts of virus was used infect SUPT1 cells. Results from each experiment were normalized to a no APOBEC3 control. Bar graph shows an average of 3 biological replicates and triplicate infections (+/- SEM). Dotted line is drawn at 100% infectivity. (B) Top: Western blot of the packaging of A3G, A3C-A3C, and A3C-A3H hap II into virions. Either HIV-1ΔEnvΔVif (no), HIV-1ΔEnv+1203 Vif (PD), or HIV-1ΔEnv+LAI Vif (LA) provirus was co-transfected into 293T cells with A3C variants. Proteins in the pelleted virions are shown in a western blot probed with anti-FLAG antibody for A3 levels and anti-p24gag as a loading control. Densitometry calculations were performed to determine the fraction of A3

remaining in the presence of Vif. Western blot shown is a representative of 3 biological replicates. Bottom: Bar graph showing quantification of relative amount of A3 protein from 3 biological replicates quantified with Image Lab. Dotted line is drawn at 0.5 and each sample is normalized to a no HIV-1 Vif control.

Discussion

Although A3G is a potent inhibitor of HIV-1 in the absence of Vif, here we show that we can create a super restriction factor with A3C/A3H double domains that are more potent than A3G. These A3C-A3H hap II and A3H hap II-A3C double domains are not only about 10-fold more antiviral against HIV-1 than A3G, but also partially resistant to HIV-1 antagonism (Figure 10 and

Figure 15). We found that A3C/A3H double domains are packaged into budding virions more than both their single domain counterparts as well as A3G, however, the increase in packaging can only partially explain the increase in antiviral activity (Figure 11). However, the ability of the A3C/A3H synthetic double domain proteins to inhibit reverse transcription after viral infection of the target cells does correlate well with their increased ability to inhibit HIV-1 (Figure 13).

Why are A3C/A3H double domains more potent than other double domains?

Previous studies have created and tested A3H double domain proteins (two A3H domains genetically linked together) and shown that they do not exhibit an increase in the antiviral activity compared to the single domain counterparts (102, 188). We recently showed that A3C-A3C proteins have a 2-fold increase in antiviral activity when compared to A3C single domain (153). However, here we show that A3C/A3H hap II double domains are over 100-fold more potent than A3C-A3C (Figure 10B). Therefore, A3C and A3H double domain proteins together are much more powerful antiviral proteins than either of the same single domains combined. These data argue that it is not just the fact that these proteins are double deaminase domain proteins that makes them more potent.

The ability of A3G to oligomerize in cells has been correlated with its antiviral activity because this oligomerization leads to increased packaging into virions (176, 177). Furthermore, the LMM fraction has been hypothesized to be the deaminase dependent form of A3G, while the HMM fraction to be the deaminase independent form. We speculate that the increased antiviral activity of A3C/A3H could be due to the ability of A3C-A3H hap II to form both high molecular weight complexes and low molecular weight complexes. Thus, since A3C-A3H hap II is has a larger percentage of protein in the LMM fraction and a much more expansive HMM fraction (spanning lanes 2-6) as well as packaged 6-fold more than A3C, this could explain why A3C/A3H double domains are better than other naturally found double domains (Figure 14).

An alternative, but not mutually exclusive hypothesis, is that A3C-A3H hap II could bind to RNA or DNA more than other double domains. This is supported from recent structural work that has shown that A3H binds to RNA to make functional dimers

(180–182). It is possible that A3C-A3H hap II packaged into virions binds either viral RNA or newly synthesized viral DNA and blocks reverse transcriptase from synthesizing full-length proviral DNA.

Why has a A3Z3 never used in a double domain A3?

Despite the potent antiviral activity of A3C/A3H double domains, nature has never recombined to form a Z3 in double domain A3. Interestingly, the amplification of *A3Z3* domains has not been sampled in nature by any mammalian groups except in Carnivora, in which duplicated *A3Z3* genes were almost entirely pseudogenized (187). These results suggest that a Z3 domain as part of a double domain protein could be selected against in nature. It is possible that these proteins are too potent to express in cells and could have alternative, harmful deaminase targets like cellular genomic DNA. A3B and A3H hap I, A3s with more nuclear localization, have been implicated in contributing to cancer, suggesting that they could be detrimental to the host (106, 107). We do see some evidence of this toxicity in that transfection of A3C-A3H constructs often result in lower yields of virus produced, although this phenomenon needs to be explored in more detail in future experiments.

Diversity within *A3* locus

The *A3* gene locus has recombined and duplicated to form the 7 *A3s* found in tandem on chromosome 22 (95, 187). This expansion of the A3 gene family has allowed for different specificities and activities for A3s. For example, A3A is a potent inhibitor of retrotransposons and retroelements (189, 190). While A3B has no activity against HIV,

A3B uses its nuclear localization to restrict dsDNA viruses like EBV and KSHV (169). A3C and A3H have sampled additional diversity through polymorphisms circulating within populations with different antiviral activity. For instance, the most common form of human A3C encodes a serine at position 188 and weakly inhibits HIV (98). There is a polymorphism, found at 10% of African populations, at position 188 that encodes for an isoleucine and is 2-fold more antiviral than the more common form of A3C (98). $A3C_{I188}$ has the ability to form dimers in solution as well as greater enzymatic activity. Additionally, A3H has 12 different haplotypes circulating in human populations. Majority of the human population encodes for A3H hap I, which is unstable, more nuclear, and weakly inhibits HIV (102, 103). However, A3H hap II encodes for a stable, more cytoplasmic, and more antiviral A3 (102). In fact, humans A3H has lost antiviral activity in two independent events to destabilize the protein, A3H hap I and A3H hap III/IV (101).

What we can learn from super restriction factors?

Through making novel tandem domain proteins that nature has not sampled, we have learned that more potent antiviral activity can be achieved with increasing packaging into budding virions, however, an increase in incorporation into virions is not enough to create an antiviral protein that is more potent than A3G. Despite these double deaminase domain proteins having twice as many catalytic domains, the increase in antiviral activity with does not come from an increase in hypermutation, but instead from deaminase independent mechanisms. Thus, our data suggest that there is an untapped mechanism of potent antiviral activity within the A3 locus that could block reverse

transcription directly rather than act through hypermutation. Lastly, these A3C/A3H double domains are partially resistant to Vif degradation. Further exploration of the ability for Vif to adapt to novel A3 targets will be interesting to determine if these super restriction factors could give rise to HIV-1 resistant cells. Together, these super restriction factors provide powerful tools to better understand prospectively evolution of this *A3* locus.

Chapter 5. Perspectives and Concluding Remarks

In my dissertation, I showed that combining novel A3 single domain proteins into double domain A3s creates super restriction factors. These super restriction factors have increased antiviral activity and are resistant to viral antagonists. By simply linking two A3C, a weak inhibitor of HIV-1, sequences together, we can improve the antiviral activity by 2-fold (Figure 4). This increase in antiviral activity is closely correlated with an increase in packaging into budding HIV-1 virions (Figure 4). Furthermore, if we inactivate the C-terminal the catalytic site, we can further improve the antiviral activity of A3C-A3C (Figure 5). We found that this increase in antiviral activity is due to a novel gain of function in deaminase independent mechanisms rather than an increase in deaminase dependent activity (Figure 7). Also, these A3C-A3C variants are resistant to viral antagonism in a single cycle infectivity assay (

Figure 9). However, the synthetic tandem domain of A3C-A3C was never as potent as A3G.

Therefore, we explored other A3 single domain combinations and found that A3C/A3H double domains are as potent or more potent than A3G (Figure 10). A3C/A3H double domains have an increase in packaging relative to their single domain counterparts and A3G, however, this increase in packaging cannot fully explain the 9 to 11-fold increase in antiviral activity when compared to A3G (Figure 11). Similar to A3C-A3C, we found that A3C/A3H double domain proteins do not have an increase in deaminase activity, but instead an increase in deaminase independent mechanisms, as seen from the dramatic reduction in late RT products (Figure 12 and Figure 13). HIV-1 Vif can partially, but not fully degrade A3C-A3H hap II (

Figure 15). Making these novel A3 double deaminase domains has provided insights about how APOBEC3 proteins function. I will use this discussion section to comment on additional questions and avenues of research.

Further characterization of A3C/A3H double domains

Here, we have shown that A3C-A3C and A3C/A3H double domains have an increase in deaminase independent inhibition of HIV-1 through quantifying the late reverse transcription products (Figure 7 and Figure 13). However, it still remains unknown where this block is occurring. Previous studies have shown that A3G is sterically hindering the process of reverse transcription through binding directly to reverse transcriptase (76). Performing CO-IPs with A3C/A3H double domains and HIV-1 reverse transcriptase will be interesting to see if A3C/A3H double domains follow a similar pattern as A3G, which serves a positive control. Additionally, it would be interesting to test if these A3C/A3H double domains inhibit early or intermediate reverse transcription products, using a RT-qPCR based assay. Alternatively, another hypothesis is that these A3C/A3H super restriction factors are binding to the nucleic acids more tightly. Making purified protein and testing both binding to DNA as well as binding to RNA in biochemistry assay would help resolve some of these questions.

A3 proteins have different localizations, but the cytoplasmic proteins have been correlated with greater antiviral activity. The best example stems from A3H polymorphisms. A3H hap I, the less stable and weakly antiviral protein, has a nuclear localization, while A3H hap II, the more stable and potent antiviral protein, has a cytoplasmic localization (102). A3C has a cell wide localization pattern (97). Therefore, it would be interesting to test if A3C/A3H double domains have a more cytoplasmic

localization through microscopy and cellular fractionation. Furthermore, there is a possibility that these super restriction factors could be toxic to cells. We have seen some evidence of this in transfection of 293T cells, which leads to lower yields of virus produced. Exploring the localization and DNA binding capability of A3C/A3H double domains could help understand if a nuclear localization and/or an increase in DNA binding is contributing to this toxicity.

Not only has an increase in packaging been correlated with cytoplasmic localization, but it has been hypothesized that there is a protein-protein interaction that could retain A3H hap II in the cytosol (102). Therefore, it would be informative to test all protein-protein interactions with mass spectrometry of super restriction factors and compare the top hits with their single domain counterparts. These comparisons would give insights to subcellular compartments and RNA granules/complexes in which these A3C/A3H double domains reside. Using RNase A treatment, we could test that the top hits of A3C/A3H double domains are most likely direct protein-protein interactions. Additionally, comparing the top protein hits of A3C/A3H hap II and A3G could provide insight as to why A3C-A3H hap II is more antiviral, as well as an understanding as to why an A3Z3 double domain has not existed in nature.

Acquiring purified A3C-A3H hap II protein not only would help elucidate the remaining biochemistry questions, but also could be used to in structural biology analyses, by crystallography. Both A3C and A3H hap II single domain proteins have been crystallized and until only recently, there has not been a full length double domain A3 structure (180–182, 191, 192). Obtaining the A3C-A3H hap II structure would help understand how both domains are interacting with each other. These finding could

complement the sucrose gradient findings and maybe help explain how higher order complexes are formed. Additionally, HIV-1 Vif is only partially able to antagonize A3C-A3H hap II (

Figure 15). Using the A3C-A3H hap II crystal structure, we can better understand the A3C-A3H hap II and Vif binding site. One hypothesis is that this Vif binding site is now partially occluded due to the folding of the two A3 domains. An alternative, but not a mutually exclusive option, is that A3C-A3H hap II is a bulkier protein and when bound to Vif, it disrupts Vif's ability to bind to other proteins like CBF-β. Crystal structures as well as HIV-1 Vif and A3C-A3H hap II CO-IP experiments are avenues to research to understand why super restriction factors cannot be fully degraded by HIV-1 Vif.

In vitro evolution experiments

There are four human A3s that have antiviral activity against HIV-1: A3D, A3F, A3G, and A3H. One single viral protein, Vif, must counteract all of these antiviral proteins to achieve maximal infectivity for the virus (193). Previous work suggests that HIV-1 Vif has evolved three separate interfaces in order to degrade A3s: one binding A3G, another one A3H, and a third able to recruit A3C/A3D/A3F (194). As shown in the single round infectivity assay in

Figure 15, both HIV-1 Vifs are able to only partially degrade A3C-A3H hap II. To follow up, I want to determine the evolutionary constraints of HIV-1 Vif to adapt to fully degrade a novel target and the subsequent consequences to target the native A3s. To do this, I would create T cell lines that stable express A3C-A3H hap II and use A3G as a positive control (195). I would challenge these cell lines to a long-term cell culture experiment to track the ability of Vif to evolve to novel A3 targets. Upon detection of super restriction factor-resistant virus, I would perform PCR amplifications of HIV-1 Vif derived from integrated proviral sequences found in the genomic DNA of infected cells. If a resistant virus arises, I plan to evaluate the ability of this adapted Vif to antagonize the naturally found A3s, with the idea that the gain of activity against A3C-A3H hap II may have come at cost of a loss of activity against other A3s. Specifically, I would passage the adapted HIV-1 virus in the presence of SUPT1 cell lines that stably express A3F, A3G, or A3H that our lab has previously created to determine if Vif is still capable of antagonizing the wild-type A3s, after gaining anti-A3C-A3H hap II activity. Together, these experiments will help prospectively explore the evolutionary potential for Vif to adapt to new antiviral specificities as a unique means to study viral evolution. Furthermore, these super restriction factors could be used to make human cells that are resistant to HIV-1 infection. Experiments where these super restriction factors are used

in cells during an infection are needed to determine whether or not HIV-1 could readily develop resistance to these novel mechanisms.

Concluding remarks

Super Restriction Factors as potential HIV therapies

It is possible that super restriction factors could serve as a scaffold to designing new therapies against HIV-1. One avenue would be genetic alteration of cells with the addition of a super restrictor gene. For example, T cell precursors are currently being altered with CCR5 deletions to make the cells resistant to HIV-1 infection. The addition of a super restrictor gene to these cells should increase their resistance to infection even further. Two potential drawbacks of this approach are the T cell mediated immune response to a novel gene product, and also the potential cytotoxicity of the super restriction factor. An alternative approach suggested by our work is perhaps a small molecule could be designed that would mimic a super restriction factor increasing the activity of one of the APOBEC3 proteins (perhaps by forced multimerization) or by inhibiting its binding to Vif. This synthetic decoy would prevent Vif from interacting with the naturally found A3s, allowing them functionally to inhibit HIV-1 without antagonism. Another targeted approach would be to increase the levels of the APOBEC3 proteins either at the transcriptional or post-transcriptional stages. Previously, attempts have been made to boost activity of all restriction factors as a therapy. This strategy involved delivering type-I interferon to induce all ISG expression in patients coinfected with HIV-1 and Hepatitis C virus. The data suggests that short-term treatment with IFN alpha could reduce HIV-1 RNA and decrease CD4+ T cell activation (196). In other experiments with

rhesus macaques infected with SIV, treatment of IFN alpha initially upregulated the expression of antiviral genes and prevented systemic infection (197). However, continued IFN alpha treatment induced IFN-I desensitization and decrease antiviral gene expression, enabling infection with increased SIV reservoir size and accelerated CD4 T-cell loss (197). Rather than stimulating all ISGs, my dissertation suggests a more targeted approach of boosting a singular restriction factor could serve as potential therapy and perhaps get around the detrimental effects of IFN-I treatments.

Why do APOBECS have enzymatic activity if they can be super potent just through blocking reverse transcription through binding?

It is interesting that the naturally found A3s restrict HIV-1 through hypermutation, but potent restriction can also be achieved by blocking reverse transcription through steric hindrance. Perhaps the naturally found A3s evolved to maintain their enzymatic activity because less A3 is needed to inhibit HIV-1 replication. Furthermore, hypermutation could cause more permanent damage than truncated reverse transcription products. A follow up experiment to evaluate if super restriction factors continue to use a deaminase-independent mechanism to restrict HIV-1 in a spreading infection context is currently being done. These results will be of interest in asking why have more potent APOBEC3 proteins that act in a cytidine-deaminase independent mechanism not evolved naturally.

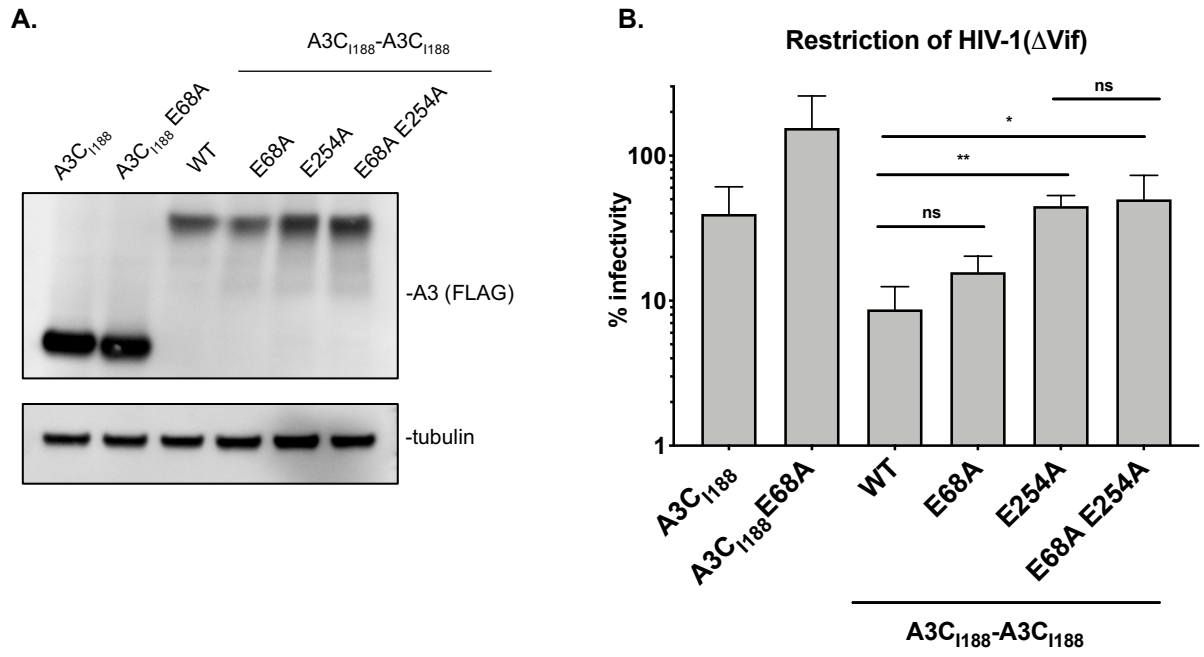
Why is it important to study restriction factors given that they do not actually inhibit HIV-1 infection in humans?

Creating super restriction factors has allowed for a retrospective understanding of how double domain A3s function. For example, through mutating the amino acids responsible for deamination in the A3C-A3C double domain, we learned that having one domain with a functional active site for deamination is better than having two functional domains. This information is useful to better understand how we can boost restriction factors to better combat HIV-1. Additionally, understanding how restriction factors work provides insight as to how viruses adapt to novel hosts.

Appendix I: Supplemental Figures

Supplemental Figure 1. Clustal Omega amino acid alignment of A3C-A3C and the closely related A3, A3F

The “RNP” amino acid sequence (Arginine-Asparagine-Proline) that links the two deaminase domains together is highlighted in yellow. The end of the first domain and the beginning of the second domain of A3C-A3C is delineated with a red arrow. The isoleucine human polymorphism in each of the domains is shown in blue text. The conserved A3 cytidine deaminase motif, His-X-Glu-X_{23-28}-Cys-Pro-X_{2-4}-Cys, is highlighted in grey. The essential glutamic acid that is necessary for deaminase activity is in green text. Asterisks indicates positions which have a conserved residue between A3C-A3C and A3F; a colon denotes conservation between groups of strongly similar properties; a period indicates conservation between groups of weakly similar properties.



Supplemental Figure 2. $A3C_{I188}$-$A3C_{I188}$ require both functioning active sites for full antiviral activity

(A) Western blot analysis of intracellular expression levels of active site point mutation in $A3C_{I188}$-$A3C_{I188}$. WT denotes wild-type $A3C_{I188}$-$A3C_{I188}$; E68A refers to a mutation in the catalytic deaminase site in either the single domain A3C or N-terminus of the synthetic tandem domain $A3C_{I188}$-$A3C_{I188}$; E254A refers to a mutation in the C-terminus of the synthetic tandem domain $A3C_{I188}$-$A3C_{I188}$; E68A E254A refers to a double catalytic deaminase site mutant in $A3C_{I188}$-$A3C_{I188}$. Antibodies to FLAG were used to detect A3s and tubulin was used as a loading control. (B) Single-cycle infectivity assay measuring the percent infectivity of each A3C variant and active site mutant against HIV-1ΔEnvΔVif. Results from each experiment were normalized to a no A3 control. Bar graph shows the mean of three biological replicates, each with triplicate infections. Error bars represent the SEM. Statistical differences were determined by unpaired *t* tests: * P≤0.05, ** P≤0.01, ns= not significant

References

1. Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, Solovyov A, Ojeda-Flores R, Arrigo NC, Islam A, Ali Khan S, Hosseini P, Bogich TL, Olival KJ, Sanchez-Leon MD, Karesh WB, Goldstein T, Luby SP, Morse SS, Mazet JAK, Daszak P, Lipkin W Ian. 2013. A Strategy To Estimate Unknown Viral Diversity in Mammals. MBio 4:1–15.
2. Littlejohn M, Locarnini S, Yuen L. 2016. Origins and Evolution of Hepatitis B Virus Hepatitis D Virus. Cold Spring Harb Perspect Med 6:a021360.
3. Suh A, Brosius J, Schmitz J, Kriegs JO. 2013. The genome of a Mesozoic paleovirus reveals the evolution of hepatitis B viruses. Nat Commun 4:1–7.
4. Gottschling M, Göker M, Stamatakis A, Bininda-Emonds ORP, Nindl I, Bravo IG. 2011. Quantifying the phylodynamic forces driving papillomavirus evolution. Mol Biol Evol 28:2101–2113.
5. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, Rambaut A, Robertson DL. 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol 1–10.
6. Caron A, Bourgarel M, Cappelle J, Liégeois F, De Nys HM, Roger F. 2018. Ebola virus maintenance: if not (Only) bats, what else? Viruses 10.
7. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. 2015. HIV infection. Nat Rev Dis Prim 1.
8. Gifford RJ. 2012. Viral evolution in deep time: Lentiviruses and mammals. Trends Genet 28:89–100.
9. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. 2004. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Rev Investig Clin 56:126–129.
10. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B, White G, Foster P, Markham PD. 1984. Frequent Detection and Isolation of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and at Risk for AIDS. Science (80-) 224:500–503.
11. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339:389–392.
12. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397:436–441.
13. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain L V., Liegeois F, Loul S, Ngole EM, Bienvenue Y, Delaporte E, Brookfield JFY, Sharp PM, Shaw GM, Peeters M, Hahn BH. 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science (80-) 313:523–526.
14. Duggal NK, Emerman M. 2012. Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol 12:687–695.
15. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, Chin CR, Feeley EM, Sims JS, Adams DJ, Wise HM, Kane L, Goulding D, Digard P, Anttila V, Baillie JK, Walsh TS, Hume DA, Palotie A, Xue Y, Colonna V, Tyler-Smith C, Dunning J,

Gordon SB, Smyth RL, Openshaw PJ, Dougan G, Brass AL, Kellam P, Everingham K, Dawson H, Hope D, Ramsay P, Campbell A, Kerr S, Harrison D, Rowan K, Addison J, Donald N, Galt S, Noble D, Taylor J, Webster N, Taylor I, Aldridge J, Dornan R, Richard C, Gilmour D, Simmons R, White R, Jardine C, Williams D, Booth M, Quasim T, Watson V, Henry P, Munro F, Bell L, Ruddy J, Cole S, Southward J, Allcoat P, Gray S, McDougall M, Matheson J, Whiteside J, Alcorn D, Rooney K, Sundaram R, Imrie G, Bruce J, McGuigan K, Moultrie S, Cairns C, Grant J, Hughes M, Murdoch C, Davidson A, Harris G, Paterson R, Wallis C, Binning S, Pollock M, Antonelli J, Duncan A, Gibson J, McCulloch C, Murphy L, Haley C, Faulkner G, Freeman T, Chaussabel D, Adamson WE, Carman WF, Thompson C, Zambon MC, Aylin P, Ashby D, Barclay WS, Brett SJ, Cookson WO, Drumright LN, Elderfield RA, Garcia-Alvarez L, Gazzard BG, Griffiths MJ, Habibi MS, Hansel TT, Herberg JA, Holmes AH, Hussell T, Johnston SL, Kon OM, Levin M, Moffatt MF, Nadel S, Warner JO, Aston SJ, Hay A, McCauley J, O'Garra A, Banchereau J, Hayward A, Kellam P, Simmonds P, McNamara PS, Semple MG, Nguyen-Van-Tam JS, Ho LP, McMichael AJ. 2012. IFITM3 restricts the morbidity and mortality associated with influenza. Nature 484:519–523.

16. Compton AA, Bruel T, Porrot F, Mallet A, Sachse M, Euvrard M, Liang C, Casartelli N, Schwartz O. 2014. IFITM proteins incorporated into HIV-1 virions impair viral fusion and spread. Cell Host Microbe 16:736–747.
17. Shi G, Schwartz O, Compton AA. 2017. More than meets the I: The diverse antiviral and cellular functions of interferon-induced transmembrane proteins. Retrovirology 14:1–11.
18. Neil SJD, Zang T, Bieniasz PD. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451:425–430.
19. Doyle T, Goujon C, Malim MH. 2015. HIV-1 and interferons: Who's interfering with whom? Nat Rev Microbiol 13:403–413.
20. Compton AA, Malik HS, Emerman M. 2013. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. Philos Trans R Soc B Biol Sci 368.
21. Perez-Caballero D, Zang T, Ebrahimi A, McNatt MW, Gregory DA, Johnson MC, Bieniasz PD. 2009. Tetherin Inhibits HIV-1 Release by Directly Tethering Virions to Cells. Cell 139:499–511.
22. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, Stephens EB, Guatelli J. 2008. The Interferon-Induced Protein BST-2 Restricts HIV-1 Release and Is Downregulated from the Cell Surface by the Viral Vpu Protein. Cell Host Microbe 3:245–252.
23. Hinz A, Miguet N, Natrajan G, Usami Y, Yamanaka H, Renesto P, Hartlieb B, McCarthy AA, Simorre JP, Göttlinger H, Weissenhorn W. 2010. Structural basis of HIV-1 tethering to membranes by the BST-2/tetherin ectodomain. Cell Host Microbe 7:314–323.
24. Venkatesh S, Bieniasz PD. 2013. Mechanism of HIV-1 Virion Entrapment by Tetherin. PLoS Pathog 9.
25. Kueck T, Neil SJD. 2012. A cytoplasmic tail determinant in HIV-1 vpu mediates targeting of tetherin for endosomal degradation and counteracts interferon-

induced restriction. PLoS Pathog 8.
26. Jia B, Serra-Moreno R, Neidermyer W, Rahmberg A, Mackey J, Fofana I Ben, Johnson WE, Westmoreland S, Evans DT. 2009. Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. PLoS Pathog 5.
27. Zhang F, Wilson SJ, Landford WC, Virgen B, Gregory D, Johnson MC, Munch J, Kirchhoff F, Bieniasz PD, Hatziioannou T. 2009. Nef Proteins from Simian Immunodeficiency Viruses Are Tetherin Antagonists. Cell Host Microbe 6:54–67.
28. Bour S, Strebel K. 1996. The human immunodeficiency virus (HIV) type 2 envelope protein is a functional complement to HIV type 1 Vpu that enhances particle release of heterologous retroviruses. J Virol 70:8285–8300.
29. Le Tortorec A, Neil SJD. 2009. Antagonism to and Intracellular Sequestration of Human Tetherin by the Human Immunodeficiency Virus Type 2 Envelope Glycoprotein. J Virol 83:11966–11978.
30. Lim ES, Malik HS, Emerman M. 2010. Ancient Adaptive Evolution of Tetherin Shaped the Functions of Vpu and Nef in Human Immunodeficiency Virus and Primate Lentiviruses. J Virol 84:7124–7134.
31. Sauter D, Schindler M, Specht A, Landford WN, Münch J, Kim KA, Votteler J, Schubert U, Bibollet-Ruche F, Keele BF, Takehisa J, Ogando Y, Ochsenbauer C, Kappes JC, Ayouba A, Peeters M, Learn GH, Shaw G, Sharp PM, Bieniasz P, Hahn BH, Hatziioannou T, Kirchhoff F. 2009. Tetherin-Driven Adaptation of Vpu and Nef Function and the Evolution of Pandemic and Nonpandemic HIV-1 Strains. Cell Host Microbe 6:409–421.
32. McNatt MW, Zang T, Hatziioannou T, Bartlett M, Fofana I Ben, Johnson WE, Neil SJD, Bieniasz PD. 2009. Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. PLoS Pathog 5:9–12.
33. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Ségéral E, Yatim A, Emiliani S, Schwartz O, Benkirane M. 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature 474:654–657.
34. Tristem M, Purvis A, Quicke DLJ. 1998. Complex evolutionary history of primate lentiviral vpr genes. Virology 240:232–237.
35. Lim ES, Fregoso OI, McCoy CO, Matsen FA, Malik HS, Emerman M. 2012. The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. Cell Host Microbe 11:194–204.
36. Etienne L, Hahn BH, Sharp PM, Matsen FA, Emerman M. 2013. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. Cell Host Microbe 14:85–92.
37. Ahn J, Hao C, Yan J, DeLucia M, Mehrens J, Wang C, Gronenborn AM, Skowronski J. 2012. HIV/Simian Immunodeficiency Virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex CRL4 DCAF1. J Biol Chem 287:12550–12558.
38. Fregoso OI, Ahn J, Wang C, Mehrens J, Skowronski J, Emerman M. 2013. Evolutionary Toggling of Vpx/Vpr Specificity Results in Divergent Recognition of the Restriction Factor SAMHD1. PLoS Pathog 9.
39. White TE, Brandariz-Nuñez A, Carlos Valle-Casuso J, Amie S, Nguyen L, Kim B,

Brojatsch J, Diaz-Griffero F. 2013. Contribution of SAM and HD domains to retroviral restriction mediated by human SAMHD1. Virology 436:81–90.
40. Carthagena L, Parise MC, Ringeard M, Chelbi-Alix MK, Hazan U, Nisole S. 2008. Implication of TRIMalpha and TRIMCyp in interferon-induced anti-retroviral restriction activities. Retrovirology 5:1–15.
41. Sakuma R, Mael AA, Ikeda Y. 2007. Alpha Interferon Enhances TRIM5α-Mediated Antiviral Activities in Human and Rhesus Monkey Cells. J Virol 81:10201–10206.
42. Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, Diaz-Griffero F, Anderson DJ, Sundquist WI, Sodroski J. 2006. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5α restriction factor. Proc Natl Acad Sci U S A 103:5514–5519.
43. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. 2004. The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. Nature 427:848–853.
44. Kaiser SM, Malik HS, Emerman M. 2007. Restriction of an Extinct Retrovirus by the Human TRIM5a Antiviral Protein. Science (80-) 316:1756–1759.
45. Hatziioannou T, Perez-Caballero D, Yang A, Cowan S, Bieniasz PD. 2004. Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5α. Proc Natl Acad Sci U S A 101:10774–10779.
46. Keckesova Z, Ylinen LMJ, Towers GJ. 2004. The human and African green monkey TRIM5α genes encode Ref1 and Lv1 retroviral restriction factor activities. Proc Natl Acad Sci U S A 101:10780–10785.
47. Perron MJ, Stremlau M, Song B, Ulm W, Mulligan RC, Sodroski J. 2004. TRIM5α mediates the postentry block to N-tropic murine leukemia viruses in human cells. Proc Natl Acad Sci U S A 101:11827–11832.
48. Yap MW, Nisole S, Lynch C, Stoye JP. 2004. Trim5α protein restricts both HIV-1 and murine leukemia virus. Proc Natl Acad Sci U S A 101:10786–10791.
49. Kirmaier A, Wu F, Newman RM, Hall LR, Morgan JS, O'Connor S, Marx PA, Meythaler M, Goldstein S, Buckler-White A, Kaur A, Hirsch VM, Johnson WE. 2010. TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. PLoS Biol 8:35–36.
50. Yap MW, Nisole S, Stoye JP. 2005. A Single Amino Acid Change in the SPRY Domain of Human Trim5α Leads to HIV-1 Restriction. Curr Biol 15:73–78.
51. Sawyer SL, Wu LI, Emerman M, Malik HS. 2005. Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci U S A 102:2832–2837.
52. Stremlau M, Perron M, Welikala S, Sodroski J. 2005. Species-Specific Variation in the B30.2(SPRY) Domain of TRIM5α Determines the Potency of Human Immunodeficiency Virus Restriction. J Virol 79:3139–3145.
53. Braaten D, Franke EK, Luban J. 1996. Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodeficiency viruses. J Virol 70:4220–4227.
54. Newman RM, Hall L, Kirmaier A, Pozzi LA, Pery E, Farzan M, O'Neil SP, Johnson

W. 2008. Evolution of a TRIM5-CypA splice isoform in old world monkeys. PLoS Pathog 4:1–8.
55. Liao CH, Kuang YQ, Liu HL, Zheng YT, Su B. 2007. A novel fusion gene, TRIM5-cyclophilin a in the pig-tailed macaque determines its susceptibility to HIV-1 infection. AIDS 21:19–26.
56. Brennan G, Kozyrev Y, Hu SL. 2008. TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis. Proc Natl Acad Sci U S A 105:3569–3574.
57. Sayah DM, Sokolskaja E, Berthoux L, Luban J. 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature 430:569–573.
58. Nisole S, Lynch C, Stoye JP, Yap MW. 2004. A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. Proc Natl Acad Sci U S A 101:13324–13328.
59. Virgen CA, Kratovac Z, Bieniasz PD, Hatziioannou T. 2008. Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. Proc Natl Acad Sci U S A 105:3563–3568.
60. Wilson SJ, Webb BLJ, Ylinen LMJ, Verschoor E, Heeney JL, Towers GJ. 2008. Independent evolution of an antiviral TRIMCyp in rhesus macaques. Proc Natl Acad Sci U S A 105:3557–3562.
61. Malfavon-Borja R, Wu LI, Emerman M, Malik HS. 2013. Birth, decay, and reconstruction of an ancient TRIMCyp gene fusion in primate genomes. Proc Natl Acad Sci U S A 110.
62. Hatziioannou T, Princiotta M, Piatak M, Yuan F, Zhang F, Lifson JD, Bieniasz PD. 2006. Generation of simian-tropic HIV-1 by restriction factor evasion. Science (80-) 314:95.
63. Kamada K, Igarashi T, Martin MA, Khamsri B, Hatcho K, Yamashita T, Fujita M, Uchiyama T, Adachi A. 2006. Generation of HIV-1 derivatives that productively infect macaque monkey lymphoid cells. Proc Natl Acad Sci U S A 103:16959–16964.
64. Ohainle M, Kim K, Komurlu Keceli S, Felton A, Campbell E, Luban J, Emerman M. 2020. TRIM34 restricts HIV-1 and SIV capsids in a TRIM5α-dependent manner. PLOS Pathog 16:e1008507.
65. Sawyer SL, Emerman M, Malik HS. 2007. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. PLoS Pathog 3:1918–1929.
66. Li X, Gold B, O'hUigin C, Diaz-Griffero F, Song B, Si Z, Li Y, Yuan W, Stremlau M, Mische C, Javanbakht H, Scally M, Winkler C, Dean M, Sodroski J. 2007. Unique features of TRIM5α among closely related human TRIM family members. Virology 360:419–433.
67. Li X, Yeung DF, Fiegen AM, Sodroski J. 2011. Determinants of the higher order association of the restriction factor TRIM5α and other tripartite motif (TRIM) proteins. J Biol Chem 286:27959–27970.
68. von Schwedler U, Song J, Aiken C, Trono D. 1993. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. J Virol 67:4945–4955.
69. Gabuzda DH, Lawrence K, Langhoff E, Terwilliger E, Dorfman T, Haseltine WA,

Sodroski J. 1992. Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. J Virol 66:6489–6495.
70. Madani N, Kabat D. 1998. An Endogenous Inhibitor of Human Immunodeficiency Virus in Human Lymphocytes Is Overcome by the Viral Vif Protein. J Virol 72:10251–10255.
71. Simon JHM, Gaddis NC, Fouchier RAM, Malim MH. 1998. Evidence for a newly discovered cellular anti-HIV-1 phenotype. Nat Med 4:1397–1400.
72. Sheehy AM, Gaddis NC, Choi JD, Malim MH. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646–650.
73. Yu Q, König R, Pillai S, Chiles K, Kearney M, Palmer S, Richman D, Coffin JM, Landau NR, Upon A. 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol 11:435–442.
74. Chelico L, Pham P, Calabrese P, Goodman MF. 2006. APOBEC3G DNA deaminase acts processively 3′ → 5′ on single-stranded DNA. Nat Struct Mol Biol 13:392–399.
75. Ara A, Love RP, Chelico L. 2014. Different Mutagenic Potential of HIV-1 Restriction Factors APOBEC3G and APOBEC3F Is Determined by Distinct Single-Stranded DNA Scanning Mechanisms. PLoS Pathog 10.
76. Pollpeter D, Parsons M, Sobala AE, Coxhead S, Lang RD, Bruns AM, Papaioannou S, Mcdonnell JM, Apolonia L, Chowdhury JA, Horvath CM, Malim MH. 2018. Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. Nat Microbiol 3.
77. Bishop KN, Verma M, Kim EY, Wolinsky SM, Malim MH. 2008. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS Pathog 4:13–20.
78. Newman ENC, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM. 2005. Antiviral Function of APOBEC3G Can Be Dissociated from Cytidine Deaminase Activity. Curr Biol 15:166–170.
79. Holmes RK, Koning FA, Bishop KN, Malim MH. 2007. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation: Comparisons with APOBEC3G. J Biol Chem 282:2587–2595.
80. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS. 2011. Human and Rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H Demonstrate a Conserved Capacity To Restrict Vif-Deficient HIV-1. J Virol 85:11220–11234.
81. 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 (80-) 302:1056–1060.
82. Anderson BD, Harris RS. 2015. Transcriptional regulation of APOBEC3 antiviral immunity through the CBF-β/RUNX axis. Sci Adv 1:e1500296.
83. Kim DY, Gross JD. 2017. CBFß and HIV Infection. Adv Exp Med Biol 962:139–147.
84. Binning JM, Smith AM, Hultquist JF, Craik CS, Caretta Cartozo N, Campbell MG, Burton L, La Greca F, McGregor MJ, Ta HM, Bartholomeeusen K, Peterlin BM, Krogan NJ, Sevillano N, Cheng Y, Gross JD. 2018. Fab-based inhibitors reveal ubiquitin independent functions for HIV Vif neutralization of APOBEC3 restriction

factors. PLoS Pathog 14:1–19.
85. Jäger S, Kim DY, Hultquist JF, Shindo K, Larue RS, Kwon E, Li M, Anderson BD, Yen L, Stanley D, Mahon C, Kane J, Franks-Skiba K, Cimermancic P, Burlingame A, Sali A, Craik CS, Harris RS, Gross JD, Krogan NJ. 2012. Vif hijacks CBF-β to degrade APOBEC3G and promote HIV-1 infection. Nature 481:371–375.
86. Desimmie BA, Delviks-Frankenberrry KA, Burdick RC, Qi D, Izumi T, Pathak VK. 2014. Multiple APOBEC3 restriction factors for HIV-1 and one vif to rule them all. J Mol Biol 426:1220–1245.
87. Han GZ, Worobey M. 2012. Endogenous lentiviral elements in the weasel family (mustelidae). Mol Biol Evol 29:2905–2908.
88. Compton AA, Hirsch VM, Emerman M. 2012. The Host Restriction Factor APOBEC3G and Retroviral Vif Protein Coevolve due to Ongoing Genetic Conflict. Cell Host Microbe 11:91–98.
89. Compton AA, Emerman M. 2013. Convergence and Divergence in the Evolution of the APOBEC3G-Vif Interaction Reveal Ancient Origins of Simian Immunodeficiency Viruses. PLoS Pathog 9.
90. VandeWoude S, Apetrei C. 2006. Going wild: Lessons from naturally occurring T-lymphotropic lentiviruses. Clin Microbiol Rev 19:728–762.
91. Allan JS, Short M, Taylor ME, Su S, Hirsch VM, Johnson PR, Shaw GM, Hahn BH. 1991. Species-specific diversity among simian immunodeficiency viruses from African green monkeys. J Virol 65:2816–2828.
92. Etienne L, Bibollet-Ruche F, Sudmant PH, Wu LI, Hahn BH, Emerman M. 2015. The Role of the Antiviral APOBEC3 Gene Family in Protecting Chimpanzees against Lentiviruses from Monkeys. PLoS Pathog 11:1–24.
93. Binning JM, Chesarino NM, Emerman M, Gross JD. 2019. Structural Basis for a Species-Specific Determinant of an SIV Vif Protein toward Hominid APOBEC3G Antagonism. Cell Host Microbe 26:739-747.e4.
94. Simon V, Bloch N, Landau NR. 2015. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat Immunol 16:546–553.
95. Münk C, Willemsen A, Bravo IG. 2012. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC Evol Biol 12:71.
96. LaRue RS, Andresdottir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jonsson SR, Landau NR, Lochelt M, Malik HS, Malim MH, Munk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu X-F, Yuhki N, Harris RS. 2008. Guidelines for Naming Nonprimate APOBEC3 Genes and Proteins. J Virol 83:494–497.
97. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS. 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol 85:11220–34.
98. Wittkopp CJ, Adolph MB, Wu LI, Chelico L, Emerman M. 2016. A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses. PLoS Pathog 12:1–24.
99. Duggal NK, Malik H, Emerman M. 2011. The Breadth of Antiviral Activity of Apobec3DE in Chimpanzees Has Been Driven by Positive Selection. J Virol

85:11361–11371.
100. Aguiar RS, Lovsin N, Tanuri A, Peterlin BM. 2008. Vpr.A3A Chimera Inhibits HIV Replication. J Biol Chem 283:2518–2525.
101. Ohainle M, Kerns JA, Li MMH, Malik HS, Emerman M. 2008. Antiretroelement Activity of APOBEC3H Was Lost Twice in Recent Human Evolution. Cell Host Microbe 4:249–259.
102. Li MMH, Emerman M. 2011. Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. J Virol 85:8197–8207.
103. Ohainle M, Kerns J a, Malik HS, Emerman M. 2006. Adaptive Evolution and Antiviral Activity of the Conserved Mammalian Cytidine Deaminase APOBEC3H. J Virol 80:3853–3862.
104. Soros VB, Yonemoto W, Greene WC. 2007. Newly Synthesized APOBEC3G Is Incorporated into HIV Virions , Inhibited by HIV RNA , and Subsequently Activated by RNase H. PLOS Pathog 3.
105. Feng Y, Baig TT, Love RP, Chelico L. 2014. Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. Front Microbiol 5:1–23.
106. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, Feng Y, Adolph MB, Chelico L, Law EK, Carpenter MA, Harris RS. 2016. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. Nat Commun 7:12918.
107. Burns MB, Temiz NA, Harris RS. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet 45:977–983.
108. McLaughlin RN, Gable JT, Wittkopp CJ, Emerman M, Malik HS. 2016. Conservation and Innovation of APOBEC3A Restriction Functions during Primate Evolution. Mol Biol Evol 33:1889–1901.
109. Li MMH, Wu LI, Emerman M. 2010. The Range of Human APOBEC3H Sensitivity to Lentiviral Vif Proteins. J Virol 84:88–95.
110. Sawyer SL, Emerman M, Malik HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol 2.
111. Ooms M, Brayton B, Letko M, Maio SM, Pilcher CD, Hecht FM, Barbour JD, Simon V. 2013. HIV-1 Vif adaptation to human APOBEC3H haplotypes. Cell Host Microbe 14:411–421.
112. Ebrahimi D, Richards CM, Carpenter MA, Wang J, Ikeda T, Becker JT, Cheng AZ, Mccann JL, Shaban NM, Salamango DJ, Starrett GJ, Lingappa JR, Yong J, Brown WL, Harris RS. 2018. Genetic and mechanistic basis for APOBEC3H alternative splicing, retrovirus restriction, and counteraction by HIV-1 protease. Nat Commun 9:1–11.
113. Adolph MB, Ara A, Feng Y, Wittkopp CJ, Emerman M, Fraser JS, Chelico L. 2017. Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. Nucleic Acids Res 45:3378–3394.
114. Anderson BD, Ikeda T, Moghadasi SA, Martin AS, Brown WL, Harris RS. 2018. Natural APOBEC3C variants can elicit differential HIV - 1 restriction activity. Retrovirology 15:1–11.
115. Goujon C, Moncorgé O, Bauby H, Doyle T, Ward CC, Schaller T, Hué S, Barclay WS, Schulz R, Malim MH. 2013. Human MX2 is an interferon-induced post-entry

inhibitor of HIV-1 infection. Nature 502:559–562.
116. Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, Cen S, Guo F, Liang C. 2013. The interferon-inducible MxB protein inhibits HIV-1 infection. Cell Host Microbe 14:398–410.
117. Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, Schoggins JW, Rice CM, Yamashita M, Hatziioannou T, Bieniasz PD. 2013. MX2 is an interferon-induced inhibitor of HIV-1 infection. Nature 502:563–566.
118. Ohainle M, Helms L, Vermeire J, Roesch F, Humes D, Basom R, Delrow JJ, Overbaugh J, Emerman M. 2018. A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV. Elife 7:1–32.
119. Mitchell PS, Young JM, Emerman M, Malik HS. 2015. Evolutionary Analyses Suggest a Function of MxB Immunity Proteins Beyond Lentivirus Restriction. PLoS Pathog 11:1–21.
120. Busnadiego I, Kane M, Rihn SJ, Preugschas HF, Hughes J, Blanco-Melo D, Strouvelle VP, Zang TM, Willett BJ, Boutell C, Bieniasz PD, Wilson SJ. 2014. Host and Viral Determinants of Mx2 Antiretroviral Activity. J Virol 88:7738–7752.
121. Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD. 2013. Extreme Genetic Fragility of the HIV-1 Capsid. PLoS Pathog 9.
122. Schmidt F, Keele BF, Del Prete GQ, Voronin D, Fennessey CM, Soll S, Kane M, Raymond A, Gifford RJ, KewalRamani V, Lifson JD, Bieniasz PD, Hatziioannou T. 2019. Derivation of simian tropic HIV-1 infectious clone reveals virus adaptation to a new host. Proc Natl Acad Sci U S A 116:10504–10509.
123. Li M, Kao E, Gao X, Sandig H, Limmer K, Pavon-Eternod M, Jones TE, Landry S, Pan T, Weitzman MD, David M. 2012. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature 491:125–128.
124. Stabell AC, Hawkins J, Li M, Gao X, David M, Press WH, Sawyer SL. 2016. Non-human Primate Schlafen11 Inhibits Production of Both Host and Viral Proteins. PLoS Pathog 12:1–19.
125. Gao G, Guo X, Goff SP. 2002. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Science (80-) 297:1703–1706.
126. Bick MJ, Carroll J-WN, Gao G, Goff SP, Rice CM, MacDonald MR. 2003. Expression of the Zinc-Finger Antiviral Protein Inhibits Alphavirus Replication. J Virol 77:11555–11562.
127. Kerns JA, Emerman M, Malik HS. 2008. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. PLoS Genet 4:0150–0158.
128. Takata MA, Gonçalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, Bieniasz PD. 2017. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature 550:124–127.
129. Zhu Y, Chen G, Lv F, Wang X, Ji X, Xu Y, Sun J, Wu L, Zheng YT, Gao G. 2011. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. Proc Natl Acad Sci U S A 108:15834–15839.
130. Muller S, Moller P, Bick MJ, Wurr S, Becker S, Gunther S, Kummerer BM. 2007. Inhibition of Filovirus Replication by the Zinc Finger Antiviral Protein. J Virol

81:2391–2400.
131. Guo X, Carroll J-WN, MacDonald MR, Goff SP, Gao G. 2004. The Zinc Finger Antiviral Protein Directly Binds to Specific Viral mRNAs through the CCCH Zinc Finger Motifs. J Virol 78:12781–12787.
132. Guo X, Ma J, Sun J, Gao G. 2007. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. Proc Natl Acad Sci U S A 104:151–156.
133. Ficarelli M, Wilson H, Galão RP, Mazzon M, Antzin-Anduetza I, Marsh M, Neil SJD, Swanson CM. 2019. KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. Elife 8:1–25.
134. Charron G, Li MMH, MacDonald MR, Hang HC. 2013. Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. Proc Natl Acad Sci U S A 110:11085–11090.
135. Schwerk J, Soveg FW, Ryan AP, Thomas KR, Hatfield LD, Ozarkar S, Forero A, Kell AM, Roby JA, So L, Hyde JL, Gale M, Daugherty MD, Savan R. 2019. RNA-binding protein isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution functions. Nat Immunol 20:1610–1620.
136. Kmiec D, Nchioua R, Sherrill-Mix S, Stürzel CM, Heusinger E, Braun E, Gondim MVP, Hotter D, Sparrer KMJ, Hahn BH, Sauter D, Kirchhoff F. 2020. CpG Frequency in the 5′ Third of the *env* Gene Determines Sensitivity of Primary HIV-1 Strains to the Zinc-Finger Antiviral Protein. MBio 11.
137. Lu J, Pan Q, Rong L, Liu S-L, Liang C. 2011. The IFITM Proteins Inhibit HIV-1 Infection. J Virol 85:2126–2137.
138. Ahi YS, Yimer D, Shi G, Majdoul S, Rahman K, Rein A, Compton AA. 2020. IFITM3 reduces retroviral envelope abundance and function and is counteracted by glycogag. MBio 11:1–15.
139. Wilkins J, Zheng YM, Yu J, Liang C, Liu SL. 2016. Nonhuman primate IFITM proteins are potent inhibitors of HIV and SIV. PLoS One 11:1–19.
140. Compton AA, Roy N, Porrot F, Billet A, Casartelli N, Yount JS, Liang C, Schwartz O. 2016. Natural mutations in IFITM 3 modulate post-translational regulation and toggle antiviral specificity. EMBO Rep 17:1657–1671.
141. Daugherty MD, Schaller AM, Geballe AP, Malik HS. 2016. Evolution-guided functional analyses reveal diverse antiviral specificities encoded by ifit1 genes in mammals. Elife 5:1–22.
142. Rosa A, Chande A, Ziglio S, De Sanctis V, Bertorelli R, Goh SL, McCauley SM, Nowosielska A, Antonarakis SE, Luban J, Santoni FA, Pizzato M. 2015. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. Nature 526:212–217.
143. Usami Y, Wu Y, Göttlinger HG. 2015. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. Nature 526:218–223.
144. Usami Y, Wu Y, Go HG. 2015. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. Nature 526:218–223.
145. Murrell B, Vollbrecht T, Guatelli J, Wertheim JO. 2016. The Evolutionary Histories of Antiretroviral Proteins SERINC3 and SERINC5 Do Not Support an Evolutionary Arms Race in Primates. J Virol 90:8085–8089.
146. Beitari S, Ding S, Pan Q, Finzi A, Liang C. 2017. Effect of HIV-1 Env on SERINC5

Antagonism. J Virol 91:1–13.
147. Li S, Ahmad I, Shi J, Wang B, Yu C, Zhang L, Zheng Y-H. 2018. Murine Leukemia Virus Glycosylated Gag Reduces Murine SERINC5 Protein Expression at Steady-State Levels via the Endosome/Lysosome Pathway to Counteract SERINC5 Antiretroviral Activity. J Virol 93:1–15.
148. Yurkovetskiy L, Guney MH, Kim K, Goh SL, McCauley S, Dauphin A, Diehl WE, Luban J. 2018. Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. Nat Microbiol 3:1354–1361.
149. Chougui G, Munir-matloob S, Matkovic R, Martin MM, Morel M, Lahouassa H, Leduc M, Ramirez BC, Etienne L, Margottin-goguet F. 2018. HIV-2 / SIV viral protein X counteracts HUSH repressor complex. Nat Microbiol 3.
150. Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassik MC, Wysocka J. 2018. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. Nature 553:228–232.
151. Robbez-Masson L, Tie CHC, Conde L, Tunbak H, Husovsky C, Tchasovnikarova IA, Timms RT, Herrero J, Lehner PJ, Rowe HM. 2018. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res 28:836–845.
152. Colón-Thillet R, Hsieh E, Graf L, McLaughlin RN, Young JM, Kochs G, Emerman M, Malik HS. 2019. Combinatorial mutagenesis of rapidly evolving residues yields super-restrictor antiviral proteins. PLOS Biol 17:e3000181.
153. McDonnell MM, Crawford KHD, Dingens AS, Bloom JD, Emerman M. 2020. APOBEC3C Tandem Domain Proteins Create Super Restriction Factors Againt HIV-1. MBio 11:1–18.
154. Verhelst J, Hulpiau P, Saelens X. 2013. Mx Proteins: Antiviral Gatekeepers That Restrain the Uninvited. Microbiol Mol Biol Rev 77:551–566.
155. Toohey K, Wehrly K, Nishio J, Perryman S, Chesebro B. 1995. Human Immunodeficiency Virus Envelope V1 and V2 Regions Influence Replication Efficiency in Macrophages by Affecting Virus Spread. Virology 213:70–79.
156. Wehrly K, Chesebro B. 1997. P24 Antigen Capture Assay for Quantification of Human Immunodeficiency Virus Using Readily Available Inexpensive Reagents. Methods A Companion to Methods Enzymol 12:288–293.
157. Vermeire J, Naessens E, Vanderstraeten H, Landi A, Iannucci V, van Nuffel A, Taghon T, Pizzato M, Verhasselt B. 2012. Quantification of Reverse Transcriptase Activity by Real-Time PCR as a Fast and Accurate Method for Titration of HIV, Lenti- and Retroviral Vectors. PLoS One 7.
158. Garcia EI, Emerman M. 2018. Recurrent Loss of APOBEC3H Activity during Primate Evolution. J Virol 92:1–18.
159. Doud MB, Bloom JD. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses 8.
160. Ara A, Love RP, Follack TB, Ahmed KA, Adolph MB, Chelico L. 2017. Mechanism of Enhanced HIV Restriction by Virion Coencapsidated Cytidine Deaminases APOBEC3F and APOBEC3G. J Virol 91:1–19.
161. Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. BMC Bioinformatics 16:1–13.

162. Gorodkin J, Heyer LJ, Brunak S, Storomo GD. 1997. Displaying the in formation contents of structural RNA alignments: the structure logos. Bioinformatics 13:583–586.
163. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190.
164. De Iaco A, Luban J. 2011. Inhibition of HIV-1 infection by TNPO3 depletion is determined by capsid and detectable after viral cDNA enters the nucleus. Retrovirology 8:98.
165. Butler SL, Hansen MST, Bushman FD. 2001. A quantitative assay for HIV DNA integration in vivo. Nat Med 7:631–634.
166. Reed JC, Molter B, Geary CD, McNevin J, McElrath J, Giri S, Klein KC, Lingappa JR. 2012. HIV-1 Gag co-opts a cellular complex containing DDX6, a helicase that facilitates capsid assembly. J Cell Biol 198:439–456.
167. Harris RS, Dudley JP. 2015. APOBECs and virus restriction. Virology 479–480:131–145.
168. 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 85:7594–7602.
169. Cheng AZ, Yockteng-Melgar J, Jarvis MC, Malik-Soni N, Borozan I, Carpenter MA, McCann JL, Ebrahimi D, Shaban NM, Marcon E, Greenblatt J, Brown WL, Frappier L, Harris RS. 2019. Epstein–Barr virus BORF2 inhibits cellular APOBEC3B to preserve viral genome integrity. Nat Microbiol 4:78–88.
170. Gillick K, Pollpeter D, Phalora P, Kim E-Y, Wolinsky SM, Malim MH. 2013. Suppression of HIV-1 Infection by APOBEC3 Proteins in Primary Human CD4 + T Cells Is Associated with Inhibition of Processive Reverse Transcription as Well as Excessive Cytidine Deamination. J Virol 87:1508–1517.
171. Iwatani Y, Chan DSB, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG. 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res 35:7096–7108.
172. Hüttenhain R, Xu J, Burton LA, Gordon DE, Hultquist JF, Johnson JR, Satkamp L, Hiatt J, Rhee DY, Baek K, Crosby DC, Frankel AD, Marson A, Harper JW, Alpi AF, Schulman BA, Gross JD, Krogan NJ. 2019. ARIH2 Is a Vif-Dependent Regulator of CUL5-Mediated APOBEC3G Degradation in HIV Infection. Cell Host Microbe 86–99.
173. Hache G, Liddament MT, Harris RS. 2005. The Retroviral Hypermutation Specificity of APOBEC3F and APOBEC3G Is Governed by the C-terminal DNA Cytosine Deaminase Domain. J Biol Chem 280:10920–10924.
174. Navarro F, Bollman B, Chen H, König R, Yu Q, Chiles K, Landau NR. 2005. Complementary function of the two catalytic domains of APOBEC3G. Virology 333:374–386.
175. Doud MB, Bloom JD. 2016. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. Viruses 8:1–17.
176. Thielen BK, Klein KC, Walker LW, Rieck M, Buckner JH, Tomblingson GW, Lingappa JR. 2007. T Cells Contain an RNase-Insensitive Inhibitor of APOBEC3G

Deaminase Activity. PLOS Pathog 3:1320–1334.
177. Huthoff H, Autore F, Gallois-Montbrun S, Fraternali F, Malim MH. 2009. RNA-dependent oligomerization of APOBEC3G Is required for restriction of HIV-1. PLoS Pathog 5.
178. Morse M, Huo R, Feng Y, Rouzina I, Chelico L, Williams MC. 2017. Dimerization regulates both deaminase-dependent and deaminase-independent HIV-1 restriction by APOBEC3G. Nat Commun 8:597.
179. Kreisberg JF, Yonemoto W, Greene WC. 2006. Endogenous factors enhance HIV infection of tissue naive CD4 T cells by stimulating high molecular mass APOBEC3G complex formation. J Exp Med 203:865–870.
180. Ito F, Yang H, Xiao X, Li SX, Wolfe A, Zirkle B, Arutiunian V, Chen XS. 2018. Understanding the Structure, Multimerization, Subcellular Localization and mC Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H. Sci Rep 8:1–15.
181. Bohn JA, Thummar K, York A, Raymond A, Brown WC, Bieniasz PD, Hatziioannou T, Smith JL. 2017. APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. Nat Commun 8:2–10.
182. Shaban NM, Shi K, Lauer K V., Carpenter MA, Richards CM, Salamango D, Wang J, Lopresti MW, Banerjee S, Levin-Klein R, Brown WL, Aihara H, Harris RS. 2018. The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism. Mol Cell 69:75-86.e9.
183. Smith JL, Pathak VK. 2010. Identification of Specific Determinants of Human APOBEC3F, APOBEC3C, and APOBEC3DE and African Green Monkey APOBEC3F That Interact with HIV-1 Vif. J Virol 84:12599–12608.
184. Schilling M, Bulli L, Weigang S, Graf L, Naumann S, Patzina C, Wagner V, Bauersfeld L, Goujon C, Hengel H, Halenius A, Ruzsics Z, Schaller T, Kochs G. 2018. Human MxB Protein Is a Pan-herpesvirus Restriction Factor. J Virol 92:1–11.
185. Tareen SU, Sawyer SL, Malik HS, Emerman M. 2009. An expanded clade of rodent Trim5 genes. Virology 385:473–483.
186. 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.
187. Ito J, Gifford RJ, Sato K. 2020. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. Proc Natl Acad Sci U S A 117:610–618.
188. Chesarino NM, Emerman M. 2020. Polymorphisms in human APOBEC3H differentially regulate ubiquitination and antiviral activity. Viruses 12.
189. Chen H, Lilley CE, Yu Q, Lee D V., Chou J, Narvaiza I, Landau NR, Weitzman MD. 2006. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol 16:480–485.
190. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. 2006. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res 34:89–95.
191. 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 19:1005–1010.
192. Yang H, Ito F, Wolfe AD, Li S, Mohammadzadeh N, Love RP, Yan M, Zirkle B, Gaba A, Chelico L, Chen XS. 2020. Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. Nat Commun 11.
193. Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, Brown WL, Reilly C, Emerman M, Harris RS. 2014. Natural Polymorphisms in Human APOBEC3H and HIV-1 Vif Combine in Primary T Lymphocytes to Affect Viral G-to-A Mutation Levels and Infectivity. PLoS Genet 10.
194. Feng Y, Love RP, Chelico L. 2013. HIV-1 viral infectivity factor (Vif) alters processive single-stranded DNA scanning of the retroviral restriction factor APOBEC3G. J Biol Chem 288:6083–6094.
195. Haché G, Shindo K, Albin JS, Harris RS. 2008. Evolution of HIV-1 Isolates that Use a Novel Vif-Independent Mechanism to Resist Restriction by Human APOBEC3G. Curr Biol 18:819–824.
196. Morón-López S, Gómez-Mora E, Salgado M, Ouchi D, Puertas MC, Urrea V, Navarro J, Jou A, Pérez M, Tural C, Clotet B, Montaner LJ, Blanco J, Crespo M, Martinez-Picado J. 2016. Short-term Treatment With Interferon Alfa Diminishes Expression of HIV-1 and Reduces CD4+ T-Cell Activation in Patients Coinfected With HIV and Hepatitis C Virus and Receiving Antiretroviral Therapy. J Infect Dis 213:1008–1012.
197. Sandler NG, Bosinger SE, Estes JD, Zhu RTR, Tharp GK, Boritz E, Levin D, Wijeyesinghe S, Makamdop KN, Del Prete GQ, Hill BJ, Timmer JK, Reiss E, Yarden G, Darko S, Contijoch E, Todd JP, Silvestri G, Nason M, Norgren RB, Keele BF, Rao S, Langer JA, Lifson JD, Schreiber G, Douek DC. 2014. Type i interferon responses in rhesus macaques prevent SIV infection and slow disease progression. Nature 511:601–605.