Structural and evolutionary analyses of host antiviral restriction factors

Caroline Langley

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Michael Emerman, Chair

Harmit Malik

Rick McLaughlin

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Structural and evolutionary analyses of host antiviral restriction factors

Caroline A. Langley

Chair of the Supervisory Committee:

Michael Emerman

Department of Microbiology

Antiviral restriction factors are critical components of the host innate immune system. This thesis investigates the structural and evolutionary dynamics of two pivotal antiviral protein families: the Mx proteins and APOBEC3 (A3) proteins.

The first section focuses on Mx proteins, which belong to the Dynamin superfamily of proteins (DSP). Comprehensive phylogenomic analyses demonstrate that Mx proteins predate the interferon signaling system in vertebrates. We elucidate the ancient monophyletic lineage of Mx proteins across diverse eukaryotes, revealing previously undescribed fungal and plant Mx orthologs. The evolutionary trajectory suggests recurrent viral appropriation of host DSPs,

underscoring an ancient co-evolutionary history of viral and antiviral functions within eukaryotes.

The second section examines the structural basis of HIV-1 Vif antagonism of human A3G, a restriction factor that induces hypermutation in viral genomes. Using cryogenic electron microscopy, we resolved the structure of A3G bound to HIV-1 Vif and hijacked cellular proteins, highlighting RNA's role as a molecular glue facilitating the Vif-A3G interaction. This interaction underpins A3G antagonism, offering insights into the molecular arms race driving host-virus co-evolution. Additionally, deep mutational scanning (DMS) of Vif proteins reveals evolutionary constraints and adaptive potential, enhancing our understanding of HIV-1 adaptation to host immune defenses.

This thesis provides a comprehensive structural and evolutionary framework for understanding how host antiviral proteins and viral antagonists co-evolve, contributing to our broader knowledge of host-virus interactions.

Table of Contents

Table of Figures

Acknowledgements

I realize that not everyone in graduate school is fortunate enough to have a mentor who inspires and pushes them to be their best, so I consider myself incredibly lucky to have had two such mentors. This work would not have been possible without their unwavering support and guidance, and the exceptional learning environments they have cultivated in their labs.

I also owe immense gratitude to my beloved coworkers past and present, who have been my biggest cheerleaders every step of the way. During the months when it seemed like every experiment failed, celebrating the successes of my brilliant lab mates was what kept me going. The broader Emerman/Malik community has also been a tremendous source of support, acting as mentors in many situations and offering reassuring smiles at conferences. I am so proud to be part of this remarkable group and aspire to continue their legacy of curiosity, innovative science, and kindness above all else.

Finally, this work would not have been possible without the love and support of my family, my perfect angel dogs Harley and Luigi, and my amazing partner John. Knowing I could come home to them provided an inner peace that kept me grounded when the stress of graduate school felt overwhelming. I love you all so very much.

1 Introduction

1.1 Evolutionary dynamics of host restriction factors and viral antagonism

The interferon (IFN) system acts as a first line of defense against viruses and other pathogens by inducing dozens of interferon-stimulated genes (ISGs) that create an antiviral environment (1). Many ISGs encode proteins with direct antiviral activity, known as restriction factors. Direct antiviral activity refers to the ability of certain proteins to directly interfere with the replication or spread of a virus within the host. These proteins can inhibit various stages of the viral life cycle, such as entry, replication, assembly, or release of new viral particles. Restriction factors are a specific subset of these antiviral proteins that are typically constitutively expressed and act to restrict viral infections by directly targeting viral components or processes. Not all antiviral proteins are considered restriction factors; those that are inducible or that operate through more indirect mechanisms, such as modulating the host immune response, do not fall under this category. Restriction factors aim to block replication at different stages of the viral lifecycle. To facilitate successful infection and replication, viruses must overcome the antiviral antagonism posed by restriction factors. These proteins are often key players in host-virus "molecular arms-races", which describes the iterative cycle of host adaptation to evade viral recognition followed by subsequent viral counteradaptation (2-4). Recurrence of these coevolutionary processes results in genetic signatures in both hosts and the viruses that infect them (5, 6). Studying the genetic diversity of host restriction factors is a powerful approach to understanding how viruses have impacted host evolution and the key mutations in viruses underscoring cross-species transmission events.

One way to distinguish the selection regimes experienced by genesis by comparing the rate at which non-synonymous mutations, which alter the amino acid sequence, fix between

orthologous genes of closely related species, relative to the rate of fixation of non-synonymous mutations within these sequences (7, 8). Genes with lower rates of non-synonymous mutations relative to synonymous mutations are said to have evolved under purifying selection. These genes often play essential roles in fundamental cellular functions and overall viability. Most mutations impair the function of such genes and are typically purged via natural selection. Conversely, genes under positive selection exhibit a higher rate of non-synonymous relative to synonymous mutations. These genes often involve cellular functions where variation or rapid divergence can be beneficial, such as immune responses against constantly evolving pathogens. In fact, host genes involved in innate antiretroviral immune responses are some of the fastest-evolving genes in human and chimpanzee genomes (9). Positive selection drives adaptive changes in both the host and the pathogen genomes, facilitating a coevolutionary dynamic between them. Such genomic analyses are a powerful technique used to identify restriction factors to characterize host determinants of viral infections.

Beyond merely identifying genes under positive selection, we can also use these analyses at the molecular level to identify specific sites in a protein involved in evolutionary conflict. Some codons or positions are under purifying selection, indicating their importance to overall protein structure or function. Other positions exhibit high variability, but this does not necessarily imply that they are dispensable for overall protein function. Instead, sites under positive selection in host restriction factors may reflect evolutionary pressures favoring changes at these positions to evade viral recognition (8, 10). For example, the antiviral activity of many host restriction factors is neutralized when they are recognized by a viral protein. Furthermore, positive selection analyses at the site level can also help pinpoint determinants of viral cross-species transmission events (2, 6, 11, 12). By identifying positions under positive selection, we

can highlight species-specific differences in the amino acid identity of restriction factors which may pose barriers to cross-species transmission. In cases where cross-species transmission has already occurred, comparing the sequence identities of the viral protein counterpart to the host restriction factor provides insight into the viral mutations that conferred antagonism (13, 14). Thus, by identifying residues in restriction factors under evolutionary pressure to mutate, we can better identify the compensatory evolution on the virus side necessary for spillover.

In addition to positive selection, the loss and gain of gene copies is another hallmark of evolutionary arms races. Gene duplication events can result in the creation of new restriction factors or enhance the existing antiviral repertoires of host species (3, 15, 16). While some viruses, such as retroviruses, can rapidly alter the functions of existing genes through error-prone replication, gene duplication and the acquisition of new genes entirely is infrequent in retroviruses (17), likely due to the limitations on viral genome sizes. In stark contrast, we frequently observe instances of gene duplication events across the evolution of host species. In fact, many restriction factors are parts of larger protein families that arose through gene duplication events. By comparing the genomes of different host species, we can identify patterns of gene gain, loss, and even occasional horizontal transfer. As discussed below, these patterns may reveal the mechanisms of viral antagonism and host defense, as well as the broader evolutionary paths that have shaped the immune landscapes of diverse organisms.

Restriction factors can be both highly specific and highly versatile in terms of the viruses they aim to antagonize (18). For example, the myxovirus resistance genes (*MxA* and *MxB* in humans) are restriction factors with a broad range of viral targets. The MxA protein restricts diverse viruses, including Influenza A (19-21), Vesicular Stomatitis Virus (VSV) (20, 21), measles (22), and hepatitis B (23), whereas the MxB paralog restricts retroviruses (24-26) and

herpesviruses (27, 28). On the other end of the spectrum, the *APOBEC3* (*A3*) genes, which encode seven proteins in primate lineages named A3A, A3B, A3C, A3D, A3F, A3G, and A3H, are more specific restriction factors against lentiviruses, retroviruses, and retroelements. Positive selection analyses have played a pivotal role in the characterization of these proteins as antiviral restriction factors, as well as the subsequent work focused on their role in molecular arms-races with viral proteins (5, 6, 12, 13, 18, 29, 30). The following sections will delve deeper into the antiviral mechanisms of Mx and A3 proteins as well as the specific viral proteins that target them. By examining the evolutionary dynamics at play, we gain a more comprehensive understanding of how viruses counteract immune defenses and the potential evolutionary tradeoffs involved.

1.2 Mx proteins as broad acting restriction factors

Mx proteins are members of the Dynamin superfamily of proteins (DSP), multi-domain GTPases that mediate many critical cellular processes within eukaryotic cells. Mx1 was identified as an antiviral protein not long after the discovery of IFN (31-33) and has subsequently become one of the most well-studied IFN-induced antiviral proteins (34). The broad antiviral activity displayed by Mx proteins suggests involvement with multiple evolutionary arms-race conflicts. However, unlike other broadly acting restriction factors that detect common signs of viral infection, Mx proteins recognize highly specific proteins from these diverse viruses (35). For example, amino acid differences in the nucleoprotein (NP) sequences between avian and human influenza virus isolates are the sole determinants of susceptibility or resistance to human MxA (14). However, MxA restriction of Semliki Forest virus (SFV), an alphavirus, is independent of the SFV NP (36). Furthermore, MxA restriction of DNA viruses such as hepatitis B virus (HBV) (37) and African swine fever virus (ASFV) (38) is determined by different viral

proteins. Investigating both the precise mechanisms of restriction and the wide antiviral range of Mx proteins, despite their target specificity, remains an ongoing area of research.

Despite limited knowledge about the precise mechanism of MxA viral counteraction, positive selection analysis identified a specific region in MxA – loop L4 –that plays an important role in the restriction of orthomyxoviruses (18). Subsequent analyses revealed that variation in loop L4 among primate MxA orthologs underscores the differences in their ability to restrict Thogoto virus (THOV) and influenza A viruses (FLUAV) (18). Identifying this genetic determinant of MxA antiviral specificity underscores the power of utilizing evolutionary methodologies to study host-virus interactions.

Mx proteins are highly conserved among vertebrates. Most mammals have two *Mx* genes, while birds have a single *Mx* gene, and fish have up to seven Mx paralogs, which evolved by gene amplification (39-41). The expansions and losses of *Mx* genes across vertebrate evolution are consistent with the profile of innate defense genes in general. To date, the evolution and diversification of the DSP, including Mx, have been only partially characterized (42-44), leaving significant gaps in our understanding of Mx's evolutionary history despite the critical role it plays in antiviral defense. Our subsequent work has sought to comprehensively examine the deep phylogenetic history of Mx within the broader DSP context.

1.3 Molecular mechanisms of Mx antiviral activity

Several mechanisms have been proposed for how Mx proteins restrict orthomyxoviruses, specifically targeting viral replication and assembly. One of the primary mechanisms involves the recognition and binding of viral nucleoproteins (NPs). This binding inhibits the transport of viral ribonucleoprotein (vRNP) complexes into the nucleus, a crucial step for viral replication (45). Additionally, Mx interferes with the assembly and budding of new viral particles by

disrupting the interaction between viral NP and other viral or host cell proteins essential for these processes (46). This disruption not only prevents the formation of functional viral particles but also triggers the degradation of viral components via cellular proteasomal pathways. Mx proteins also oligomerize, forming high-molecular-weight complexes that enhance their antiviral activity (47). These complexes create a scaffold that sequesters viral components, preventing them from participating in the viral life cycle. The ATPase activity of Mx is crucial for this oligomerization and the subsequent antiviral functions, indicating that mutations affecting ATP binding and hydrolysis significantly reduce MxA's antiviral efficacy (35). Moreover, Mx proteins exhibit localization to specific subcellular compartments where they encounter and neutralize viral components. For instance, MxA localizes to the smooth endoplasmic reticulum (ER) and Golgi apparatus, sites of viral protein synthesis and processing, potentially intercepting viral components early in the infection cycle (48). The localization is mediated by specific amino acid sequences within MxA that direct its trafficking to these organelles, facilitating efficient antiviral responses (49).

1.4 A3 proteins act as antiviral restriction factors by hypermutating the viral genome

The *A3* gene family has undergone multiple independent expansions in various lineages, most notably in primates but also in felids and equids. The current copy number diversity within these lineages results from a series of tandem duplication events in the *A3* locus, followed by positive selection and sub/neo-functionalization (50). In primates, there are seven members of the A3 protein family: A3A, A3B, A3C, A3D, A3F, A3G, and A3H. All A3 proteins are cytosine deaminases containing one or two conserved zinc-binding domains with a (Cys/His)-Xaa-Glu-$Xaa_{23–28}$-Pro-Cys-$Xaa_{2–4}$-Cys) motif (51). These domains are separated into three phylogenetic groups: Z1, Z2, and Z3 (52). The primate A3s with one zinc-binding domain are A3A (Z1), A3C

(Z2), and A3H (Z3), while A3B (Z2-Z1), A3D (Z2-Z2), A3F (Z2-Z2), and A3G (Z2-Z1) have two zinc-binding domains though only the C-terminal domain is catalytically active in most of these proteins (50). Though named for the mRNA editing activity of the founding member, APOBEC1, the relevant activity of A3s as antiviral proteins is modifying DNA (53, 54).

A3 expression is abundant in multiple immunity-associated cell types, including activated T cells, B cells, leukocytes, monocytes, macrophages, and mature dendritic cells (56-58). However, the expression of A3 enzymes is tissue and cell-type specific. Peripheral blood leukocytes express transcripts for all *A3* family members, with *A3A* and *A3G* being the most abundantly represented (58). *A3A* expression is specifically prominent in cells of the myeloid lineage, whereas *A3G* is highly expressed in both CD4+ T lymphocytes and myeloid cells. Additionally, *A3* mRNA levels are elevated in lymphoid organs such as the thymus and spleen. Transcripts for various A3 enzymes are also found in non-immune tissues, including the lung, ovary, and adipose tissue (58). In some tissues, *A3* expression may partly result from infiltrating leukocytes (56). Although multiple *A3* genes are constitutively expressed across most cell and tissue types, distinct *A3* genes are induced by activation or stimulation in a cell-type specific manner. Different studies have shown IFN-α, IFN-β, and IFN-γ enhance the expression of *A3*s within these cell types (56, 59, 60). In hematopoietic cell populations, the transcript and protein levels of certain A3 enzymes, particularly A3A and A3G, fluctuate during developmental and differentiation stages (61, 62), impacting the HIV-restrictive capacity of the specific cell type. This suggests that A3 enzymes are vital to an effective innate immune response.

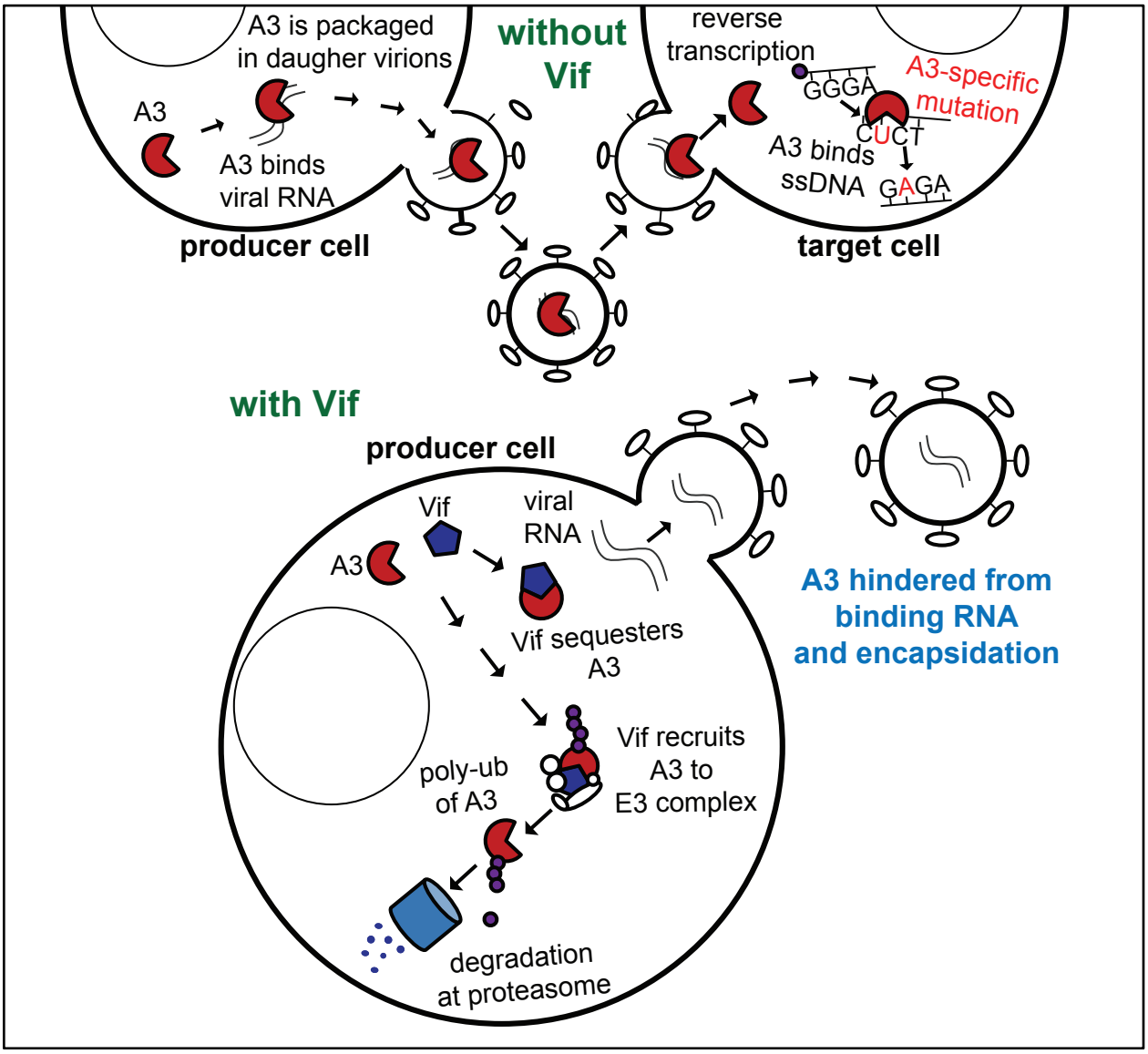
A3 antiviral restriction requires that the protein be packaged into virions and delivered to the target cell so that it is present at the site of viral reverse transcription, where it binds single-stranded viral DNA as the substrate for deamination (Fig. 1) (63). A3G is the most potent

inhibitor of HIV among the primate A3s (55). To be packaged into virions, A3 proteins must undergo RNA-dependent oligomerization (64, 65). Generally, A3 proteins are promiscuous in their RNA binding activity, exhibiting no preference for the type of RNA bound (64, 65). Interaction with either host or viral RNA can result in A3G packaging into nascent virions (64, 65). However, A3 proteins preferentially bind HIV-1 viral RNA in infected cells and do not require any viral proteins to be packaged (66). Interestingly, it has been suggested that this RNA-binding specificity mimics that of HIV nucleocapsid (NC), which also binds RNA to ensure incorporation into virions (66).

When packaged, A3G deaminates the second dC of 5′-CC dinucleotide sites in the newly synthesized viral minus-stranded ssDNA during reverse transcription (54, 63, 67, 68). Several factors contribute to the efficiency of viral ssDNA deamination, including the proximity of the target nucleotides to the 5’ end of the ssDNA, the orientation of the A3G CTD relative to the 5’ end of the ssDNA substrate, as well as the target ssDNA overall length (70). The dinucleotide preference exhibited by A3G is unique among A3 proteins (73, 74). For example, A3F prefers a 5′-TC dinucleotide substrate compared to the 5’-CC preference of A3G. This specificity influences the type and location of mutations introduced into the viral genome, thereby contributing differently to the viral mutational landscape.

The A3 signature G-to-A hypermutations can be observed in HIV sequences derived from clinical samples (75-79). In fact, studies analyzing many independent sequences isolated from the same individual have found hypermutated sequences in 43% to 100% of HIV-infected individuals (76, 80-82). Such hypermutations can significantly impact viral fitness and replication. For example, extensive G-to-A hypermutation can lead to lethal mutagenesis of the

virus, rendering it non-infectious. However, sub-lethal hypermutation levels may contribute to viral diversity and immune escape (83).



1-1 **Vif antagonizes A3 proteins by mediating their sequestration and degradation**

1.5 Deamination-independent mechanisms of A3G restriction

There is also evidence suggesting deamination-independent mechanisms of A3G-mediated HIV restriction. Introducing mutations into the active site of A3G, rendering it

catalytically dead, still hinders replication of HIV-1, mouse mammary tumor virus, and murine leukemia virus to a certain degree (65, 84, 85). However, through experimental and mathematical investigation, one group has concluded that 99.3% of the antiviral effect of A3G is dependent on its deaminase activity (86). Importantly, the A3G deamination-independent mechanism also requires A3G to be packaged into daughter virions and may be mediated through an interaction between A3G and the viral reverse transcriptase, blocking the production of viral DNA (86).

1.6 Coevolution of A3s and endogenous retroelements

Though this work focuses on A3s in the scope of retroviral restriction, it is important to recognize that the evolutionary pressure on these genes may also have been driven by mobile endogenous retroelements (16, 90). Mobile endogenous retroelements, including endogenous retroviruses (ERVs), long interspersed elements (LINEs), and short interspersed elements (SINEs), pose a significant threat to genomic integrity by copying themselves and reinserting into new locations within the genome. These elements can disrupt gene function and genomic stability through insertional mutagenesis and recombination events (91, 92). The activity of these retroelements necessitates robust defense mechanisms to maintain genomic integrity. The cytidine deaminase activity of A3 proteins plays a crucial role in their antiviral activity by introducing mutations into viral genomes, thereby inhibiting viral replication. This mutagenic activity extends beyond infectious retroviruses to include endogenous retroelements, leading to their inactivation and preventing their proliferation within the host germline genome (93-95). The presence of multiple A3 paralogs in primates, each with unique substrate specificities and expression patterns, suggests a division of labor in combating diverse retroelements. Furthermore, characterization of *A3G* retrocopies in New World monkeys revealed that the ancestral A3G and the young retrocopies created from A3G likely retained the ability to restrict

both LINE-1 retroelements and HIV-1. Over time, the intron-containing *A3G* in these species retained antiviral and anti-LINE-1 activities, but all tested *A3G* retrocopies continue to restrict HIV-1 but have lost the ability to restrict LINE-1 (16). This pattern of gene duplication via retrocopying represents another facet of the dynamic evolution of these proteins in their restriction of pathogens.

1.7 Evolutionary pressures on A3/Vif proteins

Retroviruses have evolved distinct mechanisms to overcome A3s, including the Bet protein in foamy viruses (96), the Human T-lymphotropic virus type 1 (HTLV-1) nucleocapsid (97), and the glycosylated Gag protein of murine leukemia virus (98). However, the primary focus of this work is the lentiviral accessory gene *vif*. The retention of the *vif* gene throughout lentiviral evolution, except in equine infectious anemia virus, underscores its critical role in viral survival (99, 100). This importance is attributed to the Vif protein's function as a viral antagonist of A3s during infection (101, 102).

The evolutionary arms race between A3G and Vif is ongoing. Comparative analysis of *A3G* genes across primate species revealed that *A3G* is rapidly evolving to evade Vif antagonism, which is reflected by positive selection at residues recognized by Vif (6, 12, 103). Variations in A3 proteins across different species of primates have significantly influenced the evolution of the Vif protein by acting as barriers to cross-species transmission, selecting for Vif mutations that permit the antagonism of divergent A3 proteins (32-34). For example, molecular differences between A3 proteins in hominids and Old World monkeys have been pivotal barriers to cross-species transmission of Simian Immunodeficiency Viruses (SIVs) and have shaped the evolution of HIV-1 (35, 36). The immediate precursor to the HIV-1 Group M/N *vif* gene is the *vif* gene from SIV_{cpz}, which infects chimpanzees. The SIV_{cpz} *vif* gene is derived from SIV_{rcm}, a virus

infecting the red-capped mangabey, an Old World monkey species. While the A3G proteins in chimpanzees and humans are highly conserved in sequence, A3G proteins in Old World monkeys, such as red-capped mangabeys, are more diverse. Despite the differences between Old World monkeys and hominid A3G, SIVrcm Vif can still weakly antagonize hominid A3G despite evolving under the pressure of Old World monkey A3G, an unfortunate evolutionary coincidence (35, 36). After the initial cross-species transmission event, the of SIV_{rcm} Vif evolved full antagonism of hominid A3G, resulting in the SIV_{cpz} Vif protein. Unsurprisingly, the SIV_{cpz} Vif protein can fully antagonize human A3G, contributing to the birth of HIV-1 M/N.

1.8 A3H polymorphism and Vif adaptation

Unlike other *A3* genes, *A3H* exhibits multiple polymorphic variants, encoded by different haplotypes, significantly affecting its stability and antiviral function (104). Among these, *A3H* haplotype II, referred to as *A3H*, has exerted notable evolutionary pressures on the HIV-1 Vif protein due to its robust antiviral activity. Given the simultaneous expression of various A3 proteins, the Vif proteins of HIV-1 viruses infecting individuals expressing A3H must evolve to bind this new variant while maintaining interactions with other A3 proteins. Indeed, Vif sequences isolated from individuals infected with HIV expressing A3H demonstrate unique evolutionary adaptions reflected by sequence differences exerted by the expression of this host antiviral protein (105). Furthermore, structural studies reveal a distinct interface on Vif used to facilitate interactions with A3H, which could be attributed to the fact that A3H is the most divergent of the A3 proteins (106-109).

1.9 Structural studies examining Vif antagonism of A3G

Given that A3G requires encapsidation to exert its antiviral effects, Vif must prevent A3G packaging by facilitating its degradation in the producer cell. Specifically, HIV-1 Vif recruits

various cellular proteins to assemble an E3 ubiquitin ligase complex, which ubiquitinates A3G, leading to its subsequent proteasomal degradation (101, 102, 110). The first crystal structure of HIV-1 Vif in complex with these host proteins revealed the precise binding interfaces involved in interactions with scaffold protein cullin 5 (CUL5), substrate adaptors elongin B (ELOB) and elongin C (ELOC), and core-binding factor β (CBFβ), which stabilizes Vif expression (111). Collectively, Vif, CBFβ, ELOB, and ELOC are referred to as the VCBC complex. Detailed structural analyses of the VCBC complex have elucidated the conformational dynamics and identified specific Vif residues, primarily in the N-terminal domain, that are crucial for engaging with the hydrophobic groove of CUL5 and the interaction surfaces with ELOB and ELOC (111, 112). Furthermore, the structures of both HIV-1 Vif and SIV_{rcm} Vif reveal an α-domain containing two helices, one interacting with CUL5 and the other with ELOC, mirroring interactions made by the natural substrate receptor SOCS2 (112). These structural revelations underscore the sophisticated mechanism by which Vif orchestrates the degradation of A3G, thereby neutralizing its antiviral activity. Additionally, the structure of the SIV_{rcm} Vif, combined with the conservation of VCBC interacting residues across broad SIV Vifs, suggests that the formation of this complex to restrict A3G is evolutionarily conserved across lentiviral Vifs (13).

CBFβ is generally not a component of the host CUL5 E3 complex but is necessary for Vif to antagonize A3G (113, 114). The binding interface between CBFβ and the α/β domain of Vif stabilizes the assembly of Vif with CUL5 by shielding a large hydrophobic patch on Vif from solvent exposure, thus protecting Vif from degradation (111). While this interaction is not required for Vif to bind A3G (113), it is essential for facilitating the ubiquitin-mediated degradation of A3G (113, 114). However, CBFβ is dispensable for non-primate Vifs. For example, Maedi-visna virus (MVV) Vif requires cyclophilin A (CypA) as a co-factor, while

Bovine immunodeficiency virus (BIV) Vif assembles a host E3 complex without any cofactor (115). This highlights different evolutionary paths taken by lentiviral Vif proteins to overcome different species' A3G proteins.

1.10 Using evolution, molecular virology, and structure to study restriction factors

Here, we take multiple approaches to study the evolution of two different antiviral restriction factors, Mx and A3G, in their arms races with viruses. Using phylogenetics-based approaches, we reveal an ancient lineage of potentially antiviral Mx proteins in eukaryotes, an understudied arms race for dynamin-related functions between eukaryotic viruses and their hosts. Furthermore, we comprehensively explore the evolutionary origins of the *Mx* gene family across eukaryotic evolution, highlighting novel orthologs and multiple gene family expansions driven by diverse pathogens. In addition, we elucidate the molecular interface of A3G and Vif through both structural and evolution-guided approaches and comprehensively characterize the determinants of Vif restriction of A3G using deep mutational scanning (DMS). Collectively, this work provides insights into cross-species transmission events and constraints on viral fitness by characterizing the mutational landscape of key interfaces in Vif proteins from two different strains of HIV-1 in the context of A3G antagonism.

1.11 References

1. J. W. Schoggins, Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol* **6**, 567-584 (2019).
2. M. D. Daugherty, H. S. Malik, Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet* **46**, 677-700 (2012).
3. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687-695 (2012).
4. N. R. Meyerson, S. L. Sawyer, Two-stepping through time: mammals and viruses. *Trends Microbiol* **19**, 286-294 (2011).
5. P. S. Mitchell *et al.*, Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* **12**, 598-604 (2012).
6. A. A. Compton, V. M. Hirsch, M. Emerman, The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe* **11**, 91-98 (2012).
7. Z. Yang, J. P. Bielawski, Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15**, 496-503 (2000).
8. L. D. Hurst, The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* **18**, 486 (2002).
9. R. Nielsen *et al.*, A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3**, e170 (2005).
10. E. C. Holmes, Adaptation and immunity. *PLoS Biol* **2**, E307 (2004).
11. S. L. Sawyer, L. I. Wu, J. M. Akey, M. Emerman, H. S. Malik, High-frequency persistence of an impaired allele of the retroviral defense gene TRIM5alpha in humans. *Curr Biol* **16**, 95-100 (2006).
12. S. L. Sawyer, M. Emerman, H. S. Malik, Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* **2**, E275 (2004).
13. J. M. Binning, N. M. Chesarino, M. Emerman, J. D. Gross, Structural Basis for a Species-Specific Determinant of an SIV Vif Protein toward Hominid APOBEC3G Antagonism. *Cell Host Microbe* **26**, 739-747 e734 (2019).
14. P. Zimmermann, B. Manz, O. Haller, M. Schwemmle, G. Kochs, The viral nucleoprotein determines Mx sensitivity of influenza A viruses. *J Virol* **85**, 8133-8140 (2011).
15. M. Muller, D. Sauter, The more the merrier? Gene duplications in the coevolution of primate lentiviruses with their hosts. *Curr Opin Virol* **62**, 101350 (2023).
16. L. Yang, M. Emerman, H. S. Malik, R. N. J. McLaughlin, Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. *Elife* **9** (2020).
17. E. Simon-Loriere, E. C. Holmes, Gene duplication is infrequent in the recent evolutionary history of RNA viruses. *Mol Biol Evol* **30**, 1263-1269 (2013).
18. P. S. Mitchell, M. Emerman, H. S. Malik, An evolutionary perspective on the broad antiviral specificity of MxA. *Curr Opin Microbiol* **16**, 493-499 (2013).
19. J. Dittmann *et al.*, Influenza A virus strains differ in sensitivity to the antiviral action of Mx-GTPase. *J Virol* **82**, 3624-3631 (2008).

20. J. Pavlovic, T. Zurcher, O. Haller, P. Staeheli, Resistance to influenza virus and vesicular stomatitis virus conferred by expression of human MxA protein. *J Virol* **64**, 3370-3375 (1990).
21. P. Staeheli, J. Pavlovic, Inhibition of vesicular stomatitis virus mRNA synthesis by human MxA protein. *J Virol* **65**, 4498-4501 (1991).
22. S. Schneider-Schaulies *et al.*, Cell type-specific MxA-mediated inhibition of measles virus transcription in human brain cells. *J Virol* **68**, 6910-6917 (1994).
23. E. Gordien *et al.*, Inhibition of hepatitis B virus replication by the interferon-inducible MxA protein. *J Virol* **75**, 2684-2691 (2001).
24. C. Goujon *et al.*, Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559-562 (2013).
25. M. Kane *et al.*, MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-566 (2013).
26. Z. Liu *et al.*, The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe* **14**, 398-410 (2013).
27. M. Crameri *et al.*, MxB is an interferon-induced restriction factor of human herpesviruses. *Nat Commun* **9**, 1980 (2018).
28. M. Schilling *et al.*, Human MxB Protein Is a Pan-herpesvirus Restriction Factor. *J Virol* **92** (2018).
29. N. M. Chesarino, M. Emerman, HIV-1 Vif Gained Breadth in APOBEC3G Specificity after Cross-Species Transmission of Its Precursors. *J Virol* **96**, e0207121 (2022).
30. Y. L. Li *et al.*, The structural basis for HIV-1 Vif antagonism of human APOBEC3G. *Nature* **615**, 728-733 (2023).
31. A. Isaacs, J. Lindenmann, Virus interference. I. The interferon. *Proc R Soc Lond B Biol Sci* **147**, 258-267 (1957).
32. M. A. Horisberger, P. Staeheli, O. Haller, Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. *Proc Natl Acad Sci U S A* **80**, 1910-1914 (1983).
33. P. Staeheli, O. Haller, W. Boll, J. Lindenmann, C. Weissmann, Mx protein: constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. *Cell* **44**, 147-158 (1986).
34. A. J. Sadler, B. R. Williams, Interferon-inducible antiviral effectors. *Nat Rev Immunol* **8**, 559-568 (2008).
35. O. Haller, G. Kochs, Human MxA protein: an interferon-induced dynamin-like GTPase with broad antiviral activity. *J Interferon Cytokine Res* **31**, 79-87 (2011).
36. H. Landis *et al.*, Human MxA protein confers resistance to Semliki Forest virus and inhibits the amplification of a Semliki Forest virus-based replicon in the absence of viral structural proteins. *J Virol* **72**, 1516-1522 (1998).
37. N. Li *et al.*, MxA inhibits hepatitis B virus replication by interaction with hepatitis B core antigen. *Hepatology* **56**, 803-811 (2012).
38. C. L. Netherton *et al.*, Inhibition of a large double-stranded DNA virus by MxA protein. *J Virol* **83**, 2310-2320 (2009).
39. B. Robertsen, The interferon system of teleost fish. *Fish Shellfish Immunol* **20**, 172-191 (2006).
40. Y. B. Zhang, J. F. Gui, Molecular regulation of interferon antiviral response in fish. *Dev Comp Immunol* **38**, 193-202 (2012).
41. P. Boudinot, C. Langevin, C. J. Secombes, J. P. Levraud, The Peculiar Characteristics of Fish Type I Interferons. *Viruses* **8** (2016).
42. R. Purkanti, M. Thattai, Ancient dynamin segments capture early stages of host-mitochondrial integration. *Proc Natl Acad Sci U S A* **112**, 2800-2805 (2015).
43. S. Sheikh *et al.*, A Novel Group of Dynamin-Related Proteins Shared by Eukaryotes and Giant Viruses Is Able to Remodel Mitochondria From Within the Matrix. *Mol Biol Evol* **40** (2023).
44. S. Sinha, N. Manoj, Molecular evolution of proteins mediating mitochondrial fission-fusion dynamics. *FEBS Lett* **593**, 703-718 (2019).
45. L. Graf *et al.*, Effects of allelic variations in the human myxovirus resistance protein A on its antiviral activity. *J Biol Chem* **293**, 3056-3072 (2018).
46. S. Tripathi, J. Batra, S. K. Lal, Interplay between influenza A virus and host factors: targets for antiviral intervention. *Arch Virol* **160**, 1877-1891 (2015).
47. O. Haller, S. Gao, A. von der Malsburg, O. Daumke, G. Kochs, Dynamin-like MxA GTPase: structural insights into oligomerization and implications for antiviral activity. *J Biol Chem* **285**, 28419-28424 (2010).
48. O. Haller, P. Staeheli, M. Schwemmle, G. Kochs, Mx GTPases: dynamin-like antiviral machines of innate immunity. *Trends in Microbiology* **23**, 154-163 (2015).

49. K. Faelber *et al.*, Oligomerization of dynamin superfamily proteins in health and disease. *Prog Mol Biol Transl Sci* **117**, 411-443 (2013).
50. C. Munk, A. Willemsen, I. G. Bravo, An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol Biol* **12**, 71 (2012).
51. A. Jarmuz *et al.*, An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285-296 (2002).
52. R. S. LaRue *et al.*, Guidelines for naming nonprimate APOBEC3 genes and proteins. *J Virol* **83**, 494-497 (2009).
53. R. S. Harris, S. K. Petersen-Mahrt, M. S. Neuberger, RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253 (2002).
54. H. Zhang *et al.*, The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-98 (2003).
55. C. Chaipan, J. L. Smith, W. S. Hu, V. K. Pathak, APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol* **87**, 444-453 (2013).
56. F. A. Koning *et al.*, Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**, 9474-9485 (2009).
57. H. Oliva *et al.*, Increased expression with differential subcellular location of cytidine deaminase APOBEC3G in human CD4(+) T-cell activation and dendritic cell maturation. *Immunol Cell Biol* **94**, 689-700 (2016).
58. E. W. Refsland *et al.*, Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* **38**, 4274-4284 (2010).
59. V. Mohanram, A. E. Skold, S. M. Bachle, S. K. Pathak, A. L. Spetz, IFN-alpha induces APOBEC3G, F, and A in immature dendritic cells and limits HIV-1 spread to CD4+ T cells. *J Immunol* **190**, 3346-3353 (2013).
60. V. Cobos Jimenez *et al.*, Differential expression of HIV-1 interfering factors in monocyte-derived macrophages stimulated with polarizing cytokines or interferons. *Sci Rep* **2**, 763 (2012).
61. G. Peng *et al.*, Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood* **110**, 393-400 (2007).
62. M. A. Farrow, E. Y. Kim, S. M. Wolinsky, A. M. Sheehy, NFAT and IRF proteins regulate transcription of the anti-HIV gene, APOBEC3G. *J Biol Chem* **286**, 2567-2577 (2011).
63. R. S. Harris *et al.*, DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809 (2003).
64. V. Zennou, D. Perez-Caballero, H. Gottlinger, P. D. Bieniasz, APOBEC3G incorporation into human immunodeficiency virus type 1 particles. *J Virol* **78**, 12058-12061 (2004).
65. E. S. Svarovskaia *et al.*, Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. *J Biol Chem* **279**, 35822-35828 (2004).
66. A. York, S. B. Kutluay, M. Errando, P. D. Bieniasz, The RNA Binding Specificity of Human APOBEC3 Proteins Resembles That of HIV-1 Nucleocapsid. *PLoS Pathog* **12**, e1005833 (2016).
67. B. Mangeat *et al.*, Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103 (2003).
68. Q. Yu *et al.*, Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* **11**, 435-442 (2004).
69. L. Chelico, E. J. Sacho, D. A. Erie, M. F. Goodman, A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *J Biol Chem* **283**, 13780-13791 (2008).
70. L. Chelico, P. Pham, P. Calabrese, M. F. Goodman, APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nat Struct Mol Biol* **13**, 392-399 (2006).
71. L. Chelico, C. Prochnow, D. A. Erie, X. S. Chen, M. F. Goodman, Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J Biol Chem* **285**, 16195-16205 (2010).
72. L. S. Shlyakhtenko *et al.*, APOBEC3G Interacts with ssDNA by Two Modes: AFM Studies. *Sci Rep* **5**, 15648 (2015).
73. J. F. Hultquist *et al.*, Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-11234 (2011).
74. A. Rathore *et al.*, The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. *J Mol Biol* **425**, 4442-4454 (2013).

75. J. P. Vartanian, A. Meyerhans, B. Asjo, S. Wain-Hobson, Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. *J Virol* **65**, 1779-1788 (1991).
76. M. Janini, M. Rogers, D. R. Birx, F. E. McCutchan, Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. *J Virol* **75**, 7973-7986 (2001).
77. Y. C. Ho *et al.*, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540-551 (2013).
78. K. M. Bruner *et al.*, Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043-1049 (2016).
79. K. A. Delviks-Frankenberry *et al.*, Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation. *PLoS Pathog* **12**, e1005646 (2016).
80. A. Piantadosi, D. Humes, B. Chohan, R. S. McClelland, J. Overbaugh, Analysis of the percentage of human immunodeficiency virus type 1 sequences that are hypermutated and markers of disease progression in a longitudinal cohort, including one individual with a partially defective Vif. *J Virol* **83**, 7805-7814 (2009).
81. T. L. Kieffer *et al.*, G-->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. *J Virol* **79**, 1975-1980 (2005).
82. S. K. Gandhi, J. D. Siliciano, J. R. Bailey, R. F. Siliciano, J. N. Blankson, Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. *J Virol* **82**, 3125-3130 (2008).
83. C. Munk, B. E. Jensen, J. Zielonka, D. Haussinger, C. Kamp, Running loose or getting lost: how HIV-1 counters and capitalizes on APOBEC3-induced mutagenesis through its Vif protein. *Viruses* **4**, 3132-3161 (2012).
84. R. K. Holmes, F. A. Koning, K. N. Bishop, M. H. Malim, 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 (2007).
85. K. Luo *et al.*, Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol* **81**, 7238-7248 (2007).
86. T. Kobayashi *et al.*, Quantification of deaminase activity-dependent and -independent restriction of HIV-1 replication mediated by APOBEC3F and APOBEC3G through experimental-mathematical investigation. *J Virol* **88**, 5881-5887 (2014).
87. K. Belanger, M. Savoie, M. C. Rosales Gerpe, J. F. Couture, M. A. Langlois, Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. *Nucleic Acids Res* **41**, 7438-7452 (2013).
88. D. Pollpeter *et al.*, Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nat Microbiol* **3**, 220-233 (2018).
89. E. N. Newman *et al.*, Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* **15**, 166-170 (2005).
90. M. H. Malim, M. Emerman, HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388-398 (2008).
91. H. L. Levin, J. V. Moran, Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**, 615-627 (2011).
92. J. D. Boeke, J. P. Stoye, "Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements" in Retroviruses, J. M. Coffin, S. H. Hughes, H. E. Varmus, Eds. (Cold Spring Harbor (NY), 1997).
93. H. P. Bogerd *et al.*, Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 8780-8785 (2006).
94. S. R. Richardson, I. Narvaiza, R. A. Planegger, M. D. Weitzman, J. V. Moran, APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3**, e02008 (2014).
95. R. N. McLaughlin, Jr., J. T. Gable, C. J. Wittkopp, M. Emerman, H. S. Malik, Conservation and Innovation of APOBEC3A Restriction Functions during Primate Evolution. *Mol Biol Evol* **33**, 1889-1901 (2016).
96. R. A. Russell *et al.*, Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol* **79**, 8724-8731 (2005).
97. D. Derse, S. A. Hill, G. Princler, P. Lloyd, G. Heidecker, Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. *Proc Natl Acad Sci U S A* **104**, 2915-2920 (2007).

98. A. Kolokithas *et al.*, The glycosylated Gag protein of a murine leukemia virus inhibits the antiretroviral function of APOBEC3. *J Virol* **84**, 10933-10936 (2010).
99. R. J. Gifford, Viral evolution in deep time: lentiviruses and mammals. *Trends Genet* **28**, 89-100 (2012).
100. G. Z. Han, M. Worobey, Endogenous lentiviral elements in the weasel family (Mustelidae). *Mol Biol Evol* **29**, 2905-2908 (2012).
101. A. M. Sheehy, N. C. Gaddis, M. H. Malim, The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **9**, 1404-1407 (2003).
102. S. G. Conticello, R. S. Harris, M. S. Neuberger, The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* **13**, 2009-2013 (2003).
103. M. Ortiz, G. Bleiber, R. Martinez, H. Kaessmann, A. Telenti, Patterns of evolution of host proteins involved in retroviral pathogenesis. *Retrovirology* **3**, 11 (2006).
104. M. M. Li, M. Emerman, Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. *J Virol* **85**, 8197-8207 (2011).
105. M. M. Li, L. I. Wu, M. Emerman, The range of human APOBEC3H sensitivity to lentiviral Vif proteins. *J Virol* **84**, 88-95 (2010).
106. J. Wang *et al.*, The Role of RNA in HIV-1 Vif-Mediated Degradation of APOBEC3H. *J Mol Biol* **431**, 5019-5031 (2019).
107. M. Ooms, M. Letko, V. Simon, The Structural Interface between HIV-1 Vif and Human APOBEC3H. *J Virol* **91** (2017).
108. M. Nakashima *et al.*, Mapping Region of Human Restriction Factor APOBEC3H Critical for Interaction with HIV-1 Vif. *J Mol Biol* **429**, 1262-1276 (2017).
109. F. Ito, A. L. Alvarez-Cabrera, K. Kim, Z. H. Zhou, X. S. Chen, Structural basis of HIV-1 Vif-mediated E3 ligase targeting of host APOBEC3H. *Nat Commun* **14**, 5241 (2023).
110. X. Yu *et al.*, Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060 (2003).
111. Y. Guo *et al.*, Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature* **505**, 229-233 (2014).
112. B. J. Stanley *et al.*, Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *J Virol* **82**, 8656-8663 (2008).
113. W. Zhang, J. Du, S. L. Evans, Y. Yu, X. F. Yu, T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction. *Nature* **481**, 376-379 (2011).
114. S. Jager *et al.*, Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481**, 371-375 (2011).
115. I. Taniuchi *et al.*, Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell* **111**, 621-633 (2002).

2 Antiviral Mx proteins have an ancient origin and widespread distribution among eukaryotes

This chapter is on bioRxiv (https://doi.org/10.1101/2024.08.06.606855)

2.1 Abstract

First identified in mammals, Mx proteins are potent antivirals against a broad swathe of viruses. Mx proteins arose within the Dynamin superfamily of proteins (DSP), mediating critical cellular processes, such as endocytosis and mitochondrial, plastid, and peroxisomal dynamics. And yet, the evolutionary origins of Mx proteins are poorly understood. Using a series of phylogenomic analyses with stepwise increments in taxonomic coverage, we show that Mx proteins predate the interferon signaling system in vertebrates. Our analyses find an ancient monophyletic DSP lineage in eukaryotes that includes animal Mx proteins with previously undescribed fungal MxF proteins, the relatively uncharacterized plant and algal Dynamin 4A/4C proteins, and representatives from several early-branching eukaryotic lineages. Thus, Mx-like proteins date back close to the origin of Eukarya. Our phylogenetic analyses also reveal that host-encoded and NCLDV (nucleocytoplasmic large DNA viruses)-encoded DSPs are interspersed in four distinct DSP lineages, indicating recurrent viral theft of host DSPs. Our

analyses thus reveal an ancient history of viral and antiviral functions encoded by the Dynamin superfamily in eukaryotes.

2.2 Introduction

The vertebrate interferon (IFN) system acts as a first line of defense against viruses and other pathogens by inducing dozens of interferon-stimulated genes (ISGs) that create an antiviral environment. Functional IFN systems exist in early vertebrates, including bony fishes (1-3). Among the most rapidly and highly expressed ISGs upon interferon induction in human cells are Mx proteins, identified as antiviral proteins soon after the discovery of IFN (4-6). Although Mx proteins from different species have different antiviral specificities, they have an exceptionally broad range of activity. For example, the human MxA protein restricts diverse viruses, including Influenza A (7-9), Vesicular Stomatitis Virus (VSV) (8, 9), measles (10), and hepatitis B (11), whereas the human MxB paralog restricts retroviruses (12-14) and herpesviruses (15, 16). Most mammals have two Mx genes (17) although they were lost or pseudogenized in toothed whales (18). Birds have a single Mx gene, and fish encode up to seven Mx paralogs, which evolved by gene or genome duplication (1-3). Their well-documented presence in fishes and mammalian lineages suggested that the Mx proteins may have arisen coincident with the origin of the interferon system in the common ancestor of bony fishes and mammals. However, recent findings have revealed that ISGs such as STING and cGAS predate vertebrates (19-22). In

addition, there have been two reports of Mx-like genes from invertebrate species (23, 24), suggesting an earlier origin.

Mx proteins are a member of the Dynamin superfamily of proteins (DSP), multi-domain GTPases that mediate many critical cellular processes within eukaryotic cells. Most DSPs localize to distinct cellular membranes, where they facilitate membrane remodeling. For example, Dynamin (or Dyn) proteins localize to the outer cellular membrane (25-33) and endosomes (34), whereas the Optic atrophy 1 (or Opa1) and Mitofusin (or Mfn) proteins act at mitochondrial membranes (35-41) alongside Dynamin-related proteins (or Drps) (42-46). In contrast to other studied DSPs, Mx proteins function independently of membranes (6, 47). Although Mx antiviral mechanisms are still poorly understood, one model proposes that they act by binding viral RNPs (ribonucleoproteins) and exerting a GTP hydrolysis-dependent power stroke to restrict virus replication (48), analogous to the power stroke exerted by Dynamin proteins on cellular membranes in the final step of endocytosis (49, 50).

Previous studies that deeply investigated the evolution and diversification of DSPs (51-53) included either very few or no Mx protein sequences in their analyses, leaving their evolutionary origins unclear. Conversely, studies on Mx evolution have focused exclusively on vertebrate or even mammalian Mx sequences (17, 54). Here, we analyzed the deep phylogenetic history of Mx in the context of DSPs using stepwise increments of eukaryotic phylogenetic coverage. We find unambiguous evidence that Mx-like proteins predate the birth of interferon in animals and are present within plants, fungi, and the majority of basally-branching eukaryotic lineages. Expanding our analyses to all eukaryotic DSPs, we reveal an ancient and ongoing history of lateral transfer between host genomes and nucleocytoplasmic large DNA viruses in four DSP lineages. Overall, our study reveals an ancient lineage of potentially antiviral Mx-like proteins in

eukaryotes and an understudied potential arms race for dynamin-related functions between large double-stranded DNA viruses and their hosts.

2.3 Mx predates the birth of interferon

To evaluate the evolutionary origins of Mx in the context of the broader Dynamin super-family, we carried out BLAST searches on representative metazoan (animal) species with fully sequenced genomes using different human DSPs as queries. Although many DSP genes undergo alternate splicing, we focused only on the longest isoform encoded by each DSP gene. We aligned all metazoan DSPs recovered with different query sequences using the MAFFT program (55). Because the GTPase domain is conserved across different DSPs, we manually extracted the GTPase domain from alignments of different DSPs and used these sequences to generate an all-metazoan DSP GTPase alignment which was further trimmed manually. We subsequently used these sequence alignments to generate phylogenetic trees using FastTree (56, 57) (Figure 1A) or IQ-Tree (58, 59) (Supplementary Figure 1A).

The resulting phylogeny (Figure 1A) reveals five distinct clades of DSPs with high bootstrap support, indicating high confidence in their phylogenetic relatedness. These five clades are consistent with the five DSP groups previously established in animal cells: Dyn, Drp1, Opa1, Mfn, and Mx proteins. Next, we analyzed the domain organization of the different homologs assigned to each of the five metazoan DSP clades using the NCBI conserved protein domain database (CDD) (60, 61). This analysis confirmed that all members of each DSP clade shared the characteristic domain differences previously used to distinguish the different DSPs (Figure 1A). For instance, a hallmark of canonical Dyn proteins is the Pleckstrin homology domain (PH) domain, which facilitates the recruitment of these proteins to the outer cellular membrane (62). Our analyses show that all metazoan sequences constituting the Dyn clade, and only sequences

within this clade, encode a PH domain (Figure 1A). Similarly, all members of the Mfn clade encode the Fzo (fuzzy onion) domain, which appears to be restricted to this clade (Figure 1A). In contrast, the GTPase effector domain (GED) and Middle domains (that separate the GTPase from GED domains) are found in Dyn, Drp, and Mx proteins but not in Opa1 and Mfn proteins. Thus, the GTPase domain is the only universal domain common to all five DSP clades.

Proteins from these five DSP clades localize to distinct cellular compartments in human cells. Dyn proteins, which localize to the outer cellular membrane and other internal cellular membranes (25-34), phylogenetically group with Drp1 proteins, which localize to mitochondria and peroxisomes and are critical for organelle fission/fusion (42-46). Mx antiviral proteins, which have been shown to localize to the cytoplasm, the nucleoplasm, or the nuclear pore, form an outgroup lineage to the Dyn and Drp sister clades (6, 47). Opa1 proteins, which localize to the inner mitochondrial membrane (35-37), are an outgroup to the Mx, Drp, and Dyn clades. Much more basal branching is the Mfn clade, which encodes proteins that localize to the outer mitochondrial membrane (38-41).

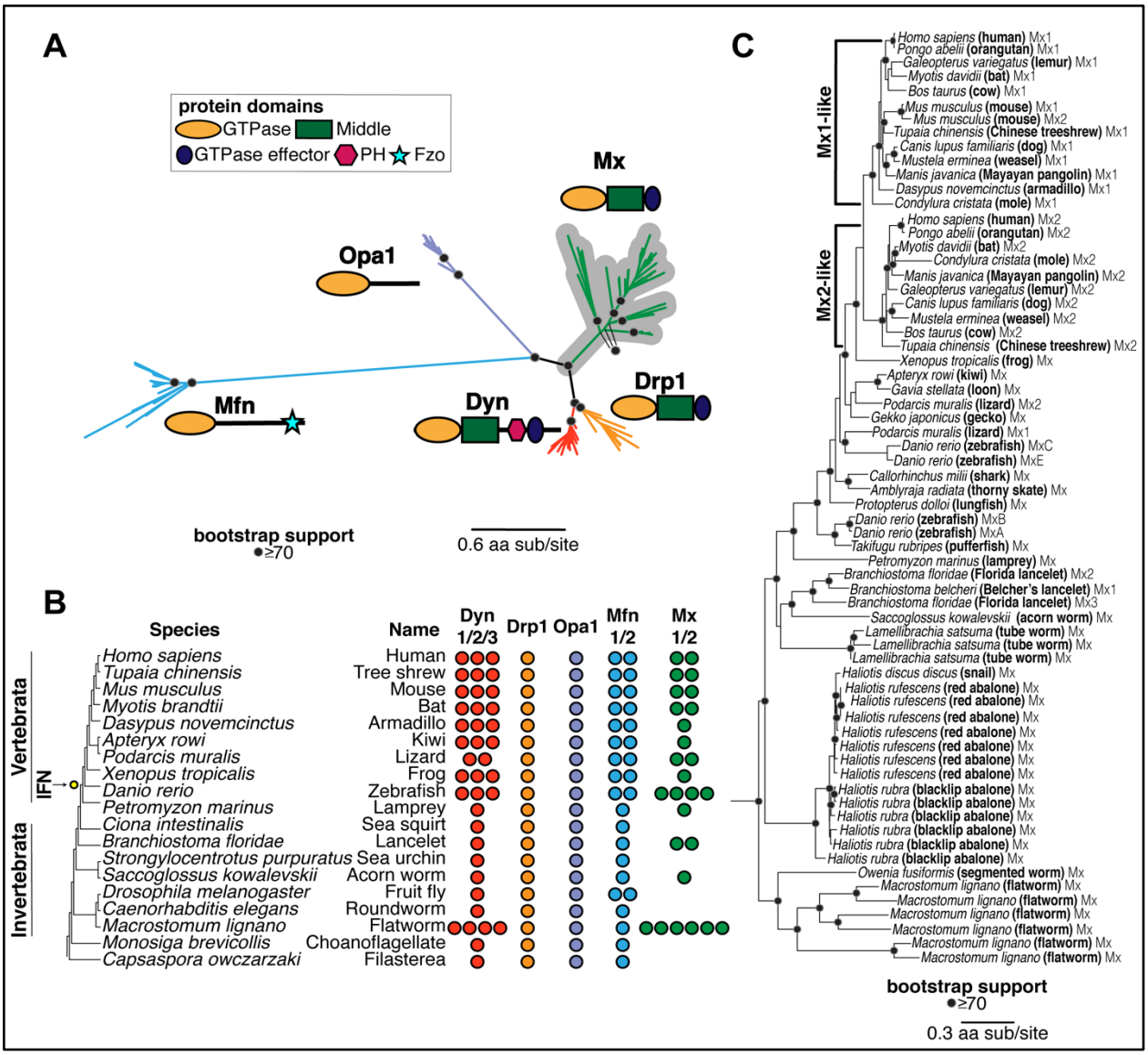
Based on their phylogenetic groupings and protein-domain analysis (Figure 1A), we assigned all DSPs from representative metazoan species to each of the five distinct DSP clades (Figure 1C) to analyze instances of gene loss or duplication. We found that several DSP paralogs from the same clade are often present within the same species. For example, the human genome encodes three Dyn paralogs (Dyn1, Dyn2, Dyn3), two Mfn paralogs (Mfn1, Mfn2), and two Mx paralogs (MxA, MxB) but only one Drp1 and Opa1 (Figure 1C). A Mfn duplication in bony vertebrates gave rise to Mfn2, which modulates antiviral immunity (63, 64),whereas an independent Mfn duplication gave rise to the Marf and Fzo proteins in the Drosophila species. In addition to Dyn duplications in bony vertebrates, there are independent Dyn duplications in at

least two invertebrate species: Macrostonum lignano (flatworm) and Spaheoforma arctica (Icthyosporea). In contrast, Drp1 and Opa1 appear to be encoded by single-copy genes in all metazoans, although we found only a partial Opa1 protein in S. arctica (indicated with a '?' in Figure 1C).

These analyses also confirm the presence of Mx proteins in the vertebrate lineage and its absence in well-studied invertebrate models like Caenorhabditis elegans and D. melanogaster. However, we also find unambiguous evidence (based on bootstrap support and domain analysis above) of the presence of Mx orthologs in many invertebrate species, including Branchistoma floridae (lancelet), Saccoglossus kowalevskii (acorn worm), and Macrostonmum lignano (flatworm), which encodes at least six distinct Mx proteins. These findings reveal that Mx proteins arose in animals much earlier than the origin of the IFN gene network (in bony vertebrates). This ancient origin was followed by recurrent loss of Mx proteins from multiple invertebrate lineages and at least one lineage of mammals (18). This pattern of recurrent gene turnover (loss and duplication) is characteristic of host-virus evolutionary arms races between host and virus, as viral evolution renders some antiviral genes obsolete or imposes pressures on host genomes to expand antiviral functions via gene duplications (65).

To rule out the alternative possibility that the invertebrate Mx homologs might have resulted from horizontal gene transfers following their origins in vertebrates, we expanded our BLAST analyses to identify additional Mx proteins in animal genomes, using invertebrate Mx proteins as queries. We performed phylogenetic analyses using an alignment of the GTPase domain for all DSPs that unambiguously group within the Mx clade (Figure 1D). Our analyses recapitulate and extend findings from previous studies of Mx proteins in mammals (66, 67). We found two lineages of Mx proteins (Mx1 and Mx2) in mammals and Mx representatives from bird,

amphibian, shark, and fish lineages. In invertebrates, we identified Mx homologs in Lamellibrachia satsuma (tube worm), multiple Haliotis species (snail, abalone), and Owenia fusiformis (segmented worm) (Figure 1D) in addition to the previously identified lancelet, acorn worm, and flatworm lineages (Figure 1B, Figure 1C). Most importantly, the topology of the Mx tree largely mirrors the topology of the species tree, consistent with the early origin of Mx proteins in the animal phylogeny. Together, these analyses show that animal Mx proteins are more ancient than the interferon system in vertebrates, have largely been subject to vertical inheritance, and have undergone several lineage-specific gene duplications and losses.



2-1 **Evolutionary origin of Mx in animals predates the interferon signaling network. (A)** Phylogenetic analyses of Dynamin-superfamily proteins (DSPs) based on their common GTPase domain in representative *Holozoa* (animals and their closest single-celled relatives) reveals five distinct DSP clades, consistent with previous analyses: Dyn, Drp1, Mfn, Opa1, and antiviral Mx (gray highlight). **(B)** Summary of localization of different DSPs in human cells. Dyn proteins localize to clathrin-coated pits on the plasma membrane, Drp proteins localize to mitochondria and peroxisomes, whereas Mfn and Opa1 proteins localize to mitochondrial membranes. In contrast, Mx proteins act independent of host membranes and localize to viral ribonucleoprotein (RNP) complexes in the cytoplasm (MxA) or proximal to the nuclear pore (MxB) or to the nucleoplasm (in other mammals, not shown). **(C)** Retention of different DSP clades in representative animals or their outgroup species (tree not drawn to scale). Drp1 and Opa1 are represented in a single copy in all animals and outgroup species, except for *S. arctica* (*Icthyosporea*), where we only recovered a partial Opa1 gene (indicated with a '?'). Mfn is also present in a single copy except in two cases. Mfn duplicated in bony vertebrates, giving rise to Mfn1 and Mfn2, coincident with the birth of the interferon system (IFN, yellow dot). Mfn

independently duplicated in the lineage leading to *D. melanogaster*. Mx proteins are present in 1-4 copies in bony vertebrates and in 1-6 copies in some invertebrate species but have also been independently lost in several lineages. **(D)** Phylogenetic analysis of Mx-like proteins reveals a phylogenetic split between Mx1-like and Mx2-like genes in mammals and independent duplications in several fish lineages. We also find unambiguous evidence of basally-branching Mx-like genes in several invertebrate species, confirming that Mx genes have been vertically inherited in animals, followed by frequent subsequent loss and duplication events. (A), (D) Alignment generated by MAFFT and tree built using FastTree. Black dots indicate nodes with bootstrap support greater than 70% based on FastTree analyses (see methods); a scale bar indicates the level of amino acid divergence.

2.4 Phylogeny of animal, fungal, and plant DSPs reveals ancient Mx orthologs

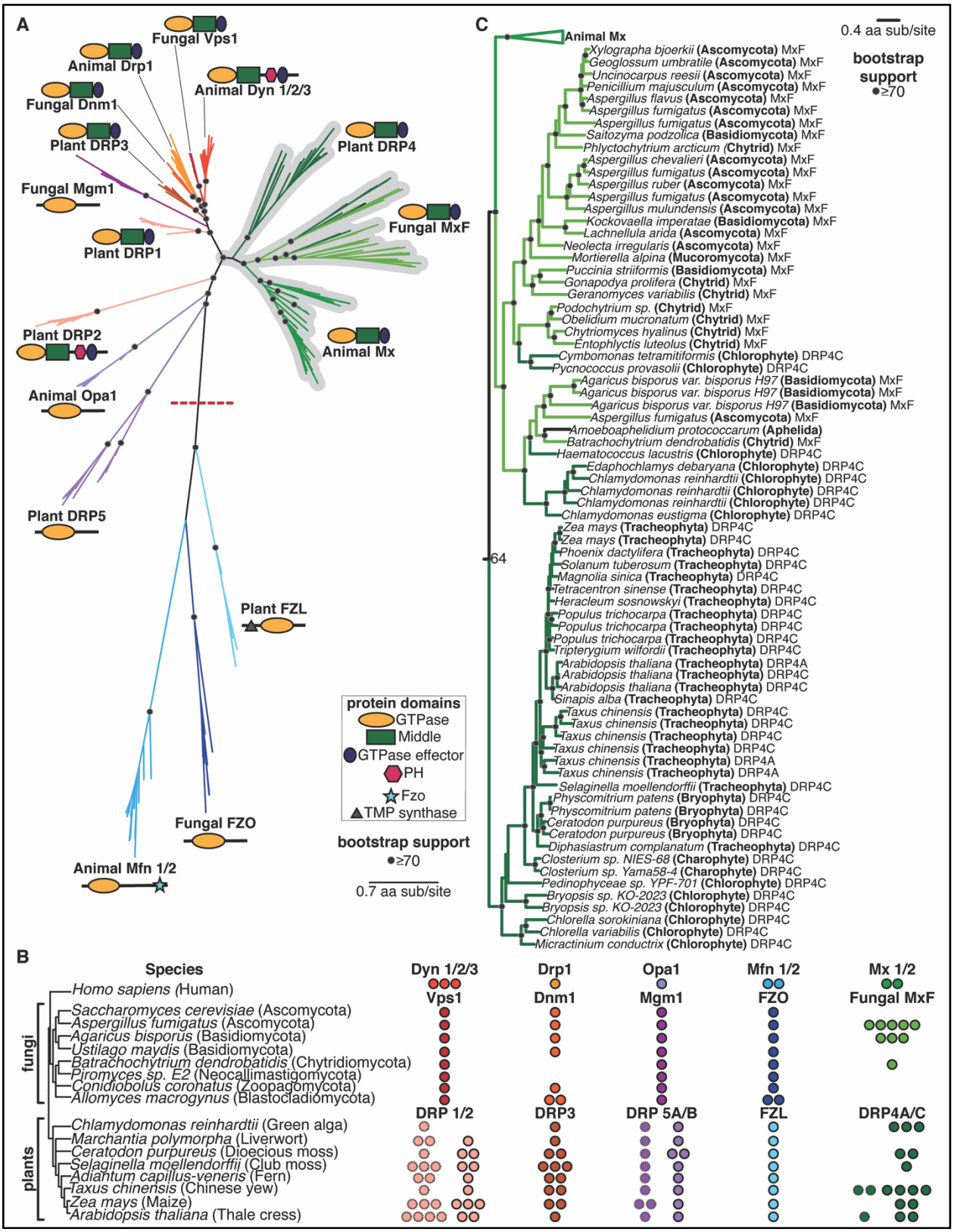
Based on our finding that the Mx clade arose early in the origins of animals, we wanted to extend our analyses of potential Mx origins to two additional lineages – fungi (a sister lineage to animals) and plants – in which DSPs have also been well-studied. Using the same approach of iterative BLAST searches of representative animal, fungal, and plant genomes using different DSP queries, we carried out phylogenetic analyses of all DSPs recovered from these genomes based on their common GTPase domain with FastTree (Figure 2A) and IQ-Tree (Suppl Figure 2A). We also analyzed their domain architecture using the CDD (Figure 2A). These analyses revealed fungal and plant orthologs of animal Mx proteins. For example, animal Mx proteins unambiguously group with uncharacterized DSPs in some fungi, which we rename MxF (for Mx-like proteins from Fungi). For example, Aspergillus fumigatus encodes five MxF proteins, Agaricus bisporus encodes three, and Batrachochytrium dendrobatidis encodes one (Figure 2B). In contrast, many different fungal lineages encode no MxF proteins at all (e.g., Saccharomyces cerevisiae, Ustilago maydis, Piromyces sp. E2, Conidiobolus coronatus, and Allomyces macrogynus) (Figure 2B). This extreme dynamism in copy number is highly reminiscent of the gene loss/ expansion seen in Mx genes in animal genomes but also explains why previous studies failed to identify MxF genes in fungi or misclassified them as Dynamin proteins. Despite their heterogeneous presence, MxF proteins are found in most major clades of fungi (Figure 2C),

including Ascomycota, Basidiomycota, Chytrid, and Mucoromycota (indicated with ‘A’, ‘B’, ‘C’, and ‘M’ in Figure 2C), and, as well as Aphelida, which are believed to the sister lineage to true fungi (68, 69).

Plant DRP4 proteins are also orthologous to animal Mx and fungal MxF proteins (Figure 2A, Figure 2C), which is consistent with a previous proposal (70). Plants encode both full-length DRP4C proteins, which resemble animal Mx proteins in length and domain architecture, and much shorter DRP4A proteins, which appear to often comprise only a GTPase domain with a truncated stalk domain. We find that DRP4A and DRP4C genes from the same species are often more closely related to each other than to orthologs in different plant species, suggesting that DRP4A genes might have arisen independently multiple times in plant evolution from full-length DRP4C proteins. Since we never find plant genomes that only encode the DRP4A proteins, we speculate that shorter DRP4A genes might have recurrently arisen to regulate the activity of full-length DRP4C proteins. Detailed phylogenetic analyses (Figure 2C) reveal that DRP4 proteins are widespread in lineages of green algae and plants and are present in many lineages, including Chlorophytes, Charophytes (also a green algae lineage), Bryophytes (non-vascular plants), and Tracheophytes (vascular plants, which include ferns, gymnosperms, and angiosperms). Our analyses also revealed two obvious instances of potential horizontal gene transfer (HGT) from fungi to chlorophytes (Figure 2C). The first instance occurred (from a chytrid MxF) into the ancestor of two Chlorophytes – Pycnococcus provasolii and Cymbomonas tetramitiformis – while the second event occurred into Haematococcus lacustris, Edaphochlamys debaryana, and Chlamydomonas reinhardtii. These are indicated as DRP4C proteins (Figure 2C) but are more likely to represent MxF proteins. Like with animal Mx and fungal MxF proteins, we identify

extremely dynamic gene turnover within plant DRP4 proteins, which is consistent with their engagement in evolutionary arms races with viruses as bona fide antiviral proteins.

Based on these findings, we conclude that the Mx lineage is much more ancient than previously believed and includes representatives of animals, fungi, and plants.



2-2 **Animal DSP clades, including Mx proteins, in fungal and plant genomes. (A)** Phylogenetic analysis of animal, fungal, and plant DSPs based on their common GTPase domain

reveals broad groupings. Animal Dyn proteins group with fungal Vps1 proteins but not with plant DRP1 or DRP2 proteins, even though only Animal Dyn and plant DRP2 proteins share a C-terminal PH domain. Animal Drp proteins group functionally with fungal Dnm1 and plant DRP3 proteins. Although Fungal Mgm1 is considered the functional equivalent of animal Opa1 and plant DRP5A/5B proteins, our phylogenetic analyses suggest that they are not true orthologs. In contrast, animal Mx proteins appear unambiguously orthologous to uncharacterized fungal MxF proteins and plant DRP4 proteins. Finally, the divergent animal Mfn lineage groups with fungal FZO and plant FZL proteins, which are very divergent from the rest of the DSPs. **(B)** Representation of various DSP classes in representative fungal and plant (and green algae) genomes. Among fungi, Vps1, Mgm1, and FZO are encoded by single-copy genes (except for an *FZO* duplication in *A. macrogynus*). *Mx*-like genes vary from zero to six copies (in *A. fumigatus*). Among plants, FZL is encoded by a single copy gene in all representative algae and plants. DRP5A and DRP5B are mostly also encoded by single-copy genes in plants (except for a DRP5A duplication in maize and a DRP5B duplication in a moss species). DRP1 (hatched outline) varies from one to four copies in plants, whereas DRP2 (solid outline) is present in one to three copies (and absent in green algae). DRP3 varies from one to three copies in all algae and plants. Finally, DRP4C is present from zero to three copies. In addition to DRP4C (solid outline), many plants also encode shorter DRP4A proteins (hatched outline). **(C)** Phylogenetic analysis reveals a monophyletic lineage consisting of animal Mx, fungal MxF (Mx-like proteins from Fungi), and plant DRP4 proteins. MxF proteins are represented in a single lineage found in a variety of fungal lineages, including Ascomycota (A), Basidiomycota (B), Chytrids (C), and Mucoromycota (M), as well as Aphelia, which are a pseudo-fungi-like sister lineage to Fungi. Most Plant DRP4C proteins from various lineages of plants – *Chlorophytes, Charophytes, Bryophytes, and Tracheophytes* are also found in a single lineage, with full-length DRP4C proteins occasionally interspersed with shorter DRP4A proteins. However, some *Chlorophyte* DRP4C proteins group with fungal MxF proteins rather than other DRP4C, which may indicate at least two MxF fungal-to-algal horizontal transfer events. (A), (C) Alignment generated by MAFFT and tree built using FastTree. Black dots indicate nodes with bootstrap support greater than 70% based on FastTree analyses (see methods); a scale bar indicates the level of amino acid divergence.

2.5 Phylogenetic relationships between other animal, fungal, and plant DSPs

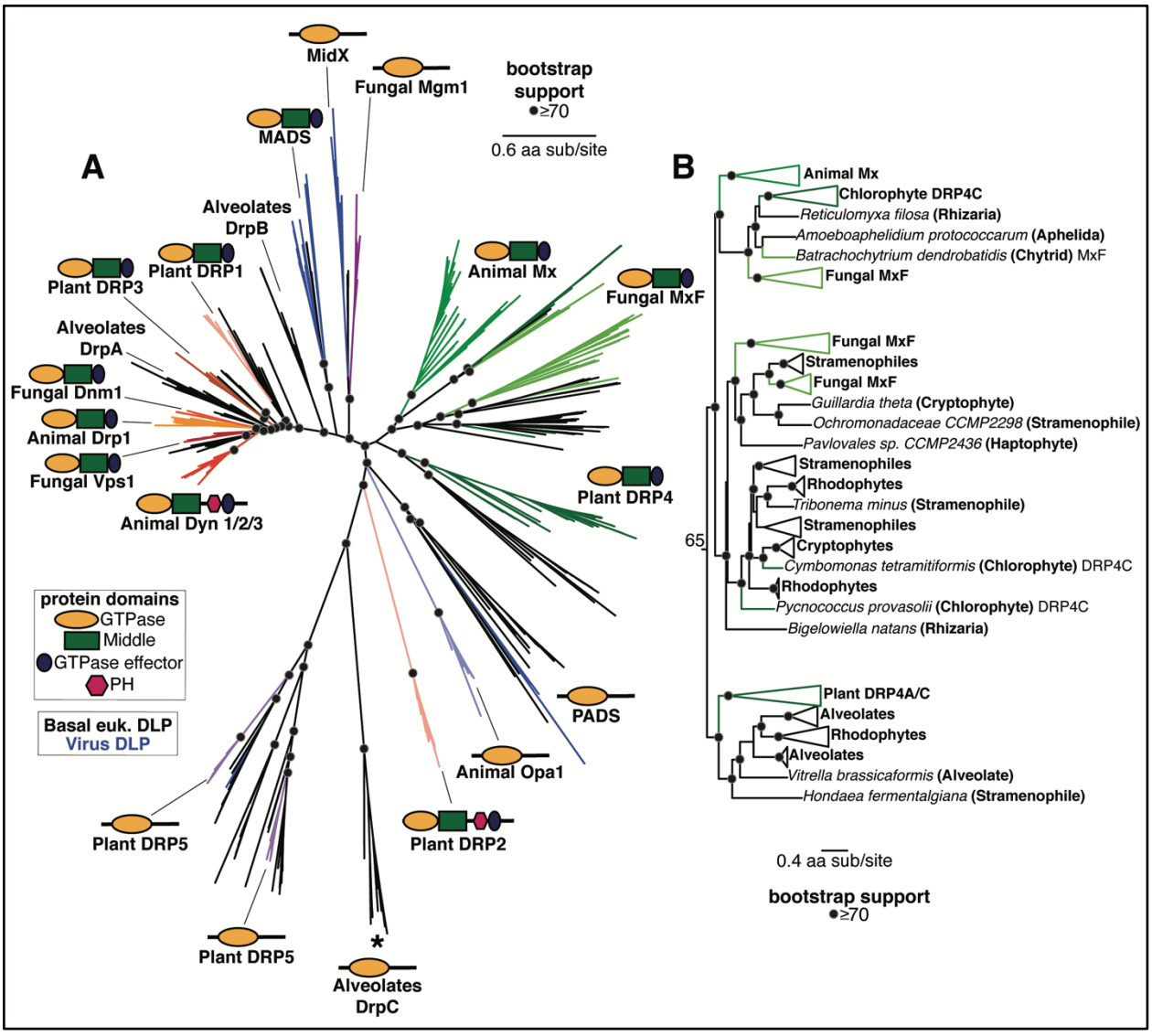
Our analyses also reveal insights into and clarify the phylogenetic relationships between the other DSPs. For example, the ‘Drp’ grouping of animal Drp proteins, fungal Dnm1 (dynamin-related GTPase), and plant DRP3 (Figure 2A) is consistent with their localization and function in mitochondria and peroxisomes (70-75). Dnm1 is encoded by a single copy gene in representative fungi, except for a duplication in Allomyces macrogynus and a loss in both Batrachochytrium dendrobatidis and Piromyces sp. E2, whereas DRP3 genes are present in 1-3 copies in all representative plant species (Figure 2B). The apparent loss of Dnm1 in some fungal

species is unexpected, given their essential roles in many organisms; this might suggest functional redundancy between different DSP clades.

In contrast, the 'Dyn' grouping is more puzzling at first glance. Fungal Vps1 (vacuolar protein sorting) proteins, present in a single copy in most fungi, are closest in sequence and considered the functional equivalent of animal Dyns (76, 77) (Figure 2A). Vps1 proteins are implicated in vacuolar fusion (78-80), membrane scission (79-82), and peroxisomal partitioning (73, 83-85). Most plants encode 1-3 copies of two Dyn-like proteins, DRP1 and DRP2, which play a role in clathrin-mediated endocytosis (70, 86-91) and at the cell plate during cytokinesis (92, 93). And yet, neither fungal Vps1 nor plant DRP1 proteins encode a pleckstrin homology (PH) domain, a defining characteristic of animal Dyn proteins. In contrast, despite being highly divergent from animal Dyn proteins, plant DRP2 proteins (Figure 2A) encode a PH domain. We addressed this apparent contradiction by making separate phylogenies of all Dyn, Drp, and Mx proteins based either on their GTPase domains (as before, Figure 2A) or their Middle and GED domains (Suppl Figure 1), which are also shared among all members of these three DSP clades. Based on the Middle and GED domain phylogeny, we find that plant DRP1 and DRP2 proteins are sister lineages (Suppl Figure 1), even though the DRP2 appears to be much more divergent than DRP1 in the GTPase phylogeny (Figure 2A). We posit that an ancestral plant DRP1/2 protein, encoding a PH domain, duplicated to give rise to DRP1, which lost the PH domain, and DRP2, which likely acquired a divergent GTPase domain via recombination (Suppl Figure 3). An alternative possibility is that plant DRP1 GTPase domains evolved much more rapidly, leading to their divergent placement in the GTPase phylogeny. Although green algae Chlamydomonas reinhardtii only encodes DRP1, most other plants encode 1-3 copies of DRP1 and DRP2 (Figure 2B, Suppl Figure 3).

Fungal Mgm1, which shares the overall domain architecture as animal Opa1, maintains mitochondrial ultrastructure and morphology and regulates mitochondrial fusion like Opa1 (94-96). And yet, the phylogenetic grouping of fungal Mgm1 and animal Opa1 proteins is not very strong (Figure 2A). Like Opa1 in animals, Mgm1 is encoded in a single copy in most fungi, while most plant genomes encode 1-2 copies of each of the DRP5A and DRP5B paralogs (Figure 2B). Based on their similar structure, plant DRP5 proteins should be excellent candidates for being functional equivalents of animal Opa1. However, unlike animal Opa1 proteins, which exclusively function in mitochondria, plant DRP5 proteins function in cytokinesis (75), chloroplast and peroxisome division (97, 98), and mitochondrial morphogenesis/division (99). Thus, the 'Opa1' grouping, consisting of animal Opa1, fungal Mgm1, and plant DRP5 proteins, does not show strong evidence of monophyly (Figure 2A) or functional similarity.

The animal Mfn, fungal FZO, and plant FZL proteins, which are mostly encoded by single-copy genes (Figure 2B), group together to the exclusion of the rest of the DSPs (Figure 2A). Animal Mfn and fungal Fzo proteins mediate the interaction between mitochondrial outer membranes to drive mitochondrial fusion (100, 101), whereas plant FZL proteins localize to chloroplasts and function in thylakoid organization (102). Thus, it is unclear whether this grouping reflects true orthology or is simply a result of their high divergence from the rest of the DSPs.



2-3 **Phylogenetic analysis of DSPs in eukaryotes reveals an ancient Mx lineage. (A) Phylogenetic analysis of the GTPase domain from all eukaryotic DSPs (except for the Mfn/FZO, FZB, and FZL clades) reveals an ancient Dyn/Drp clade that includes** Animal Drp, Plant DRP3, and Fungal Vsp1 along with fungal Dnm1, alveolate DrpA and DrpB, Plant DRP1, and DSPs from several basal branching eukaryotes (shown with black branches). Branching outside the Dyn/Drp cades are the MADS and MidX/ Fungal Mgm1 clades; blue branches indicate DSPs found encoded in viral genomes. Branching next is a single monophyletic clade of Mx-like DSPs (described in more detail in (B)) and the newly discovered PADS lineage. Finally, we find a grouping of animal OPA1, Plant DRP5 (which also contains Stramenopile, Amoebazoa, and a single virus DSP), and Alveolate DrpC. Although Plant DRP2 also groups with this final grouping, we believe its correct phylogenetic position (based on the Middle and GED domains) is as a sister to Plant DRP1 (Suppl Figure 1). **(B)** Our phylogenetic analysis delineates three deeply branching lineages of Mx-like proteins in eukaryotes. The first of these consists of representatives from animals, fungi, algae, and Rhizaria. The second lineage consists of representatives from fungi, Stramenopiles, Haptophytes, Cryptophytes, Chlorophytes, and Rhizaria. Finally, the third deep lineage consists of Mx-like sequences from plants,

alveolates, rhodophytes, and Stramenopiles. (A), (B) Alignment generated by MAFFT and tree built using FastTree. Black dots indicate nodes with bootstrap support greater than 70% based on FastTree analyses (see methods); a scale bar indicates the level of amino acid divergence.

2.6 Deep evolutionary origins of the Mx-like DSPs in eukaryotes

We expanded our survey of Mx-like and other DSP proteins beyond animals, fungi, and plants to diverse basally-branching eukaryotes and eukaryotic viruses (Figure 3A) (Suppl Figure 4A). We did not include Mfn, FZO, and FZL proteins in this analysis since they are quite divergent from the remainder of the eukaryotic DSPs. Our survey identified basally branching eukaryotic representatives in clades that were already well-established (Figure 3A). For example, we found several basal branching eukaryotes, including Stramenopiles, Alveolates, Rhizaria, Amoebazoa, Discoba, and Haptopytes, in the Dyn and Drp clades, suggesting that the Dyn/ Drp clade was already present as a fully specialized, distinct DSP clade in the last eukaryotic common ancestor (LECA).

Our analyses also revealed Mx-like proteins in basally-branching eukaryotes (Figure 3A, Figure 3B). The Mx-like proteins we have identified branch in three deep lineages (Figure 3B). The first lineage comprises animal Mx proteins, fungal MxF proteins (including Chytrids and Aphelida), chlorophyte (green algae) DRP4C, and Mx-like representatives from Rhizaria, unicellular eukaryotes that are part of the TSAR (Telonemia-Stramenopiles-Alveolates-Rhizaria) supergroup of eukaryotes (103, 104). The second lineage consists of Mx-like proteins from Stramenophiles (also referred to as Heterokonts), Cryptophytes (a group of divergent plastid-bearing algae, also referred to as Cryptomonads), Haptophytes (a distinct divergent group of algae), Rhodophytes (red algae), and Rhizaria. The third deep lineage of Mx-like proteins consists of Plant DRP4C, and Mx-like representatives from Alveolates, Rhodophytes, and Stramenopiles. Thus, Mx-like proteins are found in representatives of the majority of the early-

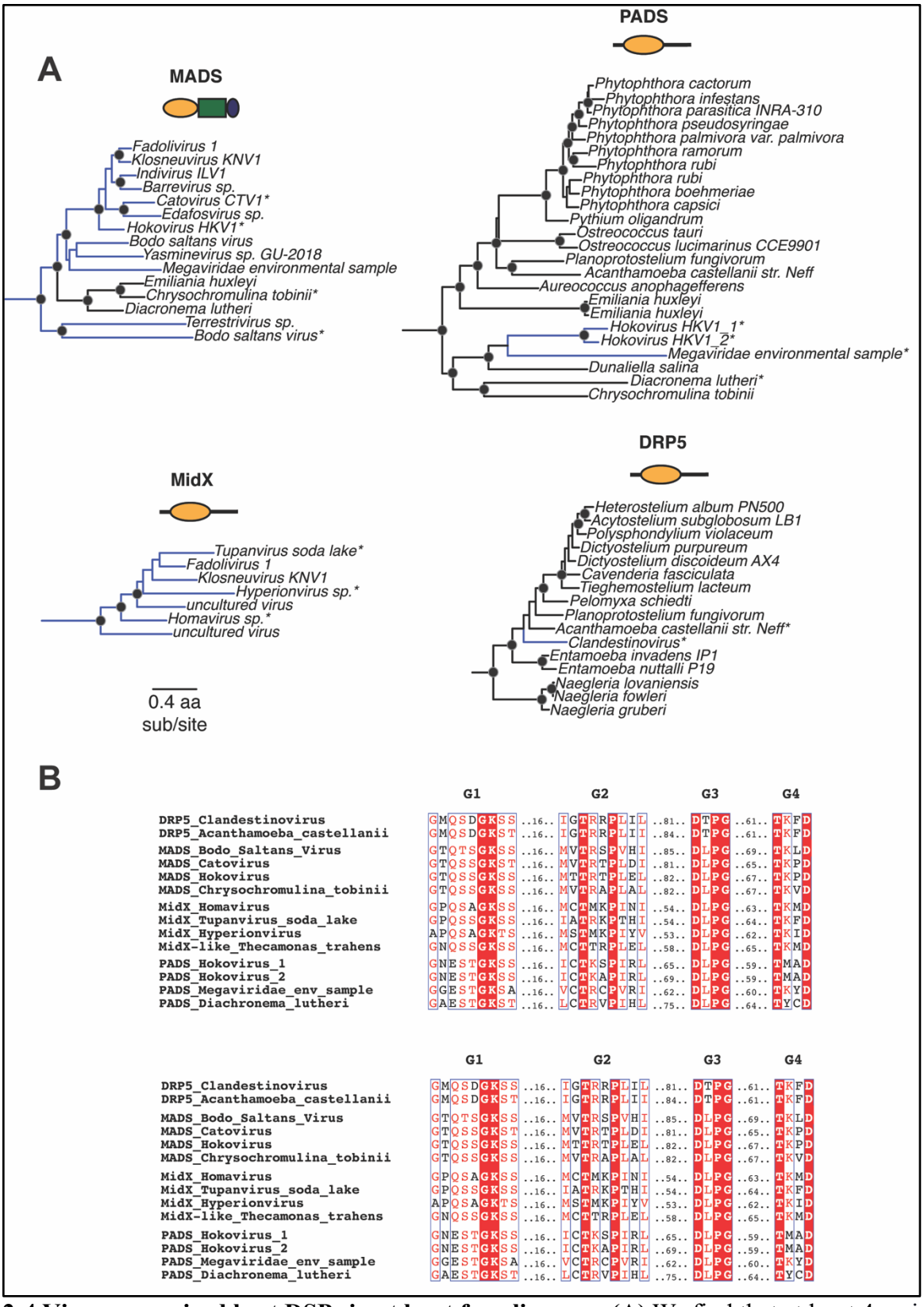
branching supergroups of extant eukaryotes, including TSAR (e.g., Stramenopiles, Alveolates), Haptists (e.g., Haptophytes), Archaeplastida (e.g., red algae, green algae, plants), Cryptista (e.g., Cryptophytes), and Amorphea (e.g., animals, fungi). Based on this representation in early branching eukaryotic supergroups, we infer that Mx proteins arose close to or shortly following LECA, much earlier than previously suspected.

Our analyses also identify eukaryotic DSP lineages that were either previously unidentified (PADS) or only recently identified (MADS, MidX) (53), which provide additional clarity about the phylogenetic relationships between different DSPs. For example, consistent with findings from a recent study (53), we also find the fungal Mgm1 clade to be more closely related to the newly identified MidX clade with high bootstrap support rather than the rest of the Dyn/Drp homologs. Thus, fungal Mgm1 proteins are phylogenetically not as closely related to the lineage that includes animal OPA1, alveolate DrpC, and plant DRP5 proteins. Similarly, the plant DRP2 GTPase domain appears to be more closely related to the MADS lineage of DSPs, consistent with our previous hypothesis that the DRP2 lineage may have swapped its original DRP1-like GTPase domain with a more divergent GTPase (Suppl. Figure 1).

Finally, we uncovered four distinct DSP clades that contain genes from both eukaryotes and viruses, strongly indicative of horizontal gene transfer (Figure 4). For example, the MADS clade (previously described as “Clade D” (53)) is found in haptophytes (Emiliana huxleyi, Chrysochromulina tobinii, and Diacronema lutheri) as well as interspersed lineages of Nucleocytoviricota (nucleocytoplasmic large DNA viruses, or NCLDV) (blue lineages, Figure 4). Similarly, the PADS lineage consists of DSPs from haptophytes, chlorophytes (green algae), oomycetes (which are part of the Stramenopiles), and NCLDVs (Figure 4). Previous studies uncovered a large insertion from an ancient giant virus in oomycete genomes (105), but we

found no evidence that the oomycete PADS sequence originates from this insertion. Our analyses also uncovered MidX sequences from NCLDV DSPs (Figure 4). Although we did not recover any MidX sequences from eukaryotic host genomes based on our analyses of the well-curated non-redundant database in NCBI, metagenomic data has revealed additional host MidX sequences in a recent study (53). Finally, we also recovered a DRP5-like DSP sequence from a Clandestinovirus NCLDV nestled within the DRP5 phylogeny from Amoebozoa (106). In nearly all cases, the viral DSPs have perfectly preserved all the catalytic residues that would indicate retention of GTPase activity (Figure 4B) (107, 108). Thus, PADS, MADS, MidX, and DRP5 DSPs represent lineages that have undergone HGT between eukaryotic host cells and their resident large viruses (Figure 4A).

Thus, in addition to the deep evolutionary origin of Mx proteins in eukaryotes, with recurrent gene turnover — a hallmark of antiviral function — we find that large DNA viruses have also recurrently coopted DSP proteins. Our phylogenetic results suggest that the critical and yet adaptable functionality of DSP proteins has placed them center-stage in the evolutionary battles between hosts and viruses.



2-4 Viruses acquired host DSPs in at least four lineages. (A) We find that at least 4 major DSP clades contain viral DSP sequences, suggesting an ancient history of HGT between host and

virus. MADS mostly contains NCLDV sequences, with three basal branching eukaryotic DSPs as outgroups. PADS consists of a few NCLDV sequences that serve as an outgroup to a diverse clade of basal branching eukaryotes. Our analysis only uncovered viral MidX DSPs. However, previous studies also uncovered eukaryotic representatives from metagenomic data (52). Finally, nestled among DRP5 sequences from *Ameoebozoa* is a single NCLDV DSP. Alignment generated by MAFFT and tree built using FastTree. Black dots indicate nodes with bootstrap support greater than 70% based on FastTree analyses (see methods); the scale bar indicates the level of amino acid divergence for all four clades shown here; asterisks indicate samples used in B. **(B)** Alignment of G Box motifs in viral DSPs indicate preservation of catalytic motifs associated with GTP hydrolysis in viral DSP sequences.

2.7 Discussion

Given their essential roles in cellular remodeling, there has been considerable interest in the evolutionary and functional diversification of Dynamin-like proteins in eukaryotes. Previous analyses have taken advantage of increased information about cellular localization, structural information, and burgeoning sequencing databases to propose schemes for how this critical gene family arose and diversified in eukaryotes (17, 29, 53, 62). We build upon these earlier studies to deeply investigate the origins of the unusual Mx antiviral DSPs in eukaryotes, which was poorly understood. We find that the Mx lineage of DSPs is much more ancient than initially believed. Some earlier reports had suggested that its evolutionary origins coincided with the birth of the interferon system in bony vertebrates (17, 54). Instead, our analyses reveal distinct Mx-like lineages in many animals (including invertebrates), fungi, plants, and basal-branching eukaryotic supergroups. It is unclear whether all of these Mx-like lineages also perform antiviral roles, just like they do in vertebrates, although there are some suggestions this might be the case. For example, some plant DRP4C proteins have been suggested to participate in antiviral function (62, 70, 75). Moreover, if Mx-related proteins were instead performing some necessary cellular membrane remodeling function in non-animal species, we would expect much less genetic turnover; instead, we find both rampant gene amplification (in animals, plants, and fungi) as well as several instances of complete gene loss (in animals and fungi). This gene turnover is more

consistent with antiviral rather than essential cellular housekeeping functions. Our findings suggest that Dynamin specialization for antiviral function occurred early in eukaryotic evolution. The origins of other interferon-induced antiviral genes, such as cGAS, STING, and Viperin, lie in bacteria (20, 21, 109). However, our analyses could not find a bacterial or archaeal ortholog to eukaryotic Mx-like genes, even though at least one bacterial Dynamin-like protein participates in antiviral defense (110).

Our analyses reveal more details about early events of Dynamin specialization, including novel DSP lineages whose cellular functions are still unknown. For example, a recent study described and characterized the MidX lineage, which can remodel mitochondrial membrane topology from inside the matrix (53), unlike Opa1 and Mfn, which remodel mitochondrial membranes from the outer membrane and intermembrane space. In addition to MidX and MADS, we identify the novel PADS lineage. Our findings indicate that much DSP diversification in extant eukaryotes must have already occurred in early eukaryotic evolution. These findings challenge previous classifications of 'modern' versus 'ancient' dynamins that might have reflected what was known at the time about their phylogenetic representation, which was heavily over-represented in well-studied model systems, i.e., animals, fungi, and plants.

Another common theme across the DSPs is the recurrence of horizontal gene transfers (HGTs), including the fungal-to-algal transfer of Mx-like genes (Figure 2C), which implies an exchange of antiviral function in some cases. Understanding the functional consequences of these inferred HGTs might provide further insight into the function of the eukaryotic DSP lineages. Most intriguing among these HGT events is the interspersing of eukaryotic and viral DSPs in at least four distinct DSP lineages – MidX, MADS, PADS, and DRP5 – which suggests an ancient, ongoing history of HGT of DSPs between multiple eukaryotic host lineages and

Nucleocytoviricota (NCLDV). NCLDV are a family of double-stranded DNA viruses, including Mimiviridae, Poxviridae, Asfarviridae, Iridoviridae, and Phycodnaviridae, typified by large genomes and viral particle sizes (111). These viruses infect many eukaryotes, including algae, amoeba, and animals (112-114). Among viruses, only NCLDVs encode their Dynamin-like proteins, suggesting some unique aspect of their cell biology may require Dynamin-like function. Unlike other lineages of double-stranded DNA viruses, NCLDV replication and assembly take place almost entirely within the host cytoplasm, with no discernible steps taking place in the host nucleus. The discovery of novel and complex membrane remodeling in Molliviridae, Mimiviridae, and Poxviridae highlights specific cell biological requirements for assembling large viral particles of NCLDVs in the host cytoplasm (115-119). Thus far, host Dynamin proteins have not been directly implicated in these processes. However, the study of cell biology of most NCLDVs is still in its infancy. Our work and others (53) suggest that NCLDV assembly of new envelope-bound virions may require or be enhanced by Dynamin-like activity, which may have spurred at least some NCLDVs to acquire and repurpose host DSPs for their function. An alternative model is that NCLDVs acquire host DSPs specifically to interfere with host DSP-mediated processes by a 'dominant-negative' model. However, if this were the case, we would not expect universal retention of all features required for GTPase catalytic activity in viral DSPs.

Our study highlights two ancient host-virus battles for Dynamin-like functions in eukaryotic lineages. First, we show that the Mx clade of DSPs is ancient, which may represent a DSP specialization for antiviral function early in eukaryotic evolution. Focused studies on characterizing representatives of this clade in fungi, plants, and basal branching eukaryotes would reveal more insight into Mx function and mechanism in animals. Second, we find a recurrent pattern of DSP acquisition by different lineages of NCLDVs, which also suggests that

host and viral DSPs may perform critical yet poorly understudied functions, at least in this widespread lineage of viruses. Understanding how these putative arms races occur for cellular and viral membrane remodeling may reveal novel aspects of viral biology and host defense. Thus, in addition to their canonical cell biological roles in membrane remodeling, our study reveals that Dynamin-superfamily proteins (DSPs) have played critical antiviral and viral functions through most of eukaryotic evolution.

2.8 Materials and Methods

Phylogenetic trees

DSPs were identified via iterative BLAST searches on representative species with fully sequenced genomes using bona fide DSPs as queries. In cases with multiple hits from the same species, only the longest isoform encoded by each DSP gene was utilized in downstream analysis. We used Clustal Omega (120) or MAFFT (55) to align sequences obtained from the same query. From these alignments, the GTPase domains were extracted and realigned using MAFFT to generate an alignment of GTPase domains from all DSPs. This alignment was then manually inspected to eliminate incomplete sequences and used to generate phylogenetic trees using FastTree (56, 57) with default settings, which was also used to perform bootstrap analyses. Trees were also subsequently built using IQ-Tree under default settings with UltraFast Bootstraps (58) using the LG+I+G4 model. Resulting trees were then manually annotated to indicate the level of bootstrap support and to highlight groupings. We expanded our search space in incremental steps of increasing phylogenetic coverage, starting first with animals, then to plants and fungi, then to all eukaryotes and viruses. More focused trees (Figure 1D, Figure 2C, Figure 3B and Figure 4) were excised from the larger trees (Figure 1A, Figure 2A, Figure 3A, respectively).

Protein domain analysis

We uploaded full-length alignments of sequences retrieved from iterative Blast searches to the NCBI Conserved Domains tool using the Batch-CD search tool (60, 61). We present schematic versions of the domain analyses in our figures. We restrict our analyses only to those domains that were reliably identified by the NCBI CD tool. However, in certain cases, we additionally searched selected full-length DSP sequences using subsequent manual BLAST searches to confirm the absence of individual domains, such as the pleckstrin-homology (PH), Fuzzy onion (Fzo), or the TMP-synthase domains. We also visually inspected the GTPase domains identified to ensure that they bore all the hallmarks of catalytically active domains, including preservation of the previously identified catalytic residues.

Cell biological localization

To represent the cytological location of different DSPs (Figure 2B), we focused on previously published cell biological studies in human cells.

DSP repertoires in representative genomes

To comprehensively identify all DSPs in representative genomes, we performed BLAST searches to specific fully sequenced genomes using the ref_seq database using representative DSPs from each of the identified clades, taking care to correctly and comprehensively identify true paralogs. We used phylogenetic analyses and (when possible) shared synteny analyses to distinguish between orthologs and paralogs. These analyses allowed us to identify cases of gene duplication within clades and confidently identify cases of specific gene losses of specific DSP clades for each representative genome sequenced.

GTPase catalytic residue alignment

Representative sequences were selected from each viral DSP clade. These were aligned and the G1, G2/Signal I, G3, and G4 boxes were manually extracted, concatemerized, and realigned using MAFFT. This alignment was run on ESPript 3 (107) with default settings and the result was divided manually to display the different G Box motifs.

2.9 References

1. P. Boudinot, C. Langevin, C. J. Secombes, J. P. Levraud, The Peculiar Characteristics of Fish Type I Interferons. *Viruses* **8** (2016).
2. B. Robertsen, The interferon system of teleost fish. *Fish Shellfish Immunol* **20**, 172-191 (2006).
3. Y. B. Zhang, J. F. Gui, Molecular regulation of interferon antiviral response in fish. *Dev Comp Immunol* **38**, 193-202 (2012).
4. A. Isaacs, J. Lindenmann, Virus interference. I. The interferon. *Proc R Soc Lond B Biol Sci* **147**, 258-267 (1957).
5. M. A. Horisberger, P. Staeheli, O. Haller, Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. *Proc Natl Acad Sci U S A* **80**, 1910-1914 (1983).
6. P. Staeheli, O. Haller, W. Boll, J. Lindenmann, C. Weissmann, Mx protein: constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. *Cell* **44**, 147-158 (1986).
7. J. Dittmann *et al.*, Influenza A virus strains differ in sensitivity to the antiviral action of Mx-GTPase. *J Virol* **82**, 3624-3631 (2008).
8. J. Pavlovic, T. Zurcher, O. Haller, P. Staeheli, Resistance to influenza virus and vesicular stomatitis virus conferred by expression of human MxA protein. *J Virol* **64**, 3370-3375 (1990).
9. P. Staeheli, J. Pavlovic, Inhibition of vesicular stomatitis virus mRNA synthesis by human MxA protein. *J Virol* **65**, 4498-4501 (1991).
10. S. Schneider-Schaulies *et al.*, Cell type-specific MxA-mediated inhibition of measles virus transcription in human brain cells. *J Virol* **68**, 6910-6917 (1994).
11. E. Gordien *et al.*, Inhibition of hepatitis B virus replication by the interferon-inducible MxA protein. *J Virol* **75**, 2684-2691 (2001).
12. C. Goujon *et al.*, Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559-562 (2013).
13. M. Kane *et al.*, MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-566 (2013).
14. Z. Liu *et al.*, The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe* **14**, 398-410 (2013).
15. M. Crameri *et al.*, MxB is an interferon-induced restriction factor of human herpesviruses. *Nat Commun* **9**, 1980 (2018).
16. M. Schilling *et al.*, Human MxB Protein Is a Pan-herpesvirus Restriction Factor. *J Virol* **92** (2018).
17. J. Verhelst, P. Hulpiau, X. Saelens, Mx proteins: antiviral gatekeepers that restrain the uninvited. *Microbiol Mol Biol Rev* **77**, 551-566 (2013).
18. B. A. Braun, A. Marcovitz, J. G. Camp, R. Jia, G. Bejerano, Mx1 and Mx2 key antiviral proteins are surprisingly lost in toothed whales. *Proc Natl Acad Sci U S A* **112**, 8036-8040 (2015).
19. X. Gui *et al.*, Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. *Nature* **567**, 262-266 (2019).
20. P. J. Kranzusch *et al.*, Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol Cell* **59**, 891-903 (2015).
21. S. R. Margolis, S. C. Wilson, R. E. Vance, Evolutionary Origins of cGAS-STING Signaling. *Trends Immunol* **38**, 733-743 (2017).
22. X. Wu *et al.*, Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res* **42**, 8243-8257 (2014).

23. M. De Zoysa *et al.*, First report of invertebrate Mx: cloning, characterization and expression analysis of Mx cDNA in disk abalone (Haliotis discus discus). *Fish Shellfish Immunol* **23**, 86-96 (2007).
24. Y. Li, X. Qiao, Z. Liu, L. Wang, L. Song, A myxovirus resistance like protein involved in CgIFNLP mediated immune response of oyster Crassostrea gigas. *Fish Shellfish Immunol* **119**, 318-328 (2021).
25. B. Antonny *et al.*, Membrane fission by dynamin: what we know and what we need to know. *EMBO J* **35**, 2270-2284 (2016).
26. S. M. Ferguson, P. De Camilli, Dynamin, a membrane-remodelling GTPase. *Nat Rev Mol Cell Biol* **13**, 75-88 (2012).
27. A. M. Gonzalez-Jamett *et al.*, Dynamin-2 function and dysfunction along the secretory pathway. *Front Endocrinol (Lausanne)* **4**, 126 (2013).
28. A. M. Gonzalez-Jamett *et al.*, Dynamin-2 in nervous system disorders. *J Neurochem* **128**, 210-223 (2014).
29. G. J. Praefcke, H. T. McMahon, The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat Rev Mol Cell Biol* **5**, 133-147 (2004).
30. M. Singh, H. Denny, C. Smith, J. Granados, R. Renden, Presynaptic loss of dynamin-related protein 1 impairs synaptic vesicle release and recycling at the mouse calyx of Held. *J Physiol* **596**, 6263-6287 (2018).
31. H. Damke, T. Baba, D. E. Warnock, S. L. Schmid, Induction of mutant dynamin specifically blocks endocytic coated vesicle formation. *J Cell Biol* **127**, 915-934 (1994).
32. D. Loerke *et al.*, Cargo and dynamin regulate clathrin-coated pit maturation. *PLoS Biol* **7**, e57 (2009).
33. J. Z. Rappoport, S. M. Simon, Real-time analysis of clathrin-mediated endocytosis during cell migration. *J Cell Sci* **116**, 847-855 (2003).
34. P. Nicoziani *et al.*, Role for dynamin in late endosome dynamics and trafficking of the cation-independent mannose 6-phosphate receptor. *Mol Biol Cell* **11**, 481-495 (2000).
35. S. Cipolat, O. Martins de Brito, B. Dal Zilio, L. Scorrano, OPA1 requires mitofusin 1 to promote mitochondrial fusion. *Proc Natl Acad Sci U S A* **101**, 15927-15932 (2004).
36. A. Olichon *et al.*, The human dynamin-related protein OPA1 is anchored to the mitochondrial inner membrane facing the inter-membrane space. *FEBS Lett* **523**, 171-176 (2002).
37. A. Olichon *et al.*, Loss of OPA1 perturbates the mitochondrial inner membrane structure and integrity, leading to cytochrome c release and apoptosis. *J Biol Chem* **278**, 7743-7746 (2003).
38. A. Santel, M. T. Fuller, Control of mitochondrial morphology by a human mitofusin. *J Cell Sci* **114**, 867-874 (2001).
39. M. Rojo, F. Legros, D. Chateau, A. Lombes, Membrane topology and mitochondrial targeting of mitofusins, ubiquitous mammalian homologs of the transmembrane GTPase Fzo. *J Cell Sci* **115**, 1663-1674 (2002).
40. F. Legros, A. Lombes, P. Frachon, M. Rojo, Mitochondrial fusion in human cells is efficient, requires the inner membrane potential, and is mediated by mitofusins. *Mol Biol Cell* **13**, 4343-4354 (2002).
41. Y. Eura, N. Ishihara, S. Yokota, K. Mihara, Two mitofusin proteins, mammalian homologues of FZO, with distinct functions are both required for mitochondrial fusion. *J Biochem* **134**, 333-344 (2003).
42. E. Smirnova, L. Griparic, D. L. Shurland, A. M. van der Bliek, Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Mol Biol Cell* **12**, 2245-2256 (2001).
43. K. L. Cerveny, Y. Tamura, Z. Zhang, R. E. Jensen, H. Sesaki, Regulation of mitochondrial fusion and division. *Trends Cell Biol* **17**, 563-569 (2007).
44. H. Chen, D. C. Chan, Mitochondrial dynamics--fusion, fission, movement, and mitophagy--in neurodegenerative diseases. *Hum Mol Genet* **18**, R169-176 (2009).
45. C. Frohlich *et al.*, Structural insights into oligomerization and mitochondrial remodelling of dynamin 1-like protein. *EMBO J* **32**, 1280-1292 (2013).
46. P. H. Reddy *et al.*, Dynamin-related protein 1 and mitochondrial fragmentation in neurodegenerative diseases. *Brain Res Rev* **67**, 103-118 (2011).
47. O. Haller, S. Gao, A. von der Malsburg, O. Daumke, G. Kochs, Dynamin-like MxA GTPase: structural insights into oligomerization and implications for antiviral activity. *J Biol Chem* **285**, 28419-28424 (2010).
48. M. L. Rennie, S. A. McKelvie, E. M. Bulloch, R. L. Kingston, Transient dimerization of human MxA promotes GTP hydrolysis, resulting in a mechanical power stroke. *Structure* **22**, 1433-1445 (2014).
49. J. S. Chappie *et al.*, A pseudoatomic model of the dynamin polymer identifies a hydrolysis-dependent powerstroke. *Cell* **147**, 209-222 (2011).
50. Y. Chen *et al.*, Conformational dynamics of dynamin-like MxA revealed by single-molecule FRET. *Nat Commun* **8**, 15744 (2017).

51. R. Purkanti, M. Thattai, Ancient dynamin segments capture early stages of host-mitochondrial integration. *Proc Natl Acad Sci U S A* **112**, 2800-2805 (2015).
52. S. Sinha, N. Manoj, Molecular evolution of proteins mediating mitochondrial fission-fusion dynamics. *FEBS Lett* **593**, 703-718 (2019).
53. S. Sheikh *et al.*, A Novel Group of Dynamin-Related Proteins Shared by Eukaryotes and Giant Viruses Is Able to Remodel Mitochondria From Within the Matrix. *Mol Biol Evol* **40** (2023).
54. F. Qi *et al.*, Birth and death of Mx genes and the presence/absence of genes regulating Mx transcription are correlated with the diversity of anti-pathogenicity in vertebrate species. *Mol Genet Genomics* **294**, 121-133 (2019).
55. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).
56. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-1650 (2009).
57. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
58. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).
59. B. Q. Minh, M. A. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **30**, 1188-1195 (2013).
60. S. Lu *et al.*, CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, D265-D268 (2020).
61. J. Wang *et al.*, The conserved domain database in 2023. *Nucleic Acids Res* **51**, D384-D388 (2023).
62. R. Ramachandran, S. L. Schmid, The dynamin superfamily. *Curr Biol* **28**, R411-R416 (2018).
63. K. Yasukawa *et al.*, Mitofusin 2 inhibits mitochondrial antiviral signaling. *Sci Signal* **2**, ra47 (2009).
64. H. Deng *et al.*, Mfn2 is responsible for inhibition of the RIG-I/IRF7 pathway and activation of NLRP3 inflammasome in Seneca Valley virus-infected PK-15 cells to promote viral replication. *Front Immunol* **13**, 955671 (2022).
65. M. D. Daugherty, H. S. Malik, Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet* **46**, 677-700 (2012).
66. P. S. Mitchell *et al.*, Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* **12**, 598-604 (2012).
67. P. S. Mitchell, M. Emerman, H. S. Malik, An evolutionary perspective on the broad antiviral specificity of MxA. *Curr Opin Microbiol* **16**, 493-499 (2013).
68. S. A. Karpov *et al.*, Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front Microbiol* **5**, 112 (2014).
69. G. Torruella *et al.*, Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Curr Biol* **25**, 2404-2410 (2015).
70. Z. Hong *et al.*, A unified nomenclature for Arabidopsis dynamin-related large GTPases based on homology and possible functions. *Plant Mol Biol* **53**, 261-265 (2003).
71. A. Legesse-Miller, R. H. Massol, T. Kirchhausen, Constriction and Dnm1p recruitment are distinct processes in mitochondrial fission. *Mol Biol Cell* **14**, 1953-1963 (2003).
72. D. Otsuga *et al.*, The dynamin-related GTPase, Dnm1p, controls mitochondrial morphology in yeast. *J Cell Biol* **143**, 333-349 (1998).
73. K. Kuravi *et al.*, Dynamin-related proteins Vps1p and Dnm1p control peroxisome abundance in Saccharomyces cerevisiae. *J Cell Sci* **119**, 3994-4001 (2006).
74. S. Arimura, N. Tsutsumi, A dynamin-like protein (ADL2b), rather than FtsZ, is involved in Arabidopsis mitochondrial division. *Proc Natl Acad Sci U S A* **99**, 5727-5731 (2002).
75. S. Y. Miyagishima, H. Kuwayama, H. Urushihara, H. Nakanishi, Evolutionary linkage between eukaryotic cytokinesis and chloroplast division by dynamin proteins. *Proc Natl Acad Sci U S A* **105**, 15202-15207 (2008).
76. E. Conibear, T. H. Stevens, Vacuolar biogenesis in yeast: sorting out the sorting proteins. *Cell* **83**, 513-516 (1995).
77. C. A. Vater, C. K. Raymond, K. Ekena, I. Howald-Stevenson, T. H. Stevens, The VPS1 protein, a homolog of dynamin required for vacuolar protein sorting in Saccharomyces cerevisiae, is a GTPase with two functionally separable domains. *J Cell Biol* **119**, 773-786 (1992).

78. A. Kulkarni, K. Alpadi, T. Sirupangi, C. Peters, A dynamin homolog promotes the transition from hemifusion to content mixing in intracellular membrane fusion. *Traffic* **15**, 558-571 (2014).
79. K. Alpadi *et al.*, Dynamin-SNARE interactions control trans-SNARE formation in intracellular membrane fusion. *Nat Commun* **4**, 1704 (2013).
80. C. Peters, T. L. Baars, S. Buhler, A. Mayer, Mutual control of membrane fission and fusion proteins. *Cell* **119**, 667-678 (2004).
81. R. J. Park *et al.*, Dynamin triple knockout cells reveal off target effects of commonly used dynamin inhibitors. *J Cell Sci* **126**, 5305-5312 (2013).
82. R. Smaczynska-de, II *et al.*, A role for the dynamin-like protein Vps1 during endocytosis in yeast. *J Cell Sci* **123**, 3496-3506 (2010).
83. D. Hoepfner, M. van den Berg, P. Philippsen, H. F. Tabak, E. H. Hettema, A role for Vps1p, actin, and the Myo2p motor in peroxisome abundance and inheritance in Saccharomyces cerevisiae. *J Cell Biol* **155**, 979-990 (2001).
84. T. Guo *et al.*, A signal from inside the peroxisome initiates its division by promoting the remodeling of the peroxisomal membrane. *J Cell Biol* **177**, 289-303 (2007).
85. F. J. Vizeacoumar, J. C. Torres-Guzman, Y. Y. Tam, J. D. Aitchison, R. A. Rachubinski, YHR150w and YDR479c encode peroxisomal integral membrane proteins involved in the regulation of peroxisome number, size, and distribution in Saccharomyces cerevisiae. *J Cell Biol* **161**, 321-332 (2003).
86. S. K. Backues, S. Y. Bednarek, Arabidopsis dynamin-related protein 1A polymers bind, but do not tubulate, liposomes. *Biochem Biophys Res Commun* **393**, 734-739 (2010).
87. C. A. Konopka, S. Y. Bednarek, Comparison of the dynamics and functional redundancy of the Arabidopsis dynamin-related isoforms DRP1A and DRP1C during plant development. *Plant Physiol* **147**, 1590-1602 (2008).
88. C. A. Konopka, S. K. Backues, S. Y. Bednarek, Dynamics of Arabidopsis dynamin-related protein 1C and a clathrin light chain at the plasma membrane. *Plant Cell* **20**, 1363-1380 (2008).
89. M. Fujimoto *et al.*, Arabidopsis dynamin-related proteins DRP2B and DRP1A participate together in clathrin-coated vesicle formation during endocytosis. *Proc Natl Acad Sci U S A* **107**, 6094-6099 (2010).
90. B. C. Lam, T. L. Sage, F. Bianchi, E. Blumwald, Regulation of ADL6 activity by its associated molecular network. *Plant J* **31**, 565-576 (2002).
91. N. G. Taylor, A role for Arabidopsis dynamin related proteins DRP2A/B in endocytosis; DRP2 function is essential for plant growth. *Plant Mol Biol* **76**, 117-129 (2011).
92. M. Fujimoto, S. Arimura, M. Nakazono, N. Tsutsumi, Arabidopsis dynamin-related protein DRP2B is co-localized with DRP1A on the leading edge of the forming cell plate. *Plant Cell Rep* **27**, 1581-1586 (2008).
93. J. B. Jin *et al.*, The Arabidopsis dynamin-like proteins ADL1C and ADL1E play a critical role in mitochondrial morphogenesis. *Plant Cell* **15**, 2357-2369 (2003).
94. C. Alexander *et al.*, OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat Genet* **26**, 211-215 (2000).
95. P. Belenguer, L. Pellegrini, The dynamin GTPase OPA1: more than mitochondria? *Biochim Biophys Acta* **1833**, 176-183 (2013).
96. Z. Song, M. Ghochani, J. M. McCaffery, T. G. Frey, D. C. Chan, Mitofusins and OPA1 mediate sequential steps in mitochondrial membrane fusion. *Mol Biol Cell* **20**, 3525-3532 (2009).
97. X. Zhang, J. Hu, The Arabidopsis chloroplast division protein DYNAMIN-RELATED PROTEIN5B also mediates peroxisome division. *Plant Cell* **22**, 431-442 (2010).
98. H. Gao, D. Kadirjan-Kalbach, J. E. Froehlich, K. W. Osteryoung, ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc Natl Acad Sci U S A* **100**, 4328-4333 (2003).
99. K. Aung, J. Hu, Differential roles of Arabidopsis dynamin-related proteins DRP3A, DRP3B, and DRP5B in organelle division. *J Integr Plant Biol* **54**, 921-931 (2012).
100. G. J. Hermann *et al.*, Mitochondrial fusion in yeast requires the transmembrane GTPase Fzo1p. *J Cell Biol* **143**, 359-373 (1998).
101. A. D. Mozdy, J. M. Shaw, A fuzzy mitochondrial fusion apparatus comes into focus. *Nat Rev Mol Cell Biol* **4**, 468-478 (2003).
102. H. Gao, T. L. Sage, K. W. Osteryoung, FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. *Proc Natl Acad Sci U S A* **103**, 6759-6764 (2006).
103. F. Burki *et al.*, Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol Evol* **1**, 231-238 (2009).

104. F. Burki *et al.*, Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC Evol Biol* **10**, 377 (2010).
105. V. Sharma, P. Colson, R. Giorgi, P. Pontarotti, D. Raoult, DNA-dependent RNA polymerase detects hidden giant viruses in published databanks. *Genome Biol Evol* **6**, 1603-1610 (2014).
106. C. Rolland *et al.*, Clandestinovirus: A Giant Virus With Chromatin Proteins and a Potential to Manipulate the Cell Cycle of Its Host Vermamoeba vermiformis. *Front Microbiol* **12**, 715608 (2021).
107. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).
108. O. Daumke, G. J. Praefcke, Invited review: Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily. *Biopolymers* **105**, 580-593 (2016).
109. E. M. Culbertson, T. C. Levin, Eukaryotic CD-NTase, STING, and viperin proteins evolved via domain shuffling, horizontal transfer, and ancient inheritance from prokaryotes. *PLoS Biol* **21**, e3002436 (2023).
110. L. Guo, L. Sattler, S. Shafqat, P. L. Graumann, M. Bramkamp, A Bacterial Dynamin-Like Protein Confers a Novel Phage Resistance Strategy on the Population Level in Bacillus subtilis. *mBio* **13**, e0375321 (2021).
111. H. A. M. Monttinen, C. Bicep, T. A. Williams, R. P. Hirt, The genomes of nucleocytoplasmic large DNA viruses: viral evolution writ large. *Microb Genom* **7** (2021).
112. N. Delaroque, W. Boland, The genome of the brown alga Ectocarpus siliculosus contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. *BMC Evol Biol* **8**, 110 (2008).
113. B. La Scola *et al.*, A giant virus in amoebae. *Science* **299**, 2033 (2003).
114. T. W. Sun *et al.*, Host Range and Coding Potential of Eukaryotic Giant Viruses. *Viruses* **12** (2020).
115. J. K. Hyun *et al.*, Membrane remodeling by the double-barrel scaffolding protein of poxvirus. *PLoS Pathog* **7**, e1002239 (2011).
116. Y. Mutsafi, E. Shimoni, A. Shimon, A. Minsky, Membrane assembly during the infection cycle of the giant Mimivirus. *PLoS Pathog* **9**, e1003367 (2013).
117. Y. Fridmann-Sirkis *et al.*, Efficiency in Complexity: Composition and Dynamic Nature of Mimivirus Replication Factories. *J Virol* **90**, 10039-10047 (2016).
118. A. Andrade *et al.*, Filling Knowledge Gaps for Mimivirus Entry, Uncoating, and Morphogenesis. *J Virol* **91** (2017).
119. E. R. Quemin *et al.*, Complex Membrane Remodeling during Virion Assembly of the 30,000-Year-Old Mollivirus Sibericum. *J Virol* **93** (2019).
120. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

3 The structural basis for HIV-1 Vif antagonism of human APOBEC3G

This chapter is currently published (https://doi.org/10.1038/s41586-023-05779-1)

Li, YL., Langley, C.A., Azumaya, C.M., Echeverria, I., Chesarino, N.M., Emerman, M., Cheng. Y., Gross, J.D. The structural basis for HIV-1 Vif antagonism of human APOBEC3G. *Nature* **615**, 728–733 (2023). https://doi.org/10.1038/s41586-023-05779-1

3.1 Abstract

The APOBEC3 (A3) proteins are host antiviral cellular proteins that hypermutate the viral genome of diverse viral families. In retroviruses, this process requires A3 packaging into viral particles[1-4]. The lentiviruses encode a protein Vif that antagonizes A3 family members by targeting them for degradation. Diversification of A3 allows host escape from Vif whereas adaptations in Vif enable cross-species transmission of primate lentiviruses. How this molecular arms race plays out at the structural level is unknown. Here, we report the cryogenic electron microscopy structure of human APOBE3G (A3G) bound to HIV-1 Vif and hijacked cellular proteins that promote ubiquitin mediated proteolysis. A small surface explains the molecular arms race, including a cross-species transmission event that led to the birth of HIV-1. Unexpectedly, we find RNA is a molecular glue for the Vif-A3G interaction, enabling Vif to repress A3G by ubiquitin dependent and independent mechanisms. Our results suggest a model where Vif antagonizes A3G by intercepting it in its most dangerous form for the virus, when bound to RNA and on pathway to packaging, to prevent viral restriction. By engaging essential surfaces required for restriction, Vif exploits a vulnerability in A3G, suggesting a general mechanism where RNA binding helps position key residues necessary for viral antagonism of a host antiviral gene.

3.2 Introduction

The APOBEC3 proteins are host cytosine deaminases with the capacity to mutate viral genomes across many different viral families (reviewed in [5-7]). APOBEC3G (A3G), in particular, is a powerful restriction factor of retroviruses that blocks viral replication by hypermutating viral cDNA and inhibiting reverse transcription[1-3]. A3G is packaged into retroviral capsids through interactions with viral genomic RNA during assembly, exerting its antiviral activity inside the capsid where reverse transcription occurs during infection[8-12]. The lentiviral protein Vif inhibits A3G packaging into virions by targeting it for ubiquitin-mediated proteolysis and through ubiquitin-independent mechanisms that are poorly understood[13-17]. On a long evolutionary timescale, A3G has undergone diversifying selection to escape antagonism by Vif, whereas adaptations in Vif allow primate lentiviruses to expand their host range and adapt to host polymorphisms[18,19]. Repeated bouts of diversifying selection and adaptation are referred to as molecular arms races[20]. Adaptations in the Vif protein encoded by an SIV that infects red-capped mangabey monkeys (SIVrcm) to antagonize the hominid version of A3G enabled cross-species transmission of a lentivirus from monkeys to chimpanzees, which underlies the ancient origin of HIV-1 and the AIDS pandemic[21]. Although it is commonly assumed sites of 'molecular arms races' report on direct protein interactions, physical evidence of this interaction site to explain the mechanisms of how Vif promotes processive ubiquitination on A3G, and how mutations in Vif or A3G promote host escape and viral adaptation, remain critical and unresolved questions.

The A3 proteins are comprised of either one or two cytidine deaminase domains (CDAs), among which A3D, A3F, and A3G (containing double domains CDA1 and CDA2) and single domain A3H inhibit replication of primate lentiviruses. To target A3 family members for ubiquitination and degradation, Vif hijacks a host Cullin-RING ubiquitin ligase (CRL) complex

and a transcription cofactor core-binding factor beta (CBFβ)[16,22,23]. Co-structures of Vif with full-length A3 family members have been a major challenge for the field due to poor solubility and difficulty obtaining homogenous protein for structural studies. Accordingly, extensive effort has been devoted to the generation of variant A3 proteins that are soluble and amenable to structural studies[24-27]. For example, a 3.9 Å cryo-EM structure of a solubility optimized variant of A3F-CDA2 covalently fused to CBFβ bound to Vif α/β domain revealed an electrostatic interface required for viral infectivity[27]. However, the interaction of CDA2 of A3F with Vif and CBFβ is weak, and the significance of the tetramer comprised of A3F-CDA2, Vif α/β domain, and CBFβ protomers for A3 antagonism is unclear[27-29]; moreover, the well-characterized evolutionary adaptations in A3 proteins in response to Vif occur in the CDA1 of A3G[19,30], leaving significant gaps in our knowledge of molecular mechanisms of Vif antagonism of A3 proteins and molecular arms races between them.

Here, we solved the structure of wild-type human A3G bound to the substrate receptor module of CRL5 containing HIV-1 Vif, CBFβ, Elongin B, and Elongin C (VCBC) using single particle cryogenic electron microscopy (Extended Data Fig. 1-5 and Supplementary Table 1). 2D classification indicates that the complex is a dimer of A3G-VCBC protomers (Extended Data Fig. 1c). Our highest resolution map, generated by focused refinement around the monomer structure of A3G-VCBC, will be presented first, followed by the structures of the dimer (Methods, Fig. 1a-b, Extended Data Fig. 2, and Supplementary Discussion).

3.3 RNA acts as a molecular glue for the Vif-A3G interaction

The crystal structure of VCBC and the AlphaFold2 predicted model of human A3G could be readily fit into a 2.7 Å resolution map of the A3G-VCBC monomer (Fig. 1a-b) [31]. Unexpectedly, well-resolved density was observed for a single-stranded RNA molecule sandwiched between A3G

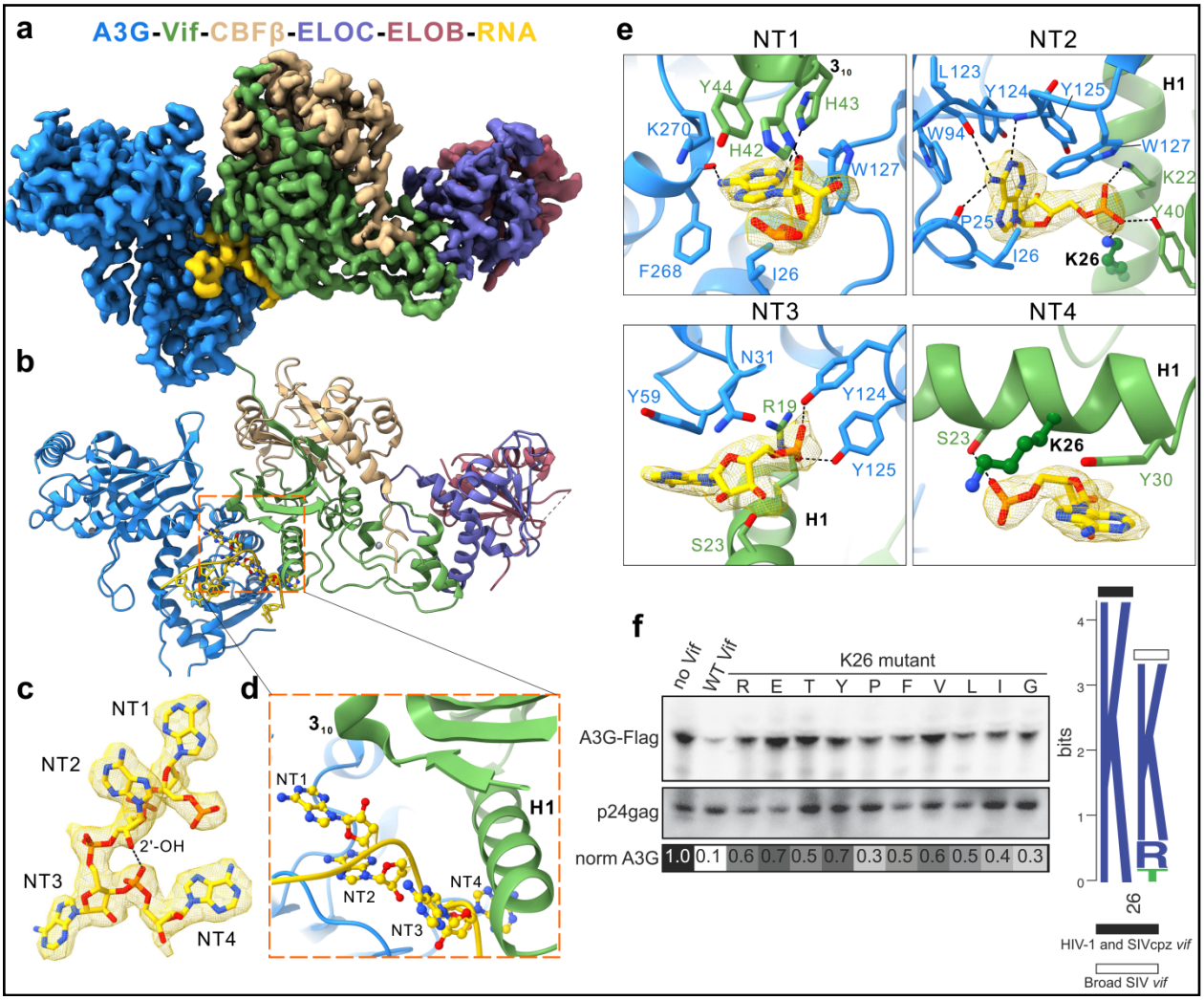
and Vif, most likely originating from the insect cells where the A3G-VCBC was co-expressed (Fig. 1a-b). Four nucleotides (named NT1-4) of single-stranded RNA are wedged deeply between A3G and Vif, with RNA forming a sharp turn mediated by a hydrogen bond between ribose 2' hydroxyl of NT2 and the phosphate backbone of NT4 (Fig. 1c-d). The majority of the interactions between A3G and Vif are mediated by CDA1, consistent with prior studies that indicate it binds RNA and is necessary and sufficient for binding to Vif[11,12,24,32-35].

RNA binding has been implicated in regulating cytoplasmic localization, self-association, and packaging of A3G into HIV-1 virus, which is essential for its antiviral activity[12,36-38]. The A3G CDA1 and CDA2 domains sandwich the RNA tetra-nucleotide at the interface formed by Vif helix 1 (residue 15-30), strand 2 (residue 39-41) and a 3_{10} helix (residues 42-46) (Fig. 1b-d). There is a division of labor between A3G and Vif for recognizing RNA. The bases of NT1-3 are bound to A3G, with aliphatic interactions, aromatic base stacking, and hydrogen bonds to NT1 and NT2 typical of sequence specific interactions (Fig. 1e). For example, the purine base of NT1 is buried in a junction formed by CDA1 (I26, W127), CDA2 (K270) of A3G and Vif (H42, H43 and Y44) and forms a hydrogen bond with main chain carbonyl of F268 on CDA2 (Fig. 1e). These interactions may explain why Vif residues that line surface of the 3_{10} helix are important for A3G degradation and viral infectivity[35,39-44]. The purine base of NT2 is buried in a hydrophobic pocket formed by A3G residues (I26, W94, Y124, Y125, W127), interacting with Y125 by T-stacking; NT2 also forms a hydrogen bond with the main chain amide of Y125 and carbonyl of P25 and L123 (Fig. 1e). Based on hydrogen bonding patterns, NT1 and NT2 are likely adenine. Sequence-specific interactions of NT1 and 2 with A3G are consistent with the enrichment of purine-rich motifs that interact with A3G in cells and virions of *vif* deficient HIV-1[45]. This result and our structure suggest A3G bound to purine-rich RNA is the substrate of the Vif E3 ligase.

In contrast to sequence specific interactions with purine NT 1 and 2 with A3G, NT3 and NT4 are stabilized by aromatic stacking interactions with Y59 of A3G and Y30 of Vif, respectively (Fig. 1e). A composite binding site for the RNA backbone is formed by Vif and A3G. Buried phosphates of NT2 and NT3 are stabilized by hydrogen bonds and salt bridges with Vif (residues Y40, K22 and K26) and A3G (Y124 and Y125), respectively (Fig. 1e). Almost all of the key contacts in the A3G-RNA-VCBC monomer interface have been mutated in prior genetic studies and result in a loss of Vif function (Supplementary Table 2 and 3), which validates their importance in our structure.

Because amino acids at positions 22, 23, 26, and 40 of Vif make exclusive interactions with RNA, we substituted them with different classes of amino acids to test the role of RNA in Vif-mediated antagonism of A3G (Fig. 1e-f). Substitution of K26 was not tolerated, supporting its role in coordinating multiple interactions with the phosphate backbone of NT2 and NT4. In contrast, K22 and S23 were tolerant of polar amino acids but refractory to nearly all aromatic and aliphatic substitutions, consistent with their role in coordinating the phosphate backbone of NT3 and NT4 (Extended Data Fig. 6). The partial loss of function of Vif substitutions at Y40 may reflect the dual role of this residue which interacts with RNA (NT2) and the 3_{10} helix of Vif. These functional results are largely reflected in the evolutionary constraints on Vif in HIV-1 and SIV sequences. For example, there is perfect conservation at residue 26 in HIV-1 and SIVcpz sequences while the only amino acids represented at position 22 are asparagine, lysine, and threonine (Fig 1f, Extended Data Fig. 6a). These positions are also enriched with polar (residue 40) or charged (residue 22, 26) amino acids in more divergent SIV Vif sequences, suggesting the binding mode for A3G, RNA and Vif is deeply conserved (Extended Data Fig. 6). These results suggest the interaction of Vif with RNA is required for the antagonism of A3G. We conclude that RNA functions as a ‘molecular

glue' to stabilize Vif-A3G interactions, much like hormones or small molecules act to recruit substrates to cellular ubiquitin E3 ligase complexes[46,47].



3-1 **Structure of the HIV-1 Vif-E3 ligase substrate receptor (VCBC) in complex with human APOBEC3G (A3G) and RNA. a,** Cryo-EM map for the A3G-RNA-VCBC monomer. **b,** Corresponding view of the refined coordinate model of A3G-VCBC complex, highlighting four nucleotide core motif (ball-and-stick) between Vif and A3G. Here and throughout, the same color code for A3G, Vif, CBFβ, ELOB, ELOC and RNA are used as indicated. **c,** Composite density map for NT1-4 of RNA with a hydrogen bond between ribose 2'-OH on NT2 and phosphate on NT4 indicated. **d,** Ribbon diagram showing NT1-4 of RNA bridging helix 1 (H1) and 3_{10} helix turn of Vif with A3G. **e,** Close-up of protein-RNA interactions between Vif and A3G for each core nucleotide of the RNA. **f,** Functional assessment of amino acid substitutions of residue K26 in HIV-1 Vif. Left: Amino acid mutants at Vif residue K26 were assessed for their ability to prevent packaging of A3G into virions. Top shows the virion incorporation of A3G and bottom shows the amount of virus ($p24^{gag}$) in the corresponding virion preparation. Below is a greyscale heatmap of the relative A3G incorporation normalized to $p24^{gag}$ based on replicate transfections (with the exception of K26Y) with the amount of A3G in the "No Vif" control set to 1.0 (darkest color). Right: Logo plot of amino acids found in the consensus of all HIV-1 clades as well as SIVcpz (black bar), and all other SIV strains using equal distribution of each SIV (white bar).

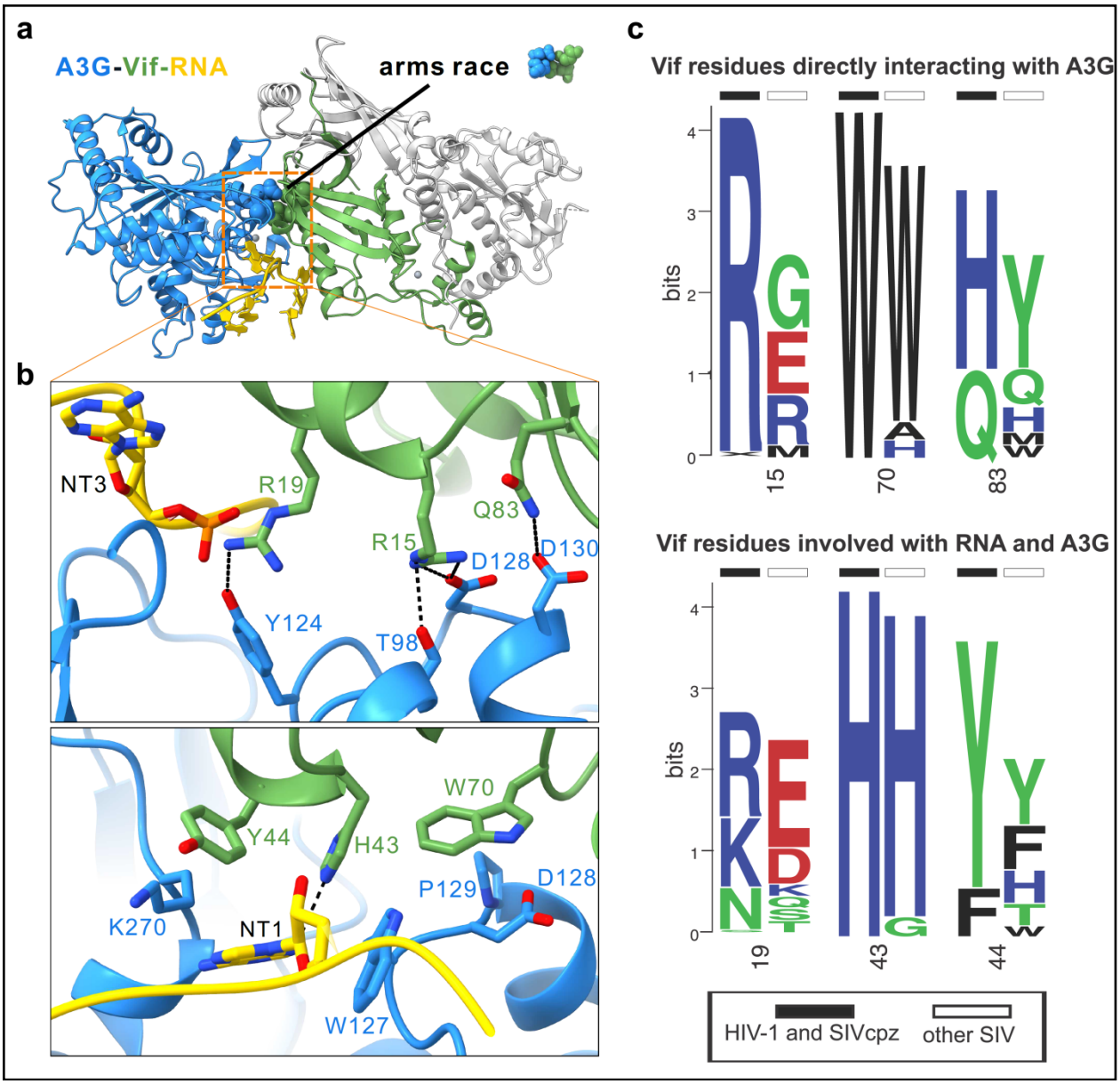
3.4 Evolution of the Vif-A3G interface and the birth of HIV-1

The identity of residues 128 and 130 in A3G have previously been shown to determine the species-specificity of the adaptation of Vif to a new host species (reviewed in [30,48]). Here, we call this interface with Vif the "arms race" interface, which has undergone diversifying selection during primate evolution[19,21,49]. The arms race interface is comprised exclusively of protein interactions between A3G and Vif and is adjacent to the RNA interface (Fig. 2a). Residues D128 and D130 of A3G are buried deep in the arms race interface forming a network of hydrogen bonds with R15 and Q83 of Vif, respectively (Fig. 2b). Q83 of Vif was previously shown to be essential for the Vif adaptation that allowed cross-species transmission from SIVrcm to chimpanzee[50] (Fig. 2b). Further, W70 of Vif interacts closely with W127, D128 and P129 through hydrophobic interactions. This intimate network of contacts explains why lysine substitution at position 128 of A3G, observed in Old World monkeys (OWM), is unable to be counteracted by SIVcpz and HIV-1 Vif; it does not contain a hydrogen bond acceptor nor the charge or shape complementarity to interact with R15 of Vif[49,51] (Fig. 2b). R15 and W70 are conserved in all HIV-1 and SIVcpz Vif sequences, consistent with them functioning as lynchpins of the HIV-1 Vif interaction with human A3G (Fig. 2c). However, as predicted from an evolutionarily dynamic interface, the Vif sequences from OWM SIV that must evolve to antagonize divergent host A3G residues in the arms race interface are themselves variable (Fig. 2c and Extended Data Fig. 7a)

Because the Vif gene that gave rise to SIVcpz, and ultimately HIV-1, is derived from SIVrcm, we asked how the interaction between rcmA3G and Vif at the arms race interface may have evolved[21]. Comparative modelling of the rcmA3G-Vif complex based on our structure indicates K128 of A3G directly interacts with Y86 of SIVrcm Vif (equivalent to residue 83 in HIV-1 Vif), suggesting adaptation of Vif to the positively selected residue 128 entailed a

remodeling of interactions that occur on the arms race interface (Extended Data Fig. 8). We suggest structural plasticity in Vif enabled amino acid substitutions, such as those occurring at position 86, to neutralize A3G and enable cross-species transmission of SIV from red-capped mangabeys to chimpanzees.

In contrast to the arms race interface, the interface between A3G and Vif that is bridged by RNA is well conserved because residues in the purine binding pocket of A3G that contact RNA, such as the L7 loop (Y124 to W127), are required for restriction in the absence of Vif (NT2, Fig. 1e, Extended Data Fig. 7b, and Supplementary Table 3) [12]. We conclude Vif binds A3G/RNA in a manner that limits A3G escape over long evolutionary timescales by engaging an essential surface required for antiviral function, explaining why genetic signatures of diversifying selection and adaptation are constrained to the direct protein interactions observed at the molecular arms race interface.

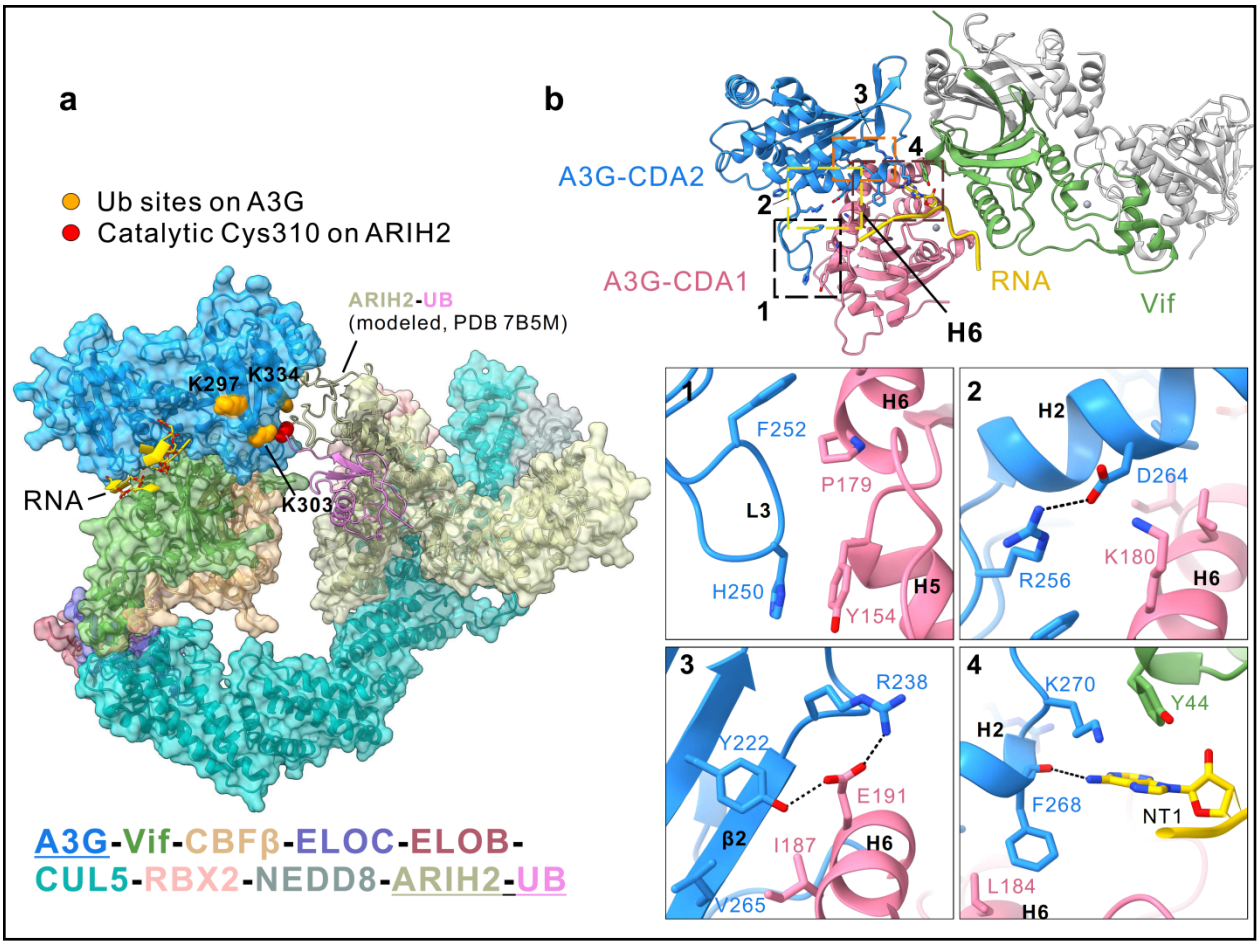


3-2 **Interplay between the molecular arms race and RNA interfaces of Vif-A3G. a,** Ribbon diagram showing position of molecular arms race interface (spheres) relative to the RNA interface (sticks). **b,** Close-up of molecular arms race interface (top) and residues that contribute to Vif-A3G binding and in contact with RNA (bottom). Residues D128 and D130 of A3G are under diversifying selection; residue Q83 is an adaptation that allowed SIVrcm Vif to neutralize hominid primate A3G to enable cross-species transmission. **c,** Logo plots of natural sequence variation in Vif residues that line the molecular arms race (top) and Vif-RNA-A3G interface (bottom). HIV-1 and SIVcpz sequences (black bars) are the consensus of all HIV-1 clades as well as SIVcpz, and SIV sequences (white bars) are all other SIV strains using equal distribution of each SIV.

3.5 Vif binds the A3G/RNA complex in a conformation compatible with ubiquitination

Cullin-RING E3 ligases conjugate ubiquitin onto substrates by orienting acceptor lysines into an 'ubiquitination zone' that is accessible by coenzymes[52]. To determine if the A3G-RNA-VCBC module is compatible with ubiquitination by CRL5, we employed comparative modeling. Lysine residues of A3G that are required for Vif-mediated ubiquitination and subsequent degradation are located within CDA2 of A3G, which is oriented towards the ARIH2 coenzyme of CRL5 that installs the first ubiquitin allowing extension of K48-linked ubiquitin chains by a ubiquitin conjugating enzyme[53-55] (Fig. 3a). The orientation of CDA2 is determined by intramolecular interactions in A3G and intermolecular interactions with Vif and RNA. Within A3G, Helix 6 (residues 178-193) of CDA1 forms interactions with CDA2 that fix domain orientations (Fig. 3b) through a salt-bridge (K180 with D264) and a series of hydrogen bonds (E191 with Y222; E191 with R238). Aromatic interactions between helix 5 of CDA1 (Y154) and the L3 loop in CDA2 (H250) as well as several hydrophobic contacts between both domains stabilize CDA domain orientations (Fig. 3b). Interactions between A3G, Vif, and RNA may fix the orientation of CDA2 in the three-way interface between K270, Y44 and NT1, respectively, suggesting RNA not only acts as a molecular glue, it may also orient A3G CDA2 acceptor lysines for ubiquitin transfer (Fig. 3b).

Most of the intramolecular interactions in A3G are consistent with the AlphaFold2 structure of human monomeric A3G but not prior crystal structures of A3G containing solubility enhancing mutations or rhesus macaque A3G, which exist as a monomer or self-associated forms[25,26]. Notably, in the crystal structure of rhesus macaque A3G, Helix 6, which was proposed to promote self-association, is buried when human A3G is bound by VCBC[26] (Fig. 3b). This observation indicates VCBC binds A3G in a manner that inhibits its self-association.



3-3 **Vif orients acceptor lysine residues on CDA2 of A3G for ubiquitin transfer**. **a,** Comparative model of A3G/RNA in complex with Vif CRL5 E3 ligase bound to coenzyme ARIH2 that transfers the first ubiquitin to CDA2 of A3G. Lysine residues identified as the A3G ubiquitination sites by mass-spectrometry and required for Vif-mediated degradation of A3G are colored in orange[53-55]. The catalytic Cys310 of ARIH2 is colored in red. **b,** Overview of interactions that stabilize the relative orientation of CDA domains in A3G. Helix 6 (H6) previously shown to be important for A3G dimerization is labeled[26]. Bottom panels show close-up of interactions within A3G CDA domains, and between A3G, RNA and Vif.

3.6 Conclusions and Implications

Vif and A3G are a paradigmatic example of a host-pathogen molecular arms race. DNA sequence analyses of primate genomes and functional studies reveal two positions in A3G that undergo diversifying selection allowing escape from antagonists such as Vif[18,19]. Yet how Vif binds A3G tightly enough to antagonize A3G was unclear. We discovered that RNA acts like a molecular glue to promote a high affinity interaction, as it increases the buried surface area of the Vif-A3G

complex and is required for viral infectivity. Our structural studies reveal a small surface of protein-protein interactions between Vif and A3G that determines cross-species transmissions of primate lentiviruses, as well as the viral adaptations in Vif underlying the origin of HIV-1. Though the buried surface area of the arms race interface is two times smaller than the RNA interface, it acts as a hot-spot controlling the fate of viral infection (Extended Data Fig.7c).

Prior biochemical and structural studies indicate A3F-CDA2 makes transient interactions with Vif and CBFβ[27-29]. Our work on wild-type A3G suggests both CDA domains and RNA make stable interactions with Vif without contacting CBFβ. In the former structural study, covalent fusion of A3F-CDA2 was employed to increase occupancy of A3F bound to Vif whereas in our study, RNA achieves this role by acting as a molecular glue. It is well established that different A3 family members engage surface exposed residues of Vif that are genetically separable (reviewed in [4]). We suggest these surfaces may be bridged by cellular cofactors as described for A3G. An alternative, but not mutually exclusive possibility is that interactions with Vif are stabilized by bipartite interactions with tandem CDA domains of A3 proteins. Structural studies of Vif-A3 complexes purified after coexpression or native purification from eukaryotic cells will allow this question to be addressed in future studies.

A new model for Vif antagonism of A3G is built on prior functional studies and two key observations from our structure. First, lentiviral genomes are enriched in purines and cross-linking immunoprecipitation sequencing studies of cells infected with *vif*-deficient HIV-1 indicate A3G preferentially binds to purine-rich sequences present in noncoding RNA, mRNA and viral genomic RNA[45,56]. Our structure reveals A3G binds to a purine-rich tetra-nucleotide motif using residues (Y124 to W127) that are essential for viral packaging in the absence of Vif[11,12] (Fig. 1c, e). We

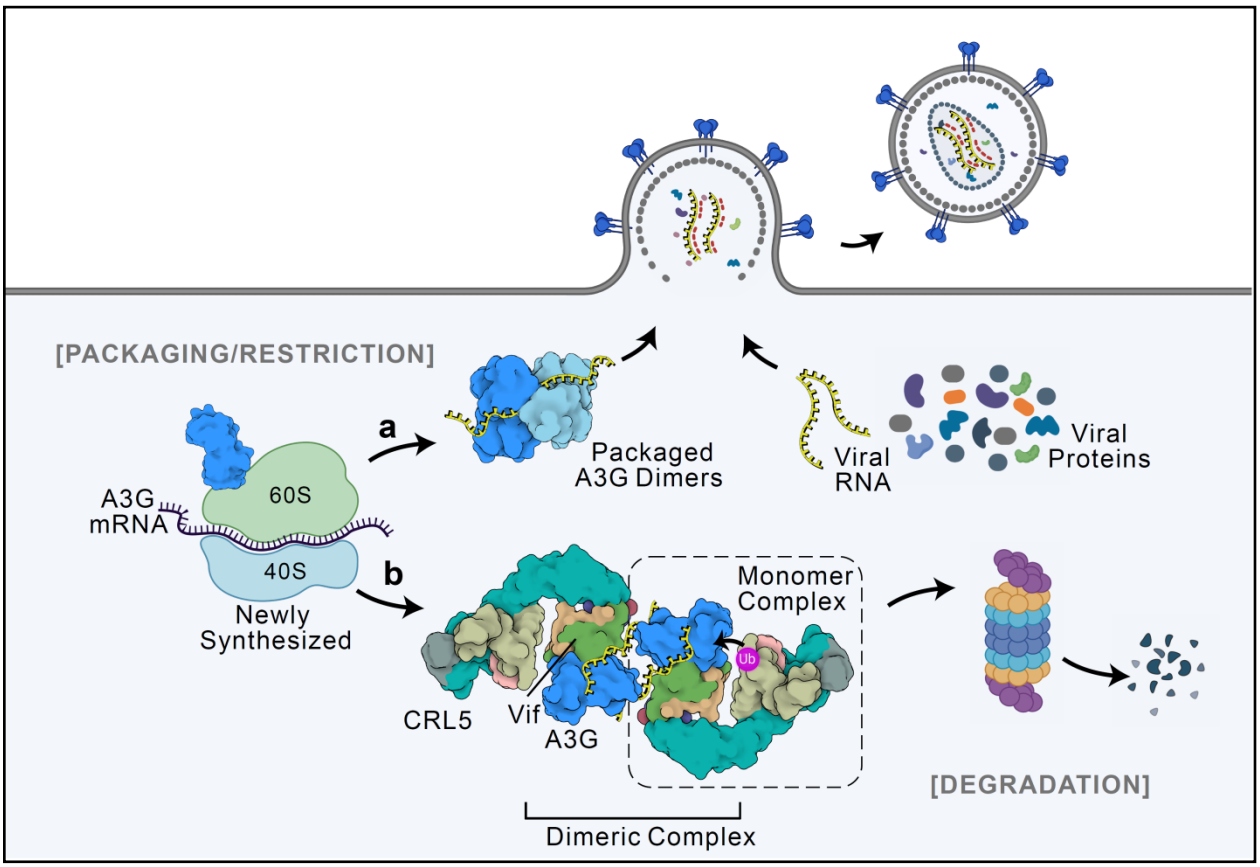
propose the substrate of the Vif E3 ligase is not A3G but instead a complex of A3G bound to purine-rich RNA, including purine-rich sequences found in the viral genome.

Second, in addition to RNA binding, A3G self-association is required for its packaging into virions[11]. Our structure indicates Vif binding to A3G is mutually exclusive with its ability to self-associate. While A3G-VCBC forms dimers, within each dimeric assembly A3G forms little or no self-association (Extended Data Fig. 9-10). This finding suggests Vif binding to A3G has the capacity to block its packaging independent of ubiquitination activity, a mechanism that may potentiate repression of restriction.

In the absence of Vif, A3G self-associates onto viral genomic RNA and is packaged into viral particles for restriction (Fig. 4). We suggest Vif antagonizes A3G early in its biosynthesis while it is a monomer in a specific complex with viral genomic RNA en route to viral packaging. Such an early intervention would ensure Vif counteracts A3G in its most dangerous form for the virus, disrupting encapsidation and promoting polyubiquitination while bound to genomic RNA. Nucleotides that are 3' to the primary Vif-A3G interaction site may template an additional copy of the Vif-E3 ligase to cooperatively reinforce Vif-A3G interactions and ubiquitination (Fig. 4 and Extended Data Fig. 11). This model is consistent with observations that newly synthesized low molecular mass forms of A3G are packaged and most sensitive to Vif-mediated degradation, explaining ubiquitin-dependent and -independent functions of Vif and how plasticity of molecular arms races can be enabled by a third party such as RNA.

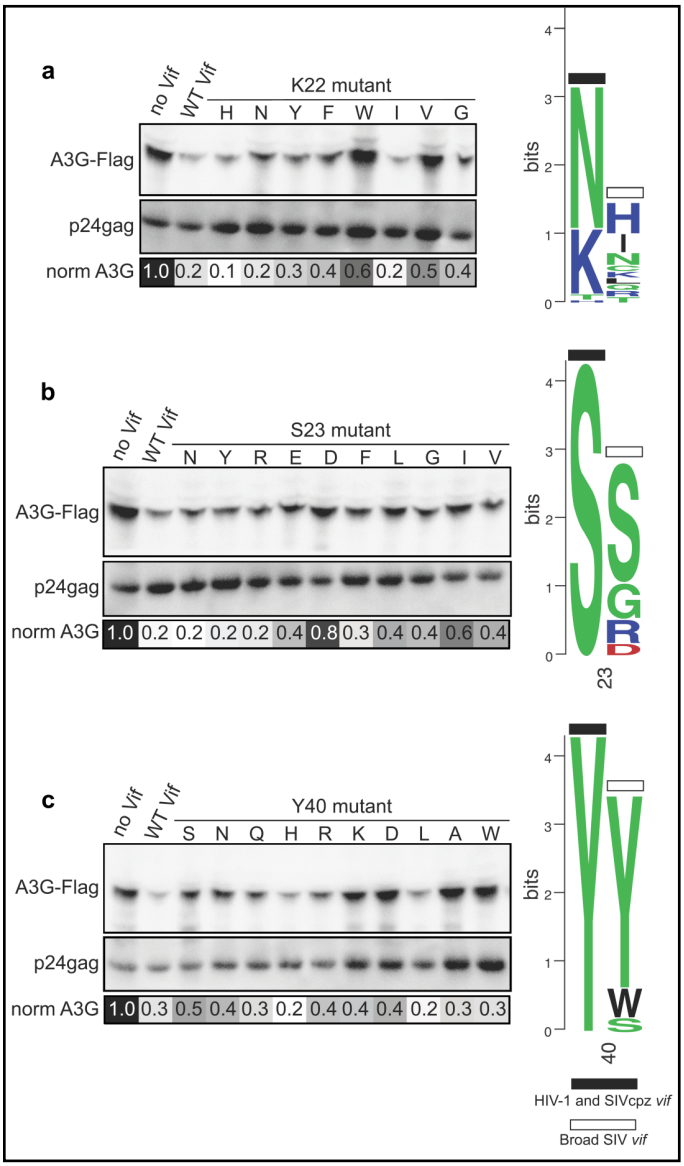
In summary, A3G binding to Vif is not restricted to the evolutionary dynamic interface subject to diversifying selection and adaptation, contrary to popular models of molecular arms races, but rather also includes a conserved interface through RNA binding that helps position key

residues necessary for viral antagonism of a host antiviral gene - a principle that may be adopted by other innate immune proteins and pathogen encoded antagonists[57,58].

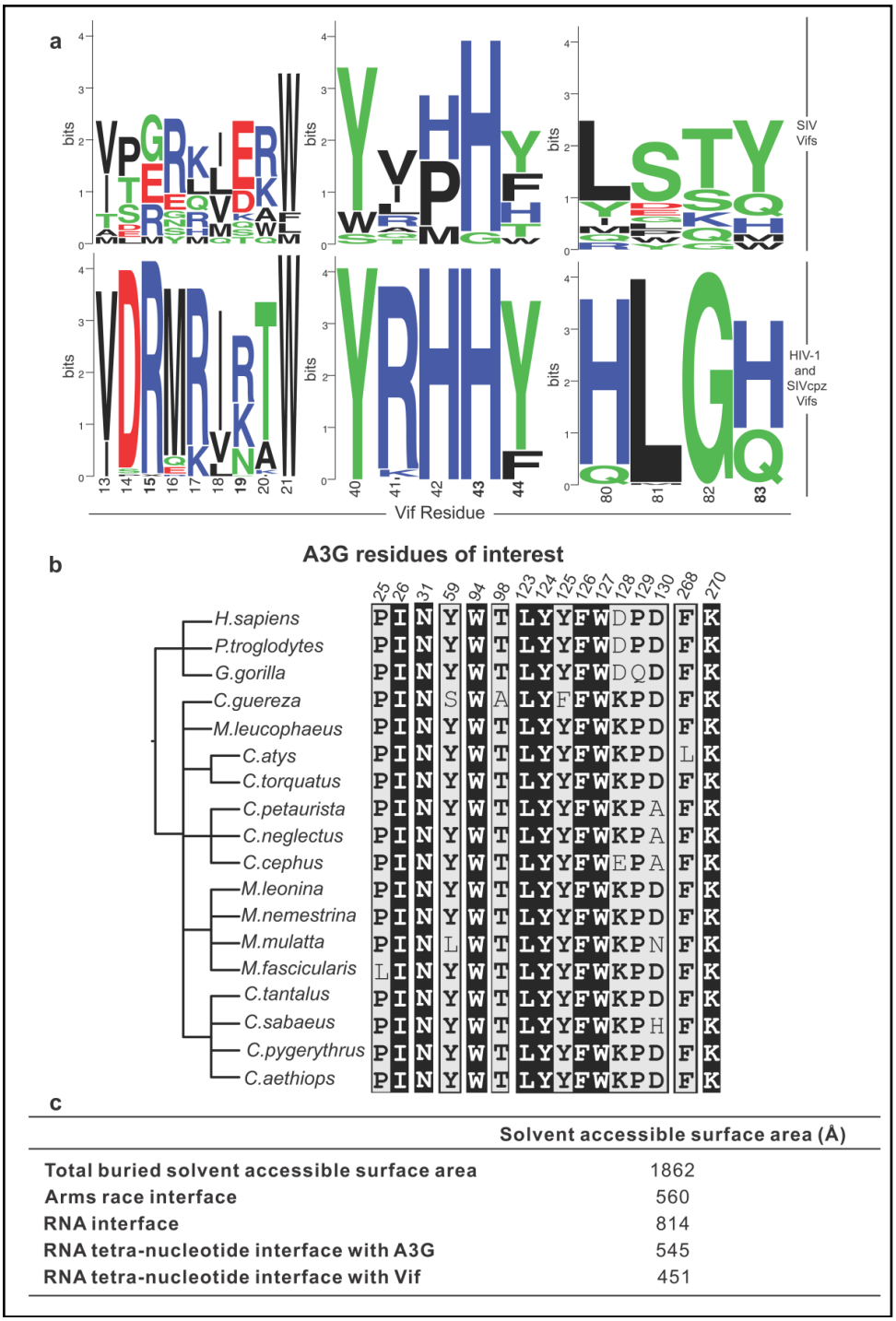


3-4 **Schematic model of A3G inhibition by HIV-1 Vif. a**, Packaging of A3G into HIV-1 virus requires A3G dimerization and its interaction with viral RNA. **b**, Vif neutralizes A3G early during its biosynthesis, by binding RNA-bound A3G, inhibiting A3G dimerization, and promoting ubiquitin-mediated proteolysis.

3.7 Selected Extended Figures



3-5 **Extended Data Fig. 6 | Functional and evolutionary assessment of Vif residues involved in RNA binding.** Left: Amino acid mutants at Vif residue (**a**) K22 (**b**) S23, and (**c**) Y40 were assessed for their ability to prevent packaging of A3G into virions. Top Western blot in each panel shows the virion incorporation of A3G, while the bottom Western blot in each panel shows the amount of virus ($p24^{gag}$) in the corresponding virion preparation. Below each panel is a greyscale heatmap of the relative A3G incorporation normalized to $p24^{gag}$ based on replicate transfections (with the exception of S23Y and S23D) with the amount of A3G in the "No Vif" control set to 1.0 (darkest color). Right of each panel: Logo plot of amino acids found in the consensus of all HIV-1 clades as well as SIVcpz (black bar), and all other SIV strains using equal distribution of each SIV (white bar)



c

	Solvent accessible surface area (Å)
Total buried solvent accessible surface area	1862
Arms race interface	560
RNA interface	814
RNA tetra-nucleotide interface with A3G	545
RNA tetra-nucleotide interface with Vif	451

3-6 **Extended Data Fig. 7 | Characterization of arms race interface. a,** Logo plots showing sequence variation of Vif residue that interacts with A3G for HIV-1 and SIVcpz (bottom) and other SIVs (top). Data is similar to that of Figure 2c except that neighboring residues are also shown for context. **b,** Sequence alignment of A3G residues that contact RNA or Vif from Old World Monkeys and hominids. Fully conserved residues are highlighted with white text on black background. **c,** Buried solvent accessible surface area for A3G-RNA-VCBC monomer structure.

3.8 Methods

Protein expression and purification

Full-length genes of HIV-1_{HXB2} Vif, CBFβ, ELOB, ELOC, human A3G with C-terminal Strep-tag, and Cullin 5 with C-terminus truncated (1-386; abbreviated as CUL5N) were cloned into a single MacroBac vector 11A (Addgene: 48294) using restriction-ligation method as described[59]. A recombinant baculovirus encoding all six proteins was generated using the Bac-to-Bac baculovirus expression system (Thermo Fisher Scientific) [60]. Suspension Sf9 insect cells were maintained in SF900^{TM} III SFM medium (Gibco) in a shaker (Innova 4430 incubator shaker) at 120 rpm at 27°C. Serially diluted recombinant baculoviruses were added to 25 ml of Sf9 cells (2 x 10^6 cells/ml) grown in polycarbonate Erlenmeyer flask (Corning), and protein expression was evaluated by SDS-PAGE and western blotting analysis to optimize the ratios of virus volume to cell volume. One liter of Sf9 insect cells (2 x 10^6 cells/ml) were infected with viruses at a ratio determined from small-scale titration experiments and cultured for 48 h before harvesting by centrifugation (1,500 x *g*, 10 min). Cell pellets were washed with PBS and resuspended in 5 times the pellet volume of lysis buffer (50 mM HEPES, 50 mM NaCl, 5% glycerol, 1% Triton X-100, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), mini cOmplete protease inhibitor cocktail (Sigma-Aldrich), 25 μg/ml DNase I (Sigma-Aldrich) and 50 μg/ml RNase A (Sigma-Aldrich), pH 8.0) and lysed by Dounce homogenizer. All purification steps were performed at 4°C. Cell lysates were clarified by centrifugation at 17,000x *g* (F15-8x50cy rotor) for 2 h. The supernatants were filtered (0.45 μm) and loaded onto a 5 ml StrepTrap HP column (GE Healthcare) pre-equilibrated in binding buffer (50 mM HEPES, 150 mM NaCl, 5 % glycerol, 1 mM TCEP, pH 8.0). The column was washed with 15 column volumes (CV) of wash buffer (50 mM HEPES, 1.5 M NaCl, 5 % glycerol, 1 mM TCEP, pH 8.0), followed by 5 CV of binding buffer. The protein was eluted in 6 CV binding buffer supplemented with 5 mM D-desthiobiotin

(Sigma-Aldrich). The eluate was dialyzed overnight against 1 liter of dialysis buffer (50 mM HEPES, 75 mM NaCl, 5 % glycerol, 1 mM TCEP, pH 7.0). The sample was applied onto a 5 ml HiTrap Heparin column (GE Healthcare) pre-equilibrated with dialysis buffer, and eluted over 0-100% linear gradient of elution buffer (50 mM HEPES, 1M NaCl, 5 % glycerol, 1 mM TCEP, pH 7). Fractions containing all six proteins were pooled, dialyzed against 1 liter of running buffer (30 mM HEPES, 150 mM NaCl, 5% glycerol, 1 mM TCEP) for at least 4 h, loaded onto a Superose 6 increase 10/300 GL column (GE Healthcare) pre-equilibrated with running buffer, and 0.3 ml fractions were collected. Size-exclusion chromatography indicates the particle is a dimer in solution consisting of 2 copies of A3G-VCBC-CUL5N (~330 kDa) (Extended Data Fig. 1a). The peak fraction was directly used for cryo-EM without further concentration (Extended Data Fig. 1a).

Cryo-EM sample preparation and data acquisition

3.5 μL purified complex (1.9 μM) was applied onto glow-discharged UltraAuFoil 300 mesh R1.2/1.3 grids (Electron Microscopy Science), incubated for 15 s at 23 °C and 100% humidity, blotted with a blot force of 0 for 12 s, and then plunge-vitrified into liquid ethane using FEI Vitrobot Mark IV (Thermo Fisher). A total of 6429 super-resolution movies were collected at a nominal magnification of 105,000x on FEI Titan Krios microscope (Thermo Fisher) equipped with a K3 direct electron detector and BioQuantum energy filter (Gatan) set to a slit width of 20 eV. The collection was performed semi-automatically using SerialEM at a dose rate of 8.0 e^- $pixel^{-1}$ s^{-1} for a total dose of 68 $e^-/Å^2$ over 118-frames[61]. Dose-fractionated image stacks were motion-corrected, dose-weighted, 2x binned to the physical pixel size of 0.835 Å by MotionCor2 in SCIPION package[62,63]. A defocus range of -0.8 μm to -2.0 μm was applied.

Image processing and 3D reconstruction

Initial processing of the resulting summed micrographs was performed in cryoSPARC v3.0[64] (Extended Data Figs. 1, 2a and Supplementary Table 1). The contrast transfer function (CTF) of dose-weighted motion-corrected micrographs was estimated by Patch CTF. Micrographs with CTF fit resolution poorer than 4 Å and excessive ice contamination were removed, resulting in final 6221 micrographs. Selected micrographs were split into two half datasets to speed up data processing. Approximately 2.3 million particles were picked using cryoSPARC circular blob with the minimum and maximum particle diameter of 150 and 200 Å respectively and minimum separation distance between particles of 108 Å, extracted, and 4x binned (3.34 Å /pixel). After 2D classification of extracted particles, class averages without proteinaceous features were discarded. The remaining particles were subjected to two rounds of 2D classification, resulting in classes with clear structural features used to generate a good initial model (Extended Data Fig. 1c). Particles saved from the first round of 2D classification underwent iterative rounds of *ab initio* reconstruction and heterogeneous refinement using 3 reference maps (1 good and 2 junk) (Extended Data Fig. 2a). Unbinned particles were re-extracted from the best reconstruction and refined with non-uniform refinement. The resulting 3D map (designated as 'consensus reconstruction') features a well-resolved top body (Extended Data Fig. 2a, black dashed box) and a bottom body of relatively poor resolution (Extended Data Fig. 2a, red dashed box).

The top body of the consensus reconstruction map displays visible helical features, in which we were able to fit the crystal structure of HIV-1 VCBC (PDB: 4N9F) and the AlphaFold 2 predicted human A3G monomeric structure (AF2: Q9HC16) [31,65] (Extended Data Fig. 2a, dashed black box). To improve the local density of the top body, particle subtraction and focused refinement were applied[66]. Using Chimera[67] and RELION[68], a mask was applied to subtract the

signal of the bottom body from particle images. These signal-subtracted particles were then re-imported into cryoSPARC and subjected to local refinement using a soft mask around the top body. The resulting focused refined map is termed 'monomer' density map as it accommodates the A3G-VCBC monomer structure well, with a nominal resolution of 2.7 Å (Extended Data Fig. 2b, green box). The local resolution of the density map is variable (Fig. 1a-b, Extended Data Fig. 3). A3G, Vif and CBFβ have well-resolved side chain density, allowing reliable model building and refinement (Extended Data Fig. 4). The resolution of ELOC and ELOB is sufficient for backbone tracing but most density for sidechains was absent. CUL5N was present in the preparation but not in our final maps presumably due to the dynamic features of VCBC-CUL5N or dissociation during freezing[15,29].

To address the conformational heterogeneity of the bottom body (Extended Data Fig. 2a, dashed red box), the 495,571 particles from this consensus refinement were subjected to 3D variability analysis (3DVA) in cryoSPARC[69] (Extended Data Fig. 2b). Particles were re-classified into six clusters using three principal components with a soft mask enclosing the bottom body and the filter resolution set to 8 Å, followed by non-uniform refinement in cryoSPARC. Four out of six selected classes were then imported into RELION and 3D classification was performed without alignment using a T value of 4 to sort out remaining low-quality particles. Classes that displayed strong density for the bottom body were further processed with 3D auto-refine in RELION. The reconstructions showed improved densities in the bottom region, where three distinct conformational states could be identified, at a nominal resolution of 3.3 Å (State 1, 57207 particles), 3.5 Å (State 1', 51055 particles), and 3.46 Å (State 2, 48310 particles), respectively. The two highest quality maps (State 1 and State 2) were refined further, and the resulting maps allowed fitting of an additional copy of A3G-VCBC complex (Extended Data Fig. 2b, 9).

Local refinement of State 1 without the top body was performed in cryoSPARC afterward, leading to a reconstruction with more complete density in this region. The State 2 reconstruction was subjected to another round of non-uniform refinement in cryoSPARC to improve anisotropy and reached a nominal resolution of 3.16 Å (Extended Data Fig. 2b). The discrete conformational states of A3G-VCBC dimer observed with 3DVA were also validated by focused classification in RELION[66] (Extended Data Fig. 5). The density of top body was subtracted from the particle images of the consensus reconstruction (Extended Data Fig. 5a, red dashed box highlights the density to keep). These signal-subtracted particle images were classified into six classes without alignment, using a soft mask focused on the bottom body (Extended Data Fig. 5b). Following refinement by 3D auto-refine, four out of six selected classes were subjected to 3D classification skipping alignment and 3D auto-refine in RELION. The resulting 3D classes are consistent with those generated with 3D variability analysis in cryoSPARC, differing only in the density levels of the bottom copy. We used the reconstruction maps generated from 3D variability analysis to build the model owing to their stronger overall density. Postprocessing of the final reconstruction was performed in RELION for estimating the global resolution using the FSC cutoff=0.143[70]. The maps were sharpened with DeepEMhancer[71] and improved by density modification without a model applied in PHENIX[72] for map interpretation and model building. Local resolution estimation was done by ResMap[73]. Directional resolution was assessed using the 3DFSC server[74]. Format conversion between software was carried out by PyEM[75].

Model building and refinement

A comparative model of HIV-1 Vif_{HXB2}-CBFβ was built with MODELLER[76,77] using the X-ray structure of HIV-1 Vif_{NL4-3}CBC-$CUL5_{NTD}$ (PDB code 4N9F) as a template[31]. The atomic model for hA3G-V_{HXB2}CBC was generated by fitting separate models of human A3G (AF2 code

Q9HC16), the aforementioned comparative model of Vif_{HXB2}-CBFβ, and ELOB/C from the SIVrcmCBC structure (PDB code 6P59) into the monomer density map using UCSF Chimera[50,65,67]. This starting model was manually rebuilt in Coot[78,79] and adjusted in ISOLDE[80] to improve local fitting. The model was then real-space refined in PHENIX[81,82]. The refined structure obtained from the monomer density map was used as a template for model building of two copies of A3G-VCBC into the EM maps for State 1, State 1', and State 2. The same model building and refinement procedure were performed for State 1 and State 2 because of their higher overall resolution. Most residues buried at the A3G-Vif and A3G-A3G interface in the monomer and dimers showed clear density for the side-chains, except Vif residues 117-154 in the bottom copy of State 2 (Extended Data Fig. 4a-b). The weak density in this region precluded precise atomic modeling, and thus A3G-Vif dimeric interface for State 2 is only interpretable on one side.

After model building the A3G-VCBC proteins, we observed unaccounted for density sandwiched between A3G and Vif in the EM maps for State 1 and 2 dimers which was annotated as oligonucleotides by Haruspex, a convolutional neural network trained to detect DNA/RNA versus protein in density maps[83]. The co-purified RNA likely originated from the insect cells where the A3G-VCBC was co-expressed. The continuous phosphate backbone density could be traced and well fit with single-stranded RNA. The densities of co-purified RNA, especially those sandwiched between A3G and Vif, are clearly resolved, showing features that allow purines and pyrimidines to be distinguishable (Fig.1a-c, Extended Data Fig. 4c). Guided by the EM density of RNA, a dummy sequence containing either adenine or uridine was manually built into the EM map in Coot. RNA geometry was improved by ERRASER[84]. The complete model including RNA was assessed using MolProbity[85] and optimized after iterative refinement in Coot and PHENIX. Model-map fit was evaluated by correlation coefficient in PHENIX and Q-score analysis[86,87]. Protein-

RNA interactions were detected by BINANA[88] and protein-only interactions were analyzed by PLIP[89] and Ligplot[90] with default settings except that the hydrogen bond distance cutoff was set to 3.5 Å. The orientation and displacement between State 1 and 2 structures were determined using PyMOL[91]. The morph movie was generated in UCSF Chimera X. Figures were created using Chimera[67], Chimera X[92], and BioRender.com. Model statistics are summarized in Supplementary Table 1.

Comparative Modeling of rcmA3G-Vif-CBFβ complex

Since the atomic structure of the rcmA3G-Vif complex is not available, we built a comparative model of rcmA3G bound to Vif-CBFβ using MODELLER v10.1[76,77]. The sequence identity between the template and model are 79%, 40%, and 100% for A3G, Vif, and CBFβ, respectively. A3G residues required for RNA binding are sequence-conserved among hominids and Old World monkeys (Extended Data Fig. 6b). The coordinates of RNA and Zn^{+2} ions were transferred from the templates into the generated model. After computing ~600 models, we used hierarchical clustering and DOPE scoring to obtain the top scoring cluster[93]. The precision of this cluster is 1.4 Å; model precision is defined as the variability among the structural models. The best scoring model was used for further analysis and Extended Data Fig. 8.

Comparative Modeling of A3G-Vif-CBFβ-CRL5-NEDD8-ARIH2-Ub

Using MODELLER we computed a comparative model of the full A3G-Vif-CBFβ in complex with neddylated CRL5 and coenzyme ARIH2. Templates included the A3G-Vif-CBFβ-ELOB/C structure presented here, the Vif-CBFβ-$CUL5_{NTD}$-ELOB/C pentameric complex (PDB code 4N9F) [31], the neddylated $CUL5_{CTD}$-RBX2-ARIH2 tetrameric complex (PDB code 7ONI) [94], and a partial structure containing CUL1-RBX1-Ub-ARIH1 (PDB code 7B5M) [52]. We computed

~100 models, which were clustered and evaluated using DOPE scoring[93]. The precision of the top scoring cluster is 1.6 Å. The best scoring model was used for further analysis and Figure 3a.

Assaying *vif* mutants for A3G degradation

We generated a library of variants at positions 22, 23, 26, and 40 using degenerate oligonucleotide mutagenesis in the HIV-1 LAI *vif* gene. Individual colonies were sequenced and ligated into a lentiviral vector. Human A3G flanked by a C-terminal 3XFLAG epitope tag in the pcDNA4/TO vector backbone (Thermo Fisher, #V102020) was transfected into HEK293T cells plated in 6-well dishes at a density of 1.5×10^5 cells/ml. The amount of A3G packaged into virions was assayed by cotransfecting 1000 ng of the Vif vector, 200 ng A3G-3XFlag, and 500 ng of psPAX2 for gag/pol production with TransIT-LT1 transfection reagent (Mirus; catalog no. MIR2304) at a reagent:plasmid DNA ratio of 3:1. Two days post transfection, 1 ml of the supernatant was filtered through a 0.2 μm syringe filter and virions were pelleted in an Eppindorf 5415R tabletop microcentrifuge for 1 h at 4°C at maximum speed. The supernatant was aspirated off, and 25 μl of NuPAGE 4× loading dye (Invitrogen; catalog no. NP0007) was added to each sample. Samples were boiled for 10 min at 95°C and loaded on an SDS-PAGE gel. Anti-FLAG (Sigma; catalog no. F3164) and anti-$p24^{gag}$ (NIH-ARP; catalog no. 3537) antibodies were used for Western blotting at a dilution of 1:5,000. Mouse IgG HRP-conjugated antibody (R&D systems; catalog no. HAF007) was used to detect primary antibodies at a dilution of 1:5,000. The chemiluminescent signals from all Western blots were imaged using a ChemiDocMP imaging system (Bio-Rad), and images were processed using ImageJ software to quantify the densitometry for each detected antibody band. Normalized A3G in virions was calculated by dividing the amount of A3G by the amount of p24gag and setting that number to 1.0 for the “No Vif” control.

Natural sequence variation analysis

To analyze natural variation in Vif at interaction sites with RNA and A3G, a curated subtype reference alignment for HIV-1 and SIVcpz sequences was downloaded from the Los Alamos HIV database (https://www.hiv.lanl.gov/content/index). This alignment includes 4 representatives from each HIV-1 group M subtype and 4 representatives from groups N, O, and P, as well as 21 SIVcpz sequences from each of the *Pan troglodytes* subspecies (*troglodytes* and *schweinfurthii)* sampled from primary isolates comprehensively encompassing the geographic range. Consensus sequences of SIVasc, SIVdeb, SIVdrl, SIVlst, SIVgsn, SIVmac, SIVmus, SIVrcm, SIVsmm, SIVgrv, SIVver, SIVtan, SIVsun, SIVsab, and SIVgor were generated from all available sequences in each respective SIV. These sequences were aligned using Clustal Omega, and logo plots were generated from these alignments using WebLogo[95].

To generate A3G sequence alignments, sequences were downloaded from NCBI (accession numbers: AGI04219.1, AAP85255.1, Q694C1.1, AGE34499.1, NP_001332845.1, XP_011887342.1, AGE34492.1, AGE34504.1, AGE34486.1, AGE34487.1, ANY26448.1, XP_011710628.1, AGX93019.1, XP_011710628.1, AEY75956.1, NP_001279005.2, AEY75955.1, Q7YR25.1). Sequences were aligned using Clustal Omega, and visualized using ESPript 3[96].

Data Availability

Cryo-EM maps and maps focused on specific regions used to guide model building are deposited in the Electron Microscopy Data Bank (EMDB) with accession codes EMD-27032 (A3G-RNA-VCBC monomer), EMD-27033 (A3G-RNA-VCBC dimer for State 1), EMD-27034 (A3G-RNA-VCBC dimer for State 2), and EMD-28667 (A3G-RNA-VCBC dimer for State 1'). The associated coordinate files are deposited in the Protein Data Bank with the accession code

8CX0 (A3G-RNA-VCBC monomer), 8CX1 (A3G-RNA-VCBC for State 1), and 8CX2 (A3G-RNA-VCBC for State 2).

3.9 Method References

15 Binning, J. M. *et al.* Fab-based inhibitors reveal ubiquitin independent functions for HIV Vif neutralization of APOBEC3 restriction factors. *PLoS Pathog* **14**, e1006830, doi:10.1371/journal.ppat.1006830 (2018).

29 Ball, K. A. *et al.* Conformational Dynamics of the HIV-Vif Protein Complex. *Biophys J* **116**, 1432-1445, doi:10.1016/j.bpj.2019.03.014 (2019).

31 Guo, Y. *et al.* Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature* **505**, 229-233, doi:10.1038/nature12884 (2014).

50 Binning, J. M., Chesarino, N. M., Emerman, M. & Gross, J. D. Structural Basis for a Species-Specific Determinant of an SIV Vif Protein toward Hominid APOBEC3G Antagonism. *Cell Host Microbe* **26**, 739-747 e734, doi:10.1016/j.chom.2019.10.014 (2019).

52 Horn-Ghetko, D. *et al.* Ubiquitin ligation to F-box protein targets by SCF-RBR E3-E3 super-assembly. *Nature* **590**, 671-676, doi:10.1038/s41586-021-03197-9 (2021).

59 Gradia, S. D. *et al.* MacroBac: New Technologies for Robust and Efficient Large-Scale Production of Recombinant Multiprotein Complexes. *Methods Enzymol* **592**, 1-26, doi:10.1016/bs.mie.2017.03.008 (2017).

60 Life_Technologies. Bac-to-Bac™ Baculovirus Expression System. (2015).

61 Mastronarde, D. N. SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microscopy and Microanalysis* **9**, 1182-1183, doi:https://doi.org/10.1017/S1431927603445911 (2003).

62 de la Rosa-Trevin, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* **195**, 93-99, doi:10.1016/j.jsb.2016.04.010 (2016).

63 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

64 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

65 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

66 Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, doi:10.7554/eLife.11182 (2015).

67 Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J Struct Biol* **157**, 281-287, doi:10.1016/j.jsb.2006.06.010 (2007).

68 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).

69 Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702, doi:10.1016/j.jsb.2021.107702 (2021).

70 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745, doi:10.1016/j.jmb.2003.07.013 (2003).

71 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, 874, doi:10.1038/s42003-021-02399-1 (2021).

72 Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat Methods* **17**, 923-927, doi:10.1038/s41592-020-0914-9 (2020).

73 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65, doi:10.1038/nmeth.2727 (2014).

74 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).

75 Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5 Zenodo (2019).

76 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815, doi:10.1006/jmbi.1993.1626 (1993).
77 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5 6 1-5 6 37, doi:10.1002/cpbi.3 (2016).
78 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
79 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
80 Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530, doi:10.1107/S2059798318002425 (2018).
81 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
82 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).
83 Mostosi, P., Schindelin, H., Kollmannsberger, P. & Thorn, A. Haruspex: A Neural Network for the Automatic Identification of Oligonucleotides and Protein Secondary Structure in Cryo-Electron Microscopy Maps. *Angew Chem Int Ed Engl* **59**, 14788-14795, doi:10.1002/anie.202000421 (2020).
84 Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods* **10**, 74-76, doi:10.1038/nmeth.2262 (2013).
85 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
86 Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol* **74**, 814-840, doi:10.1107/S2059798318009324 (2018).
87 Pintilie, G. *et al.* Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat Methods* **17**, 328-334, doi:10.1038/s41592-020-0731-1 (2020).
88 Young, J., Garikipati, N. & Durrant, J. D. BINANA 2: Characterizing Receptor/Ligand Interactions in Python and JavaScript. *J Chem Inf Model* **62**, 753-760, doi:10.1021/acs.jcim.1c01461 (2022).
89 Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res* **43**, W443-447, doi:10.1093/nar/gkv315 (2015).
90 Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-2786, doi:10.1021/ci200227u (2011).
91 Schrödinger, L. & DeLano, W. L. PyMOL. Retrieved from http://www.pymol.org/pymol. (2020).
92 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).
93 Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**, 2507-2524, doi:10.1110/ps.062416606 (2006).
94 Kostrhon, S. *et al.* CUL5-ARIH2 E3-E3 ubiquitin ligase structure reveals cullin-specific NEDD8 activation. *Nat Chem Biol* **17**, 1075-1083, doi:10.1038/s41589-021-00858-8 (2021).
95 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).
96 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324, doi:10.1093/nar/gku316 (2014).

3.10 Main Text References

1 Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650, doi:10.1038/nature00939 nature00939 [pii] (2002).
2 Harris, R. S. *et al.* DNA Deamination Mediates Innate Immunity to Retroviral Infection. *Cell* **113**, 803-809, doi:10.1016/S0092-8674(03)00423-9 (2003).
3 Pollpeter, D. *et al.* Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nat Microbiol* **3**, 220-233, doi:10.1038/s41564-017-0063-9 (2018).
4 Delviks-Frankenberry, K. A., Desimmie, B. A. & Pathak, V. K. Structural Insights into APOBEC3-Mediated Lentiviral Restriction. *Viruses* **12**, doi:10.3390/v12060587 (2020).

5 Uriu, K., Kosugi, Y., Ito, J. & Sato, K. The Battle between Retroviruses and APOBEC3 Genes: Its Past and Present. *Viruses* **13**, doi:10.3390/v13010124 (2021).
6 Cheng, A. Z. *et al.* APOBECs and Herpesviruses. *Viruses* **13**, doi:10.3390/v13030390 (2021).
7 Harris, R. S. & Dudley, J. P. APOBECs and virus restriction. *Virology* **479-480**, 131-145, doi:10.1016/j.virol.2015.03.012 (2015).
8 Burdick, R. C. *et al.* HIV-1 uncoats in the nucleus near sites of integration. *Proc Natl Acad Sci U S A* **117**, 5486-5493, doi:10.1073/pnas.1920631117 (2020).
9 Christensen, D. E., Ganser-Pornillos, B. K., Johnson, J. S., Pornillos, O. & Sundquist, W. I. Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science* **370**, doi:10.1126/science.abc8420 (2020).
10 Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V. & Campbell, E. M. Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nat Microbiol* **5**, 1088-1095, doi:10.1038/s41564-020-0735-8 (2020).
11 Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F. & Malim, M. H. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog* **5**, e1000330, doi:10.1371/journal.ppat.1000330 (2009).
12 Huthoff, H. & Malim, M. H. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J Virol* **81**, 3807-3815, doi:10.1128/JVI.02795-06 (2007).
13 Stopak, K., de Noronha, C., Yonemoto, W. & Greene, W. C. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* **12**, 591-601, doi:10.1016/s1097-2765(03)00353-8 (2003).
14 Anderson, B. D. & Harris, R. S. Transcriptional regulation of APOBEC3 antiviral immunity through the CBF-beta/RUNX axis. *Sci Adv* **1**, e1500296, doi:10.1126/sciadv.1500296 (2015).
15 Binning, J. M. *et al.* Fab-based inhibitors reveal ubiquitin independent functions for HIV Vif neutralization of APOBEC3 restriction factors. *PLoS Pathog* **14**, e1006830, doi:10.1371/journal.ppat.1006830 (2018).
16 Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060 (2003).
17 Khan, M. A. *et al.* Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. *J Virol* **75**, 7252-7265, doi:10.1128/JVI.75.16.7252-7265.2001 (2001).
18 Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* **2**, E275, doi:10.1371/journal.pbio.0020275 (2004).
19 Compton, A. A., Hirsch, V. M. & Emerman, M. The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe* **11**, 91-98, doi:10.1016/j.chom.2011.11.010 (2012).
20 Daugherty, M. D. & Malik, H. S. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet* **46**, 677-700, doi:10.1146/annurev-genet-110711-155522 (2012).
21 Etienne, L., Hahn, B. H., Sharp, P. M., Matsen, F. A. & Emerman, M. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe* **14**, 85-92, doi:10.1016/j.chom.2013.06.002 (2013).
22 Zhang, W., Du, J., Evans, S. L., Yu, Y. & Yu, X. F. T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction. *Nature* **481**, 376-379, doi:nature10718 [pii] 10.1038/nature10718 (2012).
23 Jäger, S. *et al.* Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481**, 371-375, doi:nature10693 [pii] 10.1038/nature10693 (2012).
24 Kouno, T. *et al.* Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat Struct Mol Biol* **22**, 485-491, doi:10.1038/nsmb.3033 (2015).
25 Maiti, A. *et al.* Crystal Structure of a Soluble APOBEC3G Variant Suggests ssDNA to Bind in a Channel that Extends between the Two Domains. *J Mol Biol* **432**, 6042-6060, doi:10.1016/j.jmb.2020.10.020 (2020).
26 Yang, H. *et al.* Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. *Nat Commun* **11**, 632, doi:10.1038/s41467-020-14377-y (2020).
27 Hu, Y. *et al.* Structural basis of antagonism of human APOBEC3F by HIV-1 Vif. *Nat Struct Mol Biol* **26**, 1176-1183, doi:10.1038/s41594-019-0343-6 (2019).

28 Siu, K. K., Sultana, A., Azimi, F. C. & Lee, J. E. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat Commun* **4**, 2593, doi:10.1038/ncomms3593 (2013).
29 Ball, K. A. *et al.* Conformational Dynamics of the HIV-Vif Protein Complex. *Biophys J* **116**, 1432-1445, doi:10.1016/j.bpj.2019.03.014 (2019).
30 Compton, A. A., Malik, H. S. & Emerman, M. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120496, doi:10.1098/rstb.2012.0496 (2013).
31 Guo, Y. *et al.* Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature* **505**, 229-233, doi:10.1038/nature12884 (2014).
32 Zhang, L. *et al.* Function analysis of sequences in human APOBEC3G involved in Vif-mediated degradation. *Virology* **370**, 113-121, doi:10.1016/j.virol.2007.08.027 (2008).
33 Conticello, S. G., Harris, R. S. & Neuberger, M. S. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* **13**, 2009-2013, doi:10.1016/j.cub.2003.10.034 (2003).
34 Bulliard, Y. *et al.* Functional analysis and structural modeling of human APOBEC3G reveal the role of evolutionarily conserved elements in the inhibition of human immunodeficiency virus type 1 infection and Alu transposition. *J Virol* **83**, 12611-12621, doi:10.1128/JVI.01491-09 (2009).
35 Russell, R. A. & Pathak, V. K. Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. *J Virol* **81**, 8201-8210, doi:JVI.00395-07 [pii]
10.1128/JVI.00395-07 (2007).
36 Gallois-Montbrun, S. *et al.* Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. *J Virol* **81**, 2165-2178, doi:10.1128/JVI.02287-06 (2007).
37 Stenglein, M. D., Matsuo, H. & Harris, R. S. Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *J Virol* **82**, 9591-9599, doi:10.1128/JVI.02471-07 (2008).
38 Kozak, S. L., Marin, M., Rose, K. M., Bystrom, C. & Kabat, D. The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. *J Biol Chem* **281**, 29105-29119, doi:10.1074/jbc.M601901200 (2006).
39 Yamashita, T., Kamada, K., Hatcho, K., Adachi, A. & Nomaguchi, M. Identification of amino acid residues in HIV-1 Vif critical for binding and exclusion of APOBEC3G/F. *Microbes Infect* **10**, 1142-1149, doi:10.1016/j.micinf.2008.06.003 (2008).
40 He, Z., Zhang, W., Chen, G., Xu, R. & Yu, X.-F. Characterization of Conserved Motifs in HIV-1 Vif Required for APOBEC3G and APOBEC3F Interaction. *Journal of Molecular Biology* **381**, 1000-1011, doi:10.1016/j.jmb.2008.06.061 (2008).
41 Russell, R. A., Smith, J., Barr, R., Bhattacharyya, D. & Pathak, V. K. Distinct domains within APOBEC3G and APOBEC3F interact with separate regions of human immunodeficiency virus type 1 Vif. *Journal of Virology* **83**, 1992-2003, doi:10.1128/JVI.01621-08 (2009).
42 Chen, G., He, Z., Wang, T., Xu, R. & Yu, X.-F. A patch of positively charged amino acids surrounding the human immunodeficiency virus type 1 Vif SLVx4Yx9Y motif influences its interaction with APOBEC3G. *Journal of Virology* **83**, 8674-8682, doi:10.1128/JVI.00653-09 (2009).
43 Chaipan, C., Smith, J. L., Hu, W. S. & Pathak, V. K. APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol* **87**, 444-453, doi:10.1128/JVI.00676-12 (2013).
44 Letko, M. *et al.* Vif proteins from diverse primate lentiviral lineages use the same binding site in APOBEC3G. *J Virol* **87**, 11861-11871, doi:10.1128/JVI.01944-13 (2013).
45 York, A., Kutluay, S. B., Errando, M. & Bieniasz, P. D. The RNA Binding Specificity of Human APOBEC3 Proteins Resembles That of HIV-1 Nucleocapsid. *PLoS Pathog* **12**, e1005833, doi:10.1371/journal.ppat.1005833 (2016).
46 Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640-645, doi:10.1038/nature05731 (2007).
47 Fischer, E. S. *et al.* Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49-53, doi:10.1038/nature13527 (2014).
48 Gaba, A., Flath, B. & Chelico, L. Examination of the APOBEC3 Barrier to Cross Species Transmission of Primate Lentiviruses. *Viruses* **13**, doi:10.3390/v13061084 (2021).
49 Letko, M., Booiman, T., Kootstra, N., Simon, V. & Ooms, M. Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface. *Cell Rep* **13**, 1789-1799, doi:10.1016/j.celrep.2015.10.068 (2015).

50 Binning, J. M., Chesarino, N. M., Emerman, M. & Gross, J. D. Structural Basis for a Species-Specific Determinant of an SIV Vif Protein toward Hominid APOBEC3G Antagonism. *Cell Host Microbe* **26**, 739-747 e734, doi:10.1016/j.chom.2019.10.014 (2019).
51 Schröfelbauer, B., Chen, D. & Landau, N. R. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3927-3932, doi:10.1073/pnas.0307132101 (2004).
52 Horn-Ghetko, D. *et al.* Ubiquitin ligation to F-box protein targets by SCF-RBR E3-E3 super-assembly. *Nature* **590**, 671-676, doi:10.1038/s41586-021-03197-9 (2021).
53 Albin, J. S. *et al.* Dispersed sites of HIV Vif-dependent polyubiquitination in the DNA deaminase APOBEC3F. *J Mol Biol* **425**, 1172-1182, doi:10.1016/j.jmb.2013.01.010 (2013).
54 Huttenhain, R. *et al.* ARIH2 Is a Vif-Dependent Regulator of CUL5-Mediated APOBEC3G Degradation in HIV Infection. *Cell Host Microbe* **26**, 86-99 e87, doi:10.1016/j.chom.2019.05.008 (2019).
55 Iwatani, Y. *et al.* HIV-1 Vif-mediated ubiquitination/degradation of APOBEC3G involves four critical lysine residues in its C-terminal domain. *Proceedings of the National Academy of Sciences* **106**, 19539-19544, doi:10.1073/pnas.0906652106 (2009).
56 van der Kuyl, A. C. & Berkhout, B. The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology* **9**, 92, doi:10.1186/1742-4690-9-92 (2012).
57 Soros, V. B., Yonemoto, W. & Greene, W. C. Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. *PLoS Pathog* **3**, e15, doi:10.1371/journal.ppat.0030015 (2007).
58 Goila-Gaur, R., Khan, M. A., Miyagi, E. & Strebel, K. Differential sensitivity of "old" versus "new" APOBEC3G to human immunodeficiency virus type 1 vif. *J Virol* **83**, 1156-1160, doi:10.1128/JVI.01734-08 (2009)

4 Deep Mutational Scans of HIV-1 Vif Proteins Reveal Their Evolutionary Constraints and Adaptive Potential

4.1 Introduction

Human Immunodeficiency Virus type 1 (HIV-1) belongs to the lentivirus genus of retroviruses, with origins in primates tracing back millions of years (1-3). The HIV-1 genome includes the *gag, pol,* and *env* genes, which encode core structural, enzymatic, and envelope proteins, respectively. Additionally, it includes accessory genes *vif, vpr, vpu,* and *nef*, which facilitate HIV-1's evasion of host innate immune proteins called restriction factors (4, 5). These accessory genes allow HIV-1 to circumvent host innate immune defenses, promoting viral replication and enhancing pathogenicity (6-8).

The dynamic interactions between lentiviral accessory genes and host restriction factors exert evolutionary pressures on both the virus and the host (9-11). As host restriction factors have evolved to combat these viruses, lentiviruses, including the precursors to HIV-1, have recurrently (or subsequently) evolved strategies to evade these defenses. This evolutionary arms race results in a continuous cycle of adaptation, whereby selection on the virus enhances its persistence and pathogenicity. This results in pressure on host species, over a longer timescale, to evolve mechanisms to recognize, evade, or tolerate viral infections. One of the most potent defenses against HIV-1 and other viruses in mammalian cells is mediated by the Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide-like 3 (APOBEC3, or A3) family of restriction factors. In primates, the *APOBEC3* gene family encodes seven proteins, named A3A-A3H, all of which are cytidine deaminases. Among these, A3G is the most potent inhibitor of HIV-1 (12).

The antiviral activity of A3G requires its incorporation into daughter virions and subsequent delivery to the target cell, where it is present at the site of viral reverse transcription. A3G binds

to and deaminates single-stranded viral DNA at the second cytidine (dC) of 5′-CC dinucleotide sites during reverse transcription (13, 14), introducing mutations into nascent HIV reverse transcripts (13-16), leading to missense mutations in the provirus and ultimately preventing successful infection and replication. The dinucleotide preference exhibited by A3G is unique among A3 proteins (17, 18), enabling A3G-specific G-to-A hypermutations to be readily documented in HIV sequences derived from clinical samples (19-23).

Lentiviruses, such as HIV-1, encode Vif proteins to counteract the threat imposed by A3G and other A3 proteins. Vif prevents A3 packaging by facilitating its degradation in the producer cell, thereby reducing the likelihood of A3s being incorporated into daughter virions and mitigating the threat of A3-mediated hypermutation of the viral genome (24, 25). Vif achieves this by interacting with multiple cellular proteins to assemble an E3 ubiquitin ligase complex, resulting in ubiquitination and subsequent proteasomal degradation of A3G (24-26). Due to their potent antiviral restriction, A3 proteins can act as significant barriers to cross-species transmission of viruses like HIV-1. For instance, molecular differences between A3 proteins in hominids and Old World monkeys have been pivotal barriers to cross-species transmission of Simian Immunodeficiency Viruses (SIVs) and have shaped the evolution of HIV-1 (27, 28). These barriers have, in turn, significantly influenced the evolution of the Vif protein, selecting for Vif mutations that allow antagonism of the divergent A3 proteins encoded by different species (29-31).

One of the most dramatic examples of such innovation can be observed in the evolution of the Vif gene during the adaptation of HIV-1 to facilitate transmission to humans. The ancestor of the HIV-1 *vif* gene is the *vif* gene from SIVcpz, which infects chimpanzees. The SIVcpz *vif* gene originated from SIVrcm, a virus infecting the red-capped mangabey, an Old World monkey species. Of the three sites required for full antagonism of human A3G (R15, W70, Q/H83), SIVrcm

already encoded two (R15 and W70), allowing it to weakly antagonize hominid A3G (31) despite evolving under the pressure of Old World monkey A3G (27, 28). This weak antagonism enabled SIVrcm to transmit across species and infect chimpanzees. After transmission, Vif evolved to fully antagonize chimpanzee A3G through a single amino-acid change from tyrosine (Y) to histidine (H) at site 83, conferring full resistance to hominid A3G (31). This led to the present-day SIVcpz Vif protein, which can fully antagonize both chimpanzee and human A3G (which are similar in sequence), contributing to the emergence of the HIV-1 M/N lineage capable of infecting humans (29).

Evolutionary analyses of A3G protein sequences from diverse primates and Vif from a broad panel of SIVs have provided key insights into the interacting interfaces of these proteins (30-32). These analyses predicted residues critical for Vif binding and antagonism of A3G, which were recently confirmed by the resolution of cryo-EM structures of HIV-1 Vif bound to human A3G (32-34). Specifically, the amino acid residues at sites 15 and 83 determine the binding ability of Vif proteins from diverse SIVs to A3G proteins from different primate hosts (29, 31, 35). These structures not only validated several key sites but also revealed new ones, including Vif residues that interact with RNA, which acts as a molecular glue between A3G and Vif. These and previous studies showed that the N-terminus of Vif is largely responsible for most interactions with the A3 proteins, RNA, and host PP2A (36-39), whereas the C-terminus of Vif binds the conserved host proteins Elongin B, CBFβ, and Cullin 5 (collectively called the VCBC complex). The Vif-VCBC interaction facilitates the hijacking of a host E3 ubiquitin ligase complex (40, 41) that leads to ubiquitin-mediated degradation of A3G to escape host restriction (24, 26, 42, 43). The same complex also helps target a different host protein, PPP2A, for degradation to mediate cell cycle arrest (44-46).

Despite these evolutionary and structural insights, our understanding of the selective constraints that shape HIV-1 Vif function and its evolutionary trajectory remains incomplete. Vif sequences from individuals living with HIV-1 are highly divergent and exhibit a spectrum of natural variants that affect its ability to antagonize A3 proteins (47-51). Additionally, there is considerable variation in the efficacy of Vif proteins to counteract A3 restriction, even among isolates within the same subtype (52-54). Consequently, much remains to be uncovered about the mutational landscape and evolutionary potential of HIV-1 Vif, as well as the broader implications of Vif variability for viral evolution and pathogenesis.

Deep mutational scanning (DMS) is a recently developed experimental technique that allows for the comprehensive assessment of all possible single amino-acid substitutions on a protein's function, stability, and interactions. This high-throughput method enables the simultaneous testing of all possible variants in a single pool (55-57). The resulting detailed mutational landscape offers invaluable insights into protein evolution by identifying variants that enhance or impair activity, elucidating the molecular basis for functional differences in key residues, and defining the mutational tolerance at these critical interfaces. Here, we employed DMS to generate and test variant libraries of the *vif* gene from two different Clade B viruses of HIV-1, aiming to comprehensively elucidate the structural and functional constraints acting on HIV-1 Vif evolution for A3G antagonism. By analyzing the changes in variant frequencies before and after selection, we identified mutations that confer fitness advantages or disadvantages in the context of A3G antagonism. This allowed us to delineate the constraints and potential evolutionary pathways available to HIV-1 Vif for retaining or even enhancing A3G antagonism. Our analysis identified the mutational tolerance of individual sites in Vif and highlighted differences between the two viral strains that we mutagenized. We validated the predicted fitness effects of selected Vif variants

from our DMS assay by measuring their ability to protect the viral genome from A3G-specific mutations. Thus, our *vif* DMS analyses provide a comprehensive map of the mutational landscape of HIV-1 Vif, underscoring the delicate balance between maintaining existing functionality and exploring new functionality in the context of a high-stakes host-virus arms race.

4.2 A pooled functional selection assay for *vif* variants

To characterize the selective constraints acting on HIV-1 Vif to antagonize A3G, we designed a scheme for testing Vif variants in a pooled assay (Fig. 1A) and tested it using two pilot screens. In the first screen, we used a mixture of HIV-1 LAI viruses encoding either wild-type (WT) *vif* or a mutant *vif* encoding a premature stop codon at position 40. We generated replication-competent viruses encoding either of these two *vif* variants by ligating each into an HIV-1 provirus, which was slightly modified to move the *vif* start codon out of an overlapping reading frame with Integrase (see Methods). We then infected SupT1 cells engineered to express physiological levels of A3G (17) with the mixture of WT and mutant *vif* HIV-1 viruses in independent triplicate experiments at a low multiplicity of infection (MOI). Twenty-four hours post-infection, we removed the supernatant and replaced it with fresh media to eliminate any virions produced before Vif translation. Subsequently, 72 hours post-infection, we collected the 'pre-selection' supernatant and used it to infect a second set of A3G-expressing SupT1 cells (Fig. 1A). Three days after the second-round infection, we collected the 'post-selection' supernatant for sequencing (Fig 1A). If a Vif variant encoded by the virus could not restrict A3G packaging during the first infection, the packaged A3G would introduce mutations into the viral genome during reverse transcription in the second infection, impairing the production of subsequent daughter virions. Consequently, the second viral supernatant is expected to be enriched for

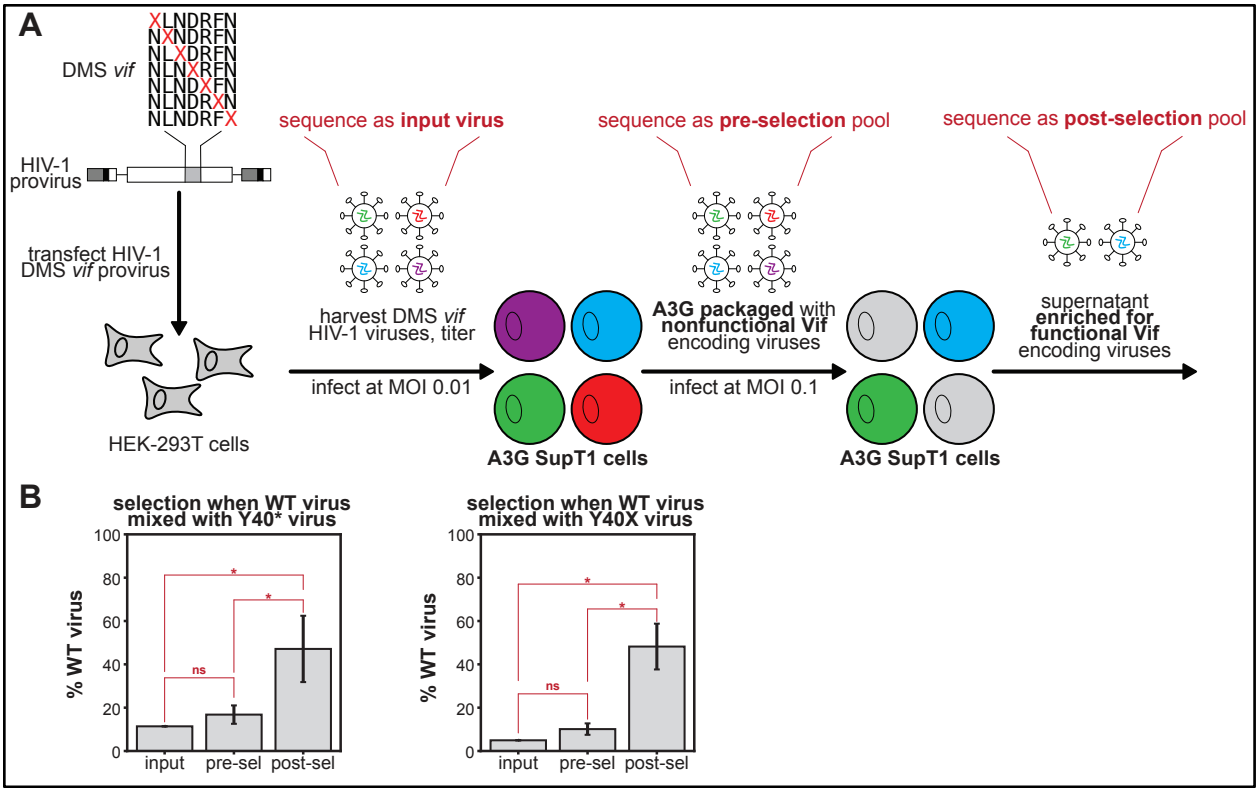
functional *vif* variants that successfully antagonized A3G and depleted for variants that failed to affect A3G packaging.

Since the restrictive effect of A3G should only be observed after two passages in A3G-expressing cells, we anticipated that the Vif variant containing the premature stop codon would be significantly depleted after selection for A3G antagonism but not before selection. As expected, we found no significant difference between the fraction of viruses encoding WT versus mutant *vif* between the input virus and the pre-selection virus replicates after the first passage (Fig. 1B). However, we did observe a significant reduction in Vif variants encoding a stop codon, translating to a selective index of -42% and -38% in post-selection replicates relative to either the input virus or the pre-selection virus (Fig. 1B).

In the second pilot screen, we tested all possible codons at position 40 of Vif. We focused on this site because the cryo-EM structure of LAI Vif in the VCBC complex bound to human A3G identified position 40 as an RNA-interacting site (32). Although this interaction is important, it is not critical for Vif function since other Vif residues also interact with the same RNA nucleotide (32). Consequently, we anticipated some constraint on residue 40 following selection for A3G antagonism but expected it to be less severe than for the stop codon in our previous experiment. We observed a selective index of -46% and -42%, comparable to the unfavorable premature stop codon in our first pilot screen when comparing the percentage of non-wild-type codons between the post-selection virus and either the input virus or the pre-selection virus (Fig. 1B).

Both our pilot screens confirmed that our selection scheme could successfully enrich functional Vif variants while depleting non-functional variants that could not successfully antagonize human A3G. Importantly, this enrichment was observed only in the post-selection

supernatant, indicating that the selective pressures were indeed imposed by A3G and were not artifacts of experimental bottlenecking or external factors. Cumulatively, these data reveal passage through A3G-expressing cells can exert discriminative selection on multiple Vif variants within an input viral pool.



4-1 **Selection of Functional Vif Variants through Passaging in A3G-Expressing Cells A.** Schematic representation of the experimental workflow for deep mutational scanning (DMS) of HIV-1 Vif. The DMS libraries were constructed by introducing mutations at codons 12-115 of the Vif protein from two HIV-1 clade B strains (LAI and 1203). These mutant libraries were inserted into a replication-competent HIV-1 provirus and transfected into HEK293T cells to produce infectious virus particles. The produced virus was used to infect SUPT1 cells engineered to express physiological levels of A3G at a multiplicity of infection (MOI) of 0.01. After 72 hours, the supernatant was harvested and used to infect a new batch of SUPT1 cells expressing A3G. Pre-selection and post-selection viral RNA were isolated from the supernatants and subjected to deep sequencing to identify and quantify the DMS Vif variants. **B.** Validation of the selection process using control libraries. One control library contained a mix of wild-type and premature stop codon mutations at position 40. Another control library included a mini-library of different variants at position 40. No significant change in the proportion of the wild-type amino acid (Y) was observed between the input and pre-selection samples, while a significant reduction in the wild-type proportion was observed in the post-selection samples. This reduction indicates

effective selection for functional Vif variants that can counteract A3G, demonstrating that the selection process is A3G-specific. The selection process effectively differentiates between highly detrimental mutations (stop codons) and a spectrum of different mutations with variable impact on the fitness of Vif against A3G.

4.3 Deep mutational scanning HIV-1 LAI *vif* variants for A3G antagonism

We synthesized DMS libraries of all possible Vif variants from the HIV-1 LAI strain at amino acid positions 12-115 to characterize the selective constraints acting on HIV-1 Vif to antagonize A3G (Fig. 1). We chose to mutagenize this region instead of the entire protein sequence because the C-terminal domain of Vif is largely involved in interactions involving host proteins in the VCBC complex, while the N-terminal domain makes interactions with A3s, RNA, and PP2A. Instead of random mutagenesis, we employed a method that specified variant codons to minimize the introduction of new A3G mutation targets. This approach ensured that we could distinguish the original DMS variants we introduced from new variants that might emerge post-selection due to A3G mutagenesis. We selected the LAI strain because its *vif* gene has been previously shown to robustly antagonize human A3G (58), and it is well-adapted to tissue culture. Additionally, the LAI Vif protein was used in the cryo-EM structure of Vif in the VCBC complex bound to human A3G (32). The resulting proviral variant library is referred to as DMS LAI *vif* HIV-1. As in the pilot screens, we used this DMS *vif* HIV-1 library to infect A3G-expressing SupT1 cells (Fig. 1).

To visualize the selective retention or depletion of all the variants from the DMS LAI *vif* HIV-1 library, we first averaged the log enrichment ratios for each variant at a given position within each of the three replicate infections. We then averaged the values for all variants at each position to generate site-level mean log enrichment ratios (Fig. 2A). Across all sites in the DMS LAI *vif* library, we found that the site-wide mean log enrichment ratio is -0.78. This log

enrichment ratio indicates that, on average, most DMS variants that deviate from the original LAI residue have a fitness disadvantage for virus replication in the presence of A3G. This low mutational tolerance may be due to intrinsic reasons, such as protein instability, or extrinsic reasons, such as impairment in A3G binding.

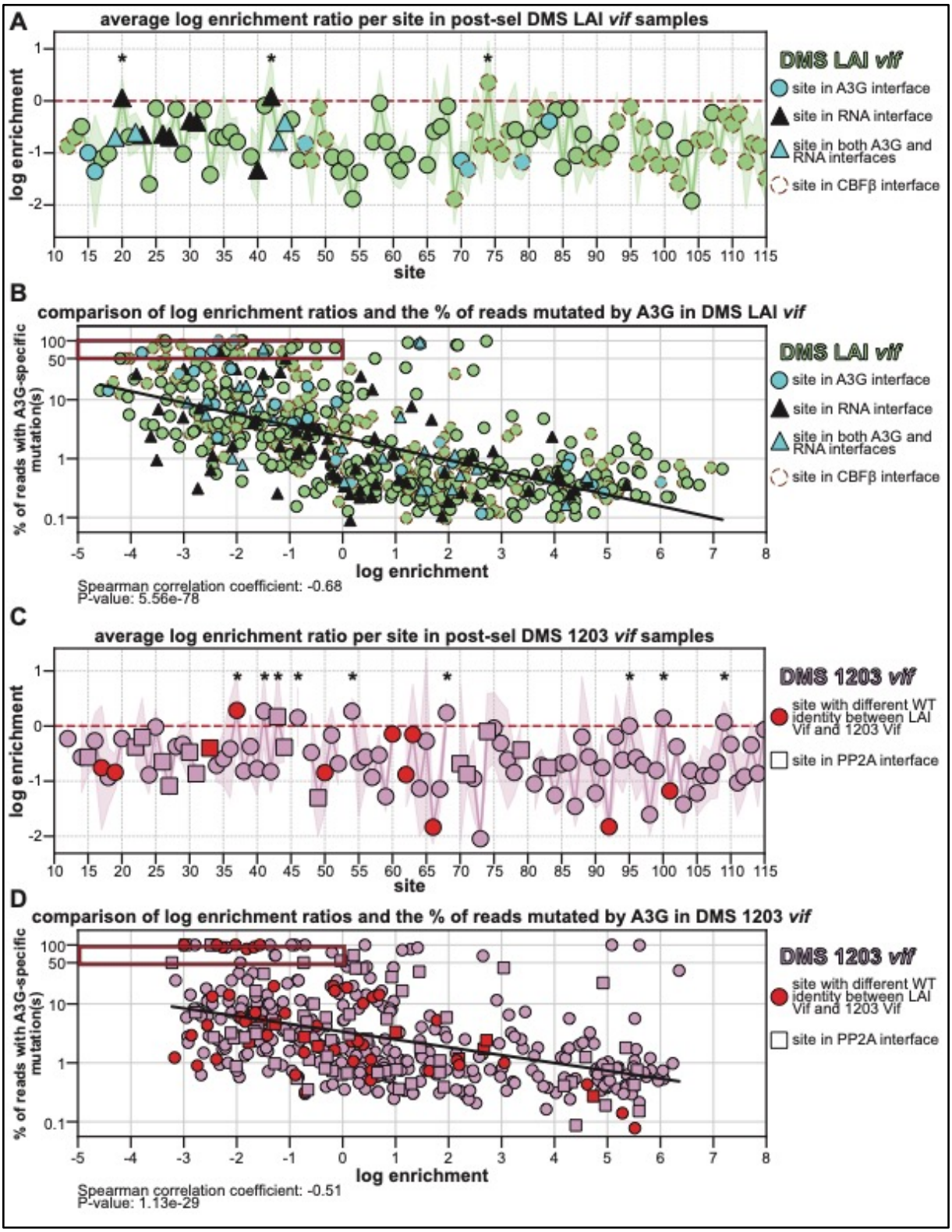
Figure 2A highlights the Vif sites that interact with A3G (blue circles), RNA (black triangles), or CBFβ (hatched brown outlines) within the VCBC complex based on the cryo-EM structure of LAI Vif bound to human A3G (32). The average log enrichment ratios for sites interacting with A3G, RNA, and CBFβ are -0.88, -0.54, and -0.83, respectively. This suggests that sites mediating Vif interactions with host proteins A3G or CBFβ are more constrained than average, whereas RNA-interacting sites are relatively more mutationally tolerant, consistent with structural studies that indicate redundancy in Vif interactions with RNA.

The term "arms-race interface" refers to sites 15, 70, and 83, where evolutionary changes enabled Vif to counter a new host species' A3G after cross-species transmission (29-31, 35, 59). Two sites in the arms-race interface directly interacting with A3G, sites 15 and 70, have average log enrichment ratios of -1.00 and -1.15, respectively. The lower-than-average log enrichment scores for sites 15 and 70 indicate that these sites are highly mutationally constrained in LAI Vif's ability to antagonize A3G. Conversely, site 83, also in the arms-race interface that directly interacts with A3H is less constrained than average, with an average log enrichment ratio of -0.39. This suggests both lower mutational constraint and the possibility that a subset of variants at site 83 might improve viral fitness by better antagonizing A3G. We explore this possibility later in this study.

Although most of the 103 sites in the DMS LAI *vif* library had depleted enrichment ratios, three sites – 20, 42, and 74 – exhibited average log enrichment ratios above 0 (asterisks, Fig.

2A). Sites 20 and 42 are involved in binding RNA, which acts as a molecular glue between Vif and A3G and is necessary for the ability of Vif to antagonize A3G, whereas site 74 binds CBFβ (32-34). Positive enrichment scores at these positions could imply that they are structurally unconstrained and can tolerate many substitutions. Alternatively, these positive enrichment scores might suggest that Vif variants with mutations at these sites outperform LAI Vif in A3G antagonism, thereby enhancing overall viral fitness.

Viral fitness is not the only readout of Vif function in our DMS assay. Failure to protect against A3G restriction would result in A3G-induced mutations in daughter viruses. To determine which *vif* variants were associated with G to A mutations in a context favored by A3G, we sequenced a 366-basepair amplicon of *vif*. This approach enabled us to identify the original DMS variant and additional mutations acquired elsewhere within the amplicon. Specific sites in the LAI *vif* template that are potential mutational targets of A3G were identified, and the percentage of reads of each DMS *vif* variant that indicated an incidence of A3G-target mutations and compared this percentage against the log enrichment ratio of that specific *vif* variant (Fig. 2B). Indeed, we found a strong negative correlation between the log enrichment ratio of a DMS *vif* variant and the percentage of reads indicating A3G-mediated mutations, with a Spearman correlation coefficient of -0.68 (Fig. 2B). Thus, the fitter (more enriched) DMS *vif* variants also have fewer reads with A3G-mediated mutations. Conversely, less-fit *vif* variants (those depleted in the post-selection supernatant) are more likely to harbor A3G-specific mutations. The convergence of these two orthogonal measures confirms that the relative fitness of any *vif* variant is correlated with its ability to suppress A3G-specific hypermutation in the viral genome.



4-2 Mutational Tolerance of Vif in the Presence of A3G A. Average log enrichment ratios for amino acid positions in post-selection DMS library of LAI Vif. The log enrichment ratio reflects the relative abundance of each variant after selection, indicating the fitness impact of mutations. Sites interacting with A3G (blue circles), CBFβ (hatched brown outlines), and RNA (black

triangles) are indicated, based on the cryo-EM structure of LAI Vif bound to human A3G (32). **B.** Scatter plot comparing log enrichment ratios and the percentage of reads with A3G-specific mutations in post-selection DMS LAI Vif variants. A strong negative correlation (Spearman correlation coefficient: -0.68) indicates that more enriched variants are better at restricting A3G packaging, resulting in fewer A3G-specific mutations. **C.** Average log enrichment ratios for amino acid positions in post-selection DMS library of 1203 Vif. Sites with different wild-type identities between LAI Vif and 1203 Vif are marked (red fill). Sites interacting with PP2A in the published structure (39) are indicated by squares**. D.** Similar scatter plot for DMS 1203 Vif variants, showing a Spearman correlation coefficient of -0.51.

4.4 Vif constraints and adaptive potential differ between two divergent HIV-1 strains

Next, we assessed how Vif evolution may have shaped its mutational constraints and adaptive potential by considering the 1203 strain, which also belongs to the same clade B of HIV-1 viruses as the LAI strain. Unlike the LAI strain, which is well-adapted to tissue culture, the 1203 strain is derived from a primary isolate of HIV-1 that evolved in the context of the antiviral A3H haplotype enabling it to antagonize both A3G and A3H (60), unlike the LAI vif. Another key difference between the LAI and 1203 Vif proteins is their ability to bind and degrade a third host factor, PP2A. LAI Vif cannot degrade PP2A, partially because it encodes a Glycine (G) instead of the Lysine or Arginine (K/R) required for PP2A degradation at the critical residue 33 (44). In contrast, 1203 Vif encodes a lysine at this critical position. Moreover, 1203 Vif is nearly identical to a closely related primary isolate Vif that can degrade PP2A (46, 61, 62), differing only at a single residue (valine at residue 31 instead of isoleucine) that reduces but does not abrogate PP2A degradation (46). Thus, whereas LAI Vif can only bind A3G, 1203 Vif can bind A3G, A3H, and PP2A. The abilities to bind A3H and PP2A are not mutually exclusive (62). Therefore, comparing LAI and 1203 *vif* genes provides an opportunity to investigate strain-specific differences in mutational tolerance towards A3G antagonism.

Using the identical strategy as outlined for the HIV-1 LAI strain, we constructed a second DMS library of variants using the *vif* gene from the 1203 strain and assessed variants for their

ability to antagonize human A3G (Fig. 2C). We also compared the mutational constraints and adaptive potential acting on the two Vif proteins. We found that the site-wide average log enrichment ratio is -0.63 for 1203 *vif*, compared to -0.78 for LAI *vif*. Like LAI *vif*, we found a strong concordance between 1203 *vif* variants with low log enrichment ratios (indicating they were depleted post-selection) and variants associated with signatures of A3G-inflicted hypermutation (Spearman correlation coefficient of -0.51, Fig. 2D).

The LAI and 1203 *vif* genes differ in sequence at 13 of 103 C-terminal codons analyzed via DMS: 17, 19, 33, 37, 47, 50, 60, 61, 62, 63, 66, 92, and 101 (Fig. 2C, indicated by red fill). The average log enrichment ratio at sites with different wild-type sequence identities in the DMS LAI and 1203 *vif* libraries is -0.95 and -0.84, respectively. The more severe than average constraint at divergent sites compared to conserved sites suggests the critical role these residues play in function and overall viral fitness. Consequently, these residues may significantly influence the mutational tolerance of other sites in the protein through epistasis. Furthermore, differences in the individual constraints of conserved sites between LAI and 1203 Vif may be attributed to the limitations on interface optimization imposed by the unique constraints resulting from non-conserved sites.

The site-level average log enrichment ratios in the DMS 1203 vif library are -0.57 at site 15, -0.68 at site 70, and -0.76 at site 83 for 1203 vif; these three residues are the arms race interface with A3G. In contrast, the log enrichment ratios were -1.00, -1.15, and -0.39, for sites 15, 70, and 83, respectively, for LAI vif. While site 83 is the least constrained of the three A3G-interacting residues in LAI vif, it is the most constrained in 1203 vif. Sites 15, 70, and 83 are all involved in binding not only A3G but also PP2A (39). Therefore, 1203 Vif may be subjected to different individual constraints at these sites compared to LAI Vif, which cannot bind PP2A (46).

We also found nine sites in the 1203 Vif library with positive log enrichment ratios: 37, 41, 43, 46, 54, 68, 95, 100, and 109 (asterisks, Fig. 2C), compared to only three in the LAI Vif library (Fig. 2A). One of these sites, residue 43, interacts with RNA, A3G (32, 33), and PP2A (39). Thus, a site constrained by three different interactions is capable of significantly improved fitness upon selection to one specific host pressure rather than three; this suggests that additional pressures on Vif functions constrain its evolution.

4.5 Individual Vif variants enhance or reduce A3G antagonism

Although the site-wide average log enrichment ratio is a useful tool to measure the overall mutational tolerance of any given site or the entire Vif protein, it does not fully capture the different fitness levels of individual variants at a given site. Examining individual log enrichment ratios allows us to identify trends in preferred biochemical properties for a specific site and compare the effects of individual variants with their representation in nature. By identifying variants at the extreme end of the threshold for A3G-specific mutations, we can highlight their critical impact on Vif functionality and more comprehensively characterize its fitness landscape. This approach moves beyond simply classifying variants as functional or non-functional, providing a nuanced understanding of why specific mutations have varying impacts on overall viral fitness.

We first investigated the most detrimental *vif* variants in each library. To do this, we identified variants with negative individual log enrichment ratios and evidence of A3G-hypermutation in 50% to 100% of reads (indicated by the red boxes in Fig. 2B and 2D). We identified some Vif variants, including those at the critical Vif zinc-coordinating sites 108 and 114, as loss-of-function in both the LAI and DMS 1203 vif datasets (Fig. 3A, Fig. 3B); this is expected since such variants would be universally detrimental to all Vif proteins. Additionally,

we identified 22 unique 'unfit' vif variants in the DMS LAI dataset (Fig. 3A). Five of these 22 variants are present at Vif's A3G or RNA binding interfaces, while ten are at the CBFβ interface. Of these 22 variants, two –F39P and M16S–appeared consistently across all three replicates. Notably, seven of 22 variants involved a mutation to a proline residue, whose unique ability to introduce kinks into protein chains likely results in significant distortions in Vif structure (63).

We identified only 11 unique 'unfit' variants in the 1203 dataset (Fig. 3B). Four of these variants –K19R, Y44R, V66G, and V98Q– are present in all three replicates, while one, T68W, is present in two. Our finding that the K19R mutation is highly deleterious to 1203 Vif is especially noteworthy because R is the wildtype residue in LAI Vif, where it interacts with both RNA and A3G (32). Thus, although R, K, and N are equally represented at residue 19 among HIV-1 Vif proteins, our findings suggest that these substitutions are not benign; their consequences on Vif function are likely driven by unique epistatic interactions conferred by other differences in LAI and 1203 Vif proteins.

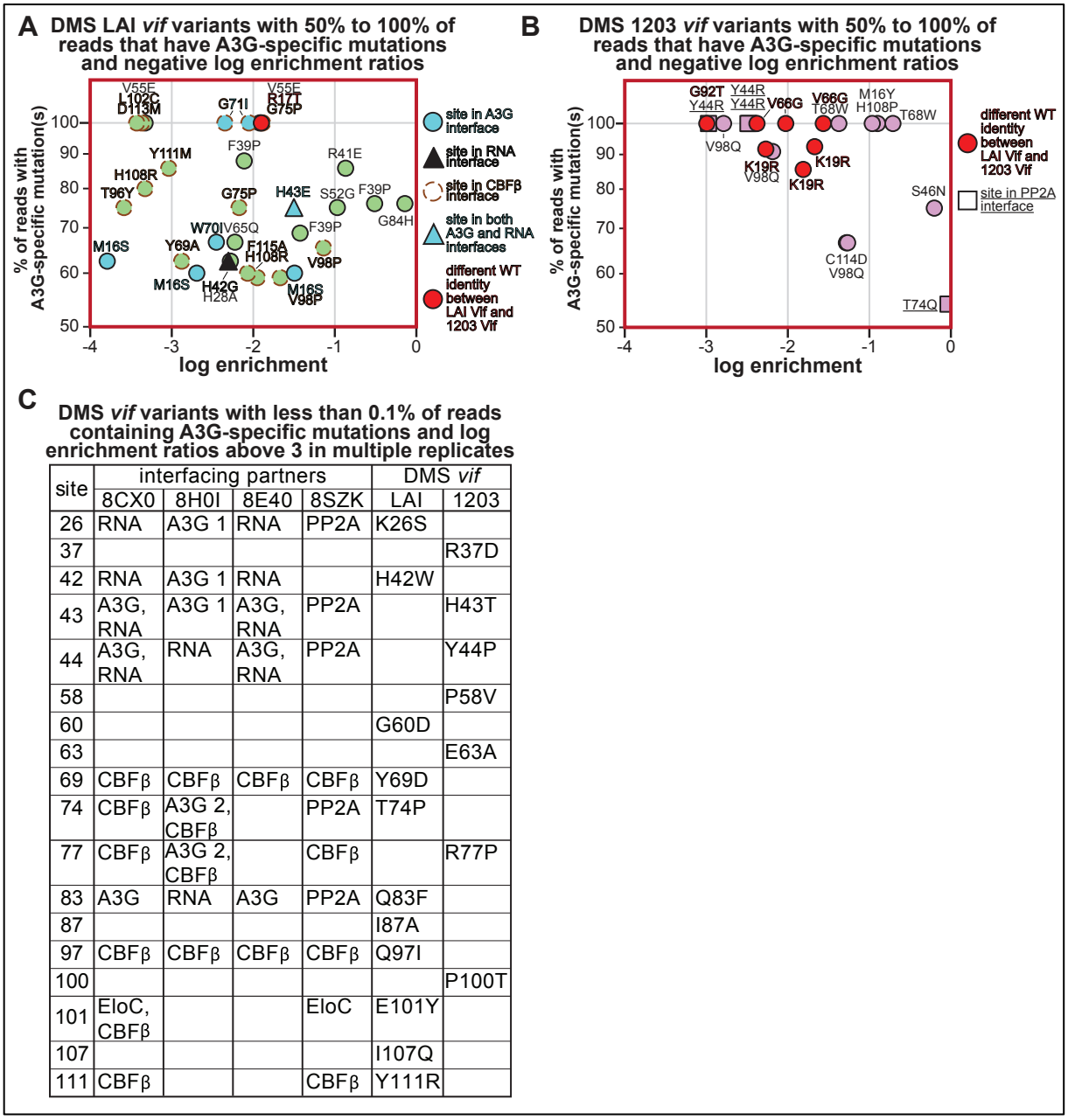
Next, we identified the fittest vif variants using the dual measures of high individual log enrichment ratios (at least two of the three replicates had log enrichment ratios above three) and a low percentage of reads (<0.1%) with signatures of A3G-hypermutation (Fig. 3C). We identified 11 and 7 'fitter' vif variants unique to the LAI and 1203 datasets, respectively. Of these 18 'fitter' variants, four– G60D and E101Y in LAI and R37D and E63A in 1203– occur in sites that differ between LAI and 1203 Vif proteins. None of these variants convert to the residue present in the other Vif. Despite all variants being tested in both 1203 and LAI vif (except for 63A, which only had sufficient coverage in the DMS 1203 vif library), their improved fitness is only specific to one strain, further emphasizing the role epistasis plays in shaping strain-specific adaptive evolution.

We mapped the positions where mutations led to fitter vif variants onto various structures in which Vif interactions with host proteins have been defined (Fig. 3C). These include a structure of LAI Vif within the VCBC complex bound to human A3G (PDB: 8CX0) (32), structures of Vif from other strains bound to RNA and human A3G (PDBs: 8E40, 8H0I) (33, 34), and a structure of Vif bound to PP2A (PDB: 8SZK) (39). These comparisons reveal some variance in the interaction of different Vif proteins with different host components. For example, site 74 interacts only with CBFβ in the structure of LAI Vif (32) but interacts with both CBFβ and A3G in the structure of Vif that features an A3G dimer (34). This site also interacts with PP2A in the structure of Vif bound to PP2A (39). These differences reveal strain-specific differences in Vif interactomes or the dynamic nature of these interactions even within the same HIV-1 strain.

Five of the eleven mutations that enhance fitness in the DMS LAI vif library occur in the CBFβ binding interface, one in the A3G binding interface, and two in the RNA binding interface. This distribution partly reflects the fact that the CBFβ interface (68 sites) is much larger than the A3G (11 sites) or RNA (13 sites) interfaces. Given that only one of these LAI mutations, Q83F, occurs at the arms-race interface, our findings suggest that there are only a few possibilities to further improve interactions at this specific interface. Therefore, Vif improvements in fitness are more likely to occur at other interfaces, such as those with RNA or components of the VCBC complex.

Surprisingly, seven of 18 ‘fitter’ variants involve mutations at sites without interaction partners in the four analyzed structures. One of these variants, I107Q, occurs at the same site as a previously described Vif mutant isolated from an individual living with HIV-1 (50). The variation at this site was previously determined to affect A3G antagonism in a cohort of HIV-1-

infected individuals from Nairobi. Thus, Vif adaptation is not limited to defined interfaces with host proteins. We hypothesize that such 'non-interface' variants may affect Vif dynamics or alternative conformational states, ultimately enabling them to indirectly influence A3G binding and antagonism.



C

DMS *vif* variants with less than 0.1% of reads containing A3G-specific mutations and log enrichment ratios above 3 in multiple replicates

site	interfacing partners				DMS *vif*	
	8CX0	8H0I	8E40	8SZK	LAI	1203
26	RNA	A3G 1	RNA	PP2A	K26S	
37						R37D
42	RNA	A3G 1	RNA		H42W	
43	A3G, RNA	A3G 1	A3G, RNA	PP2A		H43T
44	A3G, RNA	RNA	A3G, RNA	PP2A		Y44P
58						P58V
60					G60D	
63						E63A
69	CBFβ	CBFβ	CBFβ	CBFβ	Y69D	
74	CBFβ	A3G 2, CBFβ		PP2A	T74P	
77	CBFβ	A3G 2, CBFβ		CBFβ		R77P
83	A3G	RNA	A3G	PP2A	Q83F	
87					I87A	
97	CBFβ	CBFβ	CBFβ	CBFβ	Q97I	
100						P100T
101	EloC, CBFβ			EloC	E101Y	
107					I107Q	
111	CBFβ			CBFβ	Y111R	

4-3 **Individual Vif variants enhance or reduce A3G antagonism A.** Identification of DMS LAI Vif variants with negative log enrichment ratios and high percentages of A3G-specific mutations. Key variants at critical sites, such as F39P and M16S, are highlighted. **B.** Identification of DMS 1203 Vif variants with similar characteristics. Variants such as K19R and V98Q are noted for

their detrimental effects on Vif function. **C.** Most fit DMS variants, identified by high log enrichment ratios and low percentages of A3G-specific mutations in multiple replicates, are highlighted. These variants, such as G60D and E101Y in LAI Vif, suggest potential pathways for adaptive evolution.

4.6 Residue 83 is a hotspot of Vif adaptation in nature, and in our analyses

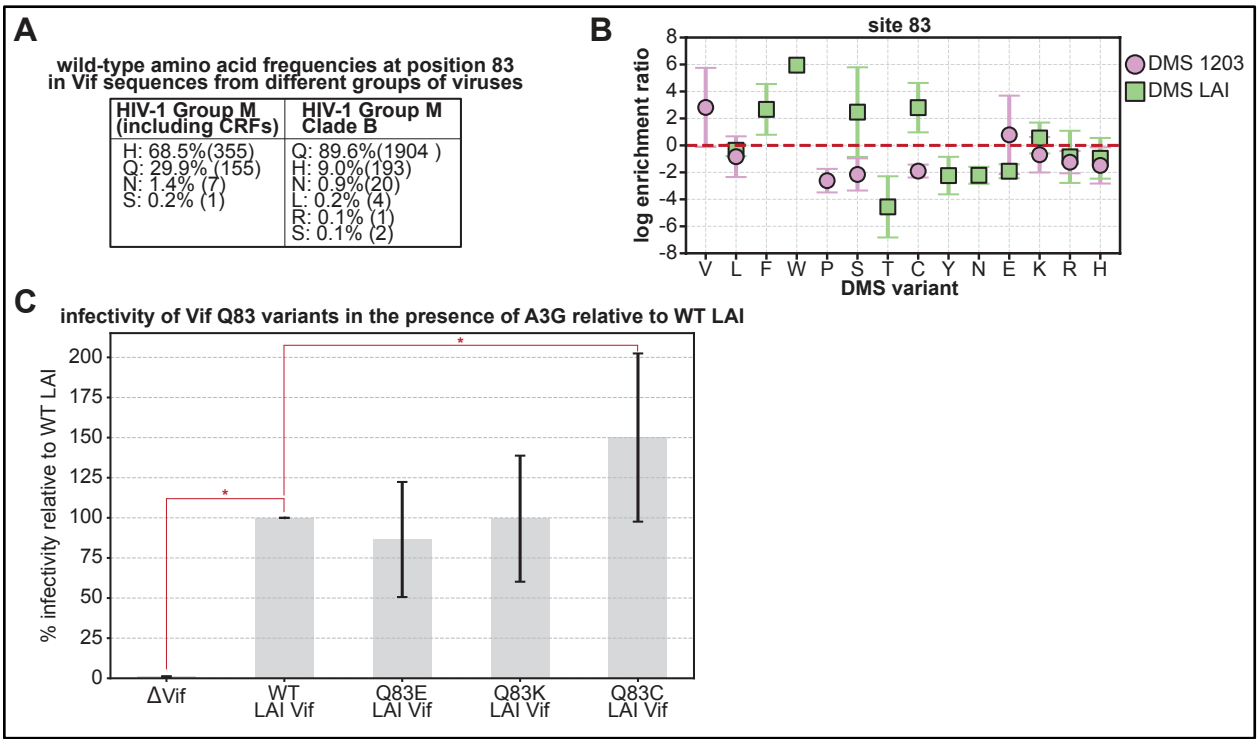
Vif from SIVrcm adapted to overcome hominid A3G via a single amino-acid change from tyrosine (Y) to histidine (H) at the site equivalent to site 83 in HIV Vif (31), making this position the primary molecular determinant governing species-specific antagonism of hominid A3G. In HIV-1 sequences, residue 83 toggles between mostly two different amino acids in HIV-1 Vif (Fig. 4A). Vif sequences from four representatives of each HIV-1 Group M subtype, including circulating recombinant forms (CRFs), encode a histidine (H) at position 83 in 68.5% of the sequences, glutamine (Q) in 29.9%, and an asparagine (N) in 1.4%. However, within clade B HIV-1 strains, position 83 is either Q (89.6%), H (9.0%), or N (0.9%). Both LAI and 1203 vif genes encode a Q at this critical residue. Substitutions to H are fitness-impaired in both DMS LAI and 1203 datasets, with an individual log enrichment ratio of -0.96 and -1.48, respectively, across three replicates. The deleterious impact of Q83H substitutions likely reflects a clade B-specific constraint. Furthermore, a Q83Y substitution, which mimics a reversion to the residue found in SIVrcm Vif, is strongly fitness-impaired, with an individual log enrichment ratio of -2.24 in the DMS LAI vif dataset (we are unable to calculate enrichment ratios in the 1203 dataset due to insufficient representation of this variant in the preselection virus). The strong depletion of tyrosine (Y) variants at site 83 in the DMS LAI vif library provides further evidence that passaging DMS libraries through A3G-expressing cells results in the depletion of non-functional variants.

Our DMS analyses also uncovered a total of seven variants with enhanced fitness at site 83 –five in LAI vif and two in 1203 vif (Fig. 4B). Intriguingly, two of these fitter variants are

83F (phenylalanine) and 83W (tryptophan), which are structurally very similar to 83Y (tyrosine), which is very unfit. This suggests that the hydroxyl group unique to Y specifically disrupts Vif binding to hominid A3G, resulting in its fitness loss compared to the other two aromatic amino acids. Similarly, 83S (serine) is one of the most enriched variants at this site, whereas 83T (threonine) is among the most depleted despite the strong biochemical similarities between serine and threonine (both have a hydroxyl domain, but threonine has an extra methyl group instead of a hydrogen atom like in serine). Cumulatively, these differences between biochemically similar amino acids highlight the nuanced biochemical landscape of the Vif-A3G interface and provide insight into the constraints imposed by different functional groups on various amino acids.

The most enriched vif variant at this site is 83C (cysteine) in the LAI DMS dataset, with an individual log enrichment ratio of 2.80. And yet, this variant is not represented in any Vif sequences from any analyzed primary HIV-1 isolates (Fig. 4A). However, this variant was previously shown to confer antagonism of a divergent A3G in an *in vivo* experimental evolution experiment, confirming its potential as an adaptive mutation (1). To further explore the fitness effects of mutations at this critical residue, we made individual point mutants at this site in a replication-competent provirus with a luciferase reporter gene and measured viral infectivity in the context of A3G (Fig. 4C). The Q83C variant exhibited a statistically significant increase in infectivity compared to the WT LAI ($p = 0.038$), underscoring its enhanced replication capability in the presence of A3G. Conversely, the less fit Q83E variant, with an average log enrichment ratio of -1.92, was less capable of antagonism of A3G than the WT LAI. Furthermore, the Q83K variant, which had an average log enrichment ratio of 0.55, exhibited no significant difference in infectivity compared to the WT LAI Vif. Our findings demonstrate that the enrichment and depletion of variants in the pooled DMS datasets represent an accurate spectrum of fitness effects

even when measured individually. However, the fact that these recovered variants are often not seen in primary isolates may imply some other constraint, for example a fitness tradeoff with antagonism of other A3 proteins.



4-4 **Residue 83 is a hotspot of Vif adaptation in nature and in our analyses**
A. Wild-type amino acid frequencies at position 83 in Vif sequences from different HIV-1 groups and subtypes. Histidine (H) and glutamine (Q) are the predominant amino acids at this position, with significant variation between groups. **B.** Log enrichment ratios for DMS variants at position 83 in LAI and 1203 Vif libraries. Variants such as cysteine (C), phenylalanine (F), and tryptophan (W) were enriched, indicating potential fitness advantages. **C.** Infectivity assays of replication-competent HIV-1 encoding Vif Q83 variants relative to the no-A3G control. Infectivity assays were performed with three biological replicates for each variant, and the data were normalized to the infectivity of the wildtype virus in the presence of A3G. The results were analyzed using statistical methods to ensure robustness and reliability. Mean infectivity values and standard deviations were calculated for each variant. Mann-Whitney U tests were conducted to compare the infectivity of each variant to that of the wildtype. The p-values obtained were adjusted using the Bonferroni correction to account for multiple comparisons.
The Q83C variant, significantly enriched in the DMS library, conferred increased replication in the presence of A3G, demonstrating its adaptive potential.

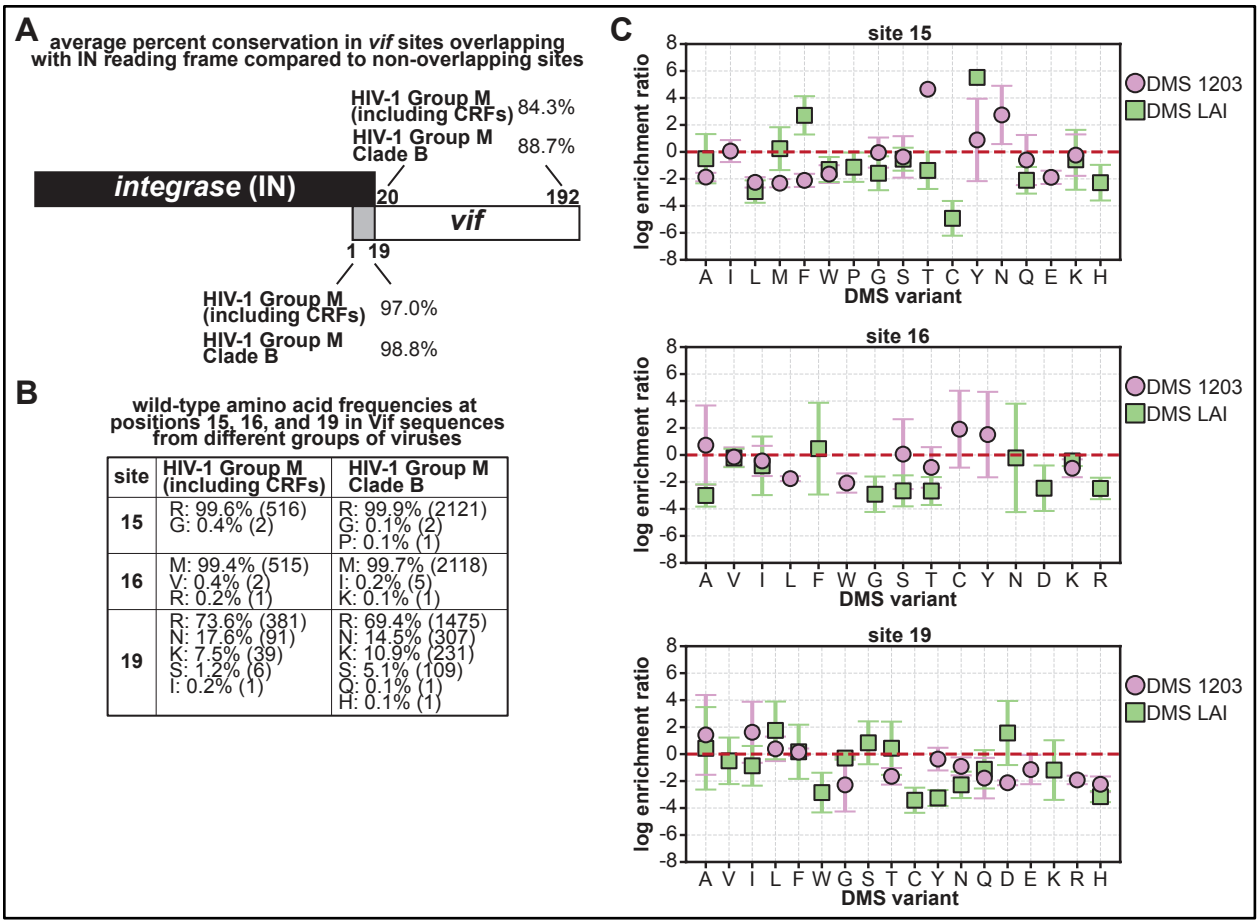
4.7 Overprinting with *integrase* (IN) in HIV-1 genomes shapes *vif* evolution

To maximize the encoding of genetic information in their minimal genomes, many RNA viruses and retroviruses employ overprinted open reading frames, wherein the same nucleotide sequence can express different proteins in different reading frames. In HIV-1 Group M genomes, the N-terminal codons for sites 1-19 in vif are overprinted with the C-terminal codons of the integrase (IN) reading frame. Unlike the Vif accessory protein, the IN protein is a core component essential for the HIV-1 life cycle within host cells, with its C-terminal tail being crucial for successful viral replication (64, 65). Across Group M Vif, the average amino acid identity conservation of C-terminal IN residues is 97%, compared to 84% at all other Vif sites. In Clade B specifically, the average amino acid identity conservation of these sites is 99% compared to 89% at all other sites (Fig. 5A).

Our DMS strategy removed the overprinting between the *vif* and *integrase* genes, potentially releasing any constraints that arose from encoding two distinct proteins from the same nucleotide sequence. Indeed, we found enriched *vif* variants at several originally overprinted codons in both DMS LAI and 1203 datasets. The overprinted sequence includes Vif residues 15, 16, and 19, which are located within the A3G binding interface. Site 15, located at the A3G and PP2A interface (32, 33, 39), and site 16, located in the A3G interface (32-34), are over 99% conserved in both Group M and specifically within Clade B (Fig. 5B). However, our DMS analysis reveals that several amino acid substitutions at either of these two residues have no fitness consequences (log enrichment ratios close to 0), *e.g.,* glycine (G) and proline (P) at site 15, and valine (V), isoleucine (I), and lysine (K) at site 16 (Fig. 5C). Moreover, phenylalanine (F) is enriched at both sites 15 and 16 in the LAI DMS dataset, whereas tyrosine (Y) is enriched at these sites in the 1203 DMS dataset (Fig. 5C). And yet, despite the fitness advantages they

confer in our analyses, neither Y nor F is ever found at either of these sites in sequenced HIV-1 isolates (Fig. 5B). Our findings suggest that *vif* adaptive potential might be restricted due to fitness tradeoffs associated with such mutations due to overprinting by *integrase* in nature.

Conversely, maintaining Vif function may have also imposed subtle constraints on the *integrase* gene. For example, Vif residue 19 overlaps with the integrase stop codon. A TGA (opal) stop codon in integrase would result in a glutamic (E) or aspartic (D) acid residue in Vif. Both E19 and D19 variants are significantly impaired for fitness in the DMS 1203 vif library (Fig. 5C). Indeed, the TGA stop codon is never sampled in primary isolates of HIV-1 or SIVcpz, presumably to maintain Vif functionality. Either of the TAA/TAG stop codons in integrase would result in R, N, K, or S at residue 19 in Vif. Asparagine (N) is depleted in both DMS datasets despite its significant representation in nature (Fig. 5C). Similarly, lysine (K) is depleted in the DMS LAI vif library (even though it is natively found in 1203 Vif), whereas arginine (R) is depleted in DMS 1203 vif library (even though it is natively found in LAI Vif) (Fig. 5C). Strain-specific selective pressures, influenced by site 19's interaction with A3G, RNA, or PP2A (32-34, 39), may determine which of the four amino acids –R, N, S, and K –are preferred in Vif and, therefore, which of the two stop codons –TAA or TAG –are preferred in the integrase gene.



site	HIV-1 Group M (including CRFs)	HIV-1 Group M Clade B
15	R: 99.6% (516) G: 0.4% (2)	R: 99.9% (2121) G: 0.1% (2) P: 0.1% (1)
16	M: 99.4% (515) V: 0.4% (2) R: 0.2% (1)	M: 99.7% (2118) I: 0.2% (5) K: 0.1% (1)
19	R: 73.6% (381) N: 17.6% (91) K: 7.5% (39) S: 1.2% (6) I: 0.2% (1)	R: 69.4% (1475) N: 14.5% (307) K: 10.9% (231) S: 5.1% (109) Q: 0.1% (1) H: 0.1% (1)

4-5 **Overprinting with *integrase* (IN) in HIV-1 genomes shapes *vif* evolution**
A. Average percent conservation of Vif sites overlapping with the integrase reading frame compared to non-overlapping sites. Sites 1-19, which overlap with the integrase coding region, exhibit higher conservation across Group M and Clade B HIV-1, suggesting functional constraints imposed by this genetic overlap. **B.** Wild-type amino acid frequencies at positions 15, 16, and 19 in Vif sequences from different groups of HIV-1 viruses. These positions, located within the A3G binding interface, are highly conserved. **C.** Log enrichment ratios for DMS variants at positions 15, 16, and 19 in LAI and 1203 Vif libraries. Variants enriched at these sites, such as glycine (G) and proline (P) at position 15, indicate potential adaptive mutations that are restricted in natural sequences due to the overlap with *integrase*.

4.8 Discussion

By using deep mutational scanning libraries, we comprehensively investigated the evolutionary constraints and adaptive potential of HIV-1 Vif in the context of A3G antagonism. We demonstrated that A3G antagonism imposes significant constraints on Vif evolution, with most variants, particularly those involved in binding A3G and CBFβ, exhibiting an overall

fitness disadvantage (as evidenced by negative log enrichment ratios). Additionally, we identified enriched (fitter) Vif variants. These findings allowed us to compare the fitness landscapes of Vif from two HIV-1 clade B strains, LAI and 1203, providing insights into the differential mutational tolerances, structural and functional constraints, and potential evolutionary trade-offs.

We found that Vif residues interacting with RNA were more mutationally tolerant than those interacting with A3G or CBFβ. Two of the three sites with average log enrichment ratios above 0 in the DMS LAI vif dataset were at RNA-interacting positions, indicating that most variants at these sites confer improved viral fitness. This higher tolerance could result from the redundancy in Vif-RNA interactions (32, 33), potentially reducing the level of constraint on each individual site. Indeed, LAI Vif residue 26, which coordinates multiple interactions with different RNA nucleotides, is less mutationally tolerant than other residues interacting simultaneously with the same RNA nucleotide (32). Higher tolerance at RNA-interacting sites could also reflect the fact that protein-RNA interactions are less dependent on specific amino acids and more on charged or polar amino acids. For example, RNA-interacting sites 22 and 23 in LAI Vif have a broad tolerance to polar amino acids (32).

Despite their similarity, comparing the DMS LAI and 1203 *vif* datasets highlights strain-specific constraints and adaptive potential. The DMS 1203 *vif* dataset demonstrated a higher global mutational tolerance than that of DMS LAI *vif* but had fewer variants with significantly higher fitness. Unlike LAI Vif, which can only antagonize human A3G, 1203 Vif can bind human A3G, A3H, and PP2A. Due to its exposure to these triple host pressures, especially the evolution towards antagonism of A3H in an individual who was homozygous for the more active A3H hapII genotype, 1203 Vif may not be optimized solely for A3G fitness. However, since our assay specifically tested for antagonism of A3G, relaxing this triple pressure may have allowed

many variants to emerge in the 1203 dataset, which might have been constrained due to interactions with either A3H or PP2A in the primary isolate.

Our analysis also identified fitter Vif variants, including some previously observed to confer improved antagonism of a divergent A3G in an *in vivo* experimental evolution experiment (1). Our experiments further confirmed the improved fitness conferred by this mutation in LAI Vif. It remains unclear whether this mutation facilitates interactions with the established arms-race interface in A3G or promotes novel interactions at a different, conserved site in A3G across species, explaining its ability to antagonize divergent A3G proteins. Despite their improved fitness, these Vif variants are not commonly observed in nature, potentially due to evolutionary trade-offs with antagonism of the other A3 proteins, such as A3F, or other host proteins, such as PP2A.

Another constraint that differs between our DMS analyses and primary HIV-1 isolates is maintaining Vif and Integrase functionality in overprinted regions in primary isolates. This constrains Vif residues 1-19 and might constrain the *integrase* stop codon. Interestingly, the size of the overprinted region is different among HIV-1 Groups M and N, as well as Groups O, P, and HIV-2, each resulting from a unique cross-species transmission event. Comparing the sequence diversity within and between these different Vif proteins may elucidate additional evolutionary constraints imposed by overprinting in distinct cross-species transmission events.

The strongest conclusion from our DMS analyses was the pervasive role of epistasis within and between different Vif proteins. For example, residues that are different between 1203 and LAI Vif proteins would be expected to be more tolerant of mutations, yet we found they are more mutationally constrained. Additionally, many variants that improve Vif fitness in both DMS datasets are in positions that do not directly interact with either A3G or RNA. Nonetheless,

single amino acid substitutions at these positions can have profoundly beneficial consequences for Vif antagonism of human A3G. Thus, many changes in the genetic backbone of the Vif proteins, often away from host interaction interfaces, both constrain the mutational landscape as well as facilitate adaptive evolution.

Our *vif* DMS analyses reveal a strong negative correlation between improved variant fitness and evidence of A3G-specific hypermutation. Vif variants that effectively restrict A3G packaging reduce A3G-mediated mutations and exhibit higher fitness, underscoring the importance of Vif's interaction with A3G in maintaining viral replication competency. Compared to merely classifying a variant as enriched or depleted, these additional analyses describe a more comprehensive spectrum of the Vif fitness landscape. Identifying Vif variants with A3G-specific mutations also offers insights into sublethal Vif mutations, which could facilitate viral adaptation, enabling the virus to explore new evolutionary trajectories while maintaining sufficient overall fitness. This nuanced understanding of the fitness landscape reveals a dynamic interplay between Vif and A3G that drives viral evolution and zoonosis.

4.9 Methods

Cell culture and transfections

SUPT1 cells expressing A3G at physiological levels were a gift from Reuben Harris (described in (17)). These along with unmodified SUPT1 (ATCC CRL-1942) cells were maintained in RPMI Medium (Gibco, #11875093), with 10% Fetal Bovine Serum (GE Healthcare, #SH30910.03), 1% Penicillin Streptomycin (Gibco, #15140122), and 1% HEPES at 37°C, referred to as RPMI complete. HEK293T cells (ATCC CRL-3216) were maintained in Dulbecco's modified Eagle's medium (Gibco, #11965092) with 10% Fetal Bovine Serum (GE Healthcare, #SH30910.03), and 1% Penicillin Streptomycin (Gibco, #15140122) at 37°C.

Plasmids

The plasmid expressing the 1203 Vif gene was previously described (60). All point mutations for infectivity experiments were introduced using PrimeSTAR HA DNA Polymerase (Takara, R010B). These variants were ligated into the HIV-1 LAI- based molecular clone pLAIΔVifΔEnvLuc2 provirus using MluI/XbaI as previously described (31, 60). pMD2.G and psPAX2 were gifts from Didier Trono (Addgene #12259/12260). The DMS vif libraries were digested with MluI and XbaI restriction enzymes and ligated into a replication competent HIV-1 proviral construct in which the start codons in the region from the original start of vif to the end of the integrase gene were mutated, as previously described in (28). The resulting proviruses are referred to as DMS LAI vif HIV-1 and DMS 1203 vif HIV-1. All constructs were confirmed by sequencing.

Construction of DMS *vif* libraries

The DMS *vif* libraries were designed to mutagenize the codons for amino acids 12-115 in LAI and 1203 Vif. Variants representing the full panel of amino acids at each site were designed,

opting against codons that would create new A3G-specific mutation targets unless there were no other options for an amino acid variant. The sequences were synthesized by Twist Biosciences (San Francisco, CA, USA). Variant codons at each position were pooled in equal proportions and delivered in a 96 well plate format with each well containing the variant pool for a different site as gene fragments flanked by MluI/XbaI digest sites. The pools for all sites were combined equally to make the complete DMS *vif* libraries.

Virus production, titering, and infection

To generate infectious virus encoding the DMS *vif* libraries, 293T (ATCC) cells were seeded at a density of 1.5×10^5 cells/ml in a 6-well plate. The following day, cells were transfected with 1000 ng of DMS *vif* HIV-1 provirus using TransIT-LT1 reagent (Mirus Bio LLC) with 3 μL of transfection reagent per μg of DNA. 24 hours later, the media was replaced with fresh media. 72 hours later, virus was harvested and filtered through a 0.2-μm-pore-size filter. We treated the supernatant with DNAse to digest residual plasmid DNA as in (66) and froze aliquots at −80 ° C. We thawed and titered aliquots using the TZM-bl assay in the presence of 20 μg/mL DEAE-Dextran (Sigma; D9885) as described in (67).

To generate VSVG pseudotyped virus for validation experiments, 293T cells (ATCC) were plated at 2×10^5 cells/mL in 2 mL in 6-well plates one day prior to transfection. 293Ts were transfected with 667 ng lentiviral plasmid, 500 ng psPAX2 and 333 ng MD2G. One day post-transfection media was replaced. 72 hours later, virus was harvested and filtered through a 0.2-μm-pore-size filter. All viral and lentiviral infections and transductions were done in the presence of 20 μg/mL DEAE-Dextran (Sigma; D9885).

Screening DMS vif viruses

For each of three replicates, $3.1e^7$ cells were infected a were infected at an MOI of 0.01 by spinoculation at 1100xg for 30 min with 20 μg/mL DEAE-Dextran. Supernatant was removed immediately following spinoculation and replaced with fresh media. 24 hours post spiculation media was replaced with fresh media. 72 hours post spinoculation, supernatant was harvested, supplemented with 20 μg/mL DEAE-Dextran, and used to infect $1.5e^7$ cells. Immediately following the second spinolcution, the media was collected as the pre-selection virus, filtered through a 0.2 μm filter, and concentrated in SW28 rotor for 1 hour at 4°C. After resuspension in PBS, viral RNA was extracted (Zymogen Quick-RNA Viral Kit, Zymogen, R1035). 24 hours post the second spinoculation, the media was replaced with fresh media. 72 hours post second spinoculation, the post-selection supernatant was harvested, filtered, and concentrated as described above, and the viral RNA was extracted.

Sequencing analysis of viral supernatants

After RNA extraction, *vif* cDNA amplicons were generated by RT-PCR using primers specific to the pLAI construct and annealing outside of the region containing the mutagenized library. Libraries were prepared for sequencing as described in (68). Pooled, multiplexed libraries were then sequenced via a NextSeq P2 250/250 paired-end read run (Fred Hutch Genomics and Bioinformatics Shared Resource).

Functional validation of DMS *vif* variants

Single-cycle infectivity assays were previously described (60). 293T cells were seeded at a density of 1.5×10^5 cells/ml in a 6-well plate. The following day, cells were transfected with 600 ng provirus, 100 ng l-VSV-G, and 400 ng A3 or pcDNA4/TOPO empty vector. 72 hours later, virus was harvested and and filtered through a 0.2-μm-pore-size filter. Filtered viruses were

normalized for virion production using an RT-qPCR assay, as described previously (69, 70). A volume of virus equivalent to 2,000 mU/ml of reverse transcriptase was used for infection of SUPT1 cells. For infectivity assays, SUPT1 cells were seeded at 2×10^4 cells per well in a 96-well plate in media containing 20 μg/ml DEAE-dextran. 72 hours later, infected cells were lysed in luciferase lysis reagent (Bright-Glo; Promega catalog no. 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.

Analysis of primary isolate Vif sequences

All primary isolate sequence alignments were obtained from the Los Alamos National Laboratory curated sequence alignments.

4.10 References

1. A. A. Compton, H. S. Malik, M. Emerman, Host gene evolution traces the evolutionary history of ancient primate lentiviruses. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120496 (2013).
2. R. J. Gifford, Viral evolution in deep time: lentiviruses and mammals. *Trends Genet* **28**, 89-100 (2012).
3. K. R. McCarthy, A. Kirmaier, P. Autissier, W. E. Johnson, Evolutionary and Functional Analysis of Old World Primate TRIM5 Reveals the Ancient Emergence of Primate Lentiviruses and Convergent Evolution Targeting a Conserved Capsid Interface. *PLOS Pathogens* **11**, e1005085 (2015).
4. K. Strebel, HIV accessory proteins versus host restriction factors. *Curr Opin Virol* **3**, 692-699 (2013).
5. L. Bergantz, F. Subra, E. Deprez, O. Delelis, C. Richetta, Interplay between Intrinsic and Innate Immunity during HIV Infection. *Cells* **8** (2019).
6. R. A. Subbramanian, E. A. Cohen, Molecular biology of the human immunodeficiency virus accessory proteins. *J Virol* **68**, 6831-6835 (1994).
7. A. Seelamgari *et al.*, Role of viral regulatory and accessory proteins in HIV-1 replication. *Front Biosci* **9**, 2388-2413 (2004).
8. T. B. Faust, J. M. Binning, J. D. Gross, A. D. Frankel, Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. *Annu Rev Virol* **4**, 241-260 (2017).
9. F. Kirchhoff, Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host Microbe* **8**, 55-67 (2010).
10. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687-695 (2012).
11. G. Boso, C. A. Kozak, Retroviral Restriction Factors and Their Viral Targets: Restriction Strategies and Evolutionary Adaptations. *Microorganisms* **8** (2020).
12. C. Chaipan, J. L. Smith, W. S. Hu, V. K. Pathak, APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol* **87**, 444-453 (2013).
13. H. Zhang *et al.*, The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-98 (2003).
14. R. S. Harris *et al.*, DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809 (2003).

15. B. Mangeat *et al.*, Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103 (2003).
16. Q. Yu *et al.*, Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* **11**, 435-442 (2004).
17. J. F. Hultquist *et al.*, Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-11234 (2011).
18. A. Rathore *et al.*, The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. *J Mol Biol* **425**, 4442-4454 (2013).
19. J. P. Vartanian, A. Meyerhans, B. Asjo, S. Wain-Hobson, Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. *J Virol* **65**, 1779-1788 (1991).
20. M. Janini, M. Rogers, D. R. Birx, F. E. McCutchan, Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. *J Virol* **75**, 7973-7986 (2001).
21. Y. C. Ho *et al.*, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540-551 (2013).
22. K. M. Bruner *et al.*, Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043-1049 (2016).
23. K. A. Delviks-Frankenberry *et al.*, Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation. *PLoS Pathog* **12**, e1005646 (2016).
24. A. M. Sheehy, N. C. Gaddis, M. H. Malim, The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **9**, 1404-1407 (2003).
25. S. G. Conticello, R. S. Harris, M. S. Neuberger, The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* **13**, 2009-2013 (2003).
26. X. Yu *et al.*, Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060 (2003).
27. L. Etienne, B. H. Hahn, P. M. Sharp, F. A. Matsen, M. Emerman, Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe* **14**, 85-92 (2013).
28. L. Etienne *et al.*, The Role of the Antiviral APOBEC3 Gene Family in Protecting Chimpanzees against Lentiviruses from Monkeys. *PLoS Pathog* **11**, e1005149 (2015).
29. N. M. Chesarino, M. Emerman, HIV-1 Vif Gained Breadth in APOBEC3G Specificity after Cross-Species Transmission of Its Precursors. *J Virol* **96**, e0207121 (2022).
30. A. A. Compton, V. M. Hirsch, M. Emerman, The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe* **11**, 91-98 (2012).
31. J. M. Binning, N. M. Chesarino, M. Emerman, J. D. Gross, Structural Basis for a Species-Specific Determinant of an SIV Vif Protein toward Hominid APOBEC3G Antagonism. *Cell Host Microbe* **26**, 739-747 e734 (2019).
32. Y. L. Li *et al.*, The structural basis for HIV-1 Vif antagonism of human APOBEC3G. *Nature* **615**, 728-733 (2023).
33. F. Ito *et al.*, Structural basis for HIV-1 antagonism of host APOBEC3G via Cullin E3 ligase. *Sci Adv* **9**, eade3168 (2023).
34. T. Kouno *et al.*, Structural insights into RNA bridging between HIV-1 Vif and antiviral factor APOBEC3G. *Nat Commun* **14**, 4037 (2023).
35. M. Letko, T. Booiman, N. Kootstra, V. Simon, M. Ooms, Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface. *Cell Rep* **13**, 1789-1799 (2015).
36. R. A. Russell, V. K. Pathak, Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. *J Virol* **81**, 8201-8210 (2007).
37. G. Chen, Z. He, T. Wang, R. Xu, X. F. Yu, A patch of positively charged amino acids surrounding the human immunodeficiency virus type 1 Vif SLVx4Yx9Y motif influences its interaction with APOBEC3G. *J Virol* **83**, 8674-8682 (2009).
38. M. Kobayashi, A. Takaori-Kondo, Y. Miyauchi, K. Iwai, T. Uchiyama, Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. *J Biol Chem* **280**, 18573-18578 (2005).
39. Y. Hu *et al.*, Structural insights into PPP2R5A degradation by HIV-1 Vif. *Nat Struct Mol Biol* 10.1038/s41594-024-01314-6 (2024).

40. Y. Guo *et al.*, Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature* **505**, 229-233 (2014).
41. B. J. Stanley *et al.*, Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. *J Virol* **82**, 8656-8663 (2008).
42. M. Marin, K. M. Rose, S. L. Kozak, D. Kabat, HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* **9**, 1398-1403 (2003).
43. A. Mehle, J. Goncalves, M. Santa-Marta, M. McPike, D. Gabuzda, Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. *Genes Dev* **18**, 2861-2866 (2004).
44. D. J. Salamango *et al.*, HIV-1 Vif Triggers Cell Cycle Arrest by Degrading Cellular PPP2R5 Phospho-regulators. *Cell Rep* **29**, 1057-1065 e1054 (2019).
45. E. J. Greenwood *et al.*, Temporal proteomic analysis of HIV infection reveals remodelling of the host phosphoproteome by lentiviral Vif variants. *Elife* **5** (2016).
46. S. Marelli *et al.*, Antagonism of PP2A is an independent and conserved function of HIV-1 Vif and causes cell cycle arrest. *Elife* **9** (2020).
47. V. Simon *et al.*, Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* **1**, e6 (2005).
48. F. A. De Maio *et al.*, Effect of HIV-1 Vif variability on progression to pediatric AIDS and its association with APOBEC3G and CUL5 polymorphisms. *Infect Genet Evol* **11**, 1256-1262 (2011).
49. M. Ooms *et al.*, HIV-1 Vif adaptation to human APOBEC3H haplotypes. *Cell Host Microbe* **14**, 411-421 (2013).
50. J. Peng *et al.*, A naturally occurring Vif mutant (I107T) attenuates anti-APOBEC3G activity and HIV-1 replication. *J Mol Biol* **425**, 2840-2852 (2013).
51. E. W. Refsland *et al.*, 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**, e1004761 (2014).
52. Y. Iwabu *et al.*, Differential anti-APOBEC3G activity of HIV-1 Vif proteins derived from different subtypes. *J Biol Chem* **285**, 35350-35358 (2010).
53. M. Binka, M. Ooms, M. Steward, V. Simon, The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. *J Virol* **86**, 49-59 (2012).
54. I. Lisovsky *et al.*, HIV-1 subtype variability in Vif derived from molecular clones affects APOBEC3G-mediated host restriction. *Intervirology* **56**, 258-264 (2013).
55. D. M. Fowler, S. Fields, Deep mutational scanning: a new style of protein science. *Nat Methods* **11**, 801-807 (2014).
56. C. L. Araya, D. M. Fowler, Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol* **29**, 435-442 (2011).
57. D. M. Fowler, J. J. Stephany, S. Fields, Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat Protoc* **9**, 2267-2284 (2014).
58. M. Ooms, M. Letko, M. Binka, V. Simon, The Resistance of Human APOBEC3H to HIV-1 NL4-3 Molecular Clone Is Determined by a Single Amino Acid in Vif. *PLOS ONE* **8**, e57744 (2013).
59. H. Huthoff, M. H. Malim, Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J Virol* **81**, 3807-3815 (2007).
60. M. M. Li, L. I. Wu, M. Emerman, The range of human APOBEC3H sensitivity to lentiviral Vif proteins. *J Virol* **84**, 88-95 (2010).
61. K. Nagata, K. Shindo, Y. Matsui, K. Shirakawa, A. Takaori-Kondo, Critical role of PP2A-B56 family protein degradation in HIV-1 Vif mediated G2 cell cycle arrest. *Biochem Biophys Res Commun* **527**, 257-263 (2020).
62. D. J. Salamango *et al.*, Functional and Structural Insights into a Vif/PPP2R5 Complex Elucidated Using Patient HIV-1 Isolates and Computational Modeling. *J Virol* **94** (2020).
63. M. P. Williamson, The structure and function of proline-rich regions in proteins. *Biochem J* **297 (Pt 2)**, 249-260 (1994).
64. M. J. Dar *et al.*, Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase. *Retrovirology* **6**, 94 (2009).
65. C. Rocchi *et al.*, The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding. *Int J Mol Sci* **23** (2022).

66. H. K. Haddox, A. S. Dingens, J. D. Bloom, Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. *PLoS Pathog* **12**, e1006114 (2016).
67. A. S. Dingens, H. K. Haddox, J. Overbaugh, J. D. Bloom, Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. *Cell Host Microbe* **21**, 777-787 e774 (2017).
68. M. B. Doud, J. D. Bloom, Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. *Viruses* **8** (2016).
69. J. Vermeire *et al.*, 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**, e50859 (2012).
70. E. I. Garcia, M. Emerman, Recurrent Loss of APOBEC3H Activity during Primate Evolution. *J Virol* **92** (2018).

5 Discussion

This body of work aims to elucidate the evolutionary and functional dynamics of two critical antiviral restriction factors, Mx and A3G. Through comprehensive phylogenetic analysis, structural determination, and deep mutational scanning, we provide new insights into the ancient origins and evolutionary adaptations of these antiviral proteins.

The discovery of Mx orthologs in non-vertebrate lineages such as fungi (MxF proteins) and plants (DRP4 proteins) underscores the deep evolutionary roots of the Mx gene family. This widespread distribution implies that the antiviral functions of Mx proteins were established early in the evolutionary timeline, likely close to the origin of eukaryotes. These proteins likely played a critical role in the initial antiviral defense strategies, long before the development of the sophisticated interferon-mediated immune responses observed in vertebrates. Interestingly, our phylogenetic analysis also identified instances of viral genes that encode Mx-like proteins, suggesting a history of viral gene theft. For example, nucleocytoplasmic large DNA viruses (NCLDVs) appear to have acquired Mx-like genes from their eukaryotic hosts on multiple occasions. These viral Mx-like proteins may confer advantages in countering host defenses or in hijacking host cellular mechanisms for viral replication. The presence of these viral genes in the Mx phylogeny underscores the complex interplay between host and viral genomes, where genetic material is exchanged and repurposed across evolutionary timelines. This observation opens avenues for exploring how these viral proteins may have adapted Mx-like functions to evade host immune responses.

The discovery that A3G and HIV-1 Vif interactions are mediated via RNA as a molecular glue provided crucial new details into the molecular underpinnings of A3G restriction. We demonstrated that Vif interactions with RNA are essential for Vif-mediated degradation of A3G

and are evolutionarily conserved. This implies that RNA has shaped Vif evolution, just as A3G from different species has. Furthermore, we comprehensively characterized the mutational landscape of two different clade B vif genes, particularly highlighting the role that epistasis plays in strain-specific constraints and adaptive potential. We identified several sites that may act as hot spots of adaptive evolution, as well as individual variants conferring improved antagonism of A3G.

5.1 Ancient Origins of Antiviral Defense Genes

Before the evolution of the interferon system, ancient antiviral genes employed a multifaceted approach to combat viral infections. These mechanisms included direct inhibition of viral replication, recognition and destruction of viral components, and activation of cellular pathways to eliminate infected cells (1). Together, these ancient defense strategies laid the foundation for the sophisticated immune responses seen in modern organisms. Given our findings indicating that Mx predates the interferon system, it remains uncertain whether the function of these Mx genes was always in immune defense or was co-opted for this purpose in certain lineages.

This discovery encourages interesting questions inspired by work on other ISGs that predate interferon, such as STING. Characterization of STING genes from different species revealed that the autophagy induction, but not the interferon-induction, function of vertebrate STING was conserved in Nematostella and Xenopus lineages (2). Additionally, work on diverse STING homologs from diverse metazoan species highlights the ancestral ability to bind cyclic dinucloetides (3). Collectively, the functional analysis of STING genes across species has shed light on their modern antiviral roles, including the conserved function of STING in autophagy,

which may represent an ancient antiviral defense mechanism that predates the interferon system. Additionally, the ability to bind cyclic dinucleotides suggests a fundamental role in pathogen detection and response, illustrating how these ancient functions have been adapted and refined in modern vertebrates to enhance antiviral immunity (4).

Similarly, the role of Viperin in antiviral defense has been significantly elucidated, revealing its ancient and conserved functions across diverse species, while also raising questions about how pathogens may have evolved to exploit these ancient genes.. Viperin can inhibit viral replication by disrupting the formation of lipid droplets and altering cellular metabolism, thereby creating an inhospitable environment for viral replication (5, 6). Additionally, Viperin's ability to interact with various cellular and viral proteins, such as NS5A of hepatitis C virus, underscores its versatility in counteracting different viral strategies (6) Pathogens may exploit these physiological functions to gain selective advantages. For example, Viperin's mitochondrial localization in adipose tissue during thermogenesis and its similar requirement for Human Cytomegalovirus (HCMV) viral replication suggest a broader role (6).

With the newfound knowledge of Mx conservation across evolution, further investigation into the role of these ancient Mx genes is needed, which could provide critical insight into the broad-acting antiviral activity of vertebrate Mx. Such studies could reveal whether the antiviral mechanisms employed by modern Mx proteins were present in their ancient counterparts or if these mechanisms evolved later in response to specific viral pressures. To achieve this, experimental approaches could include expressing ancient Mx genes in model organisms or in cell culture to assess their antiviral activity against a spectrum of viruses. Additionally, comparative functional studies of Mx proteins from representative species, including those in which Mx genes have been lost in closely related lineages, could provide valuable insights into

the evolutionary trade-offs associated with Mx-mediated antiviral defense. These experiments could help determine how different selective pressures have shaped the evolution of Mx proteins, revealing the functional diversity among Mx paralogs and the evolutionary pathways that have led to their current roles in antiviral defense.

5.2 Functional and Structural Analysis of Vif-A3G interactions

Our DMS analyses validate that A3G antagonism imposes significant constraints on Vif evolution. Most Vif variants, particularly those involved in binding A3G and the cofactor CBFβ, exhibited fitness disadvantages. However, we identified specific Vif variants with higher fitness, indicating potential adaptive mutations that could enhance viral replication despite the presence of A3G. Among all sites in the arms-race interface, site 83 in particular stood out as a site with multiple variants improving A3G restriction. Interestingly, despite our fundamental understanding of the evolutionary change from tyrosine (Y) at the equivalent of position 83 in SIVrcm Vif being crucial to gaining full antagonism of hominid A3G, HIV-2 Vif, which is the result of a different cross-species transmission event, retains a Y at position 83. While the viral precursors to HIV-1 infected other species of hominids (chimpanzees and gorillas) before spillover into humans, HIV-2 is the result of a cross-species transmission event from a virus that infects Old World monkeys. The retention of Y83 in HIV-2 Vif is noteworthy because it suggests that the evolutionary pressures and adaptive landscapes for HIV-2 Vif might be distinct from those of HIV-1 Vif, indicating that there is more than one way to evolve the ability to counteract A3G.

Another layer of complexity in the evolutionary arms race between Vif and A3 proteins is the necessity for Vif to simultaneously antagonize multiple A3 proteins. This requirement imposes additional constraints on Vif evolution, as changes that enhance the antagonism of one

A3 protein might reduce the effectiveness against others. Future work could leverage DMS libraries to weigh the constraints of individual A3s versus tandem A3s. By testing Vif variants against a combination of A3 proteins, we can potentially identify mutations that confer broad-spectrum antagonism. Additionally, testing these DMS variants against different species' A3 proteins could uncover potential new binding sites on A3G and other A3 proteins outside of the already characterized arms-race interface. Furthermore, it remains an outstanding question if A3-mediated mutations can result in mutations that improve Vif fitness. Consequently, there may be an evolutionary incentive to acquire sublethal mutations leading to less efficient A3Grestriction, facilitating the acquisition of more A3-mediated mutations.

HIV-1 crossed into the human population from chimpanzees approximately 100 years ago, around the early 20th century (7). In evolutionary terms, this is considered young, especially when compared to the millions of years over which primate lentiviruses have been co-evolving with their hosts (8). Therefore, continued characterization of HIV Vif proteins using DMS guided approaches would not only reveal the evolutionary flexibility of Vif, but also provide a better understanding of the evolutionary potential of this protein in its adaptations to the human innate immune proteins Further research using humanized mouse models or Simian-Human Immunodeficiency Viruses (SHIVs) could validate variants of interest in more physiologically relevant settings and investigate the acquisition of additional mutations, including those imposed by A3 proteins, to optimize interactions in the context of these variants. Overall, these observations highlight the complexity of the evolutionary arms race between A3G and Vif. The interplay between mutational resilience, multifunctionality, and species-specific adaptations underscores the dynamic nature of this conflict.

5.3 References

1. J. L. Imler, H. Cai, C. Meignin, N. Martins, Evolutionary immunology to explore original antiviral strategies. Philos Trans R Soc Lond B Biol Sci 379, 20230068 (2024).
2. X. Gui et al., Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. Nature 567, 262-266 (2019).
3. P. J. Kranzusch et al., Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. Mol Cell 59, 891-903 (2015).
4. S. R. Margolis, S. C. Wilson, R. E. Vance, Evolutionary Origins of cGAS-STING Signaling. Trends Immunol 38, 733-743 (2017).
5. L. P. Villarreal, Viral ancestors of antiviral systems. Viruses 3, 1933-1958 (2011).
6. E. E. Rivera-Serrano et al., Viperin Reveals Its True Function. Annu Rev Virol 7, 421-446 (2020).
7. P. M. Sharp, B. H. Hahn, The evolution of HIV-1 and the origin of AIDS. Philos Trans R Soc Lond B Biol Sci 365, 2487-2494 (2010).
8. A. A. Compton, H. S. Malik, M. Emerman, Host gene evolution traces the evolutionary history of ancient primate lentiviruses. Philos Trans R Soc Lond B Biol Sci 368, 20120496 (2013).