STUDIES ON THE MECHANISMS OF ANTI-HIV-1 FUNCTIONS OF APOBEC3F AND APOBEC3G

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Microbiology and Immunology

University of Saskatchewan

Saskatoon

By

Anjuman Ara

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor who supervised my thesis work.

Dr. Linda Chelico

Department of Microbiology and Immunology

College of Medicine

In her absence, permission may be granted by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Microbiology and Immunology

College of Medicine

University of Saskatchewan, 107, Wiggins Road, Saskatoon, Saskatchewan, CANADA, S7N 5E5

ABSTRACT

The seven human cytidine deaminases in the APOBEC3 (A3) family deaminate cytosine in single-stranded DNA to form uracil. The enzymes recognize specific di- and tri-nucleotide sequences and deaminate cytosines within them. The A3 proteins are potent antiviral restriction factors capable of inhibiting retrotransposons and both exogenous and endogenous retroviruses. The overall goal of my Ph.D. research was to biochemically characterize the A3 enzymes, APOBEC3F (A3F) and APOBEC3G (A3G) and identify the biochemical determinants of their anti-HIV-1 function. I characterized these enzymes alone and how they may act in concert to restrict HIV-1 replication. In order to inhibit HIV-1, A3 enzymes must become encapsidated into budding virions and upon infection of the next target cell, the enzymes can deaminate cytosines in HIV-1 single-stranded DNA generated during reverse transcription. The promutagenic uracils formed act as a template for second strand synthesis and result in numerous transition mutations in the double-stranded proviral DNA. These mutations inactivate the virus.

To understand how these deaminations take place, I characterized A3F and A3G biochemically. I found that like A3G, A3F is a processive enzyme that can deaminate at least two cytosines in a single enzyme-substrate encounter. Processivity is achieved through diffusional mechanisms termed sliding, jumping, and intersegmental transfer. Unlike A3G, which scans ssDNA using both sliding and jumping movements, A3F solely relies on jumping movements. Further, A3F jumping movements are distinct from A3G. We discovered that a ^{190}NPM^{192} motif in A3F prevents its sliding movement since insertion of ^{195}NPM^{197} into A3G decreased its sliding movements. Our data demonstrated that A3G is a more potent inhibitor of HIV-1 owing primarily to its unique DNA scanning mechanism and secondly to its deamination motif specificity. The data support a model in which the processive DNA scanning mechanism of an A3 enzyme can predict its mutagenic potential.

Since A3F and A3G are coexpressed in the CD4$^+$ T cells that HIV-1 infects, we undertook a study to determine if A3F and A3G were coencapsidated and could be concurrently deaminating viral DNA. First, we found that an A3F/A3G hetero-oligomer can form in cells and *in vitro*, in the absence of RNA. This hetero-oligomer has unique biochemical properties and more efficiently deaminates cytosines compared to each A3 alone. Namely, the A3F in the A3F/A3G hetero-oligomer enhances A3G-mediated deamination. Moreover, A3F and A3F/A3G caused the accumulation of shorter reverse transcripts due to decreasing the reverse transcriptase efficiency,

which would leave single-stranded (-)DNA exposed for longer periods of time enabling more deamination events to occur.

Overall my thesis research identified and characterized the mechanisms by which A3F, A3G, and A3F/A3G hetero-oligomer act as inhibitors of HIV-1. Future studies on whether hetero-oligomers of other A3s involving A3D, A3F, A3G, and A3H will be very interesting.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Linda Chelico for her excellent supervision and support and for inculcating a strong and rigorous work ethic in doing science. Thank you for your guidance and inspiration. It was a great opportunity to work with you. I will never forget what I have learned over these past few years. I have grown as a scientist through her encouragement, confidence, and high standards.

Secondly, I would like to express overwhelming thanks to my committee members, Dr. Kerri Kobryn, Dr. Joyce Wilson, Dr. Yan Zhou, and Dr. Sidney Hayes, for their valuable advice and constant support throughout my Ph.D. study. I am thankful to my committee for their inputs over the years which proved essential to succeed at a decidedly demanding project as well as my personal development as a scientist.

A sincere and heartfelt thanks to Dr. Aaron White (VIDO), for giving me an opportunity to do a summer project. His lab was the place where I started my scientific journey in Canada.

I am grateful for the camaraderie of my colleagues at Chelico lab. The good spirits of my fellow lab-mates made joining the Chelico lab in its infancy much easier. In particular, I am honored to have shared my time in the lab with Muhammad, Tyson, Robin, Nazanin, Madison, Cate, Mariam, Jonathan, Lily, and Tayyba who have helped to shape the lab environment. It has been a pleasure working with each of them. Especially, I would like to thank Robin for providing me technical help, kind words of wisdom, and keeping the lab running smoothly. Without Robin's kind assistance, this project would not have been executed in its present form. I also want to thank Cate, Madison, and Tayyba for stimulating discussions and their unwavering help.

I am grateful to the College of Medicine for providing fellowships and CHIR, NSERC and CFI for providing research funds.

And last but certainly not least I would like to thank my family for their everlasting support and belief in me and my work. I am also indebted to my daughters Alina and Jannat whose innocent demands were sometimes ignored due to spending more time in lab. I am lucky to be their mom. They supported me to discover my passion (science). I must acknowledge the patience I learned from my husband K.A. Ahmed, not only in day to day life but also in the scientific field. Thanks and acknowledgement are not enough for his patience, his sense of humor, his quiet support, and his love. I appreciative, every day, for having you by my side and for all you have done, and still do, to make life full of promise and joy.

DEDICATION

Lovingly dedicated to

my daughters
Alina and Jannat
&
my husband
Ashfaque

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A3F	APOBEC3F
A3G	APOBEC3G
AID	activation-induced cytidine deaminase
CTD	C-terminal domain
Cul5	cullin5
dNTP	deoxynucleoside triphosphate
EIAV	equine infectious anemia virus
ELISA	enzyme linked immunosorbent assay
FACS	fluorescence- activated cell sorter
GFP	green fluorescent protein
gRNA	genomic RNA
HAART	highly active antiretroviral therapy
HIV	human immunodeficiency virus
IFN	interferon
IN	integrase
IP	immunoprecipitation
IRES	internal ribosome entry site
LINEs	long interspersed elements
LRT	late reverse transcripts
LTR	long terminal repeat
NC	nucleocapsid
Nef	negative factor
NTD	N-terminal domain
ORF	open reading frames
PBMC	peripheral blood mononuclear cell
PBS	primer binding site/phosphate buffered saline
PIC	pre-integration complex
PPT	polypurine tract
PR	protease

Rev	regulator of expression of virion protein
RNA	ribonucleic acid
RNase H	ribonuclease H
RRE	rev response element
RT	reverse transcriptase
RTC	reverse transcription complex
SAMHD1	sterile alpha motif and histidine-aspartic domain 1
SINEs	short interspersed elements
SIV	simian immunodeficiency virus
TAR	Tat-acting region (transactivation response element)
Tat	transactivator
TM	transmembrane
TRIM	tripartite motif
Vif	virion infectivity factor
VLP	viral like particle
Vpr	virion protein R
Vpu	virion protein U

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

1.1. Introduction

The unifying function of the APOBEC enzyme family is that they are single-stranded polynucleotide cytidine deaminases and use this activity to either form uracil in mRNA to edit and modify function or uracil in single-stranded (ss) DNA as a promutagenic lesion to enable gene evolution or inactivation. In RNA, uracil has a coding function and as a result cytosine deamination is referred to as editing. In contrast, in ssDNA, uracil is a promutagenic lesion since if left unrepaired, it will template addition of adenine at a site that should have incorporated a guanine, thus creating C/G→T/A transition mutations. APOBEC1 (A1), an mRNA editing enzyme, is the namesake of the enzyme family, has anti-retroelement activity, but primarily acts in lipid metabolism by editing the mRNA of apolipoprotein B in gastrointestinal cells to create a smaller protein than the apolipoprotein B expressed in the liver (*1-3*). APOBEC2 (A2), whose expression is restricted to heart and skeletal muscles, is important for normal muscle development and function (*4*). Deamination by an APOBEC protein called activation-induced cytidine deaminase (AID), a lymphoid-specific ssDNA deaminase, is responsible for B cell maturation and immunoglobulin-gene diversification (*5*). The discovery of A3G in 2002 (*6*) and other cytidine deaminases (APOBEC3A-H, excluding E) shortly thereafter opened-up the field of intrinsic immunity by host encoded viral "restriction factors" (*7-10*).

Humans defend themselves against invading pathogens using innate and adaptive immune systems. The innate immune system is the first line of defense that protects the host from invading pathogens and is comprised of soluble factors like cytokines, complement proteins and various cellular factors, such as mast cells, granulocytes, natural killer cells, macrophages, and dendritic cells (*11*). A subcategory of innate immunity, called intrinsic immunity, produces specific anti-viral proteins called "restriction factors" to counteract viral replication. The members of the

cytidine deaminase family, apolipoprotein B mRNA-editing, catalytic enzymes (APOBEC), are restriction factors that provide overlapping protection against exogenous and endogenous pathogens with a single-stranded DNA intermediate, such as retroviruses and retrotransposons (*12, 13*). The APOBEC field came into the limelight after a discovery that implicated one of the family member, APOBEC3G (A3G) as a host restriction factor against Human Immunodeficiency Virus 1 (HIV-1) that causes AIDS (Acquired Immunodeficiency Syndrome) (*6*).

Out of seven A3 enzymes only four A3D, A3F, A3G and A3H, suppress HIV-1 replication. These four enzymes function by packaging into newly synthesized viral particles, which enables them to deaminate cytosines to form promutagenic uracils on nascent cDNA and directly interfere with the reverse transcription step of viral replication (*12, 13*). The end result is a hypermutated and nonfunctional virus. To counteract the deleterious effects of these enzymes, HIV encodes a protein called viral infectivity factor (Vif) that induces their polyubiquitination and proteasomal degradation (*14-16*). A3D, A3F, A3G and A3H are coexpressed in CD4+ T cells. Each are independently able to inhibit HIV-1 infection (*17-20*). A3G has been the most widely studied and appears to be the most effective restrictor of HIV-1 replication as measured by cell-based assays (*21-24*) and indicated from clinical studies (*25-30*). Clinical subjects found to have an inherent ability to express a high level of A3G are less likely to become infected with HIV-1 or convert from HIV-1 to AIDS (*25-32*). The proviruses recovered from HIV-1 patients often carry C/G→T/A mutations in an A3G preferred sequence (CCC) context, which supports the notion that cytosine deaminations in the HIV-1 genome by A3G occur most frequently (*25-27*). At a population level, these actions of A3G are sufficiently repressed by Vif so that the virus continues to replicate, although the defective viruses that build up in a person during an HIV-1 infection are in part due to A3-induced mutations (*33*).

1.2. Endogenous retroelements, endogenous retroviruses and exogenous retroviruses

Retroelements are transposable genetic elements that have integrated into mammalian genomes over millions of years (*34*). Retroelements replicate via RNA intermediates. Depending on the presence or absence of long terminal repeats (LTRs), they are divided into two groups, LTR retrotransposons (also known as endogenous retroviruses) and non-LTR retrotransposons (*35*).

Non-LTR retrotransposons comprise the majority of transposable elements in the human genome. About one third of the human genome is comprised of non-LTR retrotransposons. Non-LTR retrotransposons have been further grouped into three types; long interspersed elements

(LINEs), short interspersed elements (SINEs) and the composite hominid-specific retrotransposons. There are three types of non LTR transposable elements currently active in human genome. These are LINE-1, Alu and SINE-VNTR-Alu (SVA) (*36, 37*).

LTR retrotransposons (endogenous retroviruses) constitute about 10% of human and mice genomes. Owing to the accumulation of mutations, endogenous retroviruses have become largely inactive in humans (*38, 39*). Reconstruction of endogenous retroviruses has led to the hypothesis that they were inactivated by A3 enzyme activity (*40*). When active, endogenous retroviruses could fully bud and reinfect other cells within the same host. In particular, if endogenous retroviruses infect germ cells, virus could be vertically transferred to the offspring.

Exogenous retroviruses have a similar lifecycle to endogenous retroviruses except that they are transferred through horizontal, rather than vertical transmission. Exogenous retroviruses were originally identified as RNA tumor viruses, causing tumors in mice, birds and sheep (*41-44*). However, after discovery of reverse transcriptase, "they" were called retroviruses (*44*). Retroviruses are a family of single-strand (+) RNA viruses that reverse transcribe the genome (7-12 kb in length) into a double-stranded (ds) DNA which can then integrate into the genome of the host cell. All retroviruses contain two copies of genomic RNA (gRNA) that become non-covalently linked with their 5' ends in the mature virion (*44, 45*). Retroviruses were classified into the family *Retroviridae*. The Retroviridae family is divided into three main groups, (i) *Oncoviridae* or oncoviruses (tumour causing virus), (ii) *Spumaviridae* or spuma- or foamy viruses and (iii) *Lentiviridae* or lentiviruses. The most complex and widely studied retroviruses (*Lentiviridae)* are, HIV-1 and HIV-2 in humans and simian immunodeficiency virus (SIV) in simian species. The name lentivirus was derived from the Latin word "lentus" meaning slow, based on the slow progressive nature of the infection (*46*).

1.3. History of HIV

The retrovirus HIV, was identified in 1983 as the causative agent of Acquired Immune Deficiency Syndrome (AIDS), which was recognized as a new disease in 1981 when many homosexual men became susceptible to unusual rare malignancies and opportunistic infections (*47-49*). The reasons for HIV-1 emergence, spread, and unique pathogenicity have been a subject of intense investigation. A first clue regarding the origin came in 1986 when a morphologically related but antigenically different virus, now termed HIV-2, was found to be closely related to a simian virus causing immunodeficiency in captive macaques (*50, 51*). HIV-2 was found to cause

AIDS in patients in western Africa (*52*). Other, simian viruses were discovered in primates from sub-Saharan Africa, including, e.g, sooty mangabeys, African green monkeys, mandrills, and chimpanzees. These viruses were collectively termed simian immunodeficiency viruses (SIVs) with a suffix to denote their species of origin (*47*). Interestingly, it was found that SIVs in chimpanzees were more closely related to HIV-1 (*53*), whereas close simian relatives of HIV-2 were found in sooty mangabeys (*54*). These findings provided the first proof for the cross-species lentivirus infections in primate species that led to the emergence of AIDS in both humans and macaques (Figure 1.1) (*47*). Indeed, subsequent studies confirmed that HIV-1 and HIV-2 originated in Africa as result of zoonotic transfers of viruses infecting primates (*55*).

HIV-1 strains have been classified into four highly divergent groups based on the sequences of full-length viral genomes. HIV-1 strains are M (main), O (outlier), N (non M or O) and P (putative) (*56, 57*). Out of 4 groups, Group M represents the major group, which is responsible for more than 95% of HIV-1 infections in the world (*58*). The prevalence of other groups is extremely low and mostly restricted to West Central Africa (*59-61*). Group M is relatively evenly distributed worldwide and has been further divided into nine phylogenetic subtypes or clades, named A, B, C, D, F, G, H, J, and K. Different clades within several infected patients can create circulating recombinant forms (CRFs) by genetic recombination (*62, 63*). In the Americas and Europe, clade B is predominant, but it is rarely found in Africa (*64*). The clade C is the most prevalent HIV-1 subtype causing 50% of the infections worldwide, followed by clade A (12%), clade B (10%), CRF02-AG (5%) and CRF01-AE (4.8%) (*63, 65*).

HIV-2 is also very divergent and is classified into eight different groups (A to H). Unlike HIV-1, HIV-2 is largely geographically restricted to West Africa, but recently has been identified in Europe and India (*66, 67*). HIV-2 shares 50-60% nucleotide sequence identity with HIV-1 (*51*), but is genetically most closely related to the SIVsmm from sooty mangabey, whereas HIV-1 is most closely related to the SIVcpz from chimpanzee (*67*). Compared to HIV-1 infected patients, AIDS develops more slowly after HIV-2 infection, and the patients survive longer (*68, 69*). Furthermore, the transmissibility of HIV-2 is significantly lower than that of HIV-1 (*70, 71*). Since my research did not involve HIV-2, this literature review will mainly discuss HIV-1 henceforward.

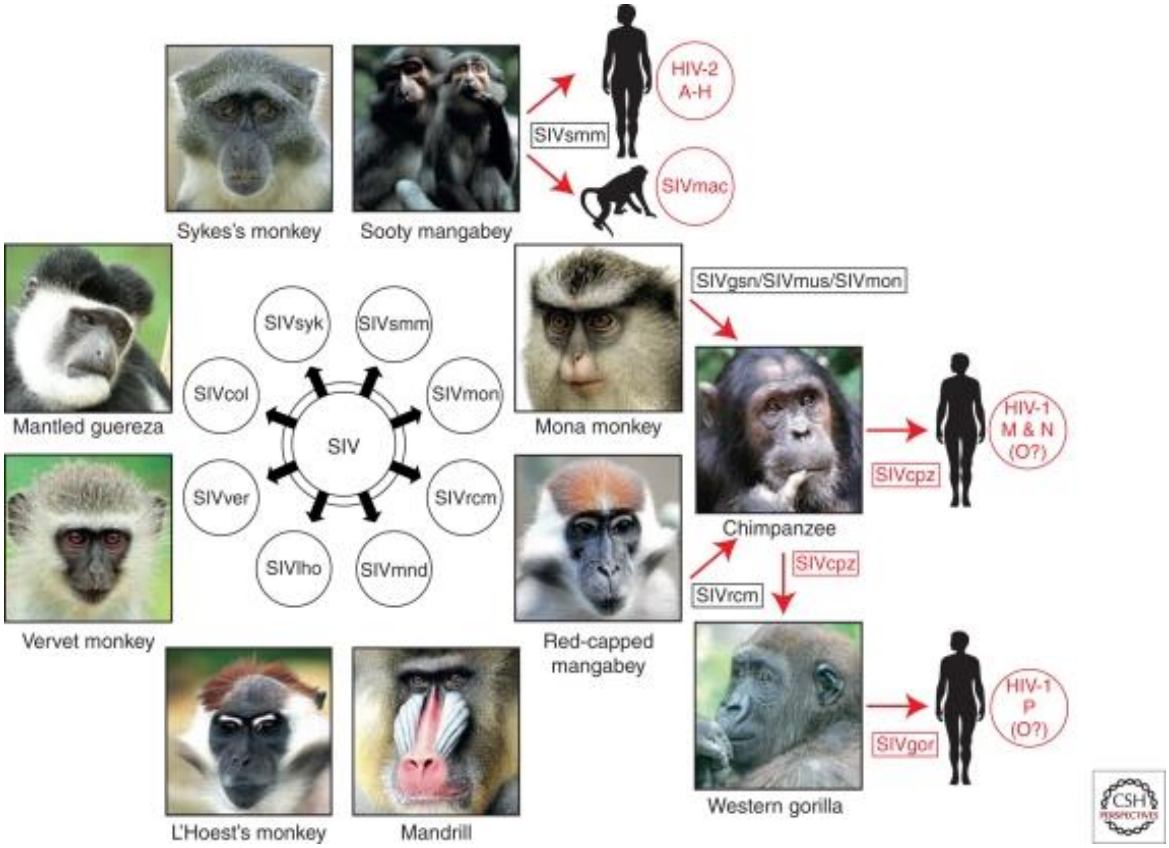

Figure 1.1. Origins of HIV viruses . More than 40 different lentiviruses termed as simian immunodeficiency viruses (SIVs) naturally infect Old World monkeys. The suffix listed denotes their primate species of origin (e.g., SIVsmm from sooty mangabeys). Humans were infected with SIV through the cross-species transmission. Zoonotic transfers of SIVs from Sooty mangabeys, Chimpanzees, and the Western gorilla has resulted in the generation of HIV-2 (A-H), HIV-1 (M & N) and HIV-1 (P) types, respectively. Reprinted with permission from (*47*).

1.4. Genome organization of HIV-1

The HIV-1 genome is a 9.2 kb positive-stranded RNA molecule. It contains nine genes that encode fifteen viral proteins (Figure 1.2). Three of the ORFs encode Gag, Pol and Env polyproteins, which are proteolytically processed into individual proteins. Gag, which is also referred to as p55, is cleaved by protease (PR) to generate structural proteins Matrix (p17), Capsid (p24), Nucleocapsid (p7), and p6 protein. The Gag-pol precursor is proteolytically processed into three enzymes PR, RT and integrase (IN). The Env precursor gp160 is cleaved by the host endoprotease furin into gp120 and gp41 (*72*). There are also two regulatory protein genes, transactivator (Tat) and regulator of expression of virion protein (Rev) and accessory proteins which are virion infectivity factor (Vif), viral proteins R (Vpr), viral proteins U (Vpu) and negative factor (Nef). These accessory proteins can exhibit multiple functions and interact with various human proteins during the viral life cycle (*73*). In addition, the HIV-1 RNA genome forms several secondary structures, including the, packaging signal (Ψ), pseudoknots, transfer RNA mimics, ribosomal frameshift motifs, *cis*-regulatory elements, Tat-acting region (TAR), primer binding site (PBS), dimer initiation site (DIS), polypurine tract (PPT), and an internal ribosome entry site (IRES) (*74*). Although HIV-1 has an IRES, the cap dependent translation is thought to be the main translation mechanism used (*75*). These RNA structures regulate viral replication, from reverse transcription initiation and reading frames manipulation to RNA nuclear export and viral RNA packaging.

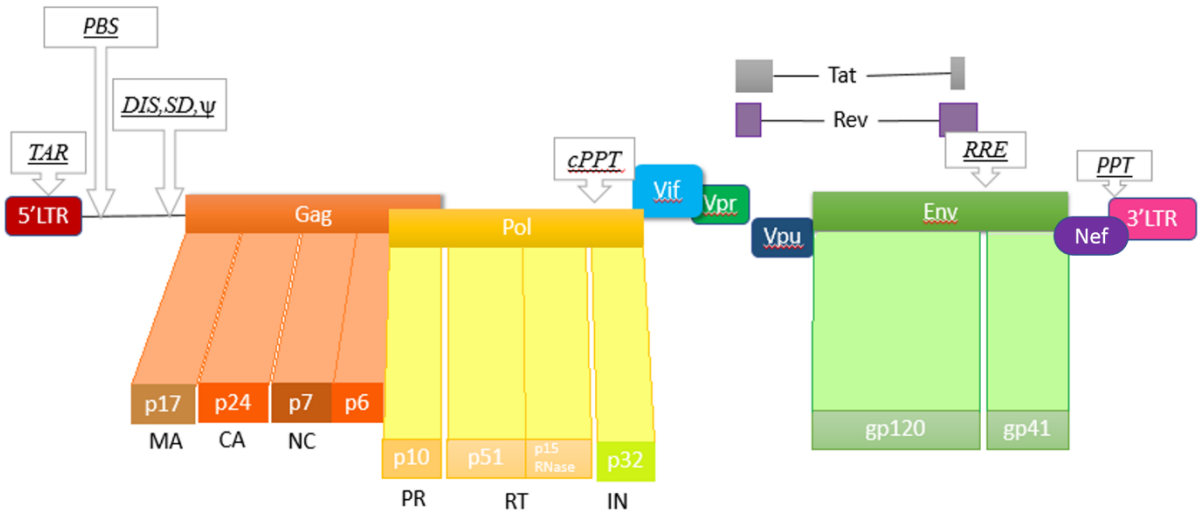

Figure 1.2. Schematic model of the HIV-1 full-length genome . HIV-1 genome contains 9 ORFs flanked by the long terminal repeats (LTRs) and several structural landmarks (italic characters). Gag and Gag-pol polyproteins are processed by viral protease. Gag is processed into matrix (MA), capsid (CA), nucleocapsid (NC), and p6. Gag-pol is processed into protease (PR), reverse transcriptase (RT), and integrase (IN). the envelope (Env) is cleaved by cellular proteases into gp120 and gp41. Landmarks include Tat-acting region (TAR), primer binding site (PBS), dimer initiation site (DIS), splice donor (SD) and central polypurine tract (cPPT), rev response element (RRE) and PPT. Modified from Guangdi *et al*. (*76*).

1.5. HIV-1 virion structure

The HIV-1 virion is approximately 100 to 120 nm in diameter with a roughly spherical shape as shown in Figure 1.3. The capsid is coated with a lipid envelope, which contains cell-derived lipid bilayers and viral envelope glycoproteins (Envs) (*77*). On the surface of the mature virion, Envs are displayed in knobbed spike structures, formed from surface protein gp120 and transmembrane protein gp 41. These two glycoproteins noncovalently bind to each other and form heterodimers (*78*). The central core of the virion consists of the Gag structural proteins: p17, p24, and p7. The matrix (MA) protein p17 forms the inner surface of the viral envelope; the capsid (CA) protein p24 forms the cone-shaped core; and the nucleocapsid (NC) protein p7 is located within the core, as a binding partner of viral RNA (*79*). Inside the core, there are two copies of the single-stranded RNA genome, which is associated with viral enzymes and accessory proteins (*80*). In addition, cellular molecules such as tRNAs, annexins, actin, and A3's from the infected cells are incorporated into virions as bystanders or through interacting with viral proteins or viral RNA (*6, 81-86*).

1.6. Viral proteins

1.6.1. Viral structural proteins

Gag is a 55 kDa precursor protein that is cleaved to produce the mature proteins: MA (P17), CA (P24), p2, NC (P7), p1 and p6 protein. The p2 and p1 proteins are spacer peptides and do not have any known function (*87*)

MA is a 17 kDa structural protein that is post-translationally myristoylated at the N-terminus (*88*). During viral assembly, MA, which is at the N-terminal domain of Gag, regulates the Gag intracellular location, and directs Gag to the plasma membrane (*88*). It also interacts with the gp41 cytoplasmic tail and promotes incorporation of Env into the virion (*89, 90*). Additionally, it contributes to nuclear migration, nuclear import and integration as a part of the reverse transcription complex (RTC) and pre-integration complex (PIC) (*91-93*).

NC is a 7 kDa nucleic acid chaperone protein that is able to promote nucleic acid rearrangement during viral assembly and replication, by binding to genomic RNA. NC enables selection, dimerization and encapsidation of genomic RNA into virions (*94-96*). NC also facilitates nucleic acid structural reorganization, primer annealing, and strand transfers which are required for efficient reverse transcription (*97, 98*).

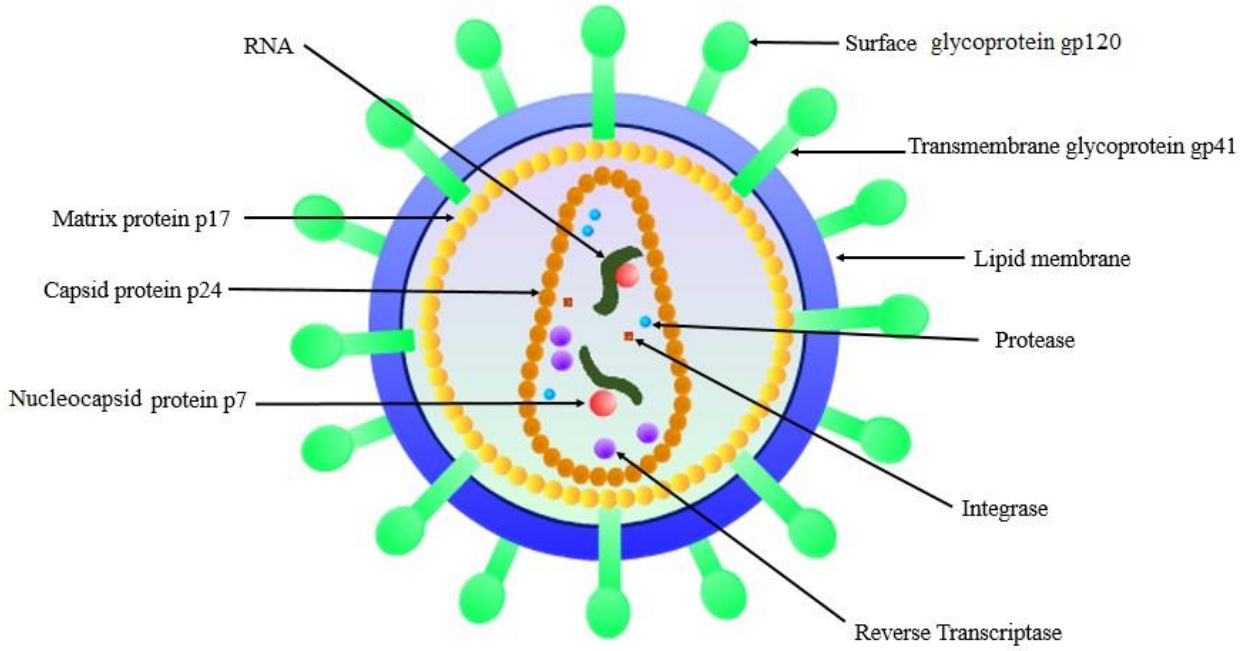

Figure 1.3. Morphologic structure of HIV-1. Lipoprotein membrane enveloped HIV-1 virus. Env glycoprotein complexes composed of gp120 and gp41 are integrated into this lipid membrane. The inner surface of viral envelope remains anchored with the matrix protein. Capsid makes the cone-shaped core. Inside the viral core, nucleocapsid is associated with two copies of HIV-1 RNA genome. Other viral enzymes like reverse transcriptase, integrase, and protease are also present in virus particles.

P6 Gag is located at the C terminal domain of Gag and regulates the final step of virus budding. It promotes separation of assembled virions from the plasma membrane through interaction with the endosomal sorting complex required for transport (ESCRT) (*99*). It is also required for the incorporation of Vpr into virions (*100, 101*).

Env encodes the polyprotein precursor gp160 that is cleaved by the cellular endoprotease furin into two subunits gp120 or SU (surface unit) and gp41 or TM (transmembrane). These two subunits non-covalently bind to each other and form the gp120/gp41 complex. HIV-1 Env mediates viral entry into target cells, in which the gp120 subunit is responsible for cell receptor (CD4) and coreceptor (CCR5 or CXCR4) binding, and gp41 embeds in the viral envelope to facilities the virus to attach to and fuse with target cell membranes (*77, 102*).

1.6.2. Viral enzymes

The Pol protein of HIV-1 is composed of three viral enzymes vital for viral replication; PR, RT, and IN (*103*). Pol is expressed as a 160 kDa protein, and the ORF partially overlaps with Gag resulting in the formation of Gag-Pol polyprotein through -1 ribosomal frameshifting (*104*). After viral release, viral PR auto-activates and cleaves the Gag-Pol polyprotein generating MA, CA, spacer peptide p2, NC, PR (p10), RT (p66/p51) and IN (p32) (*105*).

HIV-1 PR functions as a homodimer. PR is classified as an aspartic protease, since it has the sequence Asp-Thr-Gly, which is conserved in several cellular aspartic proteases (*106*). PR is known for cleaving viral precursor proteins, however HIV-1 PR also cleaves host cell proteins, such as anti-apoptotic protein Bcl-2, cytoskeletal proteins, and thereby contributes to necrotic or apoptotic cell-death (*107, 108*).

HIV-1 RT is a DNA- and RNA-dependent DNA polymerase with an RNaseH domain. The RT is capable of strand displacement synthesis and template switching (*109, 110*). RT is a heterodimer consisting of two polypeptides p66 (66 kDa) and p51 (51 kDa). RT requires hetero-dimerization for its activity. The larger subunit p66 exhibits both DNA polymerase and RNase H activities. The smaller subunit p51 is catalytically inactive and plays a primarily structural role (*111*). The p51 subunit is produced by the removal of the C-terminal RNase H domain from p66 by PR cleavage (*112*).

HIV-1 integrase (IN) functions in a tetrameric form and facilitates the integration of viral DNA into the host genome. IN is a 32 kDa protein encoded by the end of the *pol* gene. Integrase facilitates 3′ processing and strand transfer leading to the integration of processed viral DNA ends

into the host genomic DNA (*113*). Besides an enzymatic function during the integration, HIV-1 IN has additional non-enzymatic roles at steps prior to viral integration such as reverse transcription, nuclear import, polyprotein processing, assembly, and maturation (*114-116*).

1.6.3. Viral regulatory proteins

HIV-1 Rev is a 19 kDa phosphoprotein localized predominantly in the nucleolus/nucleus (*117*). Rev is a functionally conserved regulatory protein of lentiviruses and frequently cycles between the nucleus and cytoplasm. During the late stages of viral replication, Rev promotes the nuclear export of singly spliced or unspliced viral mRNAs by binding to their RRE (*118*). Stability, translation, and encapsidation of Rev-bound viral transcripts is also affected by Rev (*119*). HIV-1 Rev also regulates viral integration frequency by binding to integrase and LEDGF/p75 (*120*).

HIV-1 Tat, a 12 kDa protein, is the transactivator of HIV gene expression. Tat is localized primarily in the nucleus. Tat is one of the two viral regulatory factors essential for HIV-1 transcription from the viral 5'LTR promoter. Tat binds to a trans-activation response (TAR) RNA element, and activates transcription initiation and elongation from the 5'LTR promoter, preventing transcription termination by the 5'LTR polyadenylation signal (*121*). Tat is secreted by infected cells and acts as an extra-cellular toxin (*122*). Exogenous Tat can induce certain cytokines, neurotoxicity in the central nervous system and programmed cell-death in uninfected bystander $CD4^+$ T cells (*73, 123*).

1.6.4. Viral accessory proteins

HIV-1 Vif is a 23 kDa protein that promotes the viral infectivity. During the assembly of viral particles, Vif binds to viral genomic RNA and the NC domain of Pr55-Gag, and functions as a temporal regulator of viral assembly (*124*). Vif counteracts the antiviral activity of A3 proteins of the infected cells by inducing their proteasomal degradation in the cytoplasm. This will be discussed further in section 1.10.

HIV-1 Nef is a 27 kDa myristoylated protein, which is expressed at the early stage of viral replication (*125*). Nef induces down-regulation of cell surface receptors (CD4, CXCR4 and CCR5) that promote viral entry into the cells, thus suppressing superinfection (*126, 127*). Nef-mediates down-regulation of major histocompatibility complex class I (MHC-I) on infected cells to facilitate viral evasion from cytotoxic T lymphocyte mediated apoptosis (*128*). Moreover, Nef modulates cytoskeletal remodeling, vesicular transport, and cellular signal transduction pathways for

promoting viral replication (*73*). Recent studies identified that host transmembrane proteins serine incorporator 3 (SERINC3) and SERINC5 incorporate into budding HIV-1 virions and impair subsequent virion penetration of the susceptible target cells but this restriction is counteracted by Nef (*129, 130*). The evidence for an essential role of Nef in viral pathogenesis also comes from the observation that patients infected with Nef-deleted HIV-1 become long-term nonprogressors (*131*), who remain asymptomatic for prolonged length of time even without anti-viral therapy (*132*).

HIV-1 Vpu is a 9 kDa type 1 integral membrane protein. Vpu has at least two main biological functions. First, it degrades newly synthesized CD4 in the endoplasmic reticulum (*133-135*); second, Vpu enhances virion release from the plasma membrane of HIV-1 infected cells (*136*). Vpu counters host restriction factor tetherin (BST-2) mediated inhibition of viral release through β-TrCP and endo-lysosomal trafficking (*137*).

HIV-1 Vpr is a 14 kDa nuclear protein. Vpr participates in the nuclear import of pre-integration complexes (PIC) (*138*). It was suggested that the first α helix structure on Vpr mediates the docking of the PIC to the nuclear pore complex (*139, 140*). Interaction of Vpr with importin α, is essential for nuclear import (*141, 142*). In addition, two non-classical nuclear localizing signals (NLS) of Vpr may also help in the nuclear import of PIC (*143*). Vpr induces G2 cell cycle arrest in infected cells (*144*). Vpr also induces T-cell apoptosis through mitochondria-dependent pathway (*145-147*), transactivates cellular genes, and acts as coactivator to induce translational activation of the HIV-1 5'LTR (*147, 148*).

1.7. HIV-1 life cycle

HIV-1 replication is a complex process regulated by both viral and cellular proteins.

1.7.1. Receptor binding and entry

The first step of the HIV-1 life-cycle begins with the binding and entry into the host cells. HIV-1 attaches to the host cell by binding to the cell surface receptor CD4 via gp120 of Env. CD4 binding prompts a conformational change in gp120 protein, resulting in exposure of sites that enables coreceptor binding (CCR5 or CXCR4). Binding of both CD4 and the coreceptor induces gp41 rearrangement that exposes hydrophobic gp41 fusion peptide, which promotes virus-cell membrane fusion by inserting into the host cell membrane (*149, 150*).

1.7.2. Viral uncoating

Following the insertion of HIV-1 core into the target cell, host deoxynucleotide triphosphates begin to diffuse into the capsid and reverse transcription of the viral RNA begins (*151*). The uncoating process is believed to be regulated by viral and host cellular factors through modulating capsid stability, however the precise location and timing of uncoating remain poorly defined (*152, 153*). Previous studies suggested that viral uncoating takes about 4-8 h to complete (*154*). However, a recent study using multicolored virus and single-particle imaging demonstrated that HIV-1 viral disassembly process takes place mostly in the middle-cytoplasm region and is complete between 60-120 min post-infection (*155*).

1.7.3. Reverse transcription

Figure 1.4 represents a schematic of HIV-1 reverse transcription. HIV-1 uses the cellular $tRNA_{lys3}$, which is hybridized to the genome encoded primer-binding site at the 5' terminus through NC chaperone activity, as the primer for negative-strand DNA synthesis. Extension of the $tRNA_{lys3}$ primer creates the negative strong-stop DNA (*156*). After this is synthesized, NC facilitates a (-) strand transfer which moves the DNA/ $tRNA_{lys3}$ to the 3' end of the genome where it anneals to the complementary R region and enables DNA synthesis to continue. As RNA/DNA hybrid regions are formed the RNaseH activity of RT degrades the genomic RNA, except the two PPT regions at the 3' end and center of the RNA genome. The PPTs are more resistant to RNase H degradation and remain bound to the minus-strand DNA for longer times. This enables them to act as primers for the initiation of plus-strand DNA synthesis (*157*). The synthesis from the PPTs extends through the U3-R-U5 long terminal just formed, and on through the 18 nucleotides of the tRNA that were initially hybridized to the primer binding site on genomic RNA. The short DNA intermediate made is designated as the "plus-strand strong-stop DNA." RNase H activity removes the tRNA from the minus-strand DNA copy and exposes the primer binding site on the plus-strand strong stop DNA. Then the "second strand transfer" takes place. It could involve a circular intermediate, with the plus-strand strong-stop DNA acting as a bridge between the two ends of the minus-strand DNA. Both the minus and plus DNA strands are then extended by reverse transcriptase to the ends of their respective template strands. This results in a linear, double-stranded DNA with long terminal repeats at both ends. This DNA is called proviral DNA (*98, 157*). Reverse transcription critically relies on the nucleic acid chaperone activity of NC, which remodels nucleic acid structures to the most thermodynamically stable conformation (*79, 98, 158*).

NC plays a crucial role in almost all steps of reverse transcription such as $tRNA_{lys3}$ primer annealing and DNA synthesis initiation, minus- and plus-strand transfers, and RT-catalyzed DNA elongation (*79, 98, 158*).

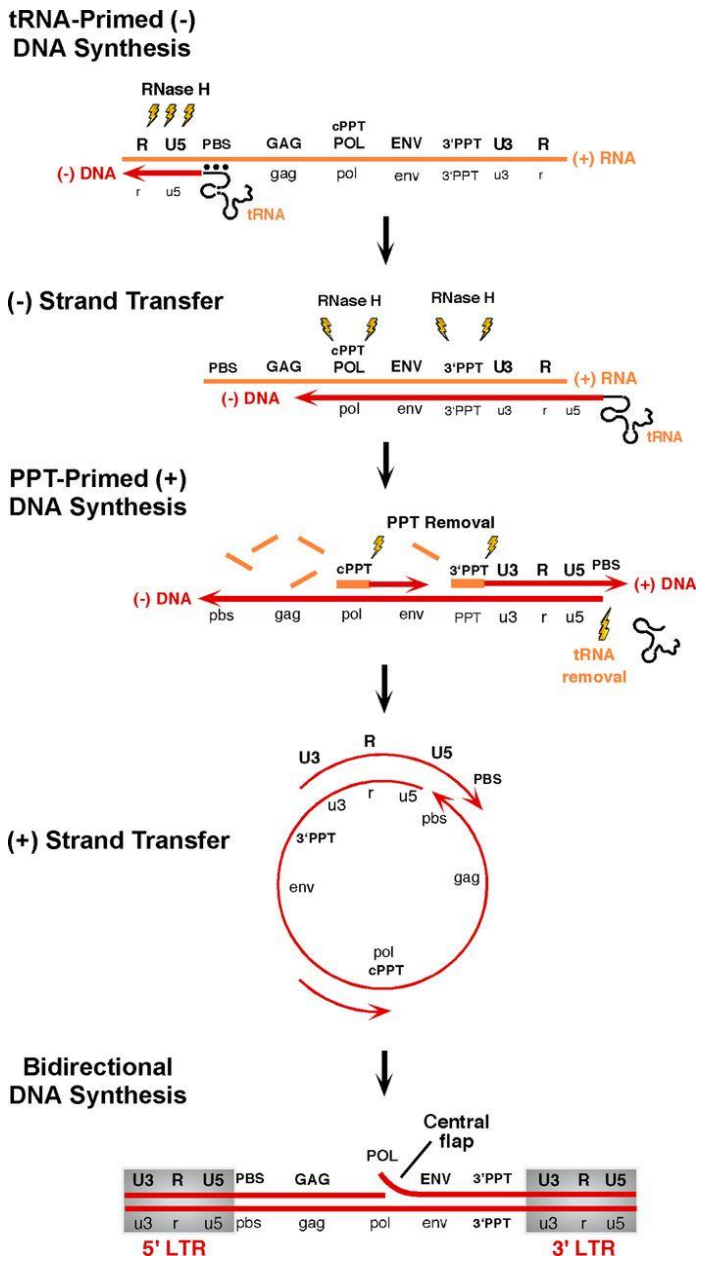

Figure 1.4. HIV Reverse Transcription. For starting reverse transcription, the HIV RNA genome absolutely requires a human $tRNA_{Lys3}$ as a primer to bind to the Primer Binding Site (PBS). HIV reverse transcriptase (RT) starts to reverse transcribe the RNA to form DNA at the 5' end of the RNA genome. The RNase H domain degrades the genomic RNA. Upon reaching the 5' end of the RNA genome, the nucleocapsid (NC) protein mediates a strand transfer of the (-) DNA strand. During (-) DNA elongation, RT degrades the RNA genome except for distinct PPT regions that are resistant to RT-mediated degradation. The (+) DNA strand is synthesized by RT from the RNA-DNA duplexes remaining at the PPT sequences. A circular intermediate is then formed, with the plus-strand strong-stop DNA acting as a bridge between the two ends of the (-) DNA strand. Finally, double-stranded DNA with long terminal repeats at both ends is formed. Reprinted with permission from (*156*).

1.7.4. Nuclear entry and integration

Following the synthesis of double stranded viral DNA, newly synthesized DNA is transported into the nucleus as part of the PIC. With the help of nuclear membrane transport machinery PICs cross the nuclear envelope (*159*). After nuclear entry, the viral DNA integrates into the chromosomal DNA with the help of virus-encoded IN protein. To facilitate viral DNA integration, IN initiates strand transfer by performing 3′ processing, wherein IN recognizes and removes a pGT dinucleotide from each 3′ end of LTR. This process generates a 3′-hydroxyl group at the viral DNA ends, which can then be covalently ligated to the 5′-phosphates of a staggered cut made by IN in chromosomal DNA (*160-162*). Then, cellular DNA repair enzymes repair the DNA gaps created between unpaired viral 5′ ends and chromosomal DNA (*163, 164*). Hence, complete integration of viral DNA into host genome creates a provirus that acts as the precursor for new virion synthesis (*165*).

1.7.5. Transcription

Following the HIV-1 provirus formation, viral mRNA synthesis starts by hijacking cellular RNA polymerase II (*166*). In the beginning, completely spliced mRNAs encoding Tat, Rev and Nef are synthesized. Then, endogenous cellular pathways export newly synthesized mRNA to the cytoplasm. Transcription is actively up-regulated after accumulation of Tat, resulting in the synthesis of incompletely spliced mRNAs encoding Env, Vpu, Vif and Vpr. Then, full-length unspliced transcripts for the Gag-Pol polyprotein is generated (*167, 168*). Rev dependent export pathways transport unspliced and incompletely spliced mRNAs through the nuclear pore to the cytoplasm (*167*).

1.7.6. Virion assembly

After the synthesis of gp160 in the endoplasmic reticulum (ER), gp160 is transported to the cell membrane for virus assembly using the vesicular transport pathway. Gag and Gag-Pol precursor are synthesized on free polyribosomes. HIV-1 virion assembly takes place on the plasma membrane. Gag acts as the "assembly machine" for virion assembly. Gag multimerization drives a number of essential events in virion assembly (*169, 170*), including recruitment of viral genomic RNA, recruitment of protein components, and self-assembly into spherical virion particles. However, post-translationally modified Env glycoproteins traffic independently to the assembly site (*171*).

1.7.7. Virion maturation

In order to promote virion release from the plasma membrane of the infected cells, HIV-1 Gag recruits the ESCRT (endosomal sorting complex required for transport) pathway through a late-assembly domain in p6 Gag. The virion release from the plasma membrane is facilitated by the ESCRT complex mediated membrane fission (*99*). After virion budding, immature virions undergo maturation via viral PR-mediated cleavage of Gag and Gag-Pol polyproteins. Then, the rearrangement of cleaved products creates a fully mature and infectious virion.

1.8. Retroviral restriction factors and their roles in HIV-1 infection

Restriction factors are germline encoded, involved in innate immunity, and can inhibit replication of specific viruses in host cells (*6, 172*). These factors target viral replication by different mechanisms and are important in determining viral host range and cross-species transmission. So far, four types of HIV-1 restriction factors have been identified (Figure 1.5 & Table 1.1). These are APOBEC3 enzymes, tripartite motif (TRIM) 5α, bone marrow stromal cell antigen 2 (BST-2, also known as tetherin) and sterile alpha motif and histidine-aspartic domain 1 (SAMHD1).

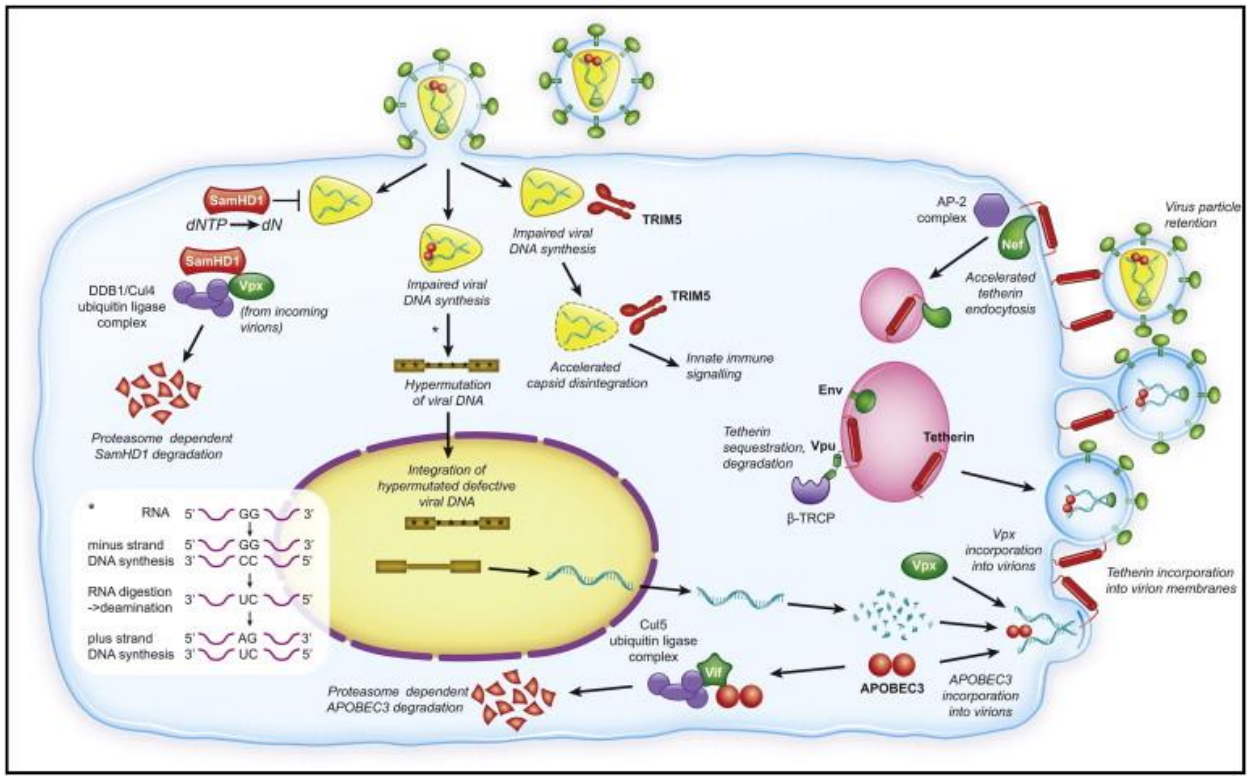

Figure 1.5. The host restriction factors and the retroviral accessory proteins. Overview of restriction factors that target HIV and their viral antagonists. The figure depicts the key mechanisms that restriction factors employ to prevent the retroviral replication and their counteraction by viral accessory proteins. The process of viral genome hypermutation by APOBEC3 is indicated in the inset panel. Reprinted with permission from (*173*).

Table 1.1. Characteristics of host restriction factors

Restriction Factor	IFN Induced	Life cycle stage inhibited	Viral antagonists	Antagonistic mechanism	Under positive selection	References
TRIM5α	Yes	Capsid uncoating	None known (escape through capsid mutations)	variation in retroviral capsid	Yes	(*174, 175*)
SAMHD1	Yes	Reverse Transcription	Vpx (some SIVs) Vpr (some SIVs)	proteasomal degradation	Yes	(*176, 177*)
Tetherin	Yes	Budding	Nef (Some SIVs) Vpu (HIV-1) Env (HIV-2)	prevent viral detachment from host cell	Yes	(*175*)
In $CD4^+$ T-cells: APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H	No	Reverse Transcription, integration, infectivity	Vif	polyubiquitination and degradation	Yes	(*18, 31, 175, 178, 179*)
In monocytes: APOBEC3A	Yes	Reverse Transcription	None	None	Yes	(*18, 31, 175, 178, 180, 181*)

1.8.1. Hallmarks of restriction factors

Restriction factors have at least four defining characteristics (Figure 1.6)

i. **Immune induction:** Restriction factor expression is often induced by interferon, but some restriction factors are also expressed constitutively due to having a role in restricting endogenous retroelement events which do not lead to IFN responses (*172*).

ii. **Virus restriction:** A restriction factor must cause a significant decrease in virus infectivity and impair viral infectivity by at least 10-fold (*173*).

iii. **Counteraction:** Restriction factors should be evolving with potent counter restriction mechanisms. These viral antagonist proteins are often encoded by 'accessory genes' that are not needed for virus replication except in the presence of restriction factors (*182*)**.** Although viral antagonists overcome antiviral effects through diverse mechanisms, antagonism is exclusively triggered by direct protein-protein interactions between the viral antagonist and corresponding restriction factor and often recruits cellular degradation or transport pathways (*154*).

iv. **Positive selection:** Positive selection results when two genetic entities such as hosts and viruses evolve in conflict with one another. In general, if mutations are advantageous it will be maintained in a population. If a host species constantly experiences pathogenic pressure, it selects for altered variants of host restriction factors that are not susceptible to the pathogen's counteraction mechanism. During virus-host interactions both the host and virus evolve to increase their fitness and this results in the host restriction factor gene having a high number of non-synonymous mutations compared with synonymous mutations, a process called positive selection. Because antagonism is exclusively triggered by direct protein-protein interactions between viral antagonist and corresponding restriction factors, the restriction factor contains a signature of positive selection primarily in the nucleotides that code for the amino acids that directly interact with the virus protein. These positive selection signatures become apparent by comparing restriction factor gene sequences between the host and related evolutionary species (*172*).

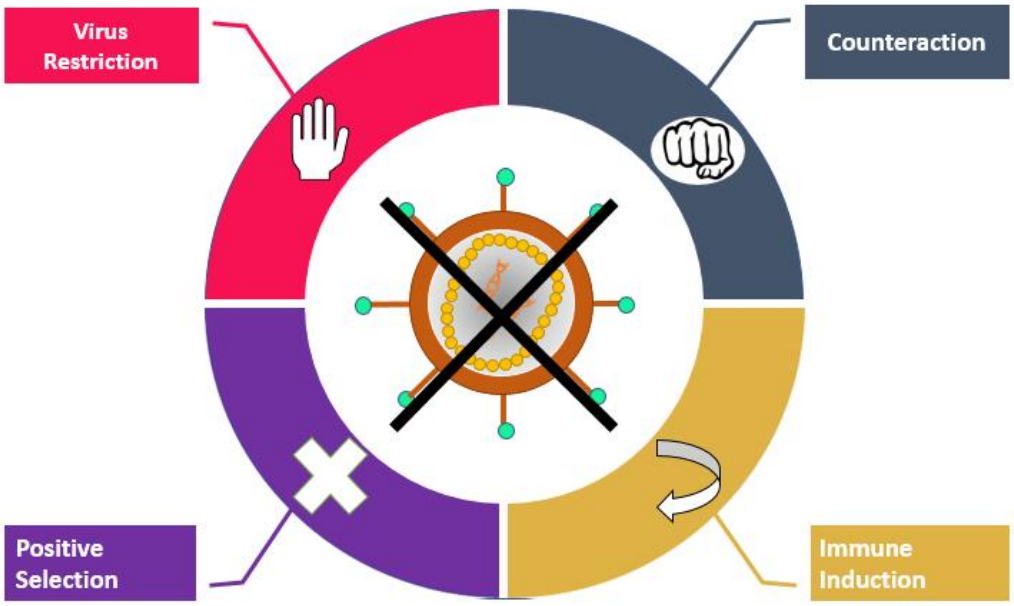

Figure 1.6. Hallmarks of a restriction factor. The four hallmarks that define an HIV restriction factor include a virus-encoded counter mechanism (fist sign), dominant restriction of viral growth (palm sign; clockwise from top left), interferon-inducible (arrow sign), and signatures of positive selection (plus sign). Modified from (*154*).

1.8.2. TRIM5α

TRIM5α (57.3 kDa,) is a member of TRIpartite Motif family of proteins that share a common architecture but have distinct functions. TRIM5α was originally discovered as an important determinant of the resistance of monkey cells to HIV-1 infection (*183*). TRIM family members bind to incoming viral capsids in the cytoplasm and lead to capsid disruption and/or degradation (*184*). TRIM5α promotes rapid and premature disassembly of the viral capsid (*184*) and promotes downstream innate immune responses (Figure 1.5 & Table 1.1) (*185*). TRIM5α contains an N-terminal domain comprising RING, B-box, and coiled-coil domains. RING is a zinc-binding domain that confers E3 ubiquitin ligase activity. B box and coiled-coil domains mediate formation of TRIM5α multimers, which is required for antiviral activity. The C-terminal domain (CTD) of TRIM5α, called B30.2 or SPRY domain, recognizes the lattice of viral CA in a species-specific manner, and determines the specificity of restriction (*173, 186*). TRIM5α activity is not antagonized by an accessory gene like other restriction factors. Instead, the HIV-1 capsid has evolved point mutations to avoid recognition by human TRIM5α, but HIV-1 capsid is still susceptible to the Rhesus Monkey version of TRIM5α (*187*). Generally, TRIM5α is ineffective against retroviruses that naturally infect the same host species, but actively blocks replication of retroviruses that are found in other species.

1.8.3. SAMHD1

SAMHD1 is 72 kDa protein, which is IFN-γ-inducible and highly expressed in myeloid cells and resting $CD4^+$ T cells. SAMHD1 depletes deoxynucleotide triphosphate (dNTP) pools in the cell to block HIV-1 reverse transcription (Figure 1.7) (*176*). Such interference obstructs the completion of viral cDNA synthesis before the capsid dissociates, and this subsequently leads to cellular protease- and nuclease- mediated virus degradation (*154*). SAMHD1 is composed of an N-terminal nuclear localization domain, SAM domain, HD domain and C-terminal variable domain (*188*). The SAM domain mediates protein-protein interactions. The HD domain contains conserved histidine and aspartate residues and is involved in nucleic acid metabolism and RNA binding (*189*).

SAMHD1 depletes the intracellular dNTP pool through dNTPase activity. After activation by dGTP, SAMHD1 hydrolyzes dNTPs substrates to deoxynucleosides and inorganic triphosphate (*176, 190*). SAMHD1 mediated restriction can be relieved by addition of exogenous dNTPs to the level required for the synthesis of viral cDNA. Moreover, HIV-2 and SIV Vpx proteins counteract

viral restriction by triggering the degradation of SAMHD1 (*191, 192*). However, HIV-1 lacks an antagonistic protein against SAMHD1 because it does not have a Vpx protein and its Vpr protein is not capable of interacting with SAMHD1. This makes myeloid cells more resistant to HIV-1 infection. Due to this reason, SAMHD1 does not show all the four hallmarks (not antagonized by the HIV-1 virus) of a restriction factor in regards HIV-1.

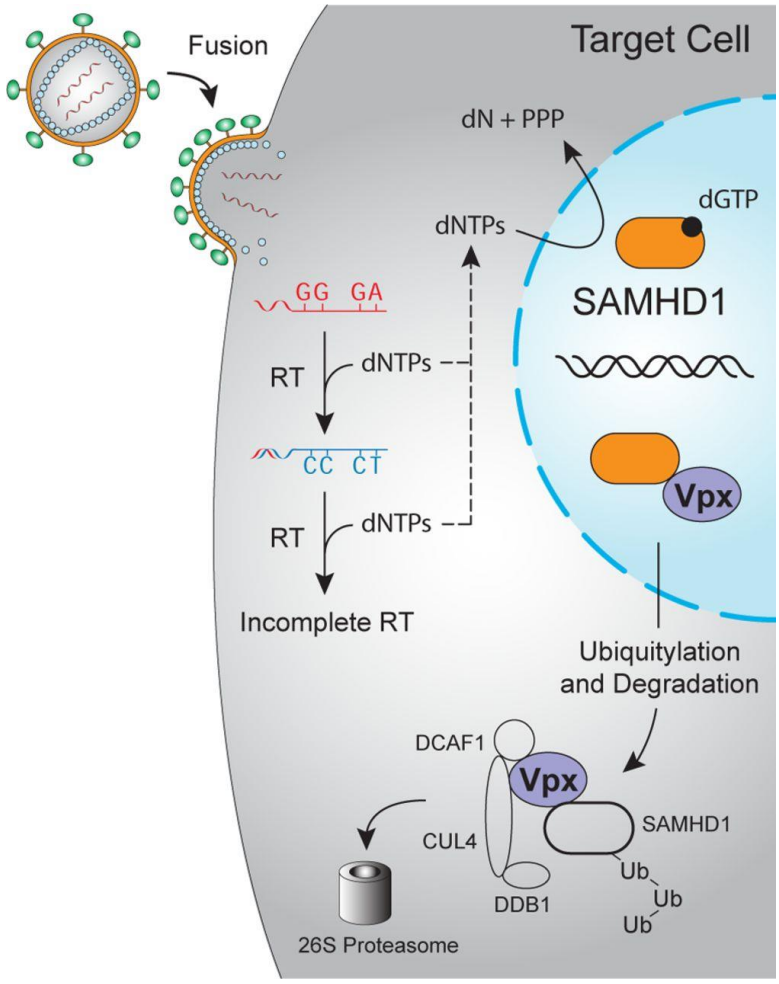

Figure 1.7. HIV restriction by SAMHDI. In myeloid cells, SAMHD1 prevents HIV-1 reverse transcriptase (RT)-mediated polymerization by depletion of cellular dNTPs. HIV-2/SIV Vpx and some Vpr variants can overcome the SAMHD1 restriction block by acting as an adaptor to an E3 ubiquitin (*Ub*) ligase complex that polyubiquitinates SAMHD1 and targets them for proteasomal degradation by 26 S proteasome. Reprinted with permission from (*154*).

1.8.4. BST-2 (tetherin)

BST-2 is a 19.7 kDa (180 amino acids) IFN-induced type 2 transmembrane protein that that acts as an inhibitor against the release of Vpu-deficient HIV-1 and other enveloped viruses (Figure 1.8). It blocks the release of virus by trapping nascent but mature virions at the cell surface (*193*, *194*) from where they may be endocytosed. This defect in release of virus particles from the cell surface was due to tetherin, which can simultaneously associate with both viral and cellular membranes. At the time of virus budding from the cell, tetherin becomes incorporated into the viral envelope and physically bridges nascent virions to the cell (*195*). However, viruses have also adapted diverse strategies to antagonize BST-2. More than one antagonist has been employed by primate lentiviruses, including Vpu, Nef, and Env. HIV-1 Vpu directly interacts with BST-2 via their transmembrane domains and transports BST-2 to the trans-Golgi network/early endosomes where BST-2 undergoes proteasomal and/or lysosomal degradation (*196*). Instead of Vpu, some SIVs use Nef to antagonize BST-2. Nef recruits adaptor Protein-2 clathrin adaptor complex to target BST-2 and induces its endocytosis (*197*). Additionally, Env gp41 cytoplasmic tail of HIV-2 and SIV_{MAC} physically interacts with BST-2 and promotes the internalization and sequestration of BST-2 from sites of viral assembly on the cell surface (*154, 198*). Thus, BST-2 is a potent restriction factor that different immunodeficiency viruses have found a way to counteract, even in the absence of Vpu.

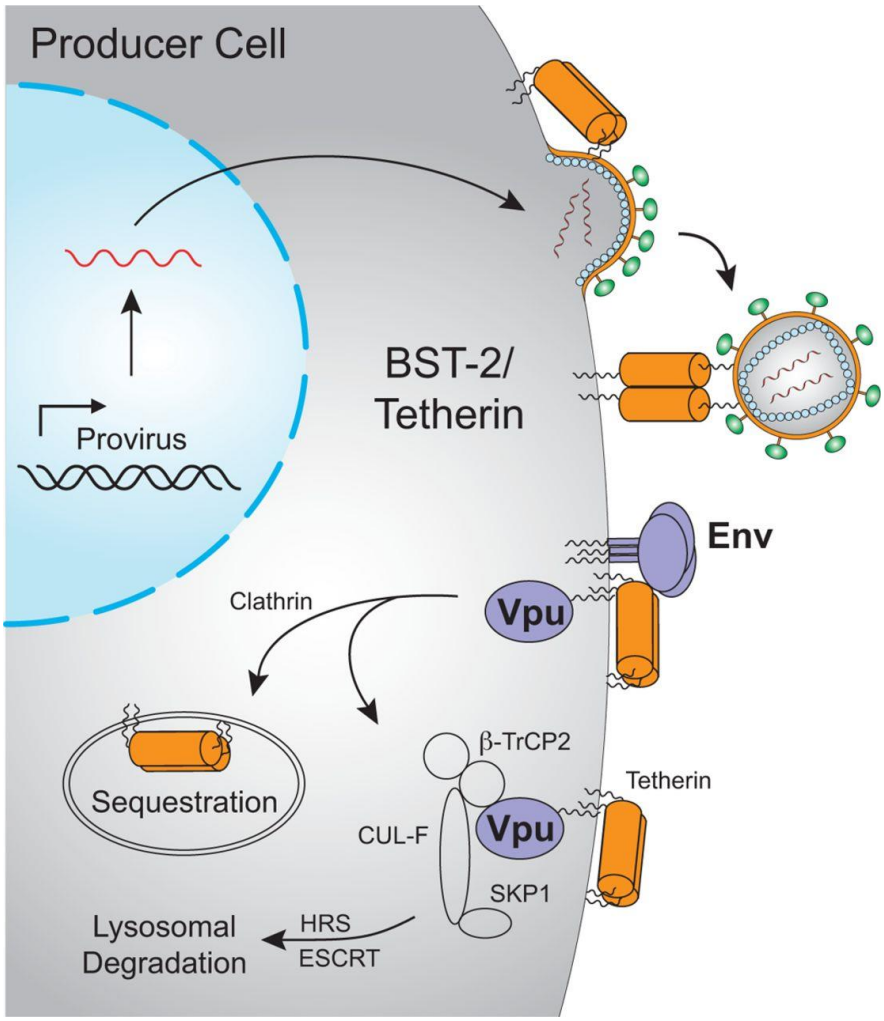

Figure 1.8. HIV restriction by BST-2/Tetherin. Tetherin blocks the release of budding virions by trapping them at the cell surface of infected cells. HIV-1 Vpu or HIV-2 Env overcomes tetherin restriction by promoting the internalization and sequestration of tetherin in compartments away from sites of viral budding. HIV-1 Vpu can also restrict tetherin by recruiting an E3 ubiquitin ligase complex that ubiquitinates tetherin and targets it for degradation in lysosomes. Reprinted with permission from (*154*).

1.8.5. APOBEC3s

The antiviral function of A3G was discovered when it was found that Vif was absolutely required for HIV-1 replication in some cell lines (non-permissive cells), but not others (permissive cells) (*6, 199, 200*). Using subtractive hybridization between a permissive and non-permissive cell lines from the same origin (CEM cell line), Sheehy *et al*., identified that the gene that encoded A3G was the non-permissive factor. Earlier research used hybrid cells constructed by fusing permissive and non-permissive cell lines to show that the non-permissive phenotype was due to a dominant-acting factor that restricted Vif-deficient HIV-1 infectivity (*6*). Subsequent genetic analysis of chromosome 22 by an unrelated research group revealed that A3G is the part of a larger subfamily of seven APOBEC3 proteins that included A3A, A3B, A3C, A3D, A3F and A3H (*201*). Each have the capacity to catalyze deamination of cytosine to uracil in ssDNA (C→U) (*202*). In the $CD4^+$ T cells that HIV-1 infects, the restriction factors A3G, A3D, A3F and A3H are expressed. For A3 enzymes to restrict HIV-1 in $CD4^+$ T cells, their encapsidation into the budding virion is a prerequisite (Figure 1.9) (*203*). It is also reported that in the absence of Vif, there is encapsidation of 7 ± 4 A3G molecules into the viral particle and this is sufficient to block viral replication (*204, 205*). Although A3A has been shown to restrict HIV-1 replication in myeloid cells, the mechanism is different since target cell A3A, not encapsidated A3A, restricts the virus (*180, 206*).

Each A3 gene encodes a protein with one or two conserved zinc (Z)-coordinating motifs (Z1, Z2 or Z3) (*207*). The defining feature of A3 family is a conserved His-X-Glu-X25-31-Pro-Cys-X2-4-Cys zinc-coordinating motif within the active site where a water molecule binds Zn^{+2} and the metal ion is coordinated by one histidine and two cysteines (*208*). The Zinc coordination motif is strictly required for deaminase activity (*209, 210*). The A3 Z domain was grouped into one of three distinct phylogenetic clusters – Z1, Z2 or Z3. In the Z-based classification system, Z1 and Z2 proteins have a Ser-Trp-Ser/Thr-Cys-X2–4-Cys motif, whereas the corresponding motif in Z3 proteins is a Thr-Trp-Ser-Cys-X2-Cys (TW-S-C-x2-C) motif. Z1 and Z2 proteins can be further distinguished by His-X1-Glu-X5- X-Val/Iso and His-X1-Glu-X5-Trp-Phe motifs, respectively (*207*). Genes for A3A (Z1), A3C (Z2), and A3H (Z3) contain a single Zn^{+2} binding domain whereas genes for A3B (Z1-Z2), A3D (Z2-Z2), A3F (Z2-Z2) and A3G (Z1-Z2) have resulted from duplications of the primordial gene (*201, 210*) and contain two putative zinc-binding motifs (Figure 1.10). However, only the C-terminal Z-domain is catalytically active (*210-212*).

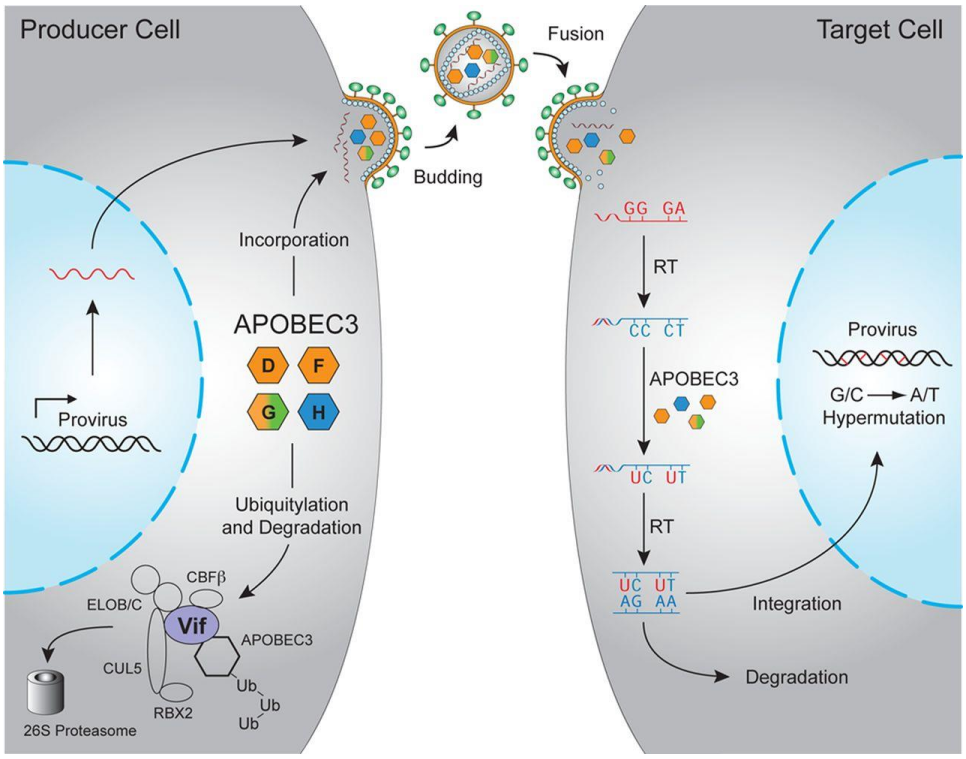

Figure 1.9. APOBEC3 (A3) mediated inhibition of HIV-1. HIV-1 Vif counteracts the A3-mediated restriction in the producer cells by binding CBFβ and recruiting an E3 ubiquitin (Ub) ligase complex to polyubiquitinate the A3 proteins and target them for degradation by the 26 S proteasome. A3s (A3D, A3F, A3G, and A3H) that avoid Vif-induced proteasomal degradation encapsidate into HIV-1 virions. In target cells, encapsidated A3s catalyze the deamination of cytosines to uracils in viral cDNA during RT-mediated polymerization. RT uses uracil as a template during positive-strand synthesis, resulting in G→A mutations. These proviral cDNAs are subsequently degraded or integrated into host genome. Reprinted with permission from (*154*).

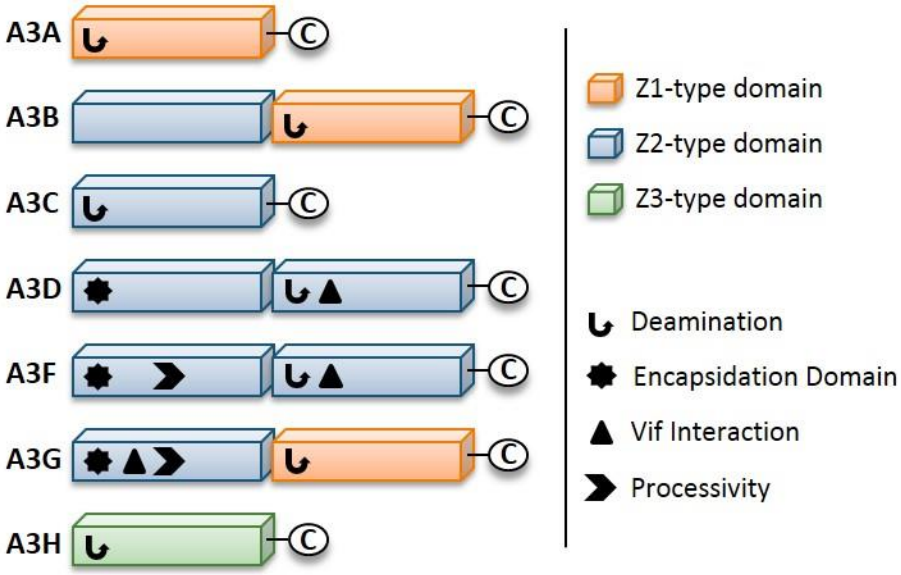

Figure 1.10. Zinc (Z) coordinating-type domains of human A3 enzymes. A3 enzymes coordinate zinc by using the motif H-X-E-X23-28-P-C-X2-4-C. Zinc-hydroxide-mediated nucleophilic attack to complete the deamination reaction is enabled by the glutamate by activating a water molecule. All A3 enzymes have deamination activity. A legend depicts known biochemical functions of each Z-type domain of the A3 enzymes. The segregation of functions in the N-terminal domain (NTD) and C-terminal domain (CTD) is common feature of A3 enzymes with two Z-type domains. The NTD and the CTD are responsible for encapsidation and deamination activity, respectively. Both domains can bind nucleic acids. The Vif binding site is in the CTD for A3D and A3F and the NTD for A3G. Reprinted with permission from (*213*).

Due to different amino acids on structural loop 7 surrounding the active site, different A3 proteins deaminate cytosines in different preferred sequence contexts (*214*, *215*). A3G preferentially deaminates the 3′C in 5′CC motifs. The deaminations of cytosine on the minus-strand DNA result in a GG→GA mutation signature on the plus-strand DNA (*216*, *217*). A3D, A3F, and A3H each deaminate at 5′TC motifs, resulting in a GA→AA mutation signature (*2*, *218*, *219*)**.** More detailed analysis of the signature mutations revealed that A3G is most active at 5′CCC motifs and A3D can be differentiated from A3F and A3H by examining a larger surrounding sequence 3′ of the cytosine (*220*). In the absence of HIV-1 Vif, A3G, A3F, A3D, and A3H molecules can package into progeny virion by binding RNA (HIV genome or cellular RNA such as 7SL or Y) that also interact with the HIV-1 nucleocapsid (NCp7) region of Gag polyprotein (*186*, *221*, *222*). Despite being present in the cytoplasm of the $CD4^+$ T cells that HIV-1 infects, the A3s have to encapsidate into the virion to restrict HIV-1 replication (*223*). This is a requirement because the capsid structure does not allow A3 enzymes to diffuse into the capsid and A3 enzymes are compartmentalized in regions of RNA processing, e.g., stress granules or P-bodies (*224*, *225*). However, the capsid does allow for dNTPs from the host to diffuse inside so that after viral maturation and entry to the next target cell, reverse transcription can begin (*151*)

When the reverse transcription starts in virions that have encapsidated A3s, the A3 cytidine deaminase activity acts upon the nascent minus-strand viral cDNAs to deaminate cytosines to uracils (Figure 1.9). The reverse transcription stage is a very competitive environment where A3 enzymes have to compete with RT for single-stranded (-)DNA before it is copied to form dsDNA and compete to bind RNA or DNA also bound by NC. As a result A3 enzymes must search within the HIV-1 (-) DNA for their specific recognition sequence efficiently in order to deaminate the most cytosines in the short window of time that the (-)DNA is exposed as single-stranded. A3 enzymes that restrict HIV-1 have been characterized to use specific DNA scanning mechanisms to search for cytosines on ssDNA and can catalyze multiple cytosines in a single enzyme–DNA substrate encounter, meaning that they are processive enzymes (*213*, *226*, *227*).

The A3 cytosine deaminase activity results in the misincorporation of adenines, instead of guanines, in the (+)DNA due to using uracil as a template (*216*). It is known that many proviral genomes undergo successful integration with these hypermutations (*228*), but other proviral genomes may not fully complete reverse transcription due to A3 interference (*229-231*). Additionally, some preintegration complexes containing U are degraded by host DNA repair

mechanisms, although there is no consensus regarding the extent to which this occurs in cells (*230, 232*). The uracils commonly cause missense and nonsense mutations that are lethal for viral replication. A3G may also hinder functional viral protein expression and progeny virus production (*233*). It was also reported that the introduction of C→U mutations in the trans-activation response (TAR) element, a key regulation factor of HIV-1 transcription elongation, can lead to an early block of viral gene expression (*234*). Previously it was proposed that A3 sublethal mutagenesis of proviral genomes has the potential to accelerate the development of drug resistance (*235-244*). However, recent studies contradict the previous hypothesis and provide some evidence for minimal contribution of A3-induced G→A hypermutation to HIV-1 evolution and emergence of drug resistance (*215, 235, 245*).

1.9. Restriction of HIV-1 by A3G, A3F, A3D, and A3H

1.9.1. Deamination-dependent restriction of HIV-1

1.9.1.1. A3G mediated restriction

A3G (46.4 kDa) was the first A3 enzyme discovered and was found to restrict HIV-1 and be responsible for the non-permissive phenotype (*6*). A3G exist in the cells as a high molecular mass ribonucleoprotein complex. Evidence suggests that only newly synthesized low molecular mass A3G is encapsidated into the virion (*246*). The amino terminal domain (NTD) of A3G is solely responsible for its virion encapsidation because the RNA binding and oligomerization of A3G is primarily mediated by the NTD of A3G (*247*). Work largely conducted by our lab has established that A3G is highly effective at restricting HIV-1, after successful virion encapsidation, because of its ability to efficiently scan ssDNA in search of its deamination target motif (*226*). Enzymes that do not use an energy source for movement on DNA use a mechanism termed facilitated diffusion to speed up the target site search (*248, 249*). These mechanisms include sliding along DNA, microscopic dissociation-reassociation events, and intersegmental transfer (Figure 1.11) (*248*). Sliding is a 1-dimensional DNA scanning mechanism by which the enzyme undertakes an in-depth search of local areas of a substrate to find their specific motif. Enzymes may also undergo a 3-dimensional DNA scanning path by jumping or intersegmental transfer to search for its motifs. An enzyme jump is a micro-dissociation from the DNA and subsequent reassociation of the enzyme with the DNA without entering the bulk solution. The enzyme stays in the negatively charged domain of the DNA and thus has a higher likelihood of reassociating with the same DNA substrate than if it had diffused into the bulk solution. Jumping helps in larger

translocations on DNA substrates, but lacks a local search mechanism. Intersegmental transfer is when a DNA-binding enzyme binds two areas of DNA simultaneously before dissociating from one area to move to another (*226, 250*). Intersegmental transfer enables even larger translocations on DNA substrates than jumping, but also lacks a local search process and requires that the enzyme have at least two DNA binding domains (*226, 251*). A combination of 1-D and 3-D movements results in a more efficient search of the DNA substrate (*226*). A3G is efficient in deamination of HIV-1 genome because it uses a balanced combination of 1-D sliding and 3-D jumping motions (*250*).

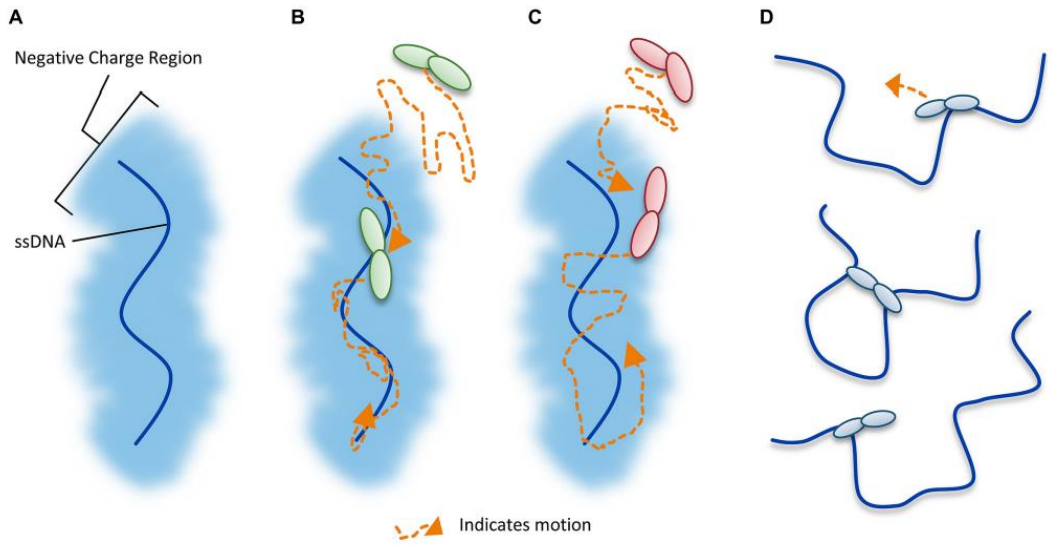

Figure 1.11. APOBEC3-mediated ssDNA scanning mechanisms. Illustration of DNA scanning by facilitated diffusion. (A) Sketch of DNA surrounded with negatively charged region that facilitates diffusion of A3 enzymes. (B–D) sketches of A3 enzyme is shown as a dimer, although the oligomerization state may vary with different A3 enzymes. (B) Sketch depicting 1-dimensional sliding movement. Path of enzyme is denoted by dotted line (orange). (C) Sketch depicting jumping motion of A3 enzyme that includes micro-dissociations and re-association of the enzyme from the DNA without diffusion into the bulk solution. The jumping motion entails a 3-dimensional DNA scanning path and enables larger translocations on DNA substrates. (D) Sketch depicting a 3-dimensional DNA scanning path by intersegmental transfer. Intersegmental transfer is a doubly bound state of an enzyme where the enzyme leaves the charged domain of the DNA and enters into the bulk solution to bind another DNA segment before releasing the first bound DNA. Reprinted with permission from (*213*).

Mechanisms of DNA scanning that support processive enzyme activity allow enzymes to catalyze multiple events in a single enzyme-DNA substrate encounter. In contrast, if an enzyme undergoes macroscopic dissociation from DNA and becomes part of the bulk solution after one turnover, it will not be able to catalyze multiple events per DNA contact and is considered to be non-processive or distributive. Most studies suggest that A3G is processive and scans ssDNA by 1-D sliding motions and 3-D jumping motions (*227, 252, 253*). A3G is a positively charged enzyme (charge of +6.5 at pH 7) that scans ssDNA by facilitated diffusion because it is retained in the negatively charged domain of the DNA and moves via electrostatic interactions (*205, 227*). The 1-D sliding motion introduces clusters of deamination in ssDNA which is evident in integrated proviral genomes (*254*). One research group has found that A3G does not efficiently deaminate closely spaced cytosines and moves between successive targets predominantly through intersegmental transfer (*205*). The efficiency imparted by 3-D movements of A3G such as jumping or intersegmental transfer during reverse transcription is that it enables the enzyme to maneuver over obstacles such as RNA/DNA hybrids that are left on the (-) DNA due to RNase H endonuclease activity (*205, 250*). The processivity determinants of A3G reside on predicted loop 7 and helix 6 of the non-catalytic NTD (*250*). Thus, despite lacking catalytic activity, the NTD contributes to A3G deamination activity by mediating the processive scanning mechanism.

1.9.1.2. A3F mediated restriction

A3F was discovered as an HIV-1 restriction factor approximately 2 years after the discovery of A3G (*2, 24*). Overexpression of A3F causes a dose-responsive drop in the infectivity of Vif-null HIV-1, and induces mutations in a GA→AA context (minus strand TC→TT) (*2, 21*). The mutational potency and antiviral effect of A3F appears to be lower than that of A3G (*226, 255*), but knockdown and knockout experiments have shown that endogenous levels of A3F can restrict HIV-1 replication and induce mutations in a $CD4^+$ T cell line, and both Human and Rhesus macaque A3F have anti HIV-1 ability (*223, 256*). It has been observed that A3F encapsidates into the core of HIV-1 particles more efficiently than A3G, which might be enabled by its ability to bind nucleic acids with seven-fold higher affinity than A3G (*22, 226, 257*).

In spite of A3F being less effective than A3G in restricting HIV-1 (*226, 257, 258*), studies indicated that antagonization by Vif is required to preserve HIV-1's replication capacity (*259*). Data has also implied that while HIV-1 can evolve a Vif-independent mechanism to evade A3G

restriction, it cannot evolve a Vif-independent mechanism to overcome A3F restriction (*260*). Thus, even if A3F does not have a strong restriction capacity for HIV-1, it does exert an important suppressive force on the virus.

Both Vif binding and cytidine deaminase activity reside within the CTD of A3F, unlike A3G where these features are split between the NTD (Vif binding) and CTD (cytidine deaminase activity) (*211, 261*). Mutagenesis and high-resolution structural studies have shown that the Vif binding surface of A3F is distinct from A3G and it is more similar to A3C and A3D (*262-264*).

1.9.1.3. A3D mediated restriction

The role of A3D as an HIV-1 restriction factor and its sensitivity to Vif was first reported by Dang *et al*. (*218*). A3D had a promiscuous deamination motif preference for the 5′TC and 5′GC contexts (*218*). Single-cycle infectivity assays (*218, 223, 265*) and spreading infections in primary cells (*266*) demonstrated less HIV-1 restriction ability compared to A3F and A3G. It was shown that a single amino acid ^{320}C in the loop7 of CTD loop 7 was responsible for the suppressed antiviral activity of A3D (*265*). Refsland *et al*. reported that A3F and A3D have similar anti-HIV function in the T cell line CEM2n (*19*). Like A3G, A3D also shows multimerization through an RNA intermediate in cells (*267*).

1.9.1.4. A3H mediated restriction

In recent years, A3H has received much attention because it was discovered to exist as multiple haplotypes in the human population (hap I-VII), each having different cellular stability and HIV-1 restriction capability. A3H is also different from other A3 enzymes because it contains an evolutionarily distinct zinc coordinating domain (Z-domain). A3H has a Z3 zinc coordinating domain whereas A3D and A3F have two Z2 domains and A3G has one Z1 (CTD) and one Z2 (NTD) domain (*210, 268*). Evidence indicates that out of seven A3H haplotypes only hap II, hap V, and hap VII can restrict Vif-deficient HIV-1 (*269, 270*). The anti-HIV-1 function of A3H haplotypes depends upon the stability of the protein (*271, 272*), their subcellular localization (*273, 274*), binding affinities to RNA (*274*), and relative levels of virion packaging (*269*). It has also been reported that A3H forms oligomers *in vitro* and multimerizes on RNA in cells (*267, 275*). The most prevalent form of A3H in the population is hap I, which is a thermodynamically unstable protein and inactive against HIV-1 (*213*). A3H hap II preferentially deaminates ssDNA at 5'TC sites, as do A3F and A3D. The mutagenic potential of A3H hap II, demonstrated by both single-

cycle infectivity assays, spreading infection experiments, and in HIV-1 infected individuals, correlates with its HIV-1 restriction ability (*213, 219, 276*). A higher frequency of the highly active and stable A3H hap II is present in populations in Africa, possibly a result of the long-term presence of the SIV/HIV in that continent (*277-279*). Inactive forms of the A3H haplotypes are mostly found in people of non-African descent (*277, 280*). Importantly, Vif from HIV-1 strains circulating outside of Africa are not very efficient in degrading stable A3H alleles (*270, 277*). It was proposed that the adaptive changes in HIV-1 Vif sequences happened depending on the presence or absence of various A3H haplotypes in a human population and that this determined the ability of A3H to act as an infection barrier in that population (*272, 276*).

1.9.2. Deamination-independent inhibition of HIV-1

Initially A3G-mediated deamination was proposed as the sole mechanism of antiviral activity against Vif-deficient HIV-1, but subsequent studies have demonstrated that A3G and other A3 proteins exert anti-viral activity that is deamination-independent (*229, 281-289*). Many reports indicated deamination-independent A3G-mediated antiviral effects at different stages of viral DNA synthesis such as the inhibition of tRNA primer binding, initiation of reverse transcription, elongation of reverse transcription, plus-strand DNA synthesis, production of aberrant viral 3′-LTR ends that suppress $tRNA_{lys3}$ cleavage and removal, and inhibition of integration (*229, 255, 284-286, 289-291*). It was also demonstrated via endogenous reverse transcription assays that A3G inhibits the elongation of reverse transcripts rather than enhancing reverse transcriptase degradation (*229*). In addition, a previous study reported that the mutation of amino acid residues in the CTD critical for deamination activity (H257R, E259Q, C291S) of A3G failed to significantly impair antiviral activity of A3G (*281*). However, the authenticity of deamination-independent A3G restriction activity has been controversial, and several other groups suggested that the deamination ability of A3G is required for HIV-1 restriction (*291-293*). However, these studies should be considered with caution as other studies have shown that the A3G E259Q catalytic mutant binds RNA about 1.4- to 2-fold less tightly than A3G and such artifacts of the catalytic mutant may result in less inhibition of RT (*229, 283*). Furthermore, using purified A3G protein, one lab suggested that A3G may inhibit RT by binding tightly to the ssDNA or RNA template. They demonstrated that when A3G encounters ssDNA it initially undergoes a rapid on and off exchange with ssDNA followed by stable binding as demonstrated by a slow dissociation mode (*294*). This slow

dissociation mode was proposed to form a roadblock that may physically obstruct viral DNA synthesis (*288, 294, 295*).

A3F was suggested to have a stronger deamination-independent mode of inhibition of HIV-1 replication than A3G, and it was presumed that the inhibition of RT polymerization was the reason for the observed inhibition (*255, 296*). A computational study suggested that the A3G deamination-independent mode contributes to <1% of its restriction activity whereas for A3F it accounted for approximately 30% of its restriction activity (*297*). However, unlike previous studies that used A3F overexpression (*255, 296*), in two recent studies that stably expressed biologically relevant levels of A3F wild-type and catalytic mutants (C280S/C283A and E251Q) there was no inhibition of RT (*258, 298*), suggesting that A3F overexpression artifacts could have influenced previous interpretations (*255, 296*). Despite the possibly inefficient mutagenic function of A3F in some studies, it appears that the deamination mode of anti-HIV activity of A3F is dominant to deamination-independent mechanisms. Mbisa *et al*. reported that A3F can reduce 3′ processing of viral DNA at the U5 and U3 ends to inhibit HIV-1 integration (*299*). Using wild-type and a catalytic mutant (E251Q) of A3F, their study found that catalytic activity is in part required to inhibit HIV-1 integration by producing the aberrant U5 and U3 ends (*299*).

1.10. Modes of virus escape from APOBEC3s

All lentiviruses except equine infectious anemia virus (EIAV) code for Vif. Vif is an A3 antagonist that protects lentiviruses from replication restriction by several members of the A3 family of DNA cytosine deaminases. Vif forms a polyubiquitin ligase complex with host proteins to induce degradation of A3 enzymes in the proteasome (*14*). Vif interacts with host protein Elongin C and Elongin C is an obligate dimer with Elongin B (referred to as EloB/C) (*16, 300-302*). After interaction with ELoB/C, Vif becomes more stable and this promotes recruitment of CBFβ (*12, 213, 303, 304*). This Vif/CBFβ/EloB/C complex interacts with the scaffolding protein Cullin 5 (Cul5) (*16, 302*) and Cul5 interacts with Rbx2. Altogether, this hexamer forms an E3 ubiquitination complex with Vif as the substrate receptor. Vif binds to an A3 enzyme and the E3 complex recruits an E2 ligase to polyubiquitinate the A3 enzyme through Lysine 48 linkages which signals the protein for proteasomal degradation (Figure 1.12) (*12, 213*).

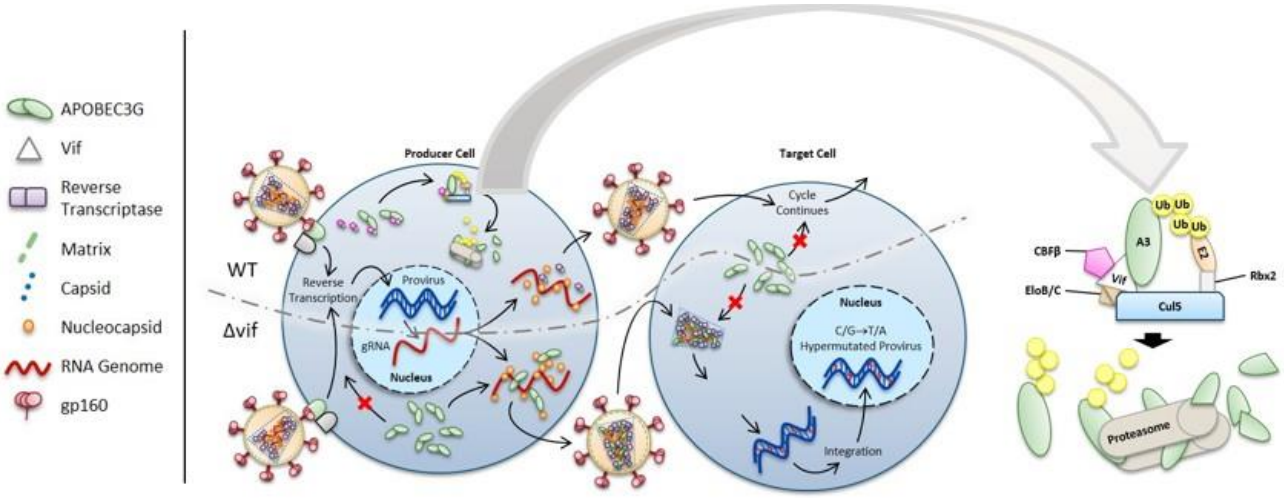

Figure 1.12. Schematic depicting HIV Vif-mediated polyubiquitination of A3 for escaping viral restriction. Vif-mediated polyubiquitination of A3. Vif interacts with Elongin C (EloC), which forms an obligate heterodimer with Elongin B (EloB), and Cul5. The transcription cofactor CBFβ stabilizes Vif. Cul5 binds to Rbx2 and subsequently recruits an E2 ubiquitin conjugating enzyme. Vif is the substrate receptor that recruits A3 enzymes. The 48K-linked ubiquitin chains result in proteasomal degradation of the A3. Modified from (*213*).

The ability of a lentiviral Vif to degrade a species' A3 enzymes influences virus host range. Early reports indicated that HIV-1 Vif cannot neutralize A3G from African green monkey (AGM), and AGM SIV Vif cannot neutralize A3G from humans (*12*). Thus, Vif inhibits the antiviral activity of A3 enzymes in a species-specific manner (*305, 306*). More recent support for this comes from evolutionary analysis of SIV from chimpanzees and HIV-1 in humans, where it was found that SIV from chimpanzee was only able to infect humans if the region of Vif was modified to antagonize human A3s (*307*). Vif interacts with A3 enzymes through a variety of motifs (Figure 1.13). Vif interacts with A3G through two positively charged regions on Vif, primarily ^{40}YRHHY^{44} and secondarily ^{21}WxSLVK^{26} (*308-310*). Similarly, various domains in Vif have been identified to interact with A3F, specifically, ^{11}WQxDRMR^{17} is the primary interface and ^{74}TGERxW^{79} is the secondary interface (*311-314*). In addition, Vif also has a ^{69}YWxL^{72} motif that interacts with both A3G and A3F (*315*). Evidence for these primary interfaces comes from experiments that mutated ^{40}YRHHY^{44} or ^{14}DRMR^{17} motifs to all alanines and found this was sufficient to block Vif-induced A3G or A3F degradation, respectively, suggesting the other domains provide a secondary stabilizing interaction (*311*). The ^{14}DRMR^{17} motif on Vif is also important to bind with A3C and A3D (*316, 317*). Vif interacts with A3H through another unique site that involves amino acid ^{39}F and ^{48}H (*318, 319*).

Each A3 has different amino acids that interact with Vif. For human A3G, the charged interface was more important for Vif association than the amino acid identity (*320*). On A3G, amino acid D128 is a determining residue since a sole mutation of ^{128}D to ^{128}K can abrogate the interaction between A3G and Vif in coimmunoprecipitation studies (*320*). For A3F, one group reported that A3F interacts with Vif through CTD amino acids ^{289}EFLARH^{294} (*321*) and another group found that ^{289}E was critical for A3F sensitivity to Vif (*322*). The key amino acid on A3H that interact with Vif is D121, similar to A3G (307).

Although Vif primarily inhibits A3G by inducing its proteasomal degradation, it can also inhibit A3G encapsidation or function through a degradation-independent route. Vif can decrease A3G mRNA translation in order to lower the steady-state levels of A3G through a Vif and A3G mRNA interaction (*323, 324*). For Vif-mediated translation inhibition of A3G in cells, two stem-loop structures within the 5′-UTR region of A3G mRNA are critical (*324*). A3G normally forms high molecular mass complexes that are less likely to be packaged into virions and Vif can induce

an even higher molecular weight form of A3G which inhibits A3G encapsidation (*325*). Vif has also been reported to inhibit the deamination activity of virion-encapsidated A3G (*326*, *327*).

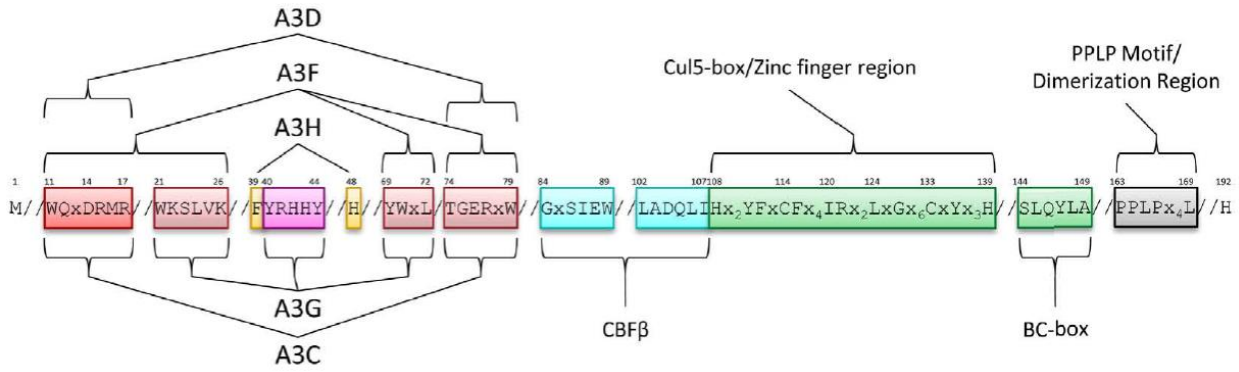

Figure 1.13. Domain organization of Vif and specific motifs that interact with host proteins. Vif uses specific motifs to interact with A3C/A3D/A3F (red, ^{11}WQxDRMR^{17} and ^{74}TGERxW^{79}), A3H (orange, ^{39}F and ^{48}H), and A3G (magenta, ^{40}YRHHY^{44}). Vif uses a common motif (pink, ^{21}WKSLVK^{26} and ^{69}YWxL^{72}) to interact with A3F and A3G. The two adjacent motifs of Vif (cyan, ^{84}GxSIEW^{89} and ^{102}LADQLI^{107}) interact with CBFβ. The Zinc is coordinated through an ^{108}H^{114}C^{133}C^{139}H motif with the Zinc finger region (green, amino acids 108-139) to stabilize the Vif structure. Vif uses the specific motif ^{120}IRxxL^{124} to directly interact with Cul5. The BC box (green ^{144}SLQYLA^{149}) of Vif mediates an interaction with Elongin C. The PPLP motif (gray, ^{163}PPLPx4L^{169}) is involved in Vif oligomerization. Slanted lines indicate intervening amino acids between the domains. Modified from (*213*).

1.11. Scope and aims of the thesis

My thesis focused on the host restriction factors A3F and A3G. The research projects were planned to carry out a more comprehensive study on the anti-HIV-1 function of A3F in comparison to A3G. The overall goal of my Ph.D. project was to gain greater insights into the anti-viral function of A3F and A3G by understanding the mechanism by which these two enzymes restrict HIV-1. There were two specific aspects that I have focused on; first, biochemical characterization of A3F and A3G and their respective anti-HIV-1 function; and second, how the simultaneous presence of both A3F and A3G affects the overall anti-HIV-1 function of these proteins. The research was undertaken with the hypothesis that *in vivo* functions of A3F can be deduced from studying A3F *in vitro*.

In this Ph.D. research the following aims were addressed:

Aim 1: Biochemical characterization of A3F in comparison to A3G (chapter 2 and chapter 3)

Aim 2: Comparison of anti-HIV-1 function of A3F and A3G (chapter 2 to chapter 5)

Aim 3: Evaluation of deamination independent mode of HIV-1 restriction by A3F and A3G (chapter 4 and chapter 5)

Aim 4: Study on the effects of simultaneous presence of both A3F and A3G on their respective deamination and anti-HIV-1 activity (chapter 4 and chapter 5)

The first part of my project (chapter 2 and chapter 3) revealed that A3F and A3G were distinctively processive enzymes and this influenced their anti-HIV function. The data supported a model in which the processive DNA scanning mechanism of an A3 enzyme could predict its mutagenic potential and suggested key differences in the impact of A3F- and A3G-induced mutagenesis on HIV-1. The second part of my project (chapter 4 and chapter 5) mainly focused on determining if coexpression of A3F and A3G affected their HIV-1 restriction ability. We used a single-cycle infectivity assay and biochemical analyses to determine if coencapsidated A3F/A3G differ in their restriction capacity than A3F or A3G alone. The data suggest that the generation of hetero-oligomeric A3 species may be a unique antiviral strategy of endogenous A3 enzymes.

CHAPTER 2. PREFACE TO CHAPTER 3

Over the last 15 years, the A3 enzymes have been extensively investigated as host-restriction factors against HIV-1 (referred to as HIV). The A3 enzymes attracted the most attention in the APOBEC field because it was assumed that an understanding of these molecules could provide important clues for developing anti-retroviral strategies (*178, 328*). However, to tip the balance in favor of host defenses, more studies were needed to understand the mechanisms of A3 function and their interaction with viral and host proteins. When my graduate research began, the A3 field was characterized by having extensive cell-based studies, and there was a paucity of biochemical data, except for A3G (*205, 227, 329*). A biochemical characterization of A3F had not been conducted and the cell-based virology studies that were available at that time suggested a disparity between A3F and A3G HIV-1 restriction activities, with A3G being more active than A3F (*258, 266, 287, 330*). However, other reports found an equal capacity of A3F and A3G to restrict HIV-1 (*2, 19, 21, 23, 24, 223, 331*). Chapter 3 constitutes the first comprehensive biochemical study that compared A3F with A3G using various synthetic ssDNA substrates. These biochemical and mechanistic studies of A3F identified that an "NPM" amino acid motif present in the interconnecting domain between the N- and C- terminal domains plays a pivotal role in A3F's ssDNA scanning mechanism. Altogether an analysis of the NPM motif was key to revealing why A3F had a different HIV restriction profile than A3G and provided a biochemical explanation for the decreased ability of A3F to inactivate HIV-1. In summary, Chapter 3 suggests a biochemical model to account for cell-based observations and proposes that the processive ssDNA scanning mechanism and the preferred deamination motif of an A3 deoxycytidine deaminases are the major determinants of HIV restriction efficiency (*226*).

CHAPTER 3. DIFFERENT MUTAGENIC POTENTIAL OF HIV-1 RESTRICTION FACTORS APOBEC3G AND APOBEC3F IS DETERMINED BY DISTINCT SINGLE-STRANDED DNA SCANNING MECHANISMS

Anjuman Ara, Robin P. Love and Linda Chelico

Department of Microbiology & Immunology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

This part of the thesis has been published in ***PLOS Pathogens***

Different Mutagenic Potential of HIV-1 Restriction Factors APOBEC3G and APOBEC3F Is Determined by Distinct Single-Stranded DNA Scanning Mechanisms. Ara A, Love RP, Chelico L *PLoS Pathog 10(3): e1004024 (2014).*

All experiments in this chapter were performed by A.A., except Figure 3.16 B-C that were performed by R.P.L. A.A. and L.C. conceived and designed the experiments. A.A., L.C., and R.P.L. analyzed the data. A.A and L.C. wrote the article. A.A., R.P.L. and L.C. revised the article.

3.1. Abstract

The A3 deoxycytidine deaminase family functions as host restriction factors that can block replication of Vif (virus infectivity factor) deficient HIV-1 virions to differing degrees by deaminating cytosines to uracils in single-stranded (-)HIV-1 DNA. Upon replication of the (-)DNA to (+)DNA, the HIV-1 reverse transcriptase incorporates adenines opposite the uracils thereby inducing C/G→T/A mutations that can functionally inactivate HIV-1. Although both A3F and A3G are expressed in cell types HIV-1 infects and are suppressed by Vif, there has been no prior biochemical analysis of A3F, in contrast to A3G. Using synthetic DNA substrates, we characterized A3F and found that similar to A3G, it is a processive enzyme and can deaminate at least two cytosines in a single enzyme-substrate encounter. However, A3F scanning movement is distinct from A3G and relies on jumping rather than both jumping and sliding. A3F jumping movements were also different from A3G. The lack of sliding movement from A3F is due to a ^{190}NPM^{192} motif since insertion of this motif into A3G decreases its sliding movements. The A3G NPM mutant induced significantly less mutations in comparison to wild-type A3G in an *in vitro* model HIV-1 replication assay and single-cycle infectivity assay, indicating that differences in DNA scanning were relevant to restriction of HIV-1. Conversely, mutation of the A3F ^{191}Pro to ^{191}Gly enables A3F sliding movements to occur. Although A3F ^{190}NGM^{192} could slide, the enzyme did not induce more mutagenesis than wild-type A3F demonstrating that the unique jumping mechanism of A3F abrogates the influence of sliding on mutagenesis. Overall, we demonstrate key differences in the impact of A3F- and A3G-induced mutagenesis on HIV-1 that supports a model in which both the processive DNA scanning mechanism and preferred deamination motif (A3F, 5'TTC; A3G 5'CCC) influences the mutagenic and gene inactivation potential of an A3 enzyme.

3.2. Author lay summary

Human cells possess a family of seven DNA-modification enzymes, termed A3, that function as part of our innate immune system. The enzymes modify cytosine in DNA which induces mutations. There are particular enzymes, A3D, A3F, A3G and A3H that appear to be most relevant to restricting HIV replication in $CD4^+$ T cells using this mutagenic mechanism, if they can avoid degradation that is induced by the HIV protein Vif. There has been little biochemical analysis of A3 enzymes other than A3G in terms of the mechanism by which these enzymes search DNA for target cytosines to deaminate. We conducted a biochemical analysis of A3F. We found

that while A3G uses 1-dimensional sliding and 3-dimensional translocations, A3F is restricted to 3-dimensional translocations. This makes the searching mechanism of A3F superficial and detrimental to the induction of a large number of mutations. In addition, gene inactivation was less likely to occur upon deamination of the target motif of A3F (5'TTC) in comparison to the target motif of A3G (5'CCC). All together the data support a model in which the way these enzymes scan DNA can predict the magnitude of mutagenesis induced and the target motif can predict ability to cause gene inactivation.

3.3. Introduction

A3F and A3G are members of a family of seven single-stranded (ss)DNA cytosine deaminases (A3A, A3B, A3C, A3D, A3F, A3G, and A3H) (*201*) and play a role in restriction of the retrovirus HIV-1 (referred to as HIV) (*154*). Research has been highly focused on primarily A3G and secondarily A3F for a number of years since they appeared to be the most efficient restrictors of HIV replication (*2, 21-24*). Although there are documented restrictive effects of A3G, and possibly A3F, at an individual level (reviewed in (*178*), the suppression of HIV by A3G and A3F at a population level is lost due to the HIV protein Vif (viral infectivity factor) (*6, 24*). Vif forms an E3 ubiquitin ligase with host proteins and causes A3G and A3F polyubiquitination and degradation through the proteasome (*15, 24, 332-335*).

The general mechanism by which A3G restricts HIV, which has been a paradigm for other A3 enzymes, requires that it be encapsidated with the ribonucleoprotein complex of HIV (*6, 217*). A3G requires its NTD, which can bind nucleic acids, for encapsidation into virions (*212*). A3G catalyzes deaminations through its CTD (*211, 212*). In the target cells that these virions infect, encapsidated A3G can deaminate cytosines to uracils in (-)DNA reverse transcribed from the RNA genome, after the reverse transcriptase associated RNaseH activity enables ssDNA regions on the (-)DNA to be accessed by the enzyme (*216, 336*). The uracils in the (-)DNA are used as a template by reverse transcriptase during (+)DNA synthesis and result in guanine to adenine mutations. If A3G can induce sufficient numbers of these mutations, the resulting proviral DNA will be functionally inactivated. The deaminases A3D, A3F, and A3H appear to follow this general mechanism of restriction in cell culture, but to differing degrees than A3G (*19, 21, 22, 223, 258, 266, 330*). The exceptions are A3A, which inhibits incoming HIV viral particles in myeloid lineage cells (*180, 206*), A3C, which does not appear to restrict HIV in cell culture, unless it contains a

rare S188I single nucleotide polymorphism (*203, 223, 337, 338*), and A3B, which can restrict HIV in 293T and HeLa cells, but not SupT1 cells (*223, 339*).

Despite a possible role for A3F, A3D and A3H Haplotype II in HIV restriction, it appears that A3G is more effective at restricting HIV replication and that perhaps the other A3 enzymes function in a collaborative way with A3G (*19, 223, 258, 266, 287, 330*). In particular there has been a recent focus on the restriction capability of A3F. A3F was initially identified as potentially being an equal contributor with A3G to the restriction of HIV (*2, 21, 23, 24, 331*), but current research demonstrates, in agreement with an earlier report (*22*), that A3F may have less antiviral activity than A3G (*258, 266, 287, 330*). Many different experimental protocols, such as analysis of stably expressed A3F from a cell line (*258*), use of primary cell lines (*266*), and A3F haplotypes from donor samples (*330*) have been applied and demonstrate that A3F has less of an effect on HIV infectivity in comparison to A3G. However, another report showed no difference in restriction efficiency of A3G and A3F beyond 2-fold using experiments that knocked-down endogenous A3 expression in a nonpermissive cell line (*19*). As a result, the role of A3F in restriction of HIV remains unclear.

Among reports demonstrating less of an effect of A3F in restricting HIV replication than A3G, there is still no identified reason for why this may occur. From some reports A3F mRNA is expressed 10-fold (*31*) or 5-fold (*330*) less than A3G mRNA, suggesting less A3F would become virion encapsidated. However another report found A3F and A3G mRNA expression was more comparable (*18*). Further, some reports have found a direct correlation with mRNA and protein levels (*18, 31*) whereas other reports have been unable to make such a correlation due to the use of different primary antibodies (*266*). Confounding the interpretation of these data are reports which demonstrated that A3F is preferentially encapsidated with the HIV ribonucleoprotein complex in comparison to A3G (*22, 257*). Song *et al*. concluded that the encapsidation difference between A3G and A3F in effect absolves any difference in cellular expression (*257*). Despite this observed more specific packaging of A3F in the ribonucleoprotein complex (*257*), studies have found a minimal contribution of A3F to the hypermutation of HIV genomes or less potency in HIV restriction (*22, 258, 266, 287, 330*). Together these data suggest that if there is a difference in restriction efficiency of A3F and A3G, that it is not the physiological conditions which cause different effects on HIV infectivity, but an inherent difference in their biochemical characteristics. However, there has been no in depth biochemical characterization of A3F to date to determine

what might be these differences between A3G and A3F. As such, we have undertaken a characterization of A3F in comparison to A3G to identify an underlying biochemical reason for these observations.

In particular, we have focused on characterizing the mechanism A3F uses to scan ssDNA. This is because it has been shown that the ssDNA scanning mechanism of A3G is important for inducing mutagenesis of (-)DNA formed during reverse transcription of RNA (*250*). A3G has been characterized to scan ssDNA through facilitated diffusion (*205, 227, 329*). Facilitated diffusion is a 3-dimensional scan of DNA by enzymes to locate their target sites for catalysis (*248, 249, 340*). The movement is characterized by sliding, jumping or intersegmental transfer motions. Sliding is used to describe short range 1-dimensional scanning motions and can enable an in-depth search of a particular area of DNA for a target motif (*249, 340*). Jumping is a term that describes micro-dissociations of the enzyme from the DNA with a reassociation on the same DNA substrate, i.e., the enzyme does not diffuse into the bulk solution (*249, 340*). The negative charge of the DNA establishes a charged radius around the DNA molecule in which a positively charged enzyme can dissociate, diffuse and still return back to the same DNA. These jumping events enable enzymes to translocate larger distances than sliding thus making the search of non-target DNA more efficient than sliding alone (*249, 340*). Intersegmental transfer is similar to jumping but describes a movement where an enzyme with two DNA binding domains interacts with two distal sites simultaneously before dissociating from one of the sites (*249, 340*). Different research groups, including our own, have found A3G to use a combined sliding and jumping search mechanism (*227, 250, 253, 329*), although one report found A3G to use intersegmental transfer (*205*). We have characterized A3G mutants and A3G in complex with different Vif variants that resulted in decreases of either sliding or jumping motions and found that the ability of these A3G forms to induce mutagenesis of nascent reverse transcribed DNA was decreased (*250, 326*). We have hypothesized that both sliding and jumping are important for inducing mutagenesis because A3G needs to conduct local searches (sliding) to effectively deaminate many cytosines, ensuring gene inactivation, and also translocate (jumping) over RNA/DNA hybrids to reach distal regions of (-)DNA (*250*). The processive scanning of other A3 enzymes has not been reported, except A3A which was found to be largely nonprocessive (*341*).

This work is the first biochemical characterization of A3F and provides a biochemical explanation for the lowered ability of A3F to inactivate HIV, as reported by numerous research

groups (*258, 266, 287, 330*) and within this report. We have found that A3F primarily uses jumping movements to scan ssDNA which is detrimental to its ability to cause numerous mutations on (-)DNA during reverse transcription. The target motif of A3G (5'CCC) also appears to cause more inactivating mutations in the HIV *protease* (*prot*) than the target motif for A3F (5'TTC), adding another level of deficiency in HIV inactivation potential. All together our data provide a model for the specific biochemical properties required for efficient restriction of HIV by A3 deaminases.

3.4. Materials and Methods

3.4.1. Protein expression and purification

Recombinant baculovirus production for expression of GST-A3G, GST-A3F (NCBI Accession BC038808), GST-A3G CTD (amino acids 197-380), GST-A3F CTD (amino acids 195-373), GST-A3G NPM, GST-A3F NGM or GST-nucleocapsid protein (NC) in *Sf*9 cells was carried out using the transfer vector pAcG2T (BD Biosciences), as previously described (*227, 247, 250*). Site directed mutagenesis was used to create the A3G NPM and A3F NGM clones. Cloning primers for A3 enzymes and the site directed mutagenesis primers were obtained from Integrated DNA Technologies and are listed in Table 3.1. *Sf*9 cells were infected with recombinant virus at a multiplicity of infection (MOI) of 1, except for GST-A3F and GST-A3F CTD which were infected at an MOI of 2. Recombinant baculovirus infected *Sf*9 cells were harvested after 72 h of infection. Cells were lysed in the presence of RNaseA and the proteins (A3G, A3G NPM, A3G CTD, and NC) were purified as described previously (*247*) to obtain protein that was cleaved from the GST tag and 95% pure. The A3F, A3F NGM, and A3F CTD enzymes were eluted from the glutathione-sepharose resin (GE Healthcare) with the GST tag, as previously described (*227*). The samples were then treated with thrombin (Merck Millipore; A3F and A3F NGM, 0.02 U/µL; A3F CTD, 0.10 U/µL) for 2-5 hours at 21 °C to cleave the GST tag. A DEAE Fast Flow column (GE Healthcare) was then used to purify the A3F, A3F NGM, and A3F CTD from the GST tag and thrombin. The proteins were loaded in low salt buffer containing 50 mM Tris pH 8.0, 50 mM NaCl, 10% glycerol, and 1 mM DTT. A linear gradient from 50 mM NaCl to 1 M NaCl was used to differentially elute the enzymes. The enzymes eluted at approximately 450 mM NaCl and were 90% pure. The SDS-PAGE gels of the purified A3 enzymes are shown in Figure 3.1. Protein fractions were stored at -80 °C. HIV RT (p66/p51) (*342*) was generously provided by Dr. Stuart F.J. Le Grice (NCI, National Institutes of Health).

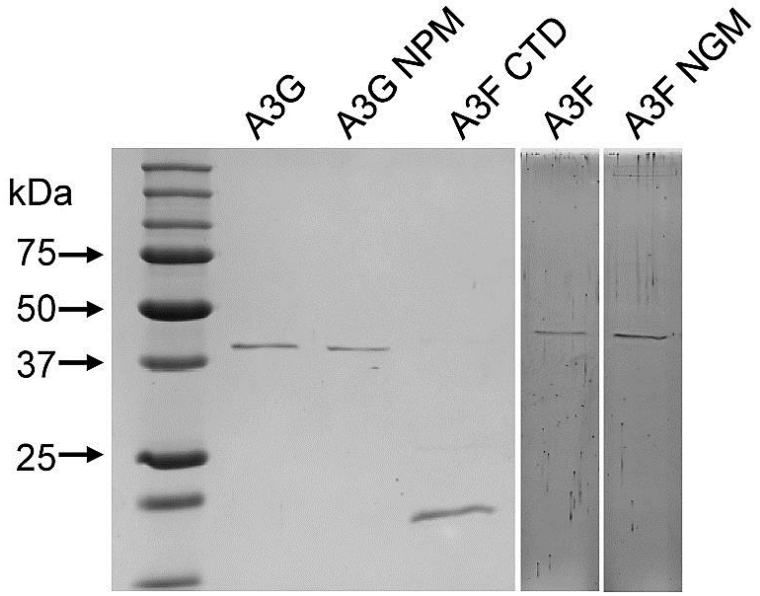

Figure 3.1. Purity of enzymes Coomassie staining (A3G, A3G NPM, A3F CTD) or Bio-Rad Oriole fluorescent stain (A3F, A3F NPM).

3.4.2. Size exclusion chromatography

The oligomerization state of A3 enzymes was determined by subjecting 10-15 µg of the purified enzymes to size exclusion chromatography using a 10 mL Superdex 200 (GE Healthcare) resin bed contained in a column with a 0.5 cm diameter and 16 cm height. The running buffer used was 50 mM Tris pH 7.5 and 200 mM NaCl. The Bio-Rad gel filtration standard set was used to generate a standard curve from which molecular masses and oligomerization states were calculated.

3.4.3. Model HIV-1 replication assay

A3-induced mutagenesis of ssDNA during reverse transcription of an RNA template was measured using an *in vitro* assay, which models reverse transcription from an RNA template and second strand synthesis, and was performed as described previously (*250*). Briefly, a synthetic (+)RNA is synthesized that contains a polypurine tract (PPT), 120 nt of the catalytic domain of the HIV protease (*prot*), and *lacZα* (248 nt). The PPT is used as a primer for (+)DNA synthesis and enables synthesis of dsDNA. The *lacZα* serves as a reporter gene for mutations by blue/white screening. The HIV protease gene was obtained by PCR using clone p93TH253.3 obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Feng Gao and Dr. Beatrice Hahn (*343*). The RNA template (50 nM) was annealed to a 24 nt DNA primer (*250*) and incubated with NC (1.5 µM), RT (1.2 µM), and dNTPs (500 µM) in RT buffer (50 mM Tris pH 7.4, 40 mM KCl, 10 mM $MgCl_2$, 1 mM DTT) in the presence or absence of 200 nM of A3G, A3F, A3G NPM or A3F NGM. Synthesized dsDNA was PCR amplified using *Pfu* C_x Turbo Hotstart (Agilent Technologies) that can use uracils as a template with high fidelity. The amplicons were cloned into a pET-Blue vector backbone that would allow the experimentally synthesized *lacZα* to be used for α-complementation (*250*). At least twenty-five mutated clones for each condition tested were analyzed. DNA sequencing was carried out at the National Research Council of Canada (Saskatoon, Saskatchewan). A t-test was used for statistical analysis of sequences.

3.4.4. Oligonucleotide deamination assays

The ssDNA substrates were obtained from Tri-Link Biotechnologies and are listed in Table 3.1. Deaminations were detected by resolving Fluorescein (F)-labeled DNA that had been treated with Uracil DNA Glycosylase (New England Biolabs) and heated under alkaline conditions on a

10%, 16%, or 20% v/v denaturing polyacrylamide gel, as described previously (*227*). The gel type was determined by fragment sizes produced by each substrate. Reactions were carried out under single hit conditions, i.e., < 15% substrate usage (*344*), to ensure that a single ssDNA substrate was interacting with at most a single enzyme. Under these conditions, a processivity factor can be determined by comparing the total number of deaminations occurring at two sites on the same DNA substrate to a calculated theoretical value of the expected deaminations that would occur at those two sites if the deaminations were not processive (see reference (*227*)). In order to obtain substrate usage within this range under steady-state conditions, the enzyme and DNA concentration were varied based on the enzyme specific activity. For calculating processivity consistently between different experiments we only used reactions where the enzyme activity resulted in 10-12% substrate usage. More ssDNA was used with A3G to ensure clear observation of all deamination bands despite the large preference for the 5'C. However, the data are not altered with ssDNA concentration (data not shown). For A3G and A3G NPM, 30, 40, or 100 nM enzyme was incubated with 300 or 500 nM fluorescein (F)-labeled ssDNA. For A3F, A3F NGM, and A3F CTD, 100 nM enzyme was incubated with 50 or 100 nM F-labeled ssDNA. For A3G CTD, 1000 nM enzyme was incubated with 500 nM F-labeled ssDNA. Reactions were incubated at 37°C for 1 - 50 min. Gel pictures were obtained using a Typhoon Trio (GE Healthcare) multipurpose scanner and analysis of integrated gel band intensities used ImageQuant software (GE Healthcare). The specific activity was calculated from single-hit condition reactions by determining the picomoles of substrate used per minute for a microgram of enzyme.

Table 3.1. Primers and DNA substrates

A3F Forward	ATG AAG CCT CAC TTC AGA AAC AC
A3F Reverse	TCA CTC GAG AAT CTC CTG CAG
A3F CTD Forward	ATG TAT CCA CAC ATA TTC TAC TTC
A3F CTD Reverse	TCA CTC GAG AAT CTC CTG CAG CTT G
A3G NPM SDM Forward	GGG GGA GAT TCT CAG AAA CCC GAT GCA CTC GAT GGA TCC ACC
A3G NPM SDM Reverse	GGT GGA TCC ATC GAG TGC ATC GGG TTT CTG AGA ATC TCC CCC
A3F NGM SDM Forward	CTA AAG GAG ATT CTC AGA AAC GGC ATG GAG GCA ATG TAT CCA CAC
A3F NGM SDM Reverse	GTG TGG ATA CAT TGC CTC CAT GCC GTT TCT GAG AAT CTC CTT TAG
A3G Stop SDM Forward	CAG AAT CAG GAA AAC TGA TCT AGA TTT TAC CCA
A3G Stop SDM Reverse	TGG GTA AAA TCT AGA TCA GTT TTC CTG ATT CTG
A3F Stop SDM Forward	CAG GAG ATT CTC GAG TGA AAG GGT CAA GAC AAT TCT G
A3F Stop SDM Reverse	CAG AAT TGT CTT GAC CCT TTC ACT CGA GAA TCT CCT G
Prot Forward	GAC AAG GAA CTG TAT CCT TTA GCT T
Prot Reverse	CTG GTA CAG TCT CAA TAG GAC TAA T
A3G – 5	AAA GAG AAA GTG ATA CCC A(dT-FAM)A CCC ATA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
A3F – 5	AAA GAG AAA GTG ATA TTC A(dT-FAM)A TTC ATA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
A3G – 14	AAA GAG AAA GTG AGA CCC AAA GAA (dT-FAM)GA AGA CCC AAA TGT TAG AAT TGT TAA TGT GTG TGA TGA TGT TGA
A3F – 14	AAA GAG TTA GTG AGA TTC AAA AT T (dT-FAM)AG AGA TTC AAA TGT TAG ATATGT TAA TGT GTG TGA TGA TGT TGA
A3G – 30	AAA GAG AAA GTG ATA CCC AAA GAG TAA AGT (dT-FAM) AGA TAG AGA GTG ATA CCC AAA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
A3F – 30	AAA GAG AAA GTG ATA TTC AAA GAG TAA AGT (dT-FAM) AGA TAG AGA GTG ATA TTC AAA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
A3G – 63	GAA TAT ATG TTG AGA CCC AAA GTA ATG AGA GAT TGA (dT-FAM)TAG ATG AGT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA CCC AAA GAG TAA AGT TGT TAA TGT GTG TAG ATA TGT TAA
A3F - 63 (TTC)	GAA TAT AGT TTT TAG TTC AAA GTA AGT GAA GAT AAT (dT-FAM) TAG AGA GTT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA TTC AAA GAG TAA AGT TGT TAA TGT GTG TAG ATA TGT TAA

Table 3.1 (continued). **Primers and DNA substrates**

A3F - 63 (ATC)	GAA TAT ATG AGT TGA ATC AAA GTA ATG AGA GAG AAT (dT-FAM) TAG ATG AGT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA ATC AAA GAG TAA AGT TGT TAA TGT GTG TAG ATA TGT TAA
Complementary RNA to A3F/A3G – 63	rCrUrU rUrCrA rUrArC rArCrA rUrArU rArUrC rArC
Trap DNA	AAA GAG AAA GTA ATA AGG AAA GAG TAA AGT ATA TTC AAA TAA ACA ATC ATT CTA CAC ATT CAT ACA ATT
A3G – 100	GGA GAT AGA TTA GAA TAC CC A AAA ATG AAT AAA AAG TGT AGT TGA ATG TAG AAA AGT GGT TAT TGA ATG ATA AGG ATG GAT GGA A (dT-FAM)G ATA TGA AAT GGA GAT AGT GTA GAT GAA AAG ACC CAA AAT GTA GTA AGT AGT TTA AGA ATA GGA GAG TAG T
A3F – 100	GGA GAT AGA TTA GAA TA TTC A AAA AAATAT AAA AAG TGT AGA ATG GTG TAG AAA AGT GGT TAT TGA ATG ATA AGG ATG GAT GGA A (dT-FAM)G ATA TGA AAT GGA GAT AGT GTA GAT GTT AAG ATT CAA AAT GTA GTA AGT AGT TTA AGA ATA GGA GAG TAG T

3.4.5. Steady state rotational anisotropy assay

Steady state fluorescence depolarization (rotational anisotropy) was used to measure enzyme-ssDNA binding affinities using the same F-labeled ssDNA substrates (with cytosines 63 nt apart) that were used for deamination reactions (Table 3.1). Reactions were 60 µL and contained F-labeled ssDNA (10 nM) in RT buffer and A3G (0 - 650 nM), A3F (0 - 80 nM), A3F CTD (0 - 600 nM), A3G NPM (0 - 350 nM), or A3F NGM (0 - 650 nM) were titrated into the reaction. A QuantaMaster QM-4 spectrofluorometer (Photon Technology International) with a dual emission channel was used to collect data and calculate anisotropy. Measurements were made at 21°C. Samples were excited with vertically polarized light at 495 nm (6 nm band pass) and vertical and horizontal emissions were measured at 520 nm (6 nm band pass). Apparent dissociation constants (K_d) were obtained by fitting to a sigmoidal curve using Sigma Plot 11.2 software.

3.4.6. Single-cycle infectivity assay

VSV-G pseudotyped HIV pNL4-3 ⊿*vif* viruses were produced by transfecting 3×10^5 293T cells per well in a 6-well plate with Qiagen Polyfect reagent. Specifically, transfections used 1100 ng of pHIV⊿*vif* (*284*), which expresses an eGFP reporter gene and 630 ng of pLTR-G (Addgene), which expresses the VSV-G protein, in the presence or absence of 220 ng of A3G, A3F or A3F NGM or 350 ng of A3G NPM in pcDNA3.1. The transfections used empty pcDNA3.1 to achieve equivalent amounts of DNA. The cotransfection molar ratio of A3 enzymes in pcDNA3.1 to the pNL4-3 ⊿*vif* was 0.33:1 (A3G, A3F, or A3F NGM) or 0.59:1 (A3G NPM). The A3G (cat# 9952) and A3F (cat # 10100) expression plasmids were obtained from the NIH AIDS Reagent program with C-terminal tags. A stop codon was introduced immediately after the A3G or A3F coding sequence to enable expression of native A3 enzymes. The amino acid sequence of the A3G and A3F clones were identical to those used in biochemical assays. Subsequently, site directed mutagenesis was used to create the A3G NPM and A3F NGM clones. The site directed mutagenesis primers were obtained from Integrated DNA Technologies and are listed in Table 3.1. Sixteen hours after the transfection, the cells were washed with PBS and the medium replaced. Virus-containing supernatants were collected 48 hours after the media change and filtered through 0.22 µm syringe filters. Virus was quantified by a p24 enzyme-linked immunosorbent assay (QuickTiter™ Lentivirus Titer Kit, Cell Biolabs Inc.). Target 293T cells were infected at an MOI of 0.5 by spinoculation at 800 x *g* for 1 h in the presence of 8 µg/ml of polybrene (*345*). Infection

levels in 293T cells was determined by flow cytometry by detecting eGFP fluorescence at 48 hours post infection and data were normalized to HIV$\triangle$*vif* infections in the absence of A3 enzymes.

3.4.7. Sequencing of integrated proviral DNA

Infected 293T cells were harvested after 48 h and the DNA was extracted using the Qiagen DNeasy Blood and Tissue kit. DNA was treated with DpnI (New England Biolabs) to remove possible contaminating plasmid DNA and the *prot* (nt 2280-2631) sequences were amplified by PCR using Phusion High Fidelity Polymerase (New England Biolabs). Primers were obtained from Integrated DNA and are listed Table 3.1. PCR products were purified and cloned with the Zero Blunt TOPO PCR cloning kit (Invitrogen). DNA sequencing was carried out at the National Research Council of Canada (Saskatoon, Saskatchewan).

3.4.8. Quantitative immunoblotting

The A3G and A3F enzymes were detected in cell lysates (40 µg total protein) and virions (130 ng of p24) used for single-cycle infectivity assays using antibodies to the native enzymes. For A3G we used the ApoC17 rabbit antiserum (Cat # 10082, NIH AIDS Reagent Program) and for A3F we used the C-18 polyclonal rabbit antibody (Cat # 11474, NIH AIDS Reagent Program). Loading controls for cell lysates (α-tubulin, Sigma) and virions (p24, Cat #3537, NIH AIDS Reagent Program) were detected using mouse monoclonal antibodies. Proteins of interest and loading controls were detected in parallel on the same gel by using the Licor/Odyssey system (IRDye 680-labeled goat anti-rabbit secondary antibody and IRDye 800-labeled goat anti-mouse secondary antibody). Visualization with an Odyssey Infrared Imaging System (Licor) and analysis of bands with Odyssey software enabled intensities of bands to be determined. Analysis of a titration of purified A3G and A3F with their respective antibodies showed that A3F detection was 9-fold less sensitive than A3G detection at a 1/1000 antibody dilution. Further, doubling the amount of antibody to A3F (1/500) resulted in a doubling of the A3F detection sensitivity in comparison to the antibody to A3G (1/1000). Therefore, an appropriate correction factor for the antibody dilution was used to adjust the integrated band intensities of A3F to enable comparison with A3G. Antibodies were used at a dilution of 1/1000 except for A3F or A3F NGM containing cell lysates which required a dilution of 1/500 for detection of A3F or A3F NGM. A t-test was used for statistical analysis.

3.5. Results

3.5.1. A3F and A3G distinctively scan ssDNA

The processive nature of A3G has been shown to be of importance for inducing mutagenesis of HIV (-)DNA in a model *in vitro* system (*250*, *326*) and in cell culture (*254*). It is not known whether A3F is processive. Since multiple lines of evidence from independent labs have shown that the effect of A3F on HIV is different than A3G (*22*, *258*, *266*, *330*), we sought to determine if there was an inherent biochemical difference between the two enzymes that could account for these observations. Specifically, we determined if there was a difference in the processive scanning mechanisms of these two enzymes with processivity being defined as the ability to deaminate more than one cytosine on an ssDNA in a single-enzyme substrate encounter. Processivity was determined using different synthetic ssDNA substrates containing two deamination motifs separated by different distances, 5'TTC for A3F and 5'CCC for A3G. This strategy was used since with A3G we have found that closely spaced deamination motifs, i.e., 5 to 15 nt are deaminated most efficiently through sliding motion and as the distance between deamination motifs increases a jumping motion facilitates processive deaminations (*250*). The substrate usage was kept below 15% to ensure single-hit conditions were maintained, which means that each ssDNA was only encountered by an enzyme at most once during the reaction (*344*).

On a substrate with the target cytosines separated by 30 nt (Figure 3.2, sketch), A3F was able to catalyze processive deaminations. The processivity factor is a ratio of the frequency of double deaminations on a single substrate to the predicted frequency of double deaminations of a nonprocessive enzyme (see Materials and Methods). Therefore, the processivity factor of 3.7 for A3F (Figure 3.2A) means that in a single enzyme-substrate encounter A3F was 3.7-fold more likely to catalyze a processive deamination than a nonprocessive deamination. On the cognate A3G substrate, A3G was 2-fold more likely than A3F to catalyze a processive deamination (compare Figure 3.2A, A3F, processivity factor of 3.7 and A3G, processivity factor of 7.9), suggesting that the processive mechanisms of A3F and A3G differ. In addition, we observed a difference in the ability of A3F and A3G to catalyze 5'-end biased deaminations. Where A3G has been found to prefer deaminations towards the 5'-end of ssDNA molecules due to a catalytic orientation specificity (*227*), A3F had a minimal 5'-end bias (Figure 3.2A, compare intensity of 5'C & 3'C bands for A3F and A3G). However, the presence or absence of a 5'-end bias does not influence the processivity calculation (*252*). Since A3G has been found to use a dual sliding and jumping

motion to scan ssDNA (*227, 250, 329*), we investigated whether the difference between A3F and A3G was due to a difference in the contributions of sliding and jumping or a different mode of scanning, e.g., intersegmental transfer.

First, we investigated the sliding ability of A3F. We conducted deamination assays on ssDNA substrates with closely spaced deamination targets, since it has been shown that sliding motions increase the frequency of closely spaced deaminations occurring processively (*250*). With cytosines 14- and 5-nt apart, A3F was unable to catalyze any detectable processive deaminations (Figure 3.2B-C, A3F, absence of 5'C & 3'C band) indicating that A3F does not use sliding motions to catalyze processive deaminations. Of note, outside of single hit conditions (>15% substrate usage) we detected the band corresponding to deamination of both 5'TTC motifs on an ssDNA (5'C & 3'C band), which demonstrated that multiple molecules of A3F were able to deaminate these substrates at both cytosine targets to near completion (data not shown). In contrast, A3G was able to processively deaminate closely spaced residues under single-hit conditions by sliding (Figure 3.2B-C, A3G, processivity factors of 4.6 and 3.5).

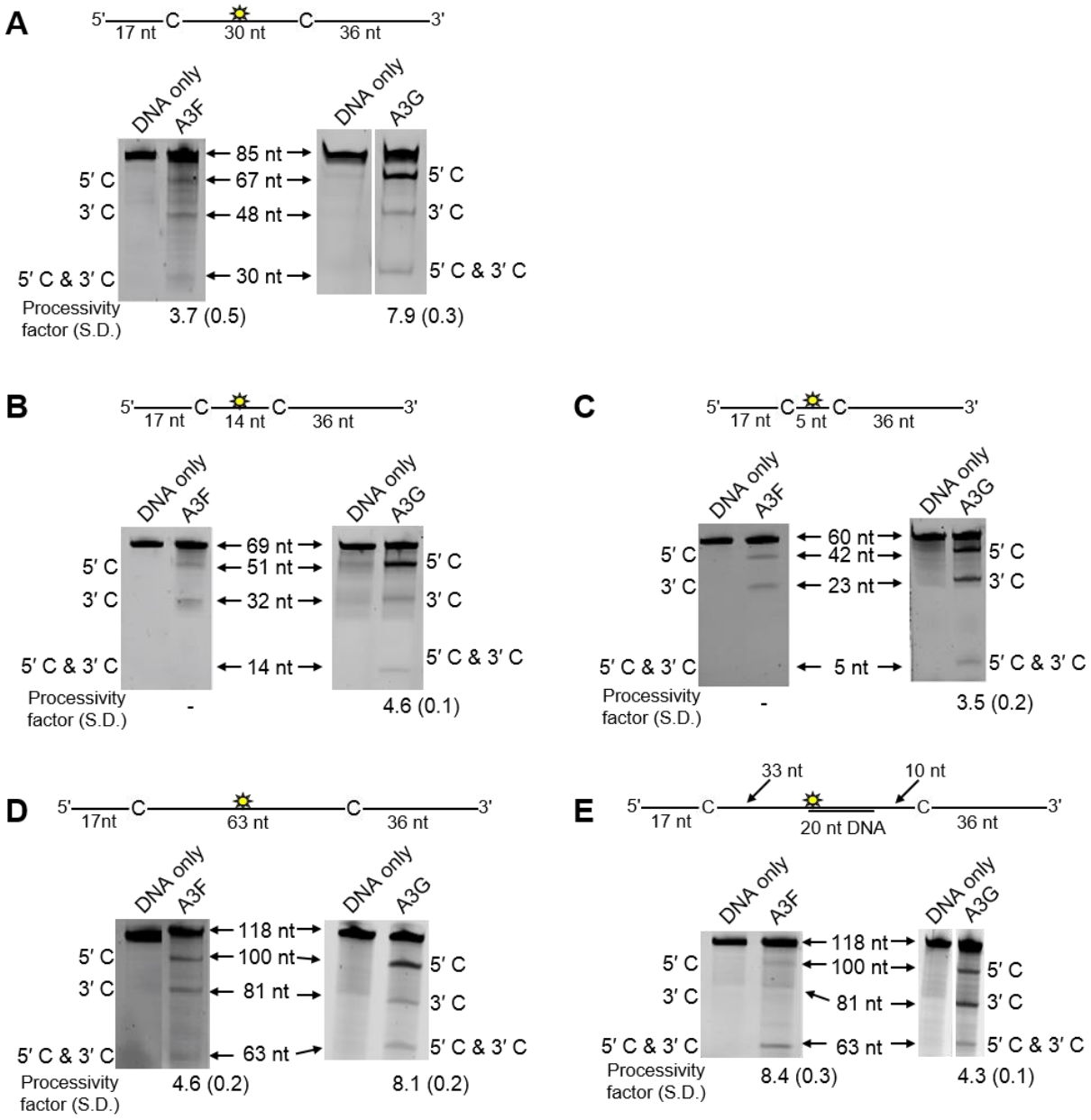

Figure 3.2. A3F and A3G are distinctively processive Processivity of A3F and A3G were tested on substrates that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by different distances. The A3F substrates had 5'TTC motifs and the A3G substrates had 5'CCC motifs. (A) The two target cytosines within the 85 nt ssDNA sequence are spaced 30 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 30 nt labeled fragment (5'C & 3'C). (B) The two target cytosines within the 69 nt ssDNA sequence are spaced 14 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 14 nt labeled fragment (5'C & 3'C). (C) The two target

cytosines within the 60 nt ssDNA sequence are spaced 5 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment (5'C & 3'C). (D) The two target cytosines within the 118 nt ssDNA sequence are spaced 63 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). (E) Deamination of the substrate described for (D), but with a 20 nt ssDNA annealed between the two target cytosines to block the sliding component of processivity. The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. The A3F: DNA ratio was 2:1 except for panel (A) in which a 1:1 ratio was used. The A3G: DNA ratio was (A-B) 1:10, (C) 1:2.5, (D-E) 1:20. Enzyme: DNA ratios were varied due to different specific activities of the enzyme on a given DNA substrate. Values are an average from three independent experiments.

Since A3F was processive on the substrate with the target cytosines separated by 30 nt (Figure 3.2A, A3F), but not on substrates with closely spaced deamination motifs (Figure 3.2B-C, A3F), the data suggested that A3F may use jumping or intersegmental transfer to processively deaminate cytosines. To investigate this further we determined the processivity of A3F on an ssDNA with the target cytosines separated by 63 nt (Figure 3.2D). On this substrate, A3F exhibited a processivity factor of 4.6 (Figure 3.2D, A3F), which is higher than the processivity factor obtained on the substrate with the target cytosines separated by 30 nt (Figure 3.2A, A3F, processivity factor of 3.7). In contrast, A3G which can slide and jump (*227*) maintained a processivity factor of ~8 (compare Figure 3.2A and D, processivity factors). To confirm that we would observe only jumping or intersegmental transfer and not sliding motions, we annealed a 20 nt complementary DNA in between the target cytosines (Figure 3.2E, sketch). The double-stranded DNA portion is not bound as tightly by A3F (Figure 3.3A-B) or A3G (Figure 3.3C and (*5, 227, 329, 346*) as ssDNA (Table 3.2) and results in the assay conditions blocking the sliding portion of the scanning activity (*227, 252*). A3G was still processive on this substrate due to the ability to translocate on DNA in 3-dimensions by jumping, but we observed a ~2-fold decrease in A3G processivity as compared to the analogous ssDNA substrate (Figure 3.2D-E, compare A3G processivity factors). We interpret that the ~2-fold decrease in A3G processivity is due to A3G molecules attempting to slide over the dsDNA which induces dissociation from the DNA substrate and diffusion into the bulk solution. For A3F we observed a 1.8-fold increase in the processivity factor when we annealed a 20 nt complementary DNA in between the target cytosines (Figure 3.2D-E, compare A3F processivity factors), despite having a reduced binding to the double-stranded (ds)DNA portion (Figure 3.3A-B). The double deaminations became so efficient that the 5'- and 3'-proximal cytosine deaminations were barely visible on the gel (Figure 3.2E, A3F). A3F bound the 118 nt ssDNA substrate (Figure 3.2D) with an apparent K_d of 20 nM (Table 3.2), which is ~7-fold lower than the apparent K_d of A3G (Table 3.2, K_d of 130 nM). This indicates that A3F is less likely to dissociate from an ssDNA substrate than A3G, but does not fully explain why we observed an increase in processivity by annealing a complementary DNA in between the target cytosines. Results were not changed by annealing a 20 nt complementary RNA molecule to the substrate (Figure 3.4A-B) or by testing A3F on a different partially dsDNA substrate which contained two 5'ATC motifs (Figure 3.5). We speculated that the processivity of A3F increased as opposed to remaining the same in the presence of the complementary DNA because the structural

change in the substrate induced by the dsDNA region made jumping events more successful. This could occur if the average jumping distance of A3F were different than A3G and the rigid dsDNA region juxtaposed the 5'TTC motifs at a distance which was highly accessible by A3F.

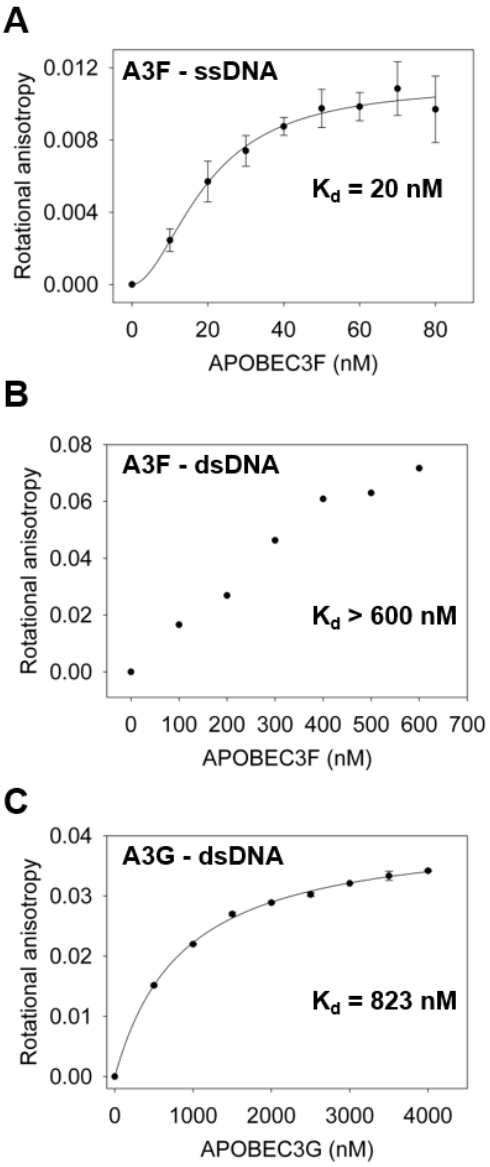

Figure 3.3. Binding affinities of A3F and A3G for single-stranded (ss) DNA or double-stranded (ds) DNA. A3F and A3G binding to fluorescein labeled DNA (10 nM) was monitored with rotational anisotropy. (A) ssDNA as shown in Figure 3.2D was used as a substrate. A3F binds this ssDNA with a high affinity (apparent K_d of 20 ± 1 nM). (B-C) The double stranded region (20 nt) created in Figure 3.2E was used as a binding substrate for (B) A3F or (C) A3G. (B) A3F was unable to bind the dsDNA to saturation in a concentration range similar to ssDNA. We were unable to concentrate A3F sufficiently to titrate in the necessary amount to saturate the dsDNA substrate. The apparent K_d is estimated to be >600 nM. (C) A saturation curve for A3G binding to dsDNA is shown for comparison. A3G binds the dsDNA with an apparent K_d of 823 ± 11 nM. Values are an average from at least two independent experiments.

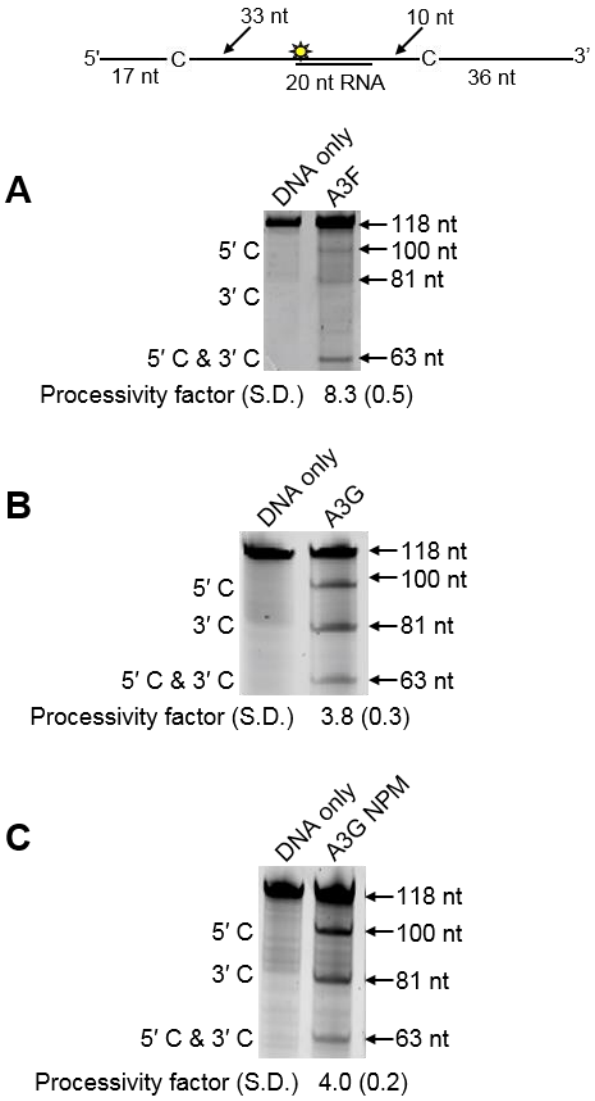

Figure 3.4. Processivity of A3F, A3G, and A3G NPM in the presence of a 20 nt RNA/DNA hybrid. Deamination was tested on a 118 nt ssDNA substrate that contained an internal fluorescein (F)-label and two deamination motifs separated by 63 nt (sketch). A 20 nt complementary RNA was annealed between the two deamination motifs. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). (A) A3F, (B) A3G, and (C) A3G NPM are able to processively deaminate the target cytosines by transversing the RNA/DNA hybrid region. A3F is 2-fold more processive than A3G and A3G NPM on this substrate. The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. The A3F: DNA ratio was 2:1 and the A3G: DNA and A3G NPM: DNA ratios were 1:20. Enzyme: DNA ratios were varied due to different specific activities of the enzyme on a given DNA substrate. Values are an average from three independent experiments.

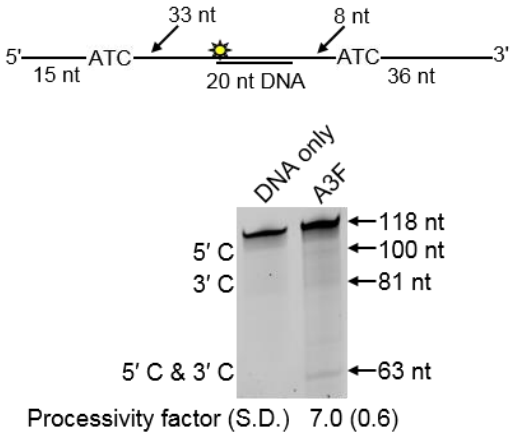

Figure 3.5. Processivity of A3F, A3G, and A3G NPM in the presence of a 20 nt ds DNA region and 5'ATC deamination motifs. Deamination was tested on a 118 nt ssDNA substrate that contained an internal fluorescein (F)-label and two 5'ATC deamination motifs separated by 63 nt (sketch). A 20 nt complementary DNA was annealed between the two deamination motifs. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). A3F is able to processively deaminate the target cytosines by transversing the dsDNA region. The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. The A3F: DNA ratio was 2:1. Values are an average from three independent experiments.

Table 3.2. Comparison of apparent dissociation constants (K_d) from ssDNA of A3G and A3F wild-type and mutants

Enzyme	K_d (nM)
A3G	130 ± 6
A3F	20 ± 1
A3F CTD	288 ± 10
A3G NPM	56 ± 4
A3F NGM	119 ± 11

To test this hypothesis, we examined the processive deaminations of A3F and A3G on an ssDNA substrate with deamination motifs separated by 100 nt. We found that as the distance between deamination motifs was increased up to 100 nt, the processivity factors of A3F also increased (Figure 3.6A). In contrast, A3G processivity exhibited a plateau when deamination motifs were 30- to 63-nt apart and the processivity factor decreased when deamination motifs were 100 nt apart (Figure 3.6B). These data demonstrate that the average jumping distance of A3F and A3G differ. Similar results were also found from analysis of deamination-induced mutations in the model HIV replication assay and are discussed later in the text (Table 3.4). To identify a possible reason for the different jumping ability of A3F we examined its oligomerization state in comparison to A3G. A3G is known to form polydisperse oligomers that are dependent on enzyme concentration and buffer conditions (*347*). Using size exclusion chromatography at low enzyme concentrations we found that A3F formed predominantly tetramers (~180 kDa) and higher order oligomers whereas A3G eluted as predominantly a monomer (~ 46 kDa) with minor dimeric species (Figure 3.6C). The finding that A3F forms more tetramers than A3G is consistent with previous sucrose gradient data (*348*) and data on the CTD portions of these enzymes. The A3F CTD can oligomerize more readily than the A3G CTD (*247, 263, 349, 350*). The A3F oligomers remained soluble as high speed centrifugation did not result in a discernable protein pellet. These data demonstrated that A3F oligomers are more stable than A3G oligomers at low protein concentration and suggest a structural difference that could account for why the A3F jumping distance is different than A3G (Figure 3.6A-B).

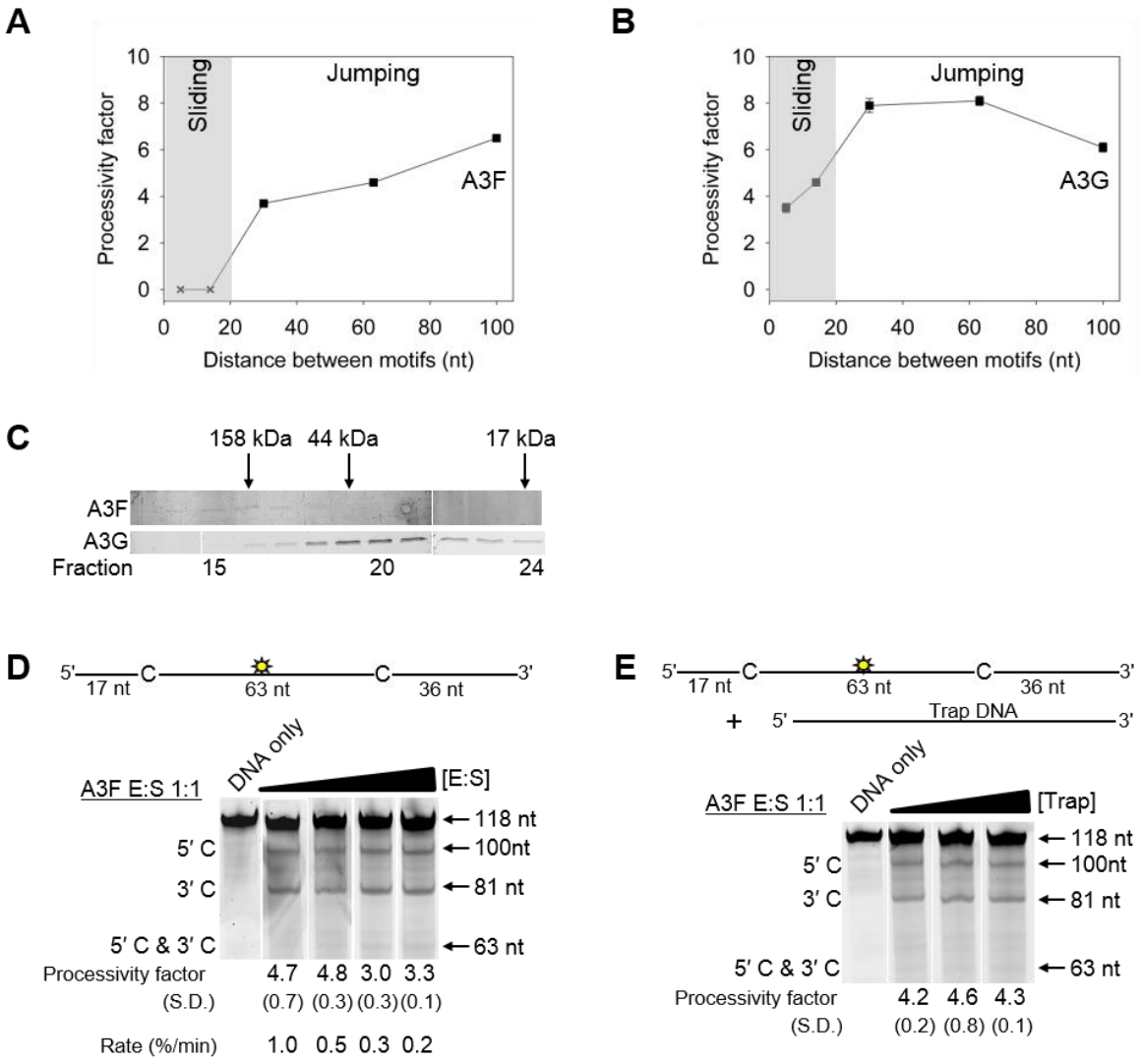

Figure 3.6. A3F translocations on ssDNA are distinct from A3G, but not due to intersegmental transfer. (A-B) Summary of processivity factors for (A) A3F and (B) A3G on ssDNA substrates where the two deamination motifs were separated by 5- to 100- nt. A3F processivity was not observable until the distance between cytosines was greater than 14 nt apart (denoted with x), but then increased until 100 nt (filled squares). This was distinct from A3G that was processive when cytosines were closely spaced (5- to 14- nt apart), reached a maximum processivity when cytosines were 30- to 63- nt apart and then decreased in processivity (filled squares). The grey area represents the region where sliding is required for processivity. Jumping has been previously defined to be translocations of ≥ 20 nt (*351*). Gels for the substrate with deamination motifs separated by 100 nt are shown in Figure 3.7. (C) Size exclusion chromatography demonstrates that A3F forms tetramers (~180 kDa) and higher order oligomers. This is in contrast to A3G which forms monomers (~46 kDa) and dimers. (D-E) Processivity of A3F was tested on a substrate that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs (5'TTC) separated by 63 nt. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled

100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). (D) The enzyme: substrate ratio of 1:1 (E:S) was kept constant, but reaction components increased (100:100, 200:200, 300:300, 400:400 nM) to investigate whether A3F could transfer between two ssDNA substrates. (E) In the presence of an unlabeled ssDNA trap (69 nt) the processivity factor of A3F (E:S of 1:1) remained the same regardless of trap concentration (1:0.5, 1:1, or 1:5 ratio of labeled ssDNA to unlabeled trap ssDNA). The measurements of processivity (Processivity factor), Standard Deviation of the mean (S.D.), and Rate (%/min) are shown below the gels. Values are an average from at least two independent experiments.

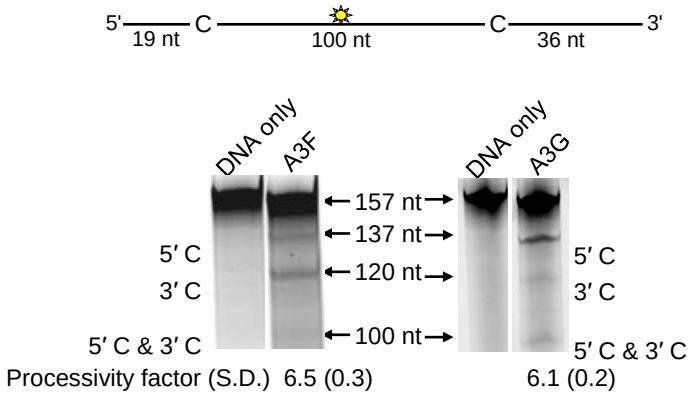

Figure 3.7. Processivity of A3F and A3G on a substrate containing deamination motifs 100 nt apart. Deamination was tested on a 157 nt ssDNA substrate that contained an internal fluorescein (F)-label and either two 5'TTC (A3F) or 5'CCC (A3G) deamination motifs (sketch). Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 137- and 120- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 100 nt labeled fragment (5'C & 3'C). A3F (left) and A3G (right) are able to processively deaminate the target cytosines. The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. The A3F: DNA ratio was 1:1 and the A3G: DNA ratio was 1:20. Values are an average from three independent experiments.

However, since the A3F ssDNA scanning mechanism is more efficient in distal translocations (Figure 3.2), we also investigated whether it was scanning ssDNA by intersegmental transfers, rather than or in addition to jumping. This mode of DNA scanning involves an enzyme molecule that binds in two distal locations on the DNA before completing the translocation by dissociating from one location (*249, 340*). The intersegmental transfer mechanism requires that the enzyme have more than one DNA binding domain. A3F could bind ssDNA with both its NTD and CTD on one or many subunits of the oligomer. This is in contrast to jumping which uses microdissociations and reassociations to scan ssDNA (*249, 340*). A key difference between jumping and intersegmental transfer is that the probability of an enzyme transferring to another DNA substrate is low for jumping but high for intersegmental transfer (*249, 340, 352*). Therefore, to observe whether A3F can scan ssDNA by intersegmental transfer we increased the enzyme and substrate concentrations, but kept their ratio constant. Crowding the reaction in this manner with enzyme and ssDNA can increase the tendency of the enzyme to translocate to a different ssDNA if intersegmental transfer is occurring (*352*). This would result in a decrease in the observed processivity with increasing reaction components. We found that A3F maintained the same processivity at a 1:1 E:S ratio at concentrations of 100 nM and 200 nM (Figure 3.6D, processivity factor of 4.7 and 4.8). At a 1:1 E:S ratio using concentrations of 300 nM and 400 nM the processivity of A3F decreased ~1.5-fold from 4.7 to 3.0 or 3.3 (Figure 3.6D), providing evidence that A3F can use intersegmental transfer to scan ssDNA. However, the decrease in A3F processivity is small (~1.5 fold), does not decrease gradually with increasing enzyme and substrate concentration, and is not completely abolished (processivity factor remains above 1) suggesting that intersegmental transfer is not the primary mechanism of DNA scanning, but can occur in a minority of ssDNA-A3F interactions. Importantly, intersegmental transfer should result in an increase in the reaction rate with increasing DNA concentration since the rate of searching is enhanced by increasing the apparent off rate, which allows more rapid sampling of DNA (*352*). However, the reaction rate of A3F decreased with increasing enzyme and substrate concentrations (Figure 3.6D, Rate) and supports the conclusion that intersegmental transfer is not a primary mode of scanning ssDNA. In further support of this interpretation is that we only observed evidence of intersegmental transfer with increasing enzyme and substrate concentration (Figure 3.6D), not when the ssDNA concentration alone was increased (Figure 3.6E, processivity factors of 4.2 to 4.6), which indicates that A3F does not readily transfer to another ssDNA without high local

concentrations of enzyme, i.e., the intersegmental transfer is not inherent to A3F but requires excessive crowding of reaction conditions. A3G showed no decrease in processivity with increasing concentration of enzyme and substrate, despite also containing both a NTD and CTD (Figure 3.8). This difference may arise since the CTD of A3G binds ssDNA in the micromolar range (*247, 350, 353*), in contrast to the CTD of A3F that can bind DNA in the nanomolar range (Table 3.2, apparent K_d of 288 nM). All together the data supported the conclusion that A3F primarily utilized jumping and not intersegmental transfer to scan ssDNA.

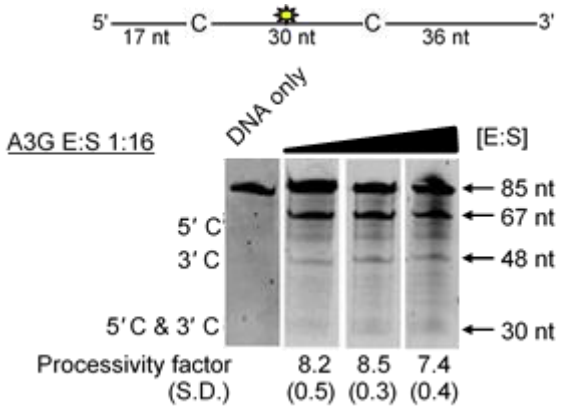

Figure 3.8. Increasing the total concentration of enzyme and substrate does not decrease the processivity of A3G. Deamination was tested on an 85 nt ssDNA substrate that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by 30 nt (sketch). Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 30 nt labeled fragment (5'C & 3'C). The processivity of A3G was not significantly changed when the enzyme: substrate (E:S) ratio (1:16) was kept constant, but reaction components increased (3: 50 nM, 30: 500 nM, 60: 1000 nM). The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. Values are an average from three independent experiments.

3.5.2. The A3F DNA scanning mechanism does not enable efficient mutagenesis of (-)DNA

Our biochemical data on synthetic substrates (Figure 3.2) predicts that A3F will not efficiently catalyze deaminations during proviral DNA synthesis due to a predominant jumping movement that would result in a superficial scan of the ssDNA (*250, 326*). Importantly, we observed this predominant jumping movement when A3F encountered an RNA/DNA hybrid (Figure 3.4A), such as would be encountered during synthesis of the HIV provirus. To test this prediction we used our model *in vitro* HIV replication system. Since this system reconstitutes reverse transcription of (-)DNA and synthesis of (+)DNA, it allows us to observe the ability of A3 enzymes to induce mutagenesis in a dynamic system, such as occurs *in vivo,* but with the advantage of controlling the amount of enzyme added to the reaction system. Specifically, this system uses an *in vitro* synthesized RNA which contains (from the 5'-end) a polypurine tract (PPT), part of the protease gene (*prot*) of HIV, and a *lacZα* reporter. The RNA is reverse transcribed to (-)DNA by reverse transcriptase and after RNaseH-mediated removal of the RNA, the PPT enables (+)DNA synthesis without the addition of an exogenous primer. In this manner we can achieve the salient properties of HIV replication that A3 enzymes must contend with, a finite time to access single-stranded (-)DNA and a heterogeneous substrate that is interspersed with RNA fragments.

The A3G data demonstrated the potential amount of mutations that could occur in this system. A3G had a clonal mutation frequency of 2.63×10^{-2} mutations/bp which is 10-fold over the background mutation frequency of reverse transcriptase (RT) (Table 3.3).

Table 3.3. A3-mediated mutation frequencies in a model HIV replication system.

Enzyme	Population mutation frequency	Base pairs sequenced	Total number of G→A mutations	Clone mutation frequency (x 10^{-2}mutations/bp)
A3G	0.89	11 040	290	2.63
A3F	0.61[a]	10 304	78	0.76[a]
A3G NPM	0.14[a]	9200	27	0.29[a]
A3F NGM	0.80	10 672	36	0.34 [a]

The ratio of white colonies to total colonies is defined as the population mutation frequency. The average number of G→A mutations per base pair in the 368 nt *prot-lacZα* construct is defined as the clone mutation frequency. For RT alone (no A3), all base changes were used to calculate the clone random mutation frequency which established a baseline of 0.27 x10^{-2} mutations/bp. The RT alone condition produced a population mutation frequency of 0.07.

[a] Significant difference was $p \leq 0.001$ versus A3G values.

[b] Significant difference was $p \leq 0.01$ versus A3G values.

[c] Significant difference was $p \leq 0.05$ versus A3G values.

Further, the A3G mutation spectra have clear hot-spots at 5'CCC or 5'CC motifs in both the *prot* and *lacZα* with some sites being mutated in 100% of clones (Figure 3.9A, e.g., 245 nt). Due to the PPT being nearest the *prot*, this region is converted to dsDNA the fastest and incurs less mutations than regions nearer the center or 3'-end of the (+)DNA (Figure 3.9A). As such, we can recover white colonies indicating a mutation in the *lacZα* reporter but upon sequencing find no mutations in the *prot*. Therefore, the number of clones with mutations in the *prot* is a measure of how efficiently an A3 enzyme can induce mutations. The *lacZα* remains single stranded longer and can therefore be visited by multiple A3 enzymes multiple times. In the *prot* region, A3G was found to induce no mutations in 13% of clones, but the majority of clones had either 1-2 mutations (47%) or 3-4 mutations (37%) (Figure 3.9C). In the *lacZα*, A3G-induced mutagenesis resulted in >7 mutations in the majority of clones (Figure 3.9D, 60%).

Addition of A3F to the model HIV replication assay resulted in a modest 2.8-fold increase over the background mutation frequency (Table 3.3). Examination of the mutation spectrum demonstrated that A3F could induce mutagenesis at a number of 5'TTC or 5'TC sites along the *prot* and *lacZα*, but that there were no clear hot-spots, except possibly at position 305 nt of *lacZα* (Figure 3.9B). This may be due to the random binding of A3F to the (-)DNA and an inefficient search of the enzyme by jumping without local scanning by sliding (Figure 3.2), which would make interaction with multiple 5'TTC or 5'TC motifs less likely to occur. Of note, the mutation frequencies induced by A3F and A3G did not increase with the addition of more enzyme to the reaction demonstrating that both A3F and A3G are present at saturating levels (data not shown). Analysis of the distances between A3F-induced mutations demonstrated that 75% of the mutations were separated by more than 20 nt (Table 3.4), confirming that A3F was using jumping this assay system, in agreement with the data on the synthetic oligonucleotide substrates (Figure 3.2).

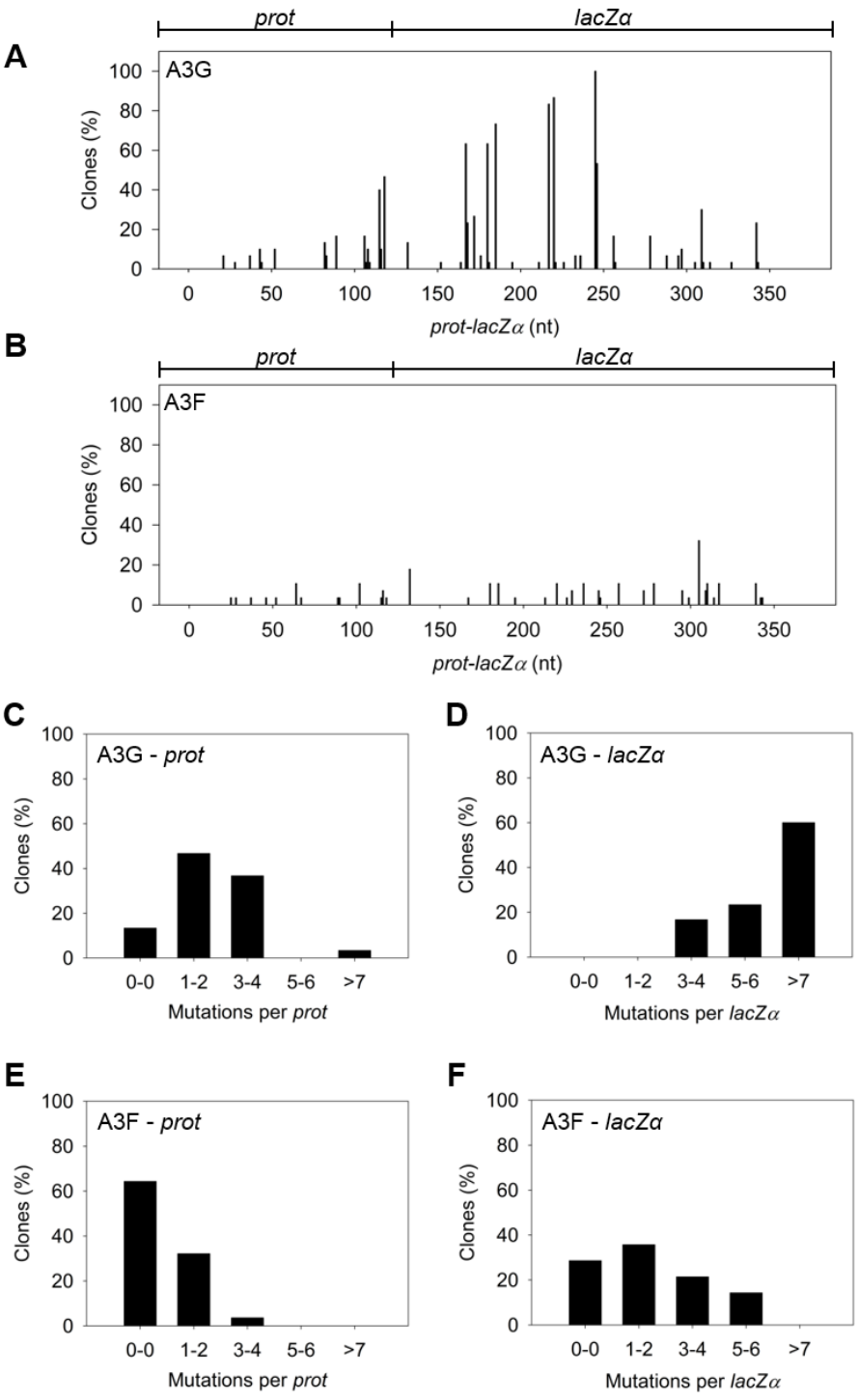

Figure 3.9. A3F exhibits low mutagenic potential in a model HIV replication system. (A-B) *S*pectra of mutations are plotted as the percentage of clones containing a mutation at a particular location (nt) in the 368 nt *prot-lacZα* construct for (*A*) A3G or (B) A3F. (C-F) Histograms illustrate the disparity between the number of mutations that can be induced by A3G versus A3F in the (C, E) prot region that is single stranded for a shorter time than the (D, F) lacZα region.

In contrast, only 50% of A3G-induced mutations were separated by more than 20 nt (Table 3.4), providing confirmation that A3G is capable of recognizing sites that are more closely spaced (Figure 3.6A-B). The analysis in Table 3.4 included all clones (highly mutated and sparsely mutated). To ensure we did not bias our analysis we also examined only sparsely mutated clones for both A3G and A3F (2-5 mutations) and obtained similar results for frequency of mutations separated by more than 20 nt (A3G, 60% and A3F, 85%). In addition, we hypothesized that the tight binding of A3F to ssDNA (Table 3.2), would prevent A3F from frequently dissociating into the bulk solution and reassociating with different (-)DNAs. In agreement with the binding data, A3F increased the population mutation frequency (frequency of white colonies) only 9-fold over the background whereas A3G caused a 12-fold increase in the population mutation frequency (Table 3.3). Although the overall level of mutagenesis induced by A3F was low, we did observe slightly more mutations in the *lacZα* than the *prot* region due to the replication kinetics (Figure 3.9E-F). In the majority of A3F clones (64%) there were no mutations in the *prot* region (Figure 3.9E). In the *lacZα* region the majority of clones only had 1-2 mutations (Figure 3.9F, 36%). However, 29% of clones did not have a G→A mutation and were recovered due to an RT induced error (Figure 3.9F, 0-0). These data demonstrated that A3F was inefficient at inducing mutagenesis during reverse transcription even in areas where the enzyme had ample time to access ssDNA (*lacZα*) and especially in regions that are single-stranded the shortest time (*prot*) (Figure 3.9E-F).

Table 3.4. Analysis of distances between G→A mutations for A3G and A3F

Enzyme	Frequency of clones with mutations >20 nt apart
A3G	0.51
A3F	0.75[a]

3.5.3. Inactivation of HIV-1 protease by A3F and A3G

The increased distance between A3F-induced mutation sites in the HIV replication assay (Table 3.4) in combination with the data on synthetic oligonucleotide DNA indicating that A3F prefers to use jumping (Figure 3.2) provides evidence that the decreased mutagenic ability observed for A3F in cell culture may be due to an inefficient search mechanism on DNA. However, these observations are inconsequential if each mutation by A3F were to inactivate the *prot* gene, which is used here as a predictor of HIV inactivation potential. We gauged the probability that the *prot* of HIV would be inactivated by A3F by determining the mutated amino acid sequences and comparing this to an extensive mutagenesis study of the *prot* conducted by Loeb *et al*. (*354*). Consistent with A3F inducing a low number of mutations (Figure 3.9 and Table 3.3), there were no A3F-induced mutations in 64% of clones (Figure 3.10B). On a per clone basis, A3F-induced mutations resulted in protease inactivation only 50% of the time (Figure 3.10B, 18% active and 18% inactive). The high number of clones remaining active was due to two reasons. First, some clones incurred mutations in regions where any amino acid is tolerated (*354*), even nonconservative changes, e.g., E21K, so the mutation was insignificant (Table 3.5). Second, some clones incurred a mutation that resulted in a conservative change to the amino acid which enabled the protease to retain full or partial activity, depending on the proximity to the active site residues (*354*) (Table 3.5). For example, the M46I mutation was induced by A3F in 11% of clones, but results an active and drug resistant protease. The remainder of the A3F-induced mutations were found only in single clones and 36% of those mutations resulted in an active protease (Table 3.5).

Altogether, A3F was not efficient at inactivating the HIV protease and could also induce resistance to protease inhibitors (Table 3.5, D30N and M46I). This was in contrast to A3G which caused inactivation of 84% of the clones and left only 3% of clones active (Figure 3.10A). A3G also did not mutate some *prot* clones, but only 13% of the time (Figure 3.10A). A3G did induce protease drug resistant mutations in 10% of the population (Table 3.5, D30N), but the examination of these clones individually demonstrated that they were inactivated by other mutations. Overall, we found that per mutation A3G was more likely to cause an inactivating mutation than A3F. This appeared to be due to the ability of a 5'CCC motif to cause more nonconservative mutations than 5'TTC in the *prot* (Table 3.5). For example, A3G had clear hot spots that caused inactivation of the protease, e.g., W42 STOP, 20% of clones; G51R, 36% of clones; G52S, 52% of clones (Table 3.5).

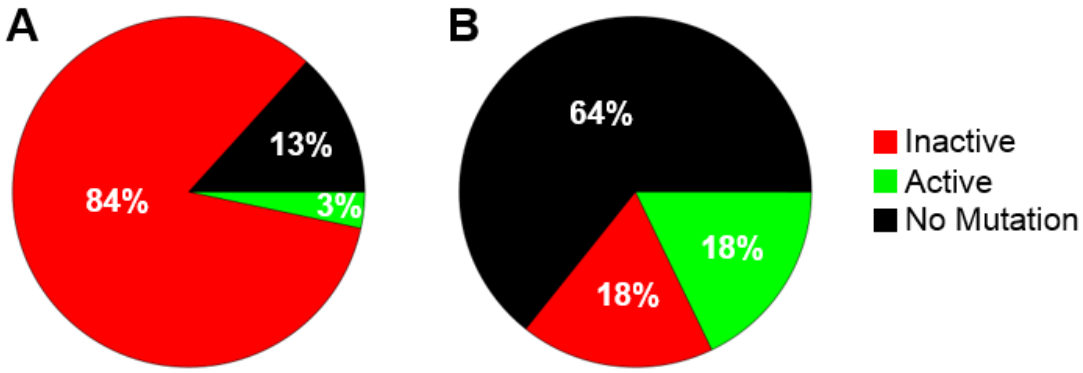

Figure 3.10. Predicted ability of A3F and A3G to inactivate HIV protease. Each prot clone was individually analyzed to determine the percentage of clones that resulted in a mutated and inactive (red) prot, mutated and active prot (green) or prot with no mutations (black) after exposure to (A) A3G or (B) A3F. (A) A3G was able to inactivate the prot in 84% of clones and left an active prot in 3% of mutated clones. A3G did not induce any mutations in the prot in 13% of clones. (B) A3F-induced mutagenesis was less effective than A3G due to no mutations being induced in 64% of clones. Of the 36% of clones with a mutation, 18% left the prot active and 18% inactivated the prot. 9, 12, 19 and 35 d of age.

Table 3.5. A3-induced mutagenesis in HIV prot region synthesized in a model HIV replication assay

Protease enzyme activity was inferred from a mutational study carried out by Loeb and colleagues (*354*) where double plus (++) is active, plus (+) is partially active and minus (–) is inactive in comparison to wild-type protease. Protease inhibitor resistance information is from http://hivdb.stanford.edu. No recorded value is used to indicate that no clones were found with a mutation at that particular site.

Protease Amino Acid Position	Nucleotide change	Amino Acid change	Predicted protease activity	Protease inhibitor resistance	Mutated A3F clones (%)	Mutated A3G clones (%)	Mutated A3G NPM clones (%)	Mutated A3F NGM clones (%)
19	CTG→CTA	L→L	++			7		
21	GAA→AAA	E→K	+		4			
22	GCT→ACT	A→T	-		4	3		
25	GAT→AAT	D→N	++		4	7		3
27	GGA→GAA GGA→AGA	G→E G→R	- -			3 10		3
28	GCA→ACA	A→T	-		4			10
29	GAT→AAT	D→N	-					
30	GAT→AAT	D→N	+	Yes	4	10		3
34	GAA→AAA	E→K	++		11			7
35	GAT→AAT	D→N	+		4			7
38	TTG→TTA	L→L	++					
40	GGG→GAG GGG→AGG GGG→GAA GGG→GGA	G→E G→R G→E G→G	- - - +			7 13		
42	TGG→TAG TGG→TGA TGG→TAA	W→STOP W→STOP W→STOP	- - -		4 4	17		
46	ATG→ATA	M→I	+	Yes	11			
48	GGG→AGG GGG→AGA GGG→GAG GGG→GGA	G→R G→R G→E G→G	++ ++ + ++			13 3 3 7		
49	GGA→AGA GGA→GAA	G→R G→E	- -			3		
51	GGA→AGA GGA→AAA GGA→GAA	G→R G→K G→E	- - -		 4 4	33 7 3		
52	GGT→AGT	G→S	-		4	47	4	

3.5.4. Determinants of processivity for APOBEC3 enzymes

To investigate the A3F DNA scanning mechanism further we made mutants in A3G and A3F to alter their processive scanning behavior. For A3G, the only other A3 double Z-domain enzyme studied with regards to processivity, the NTD domain acts as a processivity factor (*250*). The A3G CTD domain alone is non-processive (Figure 3.11A-B and (*247*, *349*)). In order to focus in on the determinants of processivity, we recombinantly expressed the CTD domain of A3F and tested its processivity using ssDNA substrates as in Figure 3.2. We found that the CTD of A3F could not processively deaminate cytosines that were spaced 63- or 30- nt apart, similar to the CTD of A3G (Figure 3.11A-B, absence of 5'C & 3'C band). The A3F CTD could also not processively deaminate target cytosines 14- or 5-nt apart (Figure 3.11C-D, absence of 5'C & 3'C band), similar to the full-length A3F enzyme (Figure 3.2B-C). These data indicated that the NTD of A3F was a processivity factor.

To determine the specific amino acids within the NTD that differentiate the processive scanning behaviors of A3F and A3G we aligned their amino acid sequences and looked for differences in the predicted helix 6 and loop 7 (Figure 3.12) since these regions have been shown to influence the scanning behavior of A3G (*250*). Specifically, it was found that helix 6 mediated sliding movements and loop 7 mediated jumping movements (*250*). Since we could not observe any scanning by sliding for A3F (Figure 3.2B-C), we hypothesized that residues within or near predicted N-terminal helix 6 would be different from A3G. For A3G, His186 was found to be essential for sliding movements (*250*). Although A3F has a His181 equivalent to A3G (His186) at the end of the predicted helix 6 in the connection domain between the NTD and CTD, A3F has an additional three amino acids, ^{190}NPM^{192}, in comparison to A3G (Figure 3.13A).

To test whether the ^{190}NPM^{192} motif prevents A3F from sliding, we inserted the NPM motif into the equivalent position in A3G (^{195}NPM^{197}) creating an A3G NPM mutant. We then tested if A3G NPM was still able to undergo scanning by sliding. Using the ssDNA substrates with target cytosines close together enables the observation of processive deaminations by sliding (*250*). On the substrate with cytosines separated by 5 nt, A3G NPM retained its processivity at an equivalent frequency to that of the wild-type A3G (compare Figure 3.6B and Figure 3.1C, processivity factors). On the substrate with cytosines separated by 14 nt, A3G NPM was essentially not processive, as evidenced by a processivity factor of 1 which means that A3G NPM double deaminations occurred at the same frequency as expected if they were uncorrelated (Figure 3.13C).

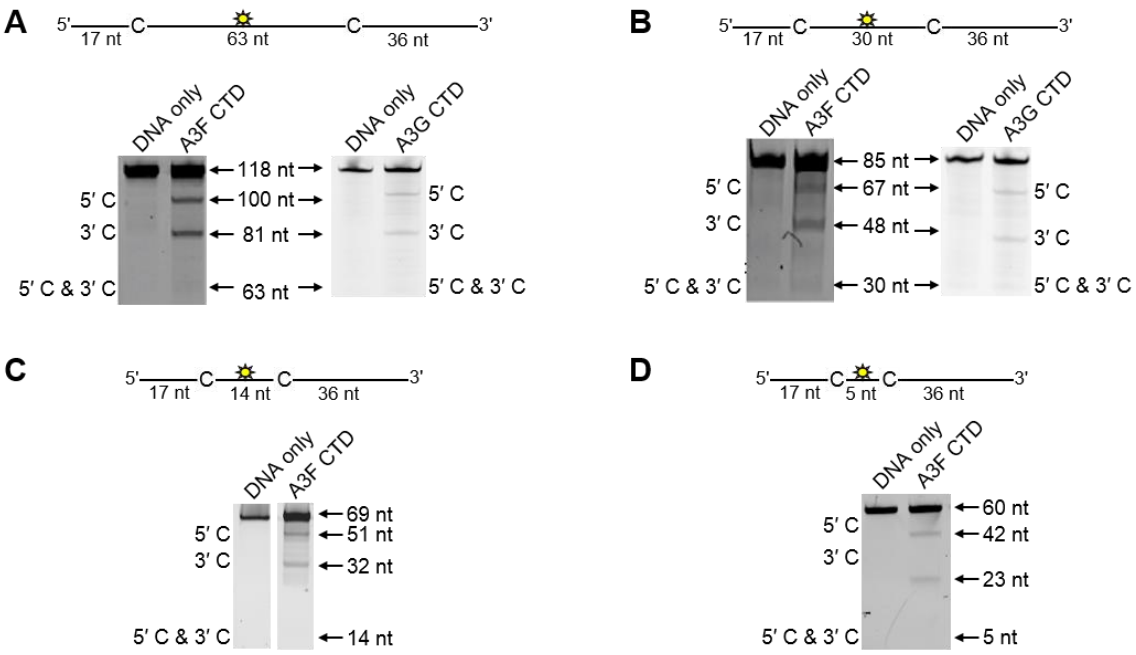

Figure 3.11. Similar to the A3G CTD, the A3F CTD does not deaminate cytosines processively. Processivity of A3F CTD and A3G CTD were tested on substrates that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by different distances. The substrates had 5'TTC motifs (A3F CTD) or 5'CCC motifs (A3G CTD). (A) The two target cytosines within the 118 nt ssDNA sequence are spaced 63 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively. Double deamination of both C residues on the same molecule, which would result in a 63 nt labeled fragment (5'C & 3'C), could not be detected indicating that A3F CTD and A3G CTD are not processive on these ssDNA substrates. (B) The two target cytosines within the 85 nt ssDNA sequence are spaced 30 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively. Double deamination of both C residues on the same molecule, which would result in a 30 nt labeled fragment (5'C & 3'C), could not be detected indicating that A3F CTD and A3G CTD are not processive on these ssDNA substrates. (C) The two target cytosines within the 69 nt ssDNA sequence are spaced 14 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively. Double deamination of both C residues on the same molecule, which would result in a 14 nt labeled fragment (5'C & 3'C), could not be detected indicating that A3F CTD is not processive on this ssDNA substrate. (D) The two target cytosines within the 60 nt ssDNA sequence are spaced 5 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively. Double deamination of both C residues on the same molecule, which would result in a 5 nt labeled fragment (5'C & 3'C), could not be detected indicating that A3F CTD is not processive on this ssDNA substrate. The A3F CTD: DNA and A3G CTD: DNA ratios were 2:1. A representative gel from three independent experiments is shown.

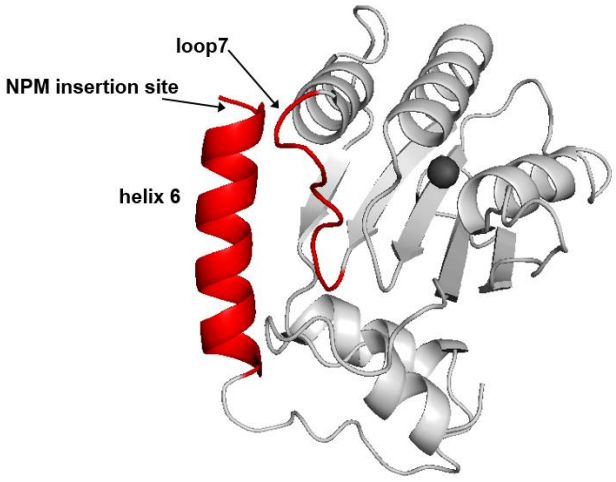

Figure 3.12. Model of the N-terminal domain (NTD) of A3G. Model (grey) shows loop 7 and helix 6 (both in red). The amino acids NPM were inserted at the end of predicted helix 6. Zinc atom is a dark grey sphere. The predicted model of A3G NTD was obtained by using the automated SWISS-MODEL program using the homologous A3G CTD (PDB: 3IQS) structure as a template. Figure was made using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.5, Schrödinger, LLC.).

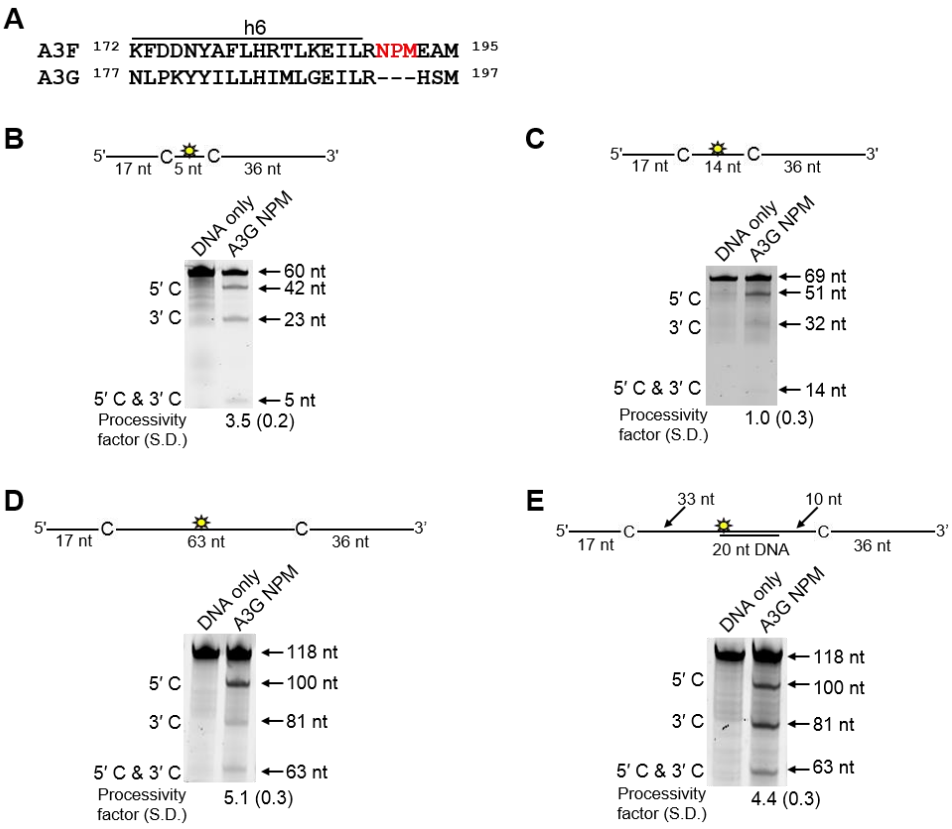

Figure 3.13. Decreased processivity of A3G NPM was attributable to changes in sliding, but not jumping movements. (A) The A3G NPM mutant was created by inserting the NPM motif found in A3F into A3G immediately after Arg 194. (B-E) Processivity of A3G NPM was tested on substrates that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by different distances. The substrates had 5'CCC motifs. (A) The two target cytosines within the 60 nt ssDNA sequence are spaced 5 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment (5'C & 3'C). (C) The two target cytosines within the 69 nt ssDNA sequence are spaced 14 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively. Double deamination of both C residues on the same molecule results in a 14 nt labeled fragment (5'C & 3'C) and were detected at a low level resulting in a processivity factor of 1 (below gel). Since the processivity factor is a ratio between the observed double deaminations and the theoretical deaminations expected to occur for a nonprocessive enzyme (see Materials and Methods), the results indicated that the A3G NPM mutant was not processive on this substrate. (D) The two target cytosines within the 118 nt ssDNA sequence are spaced 63 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). (E) Deamination of the substrate described for (D), but with a 20 nt ssDNA annealed between the two target cytosines to block the sliding component of

processivity. The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. The A3G NPM: DNA ratio was (B) 1: 2.5, (C) 1:10, (D-E) 1:20. Enzyme: DNA ratios were varied due to different specific activities of the enzyme on a given DNA substrate. Values are an average from three independent experiments.

This was in contrast to wild-type A3G that was able to processively deaminate cytosines located 14 nt apart (Figure 3.2B, processivity factor 4.6). These data indicated that the NPM insertion had decreased the sliding distance of A3G. To ensure that jumping was not affected, we tested the A3G NPM mutant on a substrate with cytosines separated by 63 nt without or with a complementary DNA or RNA annealed. First, we established the processivity on this substrate when fully single-stranded. Accordingly, the A3G NPM which had a decreased ability to slide, had a decreased processivity factor on this substrate in comparison to wild-type A3G (compare Figure 3.13D, processivity factor of 5.1 to Figure 3.2D, processivity factor of 8.1), but similar to A3F (Figure 3.2D, processivity factor of 4.6). When a complementary DNA was annealed the processivity of the A3G NPM was not decreased (Figure 3.13E, processivity factor of 4.4) demonstrating that the jumping motion of A3G NPM was not affected. Similar results were found when a complementary RNA was annealed to the substrate (Figure 3.4C). This was in contrast to the characteristic 2-fold decrease in processivity observed with A3G when a complementary DNA or RNA is annealed in between two target cytosines (Figure 3.2D-E and Figure 3.4B) consistent with the hypothesis that attempts to slide over the dsDNA region by wild-type A3G results in dissociation of the enzyme into the bulk solution. That we did not see an increase in the jumping efficiency for the A3G NPM (Figure 3.13E), in contrast to A3F (Figure 3.2E) is in agreement with published data that suggest the determinants of jumping are separate from sliding and localized to predicted loop 7 (Figure 3.12 and (*250*)). Further, the oligomerization state of A3G NPM is equivalent to wild-type A3G (data not shown), not A3F and this may influence the jumping distance of an A3 enzyme (Figure 3.6A-C).

To ensure that the effects on processivity were due to specific changes to the residues interacting with ssDNA while scanning, rather than solely due to a poor affinity for ssDNA, we examined A3G NPM by circular dichroism (CD) spectroscopy and the ability of A3G NPM to bind ssDNA using rotational anisotropy. The CD analysis confirmed that A3G NPM and A3G were structurally similar (data not shown). Interestingly, addition of the NPM residues to A3G resulted in a 2-fold increase in the binding affinity of A3G for the ssDNA, implicating these residues in the ssDNA-NTD interaction (Table 3.2, A3G, K_d of 130 nM; A3G NPM, K_d of 56 nM). The specific activity of A3G NPM was decreased ~3-fold in comparison to A3G (Table 3.6, A3G, 15 pmol/µg/min; A3G NPM, 5.5 pmol/µg/min).

Table 3.6. Specific activities of A3G and A3F wild-type and mutants

Enzyme	Specific Activity (pmol/µg/min)
A3G	15 ± 1
A3F	0.14 ± 0.03
A3G NPM	5.5 ± 1
A3F NGM	0.20 ± 0.02

The specific activity was determined using the substrate with target cytosines separated by 63 nt (118 nt substrate) and the values are shown with the standard deviation that was calculated from three independent experiments.

To further investigate the influence of the NPM motif in A3F, we attempted a reciprocal mutation, i.e., deleting the NPM motif from A3F. However, the mutant A3F did not express well in the *Sf*9 expression system indicating that the NPM deletion caused a structural instability. To circumvent this we made a conservative mutation in A3F to change the NPM motif to an NGM motif. We hypothesized that the Pro would have a significant influence on the functionality of the motif since Pro gives structural rigidity. We then tested the ability of the A3F NGM to processively deaminate two closely spaced deamination motifs by sliding. We found that A3F NGM was able to processively deaminate cytosines that were 5 nt and 14 nt apart (Figure 3.14A-B, processivity factors of 2.1 and 2.4), in contrast to A3F (Figure 3.2B-C). When the distance between the cytosines was increased to 30 nt or 60 nt apart, A3F NGM was able to undergo processive deaminations similarly to A3F (compare Figure 3.14C-D to Figure 3.2A & D). Interestingly, the apparent K_d of A3F NGM was 119 nM, which is 6-fold larger than the K_d of A3F (Table 3.2, 20 nM) further implicating these residues in the enzyme-ssDNA interaction. The specific activity of A3F NGM was ~1.5-fold higher than A3F (Table 3.6). The A3F NGM and A3G NPM results demonstrated that the presence of an NPM motif blocks the ability of both A3F and A3G to processively slide on ssDNA.

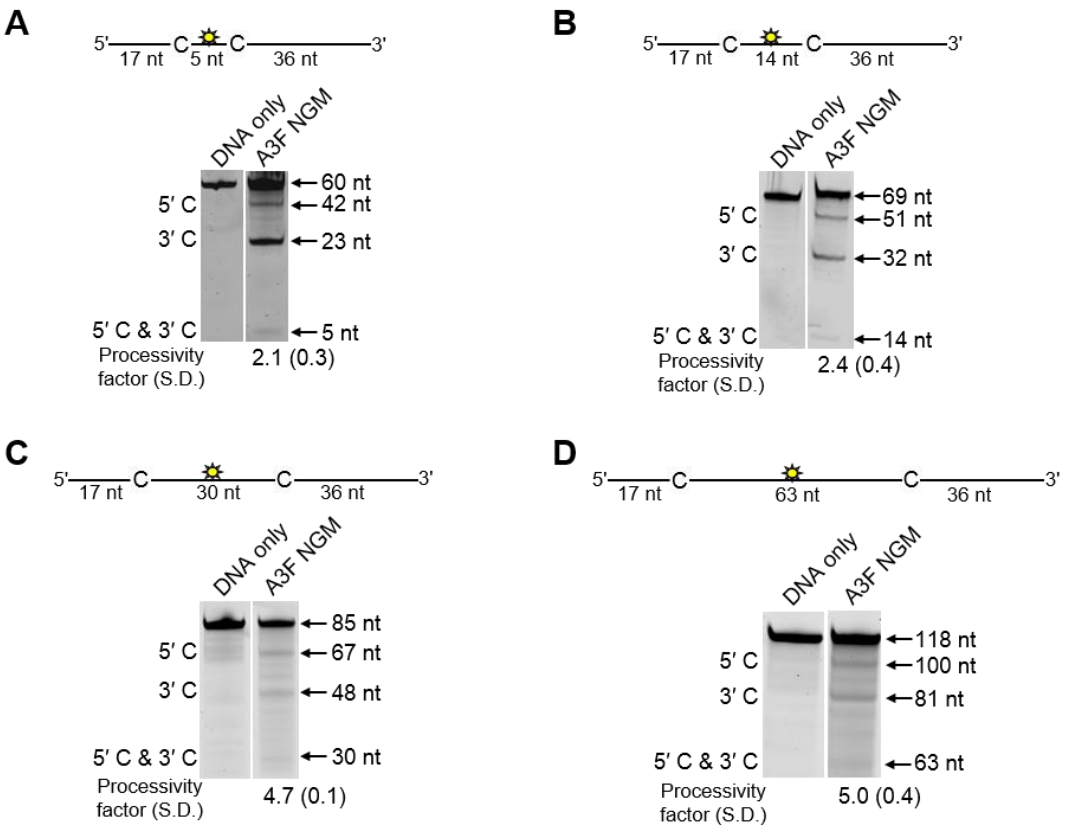

Figure 3.14. A3F NGM is able to slide on ssDNA to catalyze processive deaminations. Processivity of A3F NGM was tested on substrates that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by different distances. The substrates had 5'TTC motifs. (A) The two target cytosines within the 60 nt ssDNA sequence are spaced 5 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment (5'C & 3'C). (B) The two target cytosines within the 69 nt ssDNA sequence are spaced 14 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively. Double deamination of both C residues on the same molecule results in a 14 nt labeled fragment (5'C & 3'C). (C) The two target cytosines within the 85 nt ssDNA sequence are spaced 30 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 30 nt labeled fragment (5'C & 3'C). (D) The two target cytosines within the 118 nt ssDNA sequence are spaced 63 nt apart. Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). The measurements of processivity (Processivity factor) and the Standard Deviation of the mean (S.D.) are shown below the gel. The A3F NGM: DNA ratio was 1:1. Values are an average from three independent experiments.staining (A3G, A3G NPM, A3F CTD) or Bio-Rad Oriole fluorescent stain (A3F, A3F NPM)

2.5.5. Contribution of processivity to efficient mutagenesis of (-)DNA

Our model predicts that the A3G NPM mutant should be a poor inducer of mutagenesis during (-)DNA synthesis due the decreased ability of this mutant to slide on ssDNA (Figure 3.13). In agreement with the model, the A3G NPM induced mutagenesis poorly in the model HIV replication system (Figure 3.15A), similar to A3F (Figure 3.9B), but in contrast to wild-type A3G (Figure 3.9A). The A3G NPM mutant had a mutation frequency in the HIV replication assay (Table 3.3, 0.29 x 10^{-2} mutations/bp), which was 9-fold less than wild-type A3G (Table 3.3, 2.63 x 10^{-2} mutations/bp). The spectrum and sequence analysis demonstrated that the sparse mutations induced by A3G NPM were still in 5'GG or 5'GGG contexts, but that much fewer occurred (Figure 3.15A and Table 3.5). The A3G NPM mutant rarely induced mutations in the *prot* (Figure 3.15C) and mutations in the *lacZα* region were less than A3F (compare Figure 3.15D and Figure 3.9F). Notably, the A3G forms had a 100- (A3G) to 40- (A3G NPM) fold greater specific activity than A3F (Table 3.6). However, since A3G NPM and A3F similarly induced less mutations (Figure 3.15A and Figure 3.9B) than A3G (Figure 3.9A), the data indicated that the ssDNA searching mechanism, but not the specific activity was a primary determining factor in levels of A3-induced mutagenesis.

Our model, which is based on mutagenesis data from A3G, predicted that the mutation frequency of A3F NGM should increase in comparison to A3F. However, despite the A3F NGM mutant being able to slide (Figure 3.14), we found that A3F NGM remained inefficient at inducing mutagenesis in the *in vitro* HIV replication assay (Figure 3.15B). The induced mutagenesis of A3F NGM (Table 3.3, 0.34 x 10^{-2} mutations/bp) was more similar to A3F than A3G (Table 3.3). This could be due to A3F NGM sliding being ~2-fold less efficient than A3G (compare Figure 3.14A-B and Figure 3.2B-C) or that the recovery of sliding alone in A3F is not sufficient for increasing the levels of mutagenesis. The latter possibility suggested another determining factor specific to A3F may affect its mutagenic ability. Specifically, A3F NGM retained two distinct properties of A3F, the formation of tetramers and higher order oligomers (data not shown) and an increase in processivity with increasing distance between deamination motifs (Figure 3.14 and data not shown). Therefore, we propose that the jumping mechanism of A3F that is retained in A3F NGM and distinct from that of A3G is detrimental to efficient mutagenesis and remains as such even the presence of sliding movements. With this being considered, the contributing factors to the efficiency of A3-induced mutagenesis is not only the balance between sliding and jumping, as

exemplified by A3G, but also the type of jumping movements, as exemplified by A3F. The A3F NGM mutant also retained the 5'TTC specificity characteristic of A3F and induced similar mutations in HIV *prot* (Table 3.5).

2.5.6. APOBEC3 processive scanning mechanism determines ability to restrict HIV-1 in single-cycle replication assays

The biochemical data support the hypothesis that the processive scanning mechanism of the A3 enzyme can determine its mutagenic potential during reverse transcription. However, the *in vitro* HIV replication assay used in our experiments cannot account for how the HIV capsid environment may influence A3 enzyme-induced mutagenesis. Therefore, we used a single-cycle replication assay to test whether mutagenesis induced in the *prot* of HIV⊿*vif* proviral DNA by the deamination activity of A3G, A3F and their mutant derivatives would recapitulate the results of A3-induced *prot* mutagenesis in the model HIV replication assay. In agreement with the biochemical data, in the HIV ⊿*vif* proviral DNA the A3G-induced mutations/kb were 6- to 8-fold higher than those of A3F, A3G NPM or A3F NGM (Table 3.7). Upon analysis of codon changing mutations, we found that the A3G hotspot in the *prot* was the Trp 42 codon, which was mutated to a stop codon in all clones containing a mutation, except one clone (Table 3.8). Clones mutated by A3G-catalyzed deaminations also contained other inactivating mutations such as G51R/E or G86R (Table 3.8). In regards to hotspots, the data were similar for A3G NPM, although fewer mutations were recovered (Table 3.7 and Table 3.8).

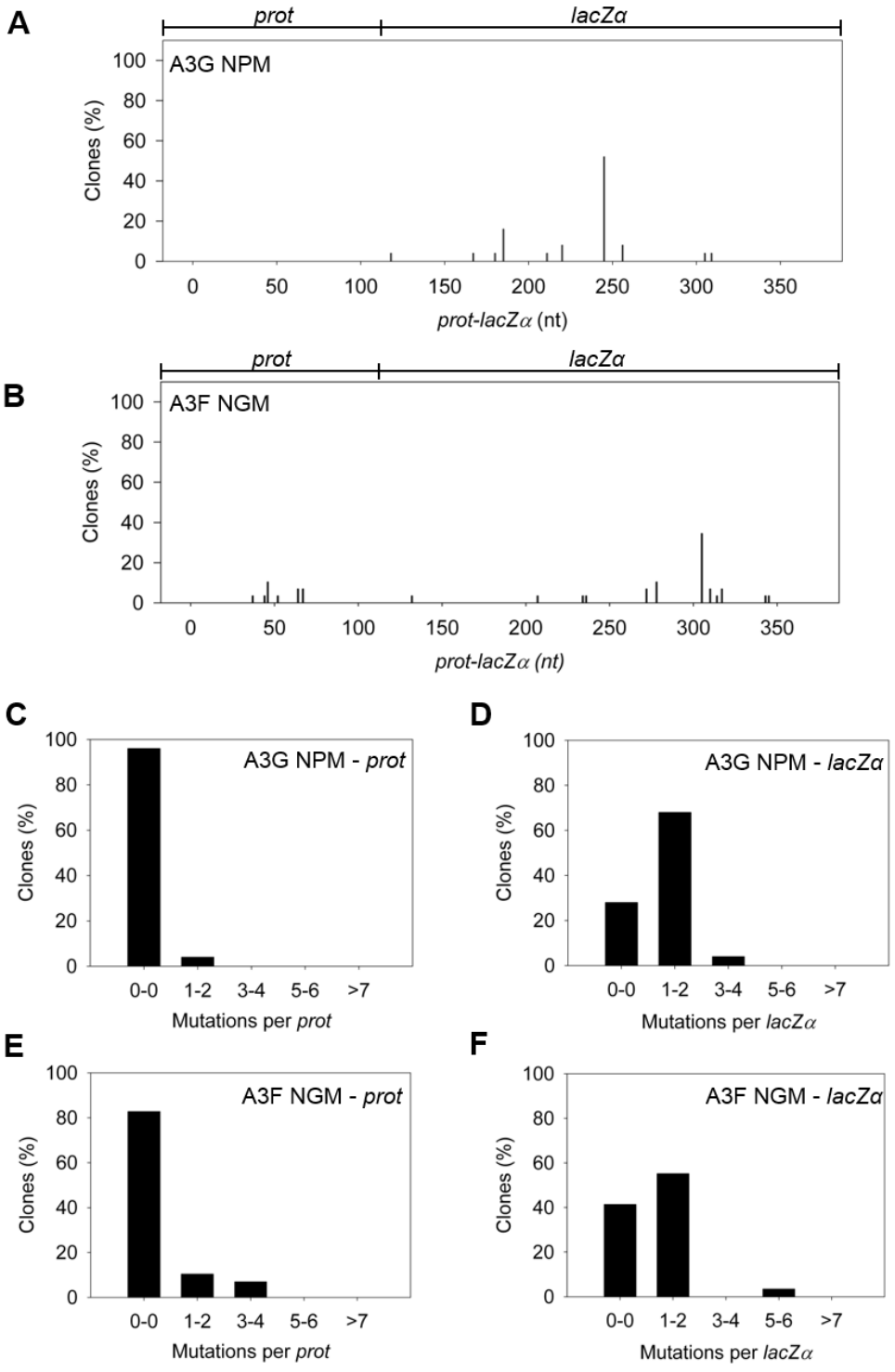

Figure 3.15. Cytosine deamination-induced mutagenesis by mutant A3F and A3G in a model HIV replication system. *(A-B) S*pectra of mutations are plotted as the percentage of clones containing a mutation at a particular location (nt) in the 368 nt *prot-lacZ*α construct for (*A*) A3G NPM or (B) A3F NGM. (C-F) Histograms show the number of mutations per prot or lacZα region for (C-D) A3G NPM or (E-F) A3F NGM.

Table 3.7. Analysis of A3-induced mutagenesis of prot DNA from integrated HIV△vif

Enzyme	Base pairs sequenced	Total number of G→A mutations	G→A mutation frequency (mutations/kb)
A3G	6318	37	5.9
A3F	7371	5	0.7
A3G NPM	7020	7	1.0
A3F NGM	9126	8	0.9

These data supported our biochemical data in which the decrease in sliding by A3G NPM in comparison to A3G resulted in a decrease of mutagenic potential (compare Figure 3.9A and Figure 3.15A). The *prot* clones exposed to A3F or A3F NGM had mutations that at best resulted in partial inhibition of protease activity, e.g., D30N or M46I, and none that resulted in complete inactivation of protease activity (Table 3.8). It is interesting that A3F NGM induced ~1.3-fold more mutations/kb than A3F, suggesting that there was a slight positive effect of the A3F NGM sliding ability on mutagenesis (Table 3.7). Overall, the *prot* sequencing data from HIV *⊿vif* proviral clones was consistent with the conclusions from the *in vitro* model HIV replication assay and many deamination hotspots were common between the two assays (compare Tables 3.5 and 3.8). Differences may have resulted from different temporal dynamics of reverse transcription (*355*) and that the *in vitro* assay used a smaller segment of the *prot* gene. The observation from *in vitro* data that the 5'TTC motif was less able to cause inactivating mutations than the 5'CCC motif was consistent with HIV *⊿vif* proviral DNA exposed to A3F or A3G (Table 3.8). Not only was the 5'CCC motif able to cause more inactivating mutations by overlapping with the Trp codon (5'TGG), which results in a stop codon, as previously observed (*216*), but also because it was more likely to cause nonconservative mutations in comparison to the 5'TTC motif (Table 3.8 and Ref (*341*)).

The impact of A3G- and A3F-induced mutations on the infectivity of the proviral DNA was also examined using the eGFP reporter gene contained in the HIV pNL4-3 *⊿vif* construct. Consistent with sequencing data from the *prot* region, the eGFP reporter gene of the integrated provirus from the same assays was inactivated 3- to 4-fold more in HIV *⊿vif* virions exposed to A3G in comparison to A3F, A3G NPM or A3F NGM (Figure 3.16A).

Table 3.8. A3-induced mutagenesis in integrated proviral HIV-1△vif prot DNA

Protease enzyme activity was inferred from a mutational study carried out by Loeb and colleagues (*354*) where double plus (++) is active, plus (+) is partially active and minus (–) is inactive in comparison to wild-type protease. Protease inhibitor resistance information is from http://hivdb.stanford.edu. No recorded value is used to indicate that no clones were found with a mutation at that particular site.

Protease Amino Acid Position	Nucleotide change	Amino Acid change	Predicted protease activity	Protease inhibitor resistance	Mutated A3F clones	Mutated A3G clones	Mutated A3G NPM clones	Mutated A3F NGM clones
20	AAG→AAA	K→K	++			2	1	
21	GAA→AAA	E→K	++		1			
30	GAT→AAT	D→N	+	Yes	1	1		
34	GAA→AAA	E→K	++		1	1		2
35	GAA→AAA	E→K	+					1
36	ATG→ATA	M→I	+		1			3
41	AGA→AAA	R→K	++		1	1		1
42	TGG→TAG TGG→TGA	W→STOP W→STOP	- -			5 1	1	
46	ATG→ATA	M→I	+	Yes		1		1
48	GGG→AGA GGG→AGG GGG→AAA GGG→GGA	G→R G→R G→K G→G	++ ++ ++ ++			1 1 1 1		
51	GGA→AGA GGA→GAA	G→R G→E	- -			3 1	2	
52	GGT→AGT	G→S	-			2		
57	AGA→AAA	R→K	++			1		
60	GAT→AAT	D→N	++			1		
65	GAA→AAA	E→K	-			2		
73	GGT→AGT	G→S	++			2		
78	GGA→AGA	G→R	-			1		
86	GGA→AGA	G→R	-			3	2	
87	AGA→AAA	R→K	++			1		
90	TTG→TTA	L→L	++			1		
94	GGC→AGC	G→S	+				1	

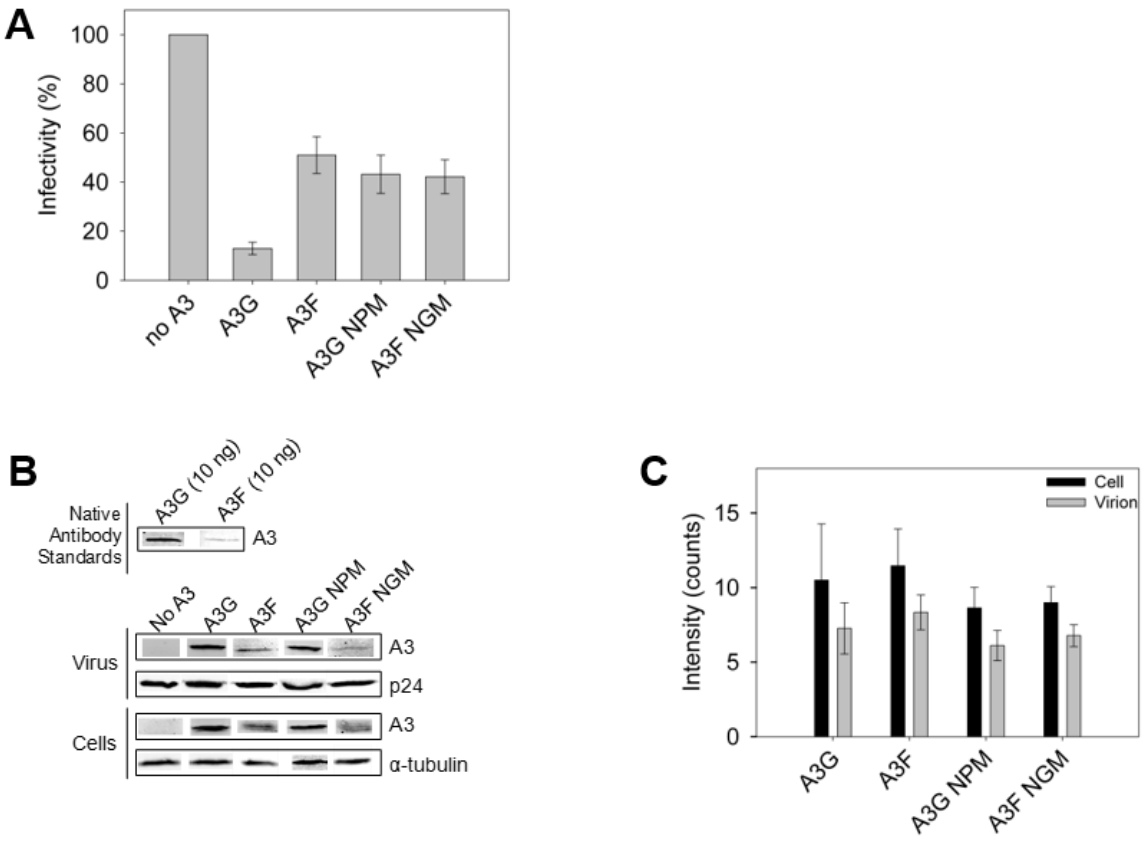

Figure 3.16. A3 enzyme processivity influences the HIV Δvif restriction efficiency. (A) Virus infectivity was measured by eGFP expression in 293T cells infected with HIVΔvif that was produced in the absence or presence of A3G, A3F, A3G NPM or A3F NGM. Results normalized to the no A3 condition are shown with the Standard Deviation of the mean calculated from at least three independent experiments. (B-C) Quantitative immunoblotting was used to determine the levels of A3G, A3F, A3G NPM, or A3F NGM expressed in cells and encapsidated into HIVΔvif virions. (B) The detection capabilities of antibodies to A3G (Apo C17, NIH AIDS Reagent Program) or A3F (C-18, NIH AIDS Reagent Program) was determined by detecting 10 ng of purified A3G or A3F with a 1/1000 dilution of the appropriate antibody. At this dilution, the antibody to A3F was found to be 9-fold less sensitive than the antibody to A3G. This data was used as a correction factor during quantitation of blots (see Materials and Methods). The loading control for cell lysates was α-tubulin and for virions was p24. (C) Blots from at least three independent experiments were analyzed using Odyssey software to determine the band intensity in the cell lysate (black bars) or virions (gray bars). Loading controls were confirmed during quantification to not be significantly different (data not shown). The error bars represent the Standard Deviation of the mean. A t-test determined that there were no significant differences in enzyme expression or encapsidation.

To ensure this was not due to differences in encapsidation efficiency between these A3 enzymes we conducted quantitative immunoblotting on virions and cell lysates. Since we had transfected untagged A3 enzymes for these experiments to avoid the potential effects a tag may have on processivity (Figure 3.17), we initially standardized the antibodies for native A3G and A3F. Using equivalent amounts of purified protein and antibody dilutions, we determined that the antibody to A3F was 9-fold less sensitive than the antibody to A3G (Figure 3.16B). To ensure this was not due to differences in encapsidation efficiency between these A3 enzymes we conducted quantitative immunoblotting on virions and cell lysates. Since we had transfected untagged A3 enzymes for these experiments to avoid the potential effects a tag may have on processivity (Figure 3.17), we initially standardized the antibodies for native A3G and A3F. Using equivalent amounts of purified protein and antibody dilutions, we determined that the antibody to A3F was 9-fold less sensitive than the antibody to A3G (Figure 3.16B). As a result, we used this as a correction factor in the calculated amounts of these enzymes in virions and cells (Figure 3.16C). The immunoblot results demonstrated that A3G and A3F were expressed in 293T cells and encapsidated into ⊿*vif* virions to a similar level (Figure 3.16B-C). Therefore, the data support that there is a *bone fide* difference in the inherent mutagenic abilities of A3G and A3F. We also confirmed that A3G and its NPM mutant and A3F and its NGM mutant were expressed in cells and encapsidated in virions similarly (Figure 3.16B-C) enabling comparisons to be made between the mutant and wild-type forms of the enzymes. The analysis of A3G or A3F mutants from single-cycle infectivity assays was consistent with biochemical data. The A3G NPM mutant that had diminished sliding ability was less able to restrict HIV replication than A3G (Figure 3.16A, 3-fold). A3F NGM was able to decrease HIV⊿*vif* infectivity 10% more than A3F, suggesting a slight positive effect of its sliding ability, but this was not statistically significant (Figure 3.16A). These data provided evidence that the processive scanning mechanism of the A3 enzyme influences the capacity to restrict HIV in a single cycle of replication. The disparity in HIV restriction efficiency was confirmed to be due to differences in mutational load by sequencing the HIV⊿*vif* integrated provirus *eGFP* reporter gene (Figure 3.18).

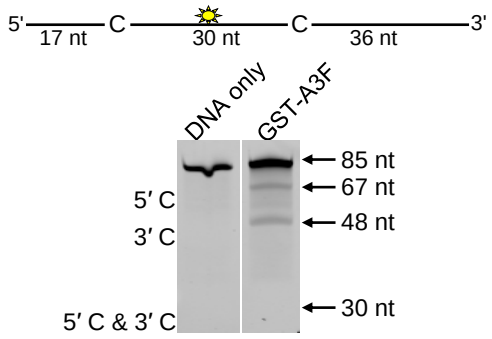

Figure 3.17. GST-A3F is not processive. Deamination was tested on an 85 nt ssDNA substrate that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by 30 nt (sketch) Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 30 nt labeled fragment (5'C & 3'C). GST-A3F is not processive on this substrate as evidenced by the absence of a double deamination band (5'C & 3'C, 30 nt). The A3F: DNA ratio was 1:1. A representative gel from three independent experiments is shown.

```
A3G             GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC 60
A3F             GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC 60
A3G NPM         GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC 60
A3F NGM         GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC 60
REFERENCE       GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC 60
                ************************************************************

A3G             GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC 120
A3F             GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC 120
A3G NPM         GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC 120
A3F NGM         GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC 120
REFERENCE       GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC 120
                ************************************************************

A3G             AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC 180
A3F             AAGCTGACCCTGAAGTTCATCTGCACCACCGACAAGCTGCCCGTGCCCTGGCCCACCCTC 180
A3G NPM         AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC 180
A3F NGM         AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC 180
REFERENCE       AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC 180
                ********************************.***************************

A3G             GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG 240
A3F             GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG 240
A3G NPM         GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG 240
A3F NGM         GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG 240
REFERENCE       GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG 240
                ************************************************************

A3G             CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC 300
A3F             CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC 300
A3G NPM         CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC 300
A3F NGM         CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC 300
REFERENCE       CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC 300
                ************************************************************

A3G             AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGACGCACACCCTGGTG 360
A3F             AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG 360
A3G NPM         AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG 360
A3F NGM         AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG 360
REFERENCE       AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG 360
                ********************************************.***************

A3G             AACCGCATCGAGCTGAAAGGCATCGACTTCAAGGAGGACGGCAACATCCTAAGGCACAAG 420
A3F             AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG 420
A3G NPM         AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG 420
A3F NGM         AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG 420
REFERENCE       AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG 420
                *****************.**********************************..********

A3G             CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC 480
A3F             CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC 480
A3G NPM         CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC 480
A3F NGM         CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC 480
REFERENCE       CTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC 480
                ************************************************************

A3G             ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC 540
A3F             ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC 540
A3G NPM         ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC 540
A3F NGM         ATCAAGATGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC 540
REFERENCE       ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC 540
                ******.*****************************************************

A3G             CACTACCAGCAGAACACCCCC-ATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTA 599
A3F             CACTACCAGCAGAACACCCCC-ATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTA 599
A3G NPM         CACTACCAGCAGAACACCCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTA 600
A3F NGM         CACTACCAGCAGAACACCCCC-ATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTA 599
REFERENCE       CACTACCAGCAGAACACCCCC-ATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTA 599
                ********************* ************************************

A3G             CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT 659
A3F             CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT 659
A3G NPM         CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT 660
A3F NGM         CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT 659
REFERENCE       CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT 659
                ************************************************************

A3G             GCTGGAGTTCGTGACCGCCGCCAGGATCACTCTCGGCATGGACGAGCTGTACAAG 714
A3F             GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG 714
A3G NPM         GCTGGAGTTCGTGACCGCCGCCAGGATCACTCTCGGCATGGACGAGCTGTACAAG 715
A3F NGM         GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG 714
```

Figure 3.18. Representative eGFP sequences of integrated proviruses. Representative eGFP sequences from the single-cycle infectivity assay (Figure 3.16A) are shown. Mutations are in bold. Alignment was made using CLUSTAL W.

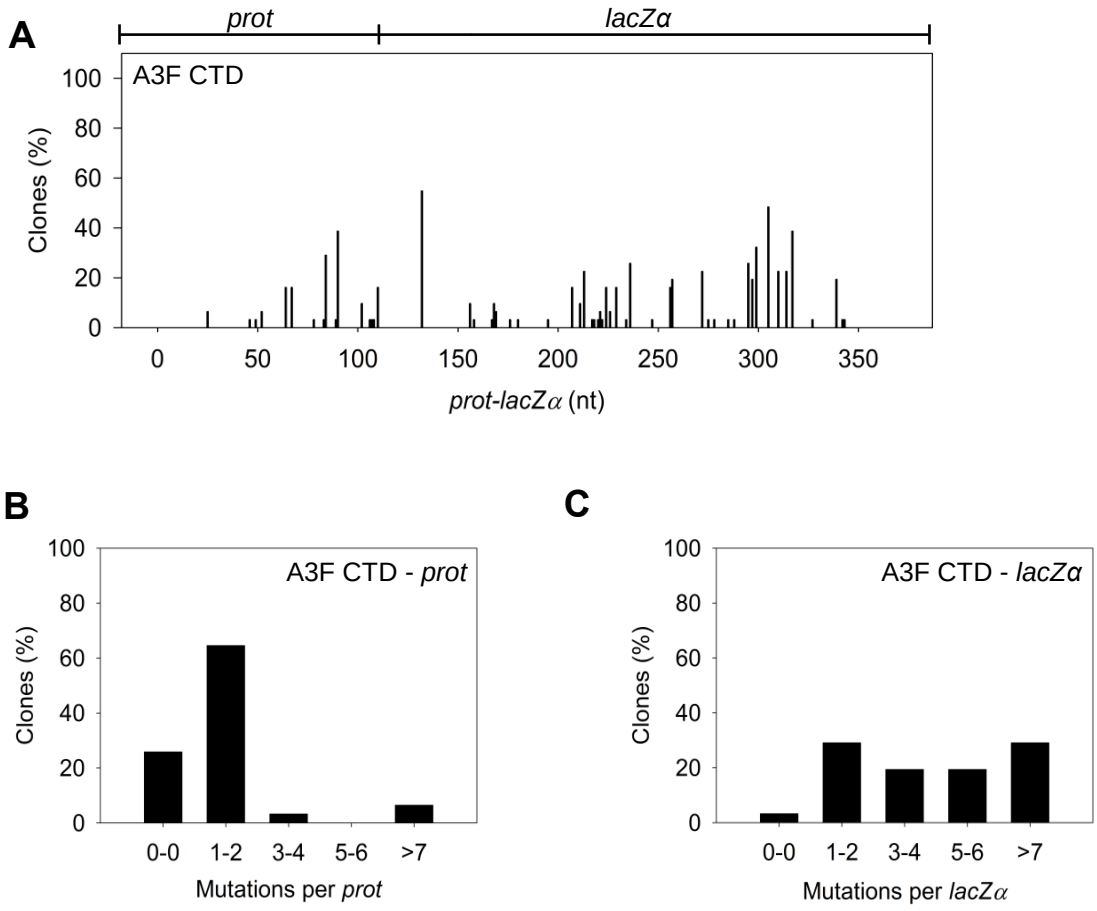

Figure 3.19. A3F CTD mutagenesis in a model HIV replication system (A) Spectrum of mutations are plotted as the percentage of clones containing a mutation at a particular location (nt) in the 368 nt prot-lacZα construct. (B-C) Analysis of the number of mutations induced by A3F-CTD in the (B) prot or (C) lacZα regions.

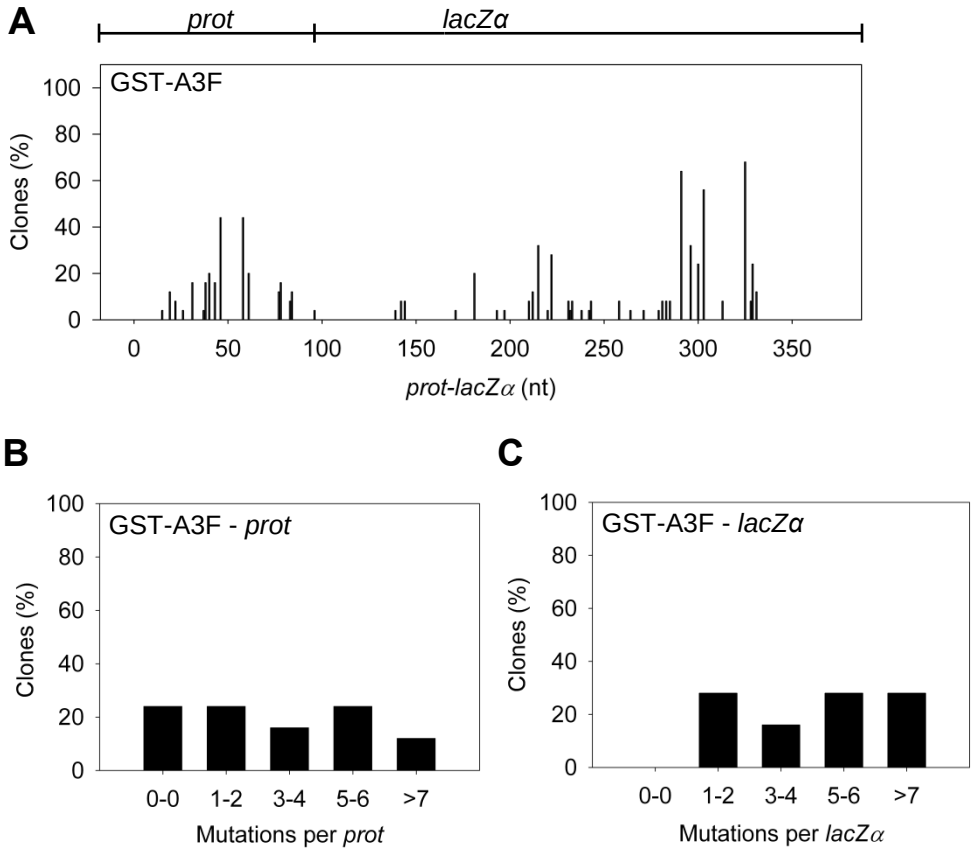

Figure 3.20. GST-A3F mutagenesis is comparable to A3F CTD in a model HIV replication system. (A) Spectrum of mutations are plotted as the percentage of clones containing a mutation at a particular location (nt) in the 368 nt prot-lacZα construct. (B-C) Analysis of the number of mutations induced by GST-A3F in the (B) prot or (C) lacZα regions.

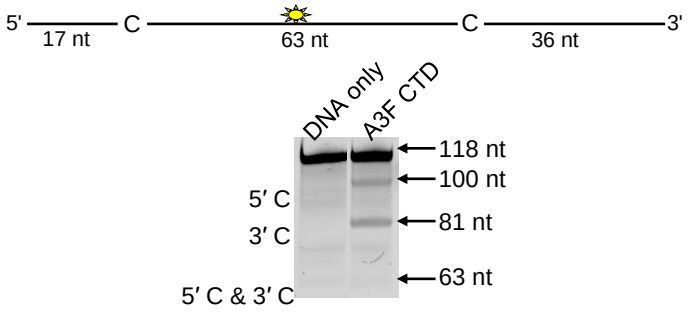

Figure 3.21. A3F CTD is not processive in the presence of NC and RT. Deamination was tested on an 118 nt ssDNA substrate that contained an internal fluorescein (F)-label (yellow star) and two deamination motifs separated by 63 nt (sketch). Single deaminations of the 5'C and 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment (5'C & 3'C). A3F CTD is unable to processively deaminate the target cytosines as evidenced by the absence of a 63 nt labeled fragment above background (5'C & 3'C). The A3F CTD: DNA ratio was 2:1. Three independent experiments were conducted.

3.6. Discussion

Reports have demonstrated that A3F is less effective than A3G at restricting HIV replication and leaves less of a mutational footprint (*258, 266, 287, 330*). This could be due to many reasons such as differences in mRNA/protein expression levels (*31, 330*), virion encapsidation levels (*22, 257*), deamination site preference (*22, 216*), or the inherent biochemical characteristics of the enzymes that govern deamination activity during proviral DNA synthesis. There is no consensus in the literature regarding whether any of the variables determined by cellular conditions, e.g., mRNA expression levels, create disparity between A3F and A3G HIV restriction activities. In addition, other reports have found an equal capacity of A3F and A3G to restrict HIV (*2, 19, 21, 23, 24, 223, 331*). To account for these differences in the literature we undertook a biochemical characterization of A3F in comparison to A3G. The data have enabled us to form a biochemical model to account for cell-based observations and propose that the processive DNA scanning mechanism and the preferred deamination motif of A3 deoxycytidine deaminases are determinants of HIV restriction efficiency.

The data support the hypothesis that a balanced sliding and jumping scanning mechanism is a major contributor to efficient restriction of HIV (*250*) and A3F has less potential to restrict HIV because it does not slide and uses a jumping translocation mechanism that is different than A3G (Figure 3.2 and Figure 3.6). Analysis of A3G and A3F mutants further support the model in which the mechanism that the enzymes scan DNA and not their specific activity can fully account for differences observed in HIV restriction (Table 3.6, Figure 3.9, Figure 3.15, and Figure 3.16). In addition, A3F-induced mutations in preferred 5'TTC /5'TC motifs were less efficient at inducing gene inactivation than the preferred A3G 5'CCC/5'CC motifs, similar to what was identified for A3A (prefers 5'TTC /5'TC) (*341*), adding another distinction in the mutagenic ability of A3F (Tables 3.5 and 2.8).

However, the data cannot support that A3F has no effect on HIV since it is suppressed by Vif (*24*), but there is evidence that the restriction abilities are distinct from A3G in regards to mutagenic load, selection pressure on HIV and contribution of deamination-independent HIV restriction (*255, 259, 260, 266, 287, 299*). It was initially recognized by Zennou and Bieniasz that per mutation, A3G could cause a much larger decrease in HIV infectivity than A3F (*22*). This early study on A3F (*22*) was in contrast to other early studies published showing A3F was similar in effectiveness to A3G (*2, 21, 23, 24, 331*). Such incongruent data still remains in the literature (*19,*

258, 266, 287, 330) and may be due to different experimental systems. Specific to our data, we observed that the HIV⊿*vif* retained 51% infectivity in the presence of A3F and 13% infectivity in the presence of A3G, suggesting that A3F is not as effective as A3G at restricting HIV (Figure 3.16A). However, Albin *et al*. found that over multiple replication cycles, A3F restricted HIV replication similarly to A3G and selected for Vif mutant revertants (*259*). It may be that A3G requires only one exposure to HIV for high level restriction compared to A3F that may require multiple cycles for strong HIV restriction, but the end point is the same. Importantly, multiple infection rounds more closely mimics how A3 enzymes would interact with HIV *in vivo*. Nonetheless, our data propose that the mechanism by which A3G and A3F reach this end is different and that A3F has the potential to cause more sequence diversification of HIV than A3G. This idea is supported by Chaipan *et al*. that found that A3F suppressed HIV in multiple rounds of replication but required a longer period of exposure to HIV before the level of suppression reached that of A3G (*266*). This is consistent with our sequence data from the *prot* of integrated proviruses (Table 3.8). As such, the role of A3F may be to supplement mutagenesis induced by A3G (*19, 287*) since their effects have been shown to be additive (*2*) or be distinct from A3G and perhaps rely on a deamination-independent mechanism, such as inhibition of reverse transcription and integration (*255, 299, 356*). A3F has been reported to exert a larger deamination-independent inhibition of HIV replication than A3G, but this is not as effective as deamination-mediated restriction of HIV (*255*, *299*).

To characterize the mechanism by which A3 enzymes induce mutagenesis we studied the A3G NPM mutant. The A3G NPM mutant demonstrated that the scanning mechanism on DNA and not specific activity is a primary determinant in mutation induction during reverse transcription. Since A3F had a lower specific activity than A3G (Table 3.6), it could be argued that this was contributing to the lower level of induced mutagenesis (Figure 3.9). However, the A3G NPM mutant, which had decreased sliding in comparison to wild-type A3G (compare Figure 3.2 and Figure 3.13), retained a specific activity more similar to that of A3G than A3F (Table 3.6), but induced a very low level of mutagenesis (Figure 3.15A, C-D) and decreased HIV infectivity only 2-fold versus A3G that decreased HIV infectivity 8-fold (Figure 3.16A). These data suggest that specific activity is not a determinant in the ability to cause mutations during reverse transcription and is supported by previous data in which the specific activity of the enzyme was inconsequential during reverse transcription (*326*). This appears to be because the activity of the

enzyme during reverse transcription is instead determined by factors such as (-)DNA synthesis and RNaseH activity (*326*). The A3G NPM mutant further confirmed that the determinants within the NTD for sliding involve residues near A3G predicted helix 6 and that this is distinct from the determinant for jumping (Figure 3.13D-E), which in A3G is loop 7 (Figure 3.12). Of note, the NPM motif is predicted to be at the end of NTD helix 6, which is a connection point between the NTD and the CTD domains (*357, 358*). Through amino acid sequence alignment we identified that A3D is the only other double domain A3 deaminase to contain an NPM motif at the end of predicted helix 6, suggesting that A3D would also lack sliding movements while scanning ssDNA, similarly to A3F. Some specific residues within helix 6 have previously been shown to affect specific activity (*250, 358*), possibly because of structural changes in the connection between the NTD and CTD that can affect the catalytic activity of the CTD or DNA binding affinity. Insertion of the NPM motif into A3G immediately after the predicted helix 6 ends (Figure 3.13A) did not cause a large disruption in structure based on CD spectra (data not shown), but did result in a ~3-fold decrease in specific activity (Table 3.6) and ~2-fold increase in binding affinity for ssDNA (Table 3.2).

To confirm a role of the NPM motif in blocking sliding movements we mutated this region in A3F to create an A3F NGM mutant with the hypothesis that removing the rigid proline residue would enable the enzyme to slide on ssDNA and deaminate closely spaced residues. Consistent with the hypothesis, closely spaced residues were processively deaminated by A3F NGM (Figure 3.14A-B). However, the ability to slide did not enable A3F NGM to induce high levels of mutagenesis similar to A3G *in vitro* (Figure 3.15) or in a single-cycle infectivity assay (Figure 3.16A and Table 3.7). This does not preclude that jumping and sliding are important for inducing mutagenesis in virus infected cells since the A3G NPM mutant that had decreased sliding restricted HIV similarly to A3F in single cycle infectivity assays (Figure 3.16A). Rather, these data indicated that the ability to slide and jump is necessary, but not sufficient to induce high levels of mutagenesis. The data supported the conclusion that the type of sliding and jumping movements, e.g., distance transversed was also important. Namely, we found that A3F processivity on ssDNA increased with increasing distance between deamination motifs, in contrast to A3G, demonstrating that the average jumping distance of A3F was larger than A3G (Figure 3.6A-B). This was confirmed with sequence analysis from the model HIV replication assay in which a larger number of deaminations were >20 nt apart for A3F than A3G (Table 3.4). Thus, the A3F NGM mutant

could slide, but was not truly a mimic of A3G. Altogether, it appears that the sliding and jumping mechanism of A3G is specifically optimal to induce a large number of deaminations during reverse transcription of DNA.

An important note regarding the study of A3F is that we found N-terminally tagged GST-A3F was not processive (Figure 3.17), despite binding ssDNA with a K_d of 46 ± 4 nM. That the binding affinity of GST-A3F was more similar to A3F (Table 3.2, K_d of 20 nM) than A3F CTD (Table 3.2, K_d of 288 nM) indicated that the GST-A3F was able to bind ssDNA with both NTD and CTD domains, despite a lack of processivity. This suggested that the GST tag caused steric hindrance on amino acid determinants for processivity in the NTD. Interestingly, we observed that nonprocessive A3F forms, both A3F CTD and GST-A3F, induced more mutations than wild-type, processive A3F (compare Figure 3.9 and Figures 3.19 and 3.20). We also found that A3A, which is largely nonprocessive, induced slightly less mutations than A3G in the *in vitro* HIV replication assay (*341*), but more than A3F. Although this initially seems difficult to reconcile, it is consistent with the overall hypothesis that processivity is related to mutagenic potential, since processive A3G is still the most efficient at inducing mutagenesis. It is only that a lack of processivity appears to be better than an "ineffective" processive enzyme such as A3F. This is not due to differences in the assay systems for characterizing processive deaminations on ssDNA oligonucleotides and the model HIV replication assay since addition of NC and RT to the ssDNA oligonucleotides in a deamination reaction did not change our observations regarding A3F CTD processivity (Figure 3.21). In comparison to the nonprocessive A3F CTD and GST-A3F, processive A3F leaves many potential deamination motifs unmodified (compare Figure 3.9 and Figures 3.19 and 3.20). Although there is inefficiency in the GST-A3F and A3F CTD having to dissociate and reassociate with the substrate many times, the reassociations can be much closer to the previous dissociation resulting in a more thorough search of the DNA. For example, we found that in the model HIV replication assay, 61% of A3F CTD-induced mutations were >20 nt apart in contrast to A3F where 75% of induced mutations were >20 nt apart (Table 3.4 and data not shown). Since the HIV replication assay is not conducted under single hit conditions, the results emphasize the inefficiency of the searching mechanism used by A3F. Since the binding affinity of A3F for ssDNA is tighter than A3G or A3F CTD (Table 3.2), it is conceivable that A3F may also have a lack of frequent movements or excursions on the ssDNA that contribute to the inefficient search for deamination motifs. However, resolution of this speculation awaits single-molecule analysis.

In sum, the data demonstrated that the interactions of A3F with ssDNA are essentially detrimental to its ability to induce a high mutation frequency.

Our data demonstrate two main points. First, the data provide a biochemical reason for the inefficiency with which A3F-induces mutagenesis of HIV$\triangle$*vif* as observed in this report and by others (*258, 266, 330*) by demonstrating that the processive scanning behavior of A3F is detrimental to its mutagenic potential. The data establish that a balanced sliding and jumping ssDNA scanning mechanism similar to A3G is required for the most efficient induction of HIV mutagenesis. Secondly, the data show that deamination of 5'CCC/5'CC has more gene inactivating potential than 5'TTC/5'TC providing an additional reason for less restriction of HIV by A3F than A3G, in agreement with previous reports (*22, 216, 341*). The data does not preclude that A3F can effectively restrict HIV and is in agreement with studies showing that A3F can restrict HIV in multiple rounds of infection (*259, 266*), but since the number of mutations induced has been correlated to HIV inactivation (*236, 254*), the data support the interpretation that A3F inactivates HIV less efficiently than A3G in a single round of infection.

CHAPTER 4. PREFACE TO CHAPTER 5

In Chapter 3, the data showed that A3F is less potent than A3G at inhibiting HIV-1 (referred to as HIV) because of its distinct scanning mechanism and that A3F also has a different preference for the sequence context surrounding the deamination site (*226*). This was eluded to an early publication by Zennou and Bieniasz (*22*), confirmed by our group (*226*) and others (*331*, *359*). A3F and A3G are commonly coexpressed in human tissues (*31*, *32*) and it has been recently reported that they can coencapsidate into HIV-1 virions (*220*). However, whether the coencapsidated A3F and A3G work independently, additively or synergistically had not been investigated in-depth (*2*, *220*, *360*). Therefore, Chapter 5 was designed to investigate if coexpressed A3F and A3G have an independent, additive, or synergistic ability to restrict HIV replication. This investigation entailed a systematic dissection of the mechanisms by which A3F and A3G interact *in vitro* and in cells. To develop an A3F/G coexpression model, previous studies used A3F and A3G containing separate plasmids to transfect cells in either an equimolar (*2*) or non-equimolar ratio (*220*). In such transfection approaches using separate plasmids, equimolar co-transfection is not in reality possible. The resulting transfected cells will have various molar ratios of A3F and A3G on a single-cell basis. As a result, HIV virions generated from such strategies will produce a mosaic (heterogeneous) population of virus with different combinations and levels of A3 proteins encapsidated. To circumvent the problem of a mosaic population of viruses and ensure equimolar expression of A3F and A3G on a single-cell basis, we applied a novel strategy that expressed A3F and A3G from the same plasmid but using separate promoters. This approach resulted in more uniform expression and virion coencapsidation of both A3F and A3G. For the first time we have shown that A3G and A3F hetero-oligomerize in the absence of RNA. Although A3G and A3F are known to function alongside each other, these data provided evidence for an A3F/G hetero-oligomeric A3 with distinct properties when compared to the individual counterparts. Thus, the generation of hetero-oligomeric A3 species may be a unique antiviral strategy of endogenous A3 enzymes.

CHAPTER 5. APOBEC3G AND APOBEC3F COENCAPSIDATED IN HIV-1 VIRIONS COOPERATE TO RESTRICT VIRAL REPLICATION

Anjuman Ara, Robin P. Love, Tyson B. Follack, Khawaja A. Ahmed, Madison B. Adolph and Linda Chelico

Department of Microbiology & Immunology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

All experiments in this chapter were performed by A.A., except Figure 5.2A that was performed in collaboration with K.A.A and Figure 5.2C that was performed by T.B.F. A.A. and L.C. conceived and designed the experiments. A.A., L.C., R.P.L., and M.B.A. analyzed the data. A.A and L.C. wrote the article. A.A., R.P.L., T.B.F., K.A.A., M.B.A., and L.C. revised the article.

This part of the thesis has been published in ***Journal of Virology***

Mechanism of enhanced HIV restriction by virion coencapsidated cytidine deaminases APOBEC3F and APOBEC3G. Ara A., Love R.P., Follack T.B., Ahmed K.A., Adolph M.B. and Chelico L. *Journal of Virology,* Jan 18;91(3). pii: e02230-16. doi: 10.1128/JVI.02230-16.

5.1. Abstract

The A3 enzymes, A3G and A3F, are coordinately expressed in $CD4^+$ T cells and can become coencapsidated into HIV-1 virions, primarily in the absence of the viral infectivity factor (Vif). A3F and A3G are deoxycytidine deaminases that inhibit HIV-1 replication by inducing guanine to adenine hypermutation through deamination of cytosine to form uracil in (-)DNA. The effect of the simultaneous presence of both A3G and A3F on HIV-1 restriction ability is not clear. Here, we used a single cycle infectivity assay and biochemical analyses to determine if coencapsidated A3G and A3F differ in their restriction capacity than A3G or A3F alone. Proviral DNA sequencing demonstrated that compared to each A3 alone, A3G and A3F when combined had a coordinate effect on hypermutation. Using size exclusion chromatography, rotational anisotropy, and *in vitro* deamination assays we demonstrate that A3F promotes A3G deamination activity by forming an A3F/G hetero-oligomer, in the absence of RNA, which is more efficient at deaminating cytosines. Further, A3F caused the accumulation of shorter reverse transcripts due to decreasing reverse transcriptase efficiency, which would leave single-stranded (-)DNA exposed for longer periods of time enabling more deamination events to occur. Although A3G and A3F are known to function alongside each other, these data provide evidence for an A3F/G hetero-oligomeric A3 with unique properties when compared to each individual counterpart.

5.2. Author lay summary

The A3 enzymes A3F and A3G act as a barrier to HIV-1 replication in the absence of the HIV-1 viral infectivity factor (Vif) protein. After A3 enzymes are encaspidated into virions they deaminate cytosines in (-) DNA which forms promutagenic uracils that induce transition mutations or proviral DNA degradation. Even in the presence of Vif, footprints of A3-catalyzed deaminations are found demonstrating that A3s still have discernable activity against HIV-1 in infected individuals. We undertook a study to better understand the activity of coexpressed A3F and A3G. The data demonstrate than an A3F/A3G hetero-oligomer can form that has unique properties compared to each A3 alone. This hetero-oligomer has increased efficiency of virus hypermutation, raising the idea that we may still not fully realize the antiviral mechanisms of endogenous A3 enzymes. Hetero-oligomerization may be a mechanism to increase their antiviral activity in the presence of Vif.

5.3. Introduction

A3 enzymes are a family of deoxycytidine deaminases that can act as host restriction factors for HIV-1 (referred to as HIV) (*213*). To restrict HIV replication, A3 enzymes must first become encapsidated in HIV virions from a producer cell (*361-363*). In the next cell that is infected, the HIV genomic RNA is reverse transcribed to form the (-) DNA. During this time, single-stranded (-) DNA is vulnerable to A3 deaminations of cytosine that form uracil (*216, 361, 362*). The reverse transcriptase is forced to use uracil as a template during (+) DNA synthesis which induces C/G $\rightarrow$ T/A mutations and can functionally inactivate the virus or lead to its degradation through DNA repair pathways that excise uracil (*13*). The effect of A3 enzymes on HIV replication in infected individuals is severely dampened by the viral infectivity factor (Vif) that interacts with A3 enzymes, recruits an E3 ubiquitin ligase complex, and induces A3 ubiquitination and degradation in the proteasome (*15, 16, 334, 335, 364, 365*). From the seven human A3 enzymes, there are four, A3D, A3F, A3G, and A3H that can inhibit HIV replication in the absence of Vif (*223*). These four enzymes are coordinately expressed in $CD4^+$ T cells and induced upon T cell activation (*31, 32*).

However, even in the presence of Vif, there is evidence of A3 deaminations inducing hypermutation of integrated proviral genomes (*26, 366-372*). There are two ways in which this occurs. In one route, the A3s are encapsidated in the presence of Vif, but at a much lower amount (*205*). Additionally, viruses can adapt to have a less fit Vif that is less efficient at inducing A3 degradation (*373*). From these proviral DNA sequences, G$\rightarrow$A mutations have been attributed to specific A3s through identification of their preferred sequence contexts surrounding the mutations by both cellular and *in vitro* studies (*2, 216-219, 362, 374*). The sequence contexts showed that A3G uniquely prefers to deaminate the 3′C in 5′CC motifs, resulting in a GG$\rightarrow$GA mutation signature (*216, 217*). A3D, A3F, and A3H all deaminate in 5′TC motifs, resulting in a GA$\rightarrow$AA mutation signature (*2, 218, 219*)**.** More detailed analysis of the signatures revealed that A3G is most active at a 5′CCC motif and A3D can be differentiated from A3F and A3H by examining a larger surrounding sequence 3′ of the cytosine (*220*). Studies that sequenced integrated proviral genomes have shown that proviral DNAs contain mutations at multiple sequence contexts suggesting that multiple A3s can mutate the same genome (*26, 366-372*). However, what was unable to be concluded from these studies was if the mutations induced from multiple A3s occurred in a single round of replication or multiple rounds of replication. If they occurred in a single round

of replication, then the A3s would need to be coencapsidated into HIV virions. Recently, it was found that A3G is primarily encapsidated into virions by binding nonspecifically to cellular RNAs (*221*). The only requirement for being virion encapsidated was that HIV Gag was also bound to the same RNA (*375*). In this model, the encapsidation is not competitive and multiple A3s should be coencapsidated. However, a computational study determined from sequence analysis of proviral DNA that A3G and A3F rarely comutate the same genome (*376*). This endpoint analysis is in contrast to cellular studies that have identified that A3 enzymes are able to coencapsidate and some studies have found that there is a synergistic effect where more mutations or more viral restriction than expected by an additive relationship are induced (*2, 220, 360*). Despite this synergistic effect being identified, a mechanism for how multiple A3 enzymes can result in more than additive mutations or virus restriction has not been determined.

Based on previous findings, there are two possible mechanisms by which coencapsidated A3s could result in higher levels of mutations. One mechanism is to slow down the reverse transcriptase to keep the (-) DNA single-stranded for longer to allow A3s more time to scan the DNA for their preferred deamination motif (*283, 287, 294*). This is thought to be accomplished by A3 enzymes binding to the template and blocking the progression of reverse transcriptase. A second mechanism is to increase the processivity of the A3 enzyme (*226, 250*). The processivity is the ability of an enzyme to deaminate multiple cytosines in a single enzyme-substrate encounter. A3 enzyme processivity is mediated by facilitated diffusion (*213*). Facilitated diffusion is a term that describes Brownian motion driven diffusion and results in enzymes moving along the DNA phosphate backbone by sliding, diffusing within the charged domain of the DNA by a mechanism termed jumping, and moving between DNA segments through a doubly bound state termed intersegmental transfer (*248, 249*). The DNA scanning enables the A3 enzyme to search and find the cytosine among the nontarget nucleotides and enables the continuation of the search after a deamination, without completely dissociating from the DNA. The sliding movements cover small distances, approximately up to 20 nt (*226, 250*). For A3G, the average sliding distance is 12 nt (*253*). The jumping and intersegmental transfer movements enable the enzyme to move distally, e.g., 100 nt or more (*226, 227*).

We initiated this study to investigate if coexpressed A3F and A3G have an additive or synergistic ability to restrict HIV replication. We focused on A3F and A3G since these enzymes are most commonly expressed together in the human population and are active against HIV (*31,*

32). While A3H Haplotypes II, V, and VII are highly restrictive of HIV, they occur less frequently in the overall population, i.e., Haplotype II, 26.5%; Haplotype V, 20%; Haplotype VII, 0.9% (*270, 277*). A3D is much less active against HIV than A3F and A3G (*266*). In the present work, the data show that A3F and A3G are able to induce a combined increase in mutations in HIV DNA and synergistic decrease in HIV infectivity in single cycle infectivity assays. The mechanism for this cooperation was two-fold. First, A3F and A3G can hetero-oligomerize in the absence of RNA and this A3F/G hetero-oligomer has improved processivity. Second, the A3F/G hetero-oligomer can decrease the efficiency of reverse transcriptase, which provides increased time for A3-catalyzed deaminations. All together the data suggest that A3F and A3G are able to function together with distinct and improved properties from the individual enzymes.

5.4. Material and Methods

5.4.1. Protein expression and purification

Recombinant baculovirus production for expression of GST-A3G, GST-A3F and GST-nucleocapsid protein in *Sf*9 cells was carried out using the transfer vector pAcG2T(BD Biosciences), as previously described (*226, 227, 250*). *Sf*9 cells were infected with recombinant virus at a multiplicity of infection (MOI) of 1 for GST-A3G and GST-NC and a MOI of 2 for GST-A3F. After 72h of infection, recombinant baculovirus infected *Sf*9 cells were harvested for protein purification. The A3G and NC were purified in the presence of RNaseA and the GST tag cleaved on the affinity column, as previously described (*247*). The A3F was purified in the presence of RNaseA and the GST tag cleaved in solution, as previously described (*226, 227*). Proteins are estimated to be approximately 95% pure by SDS-PAGE.

The *E. coli* strain containing the plasmids to express HIV reverse transcriptase and HIV protease was provided by Dr. Stuart Le Grice (National Cancer Institute). Expression of the HIV reverse transcriptase and protease were carried out as previously described (*342*). Cell lysates produced using sonication were clarified by centrifugation and then purified using a HisTrap FF crude column (GE Healthcare) and HiTrap Heparin HP (GE Healthcare) as described previously (*342*).

5.4.2. Expression of A3G and A3F in 293T cells

To express both A3F-V5 and A3G-HA on a single-cell basis, we used an expression plasmid, pVIVO2 (Invivogen), with two transcription units in a single vector. A3G and A3F were

PCR amplified from pcDNA3.1 vectors previously used in the lab with primers that contained a C-terminal 1X HA-tag sequence (for A3G) or C-terminal V5-tag sequence (for A3F) (*226*). A3F-V5 was cloned using an XbaI site in MCS1. A3G-HA was cloned using an EcoRI site in MCS2. Using this cloning strategy, plasmids containing both A3F-V5 and A3G-HA (pVIVO2 A3F-V5/A3G-HA), A3F-V5 alone (pVIVO2 A3F-V5), and A3G-HA alone (pVIVO2 A3G-HA) were constructed. These constructs were also used as a template for site directed mutagenesis to create catalytic mutants of A3F and A3G that resulted in pVIVO2 A3F-V5 E251Q/A3G-HA E259Q, pVIVO2 A3F-V5 E251Q, and pVIVO2 A3G-HA E259Q. Primer sequences are listed in Table 5.1.

5.4.3. Coimmunoprecipitation (co-IP) assay

Coimmunoprecipitation (co-IP) assays were conducted as described previously (*275*). Briefly, the 293T cells (2.5×10^6 cells per 75 cm^2 flask) were transfected with 1 µg of total DNA. Equal amounts of each plasmid pVIVO2 A3G-HA, pVIVO2-A3F-V5, or pVIVO2 A3F-V5/A3G-HA were used to transfect the cells. GeneJuice transfection reagent (EMD Millipore) was used according to the manufacturer's instructions. At 64 h post transfection, the cells were washed with PBS and lysed in co-IP buffer (50 mM Tris-Cl pH 7.4, 1% Nonidet-P40, 0.1% sodium deoxycholate, 10% glycerol, 150 mM NaCl) supplemented with RNaseA (20 µg/ml; Roche) and EDTA-free protease inhibitor (Roche). Clarified supernatants were precleared with protein A-agarose-conjugated normal rabbit IgG (2 µg, Santa Cruz Biotechnology) in the presence of RNase A (20 µg/ml; Roche). One half of the precleared supernatant was then incubated with protein A-agarose-conjugated polyclonal rabbit anti-HA antibody (2 µg; Sigma) and another half (mock) was incubated with protein A-agarose-conjugated normal rabbit IgG (2 µg, Santa Cruz Biotechnology) at 4°C for 2 h. Resin was washed and the samples were then resuspended in Laemmli sample buffer and prepared for SDS-PAGE.

After SDS-PAGE, proteins were transferred to a nitrocellulose membrane. For detection of A3G-HA and A3F-V5 in cell lysates, the nitrocellulose membrane was probed with polyclonal Rabbit HA (1:1000; Sigma) and monoclonal mouse V5 antibodies (1:1000, Sigma), respectively. For the loading control, monoclonal mouse anti-α-tubulin (1:1000; Sigma) was used. HA- and rabbit IgG- immunoprecipitated lysates were probed with anti-V5 mouse monoclonal antibodies. After incubation with Horse Radish Peroxidase (HRP) conjugated secondary antibodies, the blots

were visualized with X-ray film using Super Signal West Pico chemiluminescence substrate (Thermo-Scientific).

5.4.4. Size exclusion chromatography

To determine the oligomerization state of the A3 enzymes, we used size exclusion chromatography (SEC). A 10 mL Superdex 200 (GE Healthcare) resin bed contained in a column with a 0.5 cm diameter and 16 cm height was used. The running buffer used contained 50 mM Tris pH 8.0, 200 mM NaCl, and 1 mM DTT. The molecular masses and oligomerization states were calculated from the standard curve obtained by using Bio-Rad gel filtration standard set. For A3G or A3F alone, a total of 15 µg of the purified A3 enzyme was loaded on the column. For the combined run of A3F and A3G, 15µg of each purified enzyme was preincubated at 21°C for 3 minutes, before loading onto the column. Protein in fractions were detected by western blot analysis by using anti-A3G (ApoC17 rabbit antiserum (Cat # 10082, NIH AIDS Reagent Program) and anti-A3F (Cat # GTX47211, GeneTex) antibodies

5.4.5. Steady state rotational anisotropy assay

To measure protein-protein and protein-ssDNA binding interactions, we used steady state fluorescence depolarization (rotational anisotropy), where one of the binding partners was fluorescein (F)-labeled. Data were collected using a QuantaMaster QM-4 spectrofluorometer (Photon Technology International) with a dual emission channel. Measurements were made at 21°C. Samples were excited with vertically polarized light at 495 nm (6 nm band pass) and vertical and horizontal emissions were measured at 520 nm (6 nm band pass). Apparent dissociation constants (K_d) were obtained by fitting to a sigmoidal curve using Sigma Plot 11.2 software.

For binding to ssDNA, we measured the ability of A3F alone, A3G alone and A3F/A3G to bind to a 118 nt F-labeled ssDNA, as described previously (*226*). Reactions were performed in a 60 µl total volume, which contained F-labeled ssDNA (10 nM) in RT buffer (50 mM Tris pH 7.5, 40 mM KCl, 10 mM $MgCl_2$, 1 mM DTT) and A3G (0 - 900nM), A3F (0 - 75nM), or A3F/A3G in combination (0 - 340nM). For the A3F/A3G binding in combination, the enzymes were first mixed at an equimolar ratio and preincubated at 21 °C for 3 min before titration into the reaction.

To measure protein-protein binding, A3G was F-labeled using the Fluorescein-EX Protein Labeling Kit (Life Technologies) and used as the binding substrate for A3F. This assay was carried

out in 60 µl total volume that contained 25 nM F-labeled A3G in the presence of RT buffer and increasing amount of A3F (0 - 250nM).

5.4.6. Quantitative immunoblotting

The 293T cells expressing A3G-HA and A3F-V5 from single-cycle infectivity assays were detected using anti-HA and anti-V5 antibodies (Sigma) in cell lysates (40 µg total protein) and virions. Virions were prepared for immunoblotting by concentration using Retro-X concentrator (Clontech) and 20 µl of concentrated virus was used. Loading controls for cell lysates (α-tubulin, Sigma) and virions (p24, Cat #3537, NIH AIDS Reagent Program) were detected using mouse monoclonal antibodies. Proteins of interest and loading controls were detected in parallel on the same gel by using the Licor/Odyssey system (IRDye 680-labeled goat anti-rabbit secondary antibody and IRDye 800-labeled goat anti-mouse secondary antibody).

5.4.7. Single-cycle infectivity assay

In order to generate VSV-G pseudotyped HIV Δvif NL4-3 viruses, 3×10^5 293T cells were transfected using GeneJuice (EMD Millipore) transfection reagent as previously described (*226*). Cells were transfected with 500 ng of pNL4-3 HIV Δvif, which expresses an eGFP reporter gene (*377*), 200ng of pMDG, which expresses the VSV-G protein (*331*, *378*), in the presence of empty pVIVO2, pVIVO2 A3F-V5/A3G-HA (A3F/A3G combined), pVIVO2 A3F-V5 (A3F-V5 alone), or pVIVO2 A3G-HA (A3G-HA alone). A titration of A3 expression vector was used (25, 50, 100 ng). We used empty pVIVO2 vector to achieve a total of 800 ng DNA. Sixteen hours after the transfection, the medium (DMEM with 10% FBS) was replaced after washing the cells with PBS. Virus-containing supernatants were collected 48 h after the media change and filtered through 0.45 µm PVDF syringe filters. Virus was quantified by a p24 enzyme-linked immunosorbent assay (QuickTiterTM Lentivirus Titer Kit, Cell Biolabs Inc. or HIV-1 p24 ELISA kit XpressBio). Target 293T cells were infected with virus by spinoculation at 800 x *g* for 1 h in the presence of 8 µg/ml of polybrene (*345*). Infection levels in 293T cells were determined by flow cytometry by detecting eGFP fluorescence at 48 h post infection and normalized to HIV△vif infections in the absence of A3 enzymes (*226*). Statistical significance of results was determined using an unpaired t-test.

5.4.8. Intracellular detection of A3F-V5 and A3G-HA in 293T cells by flow cytometry

Single-cell suspensions of untransfected 293T cells and A3F-V5 and A3G-HA transfected 293T cells were prepared in PBS containing 2% FBS and 0.05% sodium azide. Intracellular staining was performed following an established procedure (*379*). Briefly, transfected and untransfected cells were fixed using BD Cytofix/Cytoperm buffer containing 4.2% paraformaldehyde. Fixed cells were permeabilized with BD 1X Perm/Wash buffer containing saponin and FBS. Cells were stained with rabbit anti-V5-CF543 and rabbit anti-HA-CF640R antibodies (Biotium Inc.) in BD Perm/Wash buffer for 30 min at 4 ºC. After three washes with BD Perm/Wash buffer, cells were resuspended in PBS containing 2% FBS and 0.05% sodium azide. Data were acquired by flow cytometry on a FACSCalibur using CellQuest software (BD Biosciences), and analyzed with FlowJo software (TreeStar).

5.4.9. Sequencing of integrated proviral DNA

After 48 h of infection, total DNA from infected 293T cells was extracted using DNAzol reagent (Life Technologies). DNA was treated with DpnI (New England Biolabs) for 1 h at 37 °C to remove possible contaminating plasmid DNA, and the *protease (prot)* (nt 2280–2631) sequences were amplified by PCR using Q5 polymerase (New England Biolabs). Primers are have been published previously (*226*). PCR products were purified and cloned with the CloneJET PCR cloning kit (Thermo). DNA was sequenced with kit-specific primers and carried out at the National Research Council of Canada (Saskatoon, Canada). Analysis of the sequence context of the mutations was done using Original Hypermut (*380*).

5.4.10. Quantification of late reverse transcript formation during HIV replication

For quantifying late reverse transcripts (LRT) by quantitative real-time PCR (qPCR) we followed the method of Belanger *et al*. (*284*). HIV Δvif infected 293T cells were harvested 8 h post-infection and total DNA was extracted, then treated with Dpn1. For each reaction, 0.9 pmol/ml of each primer (as listed in (*284*)), 0.25 pmol/ml of the probe (as listed in (*284*)) and 10 ng of template DNA were used in a 20 µl of reaction volume. Reactions were performed in triplicate with TaqMan Gene Expression master mix (Applied Biosystems). qPCR cycling conditions were 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. qPCR was carried out using StepOnePlus Real-Time PCR system (Applied Biosystems). The copy numbers in each sample were normalized for DNA input using human RNaseP copy number assay

(cat# 4403326, Applied Biosystems). Relative LRT quantitation was calculated following the comparative Ct method (ΔΔCt method).

5.4.11. *In vitro* deamination assay

All the ssDNA substrates used in this study were synthesized from Tri-Link Biotechnologies and have been published previously (*226*). In brief, substrates all contain two deamination motifs, either 5′CCC for A3G or 5′TTC for A3F, and a fluorescein labeled thymidine in between the motifs. Reactions were carried out under single-hit conditions, i.e., < 15% substrate usage, to ensure that a single ssDNA substrate was interacting with at most a single enzyme (*344*). Under these conditions, a processivity factor can be determined. The processivity factor is a ratio of the quantified total number of deaminations occurring at two sites on the same ssDNA substrate with a calculated theoretical value of deaminations that would occur at these two sites if the deamination event were not processive (see reference (*227*)). For A3F deamination reactions, 100 nM A3F was incubated with 100 nM ssDNA. For A3G, 30nM A3G was incubated with 100 nM substrate. Due to differences in specific activity on different substrates, sample times ranged from 1 to 30 min after incubation at 37°C in RT buffer. The time points where only 10-11% of the substrate was deaminated were considered in processivity factor determination. For experiments using A3F and A3G in combination, we first preincubated the enzymes at an equimolar ratio for 3 min at 21 °C before addition to the reaction. Reactions were stopped using phenol and chloroform extractions. The ssDNA was then treated with Uracil DNA Glycosylase (New England Biolabs) and heated under alkaline conditions to induce DNA breakage at deaminated motifs. The DNA fragments were then resolved on a 10% or 20% v/v denaturing polyacrylamide gel depending on the expected sizes. Gel pictures were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare) and gel band intensities were quantified by ImageQuant software (GE Healthcare).

5.4.12. Primer extension assay

An RNA template encompassing the primer binding site (PBS) and upstream region near the 5′-end of the HIV genome (nt 571–674) was generated as described previously (*283*). An 18 nt ^{32}P-labeled RNA primer to mimic $tRNA_{Lys,3}$ primer was heat annealed to the 106 nt template RNA containing the PBS as described previously (*283*). Reactions were conducted in the presence of 10 nM primer/template, RT buffer, 200 μM dNTPs, 175 nM nucleocapsid, and 480 nM reverse transcriptase in the absence or presence of A3G alone, A3F alone or A3F and A3G in combination.

Each A3 alone was used at concentrations of 40, 80, and 320 nM. For A3F and A3G in combination, an equimolar ratio was pre-incubated for 3 min at 21°C and then added to reactions to achieve a final concentration of 40, 80, or 320 nM enzyme complex. Reactions were preincubated at 37°C for 1 min before the addition of dNTPs which started the reaction. A negative control was used which contained all reaction components except reverse transcriptase to ensure there was no contaminating polymerase activity. Reactions were stopped by adding a 5-fold excess of 20 mM EDTA and 95% formamide. Primer extension was visualized by resolving samples on a 16% denaturing 8 M urea polyacrylamide gel. Gel band intensities were measured by phosphorimaging with a Bio-Rad FX scanner. The integrated gel band intensities of all bands in a lane were calculated with ImageQuant software (GE Healthcare) and used to determine the relative amounts of extended and unextended primers as well as fully extended product (82 nt). Statistical significance of primer extension assay results was determined using an unpaired t-test.

Table 5.1. Primers, probes and DNA substrate

Name	Sequence
pVIVO2-A3G-HA forward	5' TTTGAATTCATGAAGCCTCACTTCAGAAACACAGTGGAGCGAATGTATC 3'
pVIVO2-A3G-HA reverse	5'GAATTCTTAAGCATAATCTGGAACATCATATGGATAGTTTTCCTGATTCTG GAG 3’
pVIVO2-A3F-V5 forward	5' TTTTCTAGAATGAAGCCTCACTTCAGAAACACAGTGGAGCGAATGTATCGA GACAC 3’
pVIVO2-A3F-V5 reverse	5'TCTAGATCACGTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTA CCCTCGAGAATCTCCTGCAG 3'
qHIV forward	5’-CAAGTAGTGTGTGCCCGTCTGT-3’
qHIV reverse	5’-CGAGTCCTGCGTCGAGAGA-3’
qHIV probe	5’-FAM-CAGTGGCGCCCGAA-3’
PBS template primer (Forward)	TGT TAG GAC TCT GGT AAC TAG AG
PBS template primer (Reverse)	GTC CCT ATT AAC TTT CGC TTT CAA G
PBS template	UGU UAG GAC UCU GGU AAC UAG AGA UCC CUC AGA UCA CUC UAG ACU GAG UAA AAA UCU CUA GCA GUG GCG CCC GAA CAG GGA CUU GAA AGC GAA AGU UAA UAG GGA C
PBS Primer	GUC CCU GUU CGG GCG CCA

5.5. Results

5.5.1. A3F and A3G can hetero-oligomerize in an RNA independent manner

A3F and A3G have been shown to hetero-oligomerize on RNA in an RNA-dependent manner (*24*). However, A3F and A3G can also form homo-oligomers in the absence of RNA (*226, 247, 381*). Here, we examined if A3F and A3G could also form hetero-oligomers in the absence of RNA. We used A3F and A3G enzymes purified in the presence of RNaseA to remove the bound cellular RNA that causes the enzymes to form higher order oligomers, i.e., >650 kDa (*227, 348*). Based on comparison to a calibration curve, the size exclusion chromatography (SEC) confirmed the absence of higher order oligomers (Figure 5.1A-B). For the SEC, we used relatively low concentrations of enzymes (1 μM) and detected the enzymes by quantitative immunoblotting to avoid potential artifacts caused by high concentrations of A3G and A3F. A monomer of A3G or A3F is predicted to be 46.4 or 45 kDa, respectively, based on the amino acid sequences. Consistent with previous reports, A3G was present in both monomer and dimer fractions (Figure 5.1B, fraction 20, dimers, 101 kDa; fraction 22, monomers, 48 kDa). A3F was present in trimer and dimer fractions (Figure 5.1B, fraction 19, trimers, 153 kDa; fraction 20, dimers, 101 kDa) (*247*).

To examine whether A3F and A3G can hetero-oligomerize we mixed together an equal amount of each enzyme before loading onto the column. The A3F/G blots show the SEC profile for the combined run where either A3G (anti-A3G) or A3F (anti-A3F) was detected (Figure 5.1B). These bands were quantified and plotted for comparison to SEC of each enzyme alone (Figure 5.1C-D). For A3F, the peak fraction of trimers (fraction 19) shifted to a peak fraction of tetramers (fraction 18, 207 kDa) in the presence of A3G (Figure 5.1B-C). Additionally, the dimer peak (fraction 20) of A3F was decreased (Figure 5.1B-C). These data suggest that A3F, which is predominantly a trimer, can interact with one molecule of A3G to form a tetrameric A3F/G hetero-oligomer. To investigate this further, we analyzed the A3G SEC data. The A3G alone predominantly fractionated as a monomer (Figure 5.1B and D, fraction 22). In the presence of A3F, and concomitant with A3F and A3G forming a hetero-oligomer, the A3G peak fraction shifted to the tetrameric fraction (Figure 5.1B and D, fraction 18, 207 kDa). However, a significant amount of A3G monomers are still present in the presence of A3F, supporting the conclusion that there are less A3G molecules than A3F molecules in the A3F/G hetero-oligomer (Figure 5.1B). Together with the A3F SEC and immunoblotting, the data support that a trimer of A3F interacts

with a monomer of A3G and that this complex is a protein-protein interaction and not mediated by RNA.

To determine if this interaction would be relevant in cells we used two additional approaches. In one experiment, we conducted a co-IP to determine if A3G-HA expressed in 293T cells could immunoprecipitate A3F-V5. Cells transiently expressing A3G-HA, A3F-V5, or A3G-HA and A3F-V5, were lysed, incubated with RNaseA, and then incubated with anti-HA antibody or normal rabbit IgG (mock) and Sepharose A beads. Even in the absence of RNA, A3G-HA was able to immunoprecipitate A3F-V5 indicating that A3F and A3G associate through a protein-protein interaction in cells (Figure 5.1E). In another experiment, we quantified the strength of the interaction between A3F and A3G using fluorescence depolarization to measure the rotational anisotropy of fluorescein-labeled A3G (F-A3G) in the presence of increasing amounts of A3F. As the A3F formed a complex with the F-A3G, the anisotropy increased and reached saturation where all of the F-A3G was bound. From this saturation curve, an apparent dissociation constant (K_d) of 135 nM was calculated by fitting the curve to a Sigmoidal equation by regression analysis (Figure 5.1F). These data indicated that A3F and A3G interacted with a high affinity and further support that they could interact in cells. That the binding curve best fit a Sigmoidal equation rather than a rectangular hyperbola by least squares analysis further supports that multiple A3F molecules are binding to A3G, consistent with the SEC data (Figure 5.1B-D).

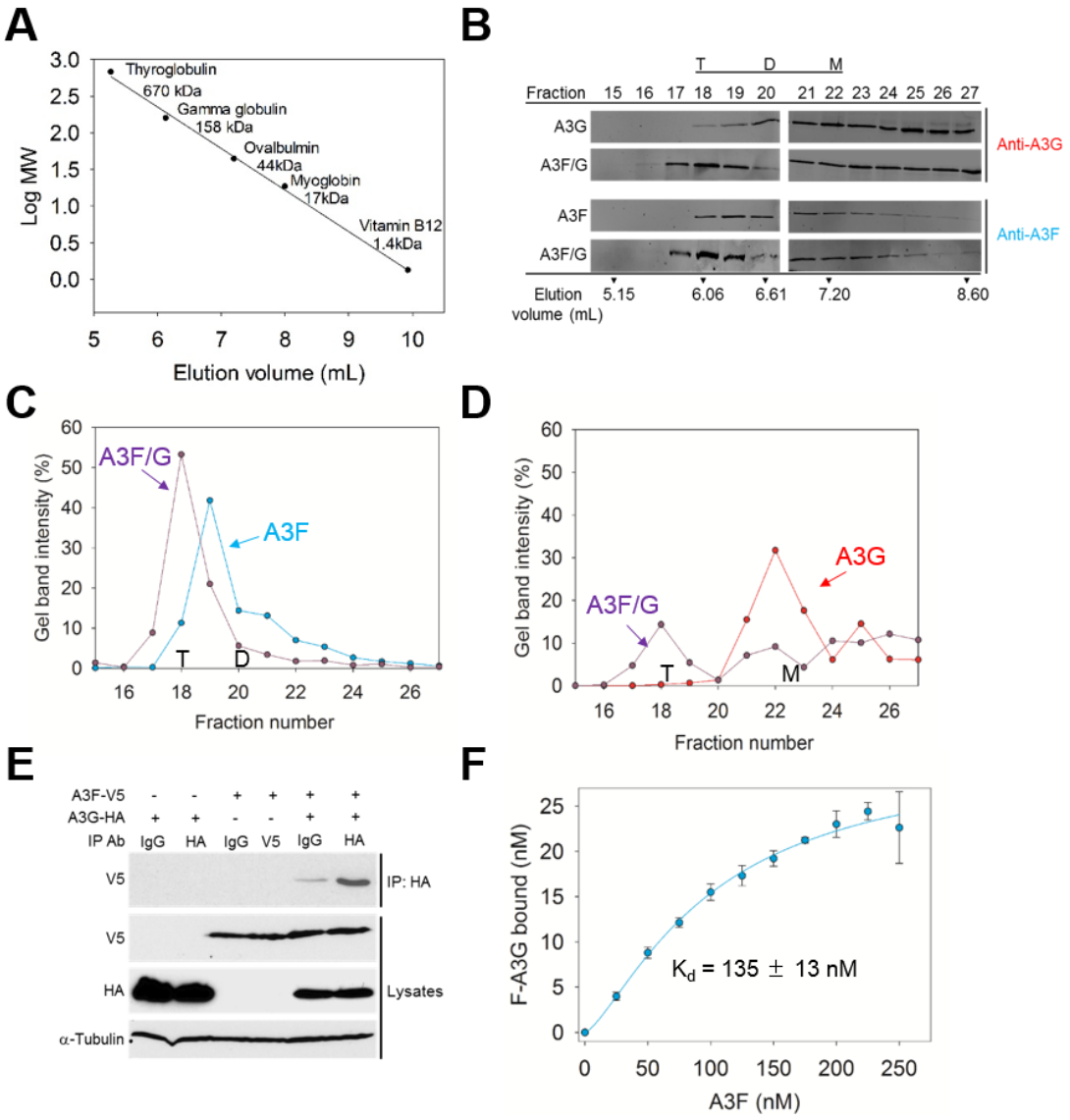

Figure 5.1. A3F and A3G hetero-oligomeric. (A-D) Size exclusion chromatography was conducted with a 10 mL G200 Superdex column. (A) A calibration curve was used to calculate the molecular weights and oligomerization states of the enzymes. (B) The A3G, A3F, and A3F/G SEC experiments used quantitative immunoblotting to detect A3G or A3F with antibodies to the native proteins. The integrated band intensities calculated using Licor/Odyssey software were used to generate chromatograms. The T, D, and M notations indicate peak fractions for tetramers, dimers, and monomers, respectively. (C) The integrated band intensities for A3F indicated that A3F alone was primarily a trimer (Fraction 19, 153 kDa) with a minority of dimers (fraction 20, 101 kDa). The A3F in the A3F/G combined run was primarily a tetramer (fraction 18, 207 kDa). (D) The integrated band intensities for A3G indicated that A3G alone was primarily a monomer (fraction 22, 46 kDa) with a minority of dimers (fraction 20, 101 kDa). The A3G in the A3F/G combined run maintained a population of monomers (fraction 22), but also was able to fractionate

with the peak fraction corresponding to tetramers (fraction 18, 207 kDa). (E) Coimmunoprecipitation of A3F-V5 with A3G-HA. The A3G-HA and A3F-V5 were transfected alone or in combination. The immunoprecipitation of cell lysates used either anti-HA antibody or Rabbit IgG (mock) and immunoblotted with antibodies against α-tubulin, HA, and V5. Cell lysates show the expression of α-tubulin, HA, and V5. (F) Steady state fluorescence depolarization was used to measure the rotational anisotropy of F-A3G interacting with A3F. Rotational anisotropy was normalized to fraction F-A3G bound. An apparent K_d was calculated by regression analysis of the saturation curve from three independent experiments. A sigmoidal fit was chosen by least squares analysis and resulted in an apparent K_d of 135 ± 13 nM.

5.5.2. A3F and A3G co-expression results in higher levels of HIV restriction

To determine if the A3F/G hetero-oligomer influences the ability of the A3 enzymes to restrict HIV replication, we conducted single-cycle infectivity assays. To model cellular conditions where A3F and A3G are coexpressed we used a plasmid that has two transcription units enabling the combined expression of both A3F-V5 and A3G-HA on a single-cell basis. Flow cytometry to detect the V5 and HA tags demonstrated a distinct population of cells that expressed both A3F-V5 and A3G-HA when A3F-V5 and A3G-HA were expressed from the same plasmid (Figure 5.2A). The population of cells expressing both A3F-V5 and A3G-HA showed a normal distribution of cells with low (L, 26%), medium (M, 52%), and high (H, 22%) coexpression on a single-cell basis (Figure 5.2A). In contrast, a cotransfection strategy using separate plasmids expressing A3F-V5 and A3

G-HA showed a skewed distribution of the cell population with predominantly a low amount of cells coexpressing A3F-V5 and A3G-HA (Figure 5.2A, 62% low (L), 34% medium (M), and 4% high (H)). As a result, the virions generated from cell populations with a normal distribution of A3F-V5 and A3G-HA coexpression will be more likely to encapsidate more homogenous levels of each A3 enzyme across the population than the virions generated from cell populations with a skewed distribution of A3F-V5 and A3G-HA coexpression. Thus, our experiments used only A3F-V5 and A3G-HA that were coexpressed from one plasmid.

Using increasing amounts of A3 plasmid for A3G-HA (A3G), A3F-V5 (A3F) and A3G-HA/A3F-V5 coexpressed (A3F/G) we observed that the maximum amount of restriction occurred at a lower level of transfected plasmid when A3F and A3G were coexpressed (Figure 5.2B). Specifically, only 25 ng of the A3F/G plasmid was required for restriction of HIV compared to 50 ng of A3G plasmid or 100 ng of A3F plasmid. A3F mediated viral restriction was consistently less than A3G, as previously reported (*2, 22, 226, 266*). However, most notable was that for the 25 ng plasmid transfection experiment the combined expression of A3F and A3G resulted in 4- to 6- fold greater restriction than either A3G or A3F alone (Figure 5.2B, 25 ng). At 50 ng and 100 ng of transfection plasmid, the restriction of A3G was not significantly different than the combined restriction of A3F/G (Figure 5.2B, 50 ng and 100 ng). The infectivity data demonstrate that at lower levels of A3G and A3F expression, when their restriction effect is not saturated, their combined action is better than their individual action in inhibiting HIV infectivity. This effect is

similar to previous reports (*2*, *360*), although the reason for this enhancement of restriction has not been previously investigated.

To investigate if the enhanced effect was a result of higher encapsidation levels of A3G or A3F we analyzed cell lysates and virions by immunoblotting (Figure 5.2C). Since the A3s have different tags they were blotted with different antibodies and as a result we are unable to compare A3G to A3F. However, we can determine if the same amount of A3G or A3F is encapsidated under the different expression conditions. The virus blot showed that the same amount of A3G was encaspsidated in both the A3G and A3F/G experiments (Figure 5.2C). Similarly, A3F was encapsidated at the same level whether it was expressed alone or with A3G (Figure 5.2C). These results indicate that the enhanced restriction was not due to higher levels of enzymes being encapsidated and suggested that coencapsidated A3G and A3F were able to synergistically restrict HIV replication.

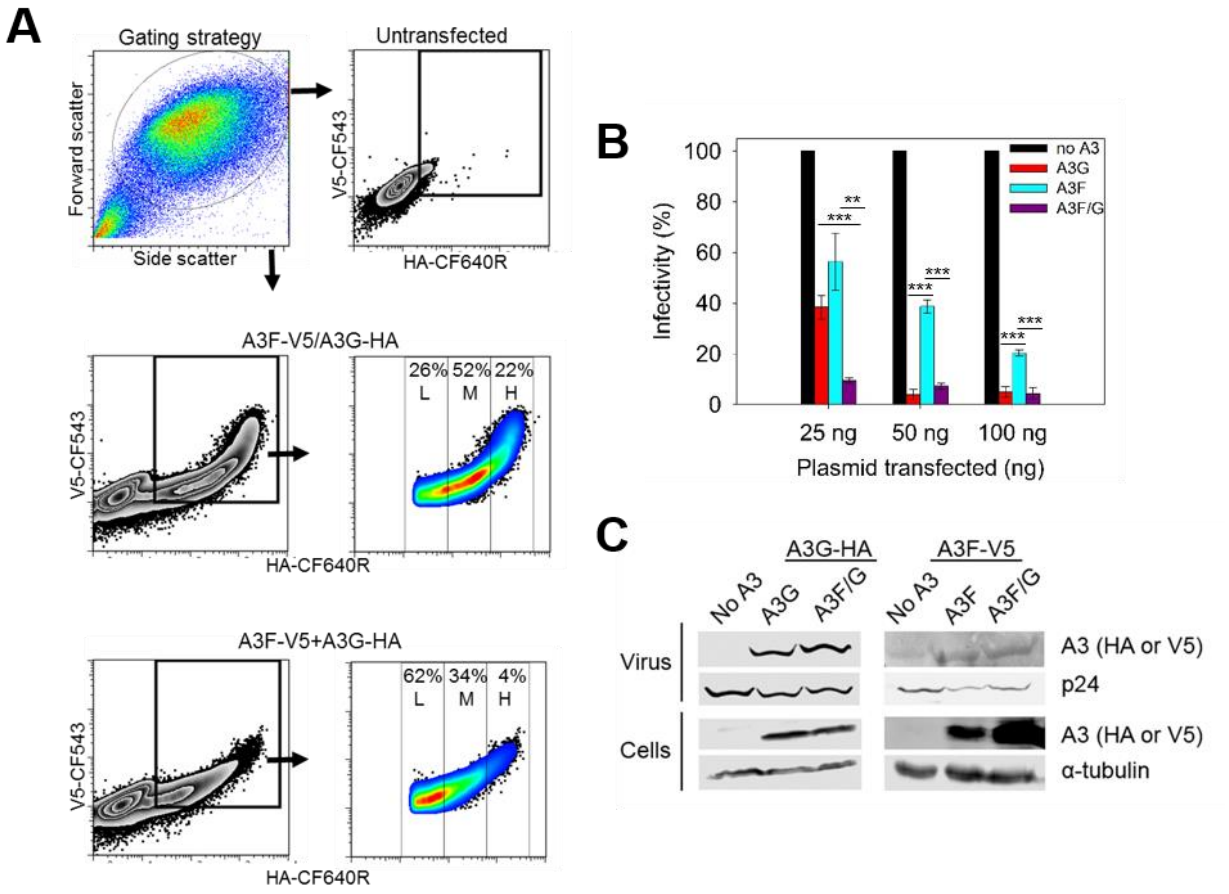

Figure 5.2. Coexpressed A3F and A3G enhances the restriction of HIV replication. (A) Flow cytometry was used to detect the A3G-HA or A3F-V5 in individual cells after transient transfection using fluorescently labeled anti-HA or anti-V5 antibodies. Cells were either transfected with one pVIVO2 vector expressing both A3G-HA and A3F-V5 or individual pVIVO2 vectors each expressing A3G-HA or A3F-V5. (B) HIV Δvif infectivity was measured by eGFP expression in 293T cells infected with HIVΔvif that was produced in the absence or presence of A3G-HA (A3G), A3F-V5 (A3F), or coexpressed A3F-V5 and A3G-HA (A3F/G). Results normalized to the no A3 condition are shown with the Standard Deviation of the mean calculated from at least three independent experiments. Designations for significant difference of values were $p \leq 0.001$ (***), $p \leq 0.01$(**), or $p \leq 0.05$ (*). (C) Immunoblotting of HA and V5 tags was used to detect A3 enzymes expressed in cells and encapsidated into HIV Δvif virions. The cell lysate and virion loading controls were α-tubulin and p24, respectively.

5.5.3. Enhanced mutagenesis of HIV Δvif induced by coencapsidated A3F and A3G

Recently it was shown that A3F and A3G are indeed able to be coencapsidated into HIV virions (*220*). To identify if one or both A3s were increasing the level of mutagenesis when coencapsidated we sequenced a portion of the proviral DNA. We used proviral DNA from the 25 ng A3 plasmid transfection experiment to avoid analyzing proviral DNA that may be saturated with mutations and PCR amplified the protease region for DNA sequencing. The mutation frequencies showed that A3G mutated the proviral DNA more than A3F (Table 5.2, A3G, 7.5 mutations/kb; A3F, 1.6 mutations/kb). When A3G and A3F were coexpressed, the mutation frequency was 14 mutations/kb (Table 5.2). If coencapsidated A3G and A3F were acting independently of each other the mutation frequency would be expected to be the sum of each mutation frequency alone (estimated 9.1 mutations/kb). That the mutations/kb when A3G and A3F were combined was 1.5-fold more than the sum of each mutation frequency alone suggests that A3G and A3F can coordinate to restrict HIV replication. When 50 ng of plasmid was transfected, we observed that for the A3F/G condition, the mutations were 2-fold greater than the sum of the mutations of A3G and A3F alone (data not shown). At 100 ng of plasmid, there was no increase in mutations in the A3F/G condition compared to the A3G and A3F alone, presumably because the restriction was reaching saturation which would make it difficult to differentiate between the conditions (Figure 5.2B, 100 ng and data not shown).

Since A3G causes mutations in the (+)DNA predominantly at GG sites and A3F predominantly at GA sites, sequence analysis can determine which A3 was inducing the increases in mutational load (*2*, *216*). Consistent with previous observations, we found that A3G alone mutated GG sites for 91% of the mutations (Table 5.2, 6.8 mutations/kb) and mutated GA sites for 9% of the mutations (Table 5.2, 0.7 mutations/kb). A3F predominantly mutated GA sites (Table 5.2, 94% of GA sites, 1.5 mutations/kb), but could also mutate GG sites at a low level (Table 5.2, 6% of mutations, 0.1 mutations/kb). This meant that we could use the sequence context to determine which A3 was inducing increases in the mutational load. For the 25 ng transfection condition, the combined mutation frequency (14 mutations/kb) was composed of 11 mutations/kb at GG sites (A3G induced) and 2.6 mutations/kb at GA sites (A3F induced) (Table 5.2). This is a 1.6-fold or 1.7-fold increase in GG mutations (Table 5.2, 6.8 mutations/kb increased to 11 mutations/kb) or GA mutations (Table 5.2, 1.5 mutations/kb increased to 2.6 mutations/kb), respectively. When A3F and A3G were coexpressed, we recovered both GG and GA mutations in

the same clone more frequently than when each A3 was expressed alone indicating that indeed both A3F and A3G were coencapsidated (Figure 5.3A). Specifically, for the A3F/G cotransfection (25 ng), 52% of mutated clones were comutated (Figure 5.3A). Of those not comutated, the mutations recovered contained only A3G-induced mutations at GG sites, consistent with there being a low mutation efficiency of A3F (Figure 5.3A and Table 5.2). For the 50 ng and 100 ng transfections, 75% of mutated clones were comutated (data not shown).

Despite only a 1.5-fold increase in mutagenesis, there was a 4-fold (A3G) to 6-fold (A3F) increase in restriction ability when the two enzymes were coexpressed in virus producer cells and coencapsidated in virions (Table 5.2 and Figure 5.2B). This inconsistency may be due to the ability of A3G to induce the formation of numerous stop codons in (+)DNA which would be able to completely ablate virus replication with only a single mutation (*216*). Stop codons arise from deamination of cytosines in (-)DNA Trp codons, e.g., 5'CCA. When the (-) DNA is deaminated to 5'UCA or 5'CUA the sequence becomes 5'TGA or 5'TAG in (+) DNA and results in a stop codon. Within the 351 nt long *protease* region that we sequenced, there was one Trp codon that was a hotspot for mutagenesis. A3G alone induced stop codon formation in 32% of the clones, A3F alone induced stop codon formation in 6.9% of the clones, and coencapsidated A3F/A3G induced stop codon formation in 50% of the clones (nucleotide positions 152 or 153). There are also numerous other A3-induced missense mutations that can inactivate protease (*226, 250, 341*). The inactivation of protease as a result of missense mutations was inferred from results from an extensive protease mutagenesis study conducted by Loeb *et al*. (*354*). Taking into account both stop codons and inactivation by missense mutations, we found that A3G alone inactivated 56% of mutated protease and left 8% percent of mutated protease active (Figure 5.3B). Due to the low A3 plasmid transfection level used in the experiment (25 ng), we recovered protease clones that were not mutated 36% of the time (Figure 5.3B). In contrast, A3F inactivated only 20% of mutated protease and left 14% mutated and active and 66% not mutated (Figure 5.3C). When A3F and A3G were coexpressed, there were no protease clones that were mutated and active (Figure 5.3D). There were 81% of protease clones that were mutated and inactivated and 19% left unmutated. This means that when A3F and A3G are coexpressed they are able to more efficiently inactivate the proviral DNA. Thus, although the combined effect of A3F/G on mutations was at most 2-fold, this enabled a larger decrease in infectivity (4- to 6-fold) and further supports that coencapsidation of A3G and A3F enhances HIV restriction from each A3 alone (Figure 5.2B and Table 5.2).

The mutations per clone were plotted to visualize the mutations across the sequenced protease region of the integrated provirus. Of the 351 nt clone, there are distinct hotspots in which A3G induces mutations (Figure 5.3E). The hotspots are deaminated in at least 20% of clones and there are 8 of these sites at positions 86, 151, 168, 170, 177, 243, 258, and 282. In the combined A3F/G condition, the same hotspots are deaminated in more clones, but no new hotspots were identified (Figure 5.3E and G, see hashed reference line). The increases in deamination at each of the sites ranged from 1.3-fold (nucleotide position 168) to 2.1-fold (nucleotide positions 86 and 243), suggesting that deamination activity across the protease gene was increased similarly (Figure 5.3E and G). For A3F, there were no distinct hotspots due to the low level of induced mutagenesis (Figure 5.3F). Increases in deamination for A3F in the A3F/G condition occurred at new sites or sites deaminated when A3F was expressed alone (Figure 5.3F-G, new sites: 106, 114, 129, 148, and 178). However, in accordance with the overall mutation frequency the increases were not high. The higher mutation frequency of A3F in the A3F/G condition was primarily due to new site mutations. All together the data indicate that for A3G and A3F, coexpression increases their mutagenic activity, but for A3G there is no change in the deamination hotspots, in contrast to A3F that is able to mutate more new sites during (-)DNA synthesis. Since A3G and A3F can form an A3F/G hetero-oligomer (Figure 5.1), we interpreted the mutation data to suggest that the A3F/G hetero-oligomer is working to increase deamination activity (observed for A3G) and accessibility to ssDNA (observed for A3F).

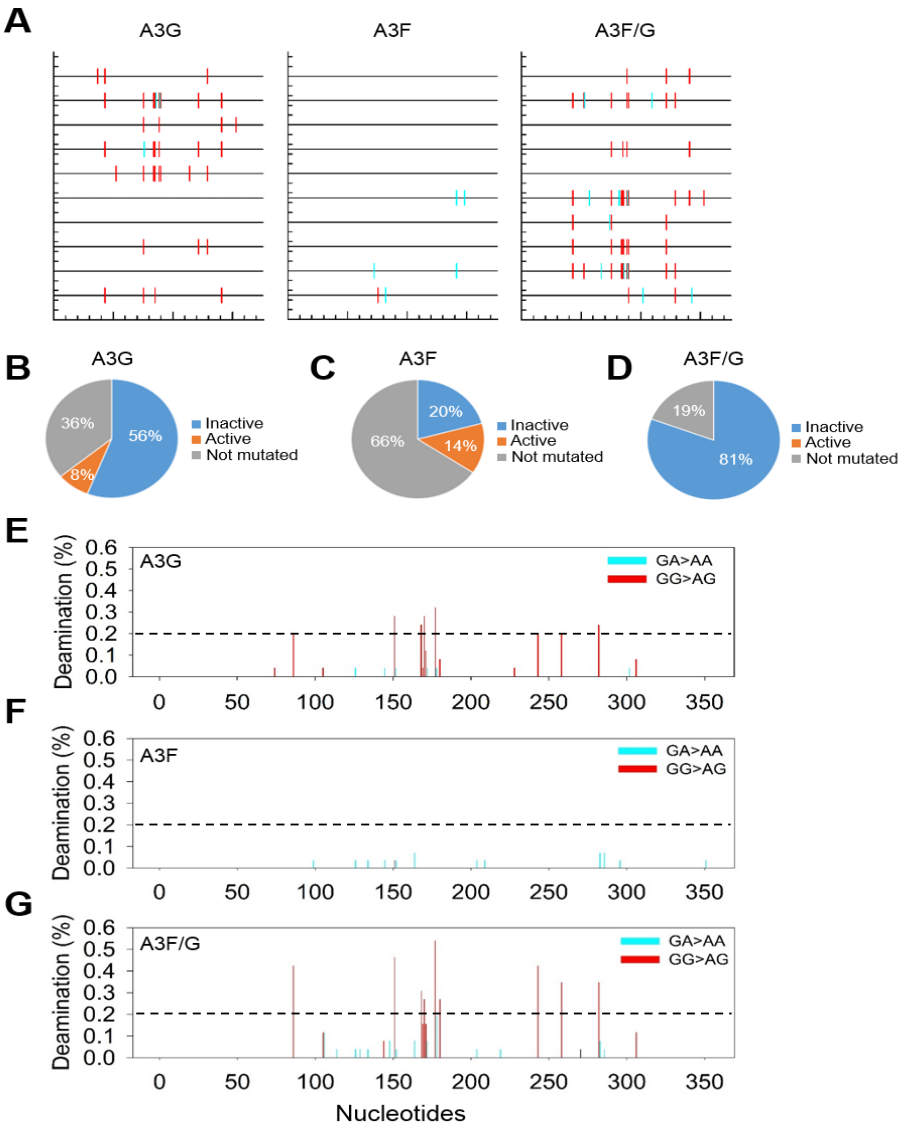

Figure 5.3. Coexpressed A3F and A3G commutate the same HIV proviral genome. (A-C) Individual analysis of each protease clone for (A) A3G, (B) A3F, and (C) A3F/G enabled determination of the percentage of clones that would result in a mutated and inactive (blue) or mutated and active (orange) protease. The percentage of clones that were not mutated are shown in grey. (D) Spectral plot generated by Hypermut (56) demonstrates the ability of coexpressed A3G and A3F to comutate the same genome. Representative samples from the 25 ng A3G, A3F, and A3F/G infectivity experiment were plotted. Across the protease gene, GG→AG (expected A3G induced mutation) are shown by a red line, GA→AA (expected A3F induced mutations) are shown by a cyan line, and GT→AT mutations (a single site) is shown in magenta. (E-G) Mutational spectra for all clones shows the percentage of clones with a mutation at a particular site in the protease gene for (E) A3G, (F) A3F, and (G) A3F/G. The GG→GA (red) and GA→AA (cyan) mutations are distinguished with color and demonstrate the specificity of the deamination targets for A3G and A3F.

Table 5.2. Analysis of A3-induced mutagenesis of protease DNA from integrated HIVΔvif

A3 enzyme	Base Pairs Sequenced	Total G→A mutations	Total GG→AG mutations	Total GA→AA mutations	Deaminations per kb	GG→AG mutations per kb	GA→AA mutations per kb
A3G	8775	66	60	6	7.5	6.8	0.7
A3F	10179	16	1	15	1.6	0.1	1.5
A3F/G	9126	129	104	24	14	11	2.6

5.5.4. A3G when part of an A3F/G hetero-oligomer has an improved ability to jump over RNA/DNA hybrids

To test our hypothesis that the A3F/G hetero-oligomer can increase the activity of A3G, we conducted *in vitro* deamination and binding assays on synthetic oligonucleotides. Using fluorescence depolarization, we found the A3F/G hetero-oligomer bound fluorescein labeled ssDNA with an apparent K_d that was unique from either A3F or A3G alone (Figure 5.4A). The A3F/G hetero-oligomer was formed by preincubating equal amounts of A3F and A3G. The A3F/G hetero-oligomer bound ssDNA with an apparent K_d of 114 nM, which is 3-fold less affinity than A3F and 2.5-fold more affinity than A3G alone. Since the binding experiment is conducted under steady state, the apparent K_d of the A3F/G hetero-oligomer is an average of all possible associations, A3F/G, A3G, and A3F. However, the data support the conclusion that the A3F/G hetero-oligomer may have distinct biochemical properties from A3G and A3F.

Previously we have shown that the processive scanning mechanism of the A3 enzymes that is used to search for the preferred deamination motifs in the DNA substrate can be a predictor of mutagenic potential (*226*). Therefore, we hypothesized that since one A3G appeared to interact with three molecules of A3F (Figure 5.1A-D), that A3F may be able to influence the processive scanning mechanism of A3G. A3F is able to jump larger distances than A3G (*226*). Jumping movements are required for overcoming in a processive manner RNA/DNA hybrid obstacles left on (-)DNA by RNaseH activity (*213*). To examine the processivity and scanning mechanism of the A3, the reactions are conducted under single-hit conditions where an ssDNA is encountered by only one enzyme during the course of the reaction (*344*). To characterize processivity and scanning movement we used an ssDNA with either two CCC motifs (for A3G activity) or two TTC motifs (for A3F activity) separated by differing distances. Under our reaction conditions we observed no deamination of CCC motifs by A3F or TTC motifs by A3G (data not shown). Thus, we examined how the A3F/G hetero-oligomer could improve activity of either A3G or A3F by using the relevant substrate. To form the hetero-oligomer, we preincubated A3F and A3G together, added the enzymes to the reaction buffer and started the reaction by the addition of ssDNA. We did not subsequently purify the A3F/G hetero-oligomer from unbound A3F or A3G to simulate cellular conditions where there is likely to be an "average" population of A3F, A3G and the A3F/G hetero-oligomer.

To examine jumping we used deamination motifs that were distantly spaced (61 nt). On the CCC motif containing ssDNA, A3G and the A3F/G hetero-oligomer had the same processivity factor of approximately 8 (Figure 5.4B). This processivity factor is calculated using the integrated gel band intensity of the 5'C & 3' C deamination band and means that the enzyme is 8-fold more likely to make a processive deamination than a nonprocessive deamination (see Material and Methods). To challenge the jumping ability of A3G and the A3F/G hetero-oligomer we annealed a complementary 20 nt RNA between the two CCC motifs (Figure 5.4C, sketch). Since A3G and A3F cannot bind to this RNA/DNA hybrid, the enzyme complex must jump over the obstacle to undergo processive deamination of both CCC motifs (*226*). A3G had a characteristic decrease in processivity due to the obstacle (Figure 5.4B-C, 2.5-fold) (*226, 326*). When in complex with A3F, the A3G processivity only decreased 1.3-fold (Figure 5.4B-C). Thus, for A3G, being in complex with A3F (A3F/G) can improve the jumping ability. We also conducted the equivalent experiment for A3F, but used the TTC motif containing ssDNA. For A3F alone, the processivity is enhanced by annealing a complementary RNA molecule, presumably due to the annealed portion enabling the TTC sites to be juxtaposed at an optimal distance for A3F jumping movements, as previously observed (Figure 5.4D-E) (*226*). There was no change in the A3F processivity factor on either DNA substrate in the presence of A3G (Figure 5.4D-E, A3F or A3F/G).

To examine processive sliding movements we used an ssDNA with closely spaced deamination motifs (3 nt apart). On the CCC containing ssDNA we found no change in the A3G processivity in the presence of A3F (Figure 5.4F). A3F is not able to processively slide to deaminate closely spaced TTC motifs, thus, on the substrate with the TTC motifs separated by 3 nt we observed no deamination of both the 5C' & 3'C (Figure 5.4G). Despite A3G being able to slide, there was no 5'C & 3'C band detected for the A3F/G hetero-oligomer on the TTC containing ssDNA (Figure 5.4G). Altogether, the data demonstrate that A3F influences the processive jumping movements of A3G when they form a hetero-oligomer, but A3F processivity is not influenced by A3G. This is consistent with there being more A3F molecules in the A3F/G hetero-oligomer that could exert more of an effect on A3G than *vice versa* (Figure 5.1).

These data are also consistent with the A3F/G hetero-oligomer having 1.7-fold higher specific activity on CCC motif ssDNA than A3G alone and no effect on the specific activity on TTC motif ssDNA. Specifically, this was tested by using an ssDNA substrate that had either a deamination motif for A3F (5'TTC) or A3G (5'CCC) and the ratio of the enzymes were varied. We

found that there was no appreciable effect of increasing A3G concentration on the deamination mediated by A3F on ssDNA containing two 5'TTC motifs (Figure 5.5A). However, for A3G, we found that the addition of A3F (ratios of A3G:A3F of 1:1 and 1:2) enhanced the rate A3G deamination on the 5'CCC containing ssDNA substrate containing up to 1.7-fold (Figure 5.5B). After A3F reached high concentrations, e.g., ratio of 1:5, the rate of A3G-mediated deamination decreased (Figure 5.5B). These results suggest that over a certain A3G:A3F ratio, A3F can impede A3G-mediated deamination. This may be due to the higher binding affinity of A3F for ssDNA that occludes A3G deamination sites (Figure 5.4A).

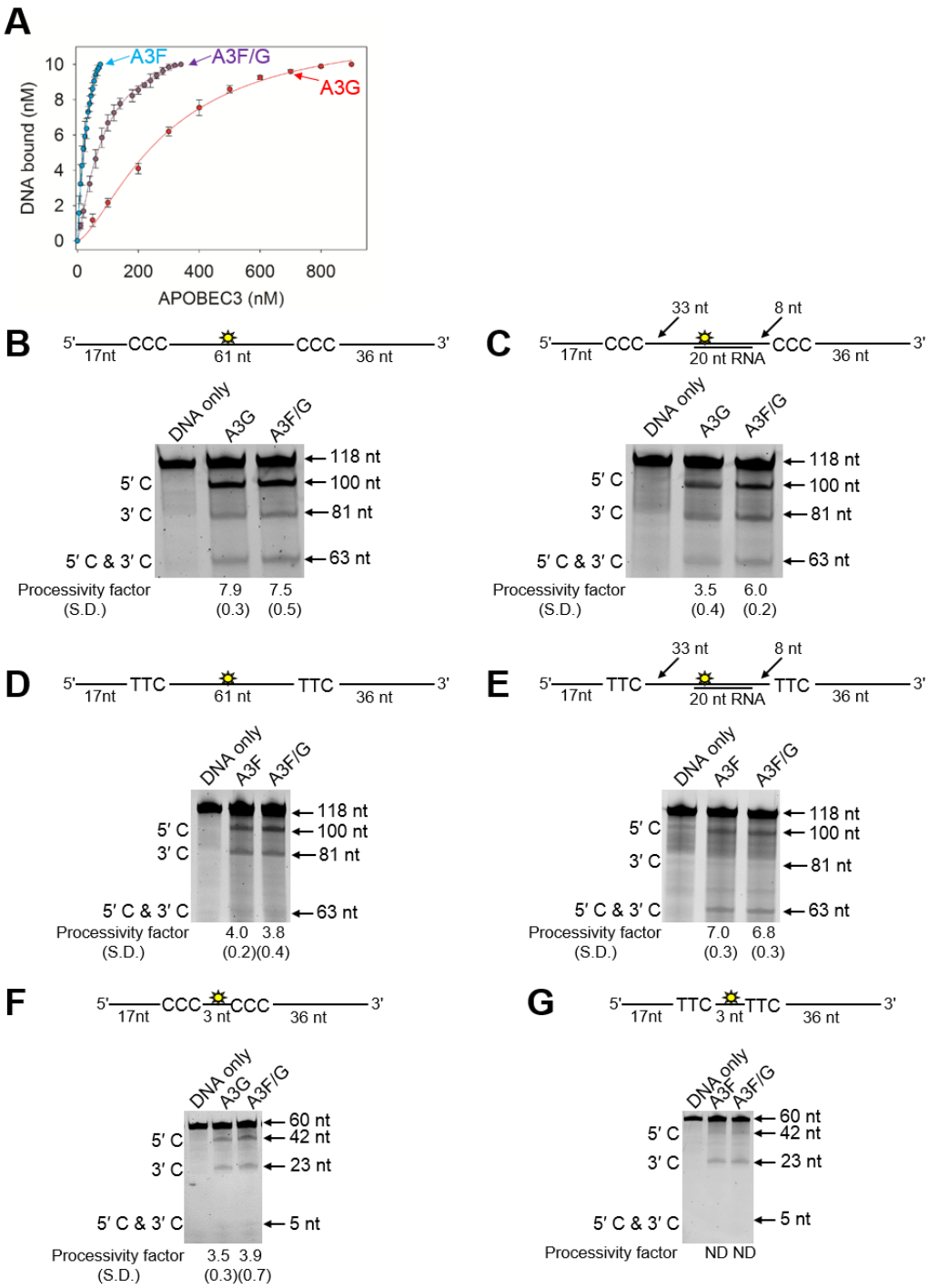

Figure 5.4. Biochemical properties of the A3F/G hetero-oligomer are distinct from A3F and A3G. Fluorescence depolarization was used to detect changes in rotational anisotropy of a F-labeled 118 nt ssDNA upon titration of A3G, A3F, or an A3F/G hetero-oligomer into the solution. The rotational anisotropy was normalized to fraction F-ssDNA bound and analyzed by regression analysis. The binding curves fit to a sigmoidal binding curve as determined by least squares analysis. The apparent K_d values were calculated to be 286 ± 17 nM for A3G, 39 ± 6 nM for A3F,

and 114 ± 20 nM for A3F/G. Error bars represent the Standard Deviation of the mean from three independent experiments. (B-G) Processivity of A3G, A3F, or A3F/G was tested on ssDNA substrates that contain a fluorescein-labeled deoxythymidine (yellow star) between two 5′CCC (for A3G) or 5′TTC (for A3F) deamination motifs separated by different distances. (B) Deamination of a 118 nt ssDNA substrate with two 5′CCC deamination motifs spaced 61 nt apart. Single deaminations of the 5′C & 3′C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. (C) Deamination of the same substrate shown in (B), but with a 20 nt complementary RNA annealed between the two 5′CCC motifs. (D) Deamination of the same substrate shown in (B), but with two 5′TTC motifs. (E) Deamination of the same substrate shown in (D), but with a 20 nt complementary RNA annealed between the two 5′TTC motifs. (F) Deamination of a 60 nt ssDNA substrate with two 5′CCC motifs spaced 3 nt apart. Single deaminations of the 5′C & 3′C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment. (G) Deamination of the same substrate shown in (F), but with two 5′TTC motifs. The notation “ND” means Not able to Determine due the absence of a detectable 5′C & 3′C band. The measurements of enzyme processivity (processivity factor) and the S.D. are shown below the gel. All values are calculated from three independent experiments.

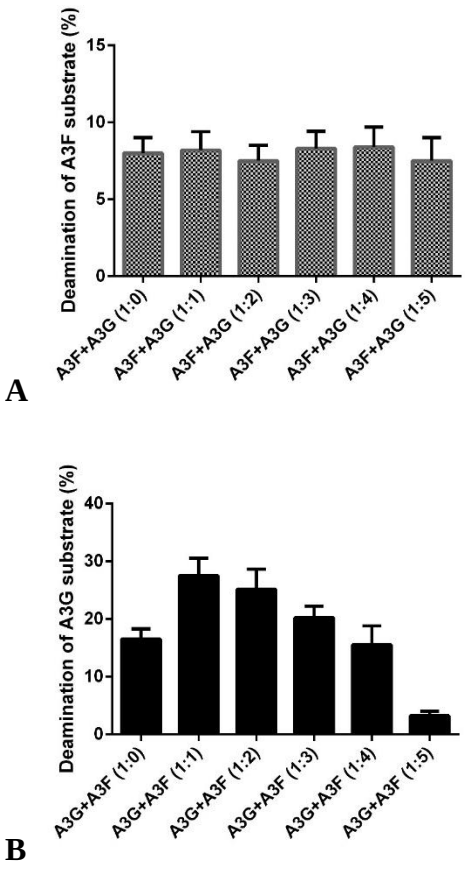

Figure 5.5. A3F influences the deamination activity of A3G. The deamination activity of A3F/G on a (A) 5'TTC or (B) 5'CCC substrate was determined in the presence of (A) 30 nM A3F with increasing concentration of A3G (30, 60, 90, 120, and 150 nM) or (B) 30 nM A3G with increasing concentrations of A3F (30, 60, 90, 120, and 150 nM). (A) A3G does not affect A3F deamination activity (gray bar), but (B) A3F can increase A3G deamination activity 1.7-fold at an equimolar ratio (black bar).

5.5.5. A3F and A3F/G decrease reverse transcripts more efficiently than A3G

Despite A3F induced mutations being increased in the presence of A3G (Table 5.2), the reason for this did not appear to be due to an increase in the processive search mechanism of A3F. Another mechanism by which A3 induced mutations can be increased is if reverse transcription was slowed down, which would leave (-)DNA single-stranded for a longer time (*355, 382, 383*). It is known that both A3G and A3F can decrease reverse transcriptase efficiency, which results in less late reverse transcript (LRT) formation and less proviral integration (*229, 255, 281, 284, 287-289, 291, 294, 296, 384, 385*). We hypothesized that the A3F/G hetero-oligomer would inhibit reverse transcriptase more efficiently and this would allow more time to search for and deaminate TC motifs in (-)DNA.

To investigate this we conducted experiments that quantified the LRT in infected 293T cells and *in vitro* primer extension of HIV reverse transcriptase from a PBS primer in the presence and absence of A3s. The relative LRT formed in the presence of the A3s was decreased for all conditions with A3G, A3F, or A3F/G. Consistent with previous findings, A3F inhibited LRT formation more than A3G (Figure 5.6A, 1.6-fold difference) and the inhibition of LRT formation was independent from deamination ability (Figure 5.6A, catalytic E→Q mutants) (*229, 255*). The A3F/G hetero-oligomer suppressed LRT formation approximately 2-fold more than A3F alone. These data were consistent with a primer extension assay in which purified HIV reverse transcriptase was used to extend the 18 nt PBS primer from a 106 nt template that had a nucleotide sequence matching the corresponding region of the HIV genome (Figure 5.6B, sketch). For the A3F/G condition, equimolar amounts of A3F and A3G were preincubated before adding to the reaction. In all reactions there was the same total moles of A3 for each titration. By titrating increasing amounts of A3 into the reaction we observed an increasing amount of inhibition of DNA polymerization (Figure 5.6B). Analysis of the primer extended showed that each A3 inhibited total primer extension similarly (Figure 5.6C). However, visual inspection of the gel for the 4:1, 8:1, and 32:1 A3:primer/template (p/t) conditions demonstrated that A3F and the A3F/G hetero-oligomer can inhibit the formation of full length products (82 nt) at the 4:1 molar ratio whereas at least 2-fold more A3G is required for the same inhibition (Figure 5.6B and D). Altogether, the data indicate that the presence of either A3F or the A3F/G hetero-oligomer can slow down RT catalyzed DNA synthesis and this probably occurs by a “road-block” mechanism

that has previously been proposed (Figure 5.7) (*294*). This may enable the A3 enzymes more time to deaminate the (-) DNA.

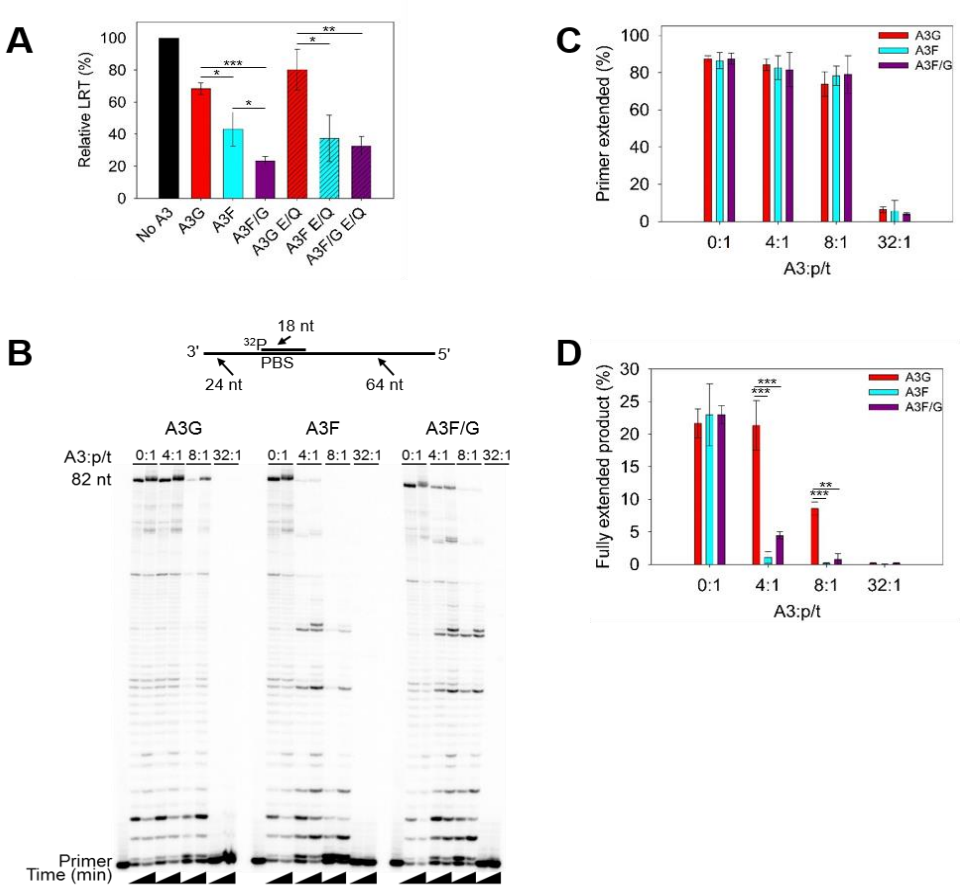

Figure 5.6. A3 enzymes can decrease reverse transcriptase efficiency. (A) Quantification of late reverse transcripts (LRT) by qPCR demonstrated that A3G, A3F, and A3F/G and their catalytic mutants can decrease LRT relative abundance. The E/Q notation means an E259Q mutation for A3G or an E251Q mutation for A3F. Error bars represent the standard deviation of the mean from three independent experiments. (B) An 18 nt ^{32}P-labeled RNA primer containing a sequence complementary to the HIV PBS was annealed to a 106 nt RNA containing the PBS (sketch). Complete extension of the primer results in a product of 82 nt (sketch). The p/t was used at a concentration of 10 nM. Primer extension by reverse transcriptase (480 nM) in the absence (0∶1) or presence (4∶1, 8∶1, 32∶1) of increasing amounts of A3G, A3F, or A3F/G relative to the p/t concentration. Reactions were sampled at 2.5 and 60 min. (C) Quantification of primer extension (%) shown in panel B from gels shown in panel B for no A3, A3G, A3F, and A3F/G at 60 min. (D) Quantification of fully extended 82 nt product (%) for no A3, A3G, A3F, and A3F/G at 60 min. (C-D) Error bars represent the Standard Deviation of the mean from three independent experiments. (A, C-D) Designations for significant difference of values were $p \leq 0.001$ (***), $p \leq 0.01$(**), or $p \leq 0.05$ (*).

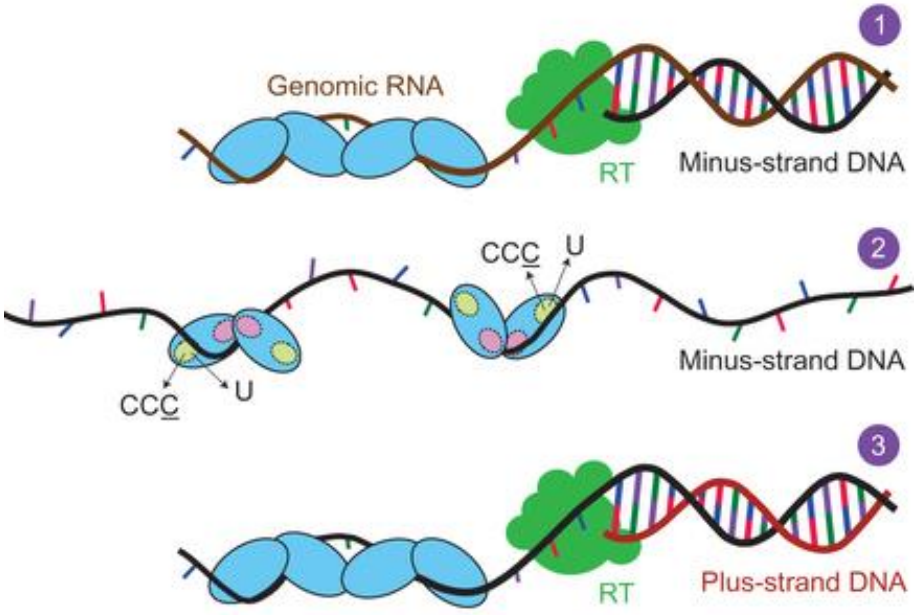

Figure 5.7. A road-block model for decreasing reverse transcriptase efficiency by A3. Step 1. A3 binds (blue) with the RNA genome (orange), blocking the synthesis of negative strand DNA (black) synthesis by RT (green). Step 2. A3 enzymes induce deaminations on exposed ssDNA. Step 3. A3 binding of ssDNA can also block RT-mediated plus-strand DNA (red) synthesis. Modified with permission from (*294*).

5.6. Discussion

A3 enzymes are coexpressed in $CD4^+$ T cells and have been shown to coordinately restrict HIV replication (*19, 24, 31, 32*). However, the mechanism of this coordinate restriction is not clear. For A3G and A3F it has been proposed that each contributes individually to HIV mutagenesis due to not being coencapsidated (*376*), that A3F and A3G enzymes are coencapsidated but act independently of one another (*2, 220*), and that A3 enzymes can enhance each other's activity (*2, 360*). Thus, we undertook a study of A3F and A3G to examine how they may coordinately restrict HIV infectivity when expressed in the same HIV producer cell. We observed that A3F and A3G can form an A3F/G hetero-oligomer in cells and *in vitro* in the absence of an RNA intermediate and inferred that they can frequently coencapsidate into HIV particles due to 52-75% of proviruses being comutated. Overall, our data shows that the A3F/G hetero-oligomer has enhanced ability to decrease the infectivity of HIV through increased mutagenic activity resulting from unique biochemical properties of the A3F/G hetero-oligomer compared to the individual enzymes.

5.6.1. A3F/G oligomerization

We determined that A3F and A3G hetero-oligomerize using several methods. Using uniquely tagged versions of A3F and A3G we conducted a co-IP similar to what was done originally by Wiegand *et al*., except that we added RNaseA to the cell lysates (*24*). Even in the absence of RNA A3G-HA could immunoprecipitate A3F-V5 indicating that they hetero-oligomerized in cells (Figure 5.1E). The complex has a high affinity association in the nanomolar range and is stable enough to be identified through SEC (Figure 5.1A-D and F). We previously determined by SEC that A3F was approximately 158 kDa in the peak fraction, which would be a trimer or tetramer (3.5-fold the molecular weight of a monomer) (*226*). However, using a similar method and column, but with quantitative immunoblotting to better visualize the bands, we identified the A3F peak fraction more exactly to be 153 kDa, which confirmed that the molecular weight is closer to that of a trimer than a tetramer. The oligomerization of A3F does not appear to be concentration dependent since it formed these trimers at protein concentrations in the low micromolar range (Figure 5.1B-C, 1 μM A3F used for SEC). In contrast the oligomerization of

A3G is concentration dependent. In the absence of RNA at low concentrations A3G is primarily a monomer and at higher concentrations it forms predominantly dimers (Figure 5.1) (*205, 226, 247, 381*). This dynamic nature of A3G oligomerization may facilitate the ability to oligomerize with A3F (Figure 5.1). Conversely, the stable oligomer of A3F may inhibit multiple

A3G molecules from binding. With 50% amino acid similarity between A3F and A3G it is reasonable to assume that they may similarly oligomerize. However, making a Y131A/W132A mutant in A3F, which is equivalent to the dimer disruption mutant F126A/W127A of A3G, did not disrupt A3F oligomerization (data not shown) suggesting that A3F oligomerizes in a manner unique form A3G. A crystal structure of rhesus A3G N-terminal domain (NTD) shows that A3G (NTD) dimerizes using helix 6 and loop 7, where F126/W127 are located, confirming previous mutational studies (*246, 247, 386*). Further this dimerization of rhesus A3G NTD is distinct from any of the existing A3F dimer models that have resulted from A3F C-terminal domain structures (*263, 314, 387, 388*). It remains to be determined how full length A3F oligomerizes in solution. However, since A3F does not tetramerize by itself (Figure 5.1B-C), this suggests that for a tetramer to form, a different association interface must be used. Since the A3F/G hetero-oligomer formed in cell lysates, it is likely to also form in virions where the apparent concentration of the components is estimated to be in the high micromolar range due to the restricted volume of the capsid (Figure 5.1E) (*389*).

5.6.2. A3G and A3F cooperate to restrict HIV replication

All studies to date on A3G and A3F coencapsidation have found different results. Most recently Desimmie *et al.* showed on a single virion level that A3G and A3F are indeed coencapsidated (*220*). In the experimental system the authors transfected an HIV plasmid and separate plasmids for each A3G and A3F. This resulted in ~50% coencapsidation (*220*). The coencapsidation in our study is implied by our data in which we recovered proviruses that had mutations at both GG and GA, representing mutation sites due to A3G and A3F deamination, respectively. However, we observed that 52-75% of the proviruses were comutated, depending on the transfection conditions. At the same approximated restriction conditions, we recovered 1.5-fold more comutated viral genomes than Desimmie *et al.*(*220*) (compare 75% in this study with 49.6%). In contrast to our study, the authors found that coencapsidated A3G and A3F had an additive, not cooperative effect (*220*). This may be due to the different A3 expression strategy. Desimmie *et al*. used separate plasmids to express the two A3s whereas our study used one plasmid to express A3G and A3F thereby ensuring each transfected cell had both enzymes expressed, which may have resulted a more defined transfection population (Figure 5.2A) (*220*).

Further, we titrated our expression plasmids to avoid saturation of the mutations that may obscure any cooperative effects (Figure 5.2B). Interestingly, Krisko *et al*. found that in a BLT

humanized mouse model that consistent HIV Δvif restriction required the combined antiviral activities of A3G and A3F and that A3F was contributing through a deamination independent mechanism (*360*). Despite more consistent HIV restriction in the presence of both A3F and A3G, the mutations were predominated by GG→GA and the GA→AA represented only ~1% (*360*). These data are in agreement with our observation that A3G deamination activity is being enhanced by A3F and that A3F deamination activity was low (Figure 5.3E-G). Our studies are in further agreement that A3F has a more robust activity to inhibit reverse transcriptase polymerization than A3G when acting alone or in the presence of A3G (Figure 5.6). However, data from Krisko *et al*. are in contrast to another humanized mouse model study in which HIV proviral DNA was found to contain mutations that resulted from only A3G deamination or predominantly A3F deamination, suggesting that A3G and A3F act individually (*390*). In the earliest study to combine A3G and A3F plasmids in the producer cell, Liddament *et al*. showed a synergistic 2.7-fold increase in the mutations caused by A3G and A3F coexpression compared to each expressed alone, but only an additive effect on infectivity (*2*). Thus the authors concluded that A3G and A3F have both an independent and synergistic (which they termed "dependent") ability to induce mutagenesis of proviral DNA (*2*). Although we observed a synergistic effect for infectivity (4- to 6- fold) and a lesser effect on mutagenesis (~2-fold), the conclusions of Liddament *et al*. are in agreement with ours (*2*).

5.6.3. Independent and dependent modes of A3F and A3G restriction

As discussed, previous studies on A3F and A3G coencapsidation and corestriction are largely complimentary to our work. However, our study identified possible mechanisms that can enable coencapsidated A3F and A3G to act cooperatively. Although A3F and A3G were identified to be coexpressed in CD4+ T cells and hetero-oligomerize in an early study, the role of RNA in mediating the hetero-oligomerization was not clear (*24*). Defining if their interaction was mediated by an RNA intermediate is important to identify because it provides information on whether they act independently while coencapsidated, such as A3F inhibiting reverse transcriptase polymerization while A3G deaminates the (-)DNA or if they are dependent on each other through forming a hetero-oligomer.

That we identified A3F and A3G to form a hetero-oligomer in the absence of RNA *in vitro* and in cells (Figure 5.1), led us to investigate a dependent mechanism of cooperation. The SEC indicated that there were likely more molecules of A3F in the hetero-oligomer than molecules of

A3G (Figure 5.1B-D). These data imply that if equal amounts of A3G and A3F are coencapsdiated, that there would be a population of free A3G molecules, A3F/G hetero-oligomers, and a minority of free A3F molecules. We tested if the A3F/G hetero-oligomer had properties that could enhance HIV restriction and found that A3G in the A3F/G hetero-oligomer was better able to processively scan ssDNA with a RNA/DNA hybrid region than A3G alone (Figure 5.4B-C). The processive scanning mechanism of an A3 has been shown to be a determinant of restriction efficiency (*226*).

The A3F/G hetero-oligomer was able to slow down the polymerization efficiency of reverse transcriptase (Figure 5.6B-D). In the primer extension assay, we kept the total moles of enzyme equal for experiments with A3G, A3F, and A3F/G. Thus in the A3F/G condition, there is half the amount of A3F compared to the A3F alone condition. Despite this, the A3F/G hetero-oligomer could inhibit synthesis of full length product as efficiently as A3F, further supporting the conclusion that the A3F/G hetero-oligomer has distinct biochemical properties and functionality. None of the conditions resulted in the inhibition of primer initiation, consistent with other studies that were focused on A3G (Figure 5.6C) (*229, 283, 288*). However, A3G, A3F and A3F/G could inhibit synthesis of full length product (Figure 5.6D). The analysis of inhibition of full length product formation are consistent with cellular data and demonstrate that the ability of A3F to inhibit reverse transcriptase is greater than A3G (Figure 5.6A and D). This may be due to A3F binding ssDNA with higher affinity than A3G (Figure 5.4A). Consistent with the A3G model that oligomers on the ssDNA block reverse transcriptase progression, both A3F and the A3F/G hetero-oligomer form trimers and tetramers in solution, respectively, and bind ssDNA cooperatively indicating that they further oligomerize on ssDNA, suggesting a mechanism of action similar to A3G (Figure 5.4A) (*294*). The slower reverse transcription that occurred *in vitro* and in cells in the presence of the A3s could provide more time to deaminate (-)DNA resulting in an independent mechanism of synergy where one A3 oligomer is inhibiting the reverse transcriptase and another is deaminating the DNA (Figure 5.6A-B). This may facilitate increased deamination of A3F in the A3F/G hetero-oligomer. The A3F/G hetero-oligomer bound ssDNA with approximately 2-fold less affinity than A3F alone (Figure 5.4A). The high affinity of A3F for ssDNA has been suggested to inhibit its movement on DNA, adding to its inefficiency with which it searches the substrate for deamination motifs (*226*). The higher K_d of the A3F/G hetero-oligomer may increase the ability of A3F in the A3F/G hetero-oligomer to search the (-)DNA before it becomes double stranded (*226*). In combination with the decrease in reverse transcriptase efficiency imposed by A3

enzymes, this provides an explanation for the increased number of A3F induced mutations when A3F and A3G are coexpressed (Figure 5.3, Figure 5.6, and Table 5.2).

5.7. Conclusion

Overall, the data support that A3F and A3G in cells can form a hetero-oligomer with unique properties from the individual enzymes. The extent to which A3F and A3G oligomerize with each other in HIV infected $CD4^+$ T cells could not be determined due to the lack of a high sensitivity antibody for A3F. However, it is likely that these interactions identified here in cell culture and *in vitro* may contribute to the hypermutation identified in proviral genomes of HIV infected people since the higher mutagenic activity could enable more efficient mutagenesis despite the presence of Vif (*178, 220*). It is also possible that other A3s may hetero-oligomerize and synergistically restrict HIV. All together our data identify a mechanism that may contribute to enhanced activity of A3s and raises the idea that we may still not fully realize all the antiviral mechanisms of endogenous A3 enzymes.

CHAPTER 6. GENERAL DISCUSSION AND FUTURE DIRECTIONS

In the past 15 years, there has been an explosion of knowledge on the ability of human antiviral restriction factors, including the A3 family of enzymes that act as the first line of host defense against retroviral infections. In the absence of Vif-mediated degradation, A3 enzymes inhibit viral replication by causing lethal levels of mutagenesis as well as through a deamination-independent mechanism that includes physical inhibition of reverse transcription and integration (*289, 291*). Out of seven, only four A3 (A3D, A3F, A3G and A3H) have been found to be relevant for HIV-1 (referred to as HIV) restriction in the majority of people. In people of African origin, it has been found that 10% of the population can carry an A3C S188I polymorphism that enables A3C to restrict HIV (*203, 338*).

When I started my graduate thesis research, the A3 proteins were characterized by extensive cell-based studies for their HIV restriction ability. Yet, there was considerable disparity among the reports regarding HIV restriction ability of A3F compared to A3G. Most of these studies compared A3F and A3G based on end-point analysis, i.e., infectivity assays and mutational analysis of proviral DNA (*258, 266, 287, 330*). However, there are three key steps that A3 enzymes must complete to be an efficient HIV restriction factor. First A3 enzymes must escape Vif antagonism, Second, A3 enzymes must be available for binding RNA that will become virion-encapsidated. Third, A3 enzymes must have a mechanism to search the nascent HIV (−) DNA. However, it was not clear whether the differential HIV restriction by A3F and A3G was the result of their distinct DNA searching ability or levels of virion encapsidation.

We undertook a biochemical study of A3F in comparison with A3G for investigating the underlying mechanisms behind the different restriction ability of these two proteins. A previous biochemical study on A3G-mediated deamination showed that A3G uses sliding and jumping motions along the ssDNA to locate and processively catalyze multiple target motifs (*227*). In this

study, using a series of synthetic ssDNA substrates, we provided the first biochemical model for A3F-catalyzed processive C→U deamination. The DNA scanning mechanisms of A3F relies heavily on jumping motions, which is different than both the jumping and sliding motions of A3G. Mechanisms of A3G-catalyzed processive deaminations of ssDNA were questioned by Nowarski *et al*., 2009 (*205*). They suggested a model in which A3G uses intersegmental transfer mechanisms to deaminate its motifs (*205*). However, using more sensitive deamination assays, we provided the evidence that A3F but not A3G may undergo intersegmental transfer to a minor extent. The discrepancy may be that Nowarski *et al*. did not use any salt in their buffers. Our buffers have 10 mM $MgCl_2$ and 40 mM KCl, which creates a more competitive electrostatic environment for enzyme-DNA interactions and is more physiologically relevant. Thus, our data support the sliding and jumping model (*227*), and refutes the intersegmental model (*205*) of A3G-mediated deamination.

Interestingly, we also found that A3F lacks sliding motions on ssDNA due to a single Pro at ^{190}NPM^{192} motif which is present in the interconnecting domain between NTD and CTD. The recent crystal structure of A3F CTD includes the ^{190}NPM^{192} sequence shows that it is a kinked region of a loop structure (*263*). Our study suggests that the Pro in the ^{190}NPM^{192} motif probably provides rigidity to the structure since the replacement of ^{191}P to ^{191}G enabled A3F sliding movements to occur. Of note, insertion of ^{195}NPM^{197} motif between NTD and CTD of A3G abrogated its sliding motion on ssDNA, suggesting that the interconnecting domain may be important in coordinating NTD and CTD for sliding. The importance of sliding is also evident from another study by our lab where Feng and Chelico., 2011 showed that the H186R mutant of A3G is less mutagenic due to lack of sliding motion (*250*). Furthermore, Browne *et al*., 2009 reported clusters of A3G-induced deaminations in integrated proviral genomes and we assume based on our data that such clustered mutations are indicative of processive deaminations accomplished by sliding motions (*254*). In our *in vitro* and cellular experiments, we did not observe A3F deaminations that were clustered, supporting data that A3F does not slide on ssDNA.

At the beginning of this study, we had the assumption that tighter enzyme binding to the substrate would correlate with higher deaminase activity. In this study, we found that the binding affinity of A3F for ssDNA was tighter than A3G, but contrary to our assumption, we found that the specific activity of A3F was 100-fold lower in comparison to A3G. Thus, the notion that associating with high affinity to the substrate will lead to high enzyme activity may not be true in

the case of A3 enzymes whose activity during HIV-1 proviral DNA synthesis largely depends on the efficient search of target motifs within a short window of time. It is conceivable that the high ssDNA affinity of A3F may be a bottleneck for excursions or frequent movements on the ssDNA that attribute to the inefficient search for deamination motifs. However, resolution of this assumption awaits single-molecule analysis.

Other than the scanning mechanisms, A3G also has the advantage of being more effective in restricting HIV because it has a preferred deamination site of 5'CC. The 5'CC in the (-)DNA overlaps with the only Trp codon (5'CCA/TGG) and results in a stop codon upon deamination of either cytosine in the motif (*216*). For example, deamination of 5'CUA would result in 5'TAG, an "amber" stop codon. In contrast, mutations induced by A3F are mostly missense mutations which may or may not lead to inactivation of the encoded protein. Overall, our findings support a model in which both the processive DNA scanning mechanism and preferred deamination motif (A3F, 5'TTC; A3G 5'CCC) influences the mutagenic and gene inactivation potential of an A3 enzyme, consistent with data from other labs (*331, 359*).

In agreement with several previous reports (*2, 19, 21, 23, 215*), our data suggest that both A3F and A3G have anti-HIV potential, but A3G is more effective. In contrast to our results in Chapter 3, and previous findings (*2, 21, 23, 215*), Miyagi *et al*., 2010 suggested that A3F has negligible antiviral activity (*258*). Miyagi *et al*. argued that studies reporting anti-HIV potential of A3F had used transient overexpression of A3F as opposed to their stable physiological expression. However, using highly sensitive flow cytometry and mutational spectra of proviral DNA, Refsland *et al*. reported that endogenous levels of A3F restrict HIV in the CEM T-cell line (*19*). Recently a study by Ping *et al*., 2016, provided the *in vivo* evidence for an anti-HIV function of A3F in HIV natural cohorts and reported that an A3F I231V variant that is resistant to Vif degradation is associated it with slow progression of AIDS in European American patients (*391*). Based on our results from Chapter 5, differences in these findings between labs may be due the amount of A3G that was expressed in the cell lines used in experiments, since A3F restricts HIV better in the presence of A3G. In our cell-based studies, we suggested that A3F may contribute to viral evolution due to the induced missense mutations. This was also supported by the literature at the time (*235-244*). However, a new study using different methods and computational analysis of hypermutated proviral DNA from HIV+ individuals suggests that A3F- and A3G- induced mutations in the presence of Vif have negligible effects on viral evolution (*245*). As a result, there

are more studies that need to be done in this area to arrive at a consensus. However, our data at this time supports that as an individual enzyme A3F can contribute to HIV-1 evolution, but A3G is unlikely to contribute.

Our knowledge of A3 enzyme structure and function is growing rapidly and research is on-going to understand how A3 enzymes interact with nucleic acids and Vif and to develop novel therapies. However, more studies are still needed to fill in some of the gaps. Based on our data and given the fact that primary cells harbour multiple A3 proteins that are coordinately coexpressed, more research should examine whether heterologous A3-A3 interactions do influence anti-HIV function in primary cells, and if this will influence development of novel therapies relying on A3 activity.

In chapter 5, using cell lines, we investigated whether coexpressed A3F and A3G have an independent, additive, or synergistic effect on anti-HIV function and what the mechanisms were for this cooperation. We focused our study on A3F and A3G since these enzymes are most commonly expressed together in the human population (*31, 32*). First, using co-IP, size exclusion chromatography (SEC), and rotational anisotropy, we determined that A3F and A3G hetero-oligomerize in the absence of RNA. We became interested in determining how the formation of A3F/G hetero-oligomeric species will affect anti-HIV function. Previous A3F/G coexpression models used either an equimolar (*2*) or non-equimolar (*220*) ratio of A3F and A3G containing plasmids to transfect cells. In such transfection strategies that use multiple plasmids, equimolar co-expression is not possible, which may result in the generation of a heterogeneous population of cells expressing A3F and A3G in various molar ratios on a single cell basis. Thus, transfection using separate plasmids will produce a mosaic population of viruses with diverse combinations and levels of A3 proteins. To ensure equimolar expression of A3F and A3G on a single-cell basis and to circumvent the problem of a mosaic population of viruses, we applied a novel strategy that expressed A3F and A3G from the same plasmid but using separate promoters. This approach resulted in more uniform expression and virion coencapsidation of both A3F and A3G.

Notably, single-cycle infectivity assays and proviral DNA sequencing revealed that the combined expression of A3F and A3G had a synergistic effect on hypermutation and resulted in 4- to 6- fold greater restriction than either A3G or A3F alone. To gain further mechanistic insights into the synergism between A3F and A3G, we systematically conducted biochemical analyses. We demonstrated that A3F promotes A3G deamination activity by forming an A3F/G hetero-oligomer,

which is more efficient at deaminating cytosines. Further, A3F/G triggered the accumulation of shorter reverse transcripts due to interference with reverse transcriptase efficiency, which would keep single-stranded (-)DNA exposed for extended periods of time allowing more deamination events to occur. Although A3F and A3G are known to function along with each other, these data provide the evidence for an A3F/G hetero-oligomeric A3 species with unique properties when compared to each counterpart.

A3F is quite distinct from A3G due to many reasons. First, it prefers 5'TC, whereas 5'CC is A3G's hot motif. The nucleotide specificity box 'YYFW' located on loop 7 of A3F is rich in aromatic amino acids and provides a highly complementary environment for the hydrophobic thymidine nucleotide (*388*). In contrast, A3G has 'YDDQ' nucleotide specificity box on loop 7, which facilitates the glutamine or aspartate acceptors forming hydrogen bonds to the 4-amino group on the cytosine base (note that a cytosine base is 3-times more hydrophilic than thymine) (*388*). Second, A3F is less potent than A3G in HIV-1 restriction. Third, both DNA deaminase activity and Vif binding reside within the CTD domain of A3F, in contrast, NTD and CTD domains of A3G are responsible for Vif binding and DNA deaminase activity, respectively. Fourth, unlike A3G, A3F is a difficult protein to purify in soluble, stable and high quantity. Fifth, in the course of our studies for chapter 3, we found that A3F CTD was 5 fold more active than A3F full-length, unlike A3G CTD that was 1000-fold less active than full-length A3G (*213, 247*). Sixth, unlike A3F, A3G is more sensitive to HIV Vif induced ubiquitin ligase complex-mediated proteasomal degradation (*213, 391*). Seventh, the dimerization interface of A3G ^{126}F-^{127}W (*247*) is adjacent to the residue ^{128}D that HIV Vif uses to interact with human A3G (*305, 320, 392*). In contrast, HIV Vif interacts with human A3F using different residues ^{289}E (*321*) and ^{324}E (*322*). Since the data presented in this thesis indicate that A3F and A3G forms a hetero-oligomeric A3 species it would be interesting to examine if Vif binding to the interface of A3F and/or A3G are concealed when A3F/G hetero-oligomers are formed and such overlapping might influence Vif induced ubiquitin ligase complex-mediated proteasomal degradation of A3F/G hetero-oligomers.

In our study (chapter 3), we tried, but could not identify the oligomerization interface of A3F. In the case of A3G, amino acids ^{126}F -^{127}W form the dimerization interface (*247*). When we mutated equivalent residues in A3F (^{131}Y-^{132}W), we could not disrupt A3F's oligomerization. This suggested that A3F might have different oligomerization interface. Moreover, we do not understand the A3F-A3G hetero-oligomerization interface. This knowledge is necessary to gain

greater insights into A3-A3 hetero-oligomerization and its anti-HIV role. Future studies are needed first to localize the A3F oligomerization interface, and second, to identify the oligomerization interface of the A3F/G hetero-oligomer. This knowledge will help in further characterizing the anti-HIV ability of hetero-oligomeric A3 species biochemically and in cell-based systems. Although there are numerous studies characterizing A3G oligomeric interfaces (*246, 247, 349, 350, 358, 393, 394*), very little is known about the oligomerization interfaces of A3F (*263, 388*). The only A3F oligomerization interface data comes from A3F CTD crystal structures (*263, 314, 388*). However, each structure has different interfaces suggesting these dimer interfaces are due to crystal packing and may not be accurate (*387*). As a result, further biochemical studies on A3F oligomerization are needed.

A remarkable observation by Ping *et al*., 2016, provided *in vivo* evidence for an anti-HIV function of A3F in natural HIV infected cohorts (*391*). The authors reported that A3F I231V variant is resistant to Vif-induced ubiquitin-mediated proteasomal degradation from the HIV-1 recombinant strains AE (recombinant of subtypes A and E) and BC (recombinant of subtypes B and C) (*391*). A3F I231V variant was associated with slow progression of AIDS in European American patients (*391*). These data suggest that the host-pathogen arms race between host factor A3F and HIV Vif is still on-going, similar to what has been found for A3H (*272, 276*). A3F and A3G differ with each other concerning their susceptibility to Vif. For example, Vif can better induce ubiquitin ligase complex-mediated proteasomal degradation of A3G than A3F (*2*) and HIV-1 cannot easily evolve a Vif-independent mechanism to evade A3F, but HIV-1 evolves to overcome A3G restriction pressure (*260*). Moreover, humans have some A3F variants that are resistant to Vif-induced degradation (*391*). Based on these data one can speculate that HIV-1 adapted to counter host factor A3G earlier than A3F. It is not clear if the distinct host-pathogen arms race (Vif susceptibility) as seen for A3F and A3G concerning HIV is the result of A3G being the predominant anti-HIV factor and A3F plays a supporting role or because A3F is a more recently evolved gene than A3G (*207*). The data presented in this thesis provide evidence in support of the first view by showing that A3F alone has less anti-HIV function compared to A3G alone and that A3F supports anti-HIV function of A3G by forming A3F/G hetero-oligomers. However, our data do not refute the possibility that A3F evolved more recently, as suggested by an 8-event model proposed by LaRue *et al*., 2008 for human A3 Z domain history (*207*). A3s have expanded and diverged throughout vertebrate evolution. This 8-event model took into account the present-day

human locus and identified the highly similar sequences within the *A3* locus to find the footprints for recent recombination events. Based on the nearly homologous nucleotide sequences in the flanking regions of the present-day human A3 locus, the authors suggested that A3A, A3B, A3D and A3F originated following recent gene duplication events.

Studies on integrated proviral genome sequences of HIV patients have shown that proviral DNA contains mutations at both 5'CC and 5'TC sequence contexts suggesting that multiple A3s can mutate the same genome (*26, 366-372*). Using shRNA-mediated knockdown of endogenous A3D or A3F or A3G in the CEM T-cell line, Refsland *et al*., 2012 provided the evidence for their anti-HIV role (*19*). Studies as this in primary $CD4^+$ T-cells are lacking, partly due to the difficulty in reliably suppressing A3 expression using shRNA. However, recently a CRISPR/Cas9 ribonucleoprotein technique has been developed for the knockout HIV-relevant molecules in human $CD4^+$ T-cells (*395*). Future studies using this tool could examine the HIV restriction of A3G in the presence and absence of A3F to determine the extent to which the interaction we characterized in cell lines contributes to HIV restriction in primary cells. It would be interesting to use Cas9 ribonucleoprotein platform to target the desired A3 for investigating hetero-oligomerization among other endogenous A3s in various combinations, including A3D/F, A3D/G, A3D/H, A3F/G, A3F/H, or A3G/H. In our preliminary study, we noted that unlike A3F, A3H did not bind with A3G when tested using rotational anisotropy (data not shown). In contrast, A3H could interact with A3F in the nanomolar range, similar to the strength of the A3F/G interaction (data not shown). Future studies with other A3s are needed to address these issues and to investigate whether such hetero-oligomerization confer any advantage *in vivo*.

Further studies will enhance our knowledge on unique hetero-oligomeric A3 species and their role in HIV restriction. Studies such as this may reconcile previous literature that found less activity for A3F. It might be that A3F was not effective if sufficient levels of A3G were not present in the cell. The study presented in this thesis suggests new avenues of inquiry on different A3s hetero-oligomers relevant to HIV restriction.

CHAPTER 7: OVERALL CONCLUSIONS

In summary, my research has furthered understanding of the mechanisms of A3F- and A3G- mediated HIV restriction, i.e., how these enzymes act within the HIV lifecycle, and how these enzymes interact with each other. We have systematically dissected the mechanisms by which A3F and A3G can collectively impact the various stages of HIV infection. We demonstrate that the processive DNA scanning mechanism and preferred deamination motif (A3F, 5'TTC; A3G 5'CCC) are responsible for differential anti-HIV function of A3F and A3G. Further, we provide evidence that A3F and A3G in cells can form a hetero-oligomer in an RNA-independent manner generating an A3 species with unique properties from the individual enzymes. We hypothesize that other A3s may also hetero-oligomerize and cooperate to restrict HIV. All together our data identify hetero-oligomerization as a potential mechanism that may contribute to enhanced activity of A3s and predicts that we may still not fully realize all the antiviral mechanisms of endogenous A3 enzymes.

REFERENCES

1. E. Britan-Rosich, R. Nowarski, M. Kotler, Multifaceted counter-APOBEC3G mechanisms employed by HIV-1 Vif. *J Mol Biol* **410**, 1065-1076.

2. M. T. Liddament, W. L. Brown, A. J. Schumacher, R. S. Harris, APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr Biol* **14**, 1385-1391 (2004).

3. T. Ikeda *et al.*, Creation of chimeric human/rabbit APOBEC1 with HIV-1 restriction and DNA mutation activities. *Sci Rep* **6**, 19035 (2016).

4. Y. Sato *et al.*, Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J Biol Chem* **285**, 7111-7118.

5. L. Chelico, P. Pham, J. Petruska, M. F. Goodman, Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem* **284**, 27761-27765 (2009).

6. A. M. Sheehy, N. C. Gaddis, J. D. Choi, M. H. Malim, Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650 (2002).

7. Y. L. Chiu, W. C. Greene, The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* **26**, 317-353 (2008).

8. V. C. Vieira, M. A. Soares, The role of cytidine deaminases on innate immune responses against human viral infections. *BioMed research international* **2013**, 683095 (2013).

9. M. Monajemi, C. F. Woodworth, J. Benkaroun, M. Grant, M. Larijani, Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment. *Retrovirology* **9**, 35 (2012).

10. M. Imahashi, M. Nakashima, Y. Iwatani, Antiviral Mechanism and Biochemical Basis of the Human APOBEC3 Family. *Frontiers in microbiology* **3**, 250 (2012).

11. G. Dranoff, Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* **4**, 11-22 (2004).

12. B. A. Desimmie *et al.*, Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol* **426**, 1220-1245 (2014).

13. R. S. Harris, J. P. Dudley, APOBECs and virus restriction. *Virology* **479-480**, 131-145 (2015).

14. R. S. Harris, B. D. Anderson, Evolutionary Paradigms from Ancient and Ongoing Conflicts between the Lentiviral Vif Protein and Mammalian APOBEC3 Enzymes. *PLoS pathogens* **12**, e1005958 (2016).

15. 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).

16. 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).

17. M. J. Wichroski, G. B. Robb, T. M. Rana, Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS Pathog* **2**, e41 (2006).

18. E. W. Refsland *et al.*, Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic acids research* **38**, 4274-4284 (2010).

19. E. W. Refsland, J. F. Hultquist, R. S. Harris, Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n. *PLoS pathogens* **8**, e1002800 (2012).

20. 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).

21. K. N. Bishop *et al.*, Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**, 1392-1396 (2004).

22. V. Zennou, P. D. Bieniasz, Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates. *Virology* **349**, 31-40 (2006).

23. Y. H. Zheng *et al.*, Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J Virol* **78**, 6073-6076 (2004).

24. H. L. Wiegand, B. P. Doehle, H. P. Bogerd, B. R. Cullen, A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J* **23**, 2451-2458 (2004).

25. A. M. Land *et al.*, Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. *J Virol* **82**, 8172-8182 (2008).

26. C. Pace *et al.*, Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. *J Virol* **80**, 9259-9269 (2006).

27. J. A. Vazquez-Perez, C. E. Ormsby, R. Hernandez-Juan, K. J. Torres, G. Reyes-Teran, APOBEC3G mRNA expression in exposed seronegative and early stage HIV infected individuals decreases with removal of exposure and with disease progression. *Retrovirology* **6**, 23 (2009).

28. X. Jin *et al.*, APOBEC3G/CEM15 (hA3G) mRNA levels associate inversely with human immunodeficiency virus viremia. *J Virol* **79**, 11513-11516 (2005).

29. M. Biasin *et al.*, Apolipoprotein B mRNA-Editing Enzyme, Catalytic Polypeptide-Like 3G: A Possible Role in the Resistance to HIV of HIV-Exposed Seronegative Individuals. *J Infect Dis* **195**, 960-964 (2007).

30. N. K. Ulenga *et al.*, Relationship between human immunodeficiency type 1 infection and expression of human APOBEC3G and APOBEC3F. *J Infect Dis* **198**, 486-492 (2008).

31. F. A. Koning *et al.*, Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *Journal of virology* **83**, 9474-9485 (2009).

32. 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*, (2010).

33. K. M. Bruner *et al.*, Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043-1049 (2016).

34. J. F. Arias, T. Koyama, M. Kinomoto, K. Tokunaga, Retroelements versus APOBEC3 family members: No great escape from the magnificent seven. *Front Microbiol* **3**, 275 (2012).

35. R. Cordaux, M. A. Batzer, The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009).

36. P. L. Deininger, M. A. Batzer, Mammalian retroelements. *Genome Res* **12**, 1455-1465 (2002).

37. E. M. Ostertag, J. L. Goodier, Y. Zhang, H. H. Kazazian, Jr., SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* **73**, 1444-1451 (2003).

38. D. Ribet, M. Dewannieux, T. Heidmann, An active murine transposon family pair: retrotransposition of "master" MusD copies and ETn trans-mobilization. *Genome Res* **14**, 2261-2267 (2004).

39. M. Dewannieux, A. Dupressoir, F. Harper, G. Pierron, T. Heidmann, Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. *Nat Genet* **36**, 534-539 (2004).

40. Y. N. Lee, P. D. Bieniasz, Reconstitution of an infectious human endogenous retrovirus. *PLoS pathogens* **3**, e10 (2007).

41. S. R. Ross, Mouse mammary tumor virus molecular biology and oncogenesis. *Viruses* **2**, 2000-2012 (2010).

42. A. Rein, Murine leukemia viruses: objects and organisms. *Adv Virol* **2011**, 403419 (2011).

43. O. C. Straub, Maedi-Visna virus infection in sheep. History and present knowledge. *Comp Immunol Microbiol Infect Dis* **27**, 1-5 (2004).

44. V. M. Vogt, *Retroviral virions and genomes. In: Retroviruses, J.M. Coffin, S.H. Hughes and H.E. Varmus (Eds)*. (Cold Spring Harbor Laboratory

Press, 1997), pp. 27-70.

45. F. H. Lin, H. Thormar, On visna virus: purification and nucleic acid content. *Virology* **42**, 1140-1143 (1970).

46. A. T. Haase, The slow infection caused by visna virus. *Curr Top Microbiol Immunol* **72**, 101-156 (1975).

47. P. M. Sharp, B. H. Hahn, Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* **1**, a006841 (2011).

48. C. Centers for Disease, Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR Morb Mortal Wkly Rep* **30**, 305-308 (1981).

49. W. C. Greene, A history of AIDS: looking back to see ahead. *Eur J Immunol* **37 Suppl 1**, S94-102 (2007).

50. L. Chakrabarti *et al.*, Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature* **328**, 543-547 (1987).

51. M. Guyader *et al.*, Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* **326**, 662-669 (1987).

52. F. Clavel *et al.*, Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343-346 (1986).

53. T. Huet, R. Cheynier, A. Meyerhans, G. Roelants, S. Wain-Hobson, Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**, 356-359 (1990).

54. V. M. Hirsch, R. A. Olmsted, M. Murphey-Corb, R. H. Purcell, P. R. Johnson, An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339**, 389-392 (1989).

55. B. H. Hahn, G. M. Shaw, K. M. De Cock, P. M. Sharp, AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607-614 (2000).

56. J. C. Plantier *et al.*, A new human immunodeficiency virus derived from gorillas. *Nat Med* **15**, 871-872 (2009).

57. F. Simon *et al.*, Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* **4**, 1032-1037 (1998).

58. P. M. Sharp, B. H. Hahn, AIDS: prehistory of HIV-1. *Nature* **455**, 605-606 (2008).

59. M. Peeters *et al.*, Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **11**, 493-498 (1997).

60. A. Vallari *et al.*, Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS Res Hum Retroviruses* **26**, 109-115 (2010).

61. A. Vallari *et al.*, Confirmation of putative HIV-1 group P in Cameroon. *Journal of virology* **85**, 1403-1407 (2011).

62. D. L. Robertson *et al.*, HIV-1 nomenclature proposal. *Science* **288**, 55-56 (2000).

63. B. S. Taylor, S. M. Hammer, The challenge of HIV-1 subtype diversity. *N Engl J Med* **359**, 1965-1966 (2008).

64. M. T. Gilbert *et al.*, The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* **104**, 18566-18570 (2007).

65. J. Hemelaar *et al.*, Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**, 679-689 (2011).

66. O. T. Campbell-Yesufu, R. T. Gandhi, Update on human immunodeficiency virus (HIV)-2 infection. *Clin Infect Dis* **52**, 780-787 (2011).

67. T. I. de Silva, M. Cotten, S. L. Rowland-Jones, HIV-2: the forgotten AIDS virus. *Trends Microbiol* **16**, 588-595 (2008).

68. R. Marlink *et al.*, Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587-1590 (1994).

69. S. J. Popper *et al.*, Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. *J Infect Dis* **180**, 1116-1121 (1999).

70. P. J. Kanki *et al.*, Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**, 943-946 (1994).

71. D. O'Donovan *et al.*, Maternal plasma viral RNA levels determine marked differences in mother-to-child transmission rates of HIV-1 and HIV-2 in The Gambia. MRC/Gambia Government/University College London Medical School working group on mother-child transmission of HIV. *AIDS* **14**, 441-448 (2000).

72. S. Hallenberger *et al.*, Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360**, 358-361 (1992).

73. L. Li, H. S. Li, C. D. Pauza, M. Bukrinsky, R. Y. Zhao, Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions. *Cell Res* **15**, 923-934 (2005).

74. J. M. Watts *et al.*, Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716 (2009).

75. R. Amorim, S. M. Costa, N. P. Cavaleiro, E. E. da Silva, L. J. da Costa, HIV-1 transcripts use IRES-initiation under conditions where Cap-dependent translation is restricted by poliovirus 2A protease. *PloS one* **9**, e88619 (2014).

76. G. Li, E. De Clercq, HIV Genome-Wide Protein Associations: a Review of 30 Years of Research. *Microbiol Mol Biol Rev* **80**, 679-731 (2016).

77. R. Wyatt, J. Sodroski, The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**, 1884-1888 (1998).

78. E. Helseth, U. Olshevsky, C. Furman, J. Sodroski, Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *Journal of virology* **65**, 2119-2123 (1991).

79. J. G. Levin, J. Guo, I. Rouzina, K. Musier-Forsyth, Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **80**, 217-286 (2005).

80. X. J. Yao, G. Kobinger, S. Dandache, N. Rougeau, E. Cohen, HIV-1 Vpr-chloramphenicol acetyltransferase fusion proteins: sequence requirement for virion incorporation and analysis of antiviral effect. *Gene Ther* **6**, 1590-1599 (1999).

81. E. Chertova *et al.*, Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *Journal of virology* **80**, 9039-9052 (2006).

82. L. Kleiman, R. Halwani, H. Javanbakht, The selective packaging and annealing of primer tRNALys3 in HIV-1. *Curr HIV Res* **2**, 163-175 (2004).

83. D. E. Ott *et al.*, Cytoskeletal proteins inside human immunodeficiency virus type 1 virions. *Journal of virology* **70**, 7734-7743 (1996).

84. S. J. Rulli, Jr. *et al.*, Selective and nonselective packaging of cellular RNAs in retrovirus particles. *Journal of virology* **81**, 6623-6631 (2007).

85. X. Wang *et al.*, Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. *The Journal of biological chemistry* **285**, 14346-14355 (2010).

86. A. Telesnitsky, S. L. Wolin, The Host RNAs in Retroviral Particles. *Viruses* **8**, (2016).

87. E. O. Freed, HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* **251**, 1-15 (1998).

88. N. M. Bell, A. M. Lever, HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol* **21**, 136-144 (2013).

89. J. Bhattacharya, A. Repik, P. R. Clapham, Gag regulates association of human immunodeficiency virus type 1 envelope with detergent-resistant membranes. *Journal of virology* **80**, 5292-5300 (2006).

90. R. H. Ghanam, A. B. Samal, T. F. Fernandez, J. S. Saad, Role of the HIV-1 Matrix Protein in Gag Intracellular Trafficking and Targeting to the Plasma Membrane for Virus Assembly. *Front Microbiol* **3**, 55 (2012).

91. M. Bukrinsky, A hard way to the nucleus. *Mol Med* **10**, 1-5 (2004).

92. R. Peytavi *et al.*, HEED, the product of the human homolog of the murine eed gene, binds to the matrix protein of HIV-1. *The Journal of biological chemistry* **274**, 1635-1645 (1999).

93. U. von Schwedler, R. S. Kornbluth, D. Trono, The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. *Proc Natl Acad Sci U S A* **91**, 6992-6996 (1994).

94. S. B. Kutluay *et al.*, Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* **159**, 1096-1109 (2014).

95. J. L. Darlix, J. L. Garrido, N. Morellet, Y. Mely, H. de Rocquigny, Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. *Adv Pharmacol* **55**, 299-346 (2007).

96. D. Muriaux, J. L. Darlix, Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol* **7**, 744-753 (2010).

97. J. S. Buckman, W. J. Bosche, R. J. Gorelick, Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *Journal of virology* **77**, 1469-1480 (2003).

98. J. G. Levin, M. Mitra, A. Mascarenhas, K. Musier-Forsyth, Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* **7**, 754-774 (2010).

99. Y. Usami *et al.*, The ESCRT pathway and HIV-1 budding. *Biochem Soc Trans* **37**, 181-184 (2009).

100. E. Kondo, F. Mammano, E. A. Cohen, H. G. Gottlinger, The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. *Journal of virology* **69**, 2759-2764 (1995).

101. X. J. Yao *et al.*, Mutagenic analysis of human immunodeficiency virus type 1 Vpr: role of a predicted N-terminal alpha-helical structure in Vpr nuclear localization and virion incorporation. *Journal of virology* **69**, 7032-7044 (1995).

102. P. M. Colman, M. C. Lawrence, The structural biology of type I viral membrane fusion. *Nat Rev Mol Cell Biol* **4**, 309-319 (2003).

103. M. Hill, G. Tachedjian, J. Mak, The packaging and maturation of the HIV-1 Pol proteins. *Curr HIV Res* **3**, 73-85 (2005).

104. A. Jacobo-Molina, E. Arnold, HIV reverse transcriptase structure-function relationships. *Biochemistry* **30**, 6351-6356 (1991).

105. S. C. Pettit, J. N. Lindquist, A. H. Kaplan, R. Swanstrom, Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* **2**, 66 (2005).

106. A. Brik, C. H. Wong, HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem* **1**, 5-14 (2003).

107. R. L. Shoeman, C. Huttermann, R. Hartig, P. Traub, Amino-terminal polypeptides of vimentin are responsible for the changes in nuclear architecture associated with human immunodeficiency virus type 1 protease activity in tissue culture cells. *Mol Biol Cell* **12**, 143-154 (2001).

108. P. R. Strack *et al.*, Apoptosis mediated by HIV protease is preceded by cleavage of Bcl-2. *Proc Natl Acad Sci U S A* **93**, 9571-9576 (1996).

109. S. P. Goff, Retroviral reverse transcriptase: synthesis, structure, and function. *J Acquir Immune Defic Syndr* **3**, 817-831 (1990).

110. S. G. Sarafianos *et al.*, Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* **385**, 693-713 (2009).

111. V. R. Prasad, S. P. Goff, Structure-function studies of HIV reverse transcriptase. *Ann N Y Acad Sci* **616**, 11-21 (1990).

112. A. Herschhorn, A. Hizi, Retroviral reverse transcriptases. *Cell Mol Life Sci* **67**, 2717-2747 (2010).

113. T. K. Chiu, D. R. Davies, Structure and function of HIV-1 integrase. *Curr Top Med Chem* **4**, 965-977 (2004).

114. T. Masuda, Non-Enzymatic Functions of Retroviral Integrase: The Next Target for Novel Anti-HIV Drug Development. *Front Microbiol* **2**, 210 (2011).

115. T. Takahata *et al.*, Critical Contribution of Tyr15 in the HIV-1 Integrase (IN) in Facilitating IN Assembly and Nonenzymatic Function through the IN Precursor Form with Reverse Transcriptase. *Journal of virology* **91**, (2017).

116. A. Engelman, G. Englund, J. M. Orenstein, M. A. Martin, R. Craigie, Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *Journal of virology* **69**, 2729-2736 (1995).

117. A. W. Cochrane, A. Perkins, C. A. Rosen, Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function. *Journal of virology* **64**, 881-885 (1990).

118. M. Fornerod, M. Ohno, M. Yoshida, I. W. Mattaj, CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051-1060 (1997).

119. H. C. Groom, E. C. Anderson, A. M. Lever, Rev: beyond nuclear export. *J Gen Virol* **90**, 1303-1318 (2009).

120. B. Grewe, K. Uberla, The human immunodeficiency virus type 1 Rev protein: menage a trois during the early phase of the lentiviral replication cycle. *J Gen Virol* **91**, 1893-1897 (2010).

121. S. Roy, U. Delling, C. H. Chen, C. A. Rosen, N. Sonenberg, A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev* **4**, 1365-1373 (1990).

122. R. C. Gallo, Tat as one key to HIV-induced immune pathogenesis and Tat (correction of Pat) toxoid as an important component of a vaccine. *Proc Natl Acad Sci U S A* **96**, 8324-8326 (1999).

123. J. M. Sabatier *et al.*, Evidence for neurotoxic activity of tat from human immunodeficiency virus type 1. *Journal of virology* **65**, 961-967 (1991).

124. S. Henriet, G. Mercenne, S. Bernacchi, J. C. Paillart, R. Marquet, Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (Vif) and the human APOBEC-3G and APOBEC-3F restriction factors. *Microbiol Mol Biol Rev* **73**, 211-232 (2009).

125. R. D. Sloan, D. A. Donahue, B. D. Kuhl, T. Bar-Magen, M. A. Wainberg, Expression of Nef from unintegrated HIV-1 DNA downregulates cell surface CXCR4 and CCR5 on T-lymphocytes. *Retrovirology* **7**, 44 (2010).

126. N. Michel, I. Allespach, S. Venzke, O. T. Fackler, O. T. Keppler, The Nef protein of human immunodeficiency virus establishes superinfection immunity by a dual strategy to downregulate cell-surface CCR5 and CD4. *Curr Biol* **15**, 714-723 (2005).

127. S. Venzke, N. Michel, I. Allespach, O. T. Fackler, O. T. Keppler, Expression of Nef downregulates CXCR4, the major coreceptor of human immunodeficiency virus, from the surfaces of target cells and thereby enhances resistance to superinfection. *Journal of virology* **80**, 11141-11152 (2006).

128. O. O. Yang *et al.*, Nef-mediated resistance of human immunodeficiency virus type 1 to antiviral cytotoxic T lymphocytes. *Journal of virology* **76**, 1626-1631 (2002).

129. A. Rosa *et al.*, HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526**, 212-217 (2015).

130. Y. Usami, Y. Wu, H. G. Gottlinger, SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526**, 218-223 (2015).

131. W. B. Dyer *et al.*, Lymphoproliferative immune function in the Sydney Blood Bank Cohort, infected with natural nef/long terminal repeat mutants, and in other long-term survivors of transfusion-acquired HIV-1 infection. *AIDS* **11**, 1565-1574 (1997).

132. M. Mikhail, B. Wang, N. K. Saksena, Mechanisms involved in non-progressive HIV disease. *AIDS Rev* **5**, 230-244 (2003).

133. K. Levesque, Y. S. Zhao, E. A. Cohen, Vpu exerts a positive effect on HIV-1 infectivity by down-modulating CD4 receptor molecules at the surface of HIV-1-producing cells. *The Journal of biological chemistry* **278**, 28346-28353 (2003).

134. F. Margottin *et al.*, A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol Cell* **1**, 565-574 (1998).

135. M. Tanaka *et al.*, Downregulation of CD4 is required for maintenance of viral infectivity of HIV-1. *Virology* **311**, 316-325 (2003).

136. S. J. Neil, S. W. Eastman, N. Jouvenet, P. D. Bieniasz, HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. *PLoS Pathog* **2**, e39 (2006).

137. J. L. Douglas *et al.*, Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/Tetherin via a {beta}TrCP-dependent mechanism. *Journal of virology* **83**, 7931-7947 (2009).

138. P. Di Marzio, S. Choe, M. Ebright, R. Knoblauch, N. R. Landau, Mutational analysis of cell cycle arrest, nuclear localization and virion packaging of human immunodeficiency virus type 1 Vpr. *Journal of virology* **69**, 7909-7916 (1995).

139. N. Morellet, S. Bouaziz, P. Petitjean, B. P. Roques, NMR structure of the HIV-1 regulatory protein VPR. *J Mol Biol* **327**, 215-227 (2003).

140. K. D. Jayappa, Z. Ao, X. Yao, The HIV-1 passage from cytoplasm to nucleus: the process involving a complex exchange between the components of HIV-1 and cellular machinery to access nucleus and successful integration. *Int J Biochem Mol Biol* **3**, 70-85 (2012).

141. M. Kamata, Y. Nitahara-Kasahara, Y. Miyamoto, Y. Yoneda, Y. Aida, Importin-alpha promotes passage through the nuclear pore complex of human immunodeficiency virus type 1 Vpr. *J Virol* **79**, 3557-3564 (2005).

142. Y. Nitahara-Kasahara *et al.*, Novel nuclear import of Vpr promoted by importin alpha is crucial for human immunodeficiency virus type 1 replication in macrophages. *J Virol* **81**, 5284-5293 (2007).

143. Y. Jenkins, M. McEntee, K. Weis, W. C. Greene, Characterization of HIV-1 vpr nuclear import: analysis of signals and pathways. *J Cell Biol* **143**, 875-885 (1998).

144. S. Mahalingam, V. Ayyavoo, M. Patel, T. Kieber-Emmons, D. B. Weiner, Nuclear import, virion incorporation, and cell cycle arrest/differentiation are mediated by distinct functional domains of human immunodeficiency virus type 1 Vpr. *J Virol* **71**, 6339-6347 (1997).

145. K. Muthumani *et al.*, HIV-1 Vpr induces apoptosis through caspase 9 in T cells and peripheral blood mononuclear cells. *J Biol Chem* **277**, 37820-37831 (2002).

146. B. Trinite, C. N. Chan, C. S. Lee, D. N. Levy, HIV-1 Vpr- and Reverse Transcription-Induced Apoptosis in Resting Peripheral Blood CD4 T Cells and Protection by Common Gamma-Chain Cytokines. *J Virol* **90**, 904-916 (2015).

147. R. Vanitharani *et al.*, HIV-1 Vpr transactivates LTR-directed expression through sequences present within -278 to -176 and increases virus replication in vitro. *Virology* **289**, 334-342 (2001).

148. L. K. Felzien *et al.*, HIV transcriptional activation by the accessory protein, VPR, is mediated by the p300 co-activator. *Proc Natl Acad Sci U S A* **95**, 5281-5286 (1998).

149. J. P. Moore, R. W. Doms, The entry of entry inhibitors: a fusion of science and medicine. *Proc Natl Acad Sci U S A* **100**, 10598-10602 (2003).

150. C. B. Wilen, J. C. Tilton, R. W. Doms, HIV: cell binding and entry. *Cold Spring Harb Perspect Med* **2**, (2012).

151. D. A. Jacques *et al.*, HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* **536**, 349-353 (2016).

152. C. Aiken, Viral and cellular factors that regulate HIV-1 uncoating. *Curr Opin HIV AIDS* **1**, 194-199 (2006).

153. N. Arhel, Revisiting HIV-1 uncoating. *Retrovirology* **7**, 96 (2010).

154. R. S. Harris, J. F. Hultquist, D. T. Evans, The restriction factors of human immunodeficiency virus. *J Biol Chem* **287**, 40875-40883 (2012).

155. Y. Ma *et al.*, Real-Time Imaging of Single HIV-1 Disassembly with Multicolor Viral Particles. *ACS Nano* **10**, 6273-6282 (2016).

156. S. F. Le Grice, Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. *The Journal of biological chemistry* **287**, 40850-40857 (2012).

157. W. S. Hu, S. H. Hughes, HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**, (2012).

158. J. A. Thomas, R. J. Gorelick, Nucleocapsid protein function in early infection processes. *Virus Res* **134**, 39-63 (2008).

159. Y. Suzuki, R. Craigie, The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol* **5**, 187-196 (2007).

160. M. J. Roth, P. L. Schwartzberg, S. P. Goff, Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* **58**, 47-54 (1989).

161. P. O. Brown, B. Bowerman, H. E. Varmus, J. M. Bishop, Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc Natl Acad Sci U S A* **86**, 2525-2529 (1989).

162. L. Krishnan, A. Engelman, Retroviral integrase proteins and HIV-1 DNA integration. *J Biol Chem* **287**, 40858-40866 (2012).

163. R. Craigie, F. D. Bushman, HIV DNA integration. *Cold Spring Harb Perspect Med* **2**, a006890 (2012).

164. B. Van Maele, Z. Debyser, HIV-1 integration: an interplay between HIV-1 integrase, cellular and viral proteins. *AIDS Rev* **7**, 26-43 (2005).

165. O. Jegede *et al.*, HIV type 1 integrase inhibitors: from basic research to clinical implications. *AIDS Rev* **10**, 172-189 (2008).

166. M. Ott, M. Geyer, Q. Zhou, The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe* **10**, 426-435 (2011).

167. J. Karn, C. M. Stoltzfus, Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med* **2**, a006916 (2012).

168. B. G. Turner, M. F. Summers, Structural biology of HIV. *J Mol Biol* **285**, 1-32 (1999).

169. B. K. Ganser-Pornillos, M. Yeager, W. I. Sundquist, The structural biology of HIV assembly. *Curr Opin Struct Biol* **18**, 203-217 (2008).

170. N. Jouvenet *et al.*, Plasma membrane is the site of productive HIV-1 particle assembly. *PLoS Biol* **4**, e435 (2006).

171. W. Muranyi, S. Malkusch, B. Muller, M. Heilemann, H. G. Krausslich, Super-resolution microscopy reveals specific recruitment of HIV-1 envelope proteins to viral assembly sites dependent on the envelope C-terminal tail. *PLoS pathogens* **9**, e1003198 (2013).

172. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687-695 (2012).

173. D. Blanco-Melo, S. Venkatesh, P. D. Bieniasz, Intrinsic cellular defenses against human immunodeficiency viruses. *Immunity* **37**, 399-411 (2012).

174. W. E. Johnson, S. L. Sawyer, Molecular evolution of the antiretroviral TRIM5 gene. *Immunogenetics* **61**, 163-176 (2009).

175. D. Wolf, S. P. Goff, Host restriction factors blocking retroviral replication. *Annu Rev Genet* **42**, 143-163 (2008).

176. H. Lahouassa *et al.*, SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* **13**, 223-228 (2012).

177. C. J. Spragg, M. Emerman, Antagonism of SAMHD1 is actively maintained in natural infections of simian immunodeficiency virus. *Proc Natl Acad Sci U S A* **110**, 21136-21141 (2013).

178. J. S. Albin, R. S. Harris, Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert Rev Mol Med* **12**, e4 (2010).

179. B. D. Anderson, R. S. Harris, Transcriptional regulation of APOBEC3 antiviral immunity through the CBF-beta/RUNX axis. *Sci Adv* **1**, e1500296 (2015).

180. G. Berger *et al.*, APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. *PLoS pathogens* **7**, e1002221 (2011).

181. B. K. Thielen *et al.*, Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. *The Journal of biological chemistry* **285**, 27753-27766 (2010).

182. M. H. Malim, M. Emerman, HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388-398 (2008).

183. M. Stremlau *et al.*, The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).

184. M. Stremlau *et al.*, Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A* **103**, 5514-5519 (2006).

185. T. Pertel *et al.*, TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361-365 (2011).

186. M. H. Malim, P. D. Bieniasz, HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb Perspect Med* **2**, a006940 (2012).

187. C. E. Jones, A. McKnight, Retroviral restriction: nature's own solution. *Curr Opin Infect Dis* **29**, 609-614 (2016).

188. Y. H. Zheng, K. T. Jeang, K. Tokunaga, Host restriction factors in retroviral infection: promises in virus-host interaction. *Retrovirology* **9**, 112 (2012).

189. T. E. White *et al.*, Contribution of SAM and HD domains to retroviral restriction mediated by human SAMHD1. *Virology* **436**, 81-90 (2013).

190. D. C. Goldstone *et al.*, HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480**, 379-382 (2011).

191. N. Laguette *et al.*, SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654-657 (2011).

192. K. Hrecka *et al.*, Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658-661 (2011).

193. S. J. Neil, T. Zang, P. D. Bieniasz, Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425-430 (2008).

194. N. Van Damme *et al.*, The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe* **3**, 245-252 (2008).

195. K. Fitzpatrick *et al.*, Direct restriction of virus release and incorporation of the interferon-induced protein BST-2 into HIV-1 particles. *PLoS pathogens* **6**, e1000701 (2010).

196. F. Kirchhoff, Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host Microbe* **8**, 55-67 (2010).

197. F. Zhang *et al.*, SIV Nef proteins recruit the AP-2 complex to antagonize Tetherin and facilitate virion release. *PLoS pathogens* **7**, e1002039 (2011).

198. A. Le Tortorec, S. J. Neil, Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. *Journal of virology* **83**, 11966-11978 (2009).

199. D. H. Gabuzda *et al.*, Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. *Journal of virology* **66**, 6489-6495 (1992).

200. U. von Schwedler, J. Song, C. Aiken, D. Trono, Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *Journal of virology* **67**, 4945-4955 (1993).

201. A. Jarmuz *et al.*, An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285-296 (2002).

202. 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).

203. M. B. Adolph *et al.*, Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. *Nucleic acids research*, (2017).

204. H. Xu *et al.*, Stoichiometry of the antiviral protein APOBEC3G in HIV-1 virions. *Virology* **360**, 247-256 (2007).

205. R. Nowarski, E. Britan-Rosich, T. Shiloach, M. Kotler, Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. *Nat Struct Mol Biol* **15**, 1059-1066 (2008).

206. F. A. Koning, C. Goujon, H. Bauby, M. H. Malim, Target cell-mediated editing of HIV-1 cDNA by APOBEC3 proteins in human macrophages. *Journal of virology* **85**, 13448-13452 (2011).

207. R. S. LaRue *et al.*, The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol Biol* **9**, 104 (2008).

208. M. S. Neuberger, R. S. Harris, J. Di Noia, S. K. Petersen-Mahrt, Immunity through DNA deamination. *Trends Biochem Sci* **28**, 305-312 (2003).

209. R. S. Harris, M. T. Liddament, Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* **4**, 868-877 (2004).

210. R. S. LaRue *et al.*, Guidelines for naming nonprimate APOBEC3 genes and proteins. *Journal of virology* **83**, 494-497 (2009).

211. G. Hache, M. T. Liddament, R. S. Harris, The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *The Journal of biological chemistry* **280**, 10920-10924 (2005).

212. F. Navarro *et al.*, Complementary function of the two catalytic domains of APOBEC3G. *Virology* **333**, 374-386 (2005).

213. Y. Feng, T. T. Baig, R. P. Love, L. Chelico, Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. *Front Microbiol* **5**, 450 (2014).

214. R. M. Kohli *et al.*, A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J Biol Chem* **284**, 22898-22904 (2009).

215. B. A. Desimmie *et al.*, APOBEC3 proteins can copackage and comutate HIV-1 genomes. *Nucleic Acids Res* **44**, 7848-7865 (2016).

216. 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).

217. R. S. Harris *et al.*, DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809 (2003).

218. Y. Dang, X. Wang, W. J. Esselman, Y. H. Zheng, Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *Journal of virology* **80**, 10522-10533 (2006).

219. A. Harari, M. Ooms, L. C. Mulder, V. Simon, Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. *J Virol* **83**, 295-303 (2009).

220. B. A. Desimmie *et al.*, APOBEC3 proteins can copackage and comutate HIV-1 genomes. *Nucleic acids research*, (2016).

221. L. Apolonia *et al.*, Promiscuous RNA binding ensures effective encapsidation of APOBEC3 proteins by HIV-1. *PLoS Pathog* **11**, e1004609 (2015).

222. N. Jouvenet, S. M. Simon, P. D. Bieniasz, Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc Natl Acad Sci U S A* **106**, 19114-19119 (2009).

223. 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).

224. Y. L. Chiu *et al.*, High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 15588-15593 (2006).

225. S. Gallois-Montbrun *et al.*, Comparison of cellular ribonucleoprotein complexes associated with the APOBEC3F and APOBEC3G antiviral proteins. *Journal of virology* **82**, 5636-5642 (2008).

226. A. Ara, R. P. Love, L. Chelico, Different mutagenic potential of HIV-1 restriction factors APOBEC3G and APOBEC3F is determined by distinct single-stranded DNA scanning mechanisms. *PLoS pathogens* **10**, e1004024 (2014).

227. 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).

228. R. A. Russell, M. D. Moore, W. S. Hu, V. K. Pathak, APOBEC3G induces a hypermutation gradient: purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA. *Retrovirology* **6**, 16 (2009).

229. K. N. Bishop, M. Verma, E. Y. Kim, S. M. Wolinsky, M. H. Malim, APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS Pathog* **4**, e1000231 (2008).

230. S. M. Kaiser, M. Emerman, Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G. *Journal of virology* **80**, 875-882 (2006).

231. H. Aydin, M. W. Taylor, J. E. Lee, Structure-guided analysis of the human APOBEC3-HIV restrictome. *Structure* **22**, 668-684 (2014).

232. B. Yang, K. Chen, C. Zhang, S. Huang, H. Zhang, Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. *The Journal of biological chemistry* **282**, 11667-11675 (2007).

233. D. Lecossier, F. Bouchonnet, F. Clavel, A. J. Hance, Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **300**, 1112 (2003).

234. R. Nowarski *et al.*, APOBEC3G inhibits HIV-1 RNA elongation by inactivating the viral trans-activation response element. *J Mol Biol* **426**, 2840-2853 (2014).

235. P. Jern, R. A. Russell, V. K. Pathak, J. M. Coffin, Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. *PLoS pathogens* **5**, e1000367 (2009).

236. H. A. Sadler, M. D. Stenglein, R. S. Harris, L. M. Mansky, APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *Journal of virology* **84**, 7396-7404 (2010).

237. U. Neogi *et al.*, Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals. *J Int AIDS Soc* **16**, 18472 (2013).

238. M. McCallum *et al.*, Basis for early and preferential selection of the E138K mutation in HIV-1 reverse transcriptase. *Antimicrob Agents Chemother* **57**, 4681-4688 (2013).

239. E. Y. Kim *et al.*, Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J Virol* **84**, 10402-10405 (2010).

240. L. C. Mulder, A. Harari, V. Simon, Cytidine deamination induced HIV-1 drug resistance. *Proc Natl Acad Sci U S A* **105**, 5501-5506 (2008).

241. A. E. Armitage *et al.*, APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. *PLoS Genet* **8**, e1002550 (2012).

242. N. Wood *et al.*, HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog* **5**, e1000414 (2009).

243. D. Ebrahimi, F. Anwar, M. P. Davenport, APOBEC3 has not left an evolutionary footprint on the HIV-1 genome. *J Virol* **85**, 9139-9146 (2011).

244. G. Hache, L. M. Mansky, R. S. Harris, Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. *AIDS Rev* **8**, 148-157 (2006).

245. K. A. Delviks-Frankenberry *et al.*, Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation. *PLoS pathogens* **12**, e1005646 (2016).

246. H. Huthoff, F. Autore, S. Gallois-Montbrun, F. Fraternali, M. H. Malim, RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog* **5**, e1000330 (2009).

247. 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).

248. O. G. Berg, R. B. Winter, P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929-6948 (1981).

249. P. H. von Hippel, O. G. Berg, Facilitated target location in biological systems. *The Journal of biological chemistry* **264**, 675-678 (1989).

250. Y. Feng, L. Chelico, Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain. *J Biol Chem* **286**, 11415-11426 (2011).

251. M. B. Adolph *et al.*, Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. *Nucleic acids research* **45**, 3378-3394 (2017).

252. L. Chelico, E. J. Sacho, D. A. Erie, M. F. Goodman, A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *The Journal of biological chemistry* **283**, 13780-13791 (2008).

253. G. Senavirathne *et al.*, Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution. *The Journal of biological chemistry* **287**, 15826-15835 (2012).

254. E. P. Browne, C. Allers, N. R. Landau, Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. *Virology* **387**, 313-321 (2009).

255. 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. *The Journal of biological chemistry* **282**, 2587-2595 (2007).

256. C. A. Virgen, T. Hatziioannou, Antiretroviral activity and Vif sensitivity of rhesus macaque APOBEC3 proteins. *Journal of virology* **81**, 13932-13937 (2007).

257. C. Song, L. Sutton, M. E. Johnson, R. T. D'Aquila, J. P. Donahue, Signals in APOBEC3F N-terminal and C-terminal deaminase domains each contribute to encapsidation in HIV-1 virions and are both required for HIV-1 restriction. *The Journal of biological chemistry* **287**, 16965-16974 (2012).

258. E. Miyagi *et al.*, Stably expressed APOBEC3F has negligible antiviral activity. *J Virol* **84**, 11067-11075 (2010).

259. J. S. Albin, G. Hache, J. F. Hultquist, W. L. Brown, R. S. Harris, Long-term restriction by APOBEC3F selects human immunodeficiency virus type 1 variants with restored Vif function. *Journal of virology* **84**, 10209-10219 (2010).

260. G. Hache, K. Shindo, J. S. Albin, R. S. Harris, Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. *Curr Biol* **18**, 819-824 (2008).

261. R. A. Russell, J. Smith, R. Barr, D. Bhattacharyya, V. K. Pathak, Distinct domains within APOBEC3G and APOBEC3F interact with separate regions of human immunodeficiency virus type 1 Vif. *Journal of virology* **83**, 1992-2003 (2009).

262. T. Kouno *et al.*, Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat Struct Mol Biol* **22**, 485-491 (2015).

263. M. F. Bohn *et al.*, Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure* **21**, 1042-1050 (2013).

264. A. M. Land *et al.*, APOBEC3F determinants of HIV-1 Vif sensitivity. *Journal of virology* **88**, 12923-12927 (2014).

265. Y. Dang *et al.*, Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. *J Virol* **85**, 5691-5695 (2011).

266. 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. *Journal of virology* **87**, 444-453 (2013).

267. J. Li *et al.*, APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *J Mol Biol* **426**, 1296-1307 (2014).

268. M. OhAinle, J. A. Kerns, H. S. Malik, M. Emerman, Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *Journal of virology* **80**, 3853-3862 (2006).

269. M. Ooms, S. Majdak, C. W. Seibert, A. Harari, V. Simon, The localization of APOBEC3H variants in HIV-1 virions determines their antiviral activity. *Journal of virology* **84**, 7961-7969 (2010).

270. X. Wang *et al.*, Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. *Journal of virology* **85**, 3142-3152 (2011).

271. Y. Dang *et al.*, Human cytidine deaminase APOBEC3H restricts HIV-1 replication. *The Journal of biological chemistry* **283**, 11606-11614 (2008).

272. 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).

273. M. M. Li, M. Emerman, Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. *Journal of virology* **85**, 8197-8207 (2011).

274. A. Zhen, J. Du, X. Zhou, Y. Xiong, X. F. Yu, Reduced APOBEC3H variant anti-viral activities are associated with altered RNA binding activities. *PloS one* **7**, e38771 (2012).

275. T. T. Baig, Y. Feng, L. Chelico, Determinants of efficient degradation of APOBEC3 restriction factors by HIV-1 Vif. *J Virol* **88**, 14380-14395 (2014).

276. M. Ooms *et al.*, HIV-1 Vif adaptation to human APOBEC3H haplotypes. *Cell Host Microbe* **14**, 411-421 (2013).

277. M. OhAinle, J. A. Kerns, M. M. Li, H. S. Malik, M. Emerman, Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **4**, 249-259 (2008).

278. D. Sakurai *et al.*, APOBEC3H polymorphisms associated with the susceptibility to HIV-1 infection and AIDS progression in Japanese. *Immunogenetics* **67**, 253-257 (2015).

279. K. Zhao *et al.*, Evolutionarily conserved pressure for the existence of distinct G2/M cell cycle arrest and A3H inactivation functions in HIV-1 Vif. *Cell Cycle* **14**, 838-847 (2015).

280. Y. Feng *et al.*, Natural Polymorphisms and Oligomerization of Human APOBEC3H Contribute to Single-stranded DNA Scanning Ability. *The Journal of biological chemistry* **290**, 27188-27203 (2015).

281. E. N. Newman *et al.*, Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* **15**, 166-170 (2005).

282. X. Wang *et al.*, The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. *Journal of virology* **86**, 3777-3786 (2012).

283. M. B. Adolph, J. Webb, L. Chelico, Retroviral restriction factor APOBEC3G delays the initiation of DNA synthesis by HIV-1 reverse transcriptase. *PloS one* **8**, e64196 (2013).

284. 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).

285. F. Guo, S. Cen, M. Niu, J. Saadatmand, L. Kleiman, Inhibition of tRNA(3)(Lys)-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. *Journal of virology* **80**, 11710-11722 (2006).

286. F. Guo *et al.*, The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA. *J Virol* **81**, 11322-11331 (2007).

287. K. Gillick *et al.*, Suppression of HIV-1 infection by APOBEC3 proteins in primary human CD4(+) T cells is associated with inhibition of processive reverse transcription as well as excessive cytidine deamination. *Journal of virology* **87**, 1508-1517 (2013).

288. Y. Iwatani *et al.*, Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res* **35**, 7096-7108 (2007).

289. 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).

290. J. L. Anderson, T. J. Hope, APOBEC3G restricts early HIV-1 replication in the cytoplasm of target cells. *Virology* **375**, 1-12 (2008).

291. J. L. Mbisa *et al.*, Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. *J Virol* **81**, 7099-7110 (2007).

292. E. Miyagi *et al.*, Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1. *J Virol* **81**, 13346-13353 (2007).

293. A. J. Schumacher, G. Hache, D. A. Macduff, W. L. Brown, R. S. Harris, The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J Virol* **82**, 2652-2660 (2008).

294. K. R. Chaurasiya *et al.*, Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid-binding protein. *Nat Chem* **6**, 28-33 (2014).

295. A. Ara *et al.*, Mechanism of Enhanced HIV Restriction by Virion Coencapsidated Cytidine Deaminases APOBEC3F and APOBEC3G. *Journal of virology* **91**, (2017).

296. K. N. Bishop, R. K. Holmes, M. H. Malim, Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *Journal of virology* **80**, 8450-8458 (2006).

297. 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. *Journal of virology* **88**, 5881-5887 (2014).

298. J. S. Albin, W. L. Brown, R. S. Harris, Catalytic activity of APOBEC3F is required for efficient restriction of Vif-deficient human immunodeficiency virus. *Virology* **450-451**, 49-54 (2014).

299. J. L. Mbisa, W. Bu, V. K. Pathak, APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. *J Virol* **84**, 5250-5259 (2010).

300. Z. Xiao, E. Ehrlich, K. Luo, Y. Xiong, X. F. Yu, Zinc chelation inhibits HIV Vif activity and liberates antiviral function of the cytidine deaminase APOBEC3G. *FASEB J* **21**, 217-222 (2007).

301. 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).

302. J. R. Bergeron *et al.*, The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. *PLoS pathogens* **6**, e1000925 (2010).

303. X. Wang *et al.*, Interactions between HIV-1 Vif and human ElonginB-ElonginC are important for CBF-beta binding to Vif. *Retrovirology* **10**, 94 (2013).

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

305. H. P. Bogerd, B. P. Doehle, H. L. Wiegand, B. R. Cullen, A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc Natl Acad Sci U S A* **101**, 3770-3774 (2004).

306. M. Letko *et al.*, Vif proteins from diverse primate lentiviral lineages use the same binding site in APOBEC3G. *Journal of virology* **87**, 11861-11871 (2013).

307. 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).

308. Y. Dang, X. Wang, T. Zhou, I. A. York, Y. H. Zheng, Identification of a novel WxSLVK motif in the N terminus of human immunodeficiency virus and simian immunodeficiency virus Vif that is critical for APOBEC3G and APOBEC3F neutralization. *Journal of virology* **83**, 8544-8552 (2009).

309. 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. *Journal of virology* **83**, 8674-8682 (2009).

310. 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).

311. 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. *Journal of virology* **81**, 8201-8210 (2007).

312. T. Yamashita, K. Kamada, K. Hatcho, A. Adachi, M. Nomaguchi, Identification of amino acid residues in HIV-1 Vif critical for binding and exclusion of APOBEC3G/F. *Microbes Infect* **10**, 1142-1149 (2008).

313. C. Richards *et al.*, The Binding Interface between Human APOBEC3F and HIV-1 Vif Elucidated by Genetic and Computational Approaches. *Cell Rep* **13**, 1781-1788 (2015).

314. M. Nakashima *et al.*, Structural Insights into HIV-1 Vif-APOBEC3F Interaction. *J Virol* **90**, 1034-1047 (2015).

315. E. Pery, K. S. Rajendran, A. J. Brazier, D. Gabuzda, Regulation of APOBEC3 proteins by a novel YXXL motif in human immunodeficiency virus type 1 Vif and simian immunodeficiency virus SIVagm Vif. *Journal of virology* **83**, 2374-2381 (2009).

316. S. Kitamura, H. Ode, Y. Iwatani, Structural Features of Antiviral APOBEC3 Proteins are Linked to Their Functional Activities. *Front Microbiol* **2**, 258 (2011).

317. S. Kitamura *et al.*, The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat Struct Mol Biol* **19**, 1005-1010 (2012).

318. M. Ooms, M. Letko, V. Simon, The Structural Interface between HIV-1 Vif and Human APOBEC3H. *J Virol* **91**, (2017).

319. 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).

320. B. Schrofelbauer, D. Chen, N. R. Landau, A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc Natl Acad Sci U S A* **101**, 3927-3932 (2004).

321. J. L. Smith, V. K. Pathak, Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. *Journal of virology* **84**, 12599-12608 (2010).

322. J. S. Albin *et al.*, A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. *The Journal of biological chemistry* **285**, 40785-40792 (2010).

323. G. Mercenne *et al.*, HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. *Nucleic acids research* **38**, 633-646 (2010).

324. S. Guerrero *et al.*, Translational regulation of APOBEC3G mRNA by Vif requires its 5'UTR and contributes to restoring HIV-1 infectivity. *Sci Rep* **6**, 39507 (2016).

325. V. B. Soros, W. Yonemoto, W. C. Greene, Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. *PLoS pathogens* **3**, e15 (2007).

326. Y. Feng, R. P. Love, L. Chelico, HIV-1 viral infectivity factor (Vif) alters processive single-stranded DNA scanning of the retroviral restriction factor APOBEC3G. *The Journal of biological chemistry* **288**, 6083-6094 (2013).

327. E. Britan-Rosich, R. Nowarski, M. Kotler, Multifaceted counter-APOBEC3G mechanisms employed by HIV-1 Vif. *J Mol Biol* **410**, 1065-1076 (2011).

328. W. C. Greene *et al.*, Novel targets for HIV therapy. *Antiviral Res* **80**, 251-265 (2008).

329. L. S. Shlyakhtenko *et al.*, Nanoscale structure and dynamics of ABOBEC3G complexes with single-stranded DNA. *Biochemistry* **51**, 6432-6440 (2012).

330. L. C. Mulder *et al.*, Moderate influence of human APOBEC3F on HIV-1 replication in primary lymphocytes. *Journal of virology* **84**, 9613-9617 (2010).

331. M. A. Langlois, R. C. Beale, S. G. Conticello, M. S. Neuberger, Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* **33**, 1913-1923 (2005).

332. 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).

333. S. Kao *et al.*, The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. *J Virol* **77**, 11398-11407 (2003).

334. K. Stopak, C. de Noronha, W. Yonemoto, W. C. Greene, HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* **12**, 591-601 (2003).

335. X. Yu *et al.*, Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060 (2003).

336. R. Suspene *et al.*, APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res* **32**, 2421-2429 (2004).

337. T. Wang *et al.*, Distinct viral determinants for the packaging of human cytidine deaminases APOBEC3G and APOBEC3C. *Virology* **377**, 71-79 (2008).

338. C. J. Wittkopp, M. B. Adolph, L. I. Wu, L. Chelico, M. Emerman, A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses. *PLoS pathogens* **12**, e1005865 (2016).

339. V. Pak, G. Heidecker, V. K. Pathak, D. Derse, The role of amino-terminal sequences in cellular localization and antiviral activity of APOBEC3B. *J Virol* **85**, 8538-8547 (2011).

340. S. E. Halford, J. F. Marko, How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* **32**, 3040-3052 (2004).

341. R. P. Love, H. Xu, L. Chelico, Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. *J Biol Chem* **287**, 30812-30822 (2012).

342. S. F. Le Grice, F. Gruninger-Leitch, Rapid purification of homodimer and heterodimer HIV-1 reverse transcriptase by metal chelate affinity chromatography. *Eur J Biochem* **187**, 307-314 (1990).

343. F. Gao *et al.*, The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *Journal of virology* **70**, 7013-7029 (1996).

344. S. Creighton, L. B. Bloom, M. F. Goodman, Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol* **262**, 232-256 (1995).

345. U. O'Doherty, W. J. Swiggard, M. H. Malim, Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. *J Virol* **74**, 10074-10080 (2000).

346. L. S. Shlyakhtenko *et al.*, Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J Biol Chem* **286**, 3387-3395 (2011).

347. L. S. Shlyakhtenko *et al.*, Atomic force microscopy studies of APOBEC3G oligomerization and dynamics. *J Struct Biol*, (2013).

348. X. Wang, P. T. Dolan, Y. Dang, Y. H. Zheng, Biochemical differentiation of APOBEC3F and APOBEC3G proteins associated with HIV-1 life cycle. *J Biol Chem* **282**, 1585-1594 (2007).

349. L. G. Holden *et al.*, Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121-124 (2008).

350. K. M. Chen *et al.*, Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116-119 (2008).

351. N. P. Stanford, M. D. Szczelkun, J. F. Marko, S. E. Halford, One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J* **19**, 6546-6557 (2000).

352. B. A. Lieberman, S. K. Nordeen, DNA intersegment transfer, how steroid receptors search for a target site. *J Biol Chem* **272**, 1061-1068 (1997).

353. A. Furukawa *et al.*, Structure and real-time monitoring of the enzymatic reaction of APOBEC3G which is involved in anti-HIV activity. *Nucleic Acids Symp Ser (Oxf)*, 87-88 (2009).

354. D. D. Loeb *et al.*, Complete mutagenesis of the HIV-1 protease. *Nature* **340**, 397-400 (1989).

355. R. Suspene, C. Rusniok, J. P. Vartanian, S. Wain-Hobson, Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucleic Acids Res* **34**, 4677-4684 (2006).

356. Y. Yang, F. Guo, S. Cen, L. Kleiman, Inhibition of initiation of reverse transcription in HIV-1 by human APOBEC3F. *Virology* **365**, 92-100 (2007).

357. R. Bransteitter, C. Prochnow, X. S. Chen, The current structural and functional understanding of APOBEC deaminases. *Cell Mol Life Sci* **66**, 3137-3147 (2009).

358. E. Harjes *et al.*, An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J Mol Biol* **389**, 819-832 (2009).

359. K. Belanger, M. A. Langlois, Comparative analysis of the gene-inactivating potential of retroviral restriction factors APOBEC3F and APOBEC3G. *J Gen Virol* **96**, 2878-2887 (2015).

360. J. F. Krisko, N. Begum, C. E. Baker, J. L. Foster, J. V. Garcia, APOBEC3G and APOBEC3F Act in Concert To Extinguish HIV-1 Replication. *J Virol* **90**, 4681-4695 (2016).

361. B. Mangeat *et al.*, Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103 (2003).

362. Y. S. Zhang, Z. X. Ning, S. Z. Yang, H. Wu, Antioxidation properties and mechanism of action of dihydromyricetin from Ampelopsis grossedentata. *Yao Xue Xue Bao* **38**, 241-244 (2003).

363. A. M. Sheehy, N. C. Gaddis, J. D. Choi, M. H. Malim, Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-650 (2002).

364. R. Mariani *et al.*, Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**, 21-31 (2003).

365. S. Jager *et al.*, Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481**, 371-375 (2012).

366. 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).

367. J. P. Vartanian, M. Henry, S. Wain-Hobson, Sustained G-->A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. *J. Gen. Virol.* **83**, 801-805 (2002).

368. 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).

369. J. P. Vartanian, A. Meyerhans, M. Sala, S. Wain-Hobson, G-->A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc Natl Acad Sci U S A* **91**, 3092-3096 (1994).

370. V. K. Pathak, H. M. Temin, Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc Natl Acad Sci U S A* **87**, 6019-6023 (1990).

371. 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).

372. 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).

373. 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).

374. B. Mangeat *et al.*, Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103 (2003).

375. 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).

376. D. Ebrahimi, F. Anwar, M. P. Davenport, APOBEC3G and APOBEC3F rarely co-mutate the same HIV genome. *Retrovirology* **9**, 113 (2012).

377. H. Zhang *et al.*, Novel single-cell-level phenotypic assay for residual drug susceptibility and reduced replication capacity of drug-resistant human immunodeficiency virus type 1. *Journal of virology* **78**, 1718-1729 (2004).

378. L. Naldini *et al.*, In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263-267 (1996).

379. K. A. Ahmed, L. Wang, P. Griebel, D. D. Mousseau, J. Xiang, Differential expression of mannose-6-phosphate receptor regulates T cell contraction. *J Leukoc Biol* **98**, 313-318 (2015).

380. P. P. Rose, B. T. Korber, Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics* **16**, 400-401 (2000).

381. R. P. Bennett, J. D. Salter, X. Liu, J. E. Wedekind, H. C. Smith, APOBEC3G subunits self-associate via the C-terminal deaminase domain. *The Journal of biological chemistry* **283**, 33329-33336 (2008).

382. S. Wurtzer *et al.*, Functional central polypurine tract provides downstream protection of the human immunodeficiency virus type 1 genome from editing by APOBEC3G and APOBEC3B. *J Virol* **80**, 3679-3683 (2006).

383. C. Hu *et al.*, The HIV-1 central polypurine tract functions as a second line of defense against APOBEC3G/F. *J Virol*, (2010).

384. J. L. Mbisa, W. Bu, V. K. Pathak, APOBEC3F and APOBEC3G Inhibit HIV-1 DNA Integration by Different Mechanisms. *J Virol*, (2010).

385. K. Shindo *et al.*, The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity. *J Biol Chem* **278**, 44412-44416 (2003).

386. X. Xiao, S. X. Li, H. Yang, X. S. Chen, Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat Commun* **7**, 12193 (2016).

387. N. M. Shaban, K. Shi, M. Li, H. Aihara, R. S. Harris, 1.92 Angstrom Zinc-Free APOBEC3F Catalytic Domain Crystal Structure. *J Mol Biol* **428**, 2307-2316 (2016).

388. K. K. Siu, A. Sultana, F. C. Azimi, J. E. Lee, Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. *Nat Commun* **4**, 2593 (2013).

389. J. A. Briggs *et al.*, The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* **11**, 672-675 (2004).

390. K. Sato *et al.*, Remarkable lethal G-to-A mutations in vif-proficient HIV-1 provirus by individual APOBEC3 proteins in humanized mice. *J Virol* **84**, 9546-9556 (2010).

391. P. An *et al.*, Role of APOBEC3F Gene Variation in HIV-1 Disease Progression and Pneumocystis Pneumonia. *PLoS Genet* **12**, e1005921 (2016).

392. B. Mangeat, P. Turelli, S. Liao, D. Trono, A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. *J Biol Chem* **279**, 14481-14483 (2004).

393. A. Furukawa *et al.*, Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J* **28**, 440-451 (2009).

394. S. M. Shandilya *et al.*, Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **18**, 28-38 (2010).

395. J. F. Hultquist *et al.*, A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. *Cell Rep* **17**, 1438-1452 (2016).