Investigation of the various modes of retroviral and endogenous retroelements restriction by APOBEC3 proteins

Kasandra Bélanger

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Microbiology and Immunology
Department of Biochemistry, Microbiology, and Immunology

Faculty of Medicine

University of Ottawa

Abstract

Mammals are constantly challenged by numerous pathogens that pose a threat to their health. Upon infection, retroviruses quickly integrate their genome into that of their host thereby permanently modifying it. Protein members of the APOBEC3 (A3) family exhibit cytidine deaminase activity that specifically acts on single-stranded DNA to deaminate deoxycytidine bases into deoxyuridines. This process is potentially mutagenic because uracil directs the insertion of adenine on the opposite DNA strand. High levels of mutations induced by A3 proteins in the retroviral genome ultimately inactivate progeny viruses. However, under conditions where low levels of A3 proteins are present, sub-lethal mutagenesis can occur and is generally believed to be beneficial for the virus. Powerful and affordable techniques designed to detect rare deamination events generated by these deaminases along the full length of retroviral genomes are therefore essential. Through the course of my studies, I developed such a new tool that I called HyperHRM which was instrumental to my project's success.

In addition to the antiretroviral affects of their catalytic activity, some members of the A3 family have the ability to hinder reverse transcription independently of their enzymatic properties. Yet, the details underlying the deamination-independent restriction by the proteins remain unclear. Through my work, I have advanced our current understanding of this elusive process by defining the essential role for RNA-binding in the inhibition of the early steps of infection by APOBEC3G (A3G). I also demonstrate that the ability to bind RNA is important for the selection of DNA dinucleotides targeted for deamination by A3 enzymes. Based on the premise that the DNA context for deamination may alter viral fitness in various ways, I then investigated the gene inactivation potency of different A3 based on

their preferred DNA substrate. My experiments showed that mutations introduced in a 5'CC context by A3G are much more lethal for the virus because of the high frequency of termination codons that are generated. I therefore clearly established that deamination target specificity has a strong influence on the overall restriction potency of A3 proteins and demonstrated that such specificity was linked to the ability of A3 proteins to bind RNA.

Finally, in addition to retroviruses, mobile elements such as retrotransposons can also lead to genomic instability if not properly controlled. The A3 protein family has been shown to play a crucial role in the restriction of these elements through a mechanism that is not believed to require the enzymatic activity of the proteins, although the details of the restriction mechanism are not yet understood. Here, I provide molecular insights on the potential mechanism of retrotransposon restriction by showing that the RNA-binding properties of the enzymes are not involved in the restriction of L1 retrotransposition. A complete elucidation of the modes of restriction employed by the A3 could lead to the development of a new generation of antiretroviral drugs.

Overall, my research has led to the design of a new research tool to detect and quantify A3-induced mutations in retroviruses, but more importantly, it has enabled a better understanding of how the RNA-binding abilities of A3 proteins play an essential role in the overall restriction potency of retroviruses and retrotransposons.

Acknowledgements

First, I would like to express my gratitude to my supervisor Dr. Marc-André Langlois. Although rough by moments, my doctoral degree has without a doubt been the most rewarding experience of my life and a big part of that is due to the great leadership and mentorship of Marc. His passion for research and the great scientist that he is have contributed to the scientist that I have become and for everything that I have learned under his supervision, I could never thank him enough. Thank you for giving me so many opportunities to improve myself. For your many advices that I always took extremely seriously. Thank you for being a mentor through a critical step in my career.

My TAC members Dr. Earl Brown, Dr. Ashok Kumar and Dr. Seung-Hwan Lee for their assistance and guidance throughout the accomplishment of my degree.

Dr. Martin Bourbonnière, Dr. Kristin Kemmerich and Dr. Jason Fernandes for technical assistance and expertise. Dr. Bourbonnière helped with the design and trouble-shooting of the HRM experiments whereas Dr. Kemmerich provided previous advices for bacterial assays performed during the course of my PhD. Dr. Fernandes offered technical expertise for the isolation of PBMCs from HIV-infected patients.

Maria Rosales Gerpe, lab mate and friend for her constant interest in my research and the many times she helped me with my experiments. Because you were the only person who really understood what I was going through and with who I could talk about it. Thanks for listening and always having time for me.

My lab mates, past and present, Tyler Milston Renner, Cindy Lam, Dr. Joanne McBane, Tara Read, Shabnam Rahimi, Halil Aydin and Olga Agah for their help and support throughout my degree. Special thanks go to Tyler Renner for many discussions that sometimes led to arguments but also made me think "outside the box". All the students who worked with me; Mathieu Savoie, Mark Campbell and Laura Rose Goodwin for all their help and for pulling through the pressure we went through together. Thank you Matt for your great sense of humour that made everything more enjoyable and totally unethical at the same time.

My mom, Suzanne Bouchard, for supporting me, worrying for me, always being so proud of me and never doubting me. Thanks for the many attentions you gave me when time was missing. My dad, André Bélanger, for many advices that may have seemed rough at first but that were always well thought and logical. Thank you both for your generosity and for making my life so much easier.

Finally, my best friend and partner in life, my husband Jean-François Duval for pulling through the last six years aside me and giving me the opportunity to fulfill my dream. For your patience, your support and your understanding. Thank you for all the times you listened to my presentations, for all the times you reassured me prior to a presentation, for all the times I had a migraine after a presentation and you took care of me. For all the sacrifices you made during the last six years, I am forever thankful.

Table of content

List of Abbreviations

A: Alanine
A1: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1
A2: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
A3A: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A
A3B: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B
A3C: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C
A3D:Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3DE
A3F: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F
A3G: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
A3H: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3H
AID: Activation- induced cytidine deaminase
AIDS: Acquired immune deficiency syndrome
ALV: Avian Leukemia Virus
AMP: Ampicillin

BAF-1: Barrier-to-autointegration-1
BLV: Bovine Leukemia Virus
bp: base-paired

C: Cysteine
CA: Capsid
Cat-α: Catalase-α
cDNA: complementary DNA
CDS: Coding DNA sequence
CMV: Cytomegalovirus
Co-IP: Co-Immunoprecipitation
cPPT: central Polypurine tract
CRPV: Cotton tail Papilloma virus
CTD: C-terminal domain
Cys: Cysteine

D: Aspartic acid
3D-PCR: Differential DNA denaturation PCR
ddPCR: droplet digital PCR
dGTPase: deoxyguanosine triphosphatase
DNA: Deoxyribonucleic acid
dNTP: deoxynucleoside triphosphate
DSB: Double-strand break
dsDNA: double-stranded DNA

E: Glutamic acid
E. coli: Escherichia Coli
EIAV: Equine Infectious Anemia virus
ELISA: Enzyme-Linked Immunosorbent Assay
Env: Envelope

FBS: Fetal bovine serum
FIV: Feline Immunodeficiency virus
FLV: Feline Leukemia Virus

G: Glycine
Gag: Group-specific antigen
GALT**:** Gut-associated lymphoid tissue
gDNA: genomic DNA
GPI: Glycosylphosphatidylinositol
GST: Glutathione S-transferase

H: Histidine
HAART: Highly active antiretroviral treatment
HEK: Human epithelial kidney
HERV: Human endogenous retrovirus
His: Histidine
HIV: Human Immunodeficiency Virus
HLA: Human leukocyte antigen
HMM: High molecular mass
hnRNP: heteregenous Ribonuceloprotein
HNSC: Head and neck squamous cell carcinoma
HRM: High resolution melting
HTLV-1: Human T-cell leukemia virus type 1

I: Isoleucine
IFN: Interferon
IL-2: Interleukin-2
IL-5: Interleukin-5
IL-7: Interleukin-7
IL-15: Interleukin-15
IN: Integrase
IP: Immunoprecipitation
IPTG: Isopropyl β-D-1-thiogalactopyranoside

K: Lysine
Kd: apparent equilibrium dissociation constant
kDa:kiloDalton
kb: kilobase

L1: LINE-1
L2: LINE-2
LINE: Long interspersed nucleotide element
LMM: Low molecular mass
LRT: Late reverse transcripts
LTR: Long terminal repeat

M: Methionine
MA: Matrix
mA3: mouse Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3
MDa: Mega Dalton
MHC: Major histocompatibility complex
mL: milliliter
MLV: Murine Leukemia Virus
mM: milimolar
MMTV: Mouse Mammary Tumor virus
M-MLV: Moloney Murine Leukemia Virus
M-PMTV: Mason-Pfizer Monkey-Type Virus
mRNA: messenger RNA

N: Asparagine
NC: Nucleocapsid
Nef: Negative regulatory factor
nm: nanometer
nM:nanomolar
NTD: N-terminal domain

O.D.: Optical density
OH: Hydroxyl
ORF: Open reading frame

PBMC: Peripheral blood mononuclear cell
P-bodies: Processing bodies
1X PBS: Phosphate buffer saline
PBS: Primer binding site
PBS-T: 1X PBS 0.1% Tween 20
PCR: Polymerase-chain-reaction
PIC: Pre-integration complex
ρmol: picomole
Pol: Polymerase
3'PPT: LTR proximal polypurine tract
PPT: Polypurine tract
PRO: Protease

qPCR: quantitative real-time PCR

R: Arginine
R: Repeat
R^2 : R-Squared
Rev: Regulator of expression of virion proteins
Rif^R: Rifampicin resistance
RNA: Ribonucleic acid
RNP: Ribonucleoprotein
RRE: Rev response element

RSV: Rous Sarcoma Virus
RT: Reverse transcription
RT-PCR: Reverse-transcription polymerase chain reaction
RTC: Reverse transcription complex

S: Serine
SAMHD1: Sterical-alpha motif and histidine-aspartic protein 1
SDM: Site-directed mutagenesis
SINE: Small interspersed nucleotide element
SG: Stress granule
SNP: Single nucleotide polymorphism
ssDNA: single-stranded DNA
-/+sssDNA: minus- or plus-strand strong-strop DNA
STD: Sexually transmitted disease
SU: Subunit
SVA: SINE-VNTR-Alu

T: Threonine
Tat: Trans-activator of transcription
TAR: Trans-activating response element
TE: Transposable element
TLR: Toll-like receptor
TM: Transmembrane
TPRT: Target-primed reverse transcription
tRNA: transfer RNA
TSD: Target site duplication

U: Unit
UDG: Uracil-DNA glycosylase
µL: microliter
µm: micrometer
µM: micromolar
UTR: Untranslated

V: Valine
Vif: Viral infectivity factor
Vpr: Viral protein r
Vpu: Viral protein u

W: Tryptophan
WHO: World Health Organization

X-Gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Y: Tyrosine

Z: Zinc

List of Figures

List of Tables

Chapter 1 General Introduction

1.1 Retroviruses

1.1.1 A brief history of retroviruses

In the early nineteenth century, bacteria were to know to scientists as the cause of many diseases. However, it was determined that bacteria were not responsible for the transmission of all illnesses and that a secondary type of invisible agent may exist. At the end of the century, joint work from Dmitri Ivanovsky and Martinus Beijerinck in the Tobacco plant led to the discovery of the first non-bacterial infectious agent (1, 2). This infectious agent was able to reproduce itself, but unlike bacteria, it could not be recovered by filtration. This small infectious agent was given the name of virus. In 1908, Ellermann and Bang from Denmark demonstrated the successful cell-free transmission of leukemia in chickens and related it to the presence of a viral agent (3). Not long after, Peyton Rous showed the similar transmissibility of sarcoma from one chicken to another (4). They had just discovered the first tumor viruses now known as Avian Leukosis virus (ALV) and Rous Sarcoma virus (RSV). At the time, their findings were highly criticised and did not receive the attention they deserved since leukemia was not yet recognized as a true form of cancer. Furthermore, chickens were not considered valuable models as they are not closely related to humans and the general belief was that no clinically-relevant conclusions could be drawn from these organisms. Only a few decades following these initial discoveries, Richard Shope and E. Western Hurst identified the first mammalian tumor virus, Shope Papillomavirus or Cottontail Rabbit Papillomavirus (CRPV) (5). They demonstrated that warts from wild cottontail rabbits could be transmitted from one animal to another through inoculation with filterable agents extracted from a wart. In 1936, John Bittner reported the presence of a viral agent, Mammary Tumor virus (MMTV), in the milk of certain mouse strains that could be

transmitted to new-borne pups and that resulted in the formation of mammary tumors (6). A few years later, in a quest to uncover more viruses at the origin of cancer in mice, Ludwik Gross isolated a virus that induced the development of leukemia, Murine Leukemia Virus (MLV). This established that leukemia was a different form of cancer, transmissible and contagious (7). At that moment, there was an important revolution in the field of virology leading to the beginning of a hunt for viruses that can cause cancer in mammals. Several tumor viruses were uncovered during that period in different animal species including but not limited to Moloney Murine Leukemia Virus (M-MLV), Bovine Leukemia Virus (BLV), Feline Leukemia Virus (FLV) and Mason-Pfizer Monkey-Type Virus (M-PMTV) (8-10).

The study of the behavioural, structural and molecular features of newly identified pathogens was however very limited at the time by the available techniques. An important leap forward came from Renato Dulbecco who developed the plaque assay in 1952 allowing the quantification of single viral particles in a sample (11). Briefly, the assay, adapted from the field of bacteriophages, consisted of establishing a cell monolayer from chicken embryos that was then infected with the desired viral culture. Melted agar was subsequently added to the samples, solidified and the embryos were then left to grow for several days. In the case of an infection, the activity of the virus resulted in the formation of a plaque, defined as a circular region on the agar originating from necrotic cells. Dulbecco and his group established that there was a linear correlation between the number of plaques and the concentration of virus used to inoculate the cells suggesting that one plaque corresponded to one virus, making it easily quantifiable. A few years later, Temin and Rubin introduced the focus assay which was an improved version of the plaque assay (12). In this case, infection of chick embryos resulted in the formation of a sarcoma. Therefore, not only did it allowed researchers to determine viral concentrations at the single-cell

to single-viral particle level but since measurements of infection depended on the oncogenic transformation of the cell, it could also determine the genetic potential of an infection. The use of cell-based assays led to significant advancements in the field through the study of viral replication, transmission and genetics. Yet, despite these breakthroughs, there were still many unresolved questions regarding tumor viruses displaying an RNA genome.

Scientists in the 1960s were challenged by observations such that no RNA replication intermediates could be detected in cells infected with RNA viruses and that inhibitors of DNA synthesis were capable of blocking the replication of RNA viruses (13). In 1964, Howard Temin proposed the provirus hypothesis as a tentative explanation (14). He suggested that, upon infection with RSV, a DNA copy of the viral RNA genome would be created and inserted into the host cell's DNA. His theory, validated today, was highly controversial at the time since it went against what was known as the Central Dogma of molecular biology, that the flow of genetic information goes from DNA to RNA to proteins. A few years later, the discovery of an RNA-dependent DNA polymerase, the reverse transcriptase (RT), by David Baltimore provided the ultimate proof for the provirus hypothesis and provided answers for the prominent unsolved observations in the field (15). Viruses possessing an RNA genome and an active RT were classified as retroviruses and only then did it come to light that many tumor viruses identified several years back were indeed part of this unique group. The joint discovery that some viruses could interact with the cellular genetic material by David Baltimore, Howard Temin and Renato Dulbecco earned them the Nobel Prize in Physiology or Medicine in 1975.

With fresh evidence that some viruses called retroviruses could insert their genetic material into the host genome, there was a rekindled interest in the 70's to uncover cancer-causing retroviruses in humans. Although initial findings were all negative and led to the general

belief that there existed no such human retroviral pathogen, in 1980 the group of Robert C. Gallo identified a retrovirus as being the cause of adult T-cell leukemia after the suspicious and recurrent observation that the disease tended to be geographically clustered suggesting the presence of a transmissible agent (16, 17). Culture of T cells from infected patients, only rendered possible after the discovery of a T cell growth factor now known as interleukin-2 (IL-2), led to the isolation of the first human retrovirus, the Human T-cell Leukemia virus 1 (HTLV-1) (18, 19). This virus was different from other oncogenic retroviruses in that its genome was more complex and encoded additional proteins, Tax and Rex (20). The gene products were later found to have a role in the regulation of transcription and gene expression respectively (21).

Almost coinciding with the discovery of HTLV-1, an important acquired immunodeficiency syndrome (AIDS) outbreak occurred in developing countries leading to intensified efforts to uncover the cause of this ravaging disease. Research efforts were soon rewarded by the identification of a virus now infamously known as the Human Immunodeficiency Virus type 1 (HIV-1) (22-24). The sequencing of the HIV-1 genome revealed the presence of supplementary proteins not found in other viruses. Viral infectivity factor (Vif), Viral protein r (Vpr), Viral protein u (Vpu), Trans-activator of transcription (Tat), Regulator of expression of virion proteins (Rev) and Negative regulatory factor (Nef) are all involved in viral replication and pathogenesis and will be discussed in more details in the following section of this chapter.

The complexity and the rapid evolution of HIV-1 have made this retrovirus an expert of immune evasion and drug resistance. More than three decades after its discovery, to this day, the main challenges in retrovirology remain the development of a curative treatment and a protective vaccine for HIV infection.

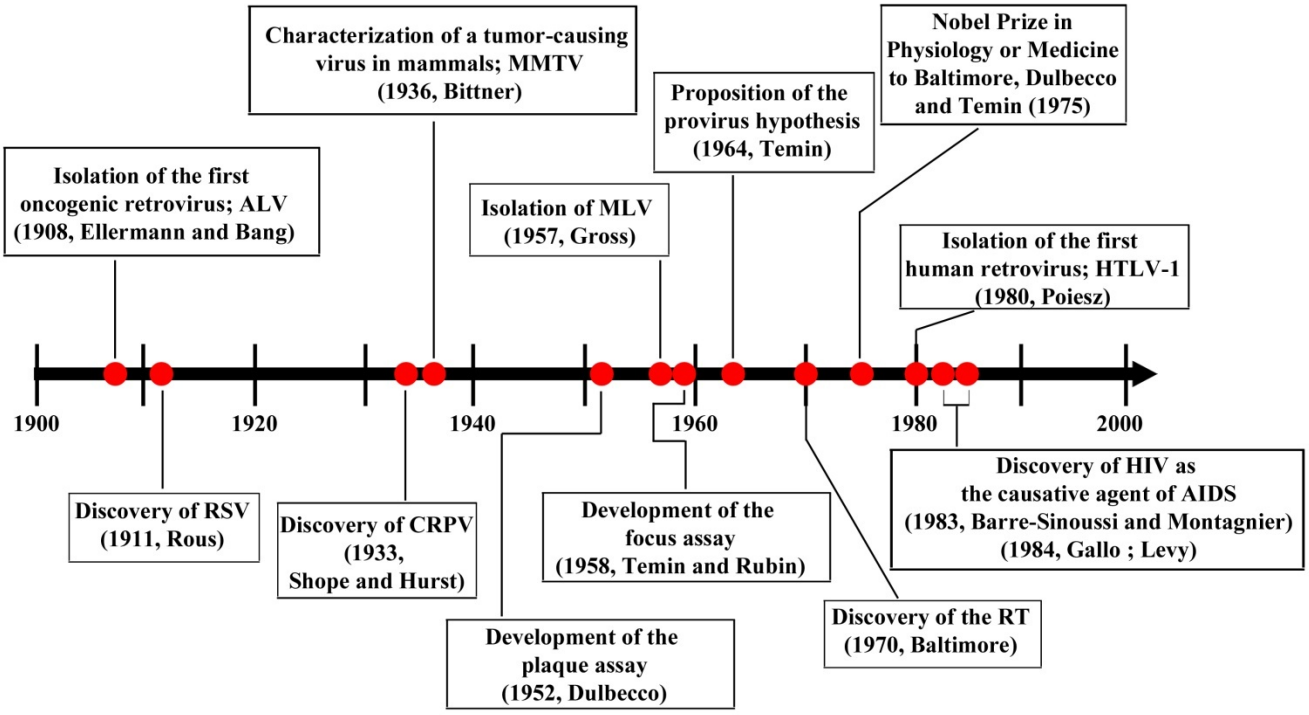

Characterization of a tumor-causing virus in mammals; MMTV (1936, Bittner)
Nobel Prize in Physiology or Medicine to Baltimore, Dulbecco and Temin (1975)
Proposition of the provirus hypothesis (1964, Temin)
Isolation of the first oncogenic retrovirus; ALV (1908, Ellermann and Bang)
Isolation of MLV (1957, Gross)
Isolation of the first human retrovirus; HTLV-1 (1980, Poiesz)
1900
1920
1940
1960
1980
2000
Discovery of RSV (1911, Rous)
Discovery of CRPV (1933, Shope and Hurst)
Development of the focus assay (1958, Temin and Rubin)
Discovery of HIV as the causative agent of AIDS (1983, Barre-Sinoussi and Montagnier) (1984, Gallo ; Levy)
Development of the plaque assay (1952, Dulbecco)
Discovery of the RT (1970, Baltimore)

Figure 1.1 Timeline depicting major breakthroughs in the discovery of retroviruses. To note, this figure does not include all viral agents isolated during the period covering 1900-2000. Only discoveries relevant to this thesis are indicated on the figure (3-7, 11, 12, 14-16, 22-24).

1.1.2 HIV-1

1.1.2.1 The viral genome

HIV-1 is a complex retrovirus from the *lentivirus* genus. Its genome is composed of a single-stranded RNA of positive orientation of approximately 9.8 kilobases (kb). It codes for 9 genes yielding 15 different proteins due to alternative splicing and protease cleavage. These include three large essential polyproteins (Gag, Pol and Env) and six accessory proteins (Vif, Vpr, Tat, Rev, Vpu and Nef) (see **Figure 6.2.3A** for schematic representation of the HIV-1 genome). The *gag* and *pol* genes are generated as unspliced messenger RNAs (mRNAs) to yield polypeptides that then get cleaved within the mature virion by the viral protease. Cleavage of Gag produces the individual matrix (MA, p17), capsid (CA, p24) and nucleocapsid (NC, p7) structural proteins whereas the protease (PRO), reverse transcriptase (RT, p51 and p66 hetero-dimer), RNase (p15) and integrase (IN, p31) proteins all originate from the *pol* gene (25). The *env* mRNA arises from a singly spliced transcript to produce a polyprotein (gp160) which gets glycosylated in the endoplasmic reticulum (ER) and cleaved by cellular proteases to yield two glycoproteins: a surface-exposed subunit (SU, gp120) and a transmembrane subunit (TM, gp41) which are then trafficked to the cell membrane as hetero-trimers (26). Similarly, Vif, Vpr and Vpu are also translated from singly spliced mRNAs whereas Tat, Rev and Nef are produced from fully spliced products (27). HIV-1's accessory genes are usually classified as regulatory (Tat and Rev) and accessory (Vif, Vpr, Vpu and Nef) proteins.

1.1.2.2 The virion

In the mature virion, two copies of the viral RNA genome are contained inside a cone-shaped shell formed by several thousand copies of the capsid protein (CA, p24) where it is

stabilized through its interaction with NC proteins together with the viral enzymes required for replication: RT, IN and PRO. Some accessory proteins are also found inside viral particles notably Vif , Vpr and Nef. The viral core is surrounded by a protective matrix containing hundreds of repetitions of the MA protein, which is in turn protected by a lipid bilayer viral envelope. The viral envelope protrudes from the host cell membrane and comprises two glycoproteins originating from the *env* gene: gp120 and gp41 (25). A schematic representation of the virion organisation of HIV-1 can be found in **Figure 1.2A**.

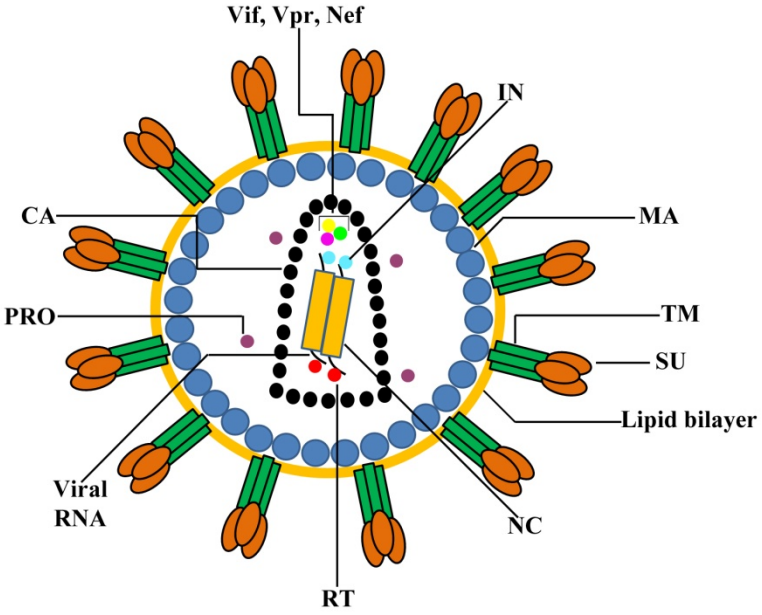
1.1.2.3 The viral replication cycle

The HIV-1 replication cycle begins with the recognition of a target cell through the interaction of the envelope glycoprotein gp120 with the host cell receptor CD4 expressed at the surface of T-helper lymphocytes, cells from the monocyte-macrophage lineage and some dendritic cells (28). Binding of CD4 by gp120 leads to a change in conformation of the variable loops of the glycoprotein to expose the binding site for a chemokine co-receptor, CXCR4 or CCR5. (29, 30). The use of either one of these co-receptors determines the cell tropism of the virus. Co-receptor binding by gp120 triggers the insertion of gp41 into the cell membrane, fusion and entry into the host cell (31). Once in the target cell, the HIV-1 core remains intact until completion of reverse transcription, which takes place inside the viral core where the viral RNA genome and the enzymes required for replication form the reverse transcription complex (RTC) (32). Uncoating of the HIV-1 capsid takes place at the nuclear pore at a later stage of the replication cycle prior to nuclear entry by a mechanism that is not completely understood (32). The integrity of the viral core until this point is required to ensure successful reverse transcription and transport to the nucleus. The containment of the viral genome inside the core

until the end of viral replication also represents a mechanism used by the virus to prevent recognition by cytosolic DNA sensors that would otherwise trigger an immune response (33).

The main feature distinguishing retroviruses from other viruses is their special mode of replication, which involves the generation of a double-stranded DNA (dsDNA) copy from their single-stranded RNA (ssRNA) genome called a provirus. It is then this provirus that integrates into the host genomic DNA with the aid of the viral encoded IN and other host proteins for transcription and translation using the host cell machinery (34). The process of reverse transcription is illustrated in **Figure 1.2B**. The latter is initiated by an influx in deoxynucleoside triphosphates (dNTPs) in the capsid following viral entry into the cell. The cellular $tRNA_3^{Lys}$, already binding to a complementary sequence near the 5' end of the HIV RNA named the primer binding site (PBS), acts as a primer for the RT which can then start the synthesis of complementary DNA (cDNA) from the RNA template in a 3' to 5' orientation. Once it reaches the 5' end of the viral genome, the polymerase together with its new product undergo a template switch, known as the minus-strand transfer, and jump to the 3' end of the same viral RNA strand or of the second RNA genome where the newly synthesized repeat region (R) associates with its complementary sequence. This step requires the digestion of the RNA in the RNA:DNA hybrid by the RNase H subunit of the RT. The first cDNA fragment generated is relatively short because of the position of the PBS on the viral genome and the orientation of reverse transcription and is termed minus- strand strong stop DNA (-sssDNA). Synthesis of the minus strand DNA (-sDNA) continues from the R region accompanied by digestion of the template by RNase H. The activity of the nuclease is however only partial and leaves behind regions of the RNA genome such as the polypurine tracts (PPT). HIV-1 has two PPTs, one near the middle of the viral genome (cPPT) and one near the 3' end of the RNA (3'PPT). The cPPT is thought to

A)


Vif, Vpr, Nef
IN
CA
MA
PRO
TM
SU
Lipid bilayer
Viral
RNA
NC
RT

B)

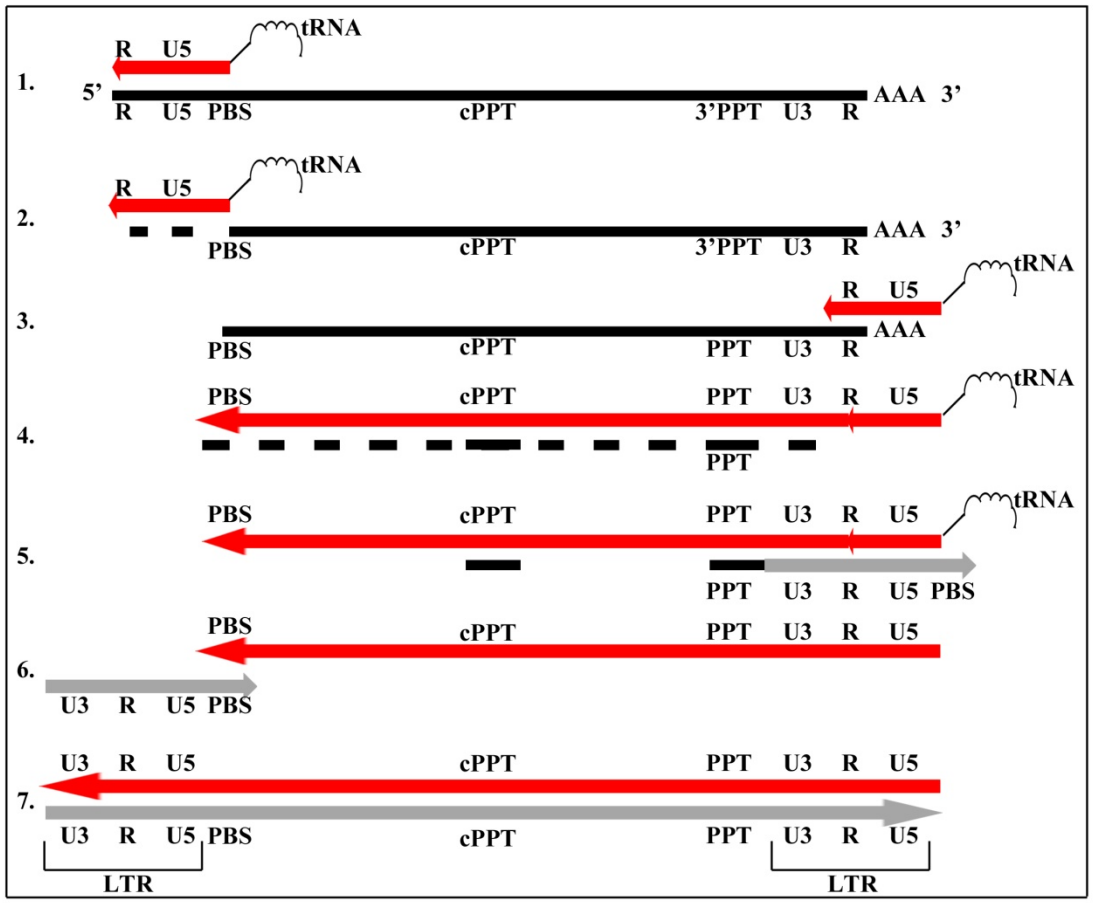

1.
tRNA
R U5
5'
R U5 PBS
cPPT
3'PPT U3 R
AAA 3'
2.
tRNA
R U5
PBS
cPPT
3'PPT U3 R
AAA 3'
3.
tRNA
R U5
PBS
cPPT
PPT U3 R
AAA
4.
tRNA
PBS
cPPT
PPT U3 R U5
PPT
5.
tRNA
PBS
cPPT
PPT U3 R U5
PPT U3 R U5 PBS
6.
PBS
cPPT
PPT U3 R U5
U3 R U5 PBS
7.
U3 R U5
cPPT
PPT U3 R U5
U3 R U5 PBS
cPPT
PPT U3 R U5
LTR
LTR

Figure 1.2 Characterization of HIV-1 (**A**) Illustration of the mature HIV-1 virion. TM and SU; transmembrane and surface subunits of envelope, NC; nucleocapsid, RT; reverse transcriptase, PR; protease, CA; capsid, IN; integrase, MA; matrix. (**B**) Schematic representation of the reverse transcription process. 1. Beginning of cDNA synthesis at the primer binding site (PBS) primed by cellular $tRNA^{Lys3}$. 2. Degradation of the RNA template by the RNase H subunit of the RT following generation of the minus-strand strong stop DNA (-sssDNA). 3. Minus strand transfer and pairing of the complementary R regions. 4. cDNA synthesis and incomplete degradation of the RNA template. 5. Synthesis of the plus strand DNA from the 3' polypurine tract (3'PPT) to generate the positive-strand strong stop DNA (+sssDNA). 6. Plus strand transfer and pairing of the two PBS regions. 7. Completion of the cDNA synthesis on both strands using each other as a template.

increase the efficiency of viral replication but is not essential (35-37). The 3'PPT is essential for the reverse transcription process and acts as a primer for the plus-strand DNA (+sDNA) synthesis, which only proceeds until the 3' end of the -sDNA, where following digestion of the tRNA primer, a second strand transfer occurs. This one is known as the plus-strand transfer and yields the plus-strand strong stop DNA (+sssDNA), where the PBS of one strand anneals to the PBS of the other strand and the sequence of each DNA strand acts as a template to complete the synthesis of the opposite strand. Completion of the process produces two identical DNA strands, which are each a copy of the RNA genome with repeats at both ends known as provirus.

At the end of reverse transcription, the core is uncoated and the dsDNA becomes part of a pre-integration complex (PIC) to be translocated into the nucleus (32). This step is believed to be highly dependent on the presence of Vpr, which is important to connect the PIC with the cellular import machinery (38). Vpr is also responsible for inducing cell cycle arrest in the G2 phase, which may be more favourable for entry of the PIC into the nucleus (39, 40). In contrast to most other retroviruses, HIV does not require cell division for proviral integration to take place. Once in the nucleus, the provirus is inserted into the host chromosomal DNA through the enzymatic activity of the IN via a highly regulated mechanism (41).

Viral gene expression, driven by the long terminal repeat (LTR) promoter, is responsible for the recruitment of several cellular transcription factors. Initially, only fully spliced mRNA transcripts are being produced. The accessory protein Tat vastly improves the efficiency of transcription (42, 43). Once synthesized, this gene product returns to the nucleus where it interacts with a stem loop element in the viral RNA (trans-activating response element (TAR)) and activates transcription of viral genes. The Rev protein also returns to the nucleus where it associates with the rev response element (RRE) of emerging transcripts to allow the export of

unspliced and singly spliced products (44). Rev activity allows the cytoplasmic translocation of the full-length RNA genome as well as remaining viral proteins Gag and Pol.

Virion assembly is then orchestrated by the Gag polyprotein (45). The packaging of the viral genome is mediated by the psi (Ψ) element located in the U5 region of its 5' LTR through its interaction with the NC domain of Gag whereas envelope glycoproteins are naturally acquired from the cell membrane as budding occurs (46). After the virus emerges from the cell, polyproteins Gag and Pol are cleaved inside particles by the viral protease through a process known as maturation to yield individual viral proteins and produce the mature virion with a conical shape which is then ready to infect a new cell and contribute to the spreading of the initial infection.

1.1.2.4 Pathogenesis of HIV-1 infection

HIV-1 infection is usually a sexually transmitted disease (STD) although transmission can also occur through the sharing of syringes by drug users, through blood transfusions and vertically from mother to child. Following the initial infection, illness is marked by an acute phase during which the virus replicates intensively and can reach viremia levels of up to 10 million copies per milliliter (mL) of plasma (47). Infected individuals experience flu-like symptoms varying in intensity (48). This stage is characterized by an important loss in CD4+ T cells in peripheral blood and a drastic depletion of this cell population in gut-associated lymphoid tissue (GALT) (49-52). Although a rebound in cell number in peripheral blood is often observed following the acute phase of HIV infection, the GALT never fully recovers. Viral levels normally peak a few weeks after the primary transmission and gradually decrease to reach a set point where they can reside for several years depending on the individuals. Spreading from

the GALT to other lymphoid tissues in periphery leads to the establishment of early latent reservoirs in CD4+ T cells (53). In those reservoirs, full-length integrated proviruses avoid the production of viral proteins and therefore prevent immune recognition (54, 55).

The second phase, or chronic phase, of HIV-1 infection is marked by a progressive loss of CD4+ T cells accompanied by several malfunctions of immune cell populations slowly leading to immunodeficiency (reviewed in (56)). This state increases the risk of opportunistic infections. The combination of a defective immune system and an increased susceptibility to opportunistic infections is defined as AIDS. Until the development of inhibitors of RT and PRO in the 1990s, this stage often ended by the patient dying. However, the use of antiretroviral agents significantly reduced the mortality rate associated with HIV-1 (57-59). Initial treatments were given as a monotherapy but quickly evolved to include a combination of anti-HIV-1 drugs. The enzyme that the virus uses to replicate, the RT, introduces on average one mutation every 1000-10,000 nucleotides synthesized resulting in up to 10 mutations per genome per round of replication (60-62). Given this extremely high mutagenic potential, it is not surprising that mutant viruses with increased resistance to treatments can be found in the existing viral population at the time of intervention (63). A way to circumvent this problem is to use a combination of antiretroviral agents therefore reducing the chances that variant viruses harbour multiple mutations giving them resistance. Mathematical modeling studies predicted that combinations of three compounds would offer an effective long-term protection defining the guidelines for the first highly active antiretroviral treatment (HAART) (63-66). Since the establishment of the first anti-HIV-1 therapies, several advancements have been made improving the composition of antiretroviral agents used for HAART to include more potent agents that are less primed to trigger resistance. HAART has the potential to reduce plasma viral load below

detection limits and suppress viral replication for several years, therefore extending the life expectancy of infected patients (67-69). Unfortunately, HAART cannot completely eliminate HIV-1 infection. The presence of viral reservoirs containing latently infected cells that are resistant to HAART and can be reactivated to productive HIV-1 once HAART is stopped prevents the complete eradication of the disease (70, 71).

1.2 Intrinsic immunity to retroviral infection

1.2.1 Restriction factors

An important feature of retroviruses is that they have extremely fast infection kinetics. From the moment when the viral envelope fuses with the cell membrane to enter into the cell, it only takes a few hours until the viral genome integrates into the host's genome and modifies it irreversibly. As a consequence, our innate and adaptive immune systems which take from a few days to several weeks to become activated, are poorly adapted at providing a good protection against retroviral infections. Surprisingly in humans, only two retroviruses out of potential hundreds are known to cause disease; HIV and HTLV (72, 73). The main reason why we are so resistant to retroviral infections is that we possess a powerful protection against these pathogenic invaders. This protection comes from our intrinsic immune defences and relies on retroviral restriction factors, representing our first line of defence against retroviruses. The most effective ones that will be introduced in this section are Trim5α, APOBEC3, SAMHDI and Tetherin (**Figure 1.2**). These proteins are constitutively expressed in our cells and are always ready to act at different stages of the retroviral life cycle at the moment of infection.

1.2.1.1 Trim5α

Trim5α is a member of the tripartite motif family, characterised by the presence of three specific domains at its N-terminus: a RING domain, a B-Box domain and a coiled coil domain (74, 75). Although retroviral restriction by this restriction factor is poorly effective against pathogens from the same host species, it plays an essential role in preventing cross-species transmission (76, 77). Indeed, human Trim5α is poorly effective against HIV-1 but efficiently restricts infection by some MLVs and by Equine Infectious Anemia virus (EIAV) (78, 79). Similarly, HIV-1 is strongly inhibited by rhesus macaque and African green monkey Trim5α (76). This restriction factor has been shown to act immediately after viral entry into the cell by inducing the premature uncoating of the capsid through a process involving the proteosomal degradation of retroviral proteins to disrupt the structure of the RTC therefore inhibiting reverse transcription (80) (**Figure 1.2a**). The exact mechanism is unclear. However, the domain organization of Trim5α induces its spontaneous assembly into hexagonal stripes, matching the conical shape of the viral capsid (81). Thus, a model can then be proposed whereby the hexagonal structure initially assembles on the capsid as it enters the cell to disrupt the viral core followed by engagement of the proteasomal machinery (82).

1.2.1.2 APOBEC3

The apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (APOBEC3) family is comprised of seven genes APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G) and APOBEC3H (A3H) located in tandem within a 150 kb region on chromosome 22 (reviewed in (83)). These proteins are cytidine deaminases that act at the level of reverse transcription by inducing deleterious

hypermutation in the provirus and will be further discussed in section 1.4 of this chapter (**Figure 1.2b**).

1.2.1.3 SAMHD1

SAMHD1 (sterical-alpha motif (SAM) and histidine-aspartic (HD) protein 1) was first described as a deoxyguanosine triphosphatase (dGTPase) triphosphohydrolase that acts specifically on deoxynucleoside triphosphate (dNTP) (84). HIV-1 is special compared to other more simple retroviruses in that it has the ability to replicate in the presence of low concentration of dNTPs as a consequence of the high affinity of its RT for nucleosides compared to other RTs (85). Resting cells such as resting CD4+ T cells, macrophages and dendritic cells have a limited pool of dNTPs and only highly specific retroviruses such as HIV-1 have the ability to replicate in those cells (86). SAMHD1 restricts retroviral infection in non-dividing cells by decreasing the concentrations of dNTPs rendering them unavailable for reverse transcription (87-89) (**Figure 1.2c**).

In addition to this function, a ribonuclease activity has recently been identified for SAMHD1 (90). This enables SAMHD1 to specifically target retroviral genomic RNA for degradation, thus conferring another level of retroviral restriction by the protein (91).

1.2.1.4 Tetherin

Tetherin is a type II transmembrane (TM) protein with a TM anchor to its N-terminus and a glycophosphatidylinositol (GPI) anchor to its C-terminus (92). This restriction factor exerts its effect at the end of the retroviral life cycle by preventing the release of virions from the cells and trapping them at the cell membrane for further internalization into endosomes and most likely

lysosomal degradation (93, 94). The current model for the tethering of viral particles by tetherin stipulates that it does so in an axial configuration in which proteins act in concert as dimers with one pair of GPI or TM anchors embedded in the virion and the other in the cell membrane (95). Although both orientations are possible, it seems that the insertion of the GPI anchors into the viral particle is favoured (**Figure 1.2d**).

1.2.2 Retrovirus versus host

Retroviruses have evolved ways to counteract host defence mechanisms exerted by antiretroviral proteins (**Figure 1.2**). In fact, several restriction factors were discovered through efforts for uncovering the function of retroviral gene products (94, 96-98). HIV-1 is probably the best example of that because unlike simple retroviruses, HIV-1 expresses six additional accessory genes with many of them involved in the pathogenesis of the virus.

A3G was discovered on the basis that some cell lines were permissive for Vif-deficient HIV-1 whereas other non-permissive cell-lines failed to produce infectious viral particles following infection with the same virus. After many years of questioning, Sheehy and her colleagues performed a series of northern blot experiments using probes derived from RNA isolated from a permissive and a non-permissive cell line. The probes were then hybridized against RNA isolated from different cell lines displaying a permissive or non-permissive phenotype. Through their manipulations, Sheehy and her group identified A3G (originally named CEM15) as the limiting component to the production of infectious viruses in non-permissive cells (96). Other A3 enzymes were identified soon after using genetic homology approaches (99, 100). The main role of Vif is to reduce cellular levels of A3 to prevent the incorporation of the restriction factor inside viral particles. Vif does so by binding to the A3

protein and to the Cul5 E3 ubiquitin ligase complex which then, assisted by CBF-β, mediates A3 ubiquitination and degradation through the proteasome (reviewed in (101)). Inhibition of A3G protein's translation by Vif has also been suggested (102). HIV-1 particles lacking a functional Vif gene are unable to produce infectious progeny in cells expressing A3G.

Tetherin was isolated following the observation that the accessory protein Vpu is required for the efficient release of viral particles from certain cell types. Microarray analyses carried out on mRNA libraries derived from phenotypically different cell lines led to the identification of the restriction factor. This protein works by downregulating the expression of tetherin at the cell surface, although the exact mechanisms by which this is achieved remain elusive (**Figure 1.2**). It may involve inhibition of its transport to the plasma membrane, its sequestration in endosomes for further lysosomal degradation or its engagement in the proteasomal machinery (103-110). All these possibilities might be true depending on the state and type of cells being infected.

Finally, HIV-2 expresses Vpx instead of the Vpu protein found in HIV-1 which is involved in the proteasomal degradation of SAMHD1 (111). Until now, no HIV-1 viral protein has been shown to inhibit the activity of SAMHD1. However, the expression of SAMHD1 is more pronounced in cells of myeloid lineage such as macrophages and dendritic cells(112). Although HIV-1 fails to replicate in dendritic cells, the virus produces a successful infection in macrophages. Kyei et al. recently identified cyclin L2 as a host protein used by HIV-1 in macrophages to overcome the restriction exerted by SAMHD1 (113). Similarly to Vpx, cyclin L2 binds to SAMHD1 and induces its proteasomal degradation. Why the same mechanism is not used by HIV-1 in other myeloid cells is not currently known.

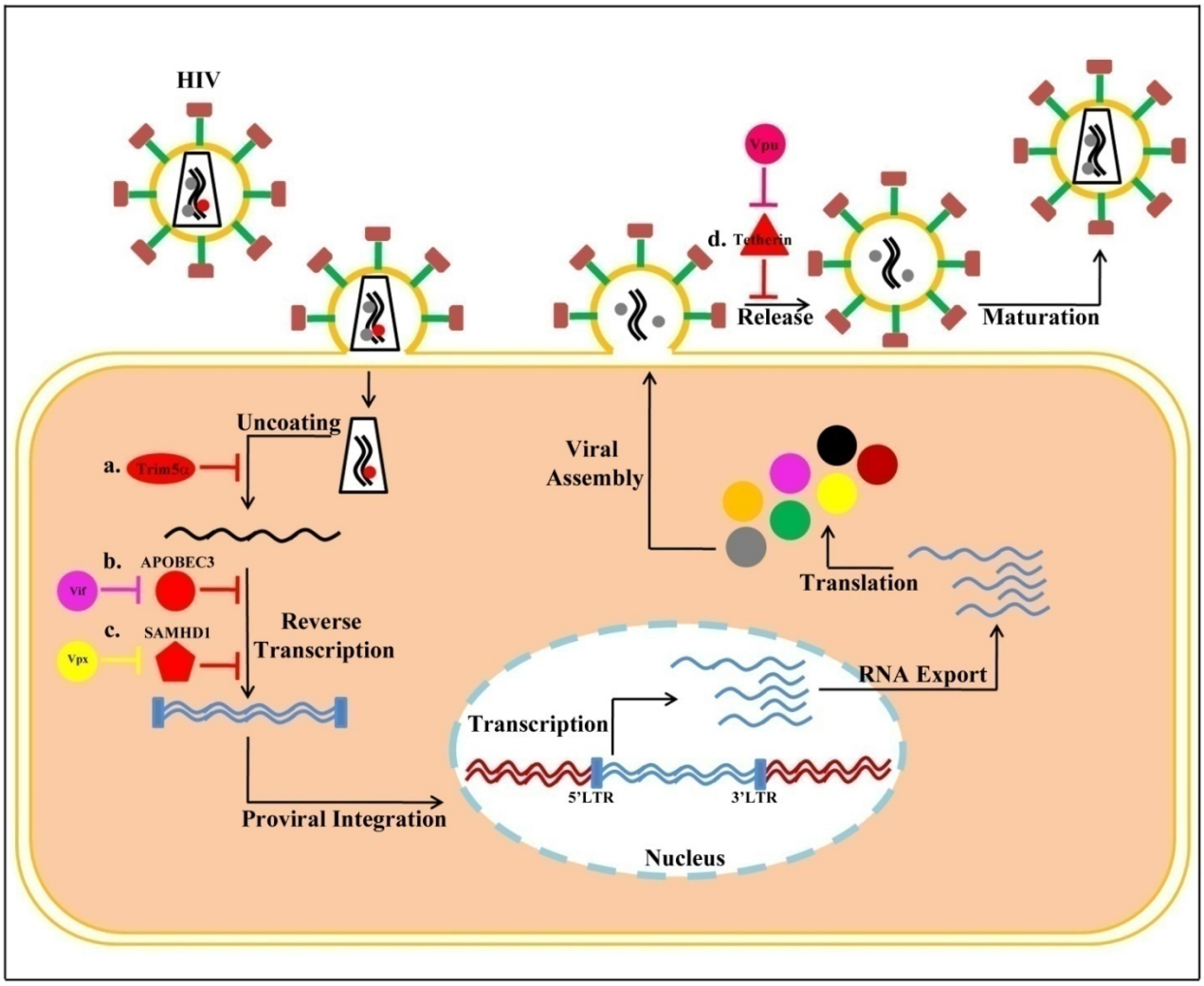

HIV
Vpu
d. Tetherin
Release
Maturation
Uncoating
a. Trim5α
Viral Assembly
b. APOBEC3
Vif
c. SAMHD1
Vpx
Reverse Transcription
Translation
RNA Export
Transcription
5'LTR
3'LTR
Proviral Integration
Nucleus

Figure 1.3 Host restriction factors against HIV and their viral antagonists. a. Trim5α acts immediately after viral entry into the cell by inducing the premature uncoating of the capsid therefore inhibiting reverse transcription. **b**. APOBEC3 or **c.** SAMHD1 affect the reverse transcription process by inducing hypermutations in the retroviral cDNA or depleting the intracellular pool of dNTPs respectively. Finally, **d**. tetherin acts at the end of the life cycle by preventing the release of virions from the cell and trapping them at the cell membrane. The viral proteins Vpx, Vif and Vpu antagonize the action of SAMHD1, APOBEC3 and tetherin respectively.

1.3 Retrotransposons

In addition to retroviruses, transposable elements (TEs) constituting approximately 45% of the human genome represent a supplementary threat that must be tightly regulated to avoid permanent damage to our genome (114). By definition, TEs are genomic DNA sequences that have the ability to move within the genome. Examples of devastating consequences related to the improperly controlled replication of these elements are well demonstrated through many genetic disorders including neurofibromatosis, choroideremia, Apert syndrome and cancer (115-119). There are three types of TEs in humans; DNA transposons, LTR retrotransposons and non-LTR retrotransposons. The first class, DNA transposons, represent close to 3% of the human genome. They encode a transposase inserted between inverted terminal repeats and required for excision of the element from the genome and insertion into a new location (120). There are no active forms of DNA transposons in humans. LTR retrotransposons account for 8% of our genome and are typical retroviruses that have become endogenized over time to become Human Endogenous Retroviruses (HERVs). These have accumulated inactivating mutations over the course of human evolution (121). Due to their defective genome, most human HERVs are replication incompetent. In contrast, non-LTR retrotransposons comprise the majority of currently active elements of the genome. This class can be further subdivided into two subtypes: the short interspersed nucleotide elements (SINEs) and the long interspersed nucleotide elements (LINES).

1.3.1 SINEs

This family includes the active Alu elements and the inactive SINE-VNTR-Alu (SVA) elements. Alu has a 300 base-paired (bp) genome comprised of two monomers originating from

the 7SL RNA gene separated by an Adenine(A)-rich linker region. Its transcription is under the control of an RNA polymerase III promoter located in the 5' region of the gene whereas a polyA tail is located at its 3' end. Since this element does not encode any protein, it is termed non-autonomous retrotransposon and relies mostly on the LINE-1 (L1) retrotransposition machinery for its replication (122). The full-length SVA genome is close to 2 kb and contains a hexamer repeat region followed by an Alu-like region, several tandem repeats regions whose number can vary depending on the element, a HERV-K10-like region and an oligo dA-rich tail at its 3' end (123). The replication of SVA is also dependent on L1 and since no promoter region is present in its genome, its transcription is believed to initiate from flanking regions of the genome.

1.3.2 LINEs

In humans, LINEs account for 17% of the genome and include L1 and LINE-2 (L2) elements although only L1 leads to productive retrotransposition events and will be the focus of this section (114). The 6 kb genome of L1 consists of an untranslated region (UTR) at both the 5' and 3' ends with an internal RNA polymerase II promoter in the 5' region, two open reading frames (ORF1 and ORF2) and a polyA tail at the 3' end (124). The L1 element is termed autonomous because it encodes its own nucleic acid binding protein from its ORF1 as well as a protein with an RT and endonuclease activity from its ORF2. These properties make L1 the only autonomous element in humans. Although the human genome contains more than 500,000 copies of L1, only approximately 80-100 are replication competent due to deletion and recombination events occurring during the retrotransposition process (see section 1.3.3 for details on the retrotransposition process of L1) (114).

1.3.3 L1 replication cycle

The replication cycle of L1 retrotransposons is illustrated in **Figure 1.4**. The cycle begins in the nucleus where L1 DNA is transcribed by the RNA polymerase II to produce an mRNA transcript which is then processed and exported from the nuclear compartment. (124). Once in the cytoplasm, the ORF1 and ORF2 proteins are translated and, due to their strong *cis*-preference for L1 mRNA, associate with the latter to form a ribonucleoprotein (RNP) complex which is known as an intermediate in the replication process (125). ORF2 bound to the mRNA then returns to the nucleus where reverse transcription and integration occur through a dual mechanism called target-primed reverse transcription (TPRT) (**Figure 1.4**) (126).

This reaction initiates with the cleavage of the non-coding strand of target DNA at a T-rich site by the L1 endonuclease. This generates a 3' hydroxyl at the site of cleavage which is used as a primer to initiate the synthesis of cDNA from the L1 mRNA template by the L1 RT. The target DNA coding strand is in turn cleaved to prime the synthesis of the second-strand through an unclear mechanism. Finally, the process ends by the integration of L1 DNA in the genome with target site duplications (TSDs) on each side.

1.3.4 Impact of L1 replication on the human genome

The persistence of L1 retroelements in the human genome for several millions of years has certainly contributed to their marked impact on its composition and structural organization. Probably the most obvious consequence of retrotransposition is the significant increase in the size of the human genome (127). The replication rate of L1 has been estimated to one new genomic insertion every twenty to two-hundred births depending on the methodology used for the estimation (128, 129). Overall, this represents a contribution of approximately two thousand

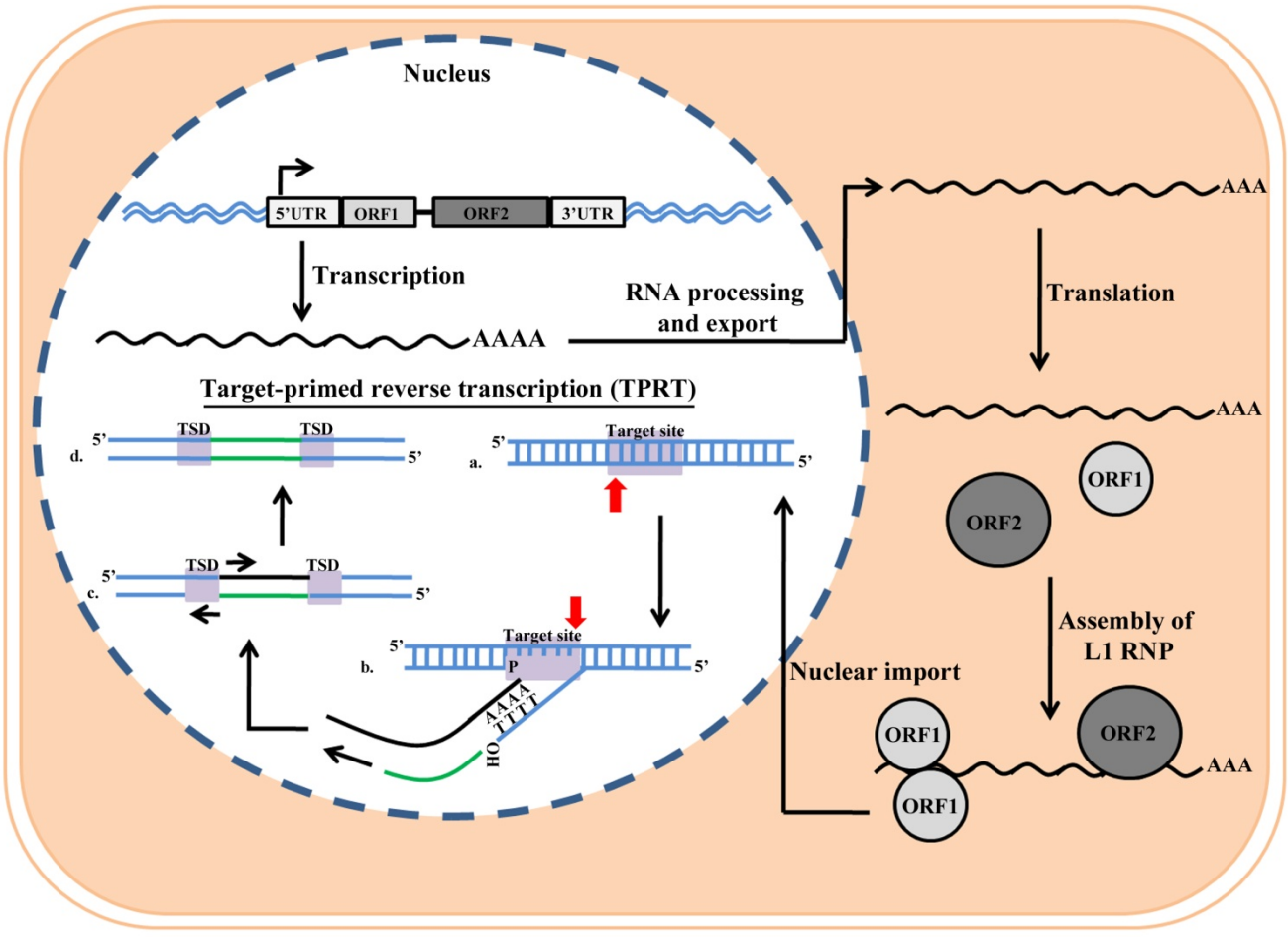

Nucleus
5'UTR
ORF1
ORF2
3'UTR
Transcription
AAAA
RNA processing and export
AAA
Translation
Target-primed reverse transcription (TPRT)
TSD
TSD
d.
a.
Target site
c.
b.
P
AAAA
TTTT
HO
ORF1
ORF2
Assembly of L1 RNP
Nuclear import
ORF1
ORF2
AAA
ORF1

Figure 1.4 Replication cycle of L1 retrotransposons. L1 DNA is transcribed from an RNA-polymerase II promoter region located in its 5' UTR (represented by an arrow) followed by processing and translocation of the mRNA in the cytoplasm. ORF1 and ORF2 are then translated and associate with L1 mRNA to form a ribonucleoprotein (RNP) complex which then gets access to the nucleus for reverse transcription and integration by target-primed reverse transcription (TPRT). **a**. L1 endonuclease (ORF2) cleaves the non-coding DNA strand in a T-rich region (red arrow). **b**. The 3' hydroxyl (OH) generated by the cleavage primes reverse transcription in which the RT (ORF2) uses L1 mRNA as a template for cDNA synthesis. The endonuclease cleaves the coding DNA strand to prime second strand DNA synthesis (inverted red arrow). **c**. DNA synthesis proceeds generating target site duplications (TSDs) on each side of L1 DNA. **d**. A copy of L1 is integrated into a new genomic location.

L1 sequences in the last six million years translating into a few hundred million bases added to the genome (114). Furthermore, evidence suggests that this genome expansion is an ongoing process because of retrotransposition events still taking place today. In addition, L1 copy number can vary between individuals and is often used as a genetic marker to study the history of a population (128, 130-132).

Retrotransposons represent an important source of genomic instability. One of the best characterized mechanism by which these pathogenic elements initiate that process is through their insertion into introns, which has been associated with several genetic disorders including heamophilia, cystic fibrosis, Apert syndrome and Duchenne muscular dystrophy to name a few (133-137). Genomic instability leading to cancer, namely breast and colon cancer, has also been documented (134-136). Furthermore, it has been suggested that DNA double-strand breaks (DSBs) generated by ORF2 in the absence of successful L1 integration could also represent a source of genomic instability (138). Those observations came however from experiments conducted in cell lines and the extent of this phenomenon *in vivo* is not well understood. Retrotransposition is also at the origin of several genomic rearrangement events including deletions and duplications many of which have been linked to disease in humans (139-143).

In addition to the impact of L1 replication on the genomic DNA, retrotransposons can affect gene expression at the RNA level. For example, L1 insertion can lead to alternative splicing through the generation of novel splice sites (144, 145). Retrotransposons also have the ability to influence transcription by many means. Firstly, their sequence is often a challenging substrate for the RNA-polymerase II therefore preventing transcription elongation of host genes upon insertion in their respective coding sequence (146). Similarly, the poly(A) tail located at the 3' end of the L1 sequence can induce transcriptional arrest (147, 148). Furthermore, the

promoter region of retroelements can induce sense or anti-sense transcription upstream or downstream of flanking genes (149). Finally, the methylation of the L1 promoter has the potential to initiate the formation of heterochromatin and repress the expression of proximal genes (150, 151).

1.4 APOBEC3 proteins

The seven members of the human A3 family play an essential role in defending humans against retroviruses and the expansion of retroelements (Reviewed in (83, 101)). These proteins are part of a larger family, the APOBEC family, that includes APOBEC1 (A1), activation-induced deaminase (AID), APOBEC2 (A2) and APOBEC4 (A4). APOBECs are all characterized by the presence of a conserved His-X-Glu-X_{25-31}-Pro-Cys-X_{2-4}-Cys motif required for the deamination process, where X can be any amino acid (99, 100, 152, 153). The histidine (His) and the two cysteines (Cys) are required for the coordination of a zinc molecule whereas the glutamic acid (Glu) is involved in proton shuttling following the nucleophilic attack by a water molecule on the carbon 4 amine group of the targeted cytidine.

A1 is the only member of the human APOBEC family to specifically act on RNA (154). This enzyme, expressed mostly in gastrointestinal cells, is involved in lipid metabolism. A1 deaminates the apolipoprotein B mRNA leading to the incorporation of an intergenic termination codon. This results in a full-length protein (ApoB100) and a truncated one (APoB48). Both proteins have distinct functions in the biology of cholesterol, triglycerides and lipids as they are expressed in different tissues (155, 156). AID is expressed specifically in activated germinal B cells and plays a crucial role in adaptive immunity by generating the diversified pool of

antibodies responsible for the humoral response (157, 158). The latter enzyme targets immunoglobulin genes for somatic hypermutation and class switch recombination.

Deaminase activity for A2 and A4 has yet to be demonstrated although they possess all the features necessary for catalysis (159). A2 is highly expressed in skeletal and cardiac muscle cells and recently, a role for this protein in muscle differentiation and early embryogenesis has been proposed (160-162). Nevertheless, deletion of this gene in mice is not lethal (162). Further studies are required to elucidate the exact physiological function of these proteins.

1.4.1 A3 domain architecture

A3 proteins possess either one (A3A, A3C and A3H) or two (A3B, A3D, A3F and A3G) zinc (Z) -coordinating domains each containing the characteristic APOBEC motif mentioned above (100, 163). Homology within this motif provides a second degree of classification for each domain into three sub-groups: Z1 (A3A, C-terminal domain (CTD) of A3B and A3G), Z2 (A3C, both domains of A3D and A3F and N-terminal domain (NTD) of A3G) and Z3 (A3H) (152).

In the case of the double-domain A3s, despite high levels of similarity at the nucleotide and amino acid level, each domain displays different functions. For most human A3 proteins with the exception of A3B in which both domains are active, the catalytic activity is exerted by the CTD exclusively (163-165). This same domain is also required for incorporation into virions, ssDNA binding as well as Vif interaction by A3F (166-170). The CTD of A3D is also responsible for its binding to Vif. (168, 170). In contrast, the packaging of A3G inside viral particles as well as nucleic acid binding by the protein rely on its NTD. (171-176). This domain

of A3G is also responsible for its interaction with Vif and dictates the species specificity of HIV-1 infection (177-180).

1.4.2 Deamination-dependent restriction of retroviruses by A3

It is now well established that members of the A3 family that most effectively act against HIV-1 are A3D, A3F, A3G and A3H (181). There is a link between the ability of these enzymes to assemble into multimeric complexes and their efficiency at restricting the virus (182). Failure of A3A and A3C to oligomerize correlates with their inability to prevent infection (182). A3B has been shown to restrict HIV-1 infection *in vitro*, however, this deaminase is not expressed in natural HIV-1 host target cells and as a consequence, does not encounter the virus under physiological conditions (183).

When antiretroviral A3 proteins are expressed in virus producing cells, in the absence of Vif, they get incorporated into emerging particles through their interaction with the NC region of the HIV-1 Gag polyprotein and viral and cellular RNAs (177, 184-199). The deaminases are then transported in the target cell and as part of the RTC they deaminate deoxycytidines into deoxyuridines in retroviral cDNA intermediates during reverse transcription leading to the incorporation of adenines in the opposite DNA strand (200-204). The net result is the accumulation of massive G-to-A mutations in the provirus, termed hypermutation. This leads to the generation of missense mutations and stop codons in newly synthesized viruses, ultimately inactivating the virus (**Figure 1.5**). An important feature of A3-mediated deamination of ssDNA is that the targeted cytosine is preferentially preceded by either a cytosine (A3G) or a thymine (other A3s) (163, 181, 205-208). Although the determinants for the substrate specificity for

deamination have not yet been fully resolved, several lines of evidence suggest the involvement of residues located in loop 7 of the CTD of A3 as well as AID (206, 209-213).

The catalytic activity of A3G is inhibited by the incorporation of the protein into RNA-dependent high molecular mass (HMM) complexes (214). This is thought to be a mechanism used by cells actively replicating to sequester the enzyme away from the nucleus. In dividing cells, A3G initially accumulates as low molecular mass (LMM) complexes, in which the enzyme is catalytically active and as it gets produced, gradually gets recruited into higher order complexes upon interaction with cellular RNAs and proteins (215-218). In those conditions, A3G fails to package into virions. It has been suggested that in addition to the proteasomal degradation of the deaminase, Vif induces the recruitment of A3G into HMM complexes to limit its activity (219). The nature of those complexes is poorly understood. The identification of numerous proteins involved in mRNA metabolism and the translocation of many components upon stress induction suggest that they might be RNA processing (P-) bodies, polysomes or stress granules (SG) (215-217).

1.4.3 Deamination-independent restriction by A3

The use of catalytically inactive mutants of A3 has led to the observation that A3G and A3F both have a strong ability to restrict retroviral infection in the absence of deamination *in vitro*, although the extent of this mode of restriction at physiological levels is still debated (220-225). Several mechanisms have been suggested to explain the deamination-independent restriction with the end result being reduced cDNA synthesis and proviral integration. It has been suggested that A3G interferes with tRNA primer annealing and reverse transcription initiation (226, 227). A direct interaction of both enzymes with the viral genomic RNA or with the viral

RT to inhibit the reverse transcription process has also been proposed (228-231). Similarly, A3 protein interaction with the viral IN could result in defects in the integration process (232). Although all the above-mentioned possibilities may not be exclusive, the exact mechanism remains to be elucidated.

Deamination-independent activities of A3 proteins extend to retrotransposons (215, 233-239). Nuclear localization of A3 proteins is not important for this restriction nor is their enzymatic activity. It is rather presumed to involve an interaction with ORF1 of L1 or the sequestration of the element from the nucleus into oligomers (215, 235). The recent identification of G-to-A mutations in L1 DNA induced by A3A suggests however, that this specific A3 family member could use a different mode of action (240). It is clear from those findings that work remains to be done in order to fully understand the restriction of retrotransposons by A3 proteins.

1.4.4 The duality of A3-mediated mutagenesis

As mentioned above, the viral protein Vif induces the degradation of A3 proteins to prevent virion incorporation and inhibit the activity of the enzymes in the target cell (see section 1.2.2 for details). Moreover, the catalytic activity of A3G in dividing cells is tightly regulated through its recruitment into HMM complexes to avoid mutagenesis of the host genomic DNA upon rupture of the nuclear membrane during cell division (see section 1.4.2 for details). Consequently, physiologically relevant low levels of A3 proteins are available for viral encapsidation at the moment of infection reducing the outcome of their mutagenic potential in the target cell. As opposed to highly mutated viral genomes, it is becoming more clear now that low levels of mutation induced by the A3 proteins can be beneficial for the virus (241-246).

A good example of the generation of more fit viruses following deamination by the A3 proteins is the emergence of drug resistance mutations. More than one hundred mutations conferring resistance to popular HIV-1 treatments have been documented with several being in an A3G hotspot (247, 248). Variant viruses resistant to 3TC (M184I), a powerful nucleoside analog RT inhibitor, have been demonstrated to be produced by A3G even in the absence of the drug (243, 249). The same mutation was also identified in peripheral blood mononuclear cells (PBMCs) from HIV-1 infected patients together with other drug resistance mutations in an A3G-specific context (250). Furthermore, viral evolution and diversification as a result of mutations induced by A3D and A3F in humanized mice leading to drug resistance and co-receptor change has recently been documented (246).

During the first weeks of HIV-1 infection, viremia is controlled by the adaptive immune system directing cytotoxic T cells (CTLs) to kill the virus (251-253). Evasion from the CTL response is therefore highly appealing for the virus as a determinant of its fate. This can be achieved by modifying viral epitopes so they can no longer be recognized by CTL upon presentation on the major histocompatibility complex (MHC) class I (254-256). In favour of a role for A3G in increasing viral fitness, it has been shown that several mutations responsible for the lack of an appropriate CTL response occur in a 5'CC target motif (257). In addition, *ex vivo* experiments were performed where the activation of $CD8^+$ T cells from HIV-1 patients was measured in response to wild type HIV epitopes or mutants in which substitutions were introduced at A3G and A3F target sites to mimic deamination by the A3 proteins. These assays revealed a decreased CTL response upon A3-mediated mutagenesis (258). In agreements with these findings, mutations in an A3-specific context have recently been observed within human

leukocyte antigen (HLA)-presented epitopes involved in CD8+ T cell-mediated response from HIV-infected individuals (243).

In contrast to A3G which is exclusively expressed in the cytoplasm sequestered away from the genetic material, some A3 proteins, notably A3A, A3B and A3C shuttle to the nucleus at varying degrees. The nuclear localization of these enzymes increases the chances of causing damage to the cell's genome. With the use of next generation sequencing technologies, two different groups were able to observe clusters of mutations in head and neck squamous cell carcinoma (HNSC), multiple myeloma, prostate and breast cancer in a specific TC consensus that they attributed to members of the A3 family (259, 260). Not long after, a correlation between A3B expression and mutation of cytosines in breast tumors was established (261). Following those initial discoveries, several researchers deployed efforts to establish a link between A3-mediated deamination of the host DNA and many types of cancer shedding light on a previously unknown harmful role for the deaminases (262-264).

1.4.5 The regulation of A3G in dividing cells

Members of the A3 family are constitutively expressed across a large variety of cells and tissues (183). Their expression is induced in response to interferons (IFNs) and some cytokines such as IL-2, IL-7 and IL-15 (183, 265). However, under these conditions, the activity of A3G, the most potent member against HIV-1, is inhibited through its transition from small 150 kDa LMM complexes into HMM complexes of up to 5-15 MDa. The same phenomenon is also observed in lymphoid CD4+ T cells where there is a constant endogenous production of cytokines (266). A3G proteins in these cells therefore accumulate into a HMM inactive form. On the other hand, resting CD4+ T cells from peripheral blood contain A3G in its LMM active form.

As mentioned above, the identity of the regulatory elements causing the formation of HMM complexes by A3G remains unsolved. Initial studies aimed at describing the components of the complexes identified a majority of mRNA-interacting proteins involved in different steps of mRNA biogenesis (216, 217). Amongst these were heterogenous ribonucleoproteins (hnRNPs), RNA helicases, translation initiation factors and ribosomal proteins. Most proteins interact with A3G indirectly through RNA-bridging. Several RNAs were also found inside HMM complexes notably many small non-coding RNAs such as 7SL and Y RNAs as well as Alu and A3G mRNAs. (215)

The recruitment of A3G into HMM complexes represents a post-translational regulation mechanism used by the cell to protect its genomic material against the mutagenic potential of the enzyme. However, the oligomeric form of the protein strongly influences the permissiveness of a cell to HIV-1 infection. Whether the existing forms of A3G can be modulated in HIV-1 target cells to promote efficient protection should be investigated.

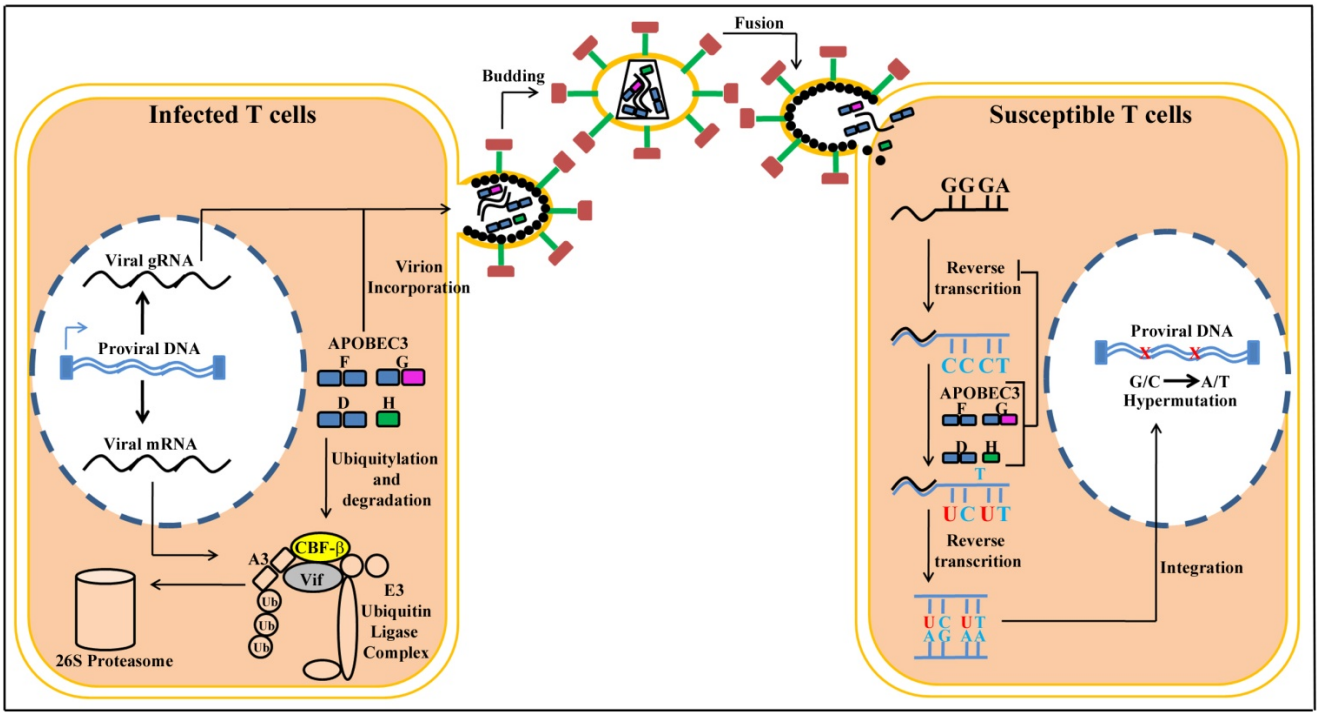

Infected T cells
Budding
Fusion
Susceptible T cells
Viral gRNA
Proviral DNA
Viral mRNA
Virion
Incorporation
APOBEC3
F
G
D
H
Ubiquitylation
and
degradation
A3
CBF-β
Vif
Ub
Ub
Ub
E3
Ubiquitin
Ligase
Complex
26S Proteasome
GGGA
Reverse
transcrition
CCCT
APOBEC3
F
G
D
H
T
UCUT
Reverse
transcrition
UC UT
AG AA
Proviral DNA
G/C → A/T
Hypermutation
Integration

Figure 1.5 Mechanism of retroviral restriction used by the A3 proteins. When expressed in viral producer cells, A3 proteins get packaged into budding particles and carried over to the target cell. Once in the target cell, they act as deoxycytidine deaminases and deaminate deoxycytidines into deoxyuridines in the retroviral cDNA during reverse transcription. The enzymatic activity of the proteins leads to the incorporation of G-to-A hypermutation in the progeny viruses and inactivation of infection. The A3 proteins also have the ability to hinder reverse transcription independently of their catalytic activity however the mechanisms are not completely understood.

1.5 Rationale, hypothesis and objectives

Rationale

HIV represents one of the most important pandemics the world has known with more than 35 million reported individuals living with HIV/AIDS in 2013 (WHO). Despite the deployment of vast resources, no effective vaccine or cure has yet been discovered that can eradicate this infection. Several HIV-1host-encoded restriction factors have been identified in recent years, but still little is known about how they work and how they are regulated. Most notably, A3G is only active when in its LMM monomeric form. Understanding how this protein is regulated and shifts between active LMM and inactive HMM states in dividing CD4+ T cells could allow the development of new anti-HIV strategies.

Hypothesis

Enforcing A3G proteins to remain in a highly active LMM state would increase their antiretroviral potency against HIV-1 and reduce the likelihood of generating sub-lethal mutagenesis that ultimately favours the virus.

Main Objective

The main objective of my thesis is to identify the regulatory elements of the A3G protein involved in modulating its active and inactive states in view of potentiating its antiretroviral activity against HIV-1.

Specific aims

First aim:

To perform a detailed structural characterization of A3G by mapping and mutating the protein domains responsible for its assembly into HMM complexes.

Second aim:

To determine whether the structural determinants that enable restriction of HIV-1 infection also inhibit LINE-1 retrotransposon activity.

Third aim:

To determine whether there is a correlation between the DNA substrate specificity for deamination by A3G and A3F and their gene-inactivating potency.

Fourth aim:

To develop a high throughput tool to identify and quantify mutations generated by A3 proteins in retroviral provirus DNA.

Chapter 2 Materials and Methods

2.1 Cell culture

Human embryonic kidney (HEK) 293T cells and NIH 3T3 mouse fibroblasts were cultured in DMEM/High Glucose medium (Multicell, Wisent Inc., St-Bruno, QC, Canada) supplemented with 10% fetal bovine serum (FBS) (CellGro, MediaTech Inc., Manassas, Virginia, USA), 1000 units (U)/mL penicillin and 100 micrograms (µg)/mL streptomycin (Wisent Inc., St-Bruno, QC, Canada). Human $CD4^+$ T lymphocytes (CEM-SS) were cultured in RPMI medium (Multicell, Wisent Inc., St-Bruno, QC, Canada) supplemented with 10% FBS (CellGro), 1000 U/mL penicillin and 100 µg/mL streptomycin (Wisent). Adherent cells were washed with 1X phosphate buffer saline (1X PBS) every 48 h, harvested with 0.05% trypsin-EDTA and put back in culture at a 1:10 dilution in DMEM. Suspension cells were collected and diluted 1:10 in fresh medium every 48 h.

2.2 Plasmids

All polymerase chain reactions (PCR) were performed in a final volume of 50 microliters (µL) containing 1x PfuTurbo buffer (Agilent Technologies, Santa Clara, California, USA), 10 nanograms (ng) template DNA, 200 nanomolar (nM) of each primer, 200 micromolar (µM) of each dNTP and 2.5 U of PfuTurbo (Agilent Technologies, Santa Clara, California, USA). The cycling conditions were as followed; initial denaturation at 94^{o}C for 2 minutes (min) followed by 30 cycles of denaturation at 94^{o}C for 45 seconds (sec), annealing at 55^{o}C for 45 sec and elongation at 72^{o}C for 1 min per 1 kb to be amplified, followed by a final elongation step of 10 min at 72 ^{o}C. All PCR reactions were carried out in an Eppendorf mastercycler Pro S series thermocycler (Eppendorf, Hamburg, Germany).

2.2.1 APOBEC expression plasmids

2.2.1.1 Flag and eGFP expression plasmids

Flag- and eGFP-tagged A2, A3C, A3F and A3G were previously described (206). All mutants of the proteins were generated by site-directed mutagenesis (SDM) using the Quick Change XL Site-directed mutagenesis kit (Agilent Technologies, Santa Clara, California, USA) following manufacturer's instructions and primers listed in **Table S1**. A3A, A3B and A3H mRNA was previously amplified from PBMCs and cloned in peGFP-C3 (Invitrogen Corp., Carlsbad, California, USA). To generate Flag-tagged A3A and A3H, the coding DNA sequence (CDS) of both proteins was amplified by PCR using a forward primer in which the Flag peptide sequence preceded by an NheI restriction site and a kozak sequence was embedded (Flag-A3A-FWD and Flag-A3H-FWD) and a reverse primer containing an HpaI site (A3-HpaI-REV) (**Table S1**). Amplicons were then inserted in the pcDNA3.1 vector (Invitrogen Corp., Carlsbad, California, USA) previously digested with NheI and HpaI (New England Biolabs (NEB), Ipswich, Massachissets, USA). The restrictive and stably expressed haplotype II of A3H (G105R) was generated by SDM as described above using primers listed in **Table S1** (267). In the case of A3B, the CDS was digested out from peGFP-C3 using NheI and HpaI (NEB) and inserted in the respective sites in a pFlag-C3 plasmid. A3D was provided by the NIH AIDS Research Reference and Reagent Program (pcDNA3.1-APOBEC3DE-V5-6xHIS, Cat. # 11433) in a pcDNA3.1 backbone. The CDS was then amplified by PCR using forward and reverse primers containing XhoI and SacII sites respectively (A3D-XhoI-FWD and A3D-SacII-REV) (**Table S1**) and the PCR product was inserted in the XhoI and SacII sites of pFlag-C3.

To generate Vpr-fusion proteins, amino acids 14 to 88 (Vpr^{14-88}) or 14 to 86 (Vpr^{14-86}) of Vpr were amplified from the HIV-1 pNL4.3 plasmid (NIH AIDS Research Reference and

Reagent Program, catalog # 114) using forward and reverse primers containing an XhoI restriction site (Vpr-XhoI-FWD and Vpr^{14-88}-XhoI-REV or Vpr^{14-86}-XhoI-REV) (**Table S1**). The amplicon was then digested with the XhoI endonuclease (NEB, Ipswich, Massachussets, USA) for 3 h at 37 ºC and ligated in frame into the XhoI site of Flag-tagged APOBECs previously digested with the same enzyme.

2.2.1.2 GST expression plasmids

A codon optimized (*Escherichia (E.) coli*) version of the A3G[W127A] mutant fused to Gluthatione-S-transferase (GST) was purchased by Dr. Jean-François Couture from the University of Ottawa who kindly provided us with an aliquot. The wild type A3G and the W94A mutant were generated from this plasmid by SDM as described earlier with primers listed in **Table S1** (A127W-GST-FWD and A127W-GST-REV and W94A-GST-FWD and W94A-GST-REV).

2.2.1.3 HA-expression plasmids

HA-expression plasmids were generated by amplification of the coding sequence of Flag-tagged proteins using a forward primer with the HA tag preceded by the NheI restriction endonuclease site embedded in its sequence (HA-NheI-FWD) and a reverse primer containing an HpaI restriction site (A3-HpaI-REV) (**Table S1**). PCR products were then digested with NheI and HpaI (NEB) for 3 h at 37 ºC, purified using the Wizard (R) SV Gel and PCR clean-up system (Promega, Madison, Wisconsin, USA) and inserted in the NheI and HpaI site of Flag-C3 previously digested with the same endonucleases and purified.

2.2.2 Viral vectors

The retroviral vectors used in this thesis have been previously described (206, 268, 269). The pMOV-eGFP expression plasmid encodes a replicative M-MLV containing an eGFP reporter gene inserted in the proline-rich region of the ecotropic Env gene (268). HIV[p8.9] self-inactivating pseudoviruses were produced by co-transfection of three expression vectors; a packaging vector for HIV-1 (pCSGW), an expression vector for Gag-Pol-Rev (p8.9) and the pDMG plasmid encoding the envelope glycoprotein of the vesicular stomatitis virus (VSV-G). The eGFP reporter gene of HIV[p8.9] is located in the CSGW vector and its expression is regulated by an internal spleen focus-forming virus (SFFV) long terminal repeat (LTR) promoter (206, 269). HIVΔVif and HIVΔVif-eGFP (pNL4-3 and pNL4-3-deltaE-eGFP) were both gratefully provided by the NIH AIDS Research Reference and Reagent Program (catalog #114 and 11100 respectively) (270). In the case of HIVΔVif-eGFP, the Env gene was replaced by an eGFP reporter gene. Upon arrival, the Vif gene from both viral vectors was inactivated through the incorporation of two consecutive termination codons by SDM as previously described (271).

2.2.3 Retrotransposons

Plasmids containing a full-length copy of the L1 (L1.3) element or the L1 element with two missense mutations in the ORF1 coding region to abolish retrotransposition (L1 mutant) and an eGFP indicator cassette were gratefully provided by Dr. Derrick Gibbings from the University of Ottawa. Both constructs have been previously described (272).

2.3 Transfection and infection

2.3.1 Transfection using GeneJuice

293T cells were seeded at 2 X 10^5 in 2 mL of DMEM in a 6-well plate or 1.5 X 10^5 in 1 mL of DMEM in a 12-well plate the day before transfection. Prior to transfection, the media was replaced with fresh DMEM. Transfections using GeneJuice transfection agent (Novagen-EMD Millipore Corp., Billerica, Massachussets, USA) were performed according to manufacturer's instructions.

2.3.2 Transfection using $CaCl_2$

293T cells were seeded at 2 X 10^6 in a 10 cm dish 24 h prior to transfection in 12 mL of DMEM. The day of transfection, media was replaced by 12 mL of fresh DMEM. The reaction was started by mixing 375 µL of water, 125 µL of $CaCl_2$ 1 molar (M) and the desired amount of DNA at a maximum of 20 µg and gently tapping. 500 µL of 2X HBS (50 mM HEPES, 10 mM KCl, 12 mM Dextrose (D-glucose), 280 mM NaCl, 1,5 mM Na_2HPO_4) pH 7.05 was then added dropwise to the mixture while vortexing. Following addition of the HBS, samples were vortexed thoroughly for 30 sec and incubated at room temperature for 20 min after what the reaction was added dropwise to the cells.

2.3.3 Infection

Viruses were produced in the presence of APOBECs by transfection with GeneJuice (Novagen-EMD Millipore) following procedures in 2.3.1 48 h prior to infection. For production of HIV[p8.9], 300 ng of pCSGW, 250 ng of p8.9 and 150 ng of VSV-G were co-transfected with 150 ng of Flag- or eGFP-tagged APOBECs with the exception of A3G[W94A/W127A] for which 1 µg was used for transfection. To produce M-MLV, 800 ng of the viral genome-encoding plasmid was co-transfected with 80 ng of Flag-APOBECs and finally, 500 ng of HIVΔVif or

HIVΔVif-eGFP were co-transfected together with 200 ng of VSV-G and 150 ng of Flag-APOBEC expression plasmids. One day before infection, 293T or 3T3 cells were seeded at 3 X 10^5 in a 6-well plate in 2 mL of DMEM. The day of infection, prior to infecting, DMEM was replaced with 3 mL of fresh DMEM containing Polybrene (Sigma-Aldrich, St. Louis, Missouri, USA) at a final concentration of 8 µg/mL. The supernatant from virus-producer cells were collected, filtered through 0.45 micrometer (µm) cartridges (EMD Millipore Corp., Billerica, Massachussets, USA) and centrifuged at 2 000 rpm for 10 min. When applicable, p30 capsid protein concentration was determined in each sample by enzyme-linked immunosorbent assay (ELISA) using the Quick Titer ^{TM} MuLV core Antigen ELISA kit (MuLV p30) (Cell Biolabs, San Diego California, USA). Following the spin, equal volume of cleared supernatants or equal amounts of viral particles as determined by ELISA were added dropwise to 293T or 3T3 target cells which were then spinoculated 1 h at 2 000 rpm. Infected cells were cultured at 37 ºC for 24 or 48 h after which infection levels were measured by flow cytometry analysis of eGFP fluorescence using the Cyan ADP flow cytometer (Beckman Coulter Inc., Brea, California, USA).

2.4 Velocity sedimentation on sucrose gradient

Cells expressing APOBEC-expression plasmids (1 to 20 µg) (section 2.3) were lysed with 800 µL of NP40 lysis buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.1% NP40, 0.1% Na-deoxycholate) supplemented with cOmplete, EDTA-free, protease inhibitor cocktail (Roche, Basel, Switzerland) for 30 min on ice prior to centrifugation at 17 000 x g for 5 min at 4 ºC to clear insoluble material. After the spin, half of the samples was treated with 1 ug/mL of RNAse A (Roche, Basel, Switzerland) for 15 min at room temperature after which all samples were

loaded on top of a 5 to 40 % sucrose gradient. The gradients were previously prepared by addition of 2.3 mL of each sucrose solution starting with 40% then 30, 20, 10 and 5% sucrose diluted in sucrose gradient buffer (10 mM Tris-HCl (pH 7.4), 25 mM KCl, 10 mM $MgCl_2$) in 13.2 mL tubes for centrifugation (Beckman Coulter Inc., Brea, California, USA) and snap-freezing in liquid nitrogen between each layer to avoid mixing. Following addition of the treated or non-treated samples on top of the 5% sucrose layer, gradients were spun in a Beckman SW41Ti rotor in an Optima^{TM} L-100 XP ultra-centrifuge (Beckman Coulter Inc., Brea, California, USA) for 6 h at 41 000 rpm at 4 ºC. Aliquots of 700 µL were then collected from the top to the bottom of each tube for a total of 9 eppendorf tubes per gradient each containing 1.4 mL which were kept on ice. 75 µL of each eppendorf was then mixed with 25 µL of 5X Laemmli loading buffer (0.5M Tris pH 6.8, 0.2% (v/v) glycerol, 20% (w/v) SDS, 0.16% (w/v)Bromophenol blue, 5% β-mercaptoethanol), boiled 5 min, incubated on ice for 2 min and stored at -20 ºC prior to Western blot analysis (section 2.9).

2.5 Fluorescence microscopy

HEK 293T cells were seeded at 2 x 10^3 cell/mL in DMEM without phenol red in 35 mm glass bottom dishes (MatTek Corp., Ashland, Massachussets, USA) one day before transfection of eGFP-APOBECs (100 ng to 1 µg) following procedures in 2.3.1. Cells were washed with 1X PBS the day after transfection and grown for an extra 24 h in fresh phenol red-free DMEM. At day 3 after transfection, imaging of live cells was performed using a Zeiss Axio Observer.Z1 inverted fluorescent microscope and Plan-Apochromat 63x/1.4 oil immersion objective (Zeiss, Jena, Germany). Images, acquired as Z-stacks, were deconvolved following the constrained

Iterative Maximum Likelihood Algorithm (http://www.zeiss.ca/) and using the AxioVision software (Zeiss, Jena, Germany).

2.6 Immunoprecipitation

2.6.1 Agarose Beads

For immunoprecipitation (IP) of Flag-tagged proteins, anti-Flag-conjugated agarose beads (Sigma-Aldrich, St. Louis, Missouri, USA) were used. HEK 293T transiently expressing Flag-tagged APOBECs (**Table 2.1**) were lysed in 800 µL of a soft lysis buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Triton 100-X) supplemented with a cOmplete, EDTA-free, protease inhibitor cocktail (Roche) on ice for 30 min prior to ultra-centrifugation at 17 000 x g for 5 min at 4 ºC to clear nuclei. Samples were then transferred in fresh eppendorfs and an aliquot of 75 µL was mixed with 25 µL of 5X Laemmli loading buffer, boiled 5 min, cooled on ice 2 min and kept at -20 ºC for analysis of protein expression in cell lysates by Western blotting. The remaining of the samples was incubated with 80 µL of a mixture of beads and 1X PBS (40 µL of each) and rotated at 4 ºC for 3 h. Prior to incubation with the samples, beads were thawed, gently mixed by inversion and 40 µL of the gelatinous solution per sample was collected in an eppendorf. Two washes with 500 µL of cold 1X PBS were then performed by spinning the beads at 8 200 x g at 4 ºC for 30 sec between each wash followed by resuspension in 1 mL of 1% BSA in 1X PBS containing 500 µg/µL of salmon DNA (Sigma-Aldrich, St. Louis, Missouri, USA) and incubation on a rotator at 4 ºC for 1 h to coat the beads to prevent unspecific binding of proteins or nucleic acids. After 1 h, the beads were washed two more times with cold 1X PBS and resuspended in an equal volume of 1X PBS for distribution to the samples. Following the 3 h period, samples were extensively washed four times with 200 µL of wash buffer 1 (50mM Tris-

HCl (pH 8.0), 150mM NaCl, 1% Igepal CA-630, 0.5% sodium deoxycholate, 0.1% SDS) intercalated by centrifugation at 8 200 x g for 30 sec at 4 ºC, once with the same buffer containing only 50 mM of NaCl and finally one last time with wash buffer 2 (20 mM Tris-HCl (pH 7.4)) prior to the elution step or papain protease treatment (Sigma-Aldrich, St. Louis, Missouri, USA).

Elution was performed by incubating the beads with 50 µL of 0.1 M triethylamine (pH 11.8) for 5 min and gently tapping to allow bound complexes to detach from anti-Flag antibodies. Eluted material was then collected in fresh eppendorfs containing 3 µL of 1 M MOPS (pH 3.0). Elution step was repeated twice for a final eluate volume of 106 µL. Samples were then used for Western blotting (section 2.9) or RNA extraction (section 2.7). For papain treatment (Sigma), aliquots of 5 µL of beads were mixed with 15 µL of 5X Laemmli loading buffer (without bromophenol blue) and analyzed by Western blotting using a fluorescent anti-Flag antibody (section 2.9) prior to enzymatic treatment to achieve approximate quantification of the APOBECs in each sample. Details on the detection and quantification methods can be found in 2.9.3. Remaining samples (approximately 30 µL) were then incubated 1 h at room temperature with 10 µL (>2 units diluted in 50 mM Tris-HCl pH 8.0) of papain on a rotator followed by centrifugation at 8 200 x g for 30 sec and collection of the supernatant for RNA extraction (section 2.7).

2.6.2 Magnetic Beads

For IP of HA- or eGFP-tagged proteins, anti-HA or anti-eGFP magnetic beads (Miltenyi Biotec, Bergisch, Gladbach, Germany) were used. Lysates from 293T cells expressing HA- or eGFP-tagged APOBECs (500 ng to 5 µg) were prepared as described in section 2.6.1 and used for IP following instructions provided with the uMACS^{TM} Epitope Tag Protein Isolation kit (Miltenyi

Biotec, Bergisch, Gladbach, Germany). For IP of HA-tagged proteins from virions, viral particles containing the APOBECs were purified by ultra-centrifugation as depicted in 2.10 and lysed in 800 µL of soft lysis buffer as in 2.6.1 prior to IP. Samples were then used for Western blot analysis (section 2.9).

2.7 RNA extraction, RT-PCR and qPCR

2.7.1 RNA extraction

RNA from immunoprecipitates was isolated using Trizol reagents (Sigma-Aldrich, St. Louis, Missouri, USA). To ensure the purity of the samples, RNA was treated with 2 units of DNase I (NEB, Ipswich, Massachussets, USA) for 10 min at 37 ºC followed by heat inactivation 10 min at 75 ºC in the presence of 0.5 M EDTA to conserve the integrity of the RNA.

A transcribed eGFP gene from an upstream T7 promoter in the linearised pcDNA3.1 vector (Invitrogen) using the MAXI® SP6/T7 In Vitro transcription kit (Ambion Inc., Austin, Texas, USA) was used to control the consistency of the extraction.

2.7.2 Reverse transcription PCR (RT-PCR)

Following the DNase I treatment, 4 µL of each sample was used for reverse transcription to generate cDNA using either random hexamers and the ImProm-II™ reverse transcription system (Promega, Madison, Wisconsin, USA) or the iScript reverse transcription supermix (Bio-RAD, Hercules, California, USA) according to manufacturer's recommendations.

2.7.3 Quantitative real-time PCR (qPCR)

Methods for quantification of RNA binding to the APOBECs was adapted from a previous report (172). A dilution of 1:100 was performed on the cDNA obtained after the RT-PCR and 8 µL of that dilution was used in each reaction for the qPCR in addition to 0.7 µL of each primer at 10

μM, 10 μL of the GoTaq® Green Master Mix (Promega, Madison, Wisconsin, USA) or the SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, Hercules, California, USA) and the necessary amount of water to achieve a final volume of 20 μL. In the case of the Promega master mix, the reference dye (ROX) was added separately at a volume of 0.2 μL per reaction whereas in the Bio-Rad kit, ROX was already provided in the supermix. The cycling conditions for the PCR were as followed; 50ºC for 2 min and 95ºC for 10 min, followed by 40 cycles of 95ºC for 15 sec and 60ºC for 1 min on an ABI ViiA7 System (Life Technologies, Carlsbad, California, USA). Primer sets against Alu, 7SL, hY1, hY3 and eGFP (Alu-FWD, Alu-REV, 7SL-FWD, 7SL-REV, hY1-FWD, hY1-REV, hY3-FWD, hY3-REV, eGFP-100-FWD and eGFP-HRM-REV) were used and are presented in **Table S1**. Relative quantities based on the cycle threshold of each sample were determined using A3G as the comparative reference using the ViiA7 software (Life Technologies, Carlsbad, California, USA).

2.8 Digital droplet PCR (ddPCR)

ddPCR was performed on cDNA obtained as described in sections 2.7.1 and 2.7.2 following a 1:2 dilution based on specialists' recommendations. Primers were the same as the ones used for qPCR analysis and are detailed in **Table S1**. For droplet generation, 5 μL of the dilution was mixed with a forward and reverse primer against the gene of interest at a final concentration of 100 nanomolar (nM) (0.2 μL of each primer per reaction), 10 μL of the EvaGreen Master Mix (Bio-Rad, Hercules, California, USA) and 4.6 μL of water to a final volume of 20μL which was then transferred to a cartridge with 70 μL of droplet generation oil (Bio-Rad, Hercules, California, USA) and placed in the QX200™ Droplet Generator (Bio-Rad, Hercules, California, USA). By carefully angling the pipette to prevent any damage to the droplets while pipetting, 40

µL of droplets (approximately 20 000 droplets) were transferred into a 96-well plate (Eppendorf, Hamburg, Germany) for PCR amplification of the target. The cycling conditions were as followed; 95 ºC for 5 min followed by 45 cycles at 95 ºC for 30 sec and 59.4 ºC for 1 min after which final steps at 4 ºC for 5 min and 90 ºC for 5 min were conducted in a C1000 Touch™ Thermal Cycler (Bio-Rad, Hercules, California, USA). Following the run, the plate was transferred to a QX200 Droplet Reader (Bio-Rad, Hercules, California, USA) to determine the number of positive and negative droplets based on their acquired fluorescence. The number of copies per µL of each target in all samples was then determined using the QuantaSoft™ software (Bio-Rad, Hercules, California, USA) provided with the instruments.

2.9 Western blotting

2.9.1 Manipulations

Cells expressing the proteins of interest were lysed 48 h following transfection in 800 µL of NP40 lysis buffer complemented with cOmplete, EDTA-free protease inhibitor cocktail (Roche) for 30 min on ice after which the samples were spun down for 5 min at 17 000 x g at 4 ºC to remove insoluble material. Lysates were then transferred to fresh eppendorfs and an aliquot of 75 µL was mixed with 25 µL of 5X Laemmli loading buffer whereas the remaining of the samples was stored at -80 ºC for further use. Aliquots were boiled 5 min at 95 ºC, incubated on ice for 2 min and loaded on a 10% SDS-PAGE gel for run 1 h at 175 V in 1X running buffer (190mM glycine, 0.1% (w/v) SDS, 25mM Tris-base). After the run, proteins were transferred to 0.45 µm PVDF membranes (EMD Millipore, Billerica, Massachussets, USA) for 1 h at 35 mA in 1X Novex transfer buffer (190mM glycine, 25mM Tris-base, 20% (v/v) methanol) followed by two quick rinses of the membranes in 1X PBS and incubation for 1 h in 5% milk in 1X PBS 0.1%

Tween 20 (PBS-T) (Acros, New Jersey, USA) on an orbital shaker. Incubation with milk was followed by rinsing in PBS-T and 1 h blotting with the primary antibody diluted in 5% milk at the indicated ratio (2.9.2). Membranes were then washed three times 10 min with PBS-T and finally quickly rinsed in 1X PBS prior to detection or if necessary, incubation 1 h with the secondary antibody diluted in 5% milk at the indicated ratio (2.9.2). Incubation with the secondary antibody was ended by three washes with PBS-T and one rinse with 1X PBS followed by detection. Detection was achieved using Clarity™ Western ECL Substrate (Bio-Rad, Hercules, California, USA) according to the instructions in the kit and either Clear Blue CL-XPosure imaging films (Thermo Scientific, Waltham, Massachussets, USA) or the ImageQuant LAS 4000 system (GE Healthcare Life Sciences, Little Chalfont, UK).

For processing of samples prior to fluorescence probing with the FITC-conjugated anti-Flag antibody, following the run, transfer was performed at 4 ºC for 1 h at 0.35 mA in chilled Towbin transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, pH 8.3) on low fluorescence PVDF membranes (Bio-Rad, Hercules, California, USA) to limit background contaminants upon detection. Following steps were as described above with the use of a reduced concentration of Tween in the 1X PBS (0.05%) and detection was achieved according to section 2.9.3.

2.9.2 Antibodies

Antibodies used for the achievement of the work described in this thesis were as followed: 1:3000 HRP-conjugated anti-Flag (A8592, Sigma-Aldrich, St. Louis, Missouri, USA), FITC-conjugated anti-Flag (F4049, Sigma-Aldrich, St. Louis, Missouri, USA), 1:1000 anti-p24 (ab9069, Abcam, Cambridge, England, UK), 1:3000 anti-eGFP (no.632381, Clontech, Mountain View, California, USA), 1:5000 HRP-conjugated anti-β-tubulin (ab21058, Abcam, Cambridge,

England, UK), 1:3000 anti-mouse (no. 7076S, Cell Signaling, Danvers, Massachussets, USA) and 1:5000 anti-HA (H3663, Sigma-Aldrich, St. Louis, Missouri, USA). The anti-p30 (1:500), anti-rabbit (1:10 000), anti-A3G (1:500) and anti-rat (1:10 000) antiserum were previously produced by Dr. Marc-André Langlois. The anti-Catalase subunit α (Cat-α) (1:1000) was gratefully provided by Dr. Stinzi from the University of Ottawa and the anti-HIV-1 integrase polyclonal (1:1000) and anti-HIV-1 reverse transcriptase polyclonal (1:1000) antibodies were obtained from the NIH AIDS Research Reference and Reagent Program (#757 and #6195 respectively) (273, 274).

2.9.3 Typhoon system for fluorescence detection and quantification

FITC-conjugated anti-FLAG antibodies were visualised using the Typhoon Trio instrument (GE Heathcare, Life Sciences, Little Chalfont, UK). Quantification of the intensity of each band was achieved using the ImageQuant software (GE Healthcare Life Sciences, Little Chalfont, UK). A 5 µg/mL N-terminal FLAG-BAP control protein (Sigma-Aldrich, St. Louis, Missouri, USA) was serially diluted to 5 ng/mL and used for western blotting using the same antibody to generate a standard curve and estimate the amount of picomole (ρmol) of each APOBEC in the samples.

2.10 Viral encapsidation assay

M-MLV, HIV[p8.9] or HIV-1ΔVif-eGFP viral particles were produced (8 µg of M-MLV or 3 µg pCSGW, 2.5 µg p8.9 and 1.5 µg VSV-G or 5 µg HIVΔVif with 2 µg VSV-g) in the presence of Flag- or eGFP- or HA- tagged APOBECs (800 ng to 1.5 µg) in 293T cells in 10 cm dishes as described in section 2.3.2 for 96 h after which supernatants were collected, filtered through a 0.45 µm cartridge (Millipore) and layered on top of a 20% sucrose cushion for ultra-centrifugation at 100 000 x g at 4 ºC for 2 h in a 70 Ti rotor (Beckman Coulter Inc, Brea, California, USA) in an Optima^{TM} L100-XP centrifuge (Beckman Coulter Inc.). Following the

spin, viral pellets were resuspended in 100 µL of RIPA lysis buffer (150mM NaCl, 1% NP-40, 0.2% sodium dodecyl sulfate (SDS), 1mM EDTA, 0.5% Na-Deoxycholate, 50mM Tris-HCL (pH 8.0)) supplemented with cOmplete, EDTA-free, protease inhibitor cocktail (Roche) and lysed for 15 min on ice. Samples were then mixed with 30 µL of 5X Laemmli loading buffer, boiled 5 min and cooled on ice for 2 min prior to analysis for protein expression by Western blotting (2.9).

2.11 Proviral DNA sequencing

M-MLV, HIV[p8.9], HIVΔVif-eGFP or HIV-1ΔVif were produced in the presence of APOBEC proteins and used to infect 293T or 3T3 cells (section 2.3.3). Target cells were harvested 24 h after infection and used for genomic DNA (gDNA) extraction using the Wizard® Genomic DNA Purification Kit (Promega, Madison, Wisconsin, USA) according to the protocol provided with the kit. DNA, diluted to 2.5 ng/µL, was then used for PCR to amplify a 717 bp region of the eGFP viral reporter gene using the eGFP-FWD and eGFP-REV primers (**Table S1**) and the following cycling conditions; 94 ºC for 1 min followed by 32 cycles at 98 ºC for 10 sec, 58 ºC for 10 sec and 72 ºC for 1 min and a final extension at 72 ºC for 1 min 30 sec in an Eppendorf Mastercycler Pro S series (Eppendorf) with the following reagents; 8 µL of diluted DNA, 1 µL of each primer at 10 µM, 4 µL of 4X dNTPs mix (Takara Bio Inc., Otsu, Shiga, Japan), 10 µL of 10X Prime Star reaction buffer (Takara Bio Inc., Otsu, Shiga, Japan), 0.5 µL (1.25 units) of Prime Star (Takara Bio Inc., Otsu, Shiga, Japan) and 28.5 µL of water to complete the volume to 50 µL. PCR products were then ran on a 1% agarose gel containing 1% ethidium bromide and visualized under UV light using an AlphaImager® Mini system (Protein Simple, San Jose, California, USA). Bands of the appropriate size were cut and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, Wisconsin, USA) following manufacturer's

specifications and eluted in 20 µL of nuclease-free water. From that volume, 17 µL was incubated with 1 µL of T4 polynucleotide kinase (PNK) (NEB, Ipswich, Massachussets, USA) and 2 µL of corresponding 10X reaction buffer for 30 min at 37 ºC followed by heat inactivation at 65 ºC for 5 min to phosphorylate PCR products. Phosphorylated products were then used for ligation at a 1:3 ratio of vector to insert in the pBlueScript vector (Adgene, Cambridge, Massachussets, USA) previously digested with 20 units of EcoRV (NEB, Ipswich, Massachussets, USA) for 3 h at 37 ºC and dephosphorylated using 5 units of Antartic phosphatase (NEB, Ipswich, Massachussets, USA) at 37 ºC for 30 min followed by heat inactivation at 65 ºC for 5 min. Ligations were performed at 16 ºC overnight using 400 units of the T4 DNA ligase (NEB, Ipswich, Massachussets, USA) and its corresponding 10X reaction buffer in a final volume of 10 µL. Total reaction volumes were then transformed in 50 µL of DH5α competent cells and plated on LB agar plates containing 100 µg/mL of ampicilin (LB-AMP), 40 µg/mL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 0.1 millimolar (mM) Isopropyl β-D-1-thiogalactopyranoside (IPTG) for white/blue screening of successful insertion of the eGFP gene in the vector and incubated overnight in a 37 ºC incubator. The next day, various numbers of colonies, depending on the assay, were picked and grown in 1 mL of LB broth supplemented with 100 µg/mL of AMP overnight at 37 ºC in a shaker incubator for DNA purification using Pure Yield^{TM} Plasmid Miniprep system (Promega, Madison, Wisconsin, USA) and sequencing at the McGill University Genome and Innovation centre using M13 forward primers.

For sequencing of mutated eGFP clones in Chapter 5, the gDNA extracted from infected cells was used for PCR to amplify the ENV-eGFP gene from M-MLV or the eGFP gene from HIV[p8.9] using primers containing NheI and HindIII restrictions sites (ENV-NheI-FWD and

ENV-HindII-REV or eGFP-NheI-FWD and eGFP-HindII-REV) (**Table S1**). After purification of the PCR products using the Wizard® SV Gel and PCR Clean-Up System (Promega), the latter were digested with NheI and HindIII (NEB, Ipswich, Massachussets, USA) for 3 h at 37 ºC, purified a second time and ligated at a 1:3 ratio of vector to insert into pcDNA3.1 (Invitrogen) also previously digested with the same restriction endonucleases for expression under the control of a Cytomegalovirus (CMV) promoter. Following transformation of the overnight ligations, random selection of clones was performed as described above and sequencing was achieved using M13 forward and BGH reverse primers.

2.12 High throughput hypermutation analysis by High Resolution melting (HyperHRM)

2.13.1 HRM samples preparation

M-MLV or HIVΔVif viruses were produced in the presence of APOBEC proteins according to section 2.3.1 and used to infect 293T or 3T3 target cells following methodology in section 2.3.3. gDNA from infected cells was then collected 24 or 48 h following infection respectively. For mutation analysis of the eGFP reporter gene, fragments of increasing length were cloned in pBlueScript (Adgene) and individual clones were screened for G-to-A mutations by Sanger sequencing. Plasmids with known number of mutations were selected for each size and diluted to 2.5 ng/µL for HRM analysis. For analysis of full-length viral genome, first round PCRs were performed on the gDNA to amplify regions up to 2.2 kb. Primers for HRM are listed in **Table S1**. Amplicons were then purified and diluted to 2.5 ng/µL for use in HRM reactions. In the case of the HIV-1 infected patient, 100 ng of gDNA extracted from PBMCs was used for a first PCR reaction in which a region of the Pol gene was amplified using the HIV-Pol-FWD and HIV-Pol-REV primers. The product was then purified and ligated in pBlueScript (Adgene) followed by

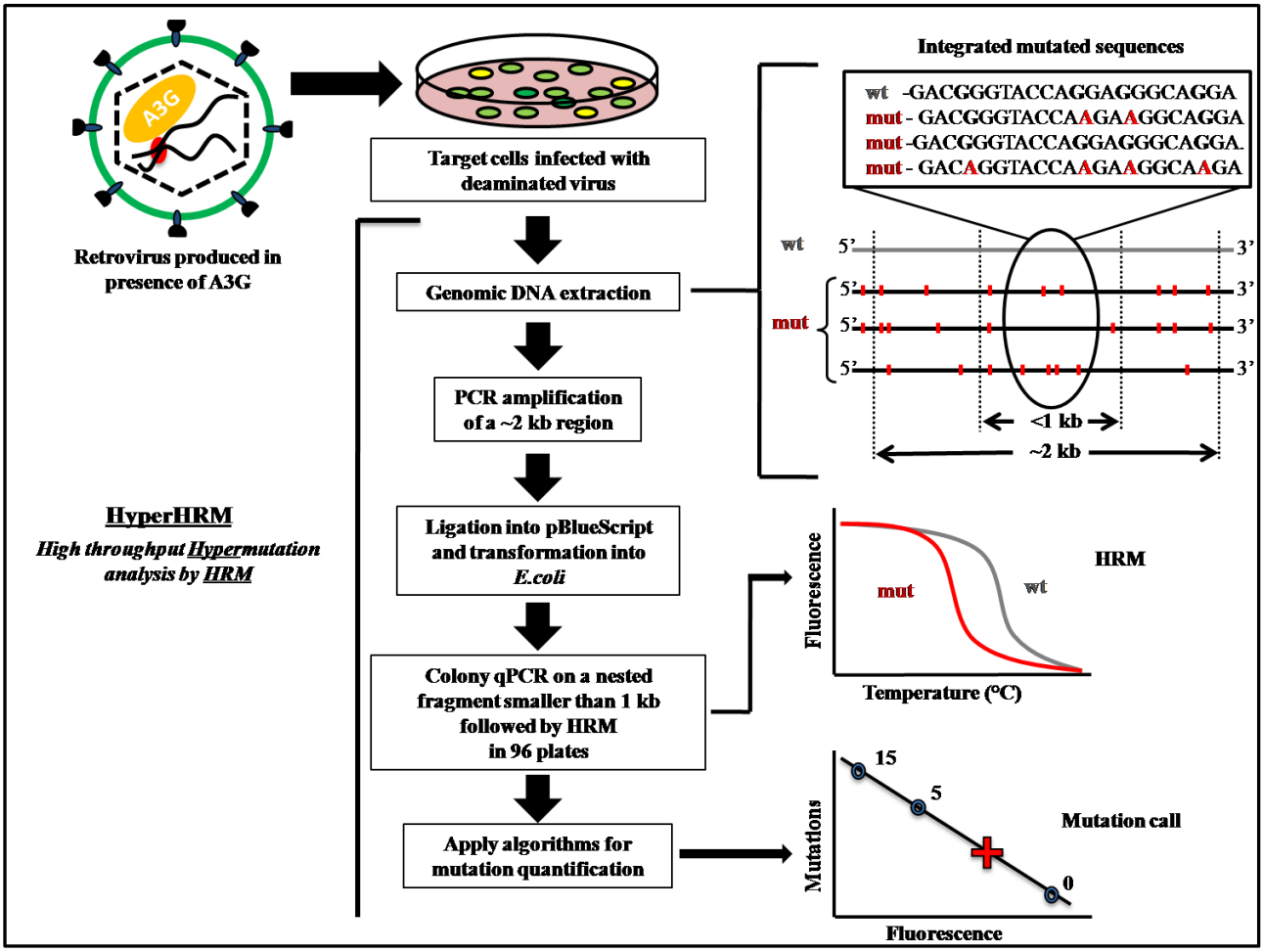

A3G
Retrovirus produced in presence of A3G
Target cells infected with deaminated virus
Genomic DNA extraction
PCR amplification of a ~2 kb region
Ligation into pBlueScript and transformation into E.coli
Colony qPCR on a nested fragment smaller than 1 kb followed by HRM in 96 plates
Apply algorithms for mutation quantification
HyperHRM
High throughput Hypermutation analysis by HRM
Integrated mutated sequences
wt -GACGGGTACCAGGAGGGCAGGA
mut - GACGGGTACCAAGAAGGCAGGA
mut -GACGGGTACCAGGAGGGCAGGA
mut - GACAGGTACCAAGAAGGCAAGA
wt
mut
5'
3'
<1 kb
~2 kb
Fluorescence
Temperature (°C)
HRM
mut
wt
Mutations
15
5
0
Mutation call
Fluorescence

Figure 2.1 Overview of the HyperHRM technique. Schematic representation of the methodology used for HRM analysis.

transformation in DH5α. Colonies were diluted in 250 µL of water and used for HRM analysis performed on the same region (**Figure 2.1**).

2.12.2 HRM reaction

DNA (plasmid, purified PCR product or diluted colony) was mixed with 10 µL of the MeltDoctor Master Mix (ABI Life Technologies, Carlsbad, California, USA, Catalog # 4415440), 0.6 µL (0.3 µM final) of each primer and 0.8 µL of water. The conditions for qPCR were 94 ºC for 2 min, followed by 40 cycles of denaturation at 95 ºC for 30 sec,

annealing at 55 ºC for 45 sec and extension at 72 ºC for 1 min. The qPCR step was followed by a melting curve analysis in which amplicons were heated up from 72 ºC to 95 ºC and fluorescence was acquired at intervals of 0.025 ºC. All reactions were achieved using a Vii7^{TM} 7 real-time PCR instrument (ABI Life Technologies).

2.13 3D-PCR

Differential denaturation DNA PCR (3D-PCR) analyses were performed in a two-step protocol (275). A first round PCR was first achieved to amplify a 717 bp fragment within the eGFP viral reporter gene using the Prime Star high fidelity polymerase (Takara) and the eGFP-FWD and eGFP-REV primers (**Table S1**). The cycling conditions were as followed; 94 ºC for 50 sec, followed by 30 cycles of denaturation at 94 ºC for 50 sec, annealing at 56 ºC for 30 sec, and an extension at 72 ºC for 1 min, with a final extension at 72 ºC for 5 min. A 1:25 dilution of the PCR product was then used for the second-round gradient PCR targeting a 279 bp nested region of the eGFP fragment using eGFP-FWD-279 and eGFP-REV-279 primers (**Table S1**). PCR cycles were; 94 ºC for 50 sec, followed by 30 cycles of a denaturation gradient from 92.2 to 86.1 ºC for 50 sec, annealing at 56 ºC for 30 sec, and an extension at 72 ºC for 30 sec, with a final

extension at 72 ºC for 5 min. PCR products were resolved on 2% agarose gel containing 1% ethidium bromide and visualized under UV light using an AlphaImager Mini system (Protein Simple).

2.14 Quantification of late reverse transcripts and proviral DNA integration

Viruses produced in the presence of the APOBEC proteins were used to infect 293T or 3T3 cells as depicted in 2.3.3. Infected cells were collected 12 h (for late reverse transcripts (LRT)) or 24 h for (proviral DNA integration) after infection and used for gDNA extraction using the Wizard® Genomic DNA Purification Kit (Promega) according to the recommendations from the manufacturer. DNA was then diluted to 2.5 ng/µL and 8 µL of that dilution was used for each reaction. For proviral DNA integration quantification, in order to select for integrated DNA, a first PCR was performed against Alu (293T) or B1 (3T3) retroelements and the eGFP reporter gene of retroviruses using primers described in

Table S1 (qAlu-1 and qAlu-2 or qB1-1 and qB1-2 and qeGFP-FWD) (276). Each PCR reaction was done in a final volume of 50 µL and contained DNA, 1 µL of each primer at 10 µM, 4 µL of 4X dNTP Mix (Takara), 10 µL of 10X PrimerStar buffer (Takara), 0.5 µL of Prime Star (1.25 units) (Takara) and 27.5 µL of water. Cycling conditions were 94 ºC for 1 min followed by 32 cycles at 98 ºC for 10 sec, 58 ºC for 10 sec and 72 ºC for 1 min and a final extension at 72 ºC for 3 min in an Eppendorf Mastercycler Pro S series (Eppendorf). PCR products were then diluted 1:40 and 8 µL was used for qPCR using nested primers and a probe inside eGFP. For both LRT and proviral DNA integration, qPCR reactions contained DNA, 1 µL of pre-mixed primers and probe (0.9 ρmol/µL of each primer and 0.25 ρmol/µL of the probe), 10 µL of the TaqMan$^{(R)}$ Gene Expression master mix (ABI, Life Technologies, Carlsbad, Massachussets, USA) and

water to a final volume of 20 µL. The sequence for all primers and probes can be found in **Table S1** (qIN-eGFP-FWD, qIN-eGFP-REV and qIN-eGFP-Probe, qHIV-FWD, qHIV-REV and qHIV-Probe, qM-MLV-FWD, qM-MLV-REV and qM-MLV-Probe). Cycling conditions for the qPCR were as followed: 10 min at 95 ºC, followed by 40 cycles of 15 sec at 95 ºC and 1 min at 60 ºC achieved on a Vii7^{TM} 7 real-time PCR instrument (ABI Life Technologies). The amount of DNA in each sample was normalised using human RNase P or mouse Tfrc copy number assays (TaqMan assays # 4 403 326 and 4 458 366 respectively, ABI Life Technologies, Carlsbad, Massachusetts, USA). Relative quantities based on the cycle threshold of each sample were determined using A2 as the comparative reference using the ViiA^{TM} 7 software (ABI Life Technologies).

2.15 Rif^{R} bacterial mutator assay

GST-tagged APOBECs were transformed in a uracil DNA glycosylase (UDG)-deficient strain of *E. coli* (BW310) and plated on LB-AMP plates. The next day, single colonies from each plate (for a total of 6 per plate) were picked and grown at 37 ºC in 1.5 mL of LB broth supplemented with 100 µg/mL of AMP and 1 mM IPTG for 16 h after which 300 µL of cultures were plated on low salt LB agar plates (10 g/L Tryptone, 5 g/L NaCl, 5 g/L Yeast extract, 15 g/L agar) containing 100 µg/mL AMP and 50 µg/mL rifampicin (Rifampicillin). Serial dilutions of the cultures were then performed to achieve a final concentration of 10^{-6}, 100 of which was plated on LB-AMP plates to control for cell viability and all plates were grown for 24 h in a 37 ºC incubator. The number of Rif resistant colonies (Rif^{R}) was compiled as the number of colonies on rifampicillin plates minus the number of colonies on LB-AMP plates after taking into account the dilution factor.

To confirm protein expression in *E. coli*, after transformation of GST-tagged APOBECs in the BW310 strain, a single transformant from each plate was picked and grown in LB broth containing 100 µg/mL of AMP until an optical density (O.D.) at 600 nanometer (nm) of around 0.6 was reached. The cells were then concentrated by centrifugation at 6 000 rpm followed by resuspension in 1 mL of bacteria lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1% Triton X-100). Samples were then lysed by sonication with 4 pulses of 10 sec each on ice after what they were centrifuged 20 min at 17 000 x g at 4 ºC to clear insoluble material. Aliquots of 75 µL were then collected from the top of each sample and mixed with 25 µL of 5X Laemmli loading buffer. Mixtures were boiled 5 min, incubated on ice 2 min and loaded on 10% SDS-PAGE gels for Western blot analysis using anti-A3G antibody (section 2.9).

2.16 Protein self-association analysis

HEK 293T cells were co-transfected with Flag- and HA- or eGFP-tagged APOBECs following procedures detailed in 2.3 and cultured for 48 h. Two days after transfection, pull-downs of HA- or eGFP-tagged proteins were performed using magnetic beads' immunoprecipitation protocol (section 2.6.2). Detection of the APOBECs in the eluates was then achieved by Western blotting (section 2.9) using anti-Flag, anti-HA or anti-eGFP antibodies (2.9.2).

2.17 HIV clinical samples

Genomic DNA of PBMCs from an HIV-1 (Group M, Subtype B) infected patient was extracted and gratefully provided by Dr. Jason Fernandes, post-doctoral fellowship in the laboratory of Dr. Jonathan Angel at the Ottawa Hospital Research Institute (OHRI).

2.18 Bioinformatics analysis

The R software (R Development Core Team) was used to test each point of the HRM melt curve of control DNA plasmids with known numbers of mutations and determine that there was a linear relationship between mutation numbers in a sample and its fluorescence at the temperature yielding the maximum difference between the fluorescence of the different control plasmids. The fluorescence data at every acquired temperature for plasmid DNA controls containing 0, 1, 4 or 9 mutations were then imported into a Microsoft Excel spread sheet to determine the temperature at which the difference between the control was the highest using the following formula:

$$\max_{1\leq j\leq n} \sum_{i=1}^{k} (Xij - Xi + 1j)$$

(where n represents the total of all temperatures tested, k the number of controls used to generate the standard curve and Xij the fluorescence acquired at temperature j for curve i) and a linear regression of the number of mutations as a function of the fluorescence was established for the controls at that specific temperature to determine the number of mutations in experimental samples using the following formula;

$$y=A + Bx$$

(where y is the number of mutations and x the fluorescence at the established temperature.)

2.19 LINE-1 Restriction assay

L1 retrotransposition assays were carried out as previously described (272). Co-transfection of 800 ng of the L1 expression plasmid in the presence of A3 was performed in HEK 293T cells in 6-well plates according to methods in section 2.3.1. Varying amounts of the APOBECs were used to get similar levels of expression (**Table 2.1**). The cells were then cultured for seven days after what they were resuspended in 1 mL of PBS-EDTA (1 mM EDTA) and analysed for eGFP

fluorescence by flow cytometry using the Cyan ADP flow cytometer (Beckman Coulter Inc., Brea, California, USA).

Table 2.1 Quantity of each A3 co-transfected for L1 restriction assay

Plasmid	Amount (ng)
Empty Vector	500
A2	75
A3A	400
A3A[W98A]	400
A3B	500
A3B[W94A]	1200
A3B[W127A]	1200
A3C	200
A3C[W94A]	200
A3D	1000
A3D[W106A]	1000
A3D[W143A]	1000
A3F	300
A3F[W93A]	1200
A3F[W126A]	1200
A3G	200
A3G[W94A]	300
A3G[W127A]	300
A3H	500
A3H[W82A]	1200

2.20 Statistical analysis

All infection, integration and late reverse transcription assays presented in this thesis were performed at least three times from three independent transfections and are represented as the mean relative value to A2 +/- standard deviation (SD). The mean fluorescence of mutated eGFP clones was the result of three independent transfections and is depicted as the relative mean fluorescence to plasmid control +/- standard deviation (SD). All statistical analysis were performed using the Student's paired *t*-test using the GraphPad Prism software.

Chapter 3 The Role of RNA-Binding in the Deamination-Independent Retroviral Restriction by APOBEC3G

Preface: The results presented in this chapter have been published in:

Kasandra Bélanger, Mathieu Savoie, Maria Rosales Gerpe, Jean-François Couture and Marc-André Langlois. **(2013)** Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. ***Nucleic Acid Research***, 41, 7438-7452.

*See Appendix II for the published article.

3.1 Introduction

A3G is the best characterized and most potent member of the A3 protein family. It is currently believed that the main mechanism by which this protein achieves its antiretroviral activity is through the deamination of target deoxycytidines into deoxyuridines in the retroviral cDNA during reverse transcription following its incorporation into budding particles from viral producer cells (Reviewed in (101, 277)). Deamination by A3G leads to the accumulation of potentially deleterious G-to-A hypermutations in the genome of progeny viruses resulting in protein malfunction and infection impairment (278). There is however, growing evidence for a deamination-independent restriction by this deaminase coming in most part from the use of catalytically inactive mutants of A3G that, although completely lacking the ability to induce mutations in the retroviral genome during replication, still significantly reduce infection levels (221, 228, 229). Defects in $tRNA_3^{Lys}$ primer annealing, plus and minus strands transfer, early and late reverse transcript synthesis and proviral DNA integration have been observed in the presence of such mutants (226, 227, 230, 232, 279). The exact mechanism for this so called deamination-independent restriction by the protein remains however unclear and the contribution of each one of the modes of action proposed above to the overall restriction by A3G is currently unknown.

RNA binding is involved in many features of A3G's biological functions. The incorporation of the protein in viral particles has been shown to be mediated by the NC domain of HIV-1 Gag and to require viral genomic RNA as well as a pol-III derived RNA, 7SL (172, 175, 186). Self-oligomerization by the protein has also been suggested to require its interaction with cellular RNA (171). In addition to oligomers, A3G is also known to assemble into larger RNA-dependent HMM complexes in replicating cells, a system that actively dividing cells have evolved to protect their genome from the potential threat that A3G's mutator activity poses to their DNA (215-217). In those large RNP complexes, the catalytic activity of A3G is inhibited through the interaction of the protein with numerous RNA species and proteomic co-factors (214). Several efforts have been made to try identifying the specific determinants for the recruitment of A3G into those HMM complexes responsible for its inactivation. As a result, numerous RNA and protein species have been identified (216, 217). Consequently, the nature of the specific cellular factors involved in the formation of the HMM complexes remains undetermined to date. We hypothesized that by inhibiting the formation of HMM complexes, we could keep A3G in an active state in dividing cells and increase its retroviral activity upon infection. We first aimed at generating mutants of the protein in order to identify the specific residues on A3G involved in the formation of HMM complexes and to further identify the cellular co-factors responsible for the inactivation of A3G in dividing cells. This led to the identification of two tryptophans in the NTD of A3G, W94 and W127, which we found to be responsible for A3G's RNA-binding ability. Although mutants of A3G harbouring alanines instead of tryptophans at these two specific locations failed to restrict retroviral infection, they still had the ability to deaminate retroviral cDNA during reverse transcription in the target cell. We further demonstrated that the lack of inhibition of infection in the case of the mutants was

due to the loss of their deamination-independent restriction defined as the inhibition of late reverse transcripts accumulation and proviral DNA integration, which was associated with the RNA-binding properties of A3G. In conclusion, we found that RNA-binding by A3G governs the deamination-independent restriction by the protein.

3.2 Results

3.2.1 Identification of RNA-binding defective mutants

In a first attempt to identify critical residues responsible for the formation of the HMM complexes by A3G, we mutated single amino acids in residues involved in critical functions of the protein such as self-oligomerization, deamination, RNA-binding, viral encapsidation and Vif interaction (**Figure 3.2.1A**) (171-173, 177, 179, 180, 185, 186, 280-282). Since a crystal structure of the full-length protein is not available, the selection of target motifs was based on homologous residues identified in the structure of the C-terminal catalytic domain of the protein (283). Residues were mutated in both domains of A3G and all mutants were analyzed for the inhibition of HMM complex formation by velocity sedimentation on 5 to 40% sucrose gradients followed by Western blotting using an anti-Flag antibody. This initial screen led to the identification of two tryptophans in the NTD of A3G, W94 and W127. When mutated to an alanine, these residues significantly reduced the RNA-binding capacity of the protein as shown by the shift in the sucrose gradients towards the middle fractions (lanes 4-7) where complexes of lower density (LMM complexes) accumulate (**Figure 3.2.1B**). Other variants of A3G did not seem to affect the properties of the protein and remained in HMM complexes aggregating in the bottom fractions (lanes 8-9) of the sucrose gradients. Treatment of the complexes formed by the W94A and W127A mutants 15 min at room temperature with RNase A resulted in a complete

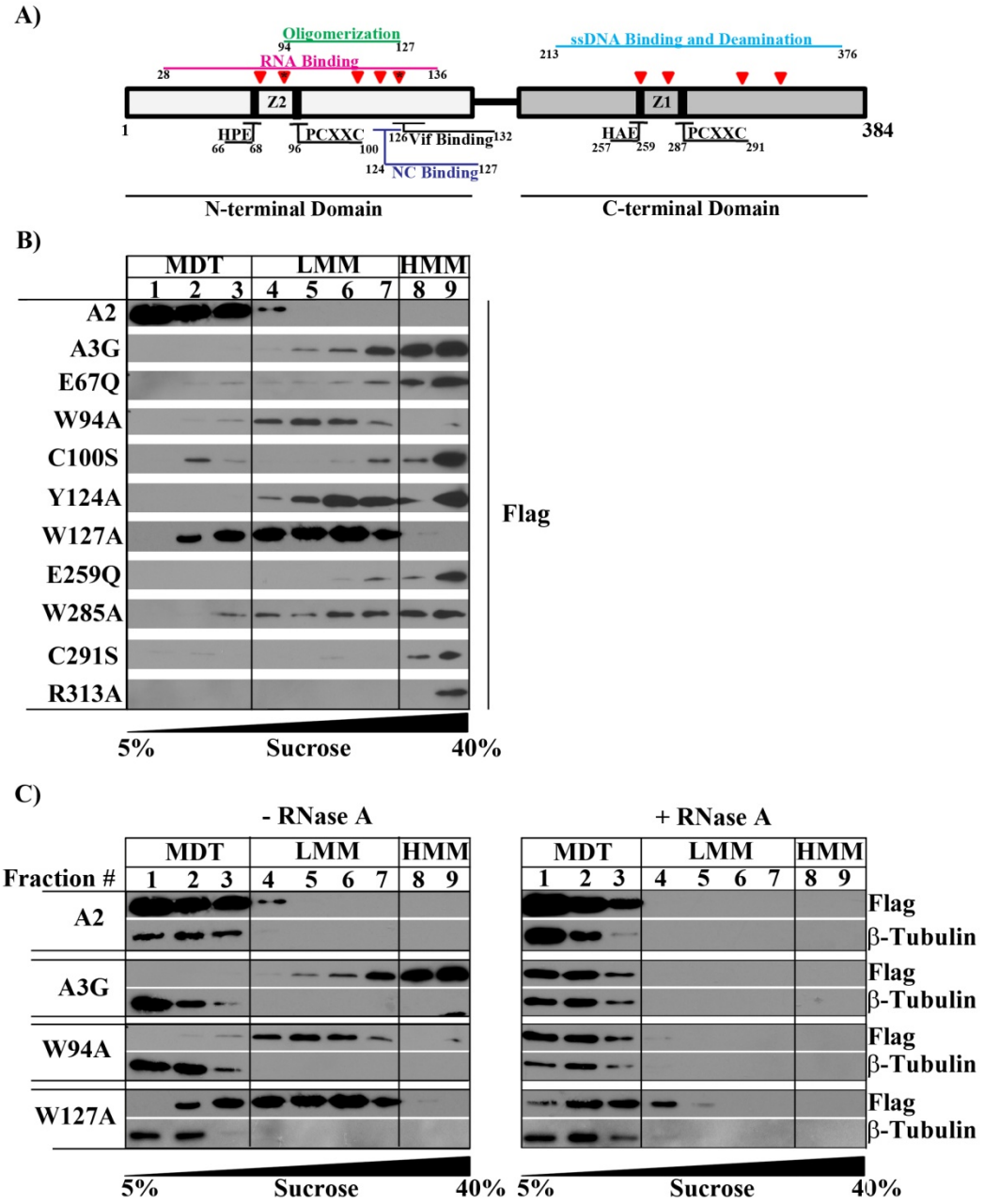

A)
Oligomerization
94
127
RNA Binding
28
136
ssDNA Binding and Deamination
213
376
Z2
Z1
1
384
HPE
66
68
PCXXC
96
100
126
Vif Binding
132
124
NC Binding
127
HAE
257
259
PCXXC
287
291
N-terminal Domain
C-terminal Domain
B)
MDT
LMM
HMM
1 2 3 4 5 6 7 8 9
A2
A3G
E67Q
W94A
C100S
Y124A
W127A
E259Q
W285A
C291S
R313A
Flag
5%
Sucrose
40%
C)
- RNase A
+ RNase A
Fraction #
MDT
LMM
HMM
1 2 3 4 5 6 7 8 9
A2
A3G
W94A
W127A
Flag
β-Tubulin
5%
Sucrose
40%
D)
Relative RNA Binding (%)
100
80
60
40
20
0
**
7SL
Alu
hY1
hY3
A3G
A2
A3G[W94A]
A3G[W127A]

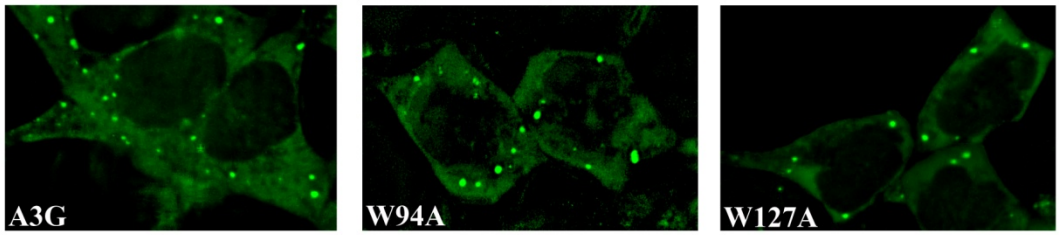
Figure 3.2.1 Identification of RNA-binding defective mutants. (**A**) Schematic representation of both domains of A3G and their main functions. The position of each mutation included in the initial screen is indicated by a red triangle. The W94 and W127 residues are identified by a little black star inside the red triangles. (**B**) All mutants of A3G were analyzed for the inhibition of HMM complex formation by velocity sedimentation on 5 to 40% sucrose gradients followed by Western blotting using an anti-Flag antibody. (**C**) Lysates from 293T cells expressing Flag-tagged A2, A3G or variants of the protein were treated for 15 min at room temperature with 1 µg/µL of RNase A (right) or untreated (left) prior to ultra-centrifugation on sucrose gradients and Western blot analysis. β-tubulin was used as a control for the consistency of the sucrose gradients as this protein does not oligomerize and should always accumulate in the top fractions (lanes 1 to 3) of the gradients. (**D**) RNA was extracted from immunoprecipitated Flag-tagged APOBECs and used for RT-PCR. The resultant cDNA was then used for the quantification of 7SL, Alu, hY1 and hY3 RNAs by qPCR. Results are represented as relative values to the wild type A3G and represent the mean +/- SD of three independent experiments. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005.

shift towards the top of the gradients (lanes 1-3) and accumulation of the proteins in monomeric, dimeric and tetrameric forms suggesting that the mutants only have a partial phenotype in terms of RNA-binding deficiency ((266, 284) and **Figure 3.2.1C**). This statement was further confirmed by performing a qPCR on RNA extracted from immunprecipitates of Flag-tagged A2, A3G or mutants of A3G using anti-Flag-conjugated agarose beads to measure their binding to 7SL, Alu, hY1 and hY3 RNAs. These RNA species have been identified as forming components of complexes formed by both A3G and A3F and have been shown to interact directly with the wild type A3G (172, 175, 215, 218). As expected, both mutants showed a significantly lower interaction with all RNAs tested although higher than the A2 negative control (**Figure 3.2.1D**).

3.2.2 Characterization of the W94A and W127A mutants

We were then curious to see the effect of the two tryptophan mutations on other features of A3G. A3G is predominantly cytoplasmic where it associates with RNA to form visible foci that are believed to be P-bodies and stress granules (216-218). The cytoplasmic localization signal of A3G has been mapped to residues within amino acids 1 to 60 and amino acids 113 to 128 (285). We were therefore interested to see whether the mutation at position 127 would affect the cell localization of A3G. In fact, the substitution of this residue for an alanine in combination with two other mutations, F126A and Y19D has been associated with the redistribution of A3G throughout the cell whereas W127A on its own had little effect on cell localization (285). In agreement with these results, our fluorescence microscopy analysis of live cells transfected with eGFP-tagged A3G, A3G[W94A] or A3G[W127A] demonstrated that the cellular distribution of A3G was unaffected by neither of the mutations and that the size and density of the foci between

A)


A3G
W94A
W127A

B)

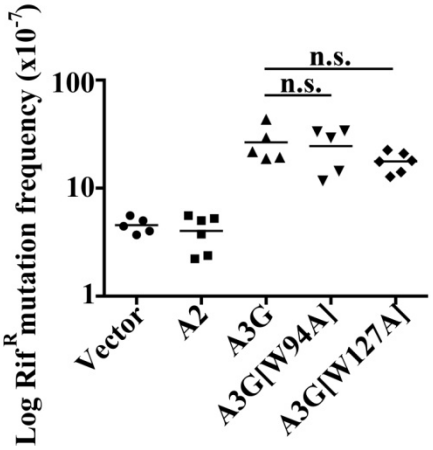

Log Rif^R mutation frequency (x10^-7)
100
10
1
n.s.
n.s.
Vector
A2
A3G
A3G[W94A]
A3G[W127A]

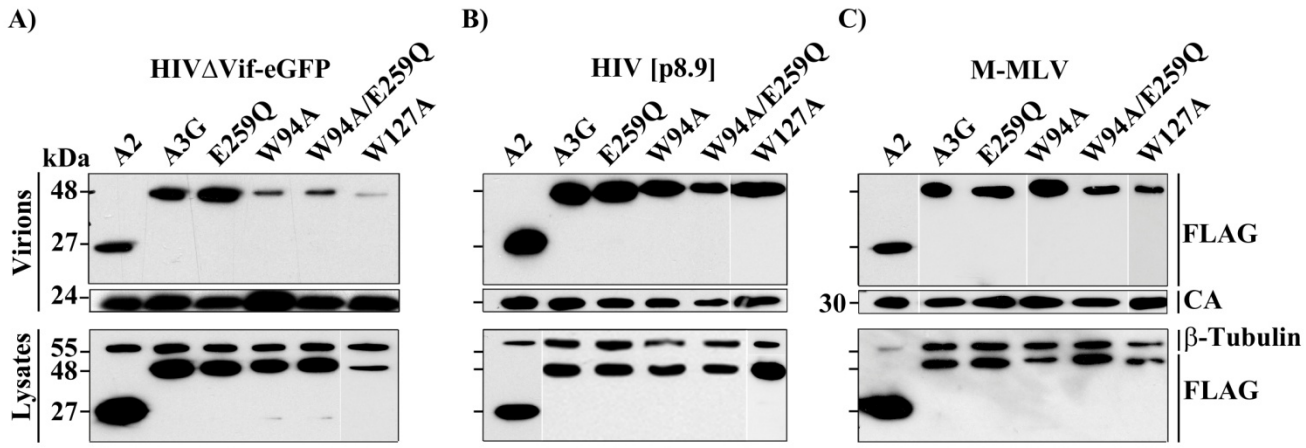
Figure 3.2.2 Characterization of the W94A and W127A mutants. (**A**) eGFP-A3G, eGFP-A3G[W94A] or eGFP-A3G[W127A] were transfected in 35 mm glass bottom dishes and cultured in clear media for 48 h. Imaging of live cells was performed using a Plan-Apochromat 63x/1.4 oil immersion objective on a Zeiss AxioObserver.Z1 microscope. (**B**) GST-expression plasmids were transformed in a UDG-deficient strain of *E.coli*. Individual colonies from each plates (each point on the graph) were then picked and grown in LB broth containing 1μg/mL of Amp overnight at 37 ºC. Cultures were then plated on Rifampicillin plates and the number of resistant colonies (Rif^R) was compiled the next day. Results are presented as the log value of Rif^R colonies per 10^7 viable cells. The bars indicate the mean value for each sample. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; n.s. non-significant.

the proteins displayed no remarkable differences (**Figure 3.2.2A**). Since the two mutants did not assemble into HMM complexes, these results suggest that those complexes are different from P-bodies and stress granules.

Seeing as the mutations could potentially destabilize the structure and affect the activity of the deaminase, we next investigated whether the tryptophan mutants were catalytically active by conducting a bacterial mutator assay (154). Briefly, GST-expression plasmids were transformed in a UDG-deficient strain of *E.coli* and the ability of the mutants to induce mutations resulting in resistance to rifampicin (Rif^R) in the bacterial DNA was measured. We found that both mutants were enzymatically active and led to the production of a significant number of Rif^R colonies (**Figure 3.2.2B**). Overall, our results demonstrated that RNA-binding deficiency did not affect the cellular localization nor the enzymatic activity of A3G.

3.2.3 Retroviral Restriction by the W94A and W127A mutants

Based on previous observations suggesting intact catalytic activity of the mutants, we then aimed to determine whether they could still restrict retroviral infection by HIVΔVif-eGFP, HIV[p8.9] and M-MLV. Viruses were produced by transfection in HEK 293T cells for 48 h after which supernatants were filtered and used to infect 293T (HIVs) or 3T3 (M-MLV) target cells by spinoculation for 1 h at 2 000 rpm in the presence of 8 μg/mL of polybrene. Levels of infection were determined as a measure of the expression of the eGFP viral reporter gene by flow cytometry 24 h (HIV[p8.9] and M-MLV) or 48 h (HIVΔVif-eGFP) after infection. Equal expression of viral proteins and APOBECs in cell and viral lysates was verified by Western blot analysis showing consistent amounts of p24 (HIVs), p30 (M-MLV) and Flag tag (APOBECs) to ensure successful viral encapsidation of the APOBECs (**Figure 3.2.3A-C**). Tryptophan 127 has


A)
HIVΔVif-eGFP
B)
HIV [p8.9]
C)
M-MLV
A2
A3G
E259Q
W94A
W94A/E259Q
W127A
kDa
Virions
48
27
24
Lysates
55
48
27
30
FLAG
CA
β-Tubulin
FLAG

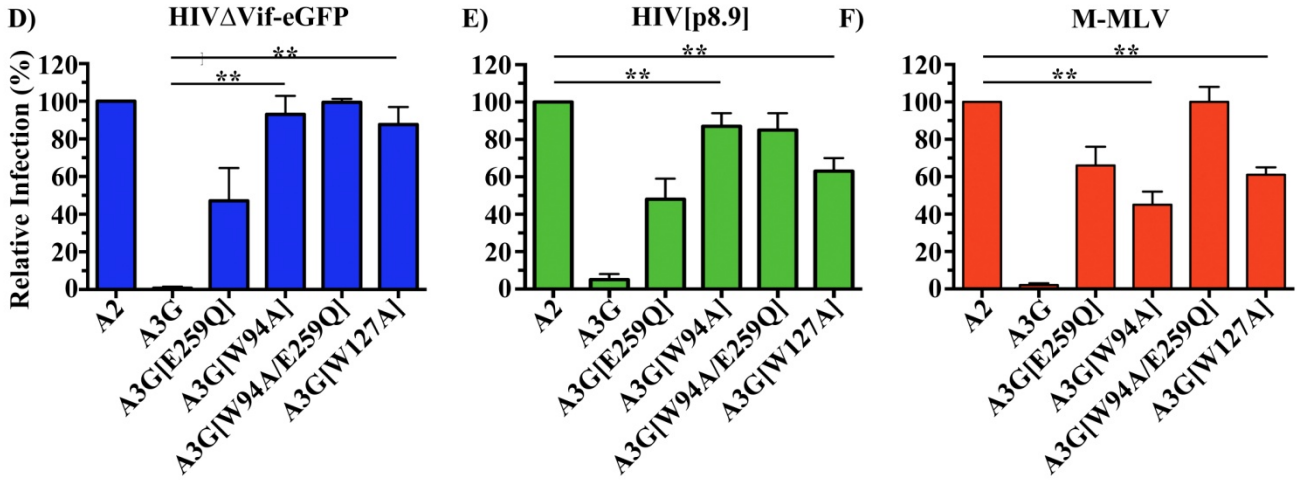

D)
HIVΔVif-eGFP
E)
HIV[p8.9]
F)
M-MLV
Relative Infection (%)
120
100
80
60
40
20
0
**
**
A2
A3G
A3G[E259Q]
A3G[W94A]
A3G[W94A/E259Q]
A3G[W127A]

Figure 3.2.3 Retroviral Restriction by the W94A and W127A mutants. (**A to C**) (**A**) HIVΔVif-eGFP, (**B**) HIV[p8.9] or (**C**) M-MLV viruses produced in the presence of Flag-APOBEC expression plasmids were purified by ultra-centrifugation and lysed using RIPA lysis buffer for Western blot analysis using an anti-Flag antibody to assess the presence of the APOBECs within virions. Expression of capsid proteins, p24 (HIVΔVif-eGFP and HIV[p8.9]) or p30 (M-MLV) was used to normalize the amounts of viruses in each sample. Viral producer cell lysates were also analysed for protein expression by western blot analysis using β-tubulin as a loading control. (**D to F**) Flow cytometry analysis of eGFP fluorescence in (**D**) HIVΔVif-eGFP, (**E**) HIV[p8.9] or (**F**) M-MLV infected cells 24 or 48 h after infection. Viruses were initially produced in 293T cells in the presence of Flag-tagged A2, A3G or variants of the protein. Data represent the mean relative values +/- SD of three independent experiments. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005.

been involved in the packaging of A3G inside HIV-1 and as a consequence, very low levels of this mutant were found inside HIVΔVif-eGFP (172, 177, 186, 280, 283). The same was also found to be true for W94A and its catalytically dead variant. Surprisingly, both tryptophan mutants were found to be incorporated into HIV[p8.9] and M-MLV at levels that were similar to the wild type A3G suggesting that the determinants for packaging into these viruses are different from HIVΔVif-eGFP. Although both the W94A and the W127A mutants appeared at similar levels into both viruses, they both showed a severe (less than 40% restriction) loss of their antiretroviral activity against HIV[p8.9] and a 45 and 60% reduction in their restriction potential respectively in the case of M-MLV (**Figure 3.2.3E and F**). Moreover, both mutants failed to restrict HIVΔVif-eGFP as a result of their inability to get incorporated into viral particles (**Figure 3.2.3D**).

Since the catalytic activity of A3G[W94A] and A3G[W127A] remained unaffected by the mutations and that the mutants retained their ability to incorporate into HIV[p8.9] and M-MLV viral particles despite significantly lower restriction potency against both viruses, we hypothesized that they may retain their ability to deaminate retroviral cDNA during reverse transcription in target cells. We therefore collected gDNA from infected cells 24 h after infection and amplified integrated eGFP reporter genes by PCR, subcloned them into a sequencing vector and analyzed them for the presence of G-to-A mutations via Sanger sequencing. We were quite surprised by the remarkably high levels of mutations induced by the two mutants in both retroviral genomes (**Tables 3.2.1 and 3.2.2**) with almost all sequences analyzed harbouring mutations. Analysis of target motifs for deamination revealed a strong preference for 5'CC dinucleotide substrates for both mutants as is observed with the wild type A3G (**Tables 3.2.3 and 3.2.4**).

Table 3.2.1 Deamination intensity induced by A3G[W94A] and A3G[W127A] in HIV[p8.9].

APOBEC	Total base pairs sequenced	Sequences mutated (%)	Total number of G-to-A mutations	Mutation rate (mutations/kb)
A3G	7887	82	82	10.4
E259Q	4302	0	0	-
W94A	18 642	92	132	7.1
W127A	12 189	82	88	7.2

Table 3.2.2 Deamination intensity induced by A3G[W94A] and A3G[W127A] in M-MLV.

APOBEC	Total base pairs sequenced	Sequences mutated (%)	Total number of G-to-A mutations	Mutation rate (mutations/kb)
A3G	11 472	88	111	9.7
E259Q	4302	0	0	-
W94A	15 057	95	85	5.6
W127A	7887	91	38	4.8

Table 3.2.3 Analysis of the preferred DNA substrate for deamination by A3G[W94A] and A3G[W127A] in HIV[p8.9]*.

APOBEC	-2	-1
A3G		
A	20	2
C	48	77
T	9	1
G	23	20
W94A		
A	12	1
C	60	90
T	4	1
G	24	8
W127A		
A	33	3
C	44	82
T	5	3
G	18	12

*Mutations were tabulated on the minus-strand retroviral DNA. Data represents the percentage of each base at position -2 and -1 relative to the deaminated cytidine at position 0.

Table 3.2.4 Analysis of the preferred DNA substrate for deamination by A3G[W94A] and A3G[W127A] in M-MLV*.

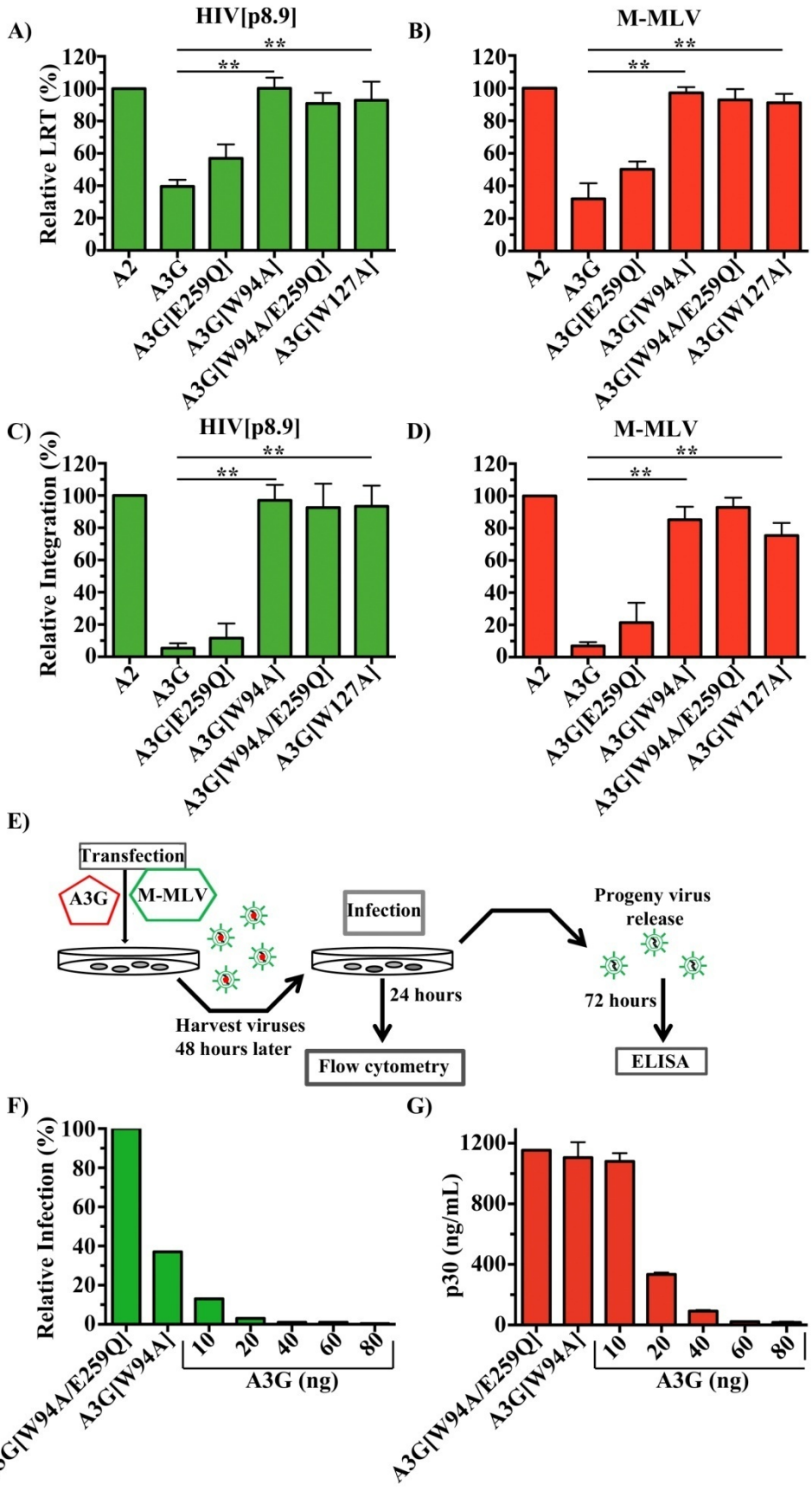
APOBEC	-2	-1
A3G		
A	20	0
C	39	68
T	4	3
G	37	29
W94A		
A	8	0
C	68	89
T	4	1
G	20	10
W127A		
A	14	2
C	53	83
T	4	0
G	29	15

*Mutations were tabulated on the minus-strand retroviral DNA. Data represents the percentage of each base at position -2 and -1 relative to the deaminated cytidine at position 0.

3.2.4 Absence of deamination-independent restriction in the presence of A3G[W94A] and A3G[W127A]

Since the loss of antiretroviral restriction observed with the W94A and W127A mutants could not be explained by a lack of deamination, we then sought to determine what could be the cause of this unusual phenotype. We turned our attention towards the deamination-independent restriction by A3G. Although the exact mechanisms behind this mode of inhibition are not currently completely understood, it is well established that it involves the inhibition of late reverse transcript accumulation and proviral DNA integration (226, 228-230, 232, 279). We thus collected gDNA from cells infected with HIV[p8.9] or M-MLV produced in the presence of A2, A3G or variants of the protein at 12 or 24 h after infection and measured the levels of late reverse transcripts or integrated eGFP, respectively by qPCR. An answer to our question emerged from the results of that experiment showing an almost complete (less than 20% restriction in all cases) loss of the inhibition of both late reverse transcript accumulation and proviral DNA integration in the presence of A3G[W94A] and A3G[W127A] suggesting that RNA-binding is required for the inhibition of the early steps of retroviral infection and that deamination by A3G only has a marginal role in preventing these two processes (**Figure 3.2.4A-D**). That was further confirmed by data showing that the catalytically inactive mutant of A3G, E259Q, displayed almost as much deamination-independent restriction (less than 10% difference) than the wild type protein.

Having found that deamination only played a minor role in the deamination-independent restriction by A3G, we then wondered about its main role. We decided to look at viral particle release using the replication competent virus M-MLV, since the RNA-binding mutants of A3G failed to be efficiently incorporated into HIVΔVif-eGFP. Viral particles were produced in the


A)
HIV[p8.9]
Relative LRT (%)
B)
M-MLV
C)
HIV[p8.9]
Relative Integration (%)
D)
M-MLV
A2
A3G
A3G[E259Q]
A3G[W94A]
A3G[W94A/E259Q]
A3G[W127A]
**
E)
Transfection
A3G
M-MLV
Harvest viruses
48 hours later
Infection
24 hours
Flow cytometry
Progeny virus
release
72 hours
ELISA
F)
Relative Infection (%)
A3G[W94A/E259Q]
A3G[W94A]
A3G (ng)
G)
p30 (ng/mL)

3.2.4 Absence of deamination-independent restriction in the presence of A3G[W94A] and A3G[W127A]. (**A to D**) gDNA from (**A and C**) HIV[p8.9] or (**B and D**) M-MLV infected cells was collected 12 or 24 h following infection for quantification of (**A and B**) late reverse transcripts (LRT) or (**C and D**) proviral DNA integration by qPCR. Values are the mean +/- SD of three independent experiments performed in triplicate and are normalized to A2. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005. (**E**) Schematic representation of the methodology used to determine the effect of A3G and its W94A mutant on M-MLV viral particle release. (**F and G**) M-MLV viruses were produced in the presence of A3G[W94A] or its catalytically inactive version or increasing amounts of the wild type A3G followed by infection of 3T3 target cells. (**F**) Levels of infection as measured by eGFP fluorescence were determined by flow cytometry analysis 24 h post-infection and supernatant was then replaced with fresh DMEM for culturing for an additional 72 h. (**G**) Three days following the initial infection, p30 capsid protein concentration in supernatants was determined by ELISA.

presence of decreasing amounts of A3G to find a quantity that would give rise to similar levels of mutations to those observed in the presence of 80 ng of the W94A mutant (**Table 3.2.1**) and these amounts were then used to infect target 3T3 cells. Infection levels were assessed by flow cytometry of eGFP 24 h after infection and viral supernatants were collected three days after infection for the quantification of p30 capsid proteins by ELISA (**Figure 3.2.4E**). The results demonstrated that at equal mutation rates (10 ng of A3G), both A3G and A3G[W94A] did not seem to affect viral particle release although they both had a significant effect on retroviral infection (60 and 80% restriction respectively) (**Figure 3.2.4F and G**). In summary, deamination by A3G does not impede late reverse transcript accumulation or proviral DNA integration nor does it have a detrimental effect on viral particle release. It is possible that mutations by A3G affects the function of genes through the introduction of termination codons or other inactivating mutations. This subject will be extensively discussed in Chapter 5 of this thesis.

3.2.5 Fusion with Vpr results in deamination-independent restriction by the RNA-binding deficient mutants.

The aim of this section was to determine the antiretroviral activity of the mutants against the more physiologically relevant HIVΔVif-eGFP by enforcing their packaging into virions through fusion with a fragment of Vpr (residues 14-88). This technique has previously been used in the field and has been shown to induce encapsidation of different proteins inside viral particles (173, 286). In fact, viral encapsidation assays performed on HIVΔVif-eGFP produced in the presence of Flag-tagged APOBECs fused to Vpr^{14-88} demonstrated similar packaging efficiency for the wild type A3G and its mutants when fused to the polypeptide (**Figure 3.2.5A**). More

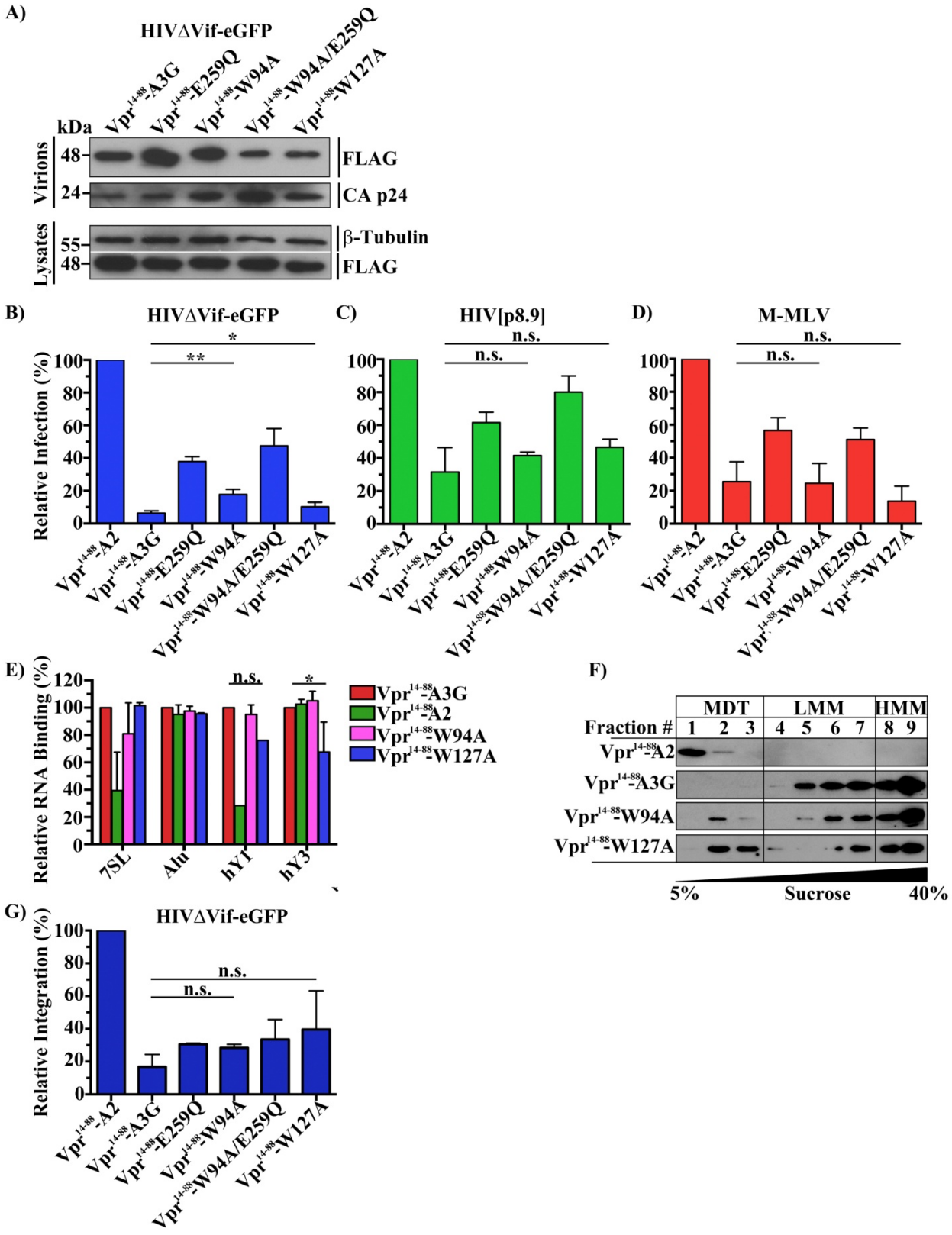

A)
HIVΔVif-eGFP
Vpr14-88-A3G
Vpr14-88-E259Q
Vpr14-88-W94A
Vpr14-88-W94A/E259Q
Vpr14-88-W127A
kDa
Virions
48
FLAG
24
CA p24
Lysates
55
β-Tubulin
48
FLAG
B)
HIVΔVif-eGFP
Relative Infection (%)
C)
HIV[p8.9]
D)
M-MLV
n.s.
E)
Relative RNA Binding (%)
Vpr14-88-A3G
Vpr14-88-A2
Vpr14-88-W94A
Vpr14-88-W127A
7SL
Alu
hY1
hY3
F)
MDT
LMM
HMM
Fraction #
1 2 3 4 5 6 7 8 9
Vpr14-88-A2
Vpr14-88-A3G
Vpr14-88-W94A
Vpr14-88-W127A
5%
Sucrose
40%
G)
HIVΔVif-eGFP
Relative Integration (%)

Figure 3.2.5 Fusion with Vpr results in deamination-independent restriction by the RNA-binding deficient mutants. (**A**) Analysis of the efficiency of the Vpr^{14-88}-fusion proteins to get incorporated into HIVΔVif-eGFP by Western blotting using an anti-Flag antibody. The expression level of p24 was used as a loading control for viral lysates whereas β-tubulin was used for cell lysates. (**B to D**) The potency of Vpr^{14-88}-A2, A3G and mutants to restrict infection by (**B**) HIVΔVif-eGFP, (**C**) HIV[p8.9] or (**D**) M-MLV was assessed by producing viruses in the presence of the vpr^{14-88}-APOBECs in 293T cells and infecting 293T or 3T3 target cells with virus-containing supernatants 48 h after transfection. Infection levels were determined by measuring the fluorescence of eGFP 24 or 48 h following infection by flow cytometry. Values were normalized to Vpr^{14-88}-A2 and reflect the mean of three independent experiments +/- SD. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; *: $P<0.05$; **$P<0.005$; n.s. non-significant. (**E**) The binding of fusion proteins to 7SL, Alu, hY1 and hY3 RNAs was quantified by qPCR performed on reverse transcribed RNA extracted after immunoprecipitation of Flag-tagged Vpr^{14-88}-A2, A3G or variants using anti-Flag-conjugated agarose beads. Results are the mean value +/- SD of three independent experiments and are relative to A3G. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; *: $P<0.05$; n.s. non-significant. (**F**) Lysates from 293T cells expressing fusion proteins were used for velocity sedimentation analysis on 5 to 40% sucrose gradients to determine the ability of the proteins to assemble into HMM complexes. Samples were then resolved on SDS-PAGE gels and detection of the proteins was achieved using an anti-Flag antibody. (**G**) Levels of proviral DNA integration in the presence of Vpr^{14-88}-APOBECs were measured by extracting gDNA from infected cells 24 h following infection and performing a qPCR on the samples. Results represent the mean value +/- SD of three independent experiments performed in triplicate and are normalized to Vpr^{14-88}-A2. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; n.s. non-significant.

interestingly, restriction assays conducted with the fusion proteins showed a strong reduction of retroviral infection in the presence of A3G[W94A] and A3G[W127A] (80 and 90% restriction respectively) although not as pronounced as the one observed with the wild type A3G (**Figure 3.2.5B**). The antiretroviral capability of the mutants fused to vpr^{14-88} was then tested against HIV[p8.9] and M-MLV and surprisingly, fusion with the viral protein appeared to rescue the antiretroviral activity of the mutants against both retroviruses (60-80% restriction) (**Figure 3.2.5C and D**). Those results were really unexpected since the initial lack of inhibition observed with the RNA-defective mutants against HIV[p8.9] and M-MLV could not be attributed to ineffective virion incorporation (**Figure 3.2.3D and E**).

One possible explanation arises from the fact that Vpr is an RNA-binding protein (287). One could therefore assume that fusion of the mutants with Vpr^{14-88} may have restored their RNA-binding properties leading to a wild type protein-like restriction phenotype. In order to investigate that issue, we immunoprecipitated Flag-tagged fusion proteins using anti-Flag-conjugated agarose beads and extracted the RNA from eluates to measure the levels of 7SL, Alu, hY1 and hY3 RNAs in each samples. We found that the W94A and W127A mutants were now able to bind all RNAs tested to similar levels to the wild A3G (**Figure 3.2.5E**). This increase in the RNA-binding ability of the mutants led to their accumulation into HMM complexes (lanes 8 and 9) as assessed by velocity sedimentation of Flag-tagged proteins (**Figure 3.2.5F**). Finally, since our main premise was that RNA-binding was required for the deamination-independent restriction by A3G, we tested the effect of the fusion proteins on the proviral DNA integration of HIVΔVif-eGFP by qPCR and determined that the mutants were now efficient (70 and 60% reduction in the levels of proviral integration respectively) at performing the deamination-

independent retroviral restriction and confirmed the necessity of RNA-binding by A3G for the restriction of the early steps of infection by the protein (**Figure 3.2.5G**).

3.2.6 RNA-binding deficient Vpr fusion mutants fail to achieve the deamination-independent restriction.

In order to confirm that the results obtained with vpr^{14-88}-fusion proteins were indeed due to their rescued RNA-binding properties, we deleted the RNA-binding residues of Vpr, amino acids 87 and 88, and fused the new polypeptide to Flag-tagged A2, A3G and its variants (287). The new RNA-binding deficient fusion proteins were then used for IP and RNA extraction followed by qPCR to confirm the lack of RNA-binding in the case of Vpr^{14-86}-W94A and Vpr^{14-86}-W127A. As shown in **Figure 3.2.6A**, the two mutants fused to the deleted Vpr protein showed reduced RNA-binding capacities similar to those observed in **Figure 3.2.1D** (**Figure 3.2.6A**). Nevertheless, fusion with Vpr^{14-86} was successful at enforcing packaging of the mutants in HIVΔVif-eGFP (**Figure 3.2.6B**). However, in this case, neither Vpr^{14-86}[W94A] or Vpr^{14-86}-[W127A] showed a significant reduction in eGFP fluorescence (**Figure 3.2.6C**) or integrated eGFP copies (**Figure 3.2.6D**). In conclusion, these results along with those of previous sections suggest that the RNA-binding properties of A3G are crucial for its deamination-independent restriction potential.

3.2.7 Absence of evidence for the requirement of a cellular co-factor in the deamination-independent restriction by A3G.

Here, we hypothesized that the requirement for RNA-binding in the deamination-independent restriction by A3G could be due to the involvement of a cellular co-factor interacting with the RNA-binding residues of the protein for its incorporation into viral particles

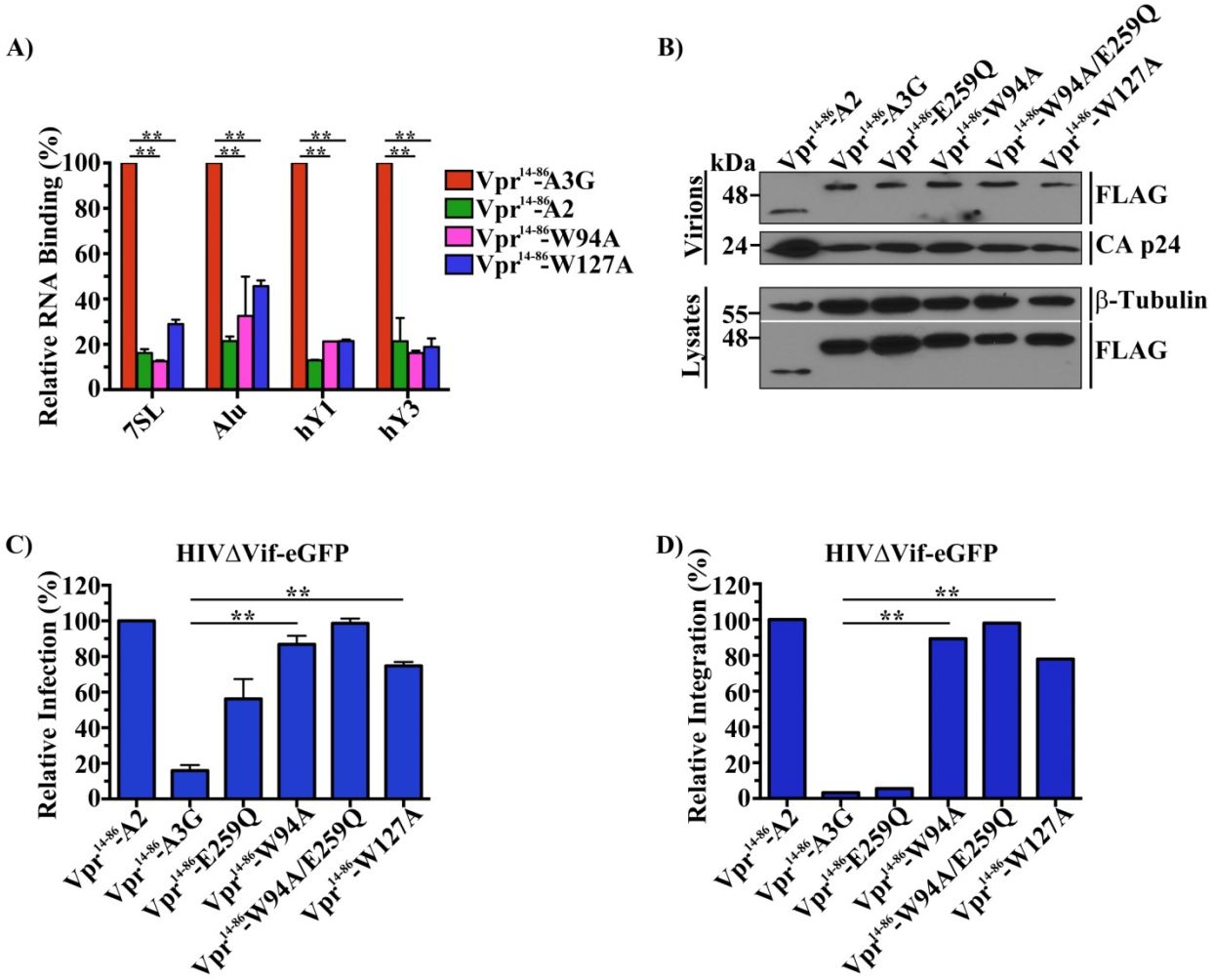
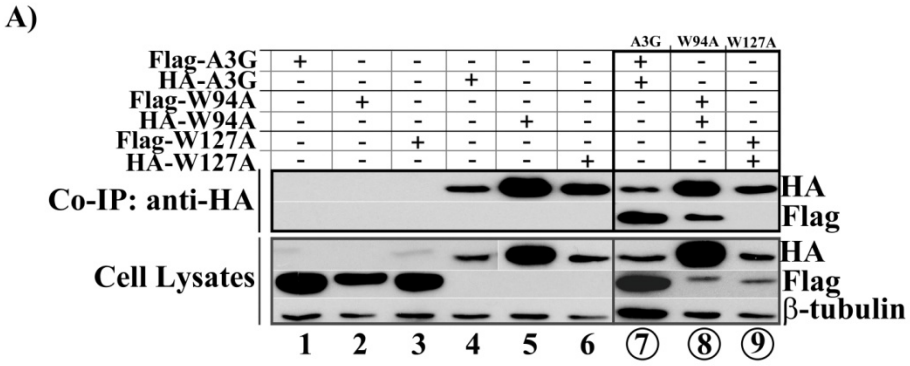
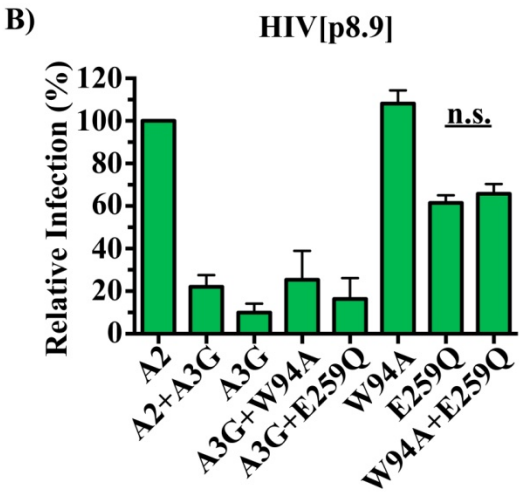
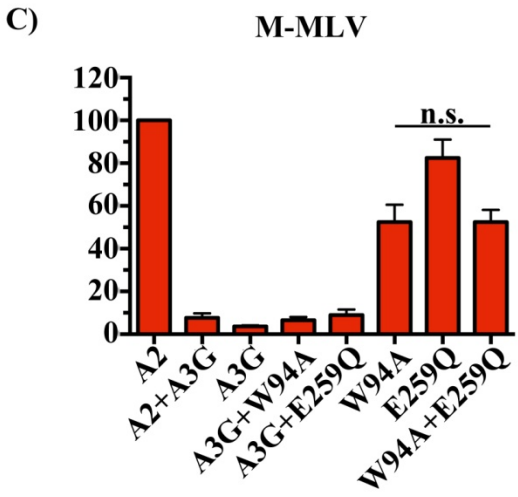

A)
Relative RNA Binding (%)
100
80
60
40
20
0
**
**
**
**
**
**
**
**
7SL
Alu
hY1
hY3
Vpr14-86-A3G
Vpr14-86-A2
Vpr14-86-W94A
Vpr14-86-W127A
B)
kDa
Vpr14-86-A2
Vpr14-86-A3G
Vpr14-86-E259Q
Vpr14-86-W94A
Vpr14-86-W94A/E259Q
Vpr14-86-W127A
Virions
48
24
FLAG
CA p24
Lysates
55
48
β-Tubulin
FLAG
C)
HIVΔVif-eGFP
Relative Infection (%)
120
100
80
60
40
20
0
**
**
Vpr14-86-A2
Vpr14-86-A3G
Vpr14-86-E259Q
Vpr14-86-W94A
Vpr14-86-W94A/E259Q
Vpr14-86-W127A
D)
HIVΔVif-eGFP
Relative Integration (%)
120
100
80
60
40
20
0
**
**
Vpr14-86-A2
Vpr14-86-A3G
Vpr14-86-E259Q
Vpr14-86-W94A
Vpr14-86-W94A/E259Q
Vpr14-86-W127A

Figure 3.2.6 RNA-binding deficient Vpr fusion mutants fail to achieve the deamination-independent restriction. (**A**) Lysates from 293T cells transfected with Vpr^{14-86}-fused A2, A3G or mutants were used for IP using anti-Flag-conjugated agarose beads followed by RNA extraction and RT-PCR. The RNA-binding ability of the proteins to 7SL, Alu, hY1 and hY3 was then determined by qPCR. Values are relative to A3G and are the mean +/- SD of three independent experiments. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005. (**B**) The incorporation of Vpr^{14-86}-fusion proteins into HIVΔVif-eGFP was determined by purifying viral particles produced in the presence of the APOBECs by ultra-centrifugation and then resuspending viral pellets in RIPA lysis buffer for Western blot analysis using an anti-Flag antibody. The level of p24 expression was used to ensure that similar amounts of viral lysates were loaded on gel. Protein expression was also assessed in cell lysates using β-tubulin as a loading control. (**C**) The restriction potency of the fusion proteins was determined 48 h after infection of 293T target cells with HIVΔVif-eGFP viruses containing the Vpr^{14-86} proteins by flow cytometry analysis of eGFP fluorescence. Results reflect the mean +/- SD of three independent experiments and are relative to A2. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005. (**D**) gDNA from HIVΔVif-eGFP-infected cells was collected 24 h after infection and used for qPCR to determine the levels of integrated eGFP. Values, relative to A2, represent the mean +/- SD of three independent experiments. . P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005.


A)
A3G
W94A
W127A
Flag-A3G
HA-A3G
Flag-W94A
HA-W94A
Flag-W127A
HA-W127A
Co-IP: anti-HA
HA
Flag
Cell Lysates
HA
Flag
β-tubulin
1 2 3 4 5 6 ⑦ ⑧ ⑨


B)
HIV[p8.9]
Relative Infection (%)
n.s.
A2
A2+A3G
A3G
A3G+W94A
A3G+E259Q
W94A
E259Q
W94A+E259Q


C)
M-MLV
n.s.
A2
A2+A3G
A3G
A3G+W94A
A3G+E259Q
W94A
E259Q
W94A+E259Q

3.2.7 Absence of evidence for the requirement of a cellular co-factor in the deamination-independent restriction by A3G. (**A**) Flag-tagged A3G, A3G[W94A] or A3G[W127A] were singly transfected or co-transfected with their HA-tagged equivalent in HEK 293T cells. Lysates were then used for the pull-down of HA-tagged proteins using anti-HA-conjugated magnetic beads and the presence of each variants in the eluates was determined by Western blot analysis using anti-Flag and anti-HA antibodies. (**B and C**) (**B**) HIV[p8.9] or (**C**) M-MLV viruses were produced in the presence of one or two APOBECs to package both proteins simultaneously and used to infect 293T or 3T3 target cells. The day after infection, levels of eGFP fluorescence were measured by flow cytometry. Data represent the mean +/- SD of three independent experiments and are relative to A2 alone. . P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; n.s. non-significant.

and its complementary role in the inhibition of late reverse transcript accumulation and proviral integration in target cells. In order to address that question, the W94A mutant was co-transfected with the catalytically inactive mutant of A3G, E259Q in 293T cells together with HIV[p8.9] or M-MLV viral expression vectors. Virus-containing supernatants were then used for infection of 293T or 3T3 target cells respectively two days after transfection and levels of infection, as measured by eGFP fluorescence, were determined by flow cytometry 24 h after infection. We wondered, since A3G[E259Q] is capable of achieving deamination-independent restriction, if it would also co-package the required cellular co-factor and complement the inhibition by the W94A mutant on retroviral infection thereby completely restoring its antiretroviral activity. The W94A mutant was chosen for this experiment because it is capable of homodimerization, unlike A3G[W127A] (**Figure 3.2.7A**). The results demonstrated no increase in the overall retstriction of A3G[W94A] against HIV[p8.9] or M-MLV in the presence of A3G[E259Q] suggesting that a co-factor for the deamination-independent restriction by A3G may not be required (**Figure 3.2.7B and C**).

3.3 Discussion

During a screen to identify the residues on A3G responsible for its recruitment into HMM complexes and its further catalytic inhibition in dividing cells, we identified two tryptophan residues in the N-terminal non-catalytic domain of the protein, W94 and W127, that when mutated to alanines, severely affected the RNA-binding properties of A3G. These residues have been the focus of previous studies looking at virion incorporation, homodimerization, nucleic acid interaction and cellular localization (171-173, 177, 179, 180, 185, 186, 280-283, 285). In agreement with our findings, both residues have previously been shown to have a significantly

reduced RNA-binding ability to Alu, 7SL and hY RNAs (172, 173, 186). In addition, mutants harbouring a mutation at amino acid position 127 were demonstrated to lack the ability to self-oligomerize and to get packaged into Vif-deficient HIV-1 particles (171, 177, 185, 283). Although we also showed defects in the encapsidation of both mutants into HIVΔVif-eGFP, significant amounts of the W94A and W127A proteins were found inside HIV[p8.9] and M-MLV virions indicating that the determinants for incorporation into these viruses are different than those required for packaging into HIVΔVif-eGFP. While the sequence of the region responsible for interacting with A3G for encapsidation is shared between the latter virus and HIV[p8.9], the pseudovirus also contains several non-HIV-1 sequences in its genome that could facilitate its association with the mutants.

An important finding of this section is that the deamination by A3G does not play a significant role in preventing late reverse transcripts accumulation and proviral DNA integration. One could then ask, 'what is the main role of deamination?' In an attempt to answer this question, we measured the levels of viral particles released in the presence or absence of A3G[W94A] and found that mutations induced by the protein did not affect the production of viral particles either. We believe that mutations generated by A3G would result in gene inactivation and that even integrated proviruses would not give rise to infectious particles. In fact, gene inactivation through A3G-mediated hypermutation is well documented in Chapter 5 of this thesis. Nonetheless, it is worth mentioning that retroviruses have a strong ability to undergo recombination events amongst themselves and mutated integrated genomes could potentially recombine with unmutated integrated genomes to produce infectious particles (288-293). This sheds light on the crucial role of the deamination-independent restriction by A3G and on the importance of uncovering the mechanisms underlying this mode of retroviral inhibition. It also

brings forward issues such as the real potency of deamination at inflicting permanent damages *in vivo* in the presence of a mixed population of viruses. We believe that a combination of both mechanisms need to act in concert in order to achieve an optimal restriction.

Finally, another important question that remains unsolved is the role of RNA in the deamination-independent restriction by A3G. One possibility is that the RNA bound to A3G could be responsible for the co-packaging of a cellular co-factor required for the inhibition of cDNA synthesis or proviral integration in the target cell. In order to address this question, we performed a viral infectivity assay in the presence of A3G[W94A] and A3G[E259Q] to try complementing the antiretroviral activity of the W94A mutant by the co-factor brought in by E259Q. However, we found no evidence for the presence of such co-factor. It is possible that structural features of both domains of A3G might be responsible for the phenotype and that the tryptophan mutations compromised the overall structure of the protein affecting the interaction with an RNA or co-factor required for restriction. The successful dimerization of A3G[W94A] and A3G[E259Q] should also be confirmed to ensure that both proteins can in fact co-package in the same viral particle. Differential proteomic analysis or analysis of RNA species present in virions containing A3G or A3G[W94A] could help solving the enigma.

In summary, this chapter established the requirement for RNA-binding by A3G for the deamination-independent restriction of retroviral infection by the protein and contributed to our understanding of this mode of restriction. Deamination induced by A3G did not prevent cDNA synthesis nor proviral DNA integration however, we believe that hypermutations in proviral genomes are necessary to inactivate gene function and limit the spread of progeny viruses. The results from this section emphasized the crucial contribution of both mechanisms of action by A3G to achieve a successful retroviral inhibition. Further studies are required to determine the

role of RNA in the deamination-independent retroviral restriction by A3G and whether or not a co-factor is required for A3G-mediated inhibition.

Chapter 4 The Role of RNA-Binding in the Deamination-Independent Restriction of L1 Retroelements by the Members of the APOBEC3 Family.

Preface: Data presented in this chapter are part of a manuscript that is currently in preparation for publication.

4.1 Introduction

Retroelements are mobile RNA sequences that produce copies of themselves that get inserted into new genomic location (124). There are two main classes of non-LTR retrotransposons: LINEs and SINEs. LINE-1 (L1) is an autonomous retrotransposon that encodes the machinery necessary for its own replication (124). This includes one protein with nucleic acid binding properties (ORF1) and another with reverse transcriptase and endonuclease activities (ORF2). Although most LINE elements in humans are inactive, it is thought that approximately 80-100 retrotransposition-competent elements are present in each human genome (294). The first step of retrotransposition takes place in the nucleus where the element is transcribed to produce an mRNA that is then processed for transport into the cytoplasm. Once in the cytoplasm, translation of ORF1 and ORF2 occurs which associate with other L1 mRNAs to form RNP complexes. These complexes are then imported back into the nucleus where target-primed reverse transcription (TPRT) is carried out. In this process, the bottom strand of the genomic DNA is cleaved by the endonuclease activity of ORF2 to yield a 3' hydroxyl group. The target DNA is used to prime reverse transcription from the L1 mRNA template. Cleavage of the coding strand of the genomic then initiates synthesis of the second-strand and the newly synthesized dsDNA is integrated at a new site into the host genome (see section 1.3.3 for details on L1 replication cycle) (124).

In humans, events associated with the replication of these elements have been linked to many genetic disorders such as hemophilia A and B and Duchenne muscular dystrophy occurring following integration of L1 DNA into germinal chromosomes (133, 295, 296). Targeting of somatic cells by L1 can also give rise to several types of cancer (Reviewed in (115)). It is therefore not surprising that we have evolved mechanisms to protect our genetic material against the threat posed by retroelements. One such mechanism is through the activity of the members of the A3 family of cytidine deaminases. The different members of this family have been shown to restrict the replication of L1 to different extent. While A3A, A3B, A3C and A3F have been shown by several independent groups to restrict L1 retrotransposition to varying degrees, restriction by A3G is less potent and somewhat even controversial (233-237, 239, 240, 297, 298). Initially, two reports demonstrated that A3G had no effect on the replication of the retroelement (239, 298). A year later, two more studies investigated the restriction of L1 by A3G and found that the latter was active against the retrotransposon (236, 297). Although the different groups differed slightly in their experimental settings, they all used similar L1- and A3-expression plasmids. Recently, Koyama and his colleagues established that the restriction of L1 retrotransposition by A3G was dependent on the oligomerization of the deaminase (299). They constructed N-terminal deletion mutants of A3G and found that the latter failed to inhibit L1 replication. Whether the oligomeric state of A3G could explain conflicting results obtained in previous studies is not known. Similarly, dimerization by A3C has been shown as a requirement for its impediment of L1 replication (235). RNA-dependent interaction of A3C dimers with the ORF1 protein was also involved in the process. Little is known about the restriction of L1 by A3D and A3H. However, restriction of this retroelement by A3H is believed to be highly dependent on the presence of two independent polymorphisms aiding the expression and stability

of the A3H protein (267, 300). Although the mechanisms underlying the restriction of retroelements by A3 proteins are not fully understood, evidences lean towards a deamination-independent mode of action (234, 235, 239). The recent discovery of G-to-A hypermutations in L1 cDNA in the presence of A3A, however suggests the involvement of the deaminase activity of the protein for its restriction (240). Nevertheless, deamination of retrotransposons by other members of the family has not been characterized.

Here, based on our previous observations that retroviral deamination-independent restriction by A3G requires the RNA-binding properties of the protein (Chapter 3 of this thesis), we hypothesized that the restriction of L1 retroelements by the A3 proteins is independent of deamination and also requires the RNA-binding properties of the deaminases. In agreement with our hypothesis, A3A which has a strong restrictive effect against L1, has been shown to interact with L1 mRNA in HMM complexes (297). In addition, the ability of A3C to inhibit L1 retrotransposition has been shown to be dependent on its RNA-dependent interaction with ORF1 (235). To test our hypothesis, we generated mutants for all members of the A3 family in which residues homologous to the RNA-binding residues identified in A3G (Chapter 3) were mutated to alanines. All the mutants were then analysed for their ability to interact with L1 mRNA and to inhibit L1 retrotransposition. Our preliminary results suggest that L1 mRNA binding is not a requirement for the restriction of the retroelement. No detectable levels of L1 mRNA-binding could be detected for any of the proteins tested, yet some members of the A3 family retained their ability to inhibit L1 replication. Other components involved in the retrotransposon's life cycle that are likely to be targetted by the A3 proteins to impede the reverse transcription process include constituents of the L1 RNP complexes such as ORF1 and ORF2 proteins.

4.2 Results

4.2.1 Schematic illustration of the mutations introduced in the seven members of the A3 family.

In Chapter 3 of this thesis, I identified two residues in the NTD of A3G that are essential for the ability of the protein to interact with RNA. This property of the protein was further demonstrated to be required for the deamination-independent restriction of HIV[p8.9] and M-MLV (**Figure 3.2.3**). Here, in view of identifying RNA-deficient mutants for each member of the human A3 family, residues homologous to the RNA-binding residues in A3G were mutated in all seven A3 proteins (**Figure 4.2.1A**). The two tryptophans identified in A3G are part of highly conserved motifs: S**W**S and ARLYY**W**. Corresponding tryptophans in the NTD of each A3 were mutated to an alanine (indicated by green arrows) by site-directed mutagenesis using primers described in **Table S1**. The W115A mutant of A3H was excluded from further analyses because of very poor expression (data not shown).

Figure 4.2.2 Characterization of the complexes formed by the wild type A3 proteins.

At least two members of the A3 family, A3F and A3G, are known to interact with RNA and protein co-factors to assemble into large ribonucleoprotein (RNP) complexes that are commonly called high molecular mass (HMM) complexes (215-218, 301). To assess the ability of each member of the family to form these HMM complexes, lysates from 293T cells transfected with Flag-tagged expression plasmids were treated for 15 min with RNase A at room temperature and spun for 6 h at 41 000 rpm through a 5 to 40 % sucrose gradient.

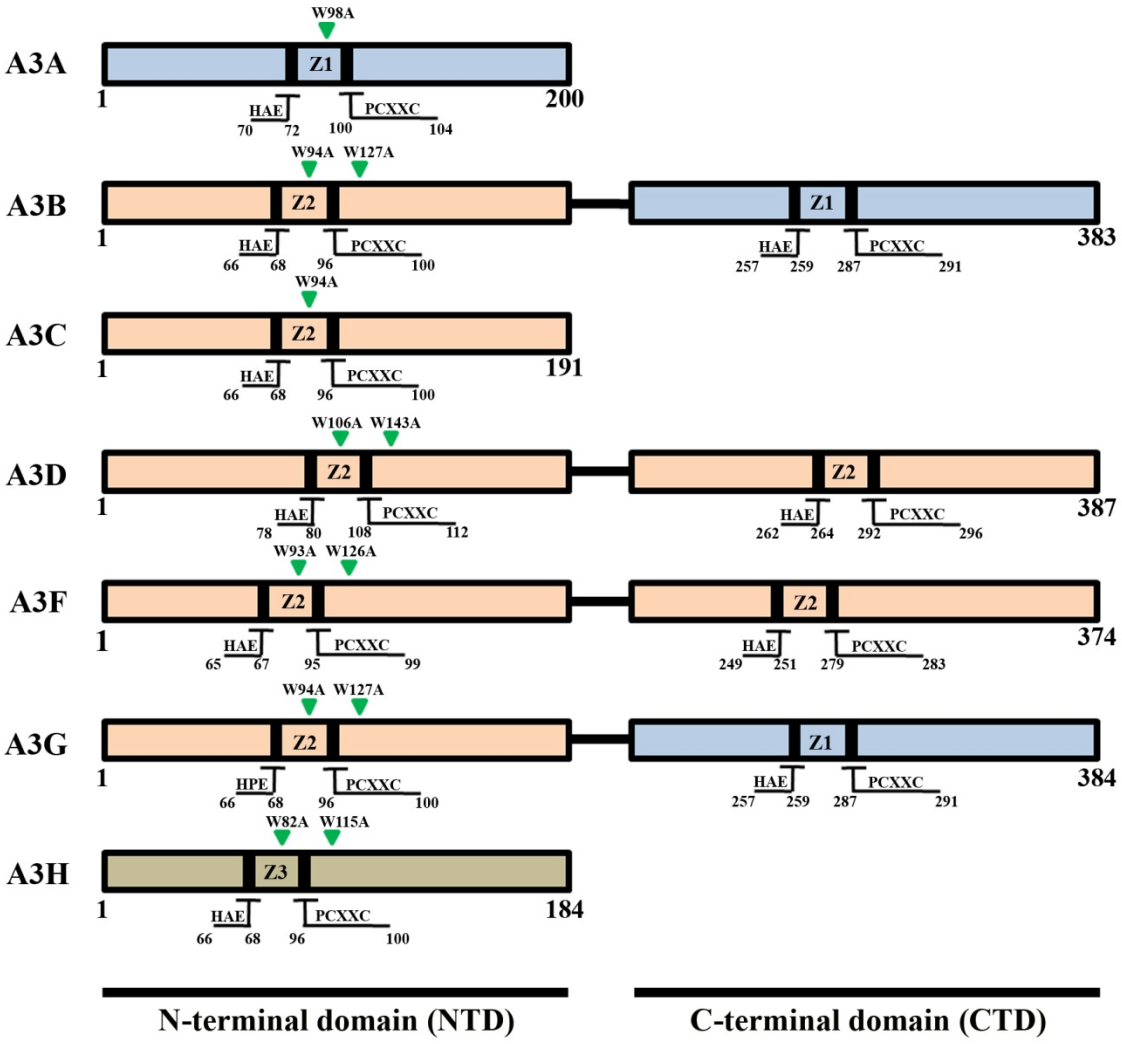

A3A
W98A
Z1
1
200
HAE
70
72
PCXXC
100
104
A3B
W94A
W127A
Z2
Z1
1
383
HAE
66
68
PCXXC
96
100
HAE
257
259
PCXXC
287
291
A3C
W94A
Z2
1
191
HAE
66
68
PCXXC
96
100
A3D
W106A
W143A
Z2
Z2
1
387
HAE
78
80
PCXXC
108
112
HAE
262
264
PCXXC
292
296
A3F
W93A
W126A
Z2
Z2
1
374
HAE
65
67
PCXXC
95
99
HAE
249
251
PCXXC
279
283
A3G
W94A
W127A
Z2
Z1
1
384
HPE
66
68
PCXXC
96
100
HAE
257
259
PCXXC
287
291
A3H
W82A
W115A
Z3
1
184
HAE
66
68
PCXXC
96
100
N-terminal domain (NTD)
C-terminal domain (CTD)

Figure 4.2.1 Schematic illustration of the mutations introduced in the seven members of the A3 family. Residues homologous to the RNA-binding tryptophans identified in the NTD of A3G were mutated to alanines in each member of the A3 family (green triangles).

Samples were then collected from the top to the bottom of the gradient and resolved by SDS-PAGE analysis followed by Western blotting using an anti-Flag antibody. In general, double-domain A3s assembled into HMM complexes (lanes 8 to 9) whereas single-domain proteins, with the exception of A3H, remained mostly in the first fractions of the sucrose gradient in monomeric, dimeric and tetrameric conformations (lanes 1 to 3) ((266, 284) and **Figure 4.2.2A**). A small proportion of A3C accumulated in all fractions of the gradient suggesting the ability of this protein to accumulate into different forms including the HMM complexes. A3H was the only A3 with only one domain that was found exclusively in HMM complexes.

Complexes formed by A3G contain several RNA species as well as RNA-binding proteins (215-217). As a consequence, the integrity of the complexes is highly dependent on RNA and treatment with RNase A results in the accumulation of the protein in its monomeric, dimeric and tetrameric forms (lanes 1 to 3) ((216, 266, 284, 301) and **Figures 3.2.2C and 4.2.2B**). The same was also observed for complexes formed by A3H suggesting the presence of RNA species in these complexes. Many RNA components have also been identified in complexes formed by A3F, however, the latter are more resistant to treatment with RNase A upon which there was only a partial shift in the sucrose gradient towards the middle fractions (lanes 4 to 7) where complexes of lower density accumulate ((216, 218, 297, 301) and **Figure 4.2.2B**). RNase A treatment had no effect on complexes formed by A3B, A3C and A3D suggesting that the composition of complexes formed by these proteins are different or that the conformation of the complexes generate a shelter protecting the RNA and preventing RNase from accessing its targets (297, 301). Further studies are required to fully understand the determinants for HMM complex formation by the different members of the A3 family.

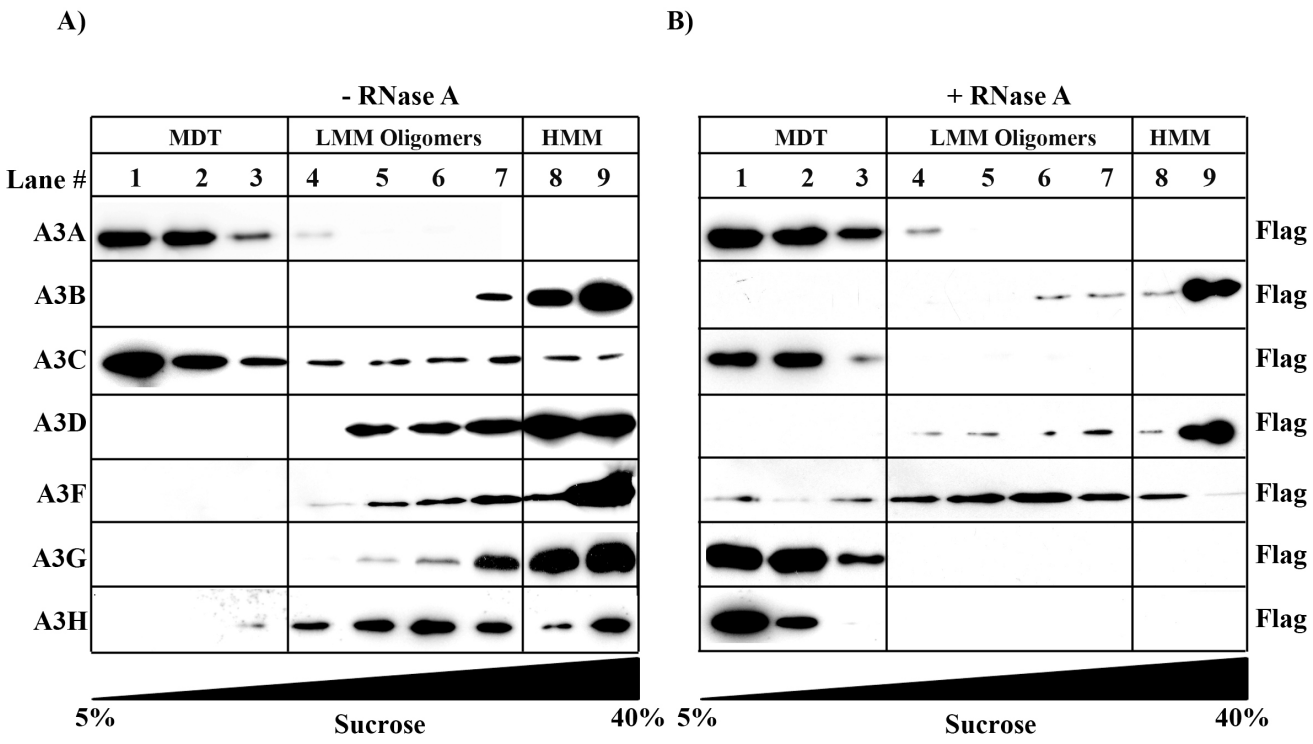

A)
B)
- RNase A
+ RNase A
MDT
LMM Oligomers
HMM
Lane #
1 2 3 4 5 6 7 8 9
A3A
A3B
A3C
A3D
A3F
A3G
A3H
Flag
5%
Sucrose
40%

Figure 4.2.2 Characterization of the complexes formed by the wild type A3 proteins. Lysates from 293T cells transfected with Flag-tagged A3 proteins were treated (right) or not (left) 15 min at room temperature with RNase A and layered on top of 5 to 40 % sucrose gradients for ultra-centrifugation 6 h at 41 000 rpm at 4 ºC. Samples were then collected from the top to the bottom of each gradient and resolved by SDS-PAGE using an anti-Flag antibody.

Figure 4.2.3 Effect of mutations on HMM complex formation by the A3 proteins.

In order to determine whether residues targeted in all seven A3 proteins were in fact involved in their RNA-binding properties, mutants were analysed for HMM complex formation by velocity sedimentation on 5 to 40 % sucrose gradients following lysis in NP40 lysis buffer and detected using an anti-Flag antibody by Western blot analysis. If the same residues involved in the RNA-binding ability of A3G were also involved in the same function in the other deaminases, we would expect to see a shift in the sucrose gradient towards the top fractions (lanes 1 to 3). This is what we observed for two mutants analyzed, A3D[W106A] and A3H[W82A] (**Figure 4.2.3A**). Mutations on conserved tryptophans in other A3s did not have an important effect on their ability to accumulate into HMM complexes.

Treatment with RNase A completely abolished the formation of complexes formed by A3D[W106A] and A3H[W82A] suggesting that, as it is the case for the W94A and W127A mutants of A3G, variants of A3D and A3H only have a partial reduction in their RNA-binding properties (**Figure 4.2.3B**). Complexes formed by A3B[W94A] appeared to be partially sensitive to the enzymatic treatment alluding to the fact that the conformation of the latter might prohibit the nuclease from reaching RNA whereas mutation of residue W94 might have induced a conformational change leading to a more open conformation allowing for some RNase to act on nucleic acids inside the complexes. This remains however to be confirmed.

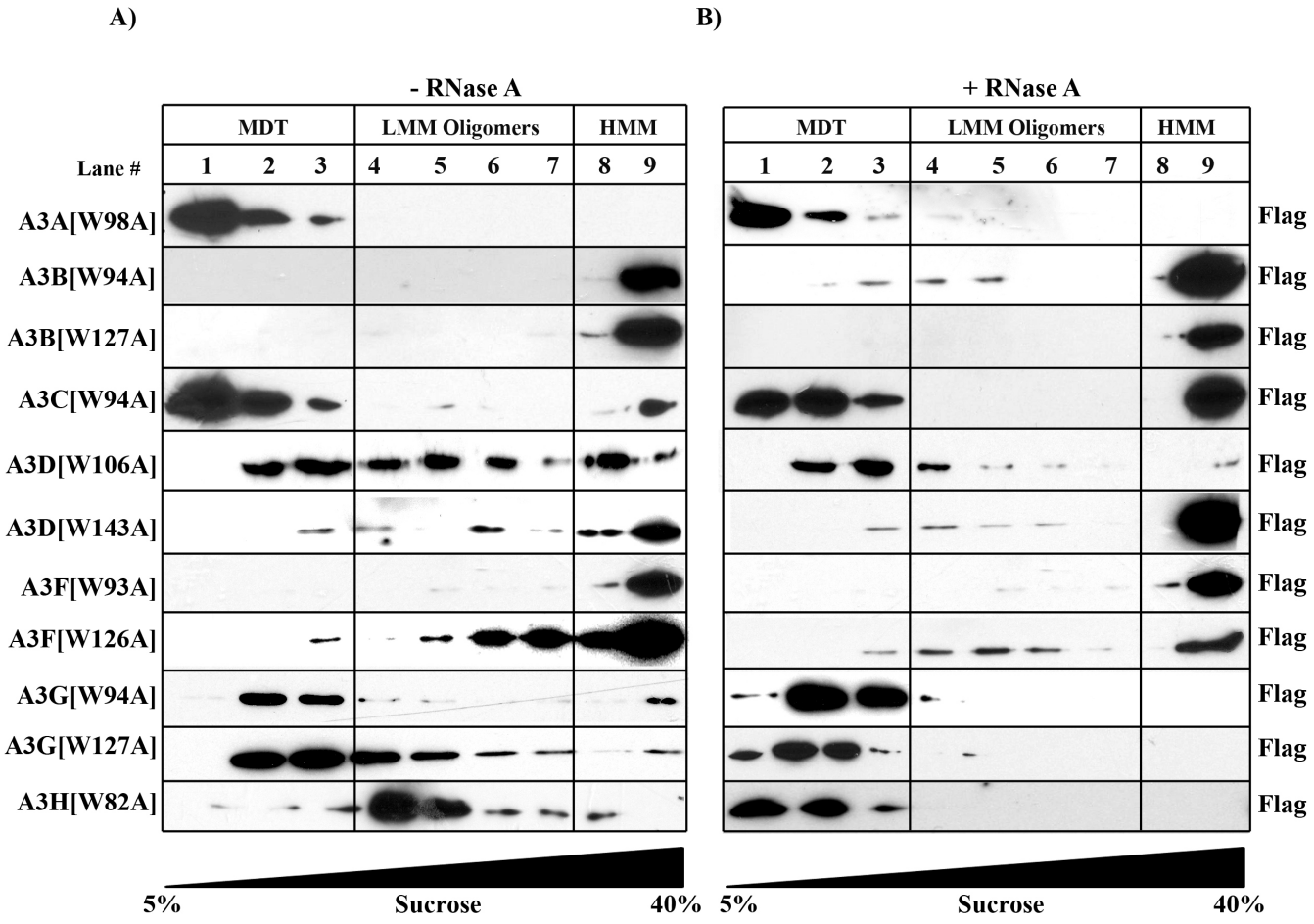

A)
B)
- RNase A
+ RNase A
MDT
LMM Oligomers
HMM
Lane #
1 2 3 4 5 6 7 8 9
A3A[W98A]
A3B[W94A]
A3B[W127A]
A3C[W94A]
A3D[W106A]
A3D[W143A]
A3F[W93A]
A3F[W126A]
A3G[W94A]
A3G[W127A]
A3H[W82A]
Flag
5%
Sucrose
40%

Figure 4.2.3 Effect of mutations on HMM complex formation by the A3 proteins. 293T cells expressing variants of each Flag-tagged A3 protein were lysed in NP40 lysis buffer for 30 min on ice followed by ultra-centrifugation for 6 h at 41 000 rpm at 4 ºC on 5 to 40 % sucrose gradients. After the spin, fractions from each gradient were collected and analyzed for the presence of Flag-tagged APOBECs by Western blotting using an anti-Flag antibody.

Figure 4.2.4 Analysis of the correlation between L1 mRNA binding and retroelement restriction.

Because an A3 protein does not assemble into HMM complexes does not necessarily imply that this protein does not have the ability to interact with a specific RNA such as L1 mRNA. Furthermore, it is not because there is no shift in the sucrose gradient with one mutant that its ability to bind a specific RNA such as L1 mRNA is not affected by the mutation. We therefore next sought to accurately measure the levels of binding of each wild type A3 proteins and variants to L1 transcripts. For that purpose, Flag-tagged A3 proteins and their mutants were transfected in 293T cells and cultured for 48 h. The amount of DNA transfected for each A3 was optimized in order to achieve similar levels of protein expression (**Table 2.1** and **Figure 4.2.4A**). Two days following transfection, the proteins were immunoprecipitated using anti-Flag-conjugated agarose beads and quantification of proteins (ρmol) in each sample was performed after Western blot analysis of the immunoprecipitates using a FITC-conjugated anti-Flag antibody. After papain treatment of the beads to release bound proteins, total supernatants were used for RNA extraction using Trizol followed by RT-PCR and absolute quantification of L1 mRNA (copies/ρmol) in each sample by digital droplet PCR (ddPCR). Since the ultimate goal was to uncover L1 RNA-binding deficient mutants, data are presented as variants compared to their wild type counterpart and are related to A2 in all cases (**Figure 4.2.4B**). In general, we found that none of the proteins that were tested had a significant level of binding to L1 transcripts and demonstrated copy numbers similar to the A2 negative control.

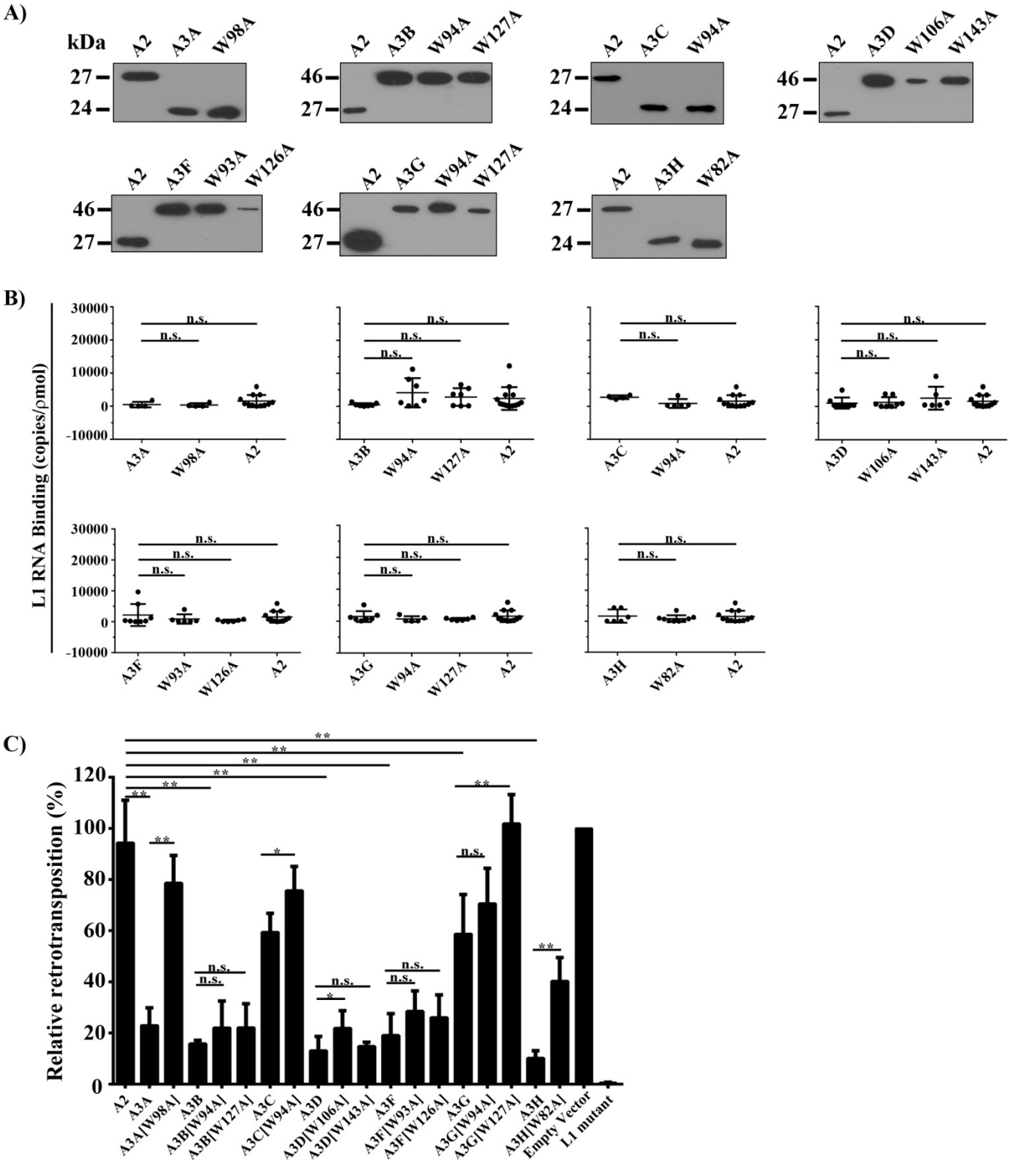

A)
kDa
A2
A3A
W98A
A3B
W94A
W127A
A3C
A3D
W106A
W143A
A3F
W93A
W126A
A3G
A3H
W82A
27
24
46
B)
L1 RNA Binding (copies/ρmol)
30000
20000
10000
0
-10000
n.s.
C)
Relative retrotransposition (%)
120
100
80
60
40
20
0
**
*
A3A[W98A]
A3B[W94A]
A3B[W127A]
A3C[W94A]
A3D[W106A]
A3D[W143A]
A3F[W93A]
A3F[W126A]
A3G[W94A]
A3G[W127A]
A3H[W82A]
Empty Vector
L1 mutant

Figure 4.2.4 Analysis of the correlation between L1 mRNA binding and retroelement restriction. (**A**) Varying amounts of Flag-tagged A3s and their mutants were transfected in 293T cells to achieve similar levels of expression and culture for 48 h. Two days after transfection, cells were harvested and lysed in NP40 lysis buffer for 30 min on ice followed by western blotting using an anti-Flag antibody. (**B**) Lysates from 293T cells expressing Flag-tagged wild type A3s or mutants of the proteins were subjected to IP using anti-Flag-conjugated agarose beads. Aliquots from immunoprecipitates were then collected for western blotting using a FITC-conjugated anti-Flag antibody and a protein standard to allow protein quantification in each sample whereas the remaining of the beads were treated with papain for complete elution of nucleic acids. Supernatants were then used for RNA extraction using Trizol reagents followed by RT-PCR and the cDNA was subjected to ddPCR analysis for absolute quantification of L1 mRNA binding. Bars represent mean values +/- SD of at least three independent transfections. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software. (**C**) 293T cells were co-transfected with the same amounts of Flag-tagged A3s and variants presented in (**A**) along with 800 ng of a L1-expression plasmid and cultured for seven days. Cells were then harvested in 1X PBS containing 1 mM EDTA and analysed for eGFP fluorescence by flow cytometry. Results represent the mean values of three independent transfections +/- SD. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; *: $P<0.05$; **$P<0.005$; n.s. non-significant.

Since some of the proteins assembled into HMM complexes whereas others remained in the first fractions of the sucrose gradient and they all demonstrated similar profiles in terms of RNA-binding ability, we concluded that there was no correlation between L1 mRNA binding and the formation of HMM complexes.

Finally, in order to determine the ability of each wild type A3 protein and their respective mutants to restrict L1 retrotransposition, a L1 expression plasmid containing an internal eGFP reporter cassette was co-transfected with varying amounts of Flag-tagged APOBECs (**Table 2.1**) (272). Co-transfection in the presence of an empty vector was used to control the effects of transfection on L1 expression whereas the L1 mutant was used as a negative control for L1 retrotransposition . Seven days following transfection, retrotransposition efficiency was assessed by flow cytometry analysis as a measure of eGFP fluorescence. Overall, all the wild type A3 were found to significantly reduce the levels of L1 retrotransposition with A3A, A3B, A3D, A3F and A3H having the strongest inhibitory effect on the element (**Figure 4.2.4C**). Although less is known about A3D and A3H, these results are in agreement with several previous reports demonstrating a potent restriction by A3A, A3B and A3F ((233, 234, 236, 237, 239, 240, 297, 298) and **Figure 4.2.4C**). The effect of A3C and A3G was more intermediate with 40% restriction in both cases. Our analyses also led to the observation that several mutants had a significantly reduced ability to restrict L1 replication when compared to their wild type counterpart (**Figure 4.2.4C**). This suggests that the mutants represent valuable tools to the elucidation of the mechanism behind the restriction of L1 retrotransposition. However, since none of the proteins showed binding to L1 mRNA but some members of the A3 family restrict the retrolement, we concluded that L1 mRNA-binding does not correlate with the restriction of retrotransposition.

4.3 Discussion

Through efforts to elucidate the mechanism for the restriction of L1 replication by the A3 proteins, we generated a panel of mutants with tryptophans mutated to alanines in two highly conserved motifs amongst members of the family (**Figure 4.2.1A**). Screening of the variants by velocity sedimentation revealed the identity of only two residues, W106 in A3D and W82 in A3H in addition to the two tryptophans previously identified in the NTD of A3G (**Figure 3.2.1B and C**), potentially involved in the RNA-binding properties of the proteins (**Figure 4.2.3A**). Mutation of conserved tryptophans in other A3s did not affect the oligomeirc state of the proteins. It is possible that the combination of more than one substitution would have a more pronounced effect on RNA-binding capability and thus, HMM complex assembly. However, based on previous observations with a double mutant of A3G (Chapter 5 of this thesis), such mutants are likely to affect the overall structure of the proteins resulting in poor expression. Residues in the CTD of members of the A3 family might also be involved in the RNA-binding properties of the proteins. In fact, contrary to A3G in which similar functions are accomplished by the NTD, Vif interaction as well as viral encapsidation are all attributed to the CTD of A3F (302). The CTD of A3D is also responsible for its interaction with Vif.

Nevertheless, more accurate measurements of the interaction of each APOBECs with L1 mRNA by ddPCR demonstrated that none of the proteins tested had the ability to bind the transcripts suggesting that there is no obvious correlation between L1 mRNA binding and HMM complex formation (**Figure 4.2.4B**). Similar observations have been made previously by another group who, in agreement with our results, showed that A3A which does not assemble into HMM complexes has the ability to interact with L1 mRNA (297). In the same report, it was established

that L1 restriction by A3 proteins was not associated with their ability to assemble into HMM complexes.

Although our initial attempt at identifying the mechanism involved in the restriction of L1 retrotransposition by the A3 family members was not as conclusive as we originally anticipated, our analysis of the L1 restriction potency of the mutants revealed the idendity of several proteins with impaired restriction properties (**Figure 4.2.4C**). These mutants can thus be used to dissect the characteristics of the proteins that are affected by the mutations and that could explain the lack of inhibition of retrotransposition. A lickely mechanism includes the interaction of A3 proteins with other components of the RNP complex formed by L1 mRNA, ORF1 and ORF2 in the cytoplasm as it has been suggested previously for A3C (235). Since the binding to ORF1 by different members of the A3 family has been demonstrated to be RNA-dependent, the requirement for RNA-binding in this mode of restriction by the enzymes cannot be excluded (235, 237). It is also possible that the different A3 proteins could each exhibit different modes of restriction against L1 retroelements. In fact, the restriction of L1 by A3C has been shown to be dependent on its interaction with ORF1 whereas binding to the protein is not required for restriction by A3A, A3B and A3F (235, 237). In addition, A3A is the only member of the family that has been demonstrated with the ability to induce G-to-A hypermutations in L1 cDNA during the reverse transcription process associated with retrotransposition (240). The different possibilities should therefore be carefully investigated for all seven members of the A3 family.

In summary, in this chapter, we provided the first characterization of the ability of all seven members of the A3 family to form HMM complexes and demonstrated that the determinants and amino acid residues governing their formation might be variable in each A3 protein. Through the generation of a large panel of A3 mutants, we identified variants with

reduced L1 restriction properties. Although L1 mRNA binding does not appear to be involved in the restriction of retrotransposition, the mutants generated in this chapter represent useful tools for the identification of the mechanisms used by the A3 proteins to restrict the retroelement.

Chapter 5 Substrate DNA Specificity of APOBEC3 Proteins influences their Gene Inactivation Potency

Preface: The results presented in this chapter have been published in two independent manuscripts. The first one covering the characterization of the RNA-binding mutant of A3G was published in:

Kasandra Bélanger and Marc-André Langlois. **(2015)** RNA-binding residues in the N-terminus of APOBEC3G influence its DNA sequence specificity and retrovirus restriction efficiency. ***Virology***, 483, 141-148.

*See Appendix V for the published article.

The second article describing the gene inactivation potency of A3 proteins based on their targeted motif for deamination was published in:

Kasandra Bélanger and Marc-André Langlois. **(2015)** Comparative Analysis of the Gene-Inactivating Potential of Human Retroviral Restriction Factors APOBEC3F and APOBEC3G. ***Journal of General Virology***, 96, 2878-2887.

* See Appendix VI for the published article.

5.1 Introduction

In general, G-to-A hypermutations caused by the A3 proteins are associated with the loss of genetic integrity and the production of non-infectious progeny viruses (reviewed in (101, 277), (278)). Nevertheless, sub-lethal levels of mutations induced by these enzymes have been shown to help the virus to evolve and increase its fitness through the acquisition of beneficial mutations leading to immune evasion and drug resistance (reviewed in (241), (242, 244, 249, 250, 258, 303)). This phenomenon is especially important *in vivo* when Vif is present since it induces the proteasomal degradation of the A3 proteins, resulting in reduced protein expression and consequently fewer proteins being incorporated into viral particles (reviewed in (101)).

A3 proteins display preferences for the cytidines they deaminate based on the local DNA sequence. Most notably, the base immediately 5' of the deaminated cytidine plays a very important role in substrate selection. A3G has a strong preference for deaminating cytidines preceded by a cytidine (5'CC), in contrast to the six other A3 proteins that prefer to deaminate in a 5'TC context, where the underlined C represents the targeted base for deamination (163, 181, 205-208). The extent of the damage induced by the A3 proteins is highly dependent on their target deamination motif (243, 245, 246, 304). On the positive DNA strand, the preferred target motif for A3G is read 5'GG as opposed to 5'CC on the minus DNA strand. Therefore, tryptophan codons (5'TGG) are highly targeted for deamination by the protein. In cases where a tryptophan is followed by another guanine 5'TGGG, two target sites are generated for A3G on the minus DNA strand (5'CCCA). If one or both cytidines are deaminated, the following termination codons TGA, TAG or TAA could potentially arise leading to gene inactivation. However, insertion of an inactivating mutation in a 5'TC context only occurs when a tryptophan codon is followed immediately by an adenine (5'TGGA). On the minus DNA strand, this sequence is read 5'TCCA and displays a single target motif for A3 proteins other than A3G. Mutation of this target site would result in the generation of a TGA stop codon. Here, we were interested in investigating the link between substrate specificity for deamination by A3 proteins and their gene inactivation potency.

This study emerged from the surprising observation that a double-mutant of A3G, A3G[W94A/W127A], completely lacking its RNA-binding capability, had an altered target motif for deamination. Firstly, we undertook an in depth characterization of the mutant in view of investigating whether the loss of RNA- binding by A3G affected other biological functions of the protein. Specifically, we asked whether the oligomeric state and the cellular localization of

the protein were affected and whether the protein could restrict retroviral infection and deaminate ssDNA. We demonstrated that mutation of tryptophans 94 and 127 resulted in the accumulation of the protein in its monomeric form whilst the cytoplasmic distribution of the protein remained unaltered. The W94A/W127A mutant failed to restrict retroviral infection while still retaining its ability to deaminate retroviral cDNA during reverse transcription. However, the intensity of deamination induced by the double-mutant of A3G was significantly reduced and more interestingly, its preferred DNA substrate was changed from 5'CC to 5'TC. Since the double-mutant was significantly less potent than A3G, in order to achieve the second goal of this study, we decided to turn our attention towards A3F which is highly mutagenic and like the RNA-binding A3G mutant, also mutates in a 5'TC context (181, 305). We found that 5'TC local dinucleotide preference resulted in the insertion of significantly fewer termination codons and that gene inactivation in this context was highly dependent on the intensity of hypermutation. In contrast, deamination of 5'CC DNA substrates frequently causes termination codons even at low mutation intensities. We conclude that at low mutation burden, mutations in 5'TC motifs are more likely to result in genetic diversification rather than gene inactivation. Thus, we provided evidence that the sequence specificity for deamination by the A3 proteins correlates with the amount of stop codons inserted in a gene and strongly influences gene inactivation potency by the enzymes.

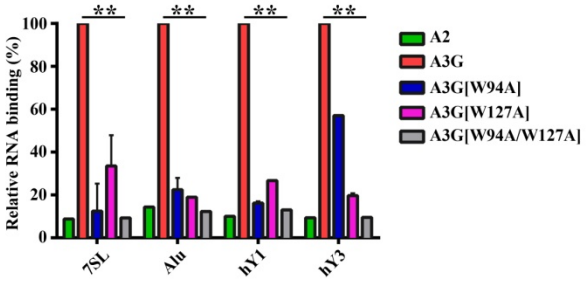
5.2 Results

5.2.1 The W94A and W127A mutations abrogate the RNA-binding properties of A3G and HMM complex formation with no effect on cellular foci.

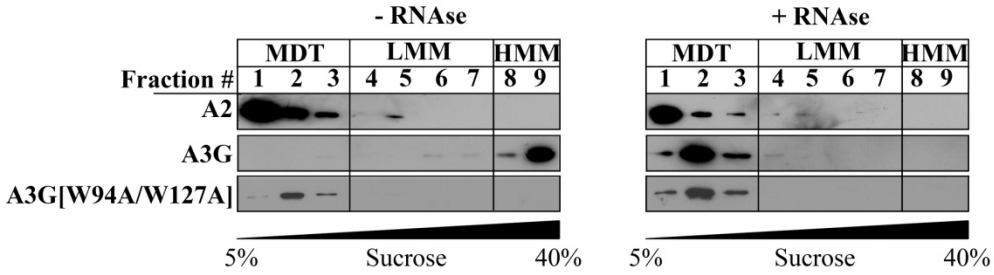
Having previously established that a mutant of A3G harbouring a single mutation at position 94 or 127 in its NTD had reduced RNA-binding abilities (**Figure 3.2.1D**), we were curious to see how a mutant containing both the W94A and the W127A mutations would interact with nucleic acids. We therefore performed a Co-IP experiment on lysates from 293T cells transfected with Flag-A2, Flag-A3G or Flag-tagged mutants of A3G using anti-Flag-conjugated agarose beads. The relative capacity of the mutants to bind 7SL, Alu, hY1 and hY3 RNAs compared to the wild type A3G was then measured by qPCR conducted on RNA isolated from immunoprecipitates. We found a 40-90% reduction in the RNA binding ability of both single mutants and a more pronounced 90-95% diminution in the case of the W94A/W127A mutant (**Figure 5.2.1A**).

In a previous chapter, we correlated the lack of RNA-binding by A3G with its inability to assemble into HMM complexes (**Figure 3.2.1B and C**). Those results were further confirmed by velocity sedimentation analysis of cells transiently expressing Flag-A2, Flag-A3G or Flag-A3G[W94A/W127A] showing the accumulation of the double-mutant in the first fractions of the sucrose gradient (lanes 1 to 3), representing the monomeric, dimeric and tetrameric forms of the protein (**Figure 5.2.1B**). Treatment of the samples with RNAse A for 15 min at room temperature led to the disruption of complexes formed by the wild type A3G whereas it had no effect on complexes containing A2 or A3G[W94A/W127A] confirming the absence of RNA-binding by the latter (**Figure 5.2.1B**). Complexes formed by A3G have been shown to accumulate as visible foci in the cytoplasm and have been postulated as P-bodies and SG (216, 217, 306). Since the double-mutant of A3G did not assemble into HMM complexes, we hypothesized that no cellular foci would be present in cells expressing the protein. In order to test our hypothesis, we transfected eGFP-tagged A3G or A3G[W94A/W127A] in 293T and cultured

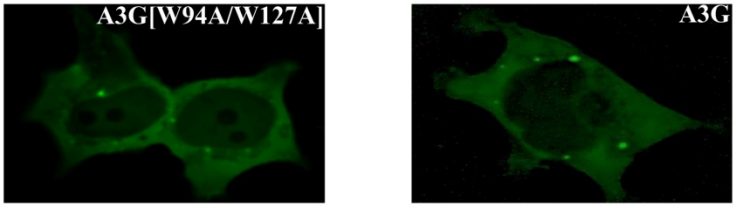
A)


**
**
**
**
100
80
60
40
20
0
Relative RNA binding (%)
7SL
Alu
hY1
hY3
A2
A3G
A3G[W94A]
A3G[W127A]
A3G[W94A/W127A]

B)


- RNAse
+ RNAse
MDT
LMM
HMM
Fraction #
1 2 3 4 5 6 7 8 9
A2
A3G
A3G[W94A/W127A]
5%
Sucrose
40%

C)


A3G[W94A/W127A]
A3G

D)

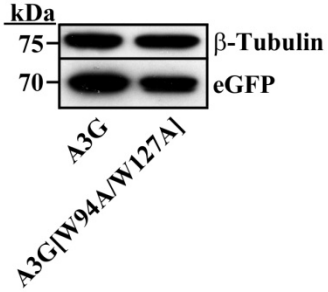

kDa
75
70
β-Tubulin
eGFP
A3G
A3G[W94A/W127A]

Figure 5.2.1 The W94A and W127A mutations abrogate the RNA-binding properties of A3G and HMM complex formation with no effect on cellular foci. (**A**) Lysates from 293T cells expressing Flag-tagged APOBECs were used for IP using anti-Flag-conjugated agarose beads. RNA was extracted from precipitates and used for qPCR analysis using primers specific for Alu, 7SL, hY1 and hY3 RNAs. Results were normalized to A2 and represent the mean +/- SD of three independent transfections. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005. (**B**) Flag-expressing plasmids were transfected in 293T cells and cultured for 48 h. Cells were then harvested and lysed in NP40 lysis buffer for 30 min on ice followed by treatment of half of the samples with RNase A for 15 min at room temperature. Lysates were submitted to ultra-centrifugation for 6 h at 4 ºC followed by resolution on SDS-PAGE gels and Western blotting using an anti-Flag antibody for detection. (**C**) eGFP-A3G or eGFP-A3G[W94A/W127A] were transfected in 293Ts in glass bottom dishes and imaging of live cells was performed using a Plan-Apochromat 63x/1.4 oil immersion objective on a Zeiss AxioObserver.Z1 microscope. (**D**) HEK 293T transfected with eGFP-tagged proteins were lysed in NP40 lysis buffer for 30 min on ice and used for Western blot analysis using anti-eGFP and anti-β-tubulin antibodies.

them for 48 h in clear media in glass bottom dishes. Imaging of live cells was then performed using an inverted fluorescence microscope. We demonstrated that for similar levels of protein expression, cellular foci could be observed in the cytoplasm of cells transfected with both the wild type A3G or the mutant of the protein with no obvious difference in size and a small reduction in abundance in the case of the W94A/W127A mutant (**Figure 5.2.1 C and D**). Since the W94A/W127A mutant did not accumulate into HMM complexes whereas cellular foci were still apparent in cells transfected with the protein, we believe that those complexes might be different from P-bodies and SG. Further studies are required to identify the nature and the components of these foci.

5.2.2 W94 and W127 residues are responsible for the creation of a charged pocket required for RNA-binding by A3G.

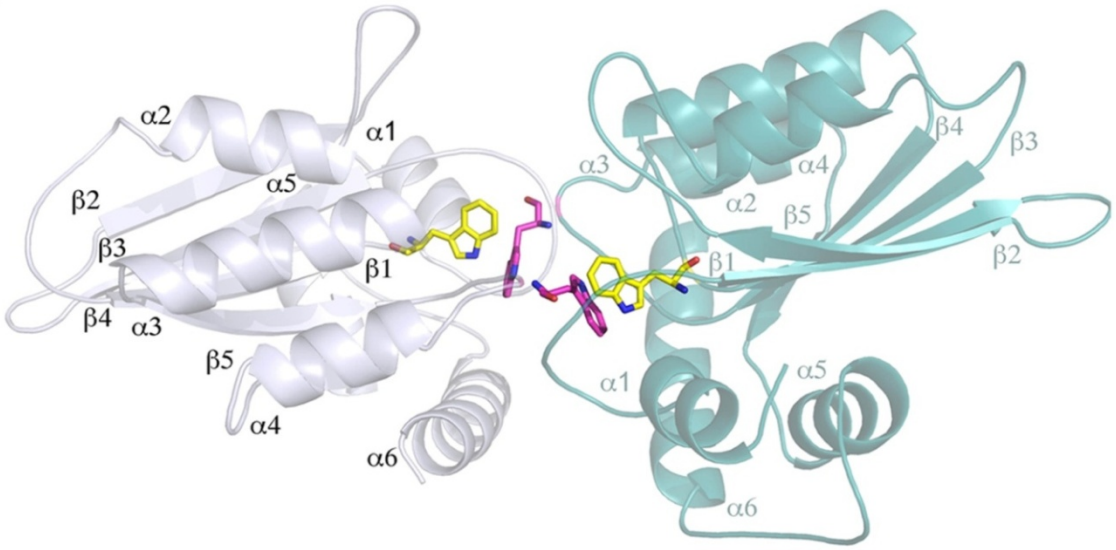
A3G's dimerization has been involved in many aspects of the protein's biology such as viral encapsidation, restriction of retroviral replication, inhibition of retroelement mobility and DNA editing (171, 175, 182, 299). Homology modeling of the A3G head-to-head NTD dimer conducted using pyMOL software revealed that, upon dimerization, the W94 residue from one monomer is in close proximity to the W127 residue from the other monomer resulting in the formation of a large positively charged surface available for RNA-binding (**Figure 5.2.2A and B**) (178). These findings involve an additional role for A3G's dimerization in RNA-binding. Since A3G[W94A/W127A] failed to bind RNA, we hypothesized that this might be due to the mutant's inability to self-associate. This prediction would be in agreement with earlier reports showing the involvement of W127 in protein's dimerization (**Figure 3.2.7A**) (171, 173, 307). To validate our hypothesis, we co-transfected Flag- and eGFP-tagged APOBECs and performed

a pull-down using anti-eGFP magnetic beads. The presence of Flag-tagged proteins in the precipitates was then determined by western blot analysis using an anti-Flag antibody. As expected, the double mutant of A3G failed to self-associate (**Figure 5.2.2B**). In summary, the results from this section indicate that the charged pocket generated upon A3G's dimerization by the proximity of the W94 and W127 residues could be required for the RNA-binding properties of the protein.

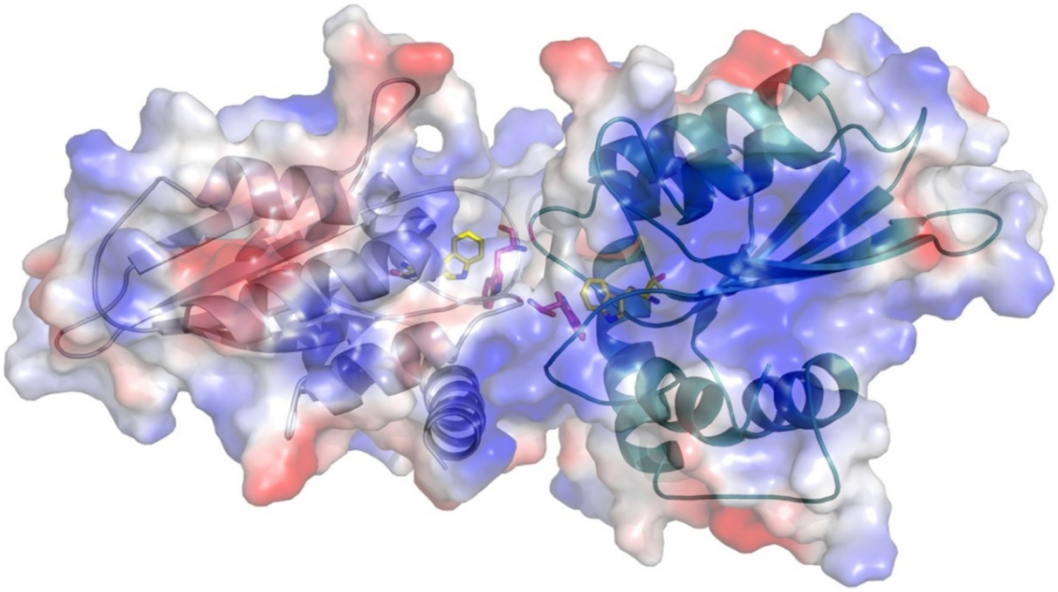
5.2.3 The binding to HIV-1 reverse transcriptase and integrase does not require binding to RNA.

A3G is known to interact with both HIV-1 reverse transcriptase and integrase in purified virions (230, 232). Although this remains to be confirmed, these interactions are thought to be responsible for the inhibition of viral DNA synthesis and proviral integration by the protein. A3G's association with HIV-1 reverse trancriptase has been shown to be independent of RNA whereas the interaction of the protein with the retroviral integrase was demonstrated to require its interaction with RNA (227, 230, 232). In Chapter 3 of this thesis, we demonstrated the requirement for RNA-binding for the deamination-independent restriction by A3G. If the binding to the reverse transcriptase and the integrase are involved in this process, both these interactions would be expected to require the RNA-binding properties of the protein. To elucidate this question, we produced HIV[p8.9] pseudotyped viruses in the presence or absence of Flag-A3G or mutants of the protein and performed a Co-IP using anti-Flag-conjugated agarose beads on whole cell lysates. Viral proteins were detected in immunoprecipitates by Western blotting. This assay revealed that a single substitution at position 94 or 127 in A3G did not affect the binding of the protein to HIV-1 reverse transcriptase and integrase nor did both mutations together (**Figure 5.2.3A**). In summary, the association of A3G with either viral protein appears to be RNA-

A)


α2
α1
α5
β2
β3
β1
β4
α3
β5
α4
α6
α3
α4
β4
β3
α2
β5
β2
β1
α1
α5
α6

B)



C)

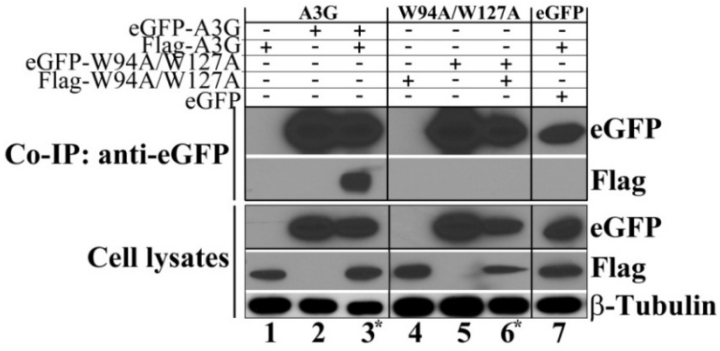

A3G
W94A/W127A
eGFP
eGFP-A3G
Flag-A3G
eGFP-W94A/W127A
Flag-W94A/W127A
eGFP
Co-IP: anti-eGFP
eGFP
Flag
Cell lysates
eGFP
Flag
β-Tubulin
1 2 3* 4 5 6* 7

Figure 5.2.2 The W94 and W127 residues are responsible for the creation of a charged pocket required for RNA-binding by A3G. (**A**) Representation of the secondary structure of the NTD of A3G in its dimeric form. The W94 (yellow) and W127 (pink) residues of each monomer are indicated. (**B**) Distribution of the charges in our model of the head-to-head dimer formed by A3G with the positively charged regions in blue and the negatively charged regions in red. The homology model was generated using the EasyPred3d and the figures were generated using PyMOL (http://pymol.sourceforge.net/). (**C**) Flag-A3G or variant were transfected alone (lanes 1, 2, 4 and 5) or in the presence of their eGFP equivalent (lanes 3 and 6) in 293T cells. Pull-downs were performed on lysates of transfected cells 48 h after transfection using anti-eGFP magnetic beads followed by Western blotting using anti-Flag and anti-eGFP antibodies. β-tubulin expression was used as a loading control. Co-transfection of Flag-A3G with an eGFP empty vector (lane 7) was included to control for unspecific binding of the APOBECs to the beads.

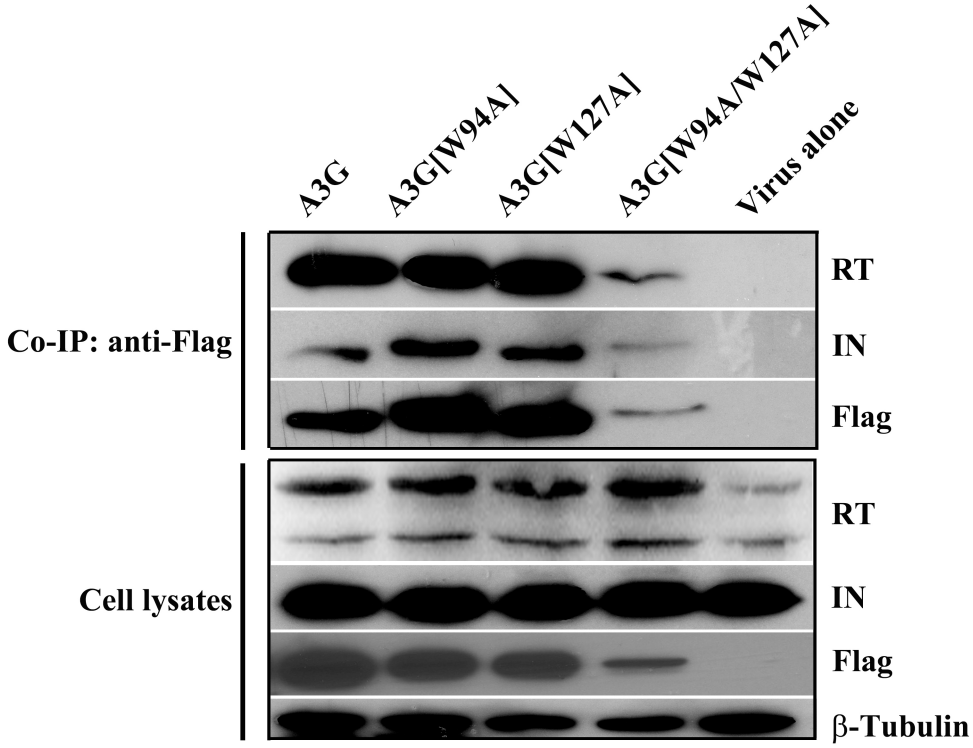
independent and therefore cannot explain the deamination-independent antiretroviral activity of the protein.

In agreement with these findings, we showed that A3G[W94A] still interacted with both retroviral proteins in purified virions although this mutant has previously been shown to fail to inhibit late reverse transcripts accumulation and proviral integration (**Figure 5.2.3B and Figure 3.2.4A to D**).

5.2.4 A3G[W94A/W127A] is incorporated into viral particles and mutates single-stranded DNA.

Finally, a feature of A3G that is essential for its antiretroviral activity is its incorporation into viral particles from virus-producing cells (reviewed in (277)). This property of the protein has been linked to its ability to interact with RNA (172, 175, 186). In order to determine whether the W94A/W127A efficiently retained its packaging into HIV[p8.9], viruses produced in the presence of eGFP-A2, eGFP-A3G or mutants of A3G were purified by ultra-centrifugation, lysed and analyzed for the presence of eGFP-expression plasmids by Western blotting. As previously described, the wild type A3G and all the single-mutants of the protein were efficiently detected inside viral particles (**Figure 5.2.4A** and **Figure 3.2.3B**). Surprisingly, the W94A/W127A mutant, despite lacking ability to bind RNA was also found to be incorporated into retroviruses at similar levels to the wild type A3G (**Figure 5.2.4A**). However, restriction assays demonstrated that just like the W94A and the W127A mutants, the double mutant had no antiretroviral activity against HIV[p8.9] (**Figure 5.2.4B**). Given that we previously found (Chapter 3) deamination without restriction to be a potential phenotype, we next asked whether A3G[W94A/W127A] could induce mutations in ssDNA (308). We first conducted a bacterial mutator assay in which the frequency of resistance to rifampicin (Rif^R) was measured in *E. coli* (154). In this assay, we

A)


A3G
A3G[W94A]
A3G[W127A]
A3G[W94A/W127A]
Virus alone
Co-IP: anti-Flag
RT
IN
Flag
Cell lysates
RT
IN
Flag
β-Tubulin

B)

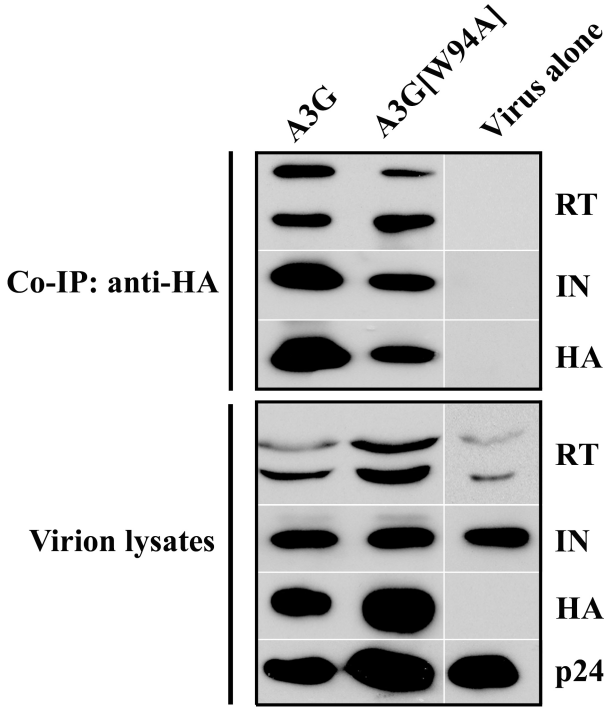

A3G
A3G[W94A]
Virus alone
Co-IP: anti-HA
RT
IN
HA
Virion lysates
RT
IN
HA
p24

Figure 5.2.3 The binding to HIV-1 reverse transcriptase and integrase is direct and does not require binding to RNA. (**A**) HIV[p8.9] particles were produced in the presence of Flag-A3G or Flag-tagged mutants of A3G by co-transfection in 293T cells. Cells were collected 48 h after transfection and lysed in NP40 lysis buffer fo 30 min on ice. Flag-tagged proteins were immunoprecipitated using anti-Flag-conjugated agarose beads and following high-pH elution of the proteins from the agarose beads, precipitates were analyzed by western blot using anti-Flag, anti-IN and anti-RT antibodies. The expression of β-tubulin was used as a loading control. (**B**) HIV[p8.9] particles were produced in the presence of HA-A3G or mutant of A3G by co-transfection in 293T cells. Viral particles were purified by ultra-centrifugation and lysed in NP40 lysis buffer for 30 min on ice. Lysates were subjected to IP using anti-HA magnetic beads. Following high-pH elution, precipitates were analyzed by Western blot using anti-HA, anti-IN and anti-RT antibodies and the expression of p24 was used as a loading control.

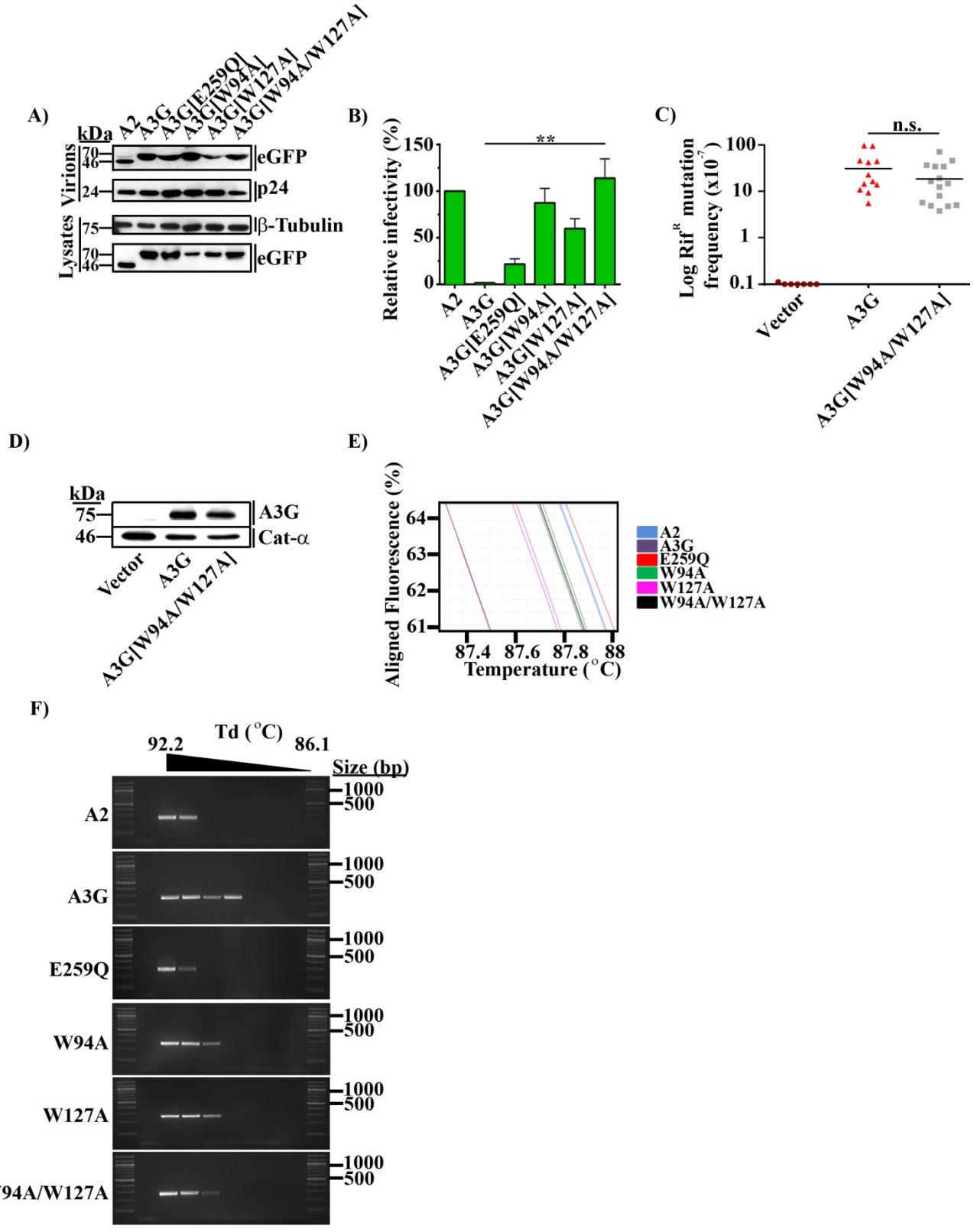

A)
kDa
A2
A3G
A3G[E259Q]
A3G[W94A]
A3G[W127A]
A3G[W94A/W127A]
Virions
70
46
eGFP
24
p24
Lysates
75
β-Tubulin
70
46
eGFP
B)
Relative infectivity (%)
150
100
50
0
**
A2
A3G
A3G[E259Q]
A3G[W94A]
A3G[W127A]
A3G[W94A/W127A]
C)
Log Rif^R mutation frequency (x10^-7)
100
10
1
0.1
n.s.
Vector
A3G
A3G[W94A/W127A]
D)
kDa
75
46
A3G
Cat-α
Vector
A3G
A3G[W94A/W127A]
E)
Aligned Fluorescence (%)
64
63
62
61
87.4
87.6
87.8
88
Temperature (°C)
A2
A3G
E259Q
W94A
W127A
W94A/W127A
F)
Td (°C)
92.2
86.1
Size (bp)
A2
A3G
E259Q
W94A
W127A
W94A/W127A
1000
500

Figure 5.2.4 A3G[W94A/W127A] is incorporated into viral particles and mutates single-stranded DNA. (**A**) HIV[p8.9] viral particles produced in the presence of eGFP-A3G or mutants of A3G were purified by ultra-centrifugation and lysed for 15 min on ice in RIPA lysis buffer. The presence of eGFP-expression plasmids in viral lysates was then assessed by Western blot analysis using an anti-eGFP antibody. Virus-producer cells were also harvested and lysed in NP40 lysis buffer for 30 min on ice prior to Western blotting. The amount of cell lysates loaded was normalized using the expression of β-tubulin whereas the amount of viral lysates in each sample was normalized using p24. (**B**) Restriction of HIV[p8.9] infection by eGFP-A3G or mutants as measured by flow cytometry analysis of eGFP fluorescence in 293T target cells 24 h after infection. Results represent the mean ± SD of triplicate values from three independent transfections. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; **P<0.005. (**C**) The ability of A3G, A3G[W94A/W127A] or an empty vector control to trigger mutations conferring resistance to rifampicin in *E. coli* were evaluated. Each point on the graph represents the log of the mutation intensity (Rif^R mutants per 10^7 viable cells) in an individual culture with the horizontal lines indicating the mean values. P-values were calculated using the Student's paired *t*-test using the GraphPad Prism software; n.s. non-significant. (**D**) pGST-expression plasmids were produced in *E.coli* in the presence of 0.1 mM IPTG inducer. Cells were then lysed and assessed for the presence of APOBEC plasmids by Western blot analysis using an anti-A3G antibody. The expression of the Cat-α bacterial housekeeping protein was used as a control to ensure that equal amounts of lysates were loaded. (**E and F**) HIV[p8.9] viral particles produced in the presence of eGFP-tagged APOBECs were used to infect 293T target cells. gDNA was extracted from infected cells 24 h after infection and used for (**E**) HRM analysis or (**F**) 3D-PCR analysis.

observed no significant difference in the number of Rif^R colonies in the presence of similar amount of the double mutant of A3G or the wild type protein, alluding to the fact that the W94A/W127A mutant fully retained its enzymatic activity (**Figure 5.2.4C and D**). We then looked for mutations in proviral DNA by extracting gDNA from 293T infected cells 24 h after infection with HIV[p8.9] viruses produced in the presence of eGFP- A2, eGFP-A3G or mutants of A3G and performing HRM or 3D-PCR analyses (**Figure 5.2.4E and F**). Our results demonstrated the presence of hypermuations with the double mutant that were comparable to those observed in the presence of the W94A mutant and slightly lower than what was observed with A3G[W127A]. A section of the eGFP reporter gene of HIV[p8.9] was sequenced and analysis of the DNA context specificity for deamination revealed a strong preference for the targeting of 5'TC dinucleotides in the case of the W94A/W127A mutant as opposed to the 5'CC observed with the wild type A3G (**Table 5.2.1**). Taken together, these findings indicate that the W94A/W127A mutant is still incorporated into viral particles and has the ability to induce hypermutations in ssDNA however, the intensity of deamination generated by the protein is reduced and most remarkably, its preferred deamination motif is altered.

Table 5.2.1 Analysis of the preferred DNA substrate for deamination by A3G[W94A/W127A] in HIV[p8.9]*.

W94A/W127A	-2	-1
A	24	2
C	26	6
T	26	**92**
G	24	-

*Mutations were tabulated on the minus-strand retroviral DNA. Data represents the percentage of each base at position -2 and -1 relative to the deaminated cytidine at position 0.

5.2.5 Mutations induced in both a 5'CC and 5'TC context have the potential to result in genetic diversification.

Here, we were curious to determine whether the new DNA motif targeted for deamination by A3G[W94A/W127A] could affect the intensity of gene inactivation. However, since the intensity of deamination induced by the mutant was lower in our viral infection assays compared to the wild type A3G, those two proteins could not be directly compared (**Figure 5.2.4E and F**). We therefore decided to use A3F since this protein also has a preference for 5'TC dinucleotide substrates and is highly mutagenic (181, 305). For that section, we turned our attention towards M-MLV in which the eGFP reporter protein is embedded within the envelope glycoprotein (Env). The Env sequence upstream of the eGFP coding sequence contains 72 potential deamination sites with 13 likely to cause termination codons whereas the Env sequence downstream of the eGFP-coding sequence contains 70 potential target sites for deamination that could yield 5 termination codons. The eGFP coding sequence that is expressed by HIV[p8.9] and M-MLV contains a single site that could generate a stop codon. If any of those sites are mutated by A3G, resulting in the generation of a premature stop codon in the Env or eGFP protein, it would prevent the expression of the eGFP reporter protein in the target cells. We therefore produced M-MLV or HIV[p8.9] viral particles in the presence of Flag-tagged A3G or A3F and infected 293T or 3T3 target cells 48 hours after transfection. gDNA from infected cells was then extracted and the Env-eGFP fusion gene from M-MLV or the eGFP gene from HIV[p8.9] was amplified by PCR. PCR products were then cloned upstream of a CMV promoter and a total of 21 clones for each conditions were analyzed for G-to-A mutations using Sanger sequencing (**Figure 5.2.5A**). Overall, we found that the mutation rate induced by A3G and A3F was highly dependent on the availability of their respective target site. In eGFP in which the number of

target sites for both proteins is equal, the mutation rate induced by both deaminases was similar with A3G being slightly more potent (**Table 5.2.2**). Whereas in Env-eGFP, since there are more target motifs for A3G than A3F, the latter had a significantly lower mutation rate. (**Table 5.2.2**).

Table 5.2.2 Analysis of the intensity of mutations induced by A3F and A3G in all clones analyzed.

A3	Gene analyzed	Number of target sites in gene analyzed	Base pairs analyzed in mutated clones	Total number of G-to-A mutations	Mutation rate (mutations/ kb)
A3F	eGFP	52	14400	118	8.2
A3F	Env-eGFP	167	54060	293	5.4
A3G	eGFP	51	14400	178	12.4
A3G	Env-eGFP	193	54060	970	18.0

Mutations compiled for all clones analyzed for both A3G and A3F were then indicated as codons affected in the Env-eGFP gene (**Figure 5.2.5C and D**). Amongst 167 5'CC and 163 5'TC possible target motifs, A3F deaminated 144 different codons with 63% of all mutations being in a 5'TC context and 13% in a 5'CC context (**Figure 5.2.5B and C**). Other mutations occurred on 5'RC substrates where R is either A or G. In the case of A3G, 187 different codons were affected by mutations generated by the protein with 65% in a 5'CC context, 19% in a 5'TC context and 16% in a 5'RC context (**Figure 5.2.5D**). In addition, we observed significantly more termination codons in the presence of A3G, as represented by red bars in the Env-eGFP gene. However, there was no noticeable difference between the number of non-synonymous mutations introduced by the two proteins, suggesting that they have an equal potential to contribute to genetic diversity (**Tables 5.2.3 and 5.2.4**).

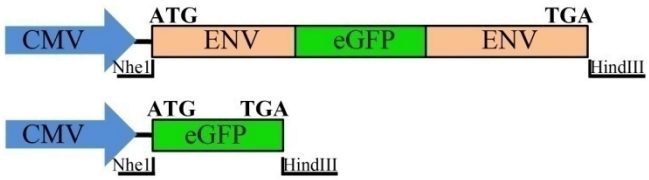
Table 5.2.3 Frequency of non-synonymous substitutions induced by A3F and A3G in eGFP.

A3	Total codons deaminated in all clones analyzed	Total non-synonymous substitutions	Proportion of non-synonymous substitutions (%)
A3F	118	38	32
A3G	184	83	45

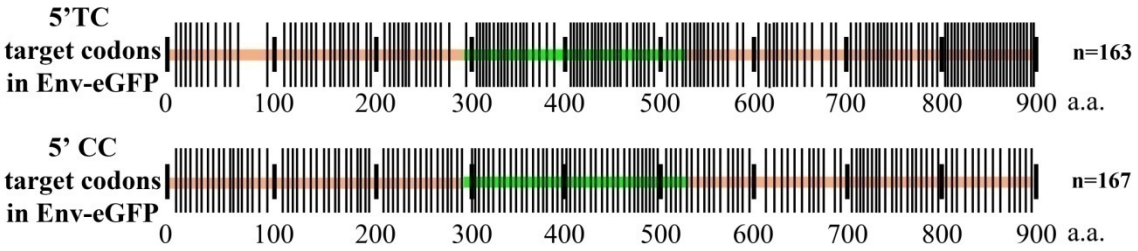
Table 5.2.4 Frequency of non-synonymous substitutions induced by A3F and A3G in Env-eGFP.

A3	Total codons deaminated in all clones analyzed	Total non-synonymous substitutions	Proportion of non-synonymous substitutions (%)
A3F	283	201	71
A3G	870	567	65

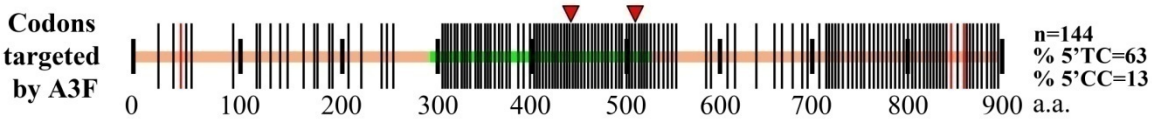
A)


ATG
TGA
CMV
ENV
eGFP
ENV
NheI
HindIII
ATG
TGA
CMV
eGFP
NheI
HindIII

B)


5'TC
target codons
in Env-eGFP
n=163
0
100
200
300
400
500
600
700
800
900
a.a.
5' CC
target codons
in Env-eGFP
n=167
0
100
200
300
400
500
600
700
800
900
a.a.

C)


Codons
targeted
by A3F
n=144
% 5'TC=63
% 5'CC=13
0
100
200
300
400
500
600
700
800
900
a.a.

D)

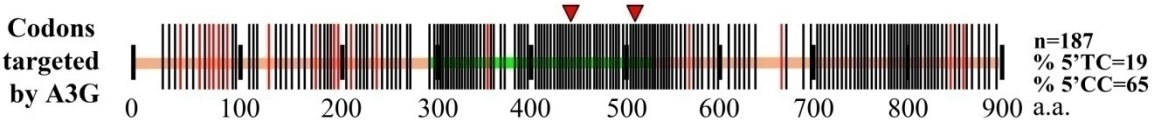

Codons
targeted
by A3G
n=187
% 5'TC=19
% 5'CC=65
0
100
200
300
400
500
600
700
800
900
a.a.

Figure 5.2.5 Mutations induced in both a 5'CC and 5'TC context have the potential to result in genetic diversification. (**A**) Schematic representation of the cloning strategy used to insert the eGFP or ENV-eGFP genes from HIV[p8.9] or M-MLV respectively into the pcDNA3.1 vector for expression under the control of an internal CMV promoter. Orientation of the arrows is representative of the sense of transcription with initiation (ATG) and termination (TGA) codons indicated. (**B**) Mapping of 5'TC (top) or 5'CC (bottom) target codons (represented as black bars) in the Env-eGFP gene. Total numbers of target sites are indicated to the right of the graph. (**C and D**) Distribution of amino acid codons affected by G-to-A mutations induced by (**C**) A3F or (**D**) A3G in the Env-eGFP gene from M-MLV. Red bars represent termination codons, inverted red triangles represent the position of inactivating mutations (E133K or E223K). The total number of targeted amino acids is represented to the right of the graph with the proportion of mutations generated in a 5'TC and a 5'CC context.

5.2.6 Mutation of 5'CC target motifs results in the incorporation of termination codons leading to gene inactivation whereas deamination of 5'TC motifs favours genetic diversification.

Mutated clones produced in the presence of A3F or A3G were transfected in 293T cells and cultured for 48 h after which the intensity of eGFP fluorescence was measured by flow cytometry. We found that deamination by both A3G and A3F did not result in the introduction of termination codons in the eGFP gene from HIV[p8.9] (**Figure 5.2.6A and Tables 5.2.5 and 5.2.6**). EGFP contains only a single TGG tryptophan codon. However, mutation of critical residues, E133K and E222K resulting from a G-to-A mutation introduced in a 5'TC context were identified in the presence of both A3s (indicated by an inverted triangle in **Figure 5.2.6**) as deleterious mutations resulting in the absence of detectable fluorescence (**Tables 5.2.9 to 5.2.12**). Other substitutions that resulted in reduced fluorescence intensity were also observed for several other clones of both A3F and A3G. Although these clones do not share mutation profiles that could explain the phenotype, some have a unique alteration including C49Y, D83N, A111T, E214K, D217N and M219I. It is conceivable that one or more of these mutations could affect the fluorescence of eGFP, alone or in combination with other substitutions. Finally, as expected, for both A3F and A3G, the intensity of fluorescence decreased as the mutation burden increased.

Deamination in a 5'CC context by A3G resulted in the insertion of significantly more termination codons in Env-eGFP with all clones mutated being inactivated through the incorporation of a stop codon whereas only 4 clones were affected by termination codons when mutations were generated by A3F in the same gene (**Figure 5.2.6B and Tables 5.2.7 and 5.2.8).** In most cases, introduction of stop codons by A3F was the result of deamination outside its usual consensus (mutation of a 5'CC dinucleotide as opposed to a 5'TC substrate). Gene inactivation in the presence of A3F was more often the result of high mutation burdens and occurred in the

Table 5.2.5 Analysis of G-to-A mutations induced by A3F in the eGFP gene from HIV[p8.9].

Clones	Total number of G-to-A mutations	Total number of termination codons	Relative Mean Fluorescence (%)*	Standard Deviation (SD)
A3F-1	0	0	98.9	3.2
A3F-2	1	0	58.9	1.8
A3F-3	2	0	9.3	3.2
A3F-4	2	0	101.8	4.3
A3F-5	2	0	1.5	0.1
A3F-6	2	0	103.0	5.5
A3F-7	3	0	114.8	5.5
A3F-8	3	0	103.3	6.6
A3F-9	4	0	103.6	5.7
A3F-10	4	0	0.0	0.0
A3F-11	5	0	71.8	7.2
A3F-12	5	0	35.2	0.1
A3F-13	5	0	1.8	0.2
A3F-14	6	0	118.3	6.9
A3F-15	7	0	73.9	7.6
A3F-16	8	0	6.0	1.4
A3F-17	8	0	2.4	0.1
A3F-18	9	0	91.9	4.4
A3F-19	10	0	79.7	7.6
A3F-20	11	0	0.0	0.0
A3F-21	21	0	0.0	0.0
Control	0	0	100.0	-

* Results are presented as the mean +/- SD relative values from three independent transfections after normalization to the non-mutated plasmid control.

Table 5.2.6 Analysis of G-to-A mutations induced by A3G in the eGFP gene from HIV[p8.9].

Clones	Total number of G-to-A mutations	Total number of termination codons	Relative Mean Fluorescence (FL1)*	Standard Deviation (SD)
A3G-1	0	0	101.0	8.1
A3G-2	2	0	95.8	4.6
A3G-3	3	0	101.2	8.8
A3G-4	3	0	96.8	2.4
A3G-5	3	0	118.4	7.6
A3G-6	4	0	1.9	0.7
A3G-7	4	0	0.0	0.0
A3G-8	4	0	76.5	8.1
A3G-9	4	0	69.2	3.7
A3G-10	5	0	50.2	3.0
A3G-11	5	0	10.9	3.3
A3G-12	6	0	2.2	0.9
A3G-13	7	0	51.0	7.9
A3G-14	8	0	0.0	0.0
A3G-15	9	0	0.0	0.0
A3G-16	11	0	12.5	2.5
A3G-17	13	0	1.0	0.2
A3G-18	18	0	0.0	0.0
A3G-19	20	0	0.0	0.0
A3G-20	23	0	0.0	0.0
A3G-21	26	0	0.0	0.0
Control	0	0	100.0	-

* Results are presented as the mean +/- SD relative values from three independent transfections after normalization to the non-mutated plasmid control.

Table 5.2.7 Analysis of G-to-A mutations induced by A3F in the Env-eGFP gene from M-MLV.

Clones	Total number of G-to-A mutations	Total number of termination codons	Relative mean Fluorescence (%)*	Standard Deviation (SD)
Env-A3F-1	0	0	94.9	8.3
Env-A3F-2	1	0	74.3	9.0
Env-A3F-3	1	0	91.0	3.8
Env-A3F-4	2	0	44.1	3.0
Env-A3F-5	3	0	0.0	0.0
Env-A3F-6	4	0	0.0	0.0
Env-A3F-7	4	0	66.2	9.3
Env-A3F-8	6	0	0.0	0.0
Env-A3F-9	6	0	57.6	5.2
Env-A3F-10	6	0	76.3	7.5
Env-A3F-11	6	0	33.7	3.7
Env-A3F-12	11	0	18.7	2.9
Env-A3F-13	15	0	0.0	0.0
Env-A3F-14	21	0	68.7	4.1
Env-A3F-15	33	0	0.0	0.0
Env-A3F-16	18	0	0.0	0.0
Env-A3F-17	28	1	0.0	0.0
Env-A3F-18	28	1	0.0	0.0
Env-A3F-19	29	1	0.0	0.0
Env-A3F-20	28	3	0.0	0.0
Env-A3F-21	43	0	0.0	0.0
Control	0	0	100	-

* Results are presented as the mean +/- SD relative values from three independent transfections after normalization to the non-mutated plasmid control.

Table 5.2.8 Analysis of G-to-A mutations induced by A3G in the Env-eGFP gene from M-MLV.

Clones	Total number of G-to-A mutations	Total number of termination codons	Relative mean Fluorescence (%)*	Standard Deviation (SD)
Env-A3G-1	0	0	95.2	8.5
Env-A3G-2	14	2	0.0	0.0
Env-A3G-3	15	3	0.0	0.0
Env-A3G-4	16	2	0.0	0.0
Env-A3G-5	21	6	0.0	0.0
Env-A3G-6	27	6	0.0	0.0
Env-A3G-7	39	7	0.0	0.0
Env-A3G-8	41	8	0.0	0.0
Env-A3G-9	42	9	0.0	0.0
Env-A3G-10	52	8	0.0	0.0
Env-A3G-11	43	6	0.0	0.0
Env-A3G-12	44	9	0.0	0.0
Env-A3G-13	47	8	0.0	0.0
Env-A3G-14	58	7	0.0	0.0
Env-A3G-15	52	7	0.0	0.0
Env-A3G-16	58	10	0.0	0.0
Env-A3G-17	64	9	0.0	0.0
Env-A3G-18	69	9	0.0	0.0
Env-A3G-19	78	11	0.0	0.0
Env-A3G-20	82	13	0.0	0.0
Env-A3G-21	109	12	0.0	0.0
Control	0	0	100.0	-

* Results are presented as the mean +/- SD relative values from three independent transfections after normalization to the non-mutated plasmid control.

Table 5.2.9 Non-synonymous substitutions generated by A3F in the eGFP gene from HIV[p8.9].

Clones	S3N	E18K	G36D	C49Y	M79I	D83N	E91K	A111T	E116K	G117N	E125K	E143K	E173K	D191N	D198N	S203N	S209N	E214K	R216H	M219I	E223K	E236K
A3F-1																						
A3F-2																						
A3F-3								X										X				
A3F-4																						
A3F-5																		X				
A3F-6											X											
A3F-7																						
A3F-8																						
A3F-9													X	X								
A3F-10				X	X	X																
A3F-11															X				X			
A3F-12							X		X							X		X				
A3F-13																		X				
A3F-14	X									X												X
A3F-15																	X	X				
A3F-16							X									X				X		
A3F-17		X																			X	
A3F-18							X															
A3F-19							X						X									
A3F-20							X					X									X	
A3F-21		X	X								X						X	X		X		

Amino acid substitutions in red represent inactivating mutations resulting in the absence of fluorescence.

Table 5.2.10 Non-synonymous substitutions generated by A3G in the eGFP gene from HIV[p8.9].

Clones	M1I	E6K	G11R/E/K	E18K	V23I	E35K	G68S	E91K	G92S	E96K	E116K	E125K	G139R	E143K	M154I	D181N	R216H	D217N	M219I	E223K	G229R/E/K	G233S	M234I	E236K
A3G-1																								
A3G-2													X								X			
A3G-3																					X			
A3G-4												X	X			X								
A3G-5			X										X								X			
A3G-6								X											X		X			
A3G-7									X				X							X	X			
A3G-8													X								X			
A3G-9											X		X											
A3G-10			X												X						X			
A3G-11																		X			X			
A3G-12													X						X		X			
A3G-13			X	X									X								X			
A3G-14																			X	X	X			
A3G-15						X							X	X				X			X			
A3G-16												X							X		X			
A3G-17													X		X				X		X			
A3G-18		X	X		X	X							X					X	X		X		X	
A3G-19	X										X	X	X							X	X		X	X
A3G-20	X		X				X				X		X		X		X	X	X		X			
A3G-21		X	X							X			X	X					X		X	X	X	

Amino acid substitutions in red represent inactivating mutations resulting in the absence of fluorescence.

Table 5.2.11 Non-synonymous substitutions generated by A3F in the eGFP region of the Env-eGFP gene from M-MLV.

Clones	E7K	C49Y	C71Y	M79I	G92D	D103N	E112K	E125K	D130N	E133K	E143K	E173K	R216H	E223K
Env-A3F-1														
Env-A3F-2														
Env-A3F-3														
Env-A3F-4														
Env-A3F-5										X				
Env-A3F-6										X				
Env-A3F-7														
Env-A3F-8								X						
Env-A3F-9														
Env-A3F-10						X								
Env-A3F-11							X							
Env-A3F-12				X										
Env-A3F-13	X				X									
Env-A3F-14														
Env-A3F-15		X							X					
Env-A3F-16											X			X
Env-A3F-17											X			X
Env-A3F-18						X			X		X		X	
Env-A3F-19											X			X
Env-A3F-20				X								X		
Env-A3F-21			X	X										

Amino acid substitutions in red represent inactivating mutations resulting in the absence of fluorescence.

Table 5.2.12 Non-synonymous substitutions generated by A3G in the eGFP region of the Env-eGFP gene from M-MLV.

Clones	G11R/E/K	E18K	D20N	W58 (STOP)	E91K	G117S	G128S	E133K	G135S	G138R	G139R	E142K	M154I	G190S	E214K	M219I	G229R/E/K	G233S	M234I
Env-A3G-1																			
Env-A3G-2																	X		
Env-A3G-3																			
Env-A3G-4																			
Env-A3G-5																			
Env-A3G-6																			
Env-A3G-7						X													
Env-A3G-8											X						X		
Env-A3G-9											X					X	X		
Env-A3G-10								X			X		X			X			X
Env-A3G-11	X																X		
Env-A3G-12			X								X					X			
Env-A3G-13					X									X	X				
Env-A3G-14											X								
Env-A3G-15	X											X				X	X		X
Env-A3G-16			X													X			
Env-A3G-17	X									X						X	X	X	
Env-A3G-18	X									X						X	X		X
Env-A3G-19							X	X	X		X		X			X			
Env-A3G-20			X	X				X			X		X		X	X	X		X
Env-A3G-21	X	X								X		X				X	X		X

Amino acid substitutions in red represent inactivating mutations resulting in the absence of fluorescence.

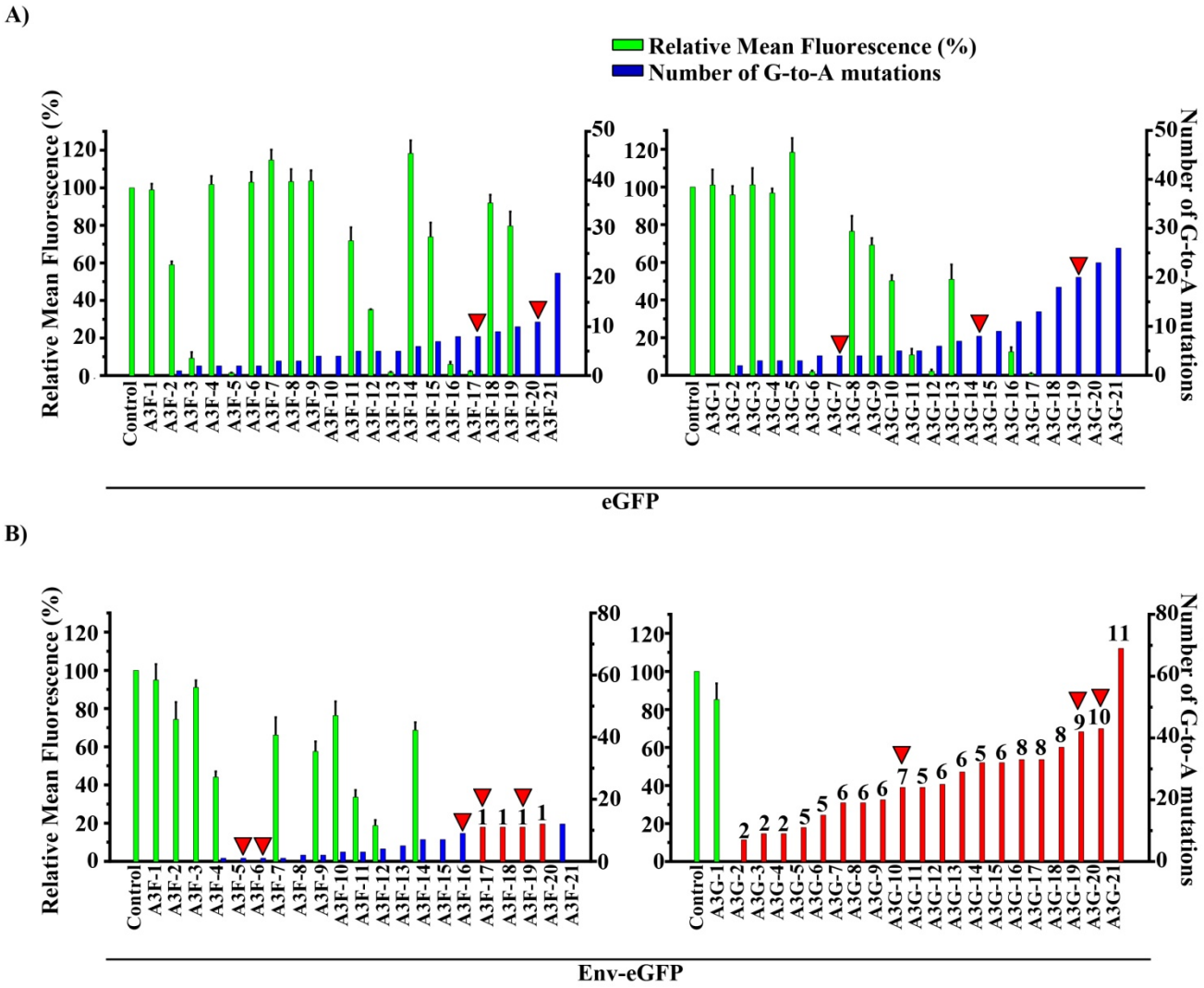

A)
Relative Mean Fluorescence (%)
Number of G-to-A mutations
Relative Mean Fluorescence (%)
Number of G-to-A mutations
Control
A3F-1
A3F-2
A3F-3
A3F-4
A3F-5
A3F-6
A3F-7
A3F-8
A3F-9
A3F-10
A3F-11
A3F-12
A3F-13
A3F-14
A3F-15
A3F-16
A3F-17
A3F-18
A3F-19
A3F-20
A3F-21
A3G-1
A3G-2
A3G-3
A3G-4
A3G-5
A3G-6
A3G-7
A3G-8
A3G-9
A3G-10
A3G-11
A3G-12
A3G-13
A3G-14
A3G-15
A3G-16
A3G-17
A3G-18
A3G-19
A3G-20
A3G-21
eGFP
B)
Relative Mean Fluorescence (%)
Number of G-to-A mutations
Env-eGFP

Figure 5.2.6 Mutations of 5'CC target motifs result in the incorporation of termination codons leading to gene inactivation whereas deamination of 5'TC motifs favours genetic diversification. (**A**) eGFP or (**B**) Env-eGFP reporter genes mutated by A3F or A3G were transfected in HEK 293T cells for measurement of eGFP fluorescence intensity by flow cytometry (green bars) and sent for Sanger sequencing for G-to-A mutations analysis (blue bars). Red bars represent the presence of at least one termination codon with the exact number indicated above the bar and inverted red triangles represent the presence of an inactivating mutation (E133K or E223K). Values for fluorescence represent the mean +/- SD from three independent transfections and were normalized to a non-mutated plasmid control. The total number of mutations for Env-eGFP was compiled including only mutations in the N-terminal Env and eGFP regions of the gene.

absence of stop codon in most (55%) cases. Overall, the results presented here suggest that the DNA sequence specificity for deamination by the A3 proteins is responsible for the generation of inactivating mutations when tryptophans are present leading to gene inactivation.

5.3 Discussion

In this chapter, we demonstrated that the deamination specificity of A3 proteins dictates the potency of retroviral gene inactivation achieved through thegenerationof termination codons or high levels of mutations. This study emerged from the surprising finding that a double-mutant of A3G, A3G[W94A/W127A], completely lacking RNA-binding capacity, had an altered target motif for deamination going from 5'CC to a 5'TC motif (**Figure 5.2.1A and Table 5.2.1**). Previous studies aiming at identifying the sequence determinants responsible for the preference for deamination by the APOBECs have mapped this property to residues in loop-7 of the CTD of A3G (206, 209-213, 309). We still do not know how RNA-binding through the NTD of the protein contributes to the selection of substrate specificity. It is possible that the two tryptophan residues identified in this study are required for proper folding of the protein allowing substrate engagement by the CTD. Nevertheless, this represents the first cell-based demonstration of the involvement of residues in the NTD of A3G for substrate specificity.

Our comparison of proviral clones mutated by A3G or A3F demonstrated that mutations induced in a 5'CC context led to significantly more inactivating mutations than 5'TC alterations induced by A3F which were only observed in 20% of the clones analyzed (**Figure 5.2.6**). However, the conditions used in the present study led to high amounts of A3 proteins being expressed resulting in several molecules being incorporated into each viral particle. The end result was a high mutation load in the viral genomes analyzed and therefore, gene inactivation

was caused either by termination codons or hypermutations. Nevertheless, it is worth mentioning that at physiological levels, lower amount of the A3 proteins are expressed and packaged in the newly synthesized viruses. We thus expect to see lower mutation rates induced by the APOBECs and in those conditions, retroviral gene inactivation would occur only if a stop codon was generated. As a consequence, A3G is expected to be the most efficient member of the family *in vivo* since it is the most likely to introduce inactivating mutations because of its unique preferred target motif for deamination (245, 246). This is further emphasized by a previous report showing almost no restriction by A3F at physiological levels (310). Mutations at 5'TC dinucleotides caused by the other members of the A3 family are most likely to be beneficial rather than detrimental to the virus (243, 246, 304).

Gene inactivation resulting from the generation of termination codons was less prominent in the case of the eGFP gene for both A3G and A3F since there is only a single 5'CC target site for deamination that can result in a stop codon (**Figure 5.2.6A**). That could account for the lack of restriction observed in the presence of mutants of A3G in the case of HIV[p8.9] (**Figure 5.2.3B**). This virus expresses eGFP as a single protein therefore presenting a single opportunity to generate a stop codon resulting in gene inactivation. Since the mutation levels induced by mutants of A3G are reduced compared to the wild type protein, the lack of termination codons could significantly impair the restriction potential of the mutants (**Table 3.2.1**). Whereas with M-MLV, since the eGFP gene is expressed as a fusion protein with the envelope, there are many more target sites for deamination that can potentially result in the incorporation of termination codons and therefore, the same mutants have been shown to restrict retroviral infection by 40-55% (**Figure 3.2.3F**). It is thus extremely important to consider the number of potential target

motifs that could yield stop codons present in a gene when using it as a reporter system since that might significantly influence the results.

This study also demonstrated that mutations induced by A3G are usually lethal for the virus either by causing gene-inactivating mutations or by extensive levels of hypermutations (278). However, the situation is different in the case of A3F and potentially other A3 proteins because they preferentially mutate 5'TC DNA substrates. As shown by the current results, there appears to be a threshold around 10 in the amount of mutations being deleterious as mutation intensities below this threshold do not appear to affect the functionality of the gene used in the current study (**Figure 5.2.6**). Sub-lethal mutations (<10), among which many are non-synonymous can potentially contribute to genetic diversification and help the progeny viruses to increase their fitness.

In general, the mutation load introduced by A3F in the Env-eGFP gene was significantly lower than that induced by A3G (**Table 5.2.1**). By comparing the availability of each enzyme's target site, we concluded that the lack of hypermutation in the N-terminal portion of the gene in the case of A3F could be explained by the fact that there were 50% fewer 5'TC motifs than 5'CC motifs (**Figure 5.2.5B**). However, in the eGFP and Env portion downstream of eGFP, both target sites are equally distributed and therefore, the absence of deamination motif could not explain the reduced hypermuation frequency. Rather, the lower mutation rate induced by A3F could be attributed to its different scanning mechanism as this enzyme exclusively jumps on the DNA substrate (304). This would explain why we observed gaps with almost no mutation intercalated within highly hypermutated regions (**Figure 5.2.5C**). The high proportion of mutations introduced at the end of the C-terminus of Env-eGFP however, is more reflective of a sliding mechanism. Yet, we believe that this is rather the result of the overwhelming amount of target

sites present in that region (70 sites within amino acids 700 to 900). Interestingly, some target motifs were consistently missed by both A3G and A3F along the Env-eGFP. Although there does not seem to be a consensus in the nucleotide sequence surrounding the non-targeted residues, the lack of deamination could depend on the secondary structure of the DNA as it is the case for the local dinucleotide preference by A3G (**Figure 5.2.5C and D**) (311). While some DNA structures may favour deamination, others could prevent the catalytic activity of the A3 proteins.

Finally, we were able to identify critical residues in the eGFP gene that confer fluorescence. Two in particular, E133K and E223K, resulted in complete loss of fluorescence (**Figure 5.2.6**). Glutamic acid to lysine (E-to-K) is a non-synonymous mutation from an acidic negatively charged to a basic positively charged amino acid. Moreover, glutamic acids are critical amino acids that are often part of the active or binding sites of proteins and this type of substitution has been proven deleterious for the envelope gene of M-MLV (312). We do not know however whether this is an epistatic mutation and whether this substitution alone is sufficient to observe the deleterious phenotype. In fact, the damaging E462K change identified in M-MLV's envelope has a more pronounced effect when accompanied by additional mutations (312). Further studies are also required to determine the effect of E133K on eGFP's function. Potential features of the protein that would strongly affect its function include folding, dimerization and wavelengths of excitation and emission (313). The other critical residue targeted by both A3G and A3F in the eGFP gene of HIV[p8.9], E223K, has been implicated in the shifted excitation spectra existing between GFP and eGFP proteins via the latter's interaction with the chromophore and its further ionization (314). The fact that this residue is now positively charged could well generate an as of yet uncharacterized shift in the excitation wavelength of

eGFP that is outside the range of wavelengths currently used in our assay or it may simply inactivate the fluorescence of the gene.

In summary, we provided here a detailed *in vitro* analysis of the spectrum of the genetic damages produced by two A3 proteins. Our study suggests that A3G-induced G-to-A mutations are extremely efficient at inactivating retroviral genes through the generation of stop codons,regardless of the mutation intensity, when tryptophan codons are present in a gene. However, mutations induced in a 5'TC context by the other A3 proteins are less apt at introducing stop codons, and gene inactivation by the latter is dependent on the total mutation burden. Finally, when A3 proteins cause intense mutagenesis, this increases the probability of targeting critical residues involved in gene function, whereas sub-lethal levels of mutations induced in a 5'TC context are potentially less harmful and even beneficial to the virus through genetic diversification.

Chapter 6 Development of a New Technology for the Screening and Profiling of Mutations Induced by the APOBEC3 Proteins in Retroviral Genomes

Preface: The results presented in this chapter have been published in:

Kasandra Bélanger, Mathieu Savoie, Halil Aydin, Tyler Renner, Zahra Montazeri and Marc-André Langlois. **(2013)** Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis. ***Virology***, 448, 168-175.

*See Appendix III for the published article.

6.1 Introduction

Cytidine deamination is a process by which an amino group is hydrolyzed from one carboxyl of the deoxycytidine base to yield a deoxyuridine which can then pair with an adenine during DNA synthesis of the opposite strand. Such DNA modifications result in C-to-T mutations on the strand being deaminated and G-to-A mutations on the complementary strand. The enzymes responsible for this process are called cytidine deaminases and play a fundamental role in both innate and adaptive immunity. One of them, AID is involved in the generation of high affinity antibodies with diversified functions whereas A3 proteins possess a potent antiretroviral activity (315). In addition to their protective role against retroviruses, it has recently been established that mutations by A3 proteins can also lead to HIV-1 drug resistance and the development of several types of cancers in humans (244, 249, 250, 261, 262, 264, 303, 316, 317). Efficient, quick and low-cost detection technologies are therefore needed to study the genetic effects of C-to-T/ G-to-A substitutions caused by A3 proteins..

To this day, the most commonly used technique to investigate the presence of G-to-A hypermutations induced by A3 proteins involves the amplification of the mutated gene of interest

by PCR, cloning it into a vector, and then sequencing it to assess the mutation burden. However, this method is highly time consuming, labour intensive, and only allows for the analysis of a selected number of clones. The recent development of next generation sequencing significantly increased the sequencing throughput and depth of sequences analyzed. Nevertheless, the complex sample preparation, the requirement of a specialized technological platform and software as well as the high costs associated with the technique make its routinely use unaffordable and unsustainable. Moreover, next generation sequencing analyses are usually limited to short fragments. 3D-PCR is another commonly used technique in the APOBEC field. It is a simple method in which a first round PCR is performed on proviral DNA to amplify the gene of interest followed by a fully nested gradient PCR to amplify a smaller fragment (275). This technique relies on the basis that GC-rich sequences melt at a higher temperature than AT-rich sequences therefore permitting the selective amplification of rare G-to-A hypermutated sequences in a population of non-mutated sequences. The technique is highly sensitive on fragments up to 300 bp, above which its efficiency decreases significantly (275, 318, 319). Moreover, this method is preferentially biased towards highly mutated sequences, and therefore does not allow for the detection of sequences with few deamination events. In addition, the gradient PCR involved in the technique does not allow for high throughput analyses.

In contrast, high resolution melting (HRM) allows for the analysis of a large number of samples in a short time frame. This method involves an initial qPCR step in which the gene of interest is amplified followed by the HRM stage in which the amplicon is gradually heated up. At some point during this process, the melting temperature of the amplicon is reached and the two strands of DNA start to separate or "melt" apart. The technique relies on a fluorescent dye that binds specifically to dsDNA. Therefore, as the temperature increases, the DNA dissociates

to become single-stranded and fluorescence gradually decreases. Similarly to 3D-PCR, variations in the GC content of the DNA sequence analyzed influences its melting temperature and the G-to-A mutations generated by the A3 proteins can therefore easily be detected (320). Until now, HRM has mostly been used for the detection of single nucleotide polymorphisms (SNPs) and analysis of DNA methylation and has been limited to fragments of less than 300 bp (321-323).

In this chapter, we optimized this technique to be able to use it as a screening tool for G-to-A mutations induced by the A3 proteins and to be able to accurately quantify the number of substitutions in fragments up to 900 bp. The complete process was named high throughput hypermutation analysis by HRM (HyperHRM). We then applied our technology to identify mutations induced by all seven A3 proteins in the entire genome of two retroviruses, Vif-deficient HIV-1 (HIVΔVif) and M-MLV, and also in peripheral blood monocytic cells (PBMCs) isolated from HIV-1 infected patients.

6.2 Results

6.2.1 Comparison of the accuracy and sensitivity of HyperHRM and 3D-PCR.

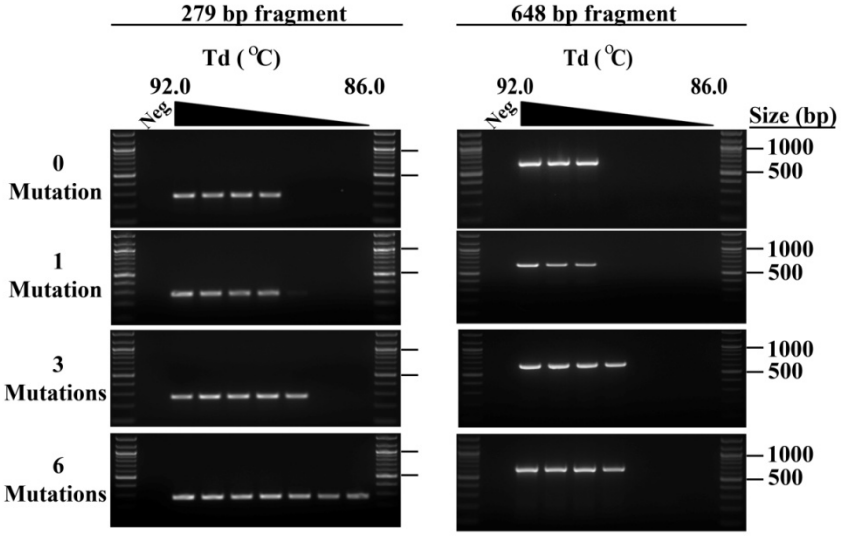
Here, we wanted to compare the detection limit of 3D-PCR and HyperHRM analysis. Plasmid DNA controls containing 0 to 9 G-to-A mutations in the eGFP reporter gene of M-MLV were used. For 3D-PCR, a first round PCR was performed to amplify the full-length eGFP (717 bp) which was then used for a gradient PCR to generate amplicons of increasing size up to 648 bp (**Figure 6.2.1A**). This method successfully allowed the detection of a single mutation in the smaller DNA fragment (279 bp) whereas only plasmids containing 3 mutations or more gave a positive signal in the 648 bp amplicon. In the case of HyperHRM, a single PCR was performed on plasmids containing different numbers of mutations to amplify fragments of increasing lengths in eGFP which were then used for melt curve analysis. To further determine

the accuracy of the technique, the temperature at which the difference between each plasmid control containing a fixed number of mutation was identified using the formula presented in Chapter 2 of this thesis (section 2.18). A linear correlation was then established between the fluorescence at the selected temperature and the number of mutation in each control for every amplicon size and the R-squared (R^2) values were calculated (**Figure 6.2.1B and Table 6.2.1**). The linear regression was used to estimate the number of G-to-A substitutions in each plasmid control based on their fluorescence at the established temperature (**Table 6.2.1**). Using this mathematical model, we found that HyperHRM could detect a single mutation in fragments as large as 648 bp and a minimum of three mutations within larger (800-900 bp) amplicons, with R^2 values above 0.94 in all cases. Finally, we established that HyperHRM's upper limit is 1000 bp, as the technique could not accurately determine the number of mutations within amplicons this size.

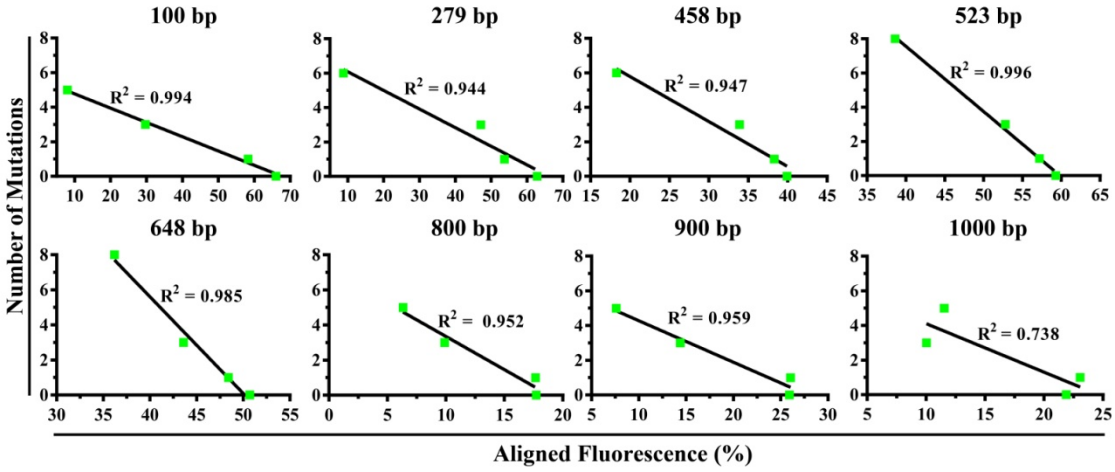
Table 6.2.1 Accuracy of HyperHRM in increasing fragments length.

Fragment length used for HyperHRM analysis	Number of mutations in control plasmids determined by HyperHRM analysis						Optimal temperature determined by HyperHRM analysis
	0	1	3	5	6	8	
100	0.0	0.8	3.6	5.7	-	-	85.2
279	0.3	1.3	2.1	-	6.2	-	87.8
428	0.3	1.0	2.2	-	6.2	-	87.9
523	0.2	1.0	2.7	-	-	8.1	86.6
648	0.0	1.0	3.6	-	-	7.7	87.9
800	0.4	0.4	3.4	4.7	-	-	89.6
900	0.5	0.4	3.2	4.9	-	-	89.7
1000	1.0	0.4	6.6	5.9	-	-	89.4

A)


279 bp fragment
648 bp fragment
Td (°C)
92.0
86.0
Neg
Size (bp)
0 Mutation
1 Mutation
3 Mutations
6 Mutations
1000
500

B)


100 bp
279 bp
458 bp
523 bp
648 bp
800 bp
900 bp
1000 bp
R² = 0.994
R² = 0.944
R² = 0.947
R² = 0.996
R² = 0.985
R² = 0.952
R² = 0.959
R² = 0.738
Number of Mutations
Aligned Fluorescence (%)

C)

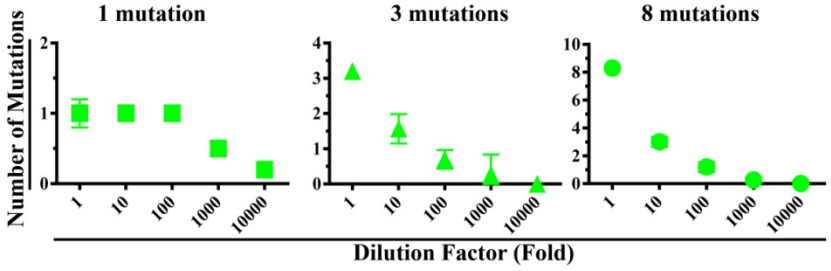

1 mutation
3 mutations
8 mutations
Number of Mutations
Dilution Factor (Fold)

Figure 6.2.1 Comparison of the accuracy and sensitivity of HyperHRM and 3D-PCR. (**A**) Plasmid DNA controls containing 0, 1, 3 or 6 G-to-A mutations were used for a first round PCR to amplify a 717 bp region of the eGFP gene followed by a thermal gradient PCR performed on a nested eGFP fragment of 279 or 648 bp. (**B**) Controls with known numbers of mutations were analyzed by HyperHRM in eGFP fragments of increasing length. A linear regression between fluorescence and number of mutations at the temperature corresponding to the maximum difference in fluorescence between the controls was established for each amplicon and the R^2 values were determined. The analysis was repeated four times and the data represent mean values. (**C**) Plasmids containing 1, 3 or 8 mutations in a 648 bp region of eGFP were diluted at the indicated ratio in a non-mutated plasmid. Dilutions were then used for HyperHRM analysis and the number of mutations in each samples was determined based on their fluorescence. Values represent the mean +/- SD of three independent experiments.

We next asked whether HyperHRM could be used for the detection of rare mutation events within a majority of non-mutated sequences. To tackle this question, we performed serial dilutions of plasmid DNA containing 1, 3 or 8 mutations in the eGFP gene and mixed these with a non-mutated plasmid. We then performed a first round PCR to amplify a fragment of 648 bp followed by HyperHRM analysis. We observed fluorescence above background for all mutations with dilutions up to 1/1000 suggesting that even a singly mutated sequence would successfully be detected in a thousand non-mutated sequences with our technique (**Figure 6.2.1C and Table 6.2.2**).

Table 6.2.2 Sensitivity of HyperHRM.

Dilution of control DNA plasmids in non-mutated plasmid	Number of mutations determined by HyperHRM analysis		
	1	3	8
No dilution	1.1	4.9	11.5
1/10	1.1	2.1	4.1
1/100	1.1	1.1	1.7
1/1000	0.7	0.5	0.5
1/10 000	0.1	0.0	0.0

6.2.2 Mutations induced by the seven A3 proteins along the full genome of M-MLV.

Having established the accuracy and the sensitivity of HyperHRM to detect G-to-A mutations, we then asked whether it could be used for high throughput analysis of mutations induced by all seven A3 proteins on the full-length genome of retroviruses. We started by investigating M-MLV by producing the virus in the presence of each member of the A3 family in 293T cells for 48 h. Two days following transfection, viruses were used to infect 3T3 target cells and 24 h after infection, gDNA was extracted from infected cells. PCR was first performed on this gDNA to amplify five large 2 kb fragments (A to E) spanning the complete viral genome. These PCR products were then purified and used for HyperHRM analysis achieved on smaller

regions of approximately 640 bp (regions 1 to 15) to cover the near full-length genome (**Figure 6.2.2A**). The 5'LTR (region 1) was excluded from the analysis because of poor PCR amplification. We found that all A3 proteins were capable of generating mutations in several regions of the M-MLV genome with A3A having the weakest deaminase activity and A3G displaying the strongest (**Figure 6.2.2B**). Different mutation profiles were observed with the different deaminases suggesting that they might act together to exert a compound effect in a natural setting (**Figure 6.2.2C**). Since HRM readouts are affected by base substitutions that alter the GC content of the DNA sequence being analyzed, we were curious to see whether the mutations observed were in fact G-to-A mutations. We therefore cloned a region showing high levels of mutations in the presence of A3G, region 12, as well as regions that appeared susceptible to deamination by the low mutator APOBECs, A3A and A3C (regions 4 and 6) in a cloning vector and analyzed them using Sanger sequencing. The results demonstrated exclusively the presence of G-to-A mutations in a 5'CC DNA context in the presence of A3G whereas several non-A3 specific substitutions were observed when viruses were produced in the presence of A3A or A3C (**Tables 6.2.4 to 6.2.6**). Using plasmids with known numbers of mutations in region 12 of M-MLV genome, a linear regression was established between the fluorescence at the temperature at which the difference between the controls was the highest and the number of mutation in each plasmid control. The regression was then used to estimate the number of mutation in 44 clones of the same region mutated by A3G. Following HyperHRM, the mutation rate obtained by combining the results for the 44 clones analyzed was approximately 11 mutations per kb which was similar to that seen with Sanger sequencing, confirming the accuracy of the technique (**Figure 6.2.2D and E and Table 6.2.3**).

Table 6.2.3 Analysis of mutations induced by A3G in region 12 of the M-MLV genome.

From / To	A	C	T	G
A	-	-	-	-
C	-	-	-	-
T	-	-	-	-
G	307	-	-	-
Number of clones analyzed	44			
Number of base pairs analyzed	28 160			
Total number of mutations	307			
Mutation rate (mutation/kb)	10.9			

Table 6.2.4 Analysis of the preferred DNA substrate for deamination by A3G in region 12 of the M-MLV genome*.

A3G	-2	-1
A	14	1
C	48	80
T	30	18
G	8	1

*Mutations were tabulated on the minus-strand retroviral DNA. Data represents the percentage of each base at position -2 and -1 relative to the deaminated cytidine at position 0.

Table 6.2.5 Mutations induced in two susceptible regions of M-MLV by A3A.

	R4				R6			
From / To	**A**	**C**	**T**	**G**	**A**	**C**	**T**	**G**
A	-	66	33	174	-	3	-	13
C	33	-	142	-	-	-	2	-
T	11	99	-	11	-	12	-	-
G	**143**	22	11	-	15	-	-	-
Number of bp sequenced	7128				6480			
Total	745				45			

Table 6.2.6 Mutations induced in two susceptible regions of M-MLV by A3C.

	R4				R6			
From / To	**A**	**C**	**T**	**G**	**A**	**C**	**T**	**G**
A	-	72	36	192	-	-	-	11
C	36	-	156	-	-	-	10	-
T	12	108	-	12	-	15	-	4
G	**156**	24	12	-	-	-	-	-
Number of bp sequenced	7776				7776			
Total	816				40			

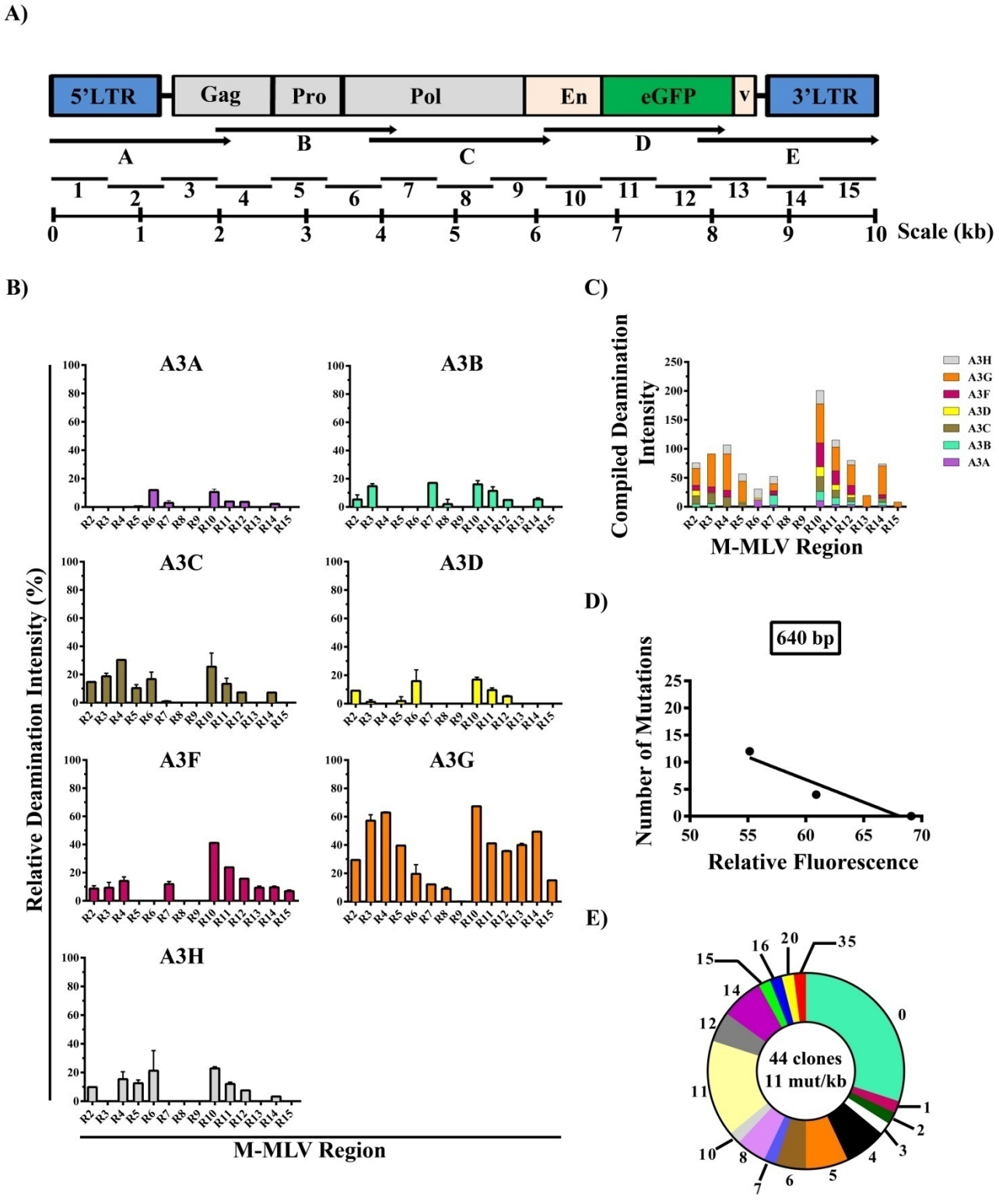

A)
5'LTR
Gag
Pro
Pol
En
eGFP
v
3'LTR
A
B
C
D
E
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
0
1
2
3
4
5
6
7
8
9
10
Scale (kb)
B)
A3A
A3B
A3C
A3D
A3F
A3G
A3H
Relative Deamination Intensity (%)
M-MLV Region
C)
Compiled Deamination Intensity
A3H
A3G
A3F
A3D
A3C
A3B
A3A
M-MLV Region
D)
640 bp
Number of Mutations
Relative Fluorescence
E)
44 clones
11 mut/kb

Figure 6.2.2 Mutations induced by the seven A3 proteins along the full genome of M-MLV. (**A**) Schematic representation of the M-MLV genome with the first round PCR products (A to E) and the 15 regions analyzed by HyperHRM. (**B**) M-MLV viruses were produced in 293T cells in the presence of the A3 proteins and used to infect 3T3 target cells. gDNA extracted from infected cells was then used for first round PCRs to amplify large regions of the genome (A to E) and PCR products were purified and used for HyperHRM analysis on smaller fragments (2-15) to cover the near full-length viral genome. Results represent the difference in fluorescence relative to the A2 negative control and are the mean +/- SD of three independent experiments performed in triplicates. (**C**) Mutation profile exerted by all seven A3s on the near full-length genome of M-MLV obtained by combining the results obtained in (**B**). (**D**) Plasmid DNA controls with fixed numbers of mutations in region 12 of the M-MLV genome were used for HyperHRM analysis and linear regression between the fluorescence and the number of mutations was established. (**E**) Region 12 produced in the presence of A3G was amplified by PCR and cloned in a sequencing vector. A total of 44 clones were analyzed by HyperHRM and the linear regression established in (**D**) was used to quantify the mutations in each clone and calculate the mutation rate.

6.2.3 Mutations induced by the seven A3 proteins within the full genome of HIVΔVif.

Similarly to M-MLV, we designed our assay by dividing the genome of HIVΔVif into five large fragments of approximately 2 kb each (A to E) for first round PCR amplification and purification followed by HRM analysis on 16 smaller regions (1 to 16) to cover the entire retroviral genome (**Figure 6.2.3A**). Mutation intensity induced by the seven A3 proteins in each region was determined relative to the A2 non-mutated control (**Figure 6.2.3B**). As previously described, A3D, A3F, A3G and A3H were found to be the most potent members of the family against this virus (181). A3B was also found to generate a significant amount of mutations in several regions of the viral genome, however, this protein is not naturally expressed in HIV target cells and therefore is not expected to play a significant role in restriction against the virus *in vivo* (183). As was observed with M-MLV, A3A showed the weakest activity against HIVΔVif. The combined deamination profile of all seven APOBECs emphasized the important and collective role of A3D, A3F, A3G and A3H in HIVΔVif restriction (**Figure 6.2.3C**). Signal above background could be detected in some regions in the presence of weakly active APOBECs, such as A3A and A3C. These regions should be analyzed by sequencing in order to determine whether these mutations are G-to-A substitutions.

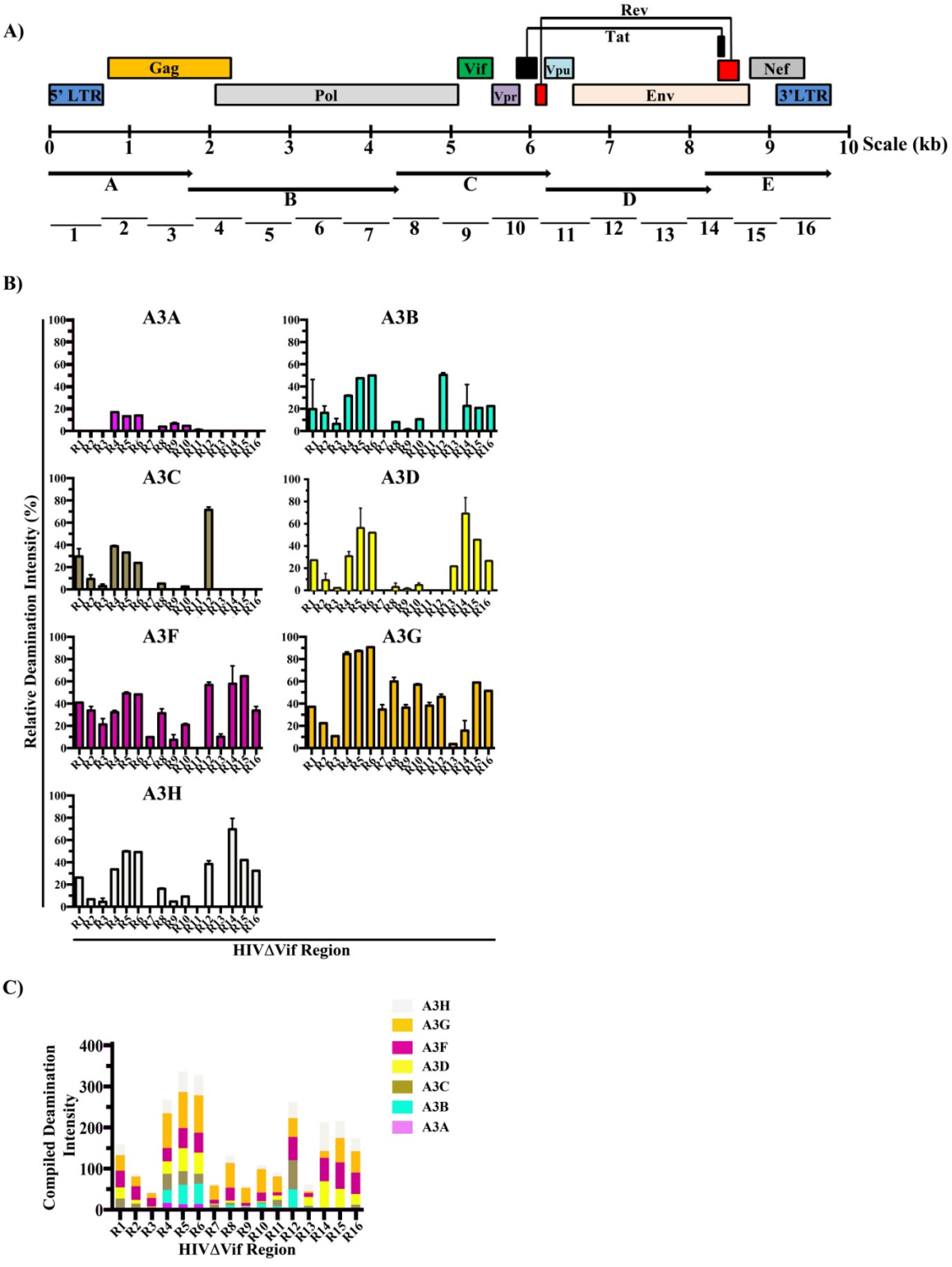

A)
Rev
Tat
Gag
Vif
Vpu
Nef
5' LTR
Pol
Vpr
Env
3'LTR
0 1 2 3 4 5 6 7 8 9 10 Scale (kb)
A
B
C
D
E
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
B)
A3A
A3B
A3C
A3D
A3F
A3G
A3H
Relative Deamination Intensity (%)
HIVΔVif Region
C)
A3H
A3G
A3F
A3D
A3C
A3B
A3A
Compiled Deamination Intensity
HIVΔVif Region

Figure 6.2.3 Mutations induced by the seven A3 proteins along the full genome of HIVΔVif. (**A**) Schematic representation of the HIVΔVif genome with the first round PCR products (A to E) and the 16 regions analyzed by HyperHRM. (**B**) HIVΔVif viruses were produced in 293T cells in the presence of the A3 proteins and used to infect 293T target cells. gDNA extracted from infected cells was then used for first round PCRs to amplify large regions of the genome (A to E) and PCR products were purified and used for HyperHRM analysis on smaller fragments (1-16) to cover the full-length viral genome. Results represent the difference in fluorescence relative to the A2 negative control and are the mean +/- SD of three independent experiments performed in triplicates. (**C**) Mutation profile exerted by all seven A3s on the full-length genome of HIVΔVif obtained by combining the results obtained in (**B**).

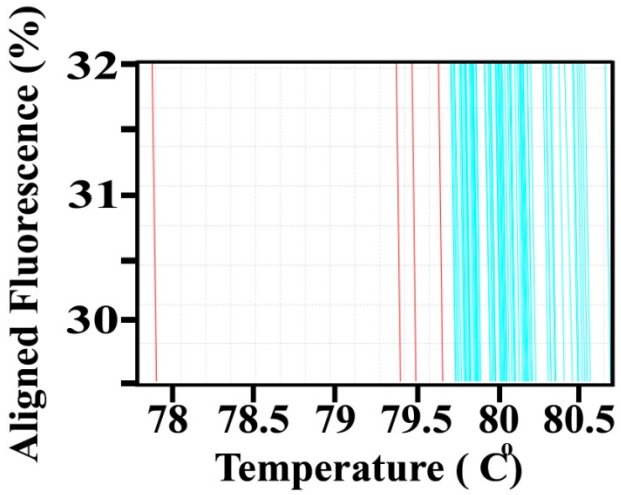
6.2.4 Clinical application of HyperHRM.

To determine whether HRM could be used to screen clinical samples, gDNA was extracted from PBMCs of an HIV-1-infected patient. We then performed PCR amplification of a 467 bp conserved region of the pol gene and subcloned the amplicon into a sequencing vector. Colonies obtained after culture at 37ºC overnight were then mixed with water and used for HRM performed on the same fragment of the pol gene. Amongst all 96 colonies analyzed, 4 showed a shift in their melting temperature towards a lower temperature and were sent for Sanger sequencing analysis of the G-to-A mutation profiles (**Figure 6.2.4A**). Clones 2, 3 and 4, which demonstrated the smaller difference in their melting temperature compared to the rest of the colonies, had less than 4 G-to-A mutations and these were not characteristic of an A3 footprint (**Figure 6.2.4B**). Nevertheless, clone 1 was found to have 25 A3-specific G-to-A mutations (**Figure 6.2.4B and Table 6.2.7**). Overall, these results suggest that HyperHRM can detect hypermutated sequences in HIV-1 clinical samples.

Table 6.2.7 Mutation analysis in an HIV-1 clinical sample using HyperHRM.

From / To	A	C	T	G
A	-	-	-	1
C	-	-	-	-
T	-	1	-	-
G	25	-	-	-
Total number of mutations		27		

A)


Aligned Fluorescence (%)
32
31
30
78
78.5
79
79.5
80
80.5
Temperature (C°)

B)

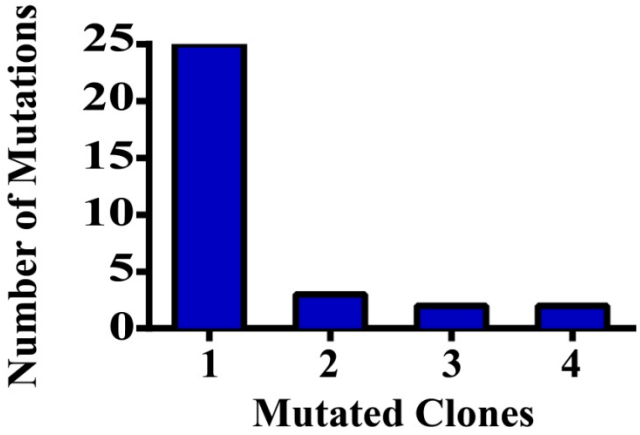

Number of Mutations
25
20
15
10
5
0
1
2
3
4
Mutated Clones

Figure 6.2.4 Clinical application of HyperHRM. (**A**) gDNA was isolated from PBMCs of an HIV-1-infected patient and used for PCR amplification of a highly conserved section of the HIV pol gene. The purified PCR product was then subcloned in a sequencing vector and colonies were used for HRM analysis using a 96-well plate. A zoomed-in section of the melting curves obtained by HRM for each clone is represented with the clones showing a reduced melting temperature indicated in red. (**B**) The four clones with a lower melting temperature (in red in (**A**)) were analyzed by sequencing. The number of G-to-A mutations in each clone is represented.

6.3 Discussion

In this study, we analyzed both the accuracy and the sensitivity of HyperHRM and extended the use of the technique for high-throughput mutation screening of full-length viral genomes. This is of significant importance since many drug resistance mutations in HIV-infected patients have been linked to A3-mediated G-to-A substitutions (244, 249, 250, 303). Moreover and of great interest, there is now robust evidence for a link between deamination of host genomic DNA by the A3 proteins and the development of many type of cancers in humans (261, 262, 264, 316, 317). In cancer as in HIV, only a small percentage of cells actually contain sequences with the mutations of interest. The need of a powerful and sensitive technique to be able to screen large DNA regions and detect rare mutation events within a majority of non-mutated sequences is therefore required. As an example, fewer than 10% of HIV-1 proviruses are hypermutated in infected individuals (324). This implies that a useful method to screen for hypermutated sequences must have the sensitivity to detect one hypermutated sequence for every ten non-hypermutated sequences. HyperHRM accurately detected the presence of hypermutated sequences even when outnumbered 1000-fold by non-hypermutated sequences (**Figure 6.2.1C**).

Analysis of the deamination profile induced by the A3 proteins along the retroviral genome of M-MLV and HIVΔVif led to the observation of a typical twin gradient as it was previously described for A3G acting on HIV-1 DNA (**Figures 6.2.2B and 6.2.3B**) (325). In that previous report, two regions in the retroviral genome were shown to be significantly less targeted for deamination by the protein. These were identified as being part of the central polypurine tract (cPPT) and the LTR proximal polypurine tract (3'PPT) which correspond to regions 9 and 13 in **Figure 6.2.3A**. In the case of M-MLV, the infrequently deaminated region 15 matches the location of the 3'PPT located just upstream of the 3' LTR, however, a cPPT has never been

documented for this retrovirus. By carefully analyzing the sequence of regions 8 and 9 we found marked similarities with the cPPT of other retroviruses and region 9 of this virus suggesting the potential existence of an as yet un-characterized cPPT in M-MLV. Further studies are required to confirm the existence of a cPPT.

Although A3A and A3C were found to have a weaker activity against both retroviruses (**Figure 6.2.2B and 6.2.3B**), surprisingly high numbers of mutations were found to be induced by these proteins in retroviral genomes. DNA sequencing analyses of the nature of the substitutions generated by these proteins in two independent regions of the viral genome of M-MLV revealed the presence of non A3-mediated mutations (**Tables 6.2.5 and 6.2.6**). It is possible that some A3 proteins may exhibit deamination-independent restriction mechanisms that affect retroviral RT fidelity. In addition, highly active A3 proteins, such as A3G and A3F, could potentially induce polymerase error in regions where their respective preferred target motifs for deamination are limited in number. This remains however to be proven. Albeit, mouse APOBEC 3 (mA3) uses a similar defence mechanism against MLVs and incorporation of this protein into virions has been shown to reduce both the activity and the fidelity of the MLV RT (326).

Finally, screening of the gDNA from HIV-1-infected PBMCs of patients demonstrated the ability of our HyperHRM technique to detect hypermutated sequences in a large population of non-mutated sequences (**Figure 6.2.4**). HyperHRM could be used directly to quantify the exact number of mutations in a given patient-derived sequence because of numerous naturally occurring base substitutions present that were not caused by A3 proteins. Nevertheless, we believe that HyperHRM is an effective tool to perform an initial selection to detect sequences that are likely to be hypermutated. These sequences can then be further analyzed by DNA sequencing to confirm the presence of A3-induced G-to-A hypermutations. HyperHRM can

therefore strategically reduce the amount of sequencing required and as a result, the overall cost of a research project.

In summary, HyperHRM is an accurate, high throughput method that allows the detection of extremely rare A3-induced mutational events in a heterogeneous sequence population. It can also accurately quantify the number of A3-induced G-to-A mutations in a homogenous PCR amplicon. HyperHRM is a method that has the potential to make significant contributions to better understanding the impact of the mutator activities of A3 proteins both in laboratory and clinical settings.

Chapter 7 Conclusion

Although the mortality rate associated with HIV-1 has significantly decreased over the last 30 years, millions of people worldwide are suffering the side effects of life-long antiretroviral treatments. Intrinsic restriction factors such as A3G represent appealing targets for the development of novel anti-HIV drugs. However, a better understanding of how this protein restricts retroviral infection, how it is regulated and how it selects its DNA substrate is necessary if it is to be exploited as a potential anti-HIV therapeutic.

The goal of my doctoral research was to characterize the mechanisms underlying the regulation of A3G. More specifically, over the past few years, we have been interested in dissecting the details of how the protein interacts with RNA. Although the deaminase activity of A3G is limited to ssDNA substrates, non-selective association of the protein with cytoplasmic RNAs leads to its accumulation into large RNP complexes in which its catalytic activity and its ability to get incorporated into viral particles are inhibited (**Figure 3.2.1**)(214, 215, 265, 266). A few years ago, we suggested that RNA-binding mutants of A3G could be used to prevent the formation of HMM complexes and increase the potency of the protein in dividing cells. However, we soon realized that the antiretroviral potency of such mutants was highly compromised.

First, in chapter 3 of this thesis, we found that the RNA-binding mutants of A3G could not restrict infection by HIV[p8.9] and had a significantly reduced antiretroviral activity against M-MLV (**Figure 3.2.3**). Our first assumption was that the mutants had lost their deaminase activity. However to our big surprise, high levels of G-to-A hypermutations were detected in the genome of both viruses in the presence of the RNA-binding mutants (**Tables 3.2.1 and 3.2.2**). We demonstrated for the first time that deamination can occur in the absence of viral restriction.

The existence of mutants with such a remarkable phenotype allowed us to assess the role played by deamination in the overall antiretroviral activity of A3G. What we found was quite intriguing. The RNA-binding mutants of A3G failed to inhibit LRTs accumulation and proviral integration (**Figure 3.2.4**). This implies that deamination, which was once suggested as the only means of restriction by A3G has no effect on the early steps of retroviral infection. This implies that a retrovirus could successfully replicate and integrate into the host genome in the presence of A3G proteins that are deamination-active and RNA-binding deficient. Nonetheless, integrated deaminated viral genomes are likely to be inactivated by mutations. In fact, in a recent report it was shown that most proviruses in PBMCs are lethally mutated and only a small fraction are released from infected cells (327). In chapter 5 of this thesis, we also demonstrated that A3G is highly effective at inactivating gene function through the incorporation of termination codons (**Figure 5.2.6**). However, as discussed in chapter 3, the possibility that retroviral genomes inactivated by hypermutations could be complemented by unmutated genomes following co-infection of the same cell isn't to be neglected. The work presented in this thesis has important implications in our understanding of how the deamination-independent and dependent modes of restriction by A3G act together to achieve optimal antiretroviral activity. It also represents the first evidence to date demonstrating a strong correlation between RNA-binding and the deamination-independent restriction by the protein.

While RNA plays an important role in the antiretroviral activity and regulation of A3G, the same does not necessarily apply to other members of the A3 family. In chapter 4, we provided a detailed analysis of the cellular organization for each of the human A3 proteins. In comparison to A3G, we showed that RNP complexes formed by A3B, A3D and A3F are much more resistant to treatment with RNase A (**Figure 4.2.2**). This suggests that RNA might not

regulate the existing forms of these proteins in the cell. Similarly, none of the APOBECs were found to interact with L1 mRNA although several proteins restricted the retroelement suggesting that L1 mRNA-binding is not involved in the restricion of retrotransposition (**Figure 4.2.4**). Nevertheless, a role for RNA in this process cannot be completely excluded since the restriction of L1 retrotransposition by A3C has been shown to require RNA-dependent binding of the protein to ORF1 (235). Therefore, while RNA may be marginal in the assembly of different members of the A3 family into HMM complexes, interaction with specific RNA species may have a role in their activity against retroviruses and retroelements. In agreement with this statement, the incorporation of A3F inside viral particles requires its interaction with 7SL RNA (194). Binding to 7SL is also necessary for the packaging and antiretroviral activity of A3H (195). Overall, these findings together with novel observations presented in chapter 4 of this thesis imply that the function occupied by RNA in the biology of each A3 protein might differ.

In chapter 5, in view of better understanding the role of RNA in the catalytic activity of A3G, we analyzed proviral DNA obtained from cells infected with viruses produced in the presence of a RNA-binding deficient A3G mutant (**Figure 5.2.1**). Our results revealed that the mutation rate induced by the double mutant of A3G was significantly reduced (**Figure 5.2.4**). Intriguingly, we also observed that the preferred DNA substrate for deamination by the mutant was altered from 5'CC to 5'TC (**Table 5.2.1**). The residues involved in ssDNA binding and substrate specificity by A3G have been mapped to its CTD (176, 206, 209-213, 309, 328, 329). Depending on the model, a group of arginines R213, R215, R313, R320, R374 and R376 as well as an asparagine N244 have been shown to be required for the binding of ssDNA by the protein (176, 309, 328, 329). Surprisingly, mutation of one or more of these residues does not affect interaction with ssDNA unless accompanied with mutations in the NTD at positions W94 and

W127 (176). In addition, a significant contribution to ssDNA binding is presumed to be through the NTD of A3G based on apparent equilibrium dissociation constant (Kd) values obtained for the full-length protein and its CTD only (174, 205, 229, 328, 330). The processivity of deamination by the enzyme is also significantly affected by mutations in its NTD implying a role for this domain in stabilizing the interaction with the ssDNA substrate (331). However, a direct involvement for residues of the NTD of A3G in ssDNA binding demonstrated in a cell-based assay had never been described until now through my research. Similarly, a single amino acid, D317 located in loop 7 of the CTD of A3G has been demonstrated to be responsible for the selection of DNA dinucleotide target motifs by the protein (213).This is in agreement with a number of previous reports also involving loop 7 in the CTD of several members of the A3 family as well as AID in their substrate specificity for deamination (206, 209-212, 309). The results presented in chapter 5 of this thesis represent the first real demonstration of the requirement for residues in the NTD of A3G in the choice of DNA dinucleotide targets. This implies that these residues from the NTD of the protein are indeed involved in its interaction with ssDNA. In agreement with our findings, Polevoda and colleagues recently mapped the region on A3G involved in its binding to ssDNA to a large surface area covering both the N- and C-terminal moieties of the protein (332). This explains why both domains individually are not sufficient for high-affinity binding to ssDNA and efficient catalytic activity (309, 328, 333, 334). Through our analysis of the RNA-binding deficient double mutant of A3G, we also demonstrated that these same two residues are involved in the interaction of the protein with both RNA and ssDNA (**Figure 5.2.1 and Table 5.2.1**). These findings led us to hypothesize that RNA-binding by A3G may act as a regulatory mechanism to prevent interaction of the protein with ssDNA until an adequate substrate is found. This raises the question as to how A3G releases RNA to

engage a ssDNA substrate. In his recent publication, Polevoda and his group confirmed our hypothesis by demonstrating that residues 181-194 from the NTD and residues 314-320 and 345-374 from the CTD of A3G were involved in the binding of the protein to both RNA and ssDNA therefore supporting the existence of a competition model for nucleic acid binding by A3G (332). The authors also extended our understanding of the regulation of ssDNA interaction by A3G by revealing the identity of RNA-binding only residues, including W94 (332). Other residues included amino acids 15-29, 41-52 and 83-99 from the N-terminus of A3G. This suggests that in addition to the competitive inhibition existing between RNA and ssDNA at residues interacting with both, binding to RNA by RNA-binding only residues represent another form of allosteric inhibition for ssDNA binding. Moreover, Polevoda and colleagues showed that although A3G interacts with RNA and ssDNA with similar affinity, ssDNA is a much better competitor. This means that A3G bound to RNA will release its substrate in the presence of ssDNA (332). Although tryptophans 94 and 127 were not identified by the authors as ssDNA and RNA-binding residues, the existence of additionnal residues cannot be excluded. In his report, Polevoda makes several references to our work since we were the first ones to predict the competitive inhibitory role played by RNA in the catalytic activity of A3G. We were pleased to see our results confirmed and extended by this group.

Also in chapter 5 of this thesis, our work with the double mutant of A3G next led us to investigate the potency for gene inactivation by A3 proteins based on their preferred DNA dinucleotide substrate. We provided a clear demonstration of the efficiency of A3G at inactivating gene function through the incorporation of stop codons even at low mutation burden (**Figure 5.2.6**). On the contrary, A3F and potentially all A3 proteins mutating in a 5'TC context require higher mutation intensities to inactivate gene function (**Figure 5.2.6**). Incorporation of

termination codons by these proteins rarely occurs (**Figure 5.2.6**). These results suggest that although mutations induced by A3G are likely to inactivate progeny viruses, mutagenesis by other members of the family could favour genetic diversification and increase viral fitness. This notion of duality of A3 deamination had been proposed by several groups in the field, however we were the first ones to directly relate it to gene function in a controlled functional assay (241-246, 249, 250, 257, 258, 303, 304, 331). An important question that emerged from these findings is whether efforts should be dedicated into the generation of A3 inhibitors rather than Vif antagonists. A recent study performed on samples from HIV-1-infected individuals revealed an extremely high mutation rate ($(4.1 \pm 1.7) \times 10^{-3}$ per base per cell) in viral genomes from PBMCs (327). A3 proteins contributed to 98% of these mutations, with A3G being responsible for the majority (88%) of all substitutions. Most viruses were inactivated by the mutations as shown by the observation that the mutation rate in plasma viruses was 44 times lower than in PBMCs (327). These results corroborate the high potency for gene inactivation of A3G depicted in our assays. Further increasing the efficiency of the protein by preventing the action of Vif is likely to be deleterious for the viral population. Since other members of the A3 family are less potent and more likely to benefit the virus, perhaps specific antagonists that target the interaction of Vif and A3G only should be the focus of our labours.

Finally, in the last chapter of this thesis (chapter 6), we developed a powerful technique that we called HyperHRM to detect and quantify single nucleotide mutations in proviral DNA. Through the use of HyperHRM, we provided the first A3-mediated mutation analysis of the full-length M-MLV retroviral genome (**Figure 6.2.2**). Intriguingly, we observed the presence of numerous substitutions different than G-to-A alterations in the presence of A3 proteins in the retroviral genome (**Tables 6.2.5 and 6.2.6**). This increase in the mutation rate by the retroviral

RT might represent a mode of restriction used by the A3 proteins independently of their deaminase activity. Indeed, some mutagenic nucleoside analogs that are known to increase the mutation rate upon retroviral replication by the RT have been shown to result in the production of non-infectious progeny viruses (335-339). These findings suggest that mutations induced by the RT are likely to be harmful for the virus above a certain threshold. In that regard, mutants of HIV-1with amino acid substitutions in the RT that decrease fidelity were also found to have a reduced fitness (340). While we previously established that RNA-binding is dispensable for the interaction of A3G with HIV-1 RT, the requirement for RNA in the reduction of the enzyme's fidelity cannot be excluded (**Figure 5.2.3**). It is possible that the A3 proteins also interact directly with the viral genomic RNA template to generate a road block and impede the activity of the RT (229, 231).

As a follow up to this thesis, it would be highly interesting to investigate interactions between members of the A3 family and RNA that could potentially lead to deleterious consequences. The recent discovery of the involvement of some A3 proteins in several types of cancer suggest that we are unaware of all the roles carried out by these proteins at a cellular level. Intriguingly, several micro-RNAs have been implicated in the development of many kinds of cancers as well as other diseases including autoimmune and neurodegenerative disorders (Reviewed in (341)). Seeing as A3 proteins have the ability to travel from one cell to another as part of extracellular vesicles such as exosomes, they might have a role in transporting nucleic acids including micro-RNAs with highly diversified functions (342-344). This could have major repercussions in recipient cells. In fact, it has been documented that there are fluctuations in the expression of certain genes following exosomal delivery of micro-RNAs (345-348). In addition, these small RNA species can serve as immune activators following engagement of Toll-like

receptors (TLRs) (349). To determine whether A3 proteins represent the driving force leading to the incorporation of micro-RNAs inside exosomes, we could take advantage of our A3-deficient mice. The exosomal content of their serum could be analyzed and compared to wild type mice in order to determine whether certain types of micro-RNAs still accumulate inside exsosomes in the absence of A3 proteins.

Since their discovery in 2002, significant advances have been made in our general understanding of the functioning of A3 proteins. However, the expression, compartmentalization and activity of the proteins greatly vary depending on the cell type and morphological state in which they are residing. In this thesis, we presented novel and unexpected roles for RNA in the regulation and modulation of the antiretroviral activity of A3 proteins. We also provided structural insights on how both domains of the full-length A3G protein interact. In addition, our work raised important issues such as the multifaceted outcome of A3-mediated retroviral mutagenesis, thereby emphasizing the need to carefully assess the relevance of either increasing or inhibiting enzymatic activities of A3 proteins in order to better fight HIV infection and spread.

References

1. **Ivanosky D.** 1892. Über die Mosaikkrankheit der Tabakspflanze. Bulletin Scientifique publié par l'Académie Impériale des Sciences de Saint-Pétersbourg / Nouvelle Serie III (in German and Russian) (St. Petersburg) **35:**67-70.
2. **Beijerinck M.** 1898. Über ein Contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter. Verhandelingen der Koninklyke akademie van Wettenschappen te Amsterdam **65:**1-22.
3. **Ellermann V. BO.** 1908. Experimentelle Leukämie bei HÜhnern. Zentralbl. Bakteriol. Parasitenkd. Infectionskr. Hyg. Abt. Orig. **46:**595-609.
4. **Rous P.** 1911. A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. The Journal of experimental medicine **13:**397-411.
5. **Shope RE, Hurst EW.** 1933. Infectious Papillomatosis of Rabbits : With a Note on the Histopathology. The Journal of experimental medicine **58:**607-624.
6. **Bittner JJ.** 1936. Some Possible Effects of Nursing on the Mammary Gland Tumor Incidence in Mice. Science **84:**162.
7. **Gross L.** 1957. Development and serial cellfree passage of a highly potent strain of mouse leukemia virus. Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine **94:**767-771.
8. **Moloney JB.** 1960. Biological studies on a lymphoid-leukemia virus extracted from sarcoma 37. I. Origin and introductory investigations. Journal of the National Cancer Institute **24:**933-951.
9. **Jarrett WF, Crawford EM, Martin WB, Davie F.** 1964. A Virus-Like Particle Associated with Leukemia (Lymphosarcoma). Nature **202:**567-569.
10. **Chopra HC, Mason MM.** 1970. A new virus in a spontaneous mammary tumor of a rhesus monkey. Cancer research **30:**2081-2086.
11. **Dulbecco R.** 1952. Production of Plaques in Monolayer Tissue Cultures by Single Particles of an Animal Virus. Proceedings of the National Academy of Sciences of the United States of America **38:**747-752.
12. **Temin HM, Rubin H.** 1958. Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture. Virology **6:**669-688.
13. **Vogt PK.** 1997. Historical Introduction to the General Properties of Retroviruses. *In* Coffin JM, Hughes SH, Varmus HE (ed.), Retroviruses, Cold Spring Harbor (NY).
14. **H.M. T.** 1964. Nature of the provirus of Rous Sarcoma. Natl. Cancer Inst. Monogr. **17:**557-570.
15. **Baltimore D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature **226:**1209-1211.
16. **Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC.** 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America **77:**7415-7419.
17. **Uchiyama T, Yodoi J, Sagawa K, Takatsuki K, Uchino H.** 1977. Adult T-cell leukemia: clinical and hematologic features of 16 cases. Blood **50:**481-492.
18. **Morgan DA, Ruscetti FW, Gallo R.** 1976. Selective in vitro growth of T lymphocytes from normal human bone marrows. Science **193:**1007-1008.

19. **Ruscetti FW, Morgan DA, Gallo RC.** 1977. Functional and morphologic characterization of human T cells continuously grown in vitro. Journal of immunology **119:**131-138.
20. **Seiki M, Hattori S, Hirayama Y, Yoshida M.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proceedings of the National Academy of Sciences of the United States of America **80:**3618-3622.
21. **Rosenblatt JD, Cann AJ, Slamon DJ, Smalberg IS, Shah NP, Fujii J, Wachsman W, Chen IS.** 1988. HTLV-II transactivation is regulated by the overlapping tax/rex nonstructural genes. Science **240:**916-919.
22. **Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220:**868-871.
23. **Levy JA, Shimabukuro J.** 1985. Recovery of AIDS-associated retroviruses from patients with AIDS or AIDS-related conditions and from clinically healthy individuals. The Journal of infectious diseases **152:**734-738.
24. **Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B, et al.** 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science **224:**500-503.
25. **Vogt VM.** 1997. Retroviral Virions and Genomes. *In* Coffin JM, Hughes SH, Varmus HE (ed.), Retroviruses, Cold Spring Harbor (NY).
26. **White TA, Bartesaghi A, Borgnia MJ, Meyerson JR, de la Cruz MJ, Bess JW, Nandwani R, Hoxie JA, Lifson JD, Milne JL, Subramaniam S.** 2010. Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure. PLoS pathogens **6:**e1001249.
27. **Turner BG, Summers MF.** 1999. Structural biology of HIV. Journal of molecular biology **285:**1-32.
28. **Freed EO, Martin MA.** 1995. The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. The Journal of biological chemistry **270:**23883-23886.
29. **Berger EA, Murphy PM, Farber JM.** 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annual review of immunology **17:**657-700.
30. **Doms RW, Peiper SC.** 1997. Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. Virology **235:**179-190.
31. **Wilen CB, Tilton JC, Doms RW.** 2012. HIV: cell binding and entry. Cold Spring Harbor perspectives in medicine **2**.
32. **Arhel N.** 2010. Revisiting HIV-1 uncoating. Retrovirology **7:**96.
33. **Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, El Marjou A, Lacabaratz C, Lelievre JD, Manel N.** 2013. The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. Immunity **39:**1132-1142.

34. **Telesnitsky A, Goff SP.** 1997. Reverse Transcriptase and the Generation of Retroviral DNA. *In* Coffin JM, Hughes SH, Varmus HE (ed.), Retroviruses, Cold Spring Harbor (NY).
35. **Charneau P, Alizon M, Clavel F.** 1992. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. Journal of virology **66:**2814-2820.
36. **Hungnes O, Tjotta E, Grinde B.** 1992. Mutations in the central polypurine tract of HIV-1 result in delayed replication. Virology **190:**440-442.
37. **Wurtzer S, Goubard A, Mammano F, Saragosti S, Lecossier D, Hance AJ, Clavel F.** 2006. Functional central polypurine tract provides downstream protection of the human immunodeficiency virus type 1 genome from editing by APOBEC3G and APOBEC3B. Journal of virology **80:**3679-3683.
38. **Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M, Emerman M.** 1994. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. Proceedings of the National Academy of Sciences of the United States of America **91:**7311-7315.
39. **Di Marzio P, Choe S, Ebright M, Knoblauch R, Landau NR.** 1995. Mutational analysis of cell cycle arrest, nuclear localization and virion packaging of human immunodeficiency virus type 1 Vpr. Journal of virology **69:**7909-7916.
40. **Jowett JB, Planelles V, Poon B, Shah NP, Chen ML, Chen IS.** 1995. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. Journal of virology **69:**6304-6313.
41. **Brown PO.** 1997. Integration. *In* Coffin JM, Hughes SH, Varmus HE (ed.), Retroviruses, Cold Spring Harbor (NY).
42. **Fisher AG, Feinberg MB, Josephs SF, Harper ME, Marselle LM, Reyes G, Gonda MA, Aldovini A, Debouk C, Gallo RC, et al.** 1986. The trans-activator gene of HTLV-III is essential for virus replication. Nature **320:**367-371.
43. **Dayton AI, Sodroski JG, Rosen CA, Goh WC, Haseltine WA.** 1986. The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. Cell **44:**941-947.
44. **Daly TJ, Cook KS, Gray GS, Maione TE, Rusche JR.** 1989. Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature **342:**816-819.
45. **Freed EO.** 1998. HIV-1 gag proteins: diverse functions in the virus life cycle. Virology **251:**1-15.
46. **Berkowitz R, Fisher J, Goff SP.** 1996. RNA packaging. Current topics in microbiology and immunology **214:**177-218.
47. **Little SJ, McLean AR, Spina CA, Richman DD, Havlir DV.** 1999. Viral dynamics of acute HIV-1 infection. The Journal of experimental medicine **190:**841-850.
48. **Gurunathan S, Habib RE, Baglyos L, Meric C, Plotkin S, Dodet B, Corey L, Tartaglia J.** 2009. Use of predictive markers of HIV disease progression in vaccine trials. Vaccine **27:**1997-2015.
49. **Mattapallil JJ, Douek DC, Hill B, Nishimura Y, Martin M, Roederer M.** 2005. Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. Nature **434:**1093-1097.

50. **Guadalupe M, Reay E, Sankaran S, Prindiville T, Flamm J, McNeil A, Dandekar S.** 2003. Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. Journal of virology **77:**11708-11717.
51. **Brenchley JM, Schacker TW, Ruff LE, Price DA, Taylor JH, Beilman GJ, Nguyen PL, Khoruts A, Larson M, Haase AT, Douek DC.** 2004. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. The Journal of experimental medicine **200:**749-759.
52. **Mehandru S, Poles MA, Tenner-Racz K, Horowitz A, Hurley A, Hogan C, Boden D, Racz P, Markowitz M.** 2004. Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. The Journal of experimental medicine **200:**761-770.
53. **Chun TW, Nickle DC, Justement JS, Meyers JH, Roby G, Hallahan CW, Kottilil S, Moir S, Mican JM, Mullins JI, Ward DJ, Kovacs JA, Mannon PJ, Fauci AS.** 2008. Persistence of HIV in gut-associated lymphoid tissue despite long-term antiretroviral therapy. The Journal of infectious diseases **197:**714-720.
54. **Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, Hermankova M, Chadwick K, Margolick J, Quinn TC, Kuo YH, Brookmeyer R, Zeiger MA, Barditch-Crovo P, Siliciano RF.** 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature **387:**183-188.
55. **Chun TW, Finzi D, Margolick J, Chadwick K, Schwartz D, Siliciano RF.** 1995. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nature medicine **1:**1284-1290.
56. **Lewin-Smith MR, Klassen MK, Frankel SS, Nelson AM.** 1998. Pathology of human immunodeficiency virus infection: infectious conditions. Annals of diagnostic pathology **2:**181-194.
57. **Collier AC, Coombs RW, Schoenfeld DA, Bassett RL, Timpone J, Baruch A, Jones M, Facey K, Whitacre C, McAuliffe VJ, Friedman HM, Merigan TC, Reichman RC, Hooper C, Corey L.** 1996. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. The New England journal of medicine **334:**1011-1017.
58. **D'Aquila RT, Hughes MD, Johnson VA, Fischl MA, Sommadossi JP, Liou SH, Timpone J, Myers M, Basgoz N, Niu M, Hirsch MS.** 1996. Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in patients with HIV-1 infection. A randomized, double-blind, placebo-controlled trial. National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. Annals of internal medicine **124:**1019-1030.
59. **Staszewski S, Miller V, Rehmet S, Stark T, De Cree J, De Brabander M, Peeters M, Andries K, Moeremans M, De Raeymaeker M, Pearce G, Van den Broeck R, Verbiest W, Stoffels P.** 1996. Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients. Aids **10:**F1-7.
60. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. Journal of virology **69:**5087-5094.

61. **Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH.** 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. Journal of virology **84:**9864-9878.
62. **O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD.** 2002. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. The Journal of biological chemistry **277:**38053-38061.
63. **Coffin JM.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science **267:**483-489.
64. **Frost SD, McLean AR.** 1994. Quasispecies dynamics and the emergence of drug resistance during zidovudine therapy of HIV infection. Aids **8:**323-332.
65. **Nowak MA, Bonhoeffer S, Shaw GM, May RM.** 1997. Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations. Journal of theoretical biology **184:**203-217.
66. **Stengel RF.** 2008. Mutation and control of the human immunodeficiency virus. Mathematical biosciences **213:**93-102.
67. **Autran B, Carcelain G, Li TS, Blanc C, Mathez D, Tubiana R, Katlama C, Debre P, Leibowitch J.** 1997. Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. Science **277:**112-116.
68. **Komanduri KV, Viswanathan MN, Wieder ED, Schmidt DK, Bredt BM, Jacobson MA, McCune JM.** 1998. Restoration of cytomegalovirus-specific CD4+ T-lymphocyte responses after ganciclovir and highly active antiretroviral therapy in individuals infected with HIV-1. Nature medicine **4:**953-956.
69. **Lederman MM, Connick E, Landay A, Kuritzkes DR, Spritzler J, St Clair M, Kotzin BL, Fox L, Chiozzi MH, Leonard JM, Rousseau F, Wade M, Roe JD, Martinez A, Kessler H.** 1998. Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir: results of AIDS Clinical Trials Group Protocol 315. The Journal of infectious diseases **178:**70-79.
70. **Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, Smith K, Lisziewicz J, Lori F, Flexner C, Quinn TC, Chaisson RE, Rosenberg E, Walker B, Gange S, Gallant J, Siliciano RF.** 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nature medicine **5:**512-517.
71. **Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF.** 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science **278:**1295-1300.
72. **Gallo RC.** 1988. HIV--the cause of AIDS: an overview on its biology, mechanisms of disease induction, and our attempts to control it. Journal of acquired immune deficiency syndromes **1:**521-535.
73. **Yoshida M, Miyoshi I, Hinuma Y.** 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proceedings of the National Academy of Sciences of the United States of America **79:**2031-2035.

74. **Nisole S, Stoye JP, Saib A.** 2005. TRIM family proteins: retroviral restriction and antiviral defence. Nature reviews. Microbiology **3:**799-808.
75. **Towers GJ.** 2007. The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology **4:**40.
76. **Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J.** 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature **427:**848-853.
77. **Kirmaier A, Wu F, Newman RM, Hall LR, Morgan JS, O'Connor S, Marx PA, Meythaler M, Goldstein S, Buckler-White A, Kaur A, Hirsch VM, Johnson WE.** 2010. TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. PLoS biology **8**.
78. **Perron MJ, Stremlau M, Song B, Ulm W, Mulligan RC, Sodroski J.** 2004. TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. Proceedings of the National Academy of Sciences of the United States of America **101:**11827-11832.
79. **Hatziioannou T, Perez-Caballero D, Yang A, Cowan S, Bieniasz PD.** 2004. Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. Proceedings of the National Academy of Sciences of the United States of America **101:**10774-10779.
80. **Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, Diaz-Griffero F, Anderson DJ, Sundquist WI, Sodroski J.** 2006. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proceedings of the National Academy of Sciences of the United States of America **103:**5514-5519.
81. **Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M.** 2011. Hexagonal assembly of a restricting TRIM5alpha protein. Proceedings of the National Academy of Sciences of the United States of America **108:**534-539.
82. **Lukic Z, Hausmann S, Sebastian S, Rucci J, Sastri J, Robia SL, Luban J, Campbell EM.** 2011. TRIM5alpha associates with proteasomal subunits in cells while in complex with HIV-1 virions. Retrovirology **8:**93.
83. **Harris RS, Dudley JP.** 2015. APOBECs and virus restriction. Virology **479-480C:**131-145.
84. **Li N, Zhang W, Cao X.** 2000. Identification of human homologue of mouse IFN-gamma induced protein from human dendritic cells. Immunology letters **74:**221-224.
85. **Skasko M, Weiss KK, Reynolds HM, Jamburuthugoda V, Lee K, Kim B.** 2005. Mechanistic differences in RNA-dependent DNA polymerization and fidelity between murine leukemia virus and HIV-1 reverse transcriptases. The Journal of biological chemistry **280:**12190-12200.
86. **Diamond TL, Roshal M, Jamburuthugoda VK, Reynolds HM, Merriam AR, Lee KY, Balakrishnan M, Bambara RA, Planelles V, Dewhurst S, Kim B.** 2004. Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. The Journal of biological chemistry **279:**51545-51553.
87. **Baldauf HM, Pan X, Erikson E, Schmidt S, Daddacha W, Burggraf M, Schenkova K, Ambiel I, Wabnitz G, Gramberg T, Panitz S, Flory E, Landau NR, Sertel S, Rutsch F, Lasitschka F, Kim B, Konig R, Fackler OT, Keppler OT.** 2012. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. Nature medicine **18:**1682-1687.

88. **Descours B, Cribier A, Chable-Bessia C, Ayinde D, Rice G, Crow Y, Yatim A, Schwartz O, Laguette N, Benkirane M.** 2012. SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4(+) T-cells. Retrovirology **9:**87.
89. **Lahouassa H, Daddacha W, Hofmann H, Ayinde D, Logue EC, Dragin L, Bloch N, Maudet C, Bertrand M, Gramberg T, Pancino G, Priet S, Canard B, Laguette N, Benkirane M, Transy C, Landau NR, Kim B, Margottin-Goguet F.** 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. Nature immunology **13:**223-228.
90. **Ryoo J, Choi J, Oh C, Kim S, Seo M, Kim SY, Seo D, Kim J, White TE, Brandariz-Nunez A, Diaz-Griffero F, Yun CH, Hollenbaugh JA, Kim B, Baek D, Ahn K.** 2014. The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. Nature medicine **20:**936-941.
91. **Choi J, Ryoo J, Oh C, Hwang S, Ahn K.** 2015. SAMHD1 specifically restricts retroviruses through its RNase activity. Retrovirology **12:**46.
92. **Kupzig S, Korolchuk V, Rollason R, Sugden A, Wilde A, Banting G.** 2003. Bst-2/HM1.24 is a raft-associated apical membrane protein with an unusual topology. Traffic **4:**694-709.
93. **Neil SJ, Sandrin V, Sundquist WI, Bieniasz PD.** 2007. An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. Cell host & microbe **2:**193-203.
94. **Neil SJ, Zang T, Bieniasz PD.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451:**425-430.
95. **Venkatesh S, Bieniasz PD.** 2013. Mechanism of HIV-1 virion entrapment by tetherin. PLoS pathogens **9:**e1003483.
96. **Sheehy AM, Gaddis NC, Choi JD, Malim MH.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418:**646-650.
97. **Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, Yatim A, Emiliani S, Schwartz O, Benkirane M.** 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature **474:**654-657.
98. **Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J.** 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature **474:**658-661.
99. **Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS.** 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Molecular biology and evolution **22:**367-377.
100. **Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N.** 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics **79:**285-296.
101. **Desimmie BA, Delviks-Frankenberrry KA, Burdick RC, Qi D, Izumi T, Pathak VK.** 2014. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. Journal of molecular biology **426:**1220-1245.
102. **Mercenne G, Bernacchi S, Richer D, Bec G, Henriet S, Paillart JC, Marquet R.** 2010. HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. Nucleic acids research **38:**633-646.

103. **Dube M, Paquay C, Roy BB, Bego MG, Mercier J, Cohen EA.** 2011. HIV-1 Vpu antagonizes BST-2 by interfering mainly with the trafficking of newly synthesized BST-2 to the cell surface. Traffic **12:**1714-1729.
104. **Dube M, Roy BB, Guiot-Guillain P, Binette J, Mercier J, Chiasson A, Cohen EA.** 2010. Antagonism of tetherin restriction of HIV-1 release by Vpu involves binding and sequestration of the restriction factor in a perinuclear compartment. PLoS pathogens **6:**e1000856.
105. **Mitchell RS, Katsura C, Skasko MA, Fitzpatrick K, Lau D, Ruiz A, Stephens EB, Margottin-Goguet F, Benarous R, Guatelli JC.** 2009. Vpu antagonizes BST-2-mediated restriction of HIV-1 release via beta-TrCP and endo-lysosomal trafficking. PLoS pathogens **5:**e1000450.
106. **Lau D, Kwan W, Guatelli J.** 2011. Role of the endocytic pathway in the counteraction of BST-2 by human lentiviral pathogens. Journal of virology **85:**9834-9846.
107. **Iwabu Y, Fujita H, Kinomoto M, Kaneko K, Ishizaka Y, Tanaka Y, Sata T, Tokunaga K.** 2009. HIV-1 accessory protein Vpu internalizes cell-surface BST-2/tetherin through transmembrane interactions leading to lysosomes. The Journal of biological chemistry **284:**35060-35072.
108. **Iwabu Y, Fujita H, Tanaka Y, Sata T, Tokunaga K.** 2010. Direct internalization of cell-surface BST-2/tetherin by the HIV-1 accessory protein Vpu. Communicative & integrative biology **3:**366-369.
109. **Mangeat B, Gers-Huber G, Lehmann M, Zufferey M, Luban J, Piguet V.** 2009. HIV-1 Vpu neutralizes the antiviral factor Tetherin/BST-2 by binding it and directing its beta-TrCP2-dependent degradation. PLoS pathogens **5:**e1000574.
110. **Goffinet C, Allespach I, Homann S, Tervo HM, Habermann A, Rupp D, Oberbremer L, Kern C, Tibroni N, Welsch S, Krijnse-Locker J, Banting G, Krausslich HG, Fackler OT, Keppler OT.** 2009. HIV-1 antagonism of CD317 is species specific and involves Vpu-mediated proteasomal degradation of the restriction factor. Cell host & microbe **5:**285-297.
111. **Ayinde D, Maudet C, Transy C, Margottin-Goguet F.** 2010. Limelight on two HIV/SIV accessory proteins in macrophage infection: is Vpx overshadowing Vpr? Retrovirology **7:**35.
112. **Schmidt S, Schenkova K, Adam T, Erikson E, Lehmann-Koch J, Sertel S, Verhasselt B, Fackler OT, Lasitschka F, Keppler OT.** 2015. SAMHD1's protein expression profile in humans. Journal of leukocyte biology **98:**5-14.
113. **Kyei GB, Cheng X, Ramani R, Ratner L.** 2015. Cyclin L2 is a critical HIV dependency factor in macrophages that controls SAMHD1 abundance. Cell host & microbe **17:**98-106.
114. **Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M,**

Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, International Human Genome Sequencing C. 2001. Initial sequencing and analysis of the human genome. Nature **409:**860-921.

115. **Cordaux R, Batzer MA.** 2009. The impact of retrotransposons on human genome evolution. Nature reviews. Genetics **10:**691-703.
116. **Bratthauer GL, Cardiff RD, Fanning TG.** 1994. Expression of LINE-1 retrotransposons in human breast cancer. Cancer **73:**2333-2336.
117. **Bratthauer GL, Fanning TG.** 1992. Active LINE-1 retrotransposons in human testicular cancer. Oncogene **7:**507-510.
118. **Bratthauer GL, Fanning TG.** 1993. LINE-1 retrotransposon expression in pediatric germ cell tumors. Cancer **71:**2383-2386.
119. **Belancio VP, Deininger PL, Roy-Engel AM.** 2009. LINE dancing in the human genome: transposable elements and disease. Genome medicine **1:**97.
120. **Munoz-Lopez M, Garcia-Perez JL.** 2010. DNA transposons: nature and applications in genomics. Current genomics **11:**115-128.
121. **Griffiths DJ.** 2001. Endogenous retroviruses in the human genome sequence. Genome biology **2:**REVIEWS1017.
122. **Batzer MA, Deininger PL.** 2002. Alu repeats and human genomic diversity. Nature reviews. Genetics **3:**370-379.

123. **Ostertag EM, Goodier JL, Zhang Y, Kazazian HH, Jr.** 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. American journal of human genetics **73:**1444-1451.
124. **Ostertag EM, Kazazian HH, Jr.** 2001. Biology of mammalian L1 retrotransposons. Annual review of genetics **35:**501-538.
125. **Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV.** 2001. Human L1 retrotransposition: cis preference versus trans complementation. Molecular and cellular biology **21:**1429-1439.
126. **Luan DD, Korman MH, Jakubczak JL, Eickbush TH.** 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72:**595-605.
127. **Liu G, Program NCS, Zhao S, Bailey JA, Sahinalp SC, Alkan C, Tuzun E, Green ED, Eichler EE.** 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome research **13:**358-368.
128. **Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, Jorde LB.** 2009. Mobile elements create structural variation: analysis of a complete human genome. Genome research **19:**1516-1526.
129. **Kazazian HH, Jr.** 1999. An estimated frequency of endogenous insertional mutations in humans. Nat Genet **22:**130.
130. **Badge RM, Alisch RS, Moran JV.** 2003. ATLAS: a system to selectively identify human-specific L1 insertions. American journal of human genetics **72:**823-838.
131. **Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE.** 2004. Natural genetic variation caused by transposable elements in humans. Genetics **168:**933-951.
132. **Boissinot S, Entezam A, Young L, Munson PJ, Furano AV.** 2004. The insertional history of an active family of L1 retrotransposons in humans. Genome research **14:**1221-1231.
133. **Kazazian HH, Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE.** 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature **332:**164-166.
134. **Deininger PL, Batzer MA.** 1999. Alu repeats and human disease. Molecular genetics and metabolism **67:**183-193.
135. **Chen JM, Stenson PD, Cooper DN, Ferec C.** 2005. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Human genetics **117:**411-427.
136. **Callinan PA, Batzer MA.** 2006. Retrotransposable elements and human disease. Genome dynamics **1:**104-115.
137. **Belancio VP, Hedges DJ, Deininger P.** 2008. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome research **18:**343-358.
138. **Gasior SL, Wakeman TP, Xu B, Deininger PL.** 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. Journal of molecular biology **357:**1383-1393.
139. **Gilbert N, Lutz-Prigge S, Moran JV.** 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell **110:**315-325.
140. **Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD.** 2002. Human l1 retrotransposition is associated with genetic instability in vivo. Cell **110:**327-338.

141. **Gilbert N, Lutz S, Morrish TA, Moran JV.** 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Molecular and cellular biology **25:**7780-7795.
142. **Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA.** 2008. L1 recombination-associated deletions generate human genomic variation. Proceedings of the National Academy of Sciences of the United States of America **105:**19366-19371.
143. **Lee J, Han K, Meyer TJ, Kim HS, Batzer MA.** 2008. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. PloS one **3:**e4047.
144. **Belancio VP, Hedges DJ, Deininger P.** 2006. LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic acids research **34:**1512-1521.
145. **Belancio VP, Roy-Engel AM, Deininger P.** 2008. The impact of multiple splice sites in human L1 elements. Gene **411:**38-45.
146. **Han JS, Szak ST, Boeke JD.** 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature **429:**268-274.
147. **Perepelitsa-Belancio V, Deininger P.** 2003. RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet **35:**363-366.
148. **Lee JY, Ji Z, Tian B.** 2008. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. Nucleic acids research **36:**5581-5590.
149. **Speek M.** 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Molecular and cellular biology **21:**1973-1985.
150. **Hata K, Sakaki Y.** 1997. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. Gene **189:**227-234.
151. **Bailey JA, Carrel L, Chakravarti A, Eichler EE.** 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proceedings of the National Academy of Sciences of the United States of America **97:**6634-6639.
152. **LaRue RS, Andresdottir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jonsson SR, Landau NR, Lochelt M, Malik HS, Malim MH, Munk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu XF, Yuhki N, Harris RS.** 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. Journal of virology **83:**494-497.
153. **Wedekind JE, Dance GS, Sowden MP, Smith HC.** 2003. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends in genetics : TIG **19:**207-216.
154. **Harris RS, Petersen-Mahrt SK, Neuberger MS.** 2002. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Molecular cell **10:**1247-1253.
155. **Chan L.** 1992. Apolipoprotein B, the major protein component of triglyceride-rich and low density lipoproteins. The Journal of biological chemistry **267:**25621-25624.
156. **Teng B, Burant CF, Davidson NO.** 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science **260:**1816-1819.
157. **Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T.** 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. The Journal of biological chemistry **274:**18470-18476.

158. **Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T.** 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell **102:**553-563.
159. **Rogozin IB, Basu MK, Jordan IK, Pavlov YI, Koonin EV.** 2005. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell cycle **4:**1281-1285.
160. **Sato Y, Probst HC, Tatsumi R, Ikeuchi Y, Neuberger MS, Rada C.** 2010. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. The Journal of biological chemistry **285:**7111-7118.
161. **Liao W, Hong SH, Chan BH, Rudolph FB, Clark SC, Chan L.** 1999. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochemical and biophysical research communications **260:**398-404.
162. **Mikl MC, Watt IN, Lu M, Reik W, Davies SL, Neuberger MS, Rada C.** 2005. Mice deficient in APOBEC2 and APOBEC3. Molecular and cellular biology **25:**7270-7277.
163. **Bogerd HP, Wiegand HL, Doehle BP, Cullen BR.** 2007. The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. Virology **364:**486-493.
164. **Hache G, Liddament MT, Harris RS.** 2005. The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. The Journal of biological chemistry **280:**10920-10924.
165. **Navarro F, Bollman B, Chen H, Konig R, Yu Q, Chiles K, Landau NR.** 2005. Complementary function of the two catalytic domains of APOBEC3G. Virology **333:**374-386.
166. **Siu KK, Sultana A, Azimi FC, Lee JE.** 2013. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. Nature communications **4:**2593.
167. **Bohn MF, Shandilya SM, Albin JS, Kouno T, Anderson BD, McDougle RM, Carpenter MA, Rathore A, Evans L, Davis AN, Zhang J, Lu Y, Somasundaran M, Matsuo H, Harris RS, Schiffer CA.** 2013. Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. Structure **21:**1042-1050.
168. **Kitamura S, Ode H, Nakashima M, Imahashi M, Naganawa Y, Kurosawa T, Yokomaku Y, Yamane T, Watanabe N, Suzuki A, Sugiura W, Iwatani Y.** 2012. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. Nature structural & molecular biology **19:**1005-1010.
169. **Albin JS, LaRue RS, Weaver JA, Brown WL, Shindo K, Harjes E, Matsuo H, Harris RS.** 2010. A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. The Journal of biological chemistry **285:**40785-40792.
170. **Smith JL, Pathak VK.** 2010. Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. Journal of virology **84:**12599-12608.
171. **Huthoff H, Autore F, Gallois-Montbrun S, Fraternali F, Malim MH.** 2009. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. PLoS pathogens **5:**e1000330.
172. **Bach D, Peddi S, Mangeat B, Lakkaraju A, Strub K, Trono D.** 2008. Characterization of APOBEC3G binding to 7SL RNA. Retrovirology **5:**54.

173. **Bulliard Y, Turelli P, Rohrig UF, Zoete V, Mangeat B, Michielin O, Trono D.** 2009. Functional analysis and structural modeling of human APOBEC3G reveal the role of evolutionarily conserved elements in the inhibition of human immunodeficiency virus type 1 infection and Alu transposition. Journal of virology **83:**12611-12621.
174. **Iwatani Y, Takeuchi H, Strebel K, Levin JG.** 2006. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. Journal of virology **80:**5992-6002.
175. **Khan MA, Goila-Gaur R, Opi S, Miyagi E, Takeuchi H, Kao S, Strebel K.** 2007. Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. Retrovirology **4:**48.
176. **Shindo K, Li M, Gross PJ, Brown WL, Harjes E, Lu Y, Matsuo H, Harris RS.** 2012. A Comparison of Two Single-Stranded DNA Binding Models by Mutational Analysis of APOBEC3G. Biology **1:**260-276.
177. **Huthoff H, Malim MH.** 2007. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. Journal of virology **81:**3807-3815.
178. **Lavens D, Peelman F, Van der Heyden J, Uyttendaele I, Catteeuw D, Verhee A, Van Schoubroeck B, Kurth J, Hallenberger S, Clayton R, Tavernier J.** 2010. Definition of the interacting interfaces of Apobec3G and HIV-1 Vif using MAPPIT mutagenesis analysis. Nucleic acids research **38:**1902-1912.
179. **Mangeat B, Turelli P, Liao S, Trono D.** 2004. A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. The Journal of biological chemistry **279:**14481-14483.
180. **Schrofelbauer B, Chen D, Landau NR.** 2004. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proceedings of the National Academy of Sciences of the United States of America **101:**3927-3932.
181. **Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS.** 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. Journal of virology **85:**11220-11234.
182. **Li J, Chen Y, Li M, Carpenter MA, McDougle RM, Luengas EM, Macdonald PJ, Harris RS, Mueller JD.** 2014. APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. Journal of molecular biology **426:**1296-1307.
183. **Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS.** 2010. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic acids research **38:**4274-4284.
184. **Burnett A, Spearman P.** 2007. APOBEC3G multimers are recruited to the plasma membrane for packaging into human immunodeficiency virus type 1 virus-like particles in an RNA-dependent process requiring the NC basic linker. Journal of virology **81:**5000-5013.
185. **Khan MA, Goila-Gaur R, Kao S, Miyagi E, Walker RC, Jr., Strebel K.** 2009. Encapsidation of APOBEC3G into HIV-1 virions involves lipid raft association and does not correlate with APOBEC3G oligomerization. Retrovirology **6:**99.

186. **Wang T, Tian C, Zhang W, Luo K, Sarkis PT, Yu L, Liu B, Yu Y, Yu XF.** 2007. 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. Journal of virology **81:**13112-13124.
187. **Stauch B, Hofmann H, Perkovic M, Weisel M, Kopietz F, Cichutek K, Munk C, Schneider G.** 2009. Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. Proceedings of the National Academy of Sciences of the United States of America **106:**12079-12084.
188. **Cen S, Guo F, Niu M, Saadatmand J, Deflassieux J, Kleiman L.** 2004. The interaction between HIV-1 Gag and APOBEC3G. The Journal of biological chemistry **279:**33177-33184.
189. **Khan MA, Kao S, Miyagi E, Takeuchi H, Goila-Gaur R, Opi S, Gipson CL, Parslow TG, Ly H, Strebel K.** 2005. Viral RNA is required for the association of APOBEC3G with human immunodeficiency virus type 1 nucleoprotein complexes. Journal of virology **79:**5870-5874.
190. **Apolonia L, Schulz R, Curk T, Rocha P, Swanson CM, Schaller T, Ule J, Malim MH.** 2015. Promiscuous RNA binding ensures effective encapsidation of APOBEC3 proteins by HIV-1. PLoS pathogens **11:**e1004609.
191. **Bogerd HP, Cullen BR.** 2008. Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. Rna **14:**1228-1236.
192. **Strebel K, Khan MA.** 2008. APOBEC3G encapsidation into HIV-1 virions: which RNA is it? Retrovirology **5:**55.
193. **Tian C, Wang T, Zhang W, Yu XF.** 2007. Virion packaging determinants and reverse transcription of SRP RNA in HIV-1 particles. Nucleic acids research **35:**7288-7302.
194. **Wang T, Tian C, Zhang W, Sarkis PT, Yu XF.** 2008. Interaction with 7SL RNA but not with HIV-1 genomic RNA or P bodies is required for APOBEC3F virion packaging. Journal of molecular biology **375:**1098-1112.
195. **Zhen A, Du J, Zhou X, Xiong Y, Yu XF.** 2012. Reduced APOBEC3H variant anti-viral activities are associated with altered RNA binding activities. PloS one **7:**e38771.
196. **Alce TM, Popik W.** 2004. APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. The Journal of biological chemistry **279:**34083-34086.
197. **Douaisi M, Dussart S, Courcoul M, Bessou G, Vigne R, Decroly E.** 2004. HIV-1 and MLV Gag proteins are sufficient to recruit APOBEC3G into virus-like particles. Biochemical and biophysical research communications **321:**566-573.
198. **Schafer A, Bogerd HP, Cullen BR.** 2004. Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. Virology **328:**163-168.
199. **Svarovskaia ES, Xu H, Mbisa JL, Barr R, Gorelick RJ, Ono A, Freed EO, Hu WS, Pathak VK.** 2004. Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. The Journal of biological chemistry **279:**35822-35828.
200. **Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH.** 2003. DNA deamination mediates innate immunity to retroviral infection. Cell **113:**803-809.

201. **Liddament MT, Brown WL, Schumacher AJ, Harris RS.** 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Current biology : CB **14:**1385-1391.
202. **Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D.** 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature **424:**99-103.
203. **Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, Richman D, Coffin JM, Landau NR.** 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nature structural & molecular biology **11:**435-442.
204. **Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L.** 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature **424:**94-98.
205. **Chelico L, Pham P, Calabrese P, Goodman MF.** 2006. APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. Nature structural & molecular biology **13:**392-399.
206. **Langlois MA, Beale RC, Conticello SG, Neuberger MS.** 2005. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. Nucleic acids research **33:**1913-1923.
207. **Love RP, Xu H, Chelico L.** 2012. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. The Journal of biological chemistry **287:**30812-30822.
208. **Rausch JW, Chelico L, Goodman MF, Le Grice SF.** 2009. Dissecting APOBEC3G substrate specificity by nucleoside analog interference. The Journal of biological chemistry **284:**7047-7058.
209. **Kohli RM, Abrams SR, Gajula KS, Maul RW, Gearhart PJ, Stivers JT.** 2009. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. The Journal of biological chemistry **284:**22898-22904.
210. **Carpenter MA, Rajagurubandara E, Wijesinghe P, Bhagwat AS.** 2010. Determinants of sequence-specificity within human AID and APOBEC3G. DNA repair **9:**579-587.
211. **Wang M, Rada C, Neuberger MS.** 2010. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. The Journal of experimental medicine **207:**141-153.
212. **Kohli RM, Maul RW, Guminski AF, McClure RL, Gajula KS, Saribasak H, McMahon MA, Siliciano RF, Gearhart PJ, Stivers JT.** 2010. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. The Journal of biological chemistry **285:**40956-40964.
213. **Rathore A, Carpenter MA, Demir O, Ikeda T, Li M, Shaban NM, Law EK, Anokhin D, Brown WL, Amaro RE, Harris RS.** 2013. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. Journal of molecular biology **425:**4442-4454.
214. **Soros VB, Yonemoto W, Greene WC.** 2007. Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. PLoS pathogens **3:**e15.

215. **Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, Heidmann T, Greene WC.** 2006. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proceedings of the National Academy of Sciences of the United States of America **103:**15588-15593.
216. **Kozak SL, Marin M, Rose KM, Bystrom C, Kabat D.** 2006. The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. The Journal of biological chemistry **281:**29105-29119.
217. **Gallois-Montbrun S, Kramer B, Swanson CM, Byers H, Lynham S, Ward M, Malim MH.** 2007. Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. Journal of virology **81:**2165-2178.
218. **Gallois-Montbrun S, Holmes RK, Swanson CM, Fernandez-Ocana M, Byers HL, Ward MA, Malim MH.** 2008. Comparison of cellular ribonucleoprotein complexes associated with the APOBEC3F and APOBEC3G antiviral proteins. Journal of virology **82:**5636-5642.
219. **Goila-Gaur R, Khan MA, Miyagi E, Kao S, Opi S, Takeuchi H, Strebel K.** 2008. HIV-1 Vif promotes the formation of high molecular mass APOBEC3G complexes. Virology **372:**136-146.
220. **Holmes RK, Koning FA, Bishop KN, Malim MH.** 2007. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. The Journal of biological chemistry **282:**2587-2595.
221. **Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM.** 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Current biology : CB **15:**166-170.
222. **Bishop KN, Holmes RK, Malim MH.** 2006. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. Journal of virology **80:**8450-8458.
223. **Browne EP, Allers C, Landau NR.** 2009. Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. Virology **387:**313-321.
224. **Holmes RK, Malim MH, Bishop KN.** 2007. APOBEC-mediated viral restriction: not simply editing? Trends in biochemical sciences **32:**118-128.
225. **Kobayashi T, Koizumi Y, Takeuchi JS, Misawa N, Kimura Y, Morita S, Aihara K, Koyanagi Y, Iwami S, Sato K.** 2014. 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.
226. **Guo F, Cen S, Niu M, Yang Y, Gorelick RJ, Kleiman L.** 2007. The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA. Journal of virology **81:**11322-11331.
227. **Adolph MB, Webb J, Chelico L.** 2013. Retroviral restriction factor APOBEC3G delays the initiation of DNA synthesis by HIV-1 reverse transcriptase. PloS one **8:**e64196.
228. **Bishop KN, Verma M, Kim EY, Wolinsky SM, Malim MH.** 2008. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS pathogens **4:**e1000231.
229. **Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG.** 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic acids research **35:**7096-7108.

230. **Wang X, Ao Z, Chen L, Kobinger G, Peng J, Yao X.** 2012. The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. Journal of virology **86:**3777-3786.
231. **Chaurasiya KR, McCauley MJ, Wang W, Qualley DF, Wu T, Kitamura S, Geertsema H, Chan DS, Hertz A, Iwatani Y, Levin JG, Musier-Forsyth K, Rouzina I, Williams MC.** 2014. Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid-binding protein. Nature chemistry **6:**28-33.
232. **Luo K, Wang T, Liu B, Tian C, Xiao Z, Kappes J, Yu XF.** 2007. Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. Journal of virology **81:**7238-7248.
233. **Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR.** 2006. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proceedings of the National Academy of Sciences of the United States of America **103:**8780-8785.
234. **Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD.** 2006. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Current biology : CB **16:**480-485.
235. **Horn AV, Klawitter S, Held U, Berger A, Vasudevan AA, Bock A, Hofmann H, Hanschmann KM, Trosemeier JH, Flory E, Jabulowsky RA, Han JS, Lower J, Lower R, Munk C, Schumann GG.** 2014. Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. Nucleic acids research **42:**396-416.
236. **Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K.** 2007. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic acids research **35:**2955-2964.
237. **Lovsin N, Peterlin BM.** 2009. APOBEC3 proteins inhibit LINE-1 retrotransposition in the absence of ORF1p binding. Annals of the New York Academy of Sciences **1178:**268-275.
238. **Wissing S, Montano M, Garcia-Perez JL, Moran JV, Greene WC.** 2011. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. The Journal of biological chemistry **286:**36427-36437.
239. **Stenglein MD, Harris RS.** 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. The Journal of biological chemistry **281:**16837-16841.
240. **Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV.** 2014. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. eLife **3:**e02008.
241. **Monajemi M, Woodworth CF, Benkaroun J, Grant M, Larijani M.** 2012. Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment. Retrovirology **9:**35.
242. **Sadler HA, Stenglein MD, Harris RS, Mansky LM.** 2010. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. Journal of virology **84:**7396-7404.
243. **Kim EY, Lorenzo-Redondo R, Little SJ, Chung YS, Phalora PK, Maljkovic Berry I, Archer J, Penugonda S, Fischer W, Richman DD, Bhattacharya T, Malim MH, Wolinsky SM.** 2014. Human APOBEC3 Induced Mutation of Human Immunodeficiency

Virus Type-1 Contributes to Adaptation and Evolution in Natural Infection. PLoS pathogens **10:**e1004281.

244. **Jern P, Russell RA, Pathak VK, Coffin JM.** 2009. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS pathogens **5:**e1000367.
245. **Armitage AE, Deforche K, Welch JJ, Van Laethem K, Camacho R, Rambaut A, Iversen AK.** 2014. Possible footprints of APOBEC3F and/or other APOBEC3 deaminases - but not APOBEC3G - on HIV-1 from patients during acute/early and chronic infection. Journal of virology.
246. **Sato K, Takeuchi JS, Misawa N, Izumi T, Kobayashi T, Kimura Y, Iwami S, Takaori-Kondo A, Hu WS, Aihara K, Ito M, An DS, Pathak VK, Koyanagi Y.** 2014. APOBEC3D and APOBEC3F Potently Promote HIV-1 Diversification and Evolution in Humanized Mouse Model. PLoS pathogens **10:**e1004453.
247. **Johnson VA, Calvez V, Gunthard HF, Paredes R, Pillay D, Shafer R, Wensing AM, Richman DD.** 2011. 2011 update of the drug resistance mutations in HIV-1. Topics in antiviral medicine **19:**156-164.
248. **Berkhout B, de Ronde A.** 2004. APOBEC3G versus reverse transcriptase in the generation of HIV-1 drug-resistance mutations. Aids **18:**1861-1863.
249. **Mulder LC, Harari A, Simon V.** 2008. Cytidine deamination induced HIV-1 drug resistance. Proceedings of the National Academy of Sciences of the United States of America **105:**5501-5506.
250. **Fourati S, Malet I, Lambert S, Soulie C, Wirden M, Flandre P, Fofana DB, Sayon S, Simon A, Katlama C, Calvez V, Marcelin AG.** 2012. E138K and M184I mutations in HIV-1 reverse transcriptase coemerge as a result of APOBEC3 editing in the absence of drug exposure. Aids **26:**1619-1624.
251. **Shankar P, Russo M, Harnisch B, Patterson M, Skolnik P, Lieberman J.** 2000. Impaired function of circulating HIV-specific CD8(+) T cells in chronic human immunodeficiency virus infection. Blood **96:**3094-3101.
252. **Heintel T, Sester M, Rodriguez MM, Krieg C, Sester U, Wagner R, Pees HW, Gartner B, Maier R, Meyerhans A.** 2002. The fraction of perforin-expressing HIV-specific CD8 T cells is a marker for disease progression in HIV infection. Aids **16:**1497-1501.
253. **Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S, Segal JP, Cao Y, Rowland-Jones SL, Cerundolo V, Hurley A, Markowitz M, Ho DD, Nixon DF, McMichael AJ.** 1998. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. Science **279:**2103-2106.
254. **Barouch DH, Letvin NL.** 2002. Viral evolution and challenges in the development of HIV vaccines. Vaccine **20 Suppl 4:**A66-68.
255. **Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, Shaw GM.** 1997. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nature medicine **3:**205-211.
256. **Geels MJ, Cornelissen M, Schuitemaker H, Anderson K, Kwa D, Maas J, Dekker JT, Baan E, Zorgdrager F, van den Burg R, van Beelen M, Lukashov VV, Fu TM, Paxton WA, van der Hoek L, Dubey SA, Shiver JW, Goudsmit J.** 2003. Identification

of sequential viral escape mutants associated with altered T-cell responses in a human immunodeficiency virus type 1-infected individual. Journal of virology **77:**12430-12440.

257. **Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, Daniels M, Ferrari G, Haynes BF, McMichael A, Shaw GM, Hahn BH, Korber B, Seoighe C.** 2009. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. PLoS pathogens **5:**e1000414.
258. **Monajemi M, Woodworth CF, Zipperlen K, Gallant M, Grant MD, Larijani M.** 2014. Positioning of APOBEC3G/F mutational hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells. PloS one **9:**e93428.
259. **Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, Klimczak LJ, Kryukov GV, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA.** 2012. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Molecular cell **46:**424-435.
260. **Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jonsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerod A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Borresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR, Breast Cancer Working Group of the International Cancer Genome C.** 2012. Mutational processes molding the genomes of 21 breast cancers. Cell **149:**979-993.
261. **Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, McDougle RM, Brown WL, Law EK, Harris RS.** 2013. APOBEC3B is an enzymatic source of mutation in breast cancer (vol 494, 366, 2013). Nature **502:**580-580.
262. **Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jager N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR, Genome APC, Consortium IBC, Consortium IM-S, PedBrain I.** 2013. Signatures of mutational processes in human cancer (vol 500, pg 415, 2013). Nature **502**.
263. **Burns MB, Temiz NA, Harris RS.** 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet **45:**977-983.
264. **Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, Getz G, Gordenin DA.** 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet **45:**970-+.

265. **Stopak KS, Chiu YL, Kropp J, Grant RM, Greene WC.** 2007. Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells. The Journal of biological chemistry **282:**3539-3546.
266. **Chiu YL, Soros VB, Kreisberg JF, Stopak K, Yonemoto W, Greene WC.** 2005. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature **435:**108-114.
267. **OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M.** 2008. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell host & microbe **4:**249-259.
268. **Sliva K, Erlwein O, Bittner A, Schnierle BS.** 2004. Murine leukemia virus (MLV) replication monitored with fluorescent proteins. Virology journal **1:**14.
269. **Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, Gage FH, Verma IM, Trono D.** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science **272:**263-267.
270. **Zhang H, Zhou Y, Alcock C, Kiefer T, Monie D, Siliciano J, Li Q, Pham P, Cofrancesco J, Persaud D, Siliciano RF.** 2004. 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.
271. **Simon JH, Malim MH.** 1996. The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes. Journal of virology **70:**5297-5305.
272. **Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV, Kazazian HH, Jr.** 2000. Determination of L1 retrotransposition kinetics in cultured cells. Nucleic acids research **28:**1418-1423.
273. **Grandgenett DP, Goodarzi G.** 1994. Folding of the multidomain human immunodeficiency virus type-I integrase. Protein science : a publication of the Protein Society **3:**888-897.
274. **Szilvay AM, Nornes S, Haugan IR, Olsen L, Prasad VR, Endresen C, Goff SP, Helland DE.** 1992. Epitope mapping of HIV-1 reverse transcriptase with monoclonal antibodies that inhibit polymerase and RNase H activities. Journal of acquired immune deficiency syndromes **5:**647-657.
275. **Suspene R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP.** 2005. Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. The Journal of general virology **86:**125-129.
276. **Brussel A, Delelis O, Sonigo P.** 2005. Alu-LTR real-time nested PCR assay for quantifying integrated HIV-1 DNA. Methods in molecular biology **304:**139-154.
277. **Albin JS, Harris RS.** 2010. Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert reviews in molecular medicine **12:**e4.
278. **Armitage AE, Deforche K, Chang CH, Wee E, Kramer B, Welch JJ, Gerstoft J, Fugger L, McMichael A, Rambaut A, Iversen AK.** 2012. APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. PLoS genetics **8:**e1002550.
279. **Li XY, Guo F, Zhang L, Kleiman L, Cen S.** 2007. APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription. The Journal of biological chemistry **282:**32065-32074.

280. **Gooch BD, Cullen BR.** 2008. Functional domain organization of human APOBEC3G. Virology **379:**118-124.
281. **Bogerd HP, Doehle BP, Wiegand HL, Cullen BR.** 2004. A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. Proceedings of the National Academy of Sciences of the United States of America **101:**3770-3774.
282. **Xu H, Svarovskaia ES, Barr R, Zhang Y, Khan MA, Strebel K, Pathak VK.** 2004. A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proceedings of the National Academy of Sciences of the United States of America **101:**5652-5657.
283. **Zhang KL, Mangeat B, Ortiz M, Zoete V, Trono D, Telenti A, Michielin O.** 2007. Model structure of human APOBEC3G. PloS one **2:**e378.
284. **Salter JD, Krucinska J, Raina J, Smith HC, Wedekind JE.** 2009. A hydrodynamic analysis of APOBEC3G reveals a monomer-dimer-tetramer self-association that has implications for anti-HIV function. Biochemistry **48:**10685-10687.
285. **Stenglein MD, Matsuo H, Harris RS.** 2008. Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. Journal of virology **82:**9591-9599.
286. **Ao Z, Yu Z, Wang L, Zheng Y, Yao X.** 2008. Vpr14-88-Apobec3G fusion protein is efficiently incorporated into Vif-positive HIV-1 particles and inhibits viral infection. PloS one **3:**e1995.
287. **Zhang S, Pointer D, Singer G, Feng Y, Park K, Zhao LJ.** 1998. Direct binding to nucleic acids by Vpr of human immunodeficiency virus type 1. Gene **212:**157-166.
288. **Gelderblom HC, Vatakis DN, Burke SA, Lawrie SD, Bristol GC, Levy DN.** 2008. Viral complementation allows HIV-1 replication without integration. Retrovirology **5:**60.
289. **Wodarz D, Levy DN.** 2009. Multiple HIV-1 infection of cells and the evolutionary dynamics of cytotoxic T lymphocyte escape mutants. Evolution; international journal of organic evolution **63:**2326-2339.
290. **Levy DN, Aldrovandi GM, Kutsch O, Shaw GM.** 2004. Dynamics of HIV-1 recombination in its natural target cells. Proceedings of the National Academy of Sciences of the United States of America **101:**4204-4209.
291. **Iwabu Y, Mizuta H, Kawase M, Kameoka M, Goto T, Ikuta K.** 2008. Superinfection of defective human immunodeficiency virus type 1 with different subtypes of wild-type virus efficiently produces infectious variants with the initial viral phenotypes by complementation followed by recombination. Microbes and infection / Institut Pasteur **10:**504-513.
292. **Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ.** 2006. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. Journal of virology **80:**2472-2482.
293. **Russell RA, Moore MD, Hu WS, Pathak VK.** 2009. 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.

294. **Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH, Jr.** 1997. Many human L1 elements are capable of retrotransposition. Nat Genet **16:**37-43.
295. **Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Careri D, Pinto MT, Kasper CK, Sommer SS.** 2001. Frequency of recent retrotransposition events in the human factor IX gene. Human mutation **17:**511-519.
296. **Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, Minami R, Nakamura H, Matsuo M.** 1993. Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. The Journal of clinical investigation **91:**1862-1867.
297. **Niewiadomska AM, Tian C, Tan L, Wang T, Sarkis PT, Yu XF.** 2007. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. Journal of virology **81:**9577-9583.
298. **Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C.** 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. The Journal of biological chemistry **281:**22161-22172.
299. **Koyama T, Arias JF, Iwabu Y, Yokoyama M, Fujita H, Sato H, Tokunaga K.** 2013. APOBEC3G oligomerization is associated with the inhibition of both Alu and LINE-1 retrotransposition. PloS one **8:**e84228.
300. **Tan L, Sarkis PT, Wang T, Tian C, Yu XF.** 2009. Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **23:**279-287.
301. **Wang X, Dolan PT, Dang Y, Zheng YH.** 2007. Biochemical differentiation of APOBEC3F and APOBEC3G proteins associated with HIV-1 life cycle. The Journal of biological chemistry **282:**1585-1594.
302. **Aydin H, Taylor MW, Lee JE.** 2014. Structure-guided analysis of the human APOBEC3-HIV restrictome. Structure **22:**668-684.
303. **Hache G, Mansky LM, Harris RS.** 2006. Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. AIDS reviews **8:**148-157.
304. **Ara A, Love RP, Chelico L.** 2014. Different mutagenic potential of HIV-1 restriction factors APOBEC3G and APOBEC3F is determined by distinct single-stranded DNA scanning mechanisms. PLoS pathogens **10:**e1004024.
305. **Zheng YH, Irwin D, Kurosu T, Tokunaga K, Sata T, Peterlin BM.** 2004. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. Journal of virology **78:**6073-6076.
306. **Wichroski MJ, Robb GB, Rana TM.** 2006. Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. PLoS pathogens **2:**e41.
307. **Chelico L, Prochnow C, Erie DA, Chen XS, Goodman MF.** 2010. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. The Journal of biological chemistry **285:**16195-16205.
308. **Belanger K, Savoie M, Rosales Gerpe MC, Couture JF, Langlois MA.** 2013. Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. Nucleic acids research **41:**7438-7452.

309. **Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS.** 2008. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature **456:**121-124.
310. **Miyagi E, Brown CR, Opi S, Khan M, Goila-Gaur R, Kao S, Walker RC, Jr., Hirsch V, Strebel K.** 2010. Stably expressed APOBEC3F has negligible antiviral activity. Journal of virology **84:**11067-11075.
311. **Holtz CM, Sadler HA, Mansky LM.** 2013. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. Nucleic acids research **41:**6139-6148.
312. **Berkowitz RD, Goff SP.** 1993. Point mutations in Moloney murine leukemia virus envelope protein: effects on infectivity, virion association, and superinfection resistance. Virology **196:**748-757.
313. **Tsien RY.** 1998. The green fluorescent protein. Annual review of biochemistry **67:**509-544.
314. **Ehrig T, O'Kane DJ, Prendergast FG.** 1995. Green-fluorescent protein mutants with altered fluorescence excitation spectra. FEBS letters **367:**163-166.
315. **Conticello SG.** 2008. The AID/APOBEC family of nucleic acid mutators. Genome biology **9:**229.
316. **Nik-Zainal S, Alexandrov L, Wedge D, Van Loo P, Raine K, Jones DR, Futreal PA, Campbell PJ, Stratton MR, Consortium ICG.** 2012. Cancer Genomics, Epigenetics and Genomic Instability. Mutational Processes Shaping the Genomes of Twenty-one Breast Cancers. Eur J Cancer **48:**S144-S144.
317. **Pavri R, Nussenzweig MC.** 2011. AID Targeting in Antibody Diversity. Adv Immunol **110:**1-26.
318. **Suspene R, Aynaud MM, Guetard D, Henry M, Eckhoff G, Marchio A, Pineau P, Dejean A, Vartanian JP, Wain-Hobson S.** 2011. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proceedings of the National Academy of Sciences of the United States of America **108:**4858-4863.
319. **Vartanian JP, Guetard D, Henry M, Wain-Hobson S.** 2008. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science **320:**230-233.
320. **Vossen RH, Aten E, Roos A, den Dunnen JT.** 2009. High-resolution melting analysis (HRMA): more than just sequence variant screening. Human mutation **30:**860-866.
321. **Wittwer CT.** 2009. High-resolution DNA melting analysis: advancements and limitations. Human mutation **30:**857-859.
322. **Reed GH, Wittwer CT.** 2004. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clinical chemistry **50:**1748-1754.
323. **Wojdacz TK, Dobrovic A.** 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic acids research **35:**e41.
324. **Piantadosi A, Chohan B, Panteleeff D, Baeten JM, Mandaliya K, Ndinya-Achola JO, Overbaugh J.** 2009. HIV-1 evolution in gag and env is highly correlated but exhibits different relationships with viral load and the immune response. Aids **23:**579-587.

325. **Suspene R, Rusniok C, Vartanian JP, Wain-Hobson S.** 2006. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucleic acids research **34:**4677-4684.
326. **Boi S, Kolokithas A, Shepard J, Linwood R, Rosenke K, Van Dis E, Malik F, Evans LH.** 2014. Incorporation of mouse APOBEC3 into murine leukemia virus virions decreases the activity and fidelity of reverse transcriptase. Journal of virology **88:**7659-7662.
327. **Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R.** 2015. Extremely High Mutation Rate of HIV-1 In Vivo. PLoS biology **13:**e1002251.
328. **Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, Harris RS, Matsuo H.** 2008. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature **452:**116-119.
329. **Harjes E, Gross PJ, Chen KM, Lu Y, Shindo K, Nowarski R, Gross JD, Kotler M, Harris RS, Matsuo H.** 2009. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. Journal of molecular biology **389:**819-832.
330. **Furukawa A, Nagata T, Matsugami A, Habu Y, Sugiyama R, Hayashi F, Kobayashi N, Yokoyama S, Takaku H, Katahira M.** 2009. Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. The EMBO journal **28:**440-451.
331. **Feng Y, Chelico L.** 2011. Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain. The Journal of biological chemistry **286:**11415-11426.
332. **Polevoda B, McDougall WM, Tun BN, Cheung M, Salter JD, Friedman AE, Smith HC.** 2015. RNA binding to APOBEC3G induces the disassembly of functional deaminase complexes by displacing single-stranded DNA substrates. Nucleic acids research.
333. **Li X, Ma J, Zhang Q, Zhou J, Yin X, Zhai C, You X, Yu L, Guo F, Zhao L, Li Z, Zeng Y, Cen S.** 2011. Functional analysis of the two cytidine deaminase domains in APOBEC3G. Virology **414:**130-136.
334. **Chen KM, Martemyanova N, Lu Y, Shindo K, Matsuo H, Harris RS.** 2007. Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. FEBS letters **581:**4761-4766.
335. **Dapp MJ, Clouser CL, Patterson S, Mansky LM.** 2009. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. Journal of virology **83:**11950-11958.
336. **Dapp MJ, Holtz CM, Mansky LM.** 2012. Concomitant lethal mutagenesis of human immunodeficiency virus type 1. Journal of molecular biology **419:**158-170.
337. **Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI.** 1999. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. Proceedings of the National Academy of Sciences of the United States of America **96:**1492-1497.
338. **Smith RA, Loeb LA, Preston BD.** 2005. Lethal mutagenesis of HIV. Virus research **107:**215-228.
339. **Harris KS, Brabant W, Styrchak S, Gall A, Daifuku R.** 2005. KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. Antiviral research **67:**1-9.

340. **Dapp MJ, Heineman RH, Mansky LM.** 2013. Interrelationship between HIV-1 fitness and mutation rate. Journal of molecular biology **425:**41-53.
341. **Ardekani AM, Naeini MM.** 2010. The Role of MicroRNAs in Human Diseases. Avicenna journal of medical biotechnology **2:**161-179.
342. **Khatua AK, Taylor HE, Hildreth JE, Popik W.** 2009. Exosomes packaging APOBEC3G confer human immunodeficiency virus resistance to recipient cells. Journal of virology **83:**512-521.
343. **Khatua AK, Taylor HE, Hildreth JE, Popik W.** 2010. Inhibition of LINE-1 and Alu retrotransposition by exosomes encapsidating APOBEC3G and APOBEC3F. Virology **400:**68-75.
344. **Madison MN, Roller RJ, Okeoma CM.** 2014. Human semen contains exosomes with potent anti-HIV-1 activity. Retrovirology **11:**102.
345. **Mittelbrunn M, Gutierrez-Vazquez C, Villarroya-Beltri C, Gonzalez S, Sanchez-Cabo F, Gonzalez MA, Bernad A, Sanchez-Madrid F.** 2011. Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. Nature communications **2:**282.
346. **Stoorvogel W.** 2012. Functional transfer of microRNA by exosomes. Blood **119:**646-648.
347. **Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO.** 2007. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature cell biology **9:**654-659.
348. **Zhang J, Li S, Li L, Li M, Guo C, Yao J, Mi S.** 2015. Exosome and exosomal microRNA: trafficking, sorting, and function. Genomics, proteomics & bioinformatics **13:**17-24.
349. **Fabbri M, Paone A, Calore F, Galli R, Gaudio E, Santhanam R, Lovat F, Fadda P, Mao C, Nuovo GJ, Zanesi N, Crawford M, Ozer GH, Wernicke D, Alder H, Caligiuri MA, Nana-Sinkam P, Perrotti D, Croce CM.** 2012. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. Proceedings of the National Academy of Sciences of the United States of America **109:**E2110-2116.

Contribution of Collaborators

The GST-W127A expression vector was provided by Dr. Jean-François Couture from the University of Ottawa who also generated the model structure of the tryptophan mutants identified in Chapter 3 of this thesis presented in Figure 5.2.2.

Maria Rosales Gerpe achieved the complementation experiment in the presence of HIV[p8.9] presented in Figure 3.2.7 of this thesis.

Tyler Milston Renner performed the 3D-PCR analysis presented in Figure 6.2.1 of this thesis.

Zahra Montazeri established the algorithm used to determine the number of mutation in a given clone based on its fluorescence at a selected temperature presented in section 2.18 of this thesis and applied throughout Chapter 6.

Cloning of the A3 proteins in Flag-C3 was performed in collaboration with Mark Campbell and used to performed experiments presented in Chapter 4 of this thesis with the help of Laura Goodwin. Both students worked under my supervision for the accomplishment of their Honour's project.

Appendices

Appendix I:

Table S1 List of primers

Primer Name	Sequence 5' to 3'
	Primers for cloning
HA-NheI-FWD	GAATTCGCTAGCACCACCATGGGATACCCATACGAC-GTCCCAGACTACGCTATGAAGCCTCACTTCAGA
Flag-A3A-FWD	GCTAGCACCACCATGGGCATGGACTACAAGGACGA-TGATGACAAGATGGAAGCCAGCCCAGCA
Flag-A3H-FWD	GCTAGCACCACCATGGGCATGGACTACAAGGACGA-TGATGACAAGATGGCTCTGTTAACAGCC
A3-HpaI-Rev	ATAAACAAGTTAACAACAACAA
A127W-GST-FWD	CGTCTGTATTACTTTTGGGATCCGGACTATCAG
A127W-GST-REV	CTGATAGTCCGGATCCCAAAAGTAATACAGACG
W94A-GST-FWD	ACCTGGTACATTAGCGCGTCTCCGTGCACCAAATG
W94A-GST-REV	CATTTGGTGCACGGAGACGCGCTAATGTACCAGGT
Vpr-XhoI-FWD	GAGACCTCGAGATGCCATACAATGAATGGACACT
Vpr^{14-88}-XhoI-REV	GAGACCTCGAGTCTCCTCTGTCGAGTAACG
Vpr^{14-86}-XhoI-REV	GAGACCTCGAGCTGTCGAGTAACG
A3D-XhoI-FWD	CTCGAGATGAATCCACAGATCAGAAATC
A3D-SacII-REV	CCGCGGTCACTGGAGAATCTCCCGTAGC
ENV-NheI-FWD	GCTAGCATGGCGCGTTCAACG
ENV-HindII-REV	AAGCTTTATGGCTCGTACTCTATAGG
eGFP-NheI-FWD	GCTAGCTATCCACCGGTCGCCACCAT
eGFP-HindII-REV	AAGCTTGCAGGTCGACTCTAGAGT

	Primers for site-directed mutagenesis (SDM)
A3A-W98A-FWD	ACTTGGTTCATCTCCGCGAGCCCCTGCTTCTCC
A3A-W98A-REV	GGAGAAGCAGGGGCTCGCGGAGATGAACCAAGT
A3B-W94A-FWD	ACCTGGTTTGTATCCGCGACCCCCTGCCCGGAC
A3B-W94A-REV	GTCCGGGCAGGGGGTCGCGGATACAAACCAGGT
A3B-W127A-FWD	CGCCTCTACTACTACGCGGAAAGAGATTACCGA
A3B-W127A-REV	TCGGTAATCTCTTTCCGCGTAGTAGTAGAGGCG
A3C-W94A-FWD	ACCTGGTACACATCTGCGAGCCCTTGCCCAGAC
A3C-W94A-REV	GTCTGGGCAAGGGCTCGCAGATGTGTACCAGGT
A3D-W106A-FWD	ACCTGGTTTGTATCAGCGAACCCCTGCCTGCCC
A3D-W106A-REV	GGGCAGGCAGGGGTTCGCTGATACAAACCAGGT
A3D-W143A-FWD	TACCGGGATAGAGATGCGCGGTGGGTGCTCCTC
A3D-W143A-REV	GAGGAGCACCCACCGCGCATCTCTATCCCGGTA
A3F-W93A-FWD	ACCTGGTTTGTATCCGCGACCCCCTGCCCGGAC
A3F-W93A-REV	GTCCGGGCAGGGGGTCGCGGATACAAACCAGGT
A3F-W126A-FWD	CGCCTCTACTACTACGCGGAAAGAGATTACCGA
A3F-W126A-REV	TCGGTAATCTCTTTCCGCGTAGTAGTAGAGGCG
A3G-E67Q-FWD	CTTAAGTACCACCCACAGATGAGATTCTTCCAC
A3G-E67Q-REV	GTGGAAGAATCTCATCTGTGGGTGGTACTTAAG
A3G-W94A-FWD	ACCTGGTACATATCCGCGAGCCCCTGCACAAAG
A3G-W94A-REV	CTTTGTGCAGGGGCTCGCGGATATGTACCAGGT
A3G-C100S-FWD	AGCCCCTGCACAAAGTCTACAAGGGATATGGCC
A3G-C100S-REV	GGCCATATCCCTTGTAGACTTTGTGCAGGGGCT
A3G-Y124A-FWD	TTTGTTGCCCGCCTCGCCTACTTCTGGGACCCA

A3G-Y124A-REV	TGGGTCCCAGAAGTAGGCGAGGCGGGCAACAAA
A3G-W127A-FWD	CGCCTCTACTACTTCGCCGACCCAGATTACCAG
A3G-W127A-REV	CTGGTAATCTGGGTCGGCGAAGTAGTAGAGGCG
A3G-E259Q-FWD	GAAGGCCGCCATGCACAGCTGTGCTTCCTGGAC
A3G-E259Q-REV	GTCCAGGAAGCACAGCTGTGCATGGCGGCCTTC
A3G-W285A-FWD	ACCTGCTTCACCTCCGCGAGCCCCTGCTTCAGC
A3G-W285A-REV	GCTGAAGCAGGGGCTCGCGGAGGTGAAGCAGGT
A3G-C291S-FWD	AGCCCCTGCTTCAGCAGTGCCCAGGAAATGGCT
A3G-C291S-REV	AGCCATTTCCTGGGCACTGCTGAAGCAGGGGCT
A3G-R313A-FWD	TGCATCTTCACTGCCGCCATCTATGATGATCAA
A3G-R313A-REV	TTGATCATCATAGATGGCGGCAGTGAAGATGCA
A3H-W82A-FWD	ACCTGTTACCTCACGGCGAGCCCCTGCTCCTCC
A3H-W82A-REV	GGAGGAGCAGGGGCTCGCCGTGAGGTAACAGGT
A3H-G105R-FWD	GACCATCTGAACCTGCGCATCTTCGCCTCCCGC
A3H-G105R-REV	GCGGGAGGCGAAGATGCGCAGGTTCAGATGGTC
	Primers for HRM analysis
M-MLV-R2-FWD	GGGGCTCGTCCGGGATC
M-MLV-R2-REV	TTGATCTTAACCTGGG
M-MLV-R3-FWD	CTTTAACCGAGACCTCATCA
M-MLV-R3-REV	TCTCCGGTCAGCAGAGTC
M-MLV-R4-FWD	GACTGTCAGCAGCTGTTGGG
M-MLV-R4-REV	TCGTTCTCCTCCCTGTCT
M-MLV-R5-FWD	GTTAGTGGACAGAAACAGGAT
M-MLV-R5-REV	GATCAGAGGAGCTTGGCGAA

M-MLV-R6-FWD	GGGGCATGGGACTGGCAG
M-MLV-R6-REV	GGCAAATTTGGGCTTTCTTG
M-MLV-R7-FWD	TCGGGTATCGGGCCTCGGC
M-MLV-R7-REV	CACCGGTCCGAACTGGAC
M-MLV-R8-FWD	TGCTTTTGGACACGGACCGG
M-MLV-R8-REV	CAGTCACTGTGTAATGAA
M-MLV-R9-FWD	CCTACACCTCAGAACATT
M-MLV-R9-REV	TAGTTAAAGTCTCCTTGATG
M-MLV-R10-FWD	GTAGAAAGAATGAATAGAAC
M-MLV-R10-REV	CCGATCTCCATTGGTTA
M-MLV-R11-FWD	ATAATATCACCTGGGAGG
M-MLV-R11-REV	CTATTGGGACGCGGGGT
M-MLV-R12-FWD	TCAGATACCAAAATCTAGG
M-MLV-R12-REV	TGTTCTGCTGGTAGTGGTC
M-MLV-R13-FWD	CGGCAGCGTGCAGCTCGCC
M-MLV-R13-REV	ACAGGCCGTAAACATAGC
M-MLV-R14-FWD	GTCACCTATCATTCCCCCA
M-MLV-R14-REV	GAGACTAAATAAAATCTTTTAT
M-MLV-R15-FWD	GAGTACGAGCCATAGATAAA
M-MLV-R15-REV	CATCGTCCTTGTAGTCCAT
EGFP-100-FWD	TGCTGCTGCCCGACAACCACTAC
EGFP-279-FWD	ACAACAGCCACAACGTCTATATCAT
EGFP-428-FWD	GCCATGCCCGAAGGCT
EGFP-523-FWD	TACGGCGTGCAGTGCTTC
EGFP-648-FWD	CGGCCACAAGTTCAGCGT

EGFP-800-FWD	TGCTTCCCGAGCTCTATAAAA
EGFP-900-FWD	TCAGATGTTTCCAGGCTCCC
EGFP-1000-FWD	ATAGCTAACGTTGGGCCAAAC
EGFP-HRM-REV	CGTCCATGCCGAGAGTGAT
M-MLV-A-FWD	TGAAAGACCCCACCTGTAGGTTTGGCAAGC
M-MLV-A-REV	ACCCGTTGTTTTTCTTCTCCGGTCAGCAGA
M-MLV-B-FWD	CTCATCACCCATCAGCCCACCTGGGACGAC
M-MLV-B-REV	ATACCCCAGATACTTGACCTGTTTCTGGCA
M-MLV-C-FWD	GGCCCTGTTACAAACCCTAGGGAACCTCG
M-MLV-C-REV	CAGTTGCAAGCGTTAATTTAGTTAAAGTCTC
M-MLV-D-FWD	AGCTCAGGCCAGGTAGAAAGAATGAATAGA
M-MLV-D-REV	CGTTCCCGCCGGTACGTACGCACCGGTGGA
M-MLV-E-FWD	CAACATCGAGGACGGCAGCGTGCAGCTCGC
M-MLV-E-REV	AGCAGATTGTACTGAGAGTGCACCATATGA
EGFP-FWD	GTGAGCAAGGGCGAGGAGCTGTTCA
EGFP-REV	CTTGTACAGCTCGTCCATGCCGAGA
HIV-R1-FWD	TGGAAGGGCTAATTCGGTCCCA
HIV-R1-REV	GCTTTACTTTCGCTTTCAAGT
HIV-R2-FWD	GTGTGGAAAATCTCTAGCAG
HIV-R2-REV	GCATTTAAAGTTCTAGGTGA
HIV-R3-FWD	ATGGTACATCAGGCCATATCA
HIV-R3-REV	CTCCCTGACATGCTGTCATCA
HIV-R4-FWD	AGGAGCGACACTAGAAGAAA
HIV-R4-REV	ATGAGTATCTGATCATACTGT
HIV-R5-FWD	TTGGAGGTTTTATCAAAGTA

HIV-R5-REV	ACACTGGAATATTGCTGGTG
HIV-R6-FWD	CAGTACAATGTGCTTCCACA
HIV-R6-REV	TCACATCATTAGTGTGGGCA
HIV-R7-FWD	CAAGAGCCATTTAAAAATCTGA
HIV-R7-REV	CCATCTAAAAATAGTACTTT
HIV-R8-FWD	TGGAGGAAATGAACAAGTAGAT
HIV-R8-REV	CTATTATGTCTACTATTCTTT
HVI-R9-FWD	CATCCACAATTTTAAAAGAA
HIV-R9-REV	CTAATATGGTATTTCTTATAG
HIV-R10-FWD	TTGATTGTTTTTCAGAATCTG
HIV-R10-REV	TTCTGATGAGCTCTCGTCGC
HIV-R11-FWD	AAAGCCTTAGGCATCTCCTAC
HIV-R11-REV	TCTCCATTATCATTCTCCCGC
HIV-R12-FWD	CCCCACTCTGTGTTAGTTTAA
HIV-R12-REV	ACAATTAAAACTGTGCGTTAC
HIV-R13-FWD	ATAATCTTTAAGCAATCCTCA
HIV-R13-REV	CTCGCCTGGAGCTGTTTGATG
HIV-R14-FWD	CAGCATCTGTTGCAACTCACA
HIV-R14-REV	AGTAAGTCTCTCAAGCGGTGG
HIV-R15-FWD	GAGCCTGTGCCTCTTCAGCTA
HIV-R15-REV	TCTTCTTTGGGAGTGAATTAG
HIV-R16-FWD	ACAAGGCAGCTGTAGATCTTA
HV-R16-REV	TGCTAGAGATTTTCCACACTG
HIV-A-REV	AACTCTTGCTTTATGGCCGGG
HIV-B-FWD	CCAAAATGCGAACCCAGATTG

HIV-B-REV	TTGGGCCTTATCTATTCCATC
HIV-C-FWD	GGCATGGGTACCAGCACACAA
HIV-C-REV	TGATGAGTCTGACTGTTCTGA
HIV-D-FWD	GTGTTGCTTTCATTGCCAAGT
HIV-D-REV	TCCACAGCCAGGATTCTTGCC
HIV-E-FWD	ATTTGCTGAGGGCTATTGAGG
HIV-POL-FWD	ATTTTAACCTGCCACCTGTAGTAGC
HIV-POL-REV	ATACTGCCATTTGTACTGCTG
	Primers for qPCR and ddPCR
qeGFP-FWD	GTGAGCAAGGGCGAGGAGCTGTTCA
qIN-eGFP-FWD	TGCTGCTGCCCGACAAC
qIN-eGFP-REV	CGTCCATGCCGAGAGTGAT
qIN-eGFP-Probe	FAM-CACCCAGTCCGCC
qAlu-1	TCCCAGCTACTGGGGAGGCTGAGG
qAlu-2	GCCTCCCAAAGTGCTGGGATTACAG
qB1-1	CCAGGACACCAGGGCTACAGAG
qB1-2	CCCGAGTGCTGGGATTAAAG
qHIV-FWD	CAAGTAGTGTGTGCCCGTCTGT
qHIV-REV	CGAGTCCTGCGTCGAGAGA
qHIV-Probe	FAM-CAGTGGCGCCCGAA
qM-MLV-FWD	TTGGGAGGGTCTCCTCTGAGT
qM-MLV-REV	GAGTGGTAACAGTCTGGCCCATA
qM-MLV-Probe	ATTGACTACCCGTCAGCGG
7SL-FWD	ATCGGGTGTCCGCACTAAG

7SL-REV	CACCCCTCCTTAGGCAACCT
Alu-FWD	TCACGCCTGTAATCCCAGC
Alu-REV	GATCTCGGCTCACTGCAAG
hY1-FWD	CTGGTCCGAAGGTAGTGAG
hY1-REV	CTAGTCAAGTGCAGTAGTG
hY3-FWD	GCTGGTCCGAGTGCAGTGGTGTTTAC
hY3-REV	AGGCTAGTCAAGTGAAGCAGTG

Appendix II: Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses.

Kasandra Bélanger Mathieu Savoie, Maria Carla Rosales Gerpe, Jean-François Couture and Marc-André Langlois.

Contribution:

I performed most of the experiments presented in the manuscript. Maria Rosales Gerpe assisted with the complementation assay in the presence of HIV[p8.9]. I contributed to the assembly of the manuscript, proofread the manuscript, and assisted in the response to reviewers.

Published: *Nucleic Acid Research*, June 2013

7438–7452 Nucleic Acids Research, 2013, Vol. 41, No. 15 Published online 12 June 2013
doi:10.1093/nar/gkt527

Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses

Kasandra Bélanger[1], Mathieu Savoie[1], María Carla Rosales Gerpe[1], Jean-François Couture[1,2] and Marc-André Langlois[1,3,4,*]

[1]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5, [2]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5, [3]Emerging Pathogens Research Centre, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5 and [4]Department of Pathology and Laboratory Medicine, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5

Received January 24, 2013; Revised May 19, 2013; Accepted May 21, 2013

ABSTRACT

APOBEC3G (A3G) is a host-encoded protein that potently restricts the infectivity of a broad range of retroviruses. This can occur by mechanisms dependent on catalytic activity, resulting in the mutagenic deamination of nascent viral cDNA, and/or by other means that are independent of its catalytic activity. It is not yet known to what extent deamination-independent processes contribute to the overall restriction, how they exactly work or how they are regulated. Here, we show that alanine substitution of either tryptophan 94 (W94A) or 127 (W127A) in the non-catalytic N-terminal domain of A3G severely impedes RNA binding and alleviates deamination-independent restriction while still maintaining DNA mutator activity. Substitution of both tryptophans (W94A/W127A) produces a more severe phenotype in which RNA binding and RNA-dependent protein oligomerization are completely abrogated. We further demonstrate that RNA binding is specifically required for crippling late reverse transcript accumulation, preventing proviral DNA integration and, consequently, restricting viral particle release. We did not find that deaminase activity made a significant contribution to the restriction of any of these processes. In summary, this work reveals that there is a direct correlation between A3G's capacity to bind RNA and its ability to inhibit retroviral infectivity in a deamination-independent manner.

INTRODUCTION

APOBEC3G (A3G) is one of several cell-intrinsic host retroviral restriction factors in humans that potently inhibit the replication of a broad range of viruses, retroviruses and retroelements [reviewed in (1)]. It is currently believed that A3G's striking ability to deaminate cytidines into uridines in single-stranded retroviral DNA replication intermediates represents the major mechanism responsible for its antiretroviral activity. Extensive mutations, also called hypermutation, can potentially lead to the generation of premature termination codons and dysfunctional proteins resulting in non-infectious viral progeny (2–5). A3G can, however, also restrict the infectivity of retroviruses by means that do not rely on deamination, but these have yet to be clearly understood (6,7).

A3G proteins expressed in retrovirus-infected cells are packaged into the capsids of progeny virions and exert their enzymatic activity during proviral cDNA synthesis in newly infected target cells (1). Packaging of A3G into human immunodeficiency virus type I (HIV-1) virions is RNA dependent and mediated by the interaction of residues in the N-terminal domain (NTD) of A3G and the nucleocapsid region of the retroviral structural protein Gag (8,9). To counteract the deleterious effects of A3G, HIV-1 acquired the ability to prevent its packaging into virions. The viral infectivity factor (Vif) is an HIV-1 accessory protein that binds to A3G before its incorporation into virions and quickly promotes its degradation by the proteasome [reviewed in (10)]. HIV-1 particles that are released from infected cells expressing Vif are devoid of A3G and are thus fully infectious.

A3G can directly bind RNA via its non-catalytic NTD (11–13). Newly translated monomeric A3G rapidly assembles not only in the cytoplasm into RNA-independent dimeric and tetrameric structures but also into larger oligomeric assemblies that require RNA (11,14–17). In actively dividing cells such as activated T cells and cell lines, these oligomeric complexes will further aggregate into large high molecular mass (HMM) ribonucleoprotein complexes, which are estimated to be between 5 and 15 MDa in size (11,18). A3G proteins in these HMM complexes no longer exhibit enzymatic activity and

*To whom correspondence should be addressed. Tel: +1 613 562 5800 (ext. 7110); Email: langlois@uottawa.ca

cannot be packaged into HIV-1 virions (19). Thus, only low molecular mass (LMM) oligomeric A3G complexes that have not yet aggregated into HMM complexes are packaged into virions and exert cytidine deaminase activity during proviral DNA synthesis (19). It is still unclear what triggers the formation of HMM complexes in cell lines and activated lymphocytes. Understanding how these large oligomeric structures assemble is of significant importance because binding to RNA is deemed to be required for HIV-1 virion packaging. Paradoxically, RNA also appears to act as a negative regulator of A3G's catalytic activity by causing its aggregation into ribonucleic complexes (19). A3G binds various RNAs including those coding for itself, GAPDH and HIV-1, as well as several species of non-coding RNAs such as 7SL, hY1, hY3, hY4, hY5 and *Alu* (18,20–24). It is currently unknown whether binding to any of these RNAs is specifically required for A3G's antiviral activity.

The catalytic activity of A3G is currently thought to play a dominant role in the inhibition of retroviral infectivity. Notably, in addition to inflicting genetic damage, poor plus-strand transfer and defective proviral integration have also been reported to be caused by DNA editing (25–28). In parallel, several reports show that significant deamination-independent antiretroviral activity is displayed by catalytically inactive A3G enzymes (6,7,28–30). Disruptions in the zinc-binding motif of the C-terminal domain inactivate the catalytic activity of A3G. Deamination-independent mechanisms such as the inhibition of primer annealing, strand transfer, viral transcript accumulation and proviral integration have been described to collectively partake in the overall restriction of infection (28,31). An important component contributing to the deaminase-independent antiretroviral activity appears to be the inhibition of reverse transcript synthesis. This could occur by the direct interaction of A3G with the reverse transcriptase or by creating roadblocks to the processivity of the reverse transcriptase through binding to ssDNA replication intermediates (6,32). In this context, reduced retroviral cDNA synthesis would be one of the causative factors for impaired proviral integration and infection. Despite the identification of numerous antiretroviral mechanisms, it has not yet been established to what extent G-to-A hypermutation and deamination-independent mechanisms contribute to the overall inhibition of infection.

In this study, we investigated the roles of A3G in RNA-binding, HMM complex assembly and cytidine deamination at different stages of the retroviral infection cycle. We found that tryptophans 94 and 127, which are located in the non-catalytic NTD of human A3G, regulate RNA-binding and HMM complex assembly. Interestingly, both W94A and W127A mutants retain the capacity to intensely deaminate proviral DNA but no longer restrict proviral DNA synthesis, integration or viral particle release. These unique features of the mutants have allowed us to measure the direct contribution of deamination and deamination-independent restriction mechanisms on various steps of the infection cycle of three commonly studied retroviruses.

MATERIALS AND METHODS

Velocity sedimentation

The 293T cells transiently expressing FLAG-A2, FLAG-A3G or the various A3G mutants were lysed for 30 min on ice with NP40 lysis buffer [50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 0.1% NP40, 0.1% Na-deoxycholate] supplemented with cOmplete, EDTA-free, protease inhibitor cocktail (Roche). After removing insoluble material by centrifugation at 17 000*g* for 5 min at 4°C, half the samples were treated with 1 μg/ml of RNase A for 15 min at room temperature. RNase-treated and -untreated samples were then loaded on top of 5–40% sucrose step gradient (5, 10, 20, 30 and 40% in 2.3 ml of fractions) in the following buffer: 10 mM Tris–HCl (pH 7.4), 25 mM KCl and 10 mM $MgCl_2$. Gradients were spun for 6 h at 41 000 rpm (288 000*g*) at 4°C in a Beckman SW41 Ti rotor. After the spin, 1.4 ml of fractions were sequentially collected from the top of the gradient, and a 75 μl of aliquot of each fraction was mixed with 25 μl of 4× Laemmli loading buffer. Finally samples were resolved by SDS–PAGE and analyzed by immunoblotting.

Retroviral vectors

The pMOV-eGFP expression vector encodes a replicative Moloney murine leukemia virus (MoMLV) with the eGFP reporter gene inserted in the proline-rich region of the ecotropic *env* gene (33). The single-cycle HIV[p8.9] pseudovirus is generated by a multi-plasmid expression system using a packaging vector for HIV-1 (pCSGW), an expression vector for Gag-Pol-Rev (p8.9) and the pMDG plasmid encoding the envelope glycoprotein of the vesicular stomatitis virus (34,35). The eGFP reporter gene is located in the pCSGW packaging vector and is expressed from an internal spleen focus-forming virus (SFFV) long terminal repeat (LTR) promoter. The pNL4-3-deltaE-eGFP plasmid was obtained from the NIH AIDS Research Reference and Reagent Program (catalog #11100) (36). Two termination codons were introduced in the *vif* gene to prevent expression as previously described and renamed pHIVΔVif for simplicity (37).

Transfections and virus production

Viruses were produced by transfecting 3×10^5 293T cells seeded in a 6-well plate with 800 ng of pMOV-eGFP (to produce MoMLV) or 500 ng of pHIVΔVif with 200 ng of pMDG (to produce HIVΔVif) using the GeneJuice transfection reagent (Merck). HIV[p8.9] pseudoviruses were produced by co-transfecting 250 ng of p8.9, along with 300 ng of pCSGW and 150 ng of pMDG.

APOBEC virion packaging was done by co-transfecting virus expression plasmids (in the quantities described earlier in the text) along with APOBEC expression vectors: 80 ng for MoMLV and 150 ng for HIV[p8.9] and HIVΔVif. Cells were washed with phosphate buffered saline 16 h following transfection and grown in culture for an additional 48 h. Virus-containing supernatants were then collected, cleared by centrifugation at 800*g*, and filtered through 0.45 μ cartridge filters. Virus

production was quantified and normalized by enzyme-linked immunosorbent assay for p24 or p30 (QuickTiterTM Lentivirus Titer Kit and QuickTiterTM MuLV Core Antigen ELISA Kit, Cell Biolabs Inc.). Viral stocks produced in presence of APOBEC3 proteins were normalized to the p24 or p30 amounts of their respective A2 control.

Optimization of infection assays

Transfection and infection assays were carefully optimized for optimal dynamic range and single transduction events using a multiplicity of infection (MOI) of 0.5 (Supplementary Figure S1). During careful optimization of our infection assays, we found that viral infectivity is most adversely affected when virus-producing vectors are co-transfected along with control vectors expressing a protein (e.g. eGFP) rather than an 'empty vector' (e.g. pcDNA 3.1), particularly when large quantities are co-transfected (Supplementary Figure S1A and B). For this reason, we chose human A2 as a negative control to establish the maximum infection in all our assays. A2 is a member of the broader APOBEC3 family whose crystal structure enabled the early structural homology models of human A3G (38). A2 harbors a single zinc-binding pseudo-catalytic domain, does not form RNA-dependent oligomers and has no detectable deaminase or antiviral activity (39–42).

Infection assays

Target cells were infected at an MOI of 0.5 with respect to viruses produced in the presence of A2. Infections were carried out by spinoculation at 800*g* for 1 h in the presence of 8 μg/ml of polybrene (43). Infection levels in human 293T and murine NIH 3T3 cells were monitored by flow cytometry analysis by measuring eGFP fluorescence at 24 h (for MoMLV and HIV[p8.9]) or 48 h (for HIVΔVif) after infection. Infection levels were normalized to A2.

Late reverse transcript and proviral integration quantitation

Methods for quantifying late reverse transcript (LRT) accumulation and proviral integration by quantitative real-time PCR (qPCR) were adapted from the following references (44,45). DNA extractions were performed 12 h post-infection for late LRT analysis and at 24 h post-infection for integration analysis. Primer and probes sequences are given in Supplementary Table S1. For LRT analysis, the primers and probe sets for HIVΔVif or HIV[p8.9] and MoMLV are respectively as follows: qHIV-FWD, qHIV-REV and the qHIV-Probe; qMoMLV-FWD, qMoMLV-REV and the qMLV-Probe. Integrated proviruses were amplified by a first round PCR using primer pairs against eGFP and either *Alu* (293T) or *B1* (NIH 3T3) retroelements. The primers used to amplify integrated sequences were as follows: qeGFP-FWD, qAlu-1 and qAlu-2 for 293T; qeGFP-FWD, qB1-1 and qB1-2 for NIH 3T3. PCR products from the first round were then diluted 1:40 and used for qPCR using nested primers qIN-eGFP-FWD and qIN-eGFP-REV, and the qIN-eGFP-Probe. For each reaction, 0.9 pmol/ml of each primer, 0.25 pmol/ml of the probe and 10 ng of template DNA were used in a 20 ml of reaction volume. Reactions were performed in quadruplicate with TaqMan® Gene Expression master mix, Applied Biosystems (AB). Cycling conditions were 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C carried out on an AB Viii7 System. The copy numbers in each sample were normalized for DNA input using human RNase P or mouse Tfrc copy number assays (TaqMan assays 4 403 326 and 4 458 366, AB). Relative quantitation (RQ) was computed by the ViiA7 data analysis software (AB). Normalizations were performed using corresponding A2 levels for each experiment.

Statistical analysis

Restriction, integration and LRT data were expressed as mean relative values ± SD. All experiments were performed using triplicates values for restriction experiments and quadruplicates for integration and LRT analysis. Experiments were repeated at least three times from completely independent transfection assays. All statistical analyses were performed using Student's paired *t*-test using GraphPad Prism software.

RESULTS

Tryptophans 94 and 127 are involved in HMM complex assembly and RNA binding

A3G is an RNA-binding enzyme that aggregates into HMM complexes in the cytoplasm of activated CD4 + T lymphocytes and immortalized cell lines (11,18). Here, we have optimized the conditions of velocity sedimentation assays so that HMM complexes consistently accumulate in the bottom two fractions of 5–40% sucrose gradients (fractions 8 and 9), and that RNA-dependent LMM oligomeric forms of A3G consistently accumulate in fractions 4–7. Pretreatment of cell extracts with RNase dissolves HMM and LMM complexes and causes A3G to localize in fractions of the gradient that represent the predicted monomeric, dimeric and tetrameric forms of the protein (labeled as MDT in our figures). The assays were designed so that these RNA-independent forms of A3G consistently accumulate in fractions 1–3. We used endogenous β-tubulin in all our sedimentation assays as a marker for gradient quality control because it exclusively assembles into RNA-independent heterodimers that are consistently detected in fractions 1–3 only.

During the course of a screen to identify the amino acids of A3G that govern its assembly into HMM complexes, we discovered that mutation of tryptophans 94 and 127 to alanine (W94A; W127A) prevented the formation of these complexes (Figure 1A and Supplementary Figure S2). Despite the absence of HMM complexes in fractions 8 and 9, RNA-dependent LMM oligomeric complexes were present throughout the middle fractions of the sucrose gradient (fractions 4–7). Pretreatment of the extracts with RNase resulted in a complete shift toward the top of the gradient populated by the monomeric, dimeric and tetrameric forms of the A3G protein (fractions 1–3).

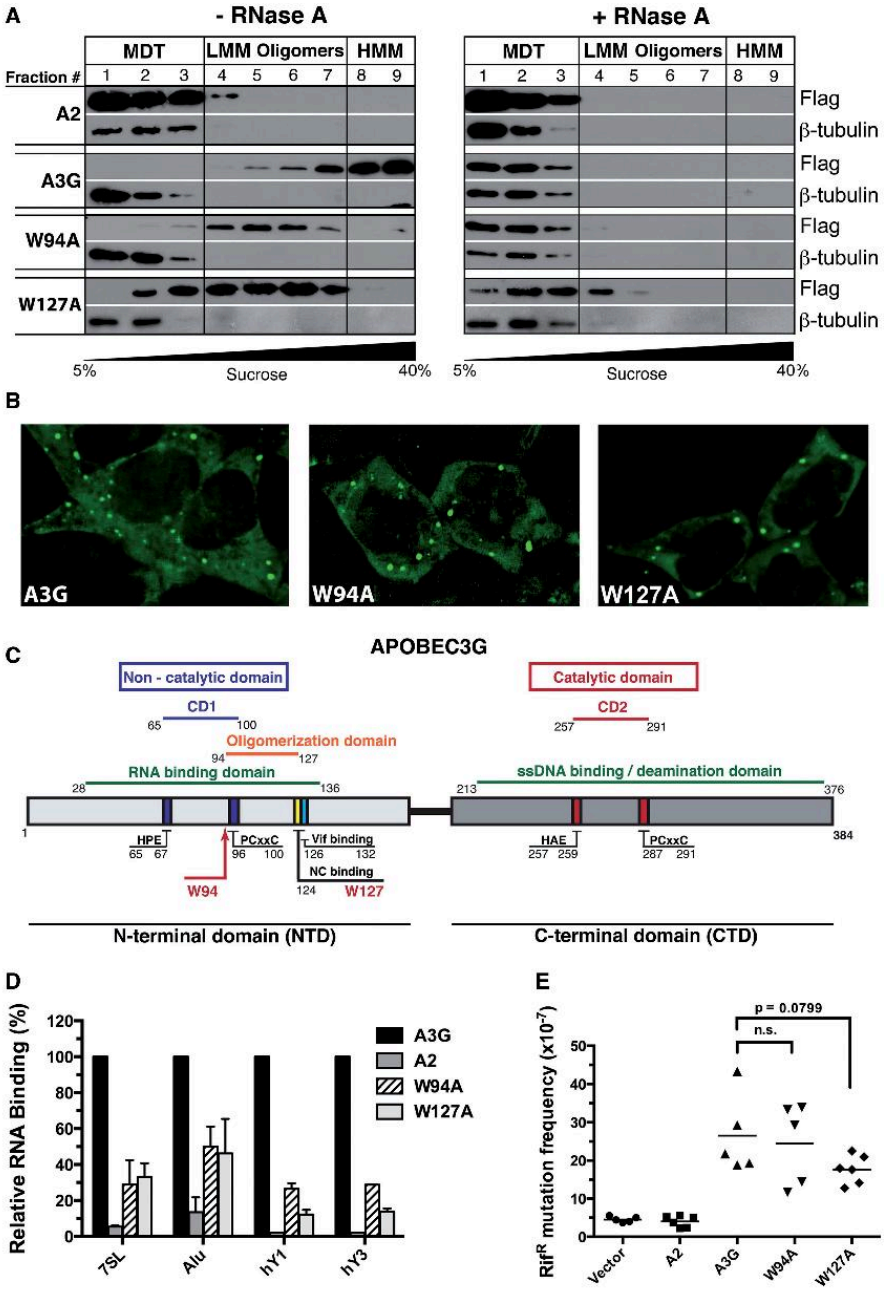


Figure 1. Effects of the W94A and W127A mutations on HMM complex assembly, subcellular distribution, RNA-binding and DNA deaminase activity. (**A**) Lysates of transfected 293T cells, treated (*right panels*) or untreated (*left panels*) with RNase, were resolved by velocity sedimentation over a non-denaturing 5–40% sucrose gradient and analyzed by western blot using anti-FLAG and anti-β-tubulin antibodies. (**B**) Fluorescent imaging of the subcellular localization of 293T cells expressing eGFP-A3G, eGFP-W94A and eGFP-W127A. (**C**) Schematic representation of the location of important residues and binding domains in the A3G protein sequence. (**D**) Binding of FLAG-tagged A2, A3G, W94A and W127A to 7SL, *Alu*, hY1 and hY3 RNAs were determined by qPCR. Relative binding to A3G is depicted. Results represent the mean ± SD of triplicate values from three independent transfection experiments. (**E**) Evaluation of the intrinsic DNA cytidine deaminase activity using a bacterial mutator assay. Each point represents the mutation frequency (Rif^R mutants per 10^7 viable cells) of an independent bacterial culture; median values are indicated.

These particular features of the W94A and W127A mutants were not observed with any of the other A3G point mutants that were tested (Supplementary Figure S2).

A3G is a cytoplasmic protein that forms numerous foci. These structures are believed to associate with RNA-processing bodies (P-bodies), which are sites of RNA storage, turnover and decapping (46). We were concerned that altering HMM complex assembly would also affect the cellular localization of the mutant proteins. We therefore transiently expressed eGFP-fusions of the mutant proteins in 293T cells and analyzed their intracellular distribution using fluorescence microscopy. We did not detect any obvious differences in size, intensity or abundance of cellular foci between wild-type A3G and the W94A and W127A mutants (Figure 1B).

Tryptophans 94 and 127 are located in the NTD of the protein in a region predicted to be involved in RNA binding, protein oligomerization, Vif interaction and cellular localization (Figure 1C) (47). W127 was first identified as a residue essential for the packaging of A3G into HIV virions (9). It is also required for binding to *Alu*, 7SL and various hY RNAs, and these RNA-binding features of A3G correlate with its ability to inhibit *Alu* retrotransposition (20,24). Direct *in vitro* binding assays performed using purified protein also confirmed the reduced affinity of the W127A mutant for RNA (48). Other studies revealed that this residue was critical for cytoplasmic localization and N-terminal oligomerization (17,47). W94 was also reported to influence A3G packaging into HIV virions, but to a lesser extent than W127 (24). There are however discordant reports as to whether W94 can bind 7SL RNA (13,22).

W94A and W127A have reduced RNA-binding capacity, but DNA editing is mostly unaffected

Here, we independently investigated the binding of the A3G mutants to a selection of RNAs: *Alu*, 7SL, hY1, hY3 and β-actin (Figure 1D). We measured the relative capacity of the mutants to bind RNA compared with wild-type A3G by carrying out qPCR analysis on RNA isolated from immunoprecipitates of the A3G variants transiently expressed in 293T cells. We found that in agreement with earlier studies (20,24), the W94A and W127A mutants associated 50–90% less efficiently with *Alu*, 7SL, hY1 and hY3 RNAs compared with A3G (Figure 1D). A2 non-specifically bound RNA to similar levels as the bead-only control and was thus used as a negative binding control in all our subsequent assays (Supplementary Figure S3A). β-actin mRNA did not significantly bind to any of the APOBEC proteins, which is in line with previous studies (18,21), and was excluded from the graphs to improve clarity.

Before further characterizing these mutants, we wanted to ascertain whether they retained enzymatic activity on DNA by using a bacterial mutator assay commonly used to measure the catalytic activity of cytidine deaminases (49). The results of this assay revealed that the enzymatic activities of the mutants were similar to the wild-type A3G protein, whereby both these proteins were capable of mutating *Escherichia coli* genomic DNA and giving rise to a relatively large number of rifampicin-resistant colonies (Rif^R) (Figure 1E). In summary, our results show that the W94A and W127A mutants both have severely diminished RNA-binding properties compared with wild-type A3G, but this had no significant impact on the catalytic activity of the proteins.

RNA-binding mutants are packaged with different efficiencies into HIV-1 ΔVif and MoMLV virions

Here, we compared the virion packaging efficiency of the wild-type A3G protein to that of W94A, W127A, an inactive catalytic mutant E259Q, and corresponding W94A or W127A compound mutants: W94A/E259Q and W127A/E259Q. Three retroviruses were tested: HIVΔVif (NL4-3-derived with the env gene substituted for eGFP), HIV[p.8.9] (a self-inactivating HIV-derived pseudovirus expressing eGFP from an internal SFFV promoter) and replicative ecotropic MoMLV expressing an Env-eGFP fusion protein (see 'Materials and Methods' section for details). As previously described by others, we found that W94A and W127A were poorly packaged into HIVΔVif particles (Figure 2A). Surprisingly, all A3G variants were packaged efficiently into HIV[p8.9] and MoMLV virions (Figure 2B and C). These results indicate that the factors that govern virion encapsidation are different for HIV-1ΔVif and MoMLV. Our reasoning as to why the mutants proteins are packaged efficiently into HIV[p8.9] virions is presented in the discussion.

RNA binding is required for retroviral restriction

Infection assays show that both W94A and W127A mutants displayed little or no antiretroviral activity on HIVΔVif as would be expected because of the packaging defect, whereas the catalytic mutant, E259Q, reduced the relative number of eGFP-positive target cells by ~40–50% for all viruses tested (Figure 2D, E and F). Although the W94A and W127A mutants were ineffective in restricting the infection of HIV[p8.9] (Figure 2E), they reduced the infectivity of MoMLV by 55 and 40%, respectively (Figure 2F). Double mutants for both RNA-binding and catalytic activity, W94A/E259Q and W127A/E259Q, were completely ineffective in restricting the infection of all the viruses tested.

We next asked whether W94A and W127A could mutate HIV[p8.9] and MoMLV, despite having defective RNA-binding properties. As predicted by the bacterial mutator assay, both W94A and W127A mutants introduced high levels of hypermutation in both retroviruses tested, with the vast majority (82–92%) of all sequences analyzed being mutated (Table 1). Also, we found no evidence of DNA editing by mutant proteins containing the E259Q substitution. Analysis of the DNA context specificity for the deamination revealed a strong preference for the targeting of 5′-CCC-′3 trinucleotides for all A3G variants, indicating that reduced RNA binding did not impact DNA-targeting specificity (Table 2).

Because wild-type and mutant A3G proteins appear to be packaged with the same efficiency in MoMLV and HIV[p8.9] virions, differences in mutation rates could be explained by reduced deamination efficiency. To evaluate

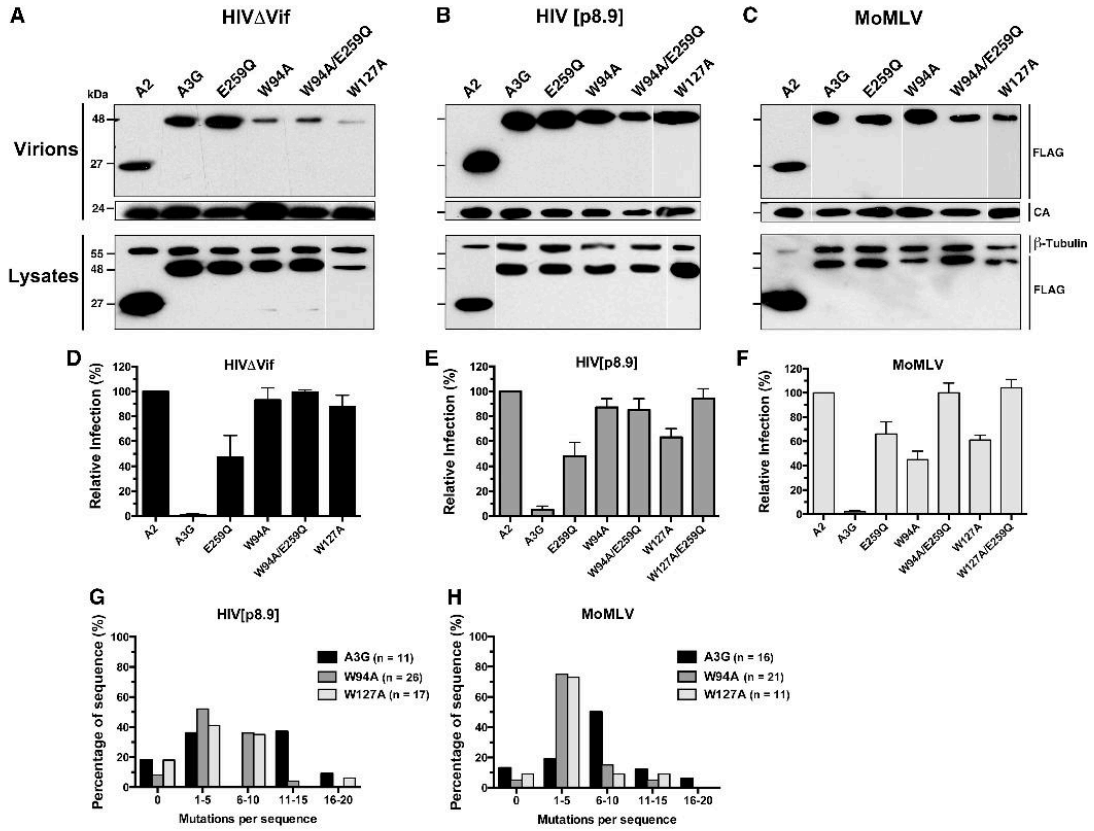


Figure 2. Virus-dependent packaging of W94A and W127A, and retroviral restriction analysis. (**A–C**) Viral particles were purified from cell supernatants, lysed and assayed for the presence of FLAG-tagged APOBEC proteins by western blot analysis (*top panels*). Lysates of virus-producing cells are shown in the *bottom panels*. (**D–F**) Restriction analysis of virus infectivity as measured by eGFP expression in target cells infected with HIVΔVif (D), HIV[p8.9] (E) or MoMLV (F). Results represent the mean ± SD of triplicate values from at least three independent transfection experiments. (**G** and **H**) Analysis of mutations in proviral DNA. Histograms depict the proportion of total sequences containing the indicated number of mutations. The total number of clones sequenced is indicated in the chart legend. Sequences were compiled from one experiment for A3G and W127A and from two independent experiments for W94A.

Table 1. Editing of HIV[p8.9] and MoMLV by A3G and variants

Virus	APOBEC3 protein	Base pairs sequenced*	Sequences mutated (%)	Total number of G-to-A mutations	G-to-A mutation frequency (mutations/kb)
HIV[p8.9]					
	A3G	7887	82	82	12.7
	E259Q	4302	0	0	
	W94A	18 642	92	132	7.1
	W94A/E259Q	4302	0	0	
	W127A	12 189	82	88	7.2
MoMLV					
	A3G	11 472	88	111	9.7
	E259Q	4302	0	0	
	W94A	15 057	95	85	5.6
	W94A/E259Q	4302	0	0	
	W127A	7887	91	38	4.8

*A 717-bp segment of integrated viral DNA was amplified by PCR and cloned. Independent clones were sequenced, and mutations computed on the plus-strand DNA were analyzed.

Table 2. Local sequence preference for deamination on the minus-strand retroviral DNA*

APOBEC3	HIV[p8.9]		MoMLV	
	−2	−1	−2	−1
A3G				
A	20	2	20	0
C	48	77	39	68
G	9	1	7	3
T	23	20	37	29
W94A				
A	12	1	8	0
C	60	90	68	89
G	4	1	4	1
T	24	8	20	10
W127A				
A	33	3	14	2
C	44	82	53	83
G	5	3	3	0
T	18	13	29	16

*Values were computed with respect to the deaminated cytidine (position zero) on the viral minus strand. Values represent the percentage of occurrence of each base at positions −1 and −2 respective to the deaminated cytidine. Sequences compiled in Table 1 were used for the analysis.

this, we calculated the mutation frequency in each individual sequence examined (Figure 2G and H). Our analysis shows that W94A and W127A introduced on average between 1 and 10 mutations per sequence for HIV[p8.9] and 1–5 for MoMLV. Wild-type A3G introduced slightly more mutations per sequence on both viruses, which explains the results of Table 1.

RNA binding is required for the inhibition of proviral DNA accumulation and integration

Retroviruses produced in the presence of A3G display reduced levels of LRT and proviral integration (50). Here, we sought to examine how the W94A and W127A mutations impact LRT accumulation and proviral integration. Results show that neither W94A nor W127A significantly hinder LRT accumulation, whereas wild-type A3G and E259Q reduced these levels by ~40–60% for both viruses (Figure 3A and B). A3G and E259Q had much more dramatic effects on integration with measured reductions of 94 and 89% for HIV[p8.9] and 92 and 81% for MoMLV, respectively (Figure 3C and D). These results clearly reveal the marginal role of deamination in preventing these two early steps of the infection. On the other hand, W94A had no significant effect on reducing the proviral integration of either MoMLV or HIV[p8.9]. Equally, W127A did not reduce the integration of HIV[p8.9], but appeared to have a slight effect on MoMLV. Inactivation of the deaminase activity of the W94A RNA-binding mutant had no detectable impact on LRT accumulation or integration, which again supports that deamination is not a major contributor in preventing these specific processes.

Hypermutation does not affect MoMLV particle release

We were curious to determine whether viral particle release was affected by the DNA mutator activity of the RNA-binding mutant W94A. Because W94A is not adequately packaged into HIVΔVif particles, we performed this experiment on MoMLV. Deamination-induced damage that can affect particle release includes: mutational damage to the retroviral promoter, loss of protein function or localization, or the generation of stop codons that halt protein synthesis. A3G variants and MoMLV expression plasmids were co-transfected at increasing A3G-to-virus ratios into 293T cells, and NIH 3T3 target cells were infected with MoMLV particles at an MOI of 0.5. Virus-containing supernatants were then collected 72 h later, and p30 levels were measured by enzyme-linked immunosorbent assay (Figure 4A).

Despite W94A reducing the apparent infection by ~60% (Figure 4B), the amount of p30 particles released was nearly identical compared with the enzymatically inactive W94A/E259Q control (Figure 4C). On the other hand, A3G had a dramatic effect on MoMLV infection at all co-transfection ratios tested. However, reduction of particle release was only observed at a 1:40 ratio (20 ng) and above (Figure 4C). Overall, these experiments indicate that mutations inflicted by W94A had no detectable impact on MoMLV particle release.

$Vpr_{14–88}$ polypeptide fusions rescue RNA-binding and deamination-independent restriction

Here, we returned our attention to HIVΔVif restriction by the A3G RNA-binding mutants. It has been shown that fusing a Vpr polypeptide to proteins of interest can enable their packaging into HIV virions (13,51). Improving virion packaging of the A3G mutant proteins would allow us to determine whether RNA binding is also required for HIVΔVif restriction. To investigate this issue, we generated fusion proteins with the $Vpr_{14–88}$ polypeptide (referred to as Vpr for simplicity) with all our A3G variants and performed virion packaging and restriction assays. As expected, we observed vastly improved packaging of both Vpr-W94A and Vpr-W127A into HIVΔVif, and a recovery of the antiretroviral activities of both mutants (Figure 5A and B). Surprisingly, both Vpr-W94A and Vpr-W127A now restricted HIV[p8.9] and MoMLV to levels comparable with Vpr-A3G (Figure 5C and D). This was unexpected because we had not detected a packaging defect with the mutants on HIV[p8.9] and MoMLV (Figure 2B and C).

Vpr is an HIV-1 accessory protein known to directly bind RNA (52). Proteins fused to Vpr would hence be expected to display overall increased RNA-binding properties. To determine whether RNA binding is restored with the mutants, we measured the binding of Vpr fusion proteins to *Alu*, 7SL, hY1, hY3 and β-actin RNAs using a similar approach as in Figure 1D. We found that binding to RNA was vastly improved in all cases except for β-actin that remained at background levels and again was not plotted on the graph. Strikingly, Vpr-A2 also displayed RNA-binding properties similar to Vpr-A3G for *Alu* and hY3 and greatly improved binding to 7SL and hY1 (Figure 5E). Comparative binding of the RNAs with Vpr-A3G, Vpr-A2 and agarose beads is shown in Supplementary Figure S3B. To verify whether

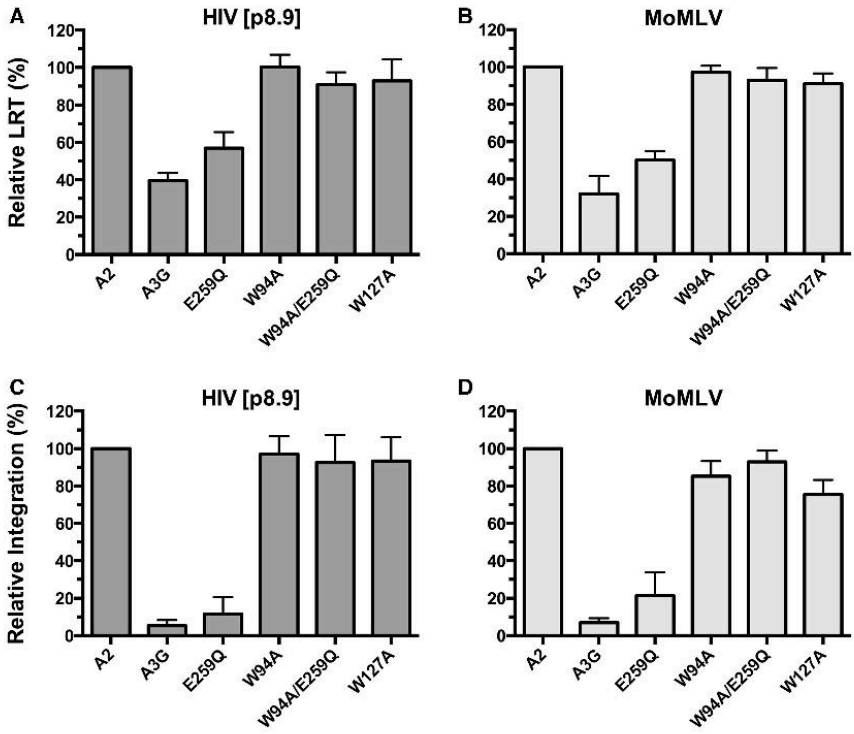


Figure 3. Effects of W94A and W127A on proviral DNA synthesis and integration. (**A** and **B**) Analysis of LRT accumulation. (**C** and **D**) Analysis of proviral DNA integration. DNA from infected 293T cells was collected at 12 h post-infection for LRT analysis and at 24 h for integration analysis. Infections were performed on 293T cells with HIV[p8.9] (A and C), or on NIH 3T3 cells with MoMLV (B and D). The results reflect the mean RQ ± SD of three independent experiments each performed in quadruplicate. Data were normalized to A2 values.

increased RNA binding also impacted the intracellular oligomeric forms of the mutant APOBEC3 proteins, velocity sedimentation assays were carried out on the Vpr fusion proteins. Both Vpr-W94A and Vpr-W127A had their ability to assemble into HMM complexes restored (Figure 5F). Finally, we tested whether the Vpr fusion proteins restricted proviral integration of HIVΔVif (Figure 5G). Integration was compromised to a similar extent by all Vpr-A3G variants, but not by Vpr-A2 that was now also capable of binding RNA.

To ensure that the observed phenotype of the Vpr fusion proteins was conferred specifically by the RNA-binding properties of Vpr, we deleted amino acids 87 and 88 of the $Vpr_{14–88}$ polypeptide that have been previously been shown to mediate RNA binding (52) and repeated experiments depicted in Figure 5. We found that W94A and W127A fusions with the $Vpr_{14–86}$ polypeptide defective in RNA binding, Vpr(ΔRNA), were efficiently packaged into HIVΔVif virions (Figure 6A), did not restrict the infection (Figure 6B), were unable to inhibit proviral integration (Figure 6C) and displayed RNA-binding defects (Figure 6D). In summary, these data show that RNA binding is an essential property for A3G to be able to restrict Vif-deficient HIV-1 infection.

Residues W94 and W127 cooperate to bind RNA

To gain further insight into how W94 and W127 enable A3G to interact with RNA, we conducted homology modeling of the A3G head-to-head NTD dimer (Supplementary Figure S4). In our model, the two A3G NTD monomers make extensive contacts, including the loops connecting the α1-β1 and β4-α4 with the corresponding β4-α4 and α1-β1 loops of the reciprocal protomer. Interestingly, close inspection of the NTD dimer shows that W94 of the first monomer is in close proximity to W127 in the other monomer (Figure 7A and Supplementary Figure S4A), a result also observed by Lavens *et al.* (53). The structure shows that on dimerization, there is a significant increase in the size of the positively charged patch that extends to the C-terminal end of α6 of the reciprocal dimer's subunit (Supplementary Figure S4B). Overall, our modeling study suggests that A3G dimerization generates a large surface for RNA binding, and that W94A and W127A substitutions would strongly disfavor the binding of RNA. An A3G mutant carrying a double W94A/W127A substitution should therefore potentiate the RNA-binding defect. To validate this prediction, we generated the double mutant and analyzed its RNA-binding properties.

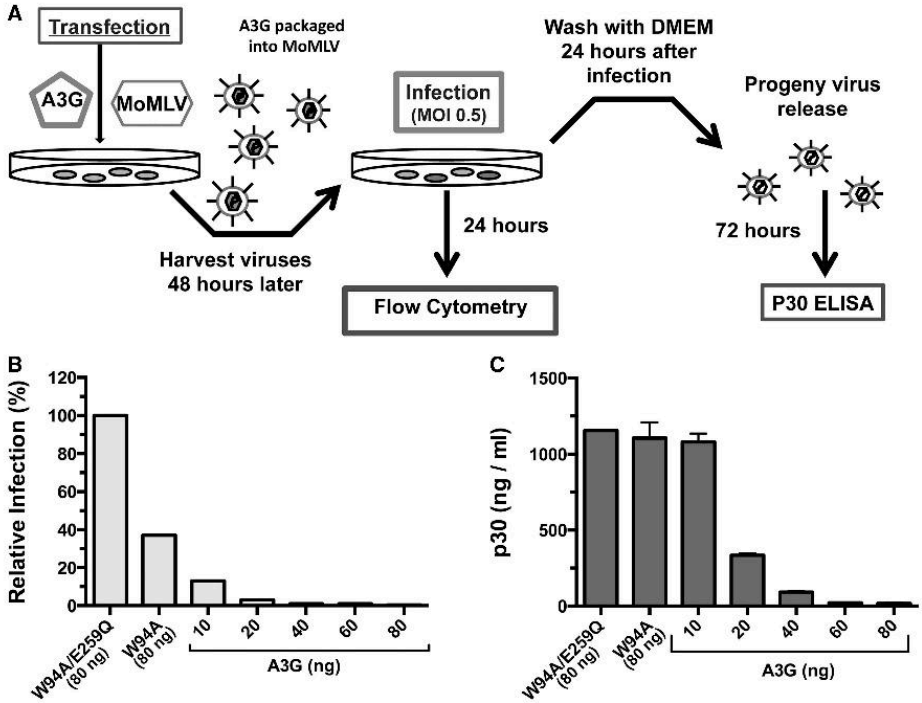


Figure 4. Effect of deamination on MoMLV particle release. (**A**) Diagram of the experimental method. (**B**) Relative infection of NIH 3T3 cells by MoMLV measured after 24 h. Viral particles were produced in presence of a 1:10 A3G (80 ng)-to-MoMLV (800 ng) proviral plasmid ratio for W94A and W94A/E259Q, and increasing amounts of wild-type A3G as indicated. (**C**) p30 levels in cell supernatants were measured by enzyme-linked immunosorbent assay 72 h after infection. Data represent the mean ± SD of three independent protein measurements.

We found that the RNA-dependent oligomerization of W94A/W127A was completely abolished (Figure 7B). Additionally, the double mutant did not significantly bind to any of the RNAs tested (Figure 7C).

Co-expression of W94A with E259Q does not restore the restriction defect

The inability of the W94A and W127A mutants to prevent viral cDNA accumulation and integration could potentially be explained by the absence of a cofactor that normally binds to these tryptophan residues on wild-type A3G. Here, we sought to establish whether we could restore restriction to its full potency by producing viruses in the presence of equal quantities of W94A and E259Q. Only the W94A mutant was used in these assays because it has the ability to self-associate as opposed to W127A that does not (Supplementary Figure S5A and B). E259Q is efficiently packaged into HIV[p8.9] and MoMLV virions and can assemble into HMM complexes (Supplementary Figure S2). Our complementation assays on HIV[p8.9] and MoMLV indicate that E259Q and W94A do not complement each other's function, which would have resulted in an increase of the overall restriction (Figure 7D and E). These results weigh against the possibility that a virion-packaged *trans*-acting cofactor is required for enabling A3G to restrict retroviral infection.

DISCUSSION

We initially set out to identify the residues in A3G that are responsible for HMM complex assembly to gain further insight into the protein's regulation. Careful optimization of velocity sedimentation assays facilitated consistent and well-defined separation between HMM complexes, oligomeric LMM complexes and the RNA-independent forms of the protein (Figure 1A). We used this method to screen A3G point mutants and identified W94 and W127 as critical residues for HMM assembly. These two amino acids have been the object of previous studies that have focused on RNA binding, protein oligomerization and packaging of A3G into HIV-1 virions. In agreement with previous work, we found that both mutants associated much less efficiently with various RNAs (Figure 1D) (17,22,24,54,55).

Although W94A or W127A substitutions are known to have detrimental effects on HIV-1 virion packaging, we were surprised to see that the packaging of these mutants into MoMLV and HIV[p8.9] was largely unaffected (Figure 2B and C). We do not fully understand why

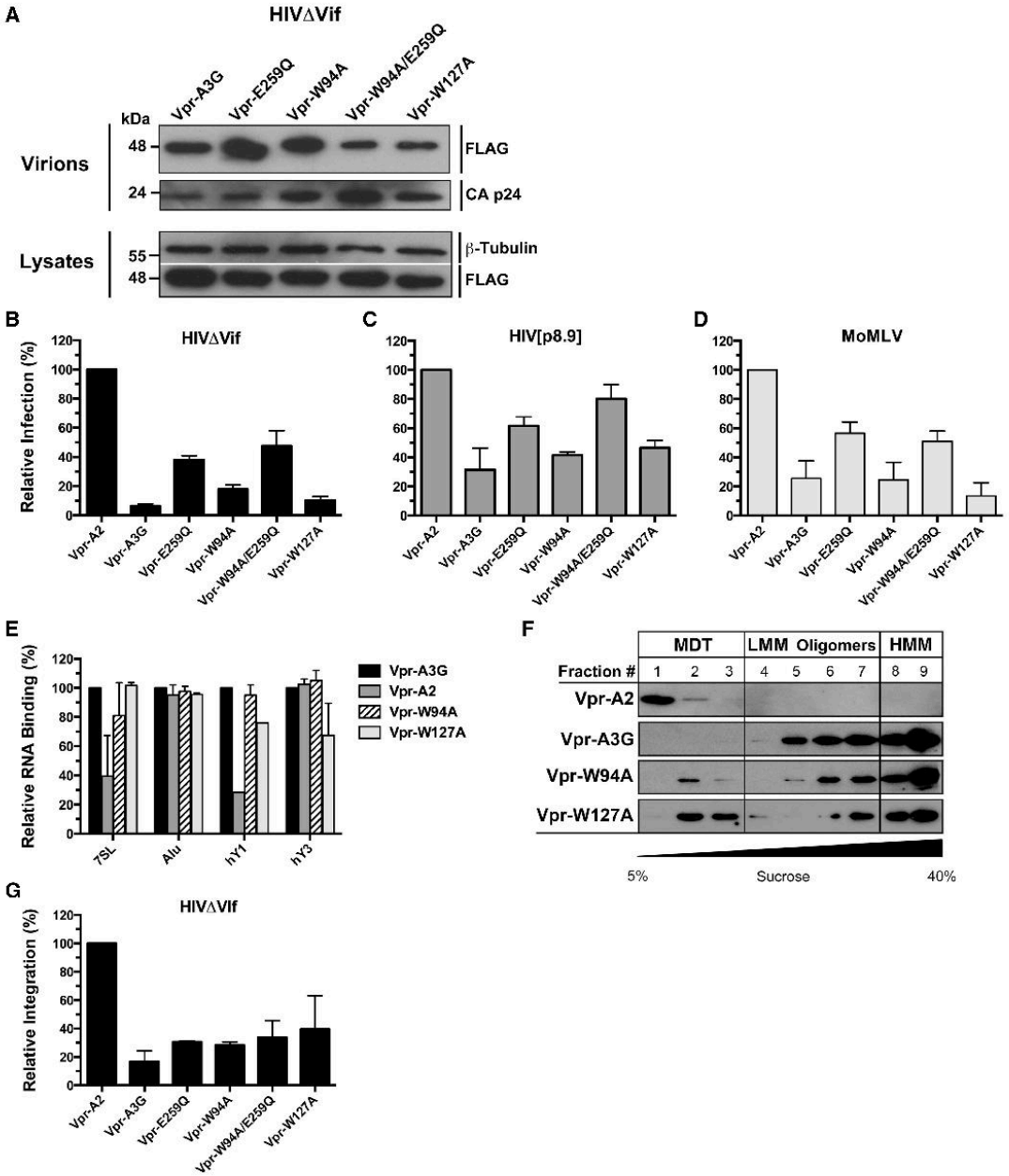


Figure 5. Vpr_{14-88} polypeptide (Vpr) fusions restore virion packaging, RNA-binding, HMM complex assembly and antiretroviral properties of W94A and W127A. (**A**) Packaging of all Vpr-A3G fusion proteins into HIVΔVif virions. (**B–D**) Antiretroviral activities of Vpr-A3G fusion proteins on HIVΔVif (B), HIV[p8.9] (C) and MoMLV (D). (**E**) Evaluation of the RNA-binding properties of Vpr fusion proteins. Data represent the mean ± SD of triplicate values from three independent experiments. (**F**) Lysates of 293T cells transfected with Vpr expression vectors analyzed by velocity sedimentation over non-denaturing sucrose gradients. (**G**) Effect of Vpr fusion proteins on HIVΔVif proviral integration. The results reflect the mean RQ ± SD of three independent experiments each performed in quadruplicate. Data were normalized to Vpr-A2 values.

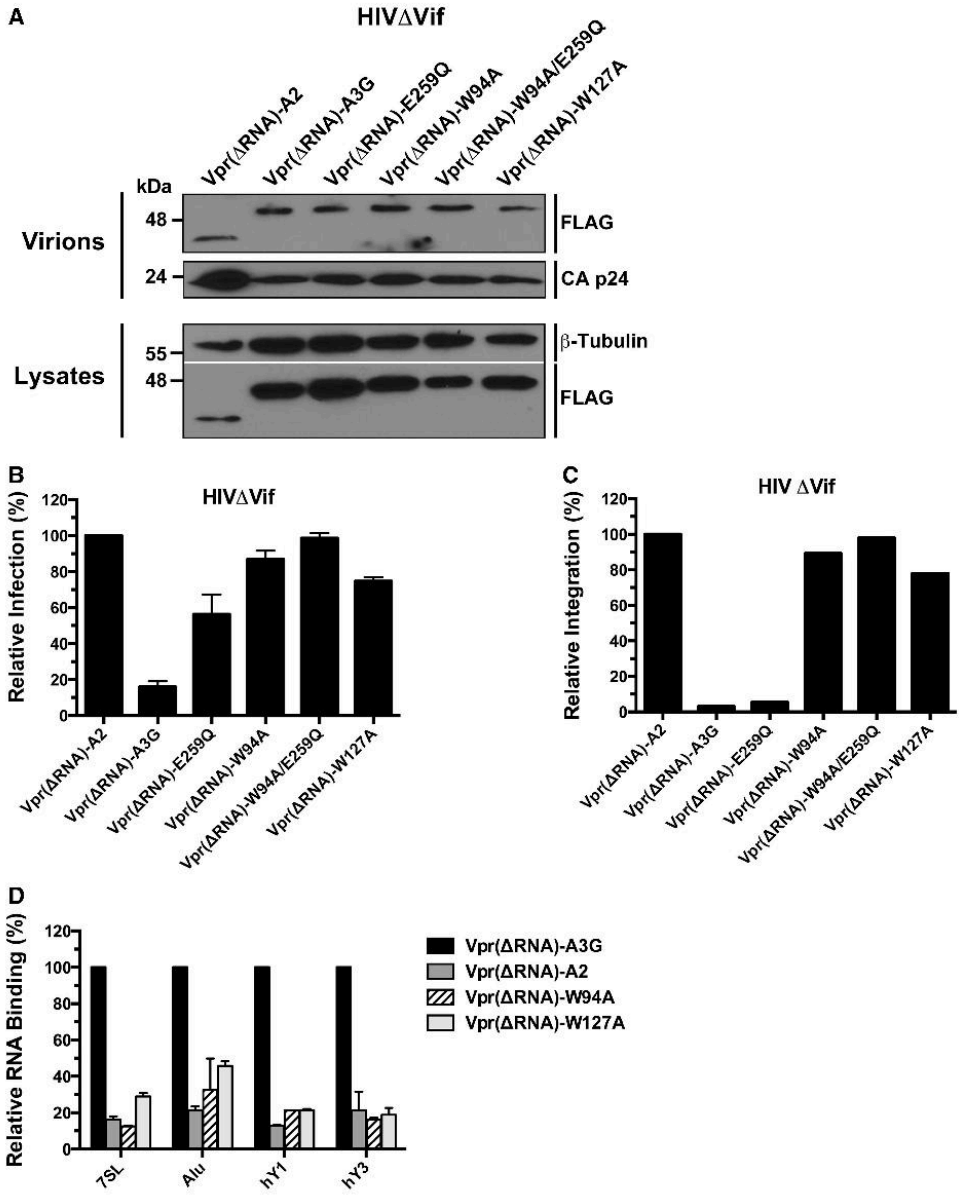


Figure 6. Fusion to the RNA-binding defective Vpr_{14-86} polyprotein [Vpr(ΔRNA)] does not restore the restriction potential of the W94A and W127A mutants. (**A**) Analysis of the packaging of all Vpr(ΔRNA)-A3G fusion proteins into HIVΔVif virions. (**B**) Antiretroviral activities of Vpr(ΔRNA)-A3G fusion proteins on HIVΔVif. (**C**) Evaluation of the RNA-binding properties of VprΔRNA) fusion proteins. Data represent the mean ± SD of triplicate values from three independent experiments. (**D**) Effect of Vpr(ΔRNA) fusion proteins on HIVΔVif proviral integration. The results reflect the mean RQ of one experiment performed in quadruplicate. Data were normalized to Vpr-A2 values.

HIV[p8.9] did not also suffer from the same packaging defects as HIVΔVif because both viruses express identical Gag NC sequences. HIV[p8.9] does however contain numerous non-HIV elements in its genome that we believe were responsible for rescuing the packaging defect. Further studies are required to identify the determinants that govern the packaging of A3G into MoMLV and HIV[p8.9].

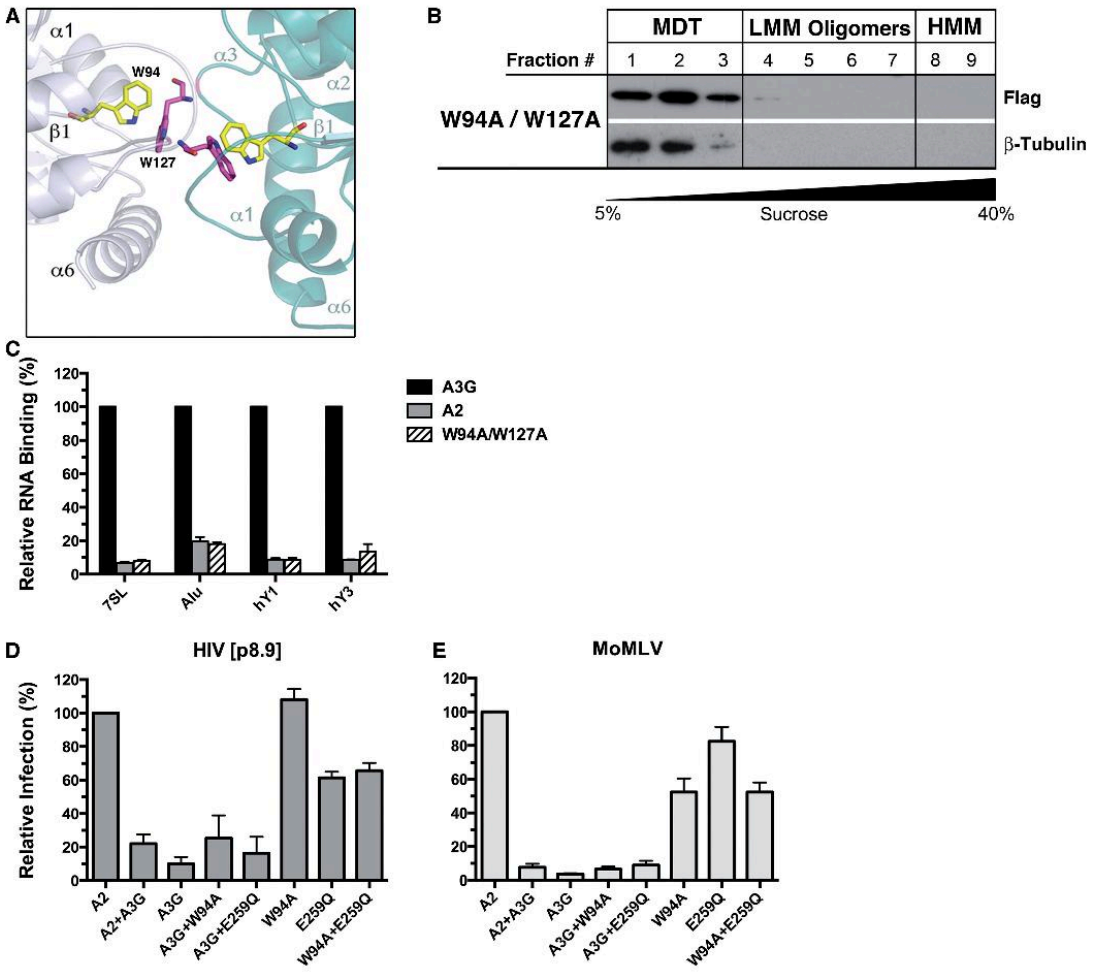


Figure 7. Retroviral restriction by A3G requires RNA but not a protein co-factor. (**A**) Cropped field from the homology model of the head-to-head NTD dimer of A3G (see Supplementary Figure S4 for the complete model and details). The W94 and W127 interaction domain at the interface of the monomers is shown, W94 (yellow); W127 (pink). (**B**) Lysates of 293T cells transfected with the W94A/W127A mutant resolved by non-denaturing velocity sedimentation. (**C**) Evaluation of the RNA-binding properties of the W94A/W127A mutant. (**D** and **E**) Viruses for the complementation assays were produced by co-transfecting a total of 80 ng of APOBEC expression plasmids (40 ng each) for MoMLV restrictions assays, or 150 ng for HIV[p8.9] (75 ng each). Data represent the mean ± SD of triplicate values from three independent experiments.

Perhaps the most controversial part of this study was the finding that A3G's deaminase activity had little or no detectable impact on LRT accumulation and proviral integration for all three viruses tested. Even MoMLV progeny virus release was unaffected by the inactivation of the catalytic activity of the W94A mutant. These results do not however mean that A3G-mutated viruses are infectious; they are most likely highly compromised in their replicative fitness. But yet again, it is well documented that the infected cells of HIV-1 patients contain multiple copies of integrated virus that can potentially recombine and complement each other's function (56–61). In a similar manner, sublethally mutated proteins from one provirus could complement the function of lethally mutated proteins from another (61–63). This therefore raises questions about long-term protection that is actually conferred by hypermutation during the course of a natural retroviral infection.

Although the deaminase activity of the W94A and W127A mutants did not impair the early stages of HIVΔVif or HIV[p8.9] infection, it reduced the infection levels of MoMLV by ~50–60% (Figures 2E and F and 6B). Infection levels measured in our single-round assays reflect the number of target cells expressing a reporter protein (eGFP) driven by the promoter of the integrated provirus. Reporter gene expression is only possible if the

provirus has successfully integrated into the target genome. Apparent antiretroviral activity in these systems is therefore a reflection that processes such as eGFP mRNA expression, splicing, translation and protein fluorescence have been affected by the mutations. In the case of HIVΔVif and HIV[p8.9], eGFP is expressed from a monocistronic mRNA driven by either the LTR promoter (HIVΔVif) or an internal SFFV promoter (HIV[p8.9]). In MoMLV, however, eGFP is expressed as a fusion protein with Env. Upstream of the eGFP coding sequence are 296 amino acids of the N-terminus of Env. This sequence contains 72 putative deamination target sites that can potentially yield 13 termination codons (Supplementary Table S2). The eGFP coding sequence within all three viruses is identical and contains only a single site that can generate a termination codon. We therefore believe that the reduced apparent infection of MoMLV by W94A and W127A could be caused in part by the generation of premature termination codons in the N-terminal Env segment thereby preventing the expression of the eGFP reporter protein. Another possibility that may contribute to explain our observations is that a portion of deaminated proviral cDNA is degraded before integration via a uracil DNA glycosylase base excision pathway (64).

It has been debated whether the E259Q substitution, which eliminates the proton donor in the catalytic site required for the deamination process, could affect intrinsic properties of the A3G protein other than catalytic activity alone, such as DNA binding for instance. To address this controversial issue, we compared the effects of the E259Q mutant with that of the C-terminal domain DNA-binding mutant R313A (65). We found no differences between the two mutants in their capacities to assemble into HMM complexes, restrict HIVΔVif infection or inhibit proviral integration (Supplementary Figure S6). Equally, we did not find any hypermutated proviral sequences when HIVΔVif was produced with either mutant (data not shown).

Another important question that emerged from this study was whether the W94 and W127 residues of A3G recruit a virion-packaged cofactor required for deamination-independent viral restriction. To answer this, we co-expressed the W94A and E259Q mutants and performed retroviral restriction assays. Our hypothesis was that if a co-factor was involved, association of W94A and E259Q mutants would improve overall restriction levels. Our results showed however that restriction was not restored, therefore weighing against the existence of such a co-factor (Figure 7D and E). Nonetheless, although RNA binding is essential for deamination-independent restriction, it is not alone sufficient to provide maximum restriction potential. Specific RNA species that bind to A3G may be required as supported by the absence of detectable restriction of infection with the RNA-binding Vpr-A2 fusion protein (Figure 5). Clues to the identity of these RNAs could be obtained from differential analyses of the RNA content of HMM and LMM complexes. Additionally, the RNA-binding affinity of A3G and the manner by which both its protein domains interact with RNA may also be of capital importance to prevent retroviral cDNA synthesis and integration.

In summary, the current work illustrates the essential and direct role of RNA in the deamination-independent restriction of retroviruses by A3G. Proviral DNA synthesis and integration are potently inhibited by processes that do not require the cytidine deaminase activity of the protein. Deamination-independent restriction mechanisms therefore appear to be important contributors in preventing irreversible and potentially harmful proviral integration into the host's genomic DNA. Although abundant A3G-induced G-to-A mutations had only a minor impact on restricting the early stages of the infection, they most likely play a major role in limiting the infectivity, fitness and spread of progeny retroviruses in physiological conditions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–6, Supplementary Materials and Methods and Supplementary References [66,67].

ACKNOWLEDGEMENTS

The authors thank Tara Read, Lionel Filion, Raed Hanania and Kristin Kemmerich for technical support. They are especially grateful to Reuben Harris, Linda Chelico and Ilona Skerjanc for helpful discussions and comments on the manuscript, and the NIH AIDS Research and Reference Reagent Program for reagents. K.B. holds an Ontario Graduate Scholarship. J.-F.C. holds a Canada Research Chair in Structural Biology and Epigenetics. M.-A.L. holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity.

FUNDING

Canadian Institutes of Health Research [#89774]; Early Researcher Award from the Ontario Ministry of Research and Innovation (to M.-A.L.). Funding for open access charge: Canadian Institutes of Health Research [#89774 to M.-A.L.].

Conflict of interest statement. None declared.

REFERENCES

1. Albin,J.S. and Harris,R.S. (2010) Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert Rev. Mol. Med.*, **12**, e4.
2. Jern,P., Russell,R.A., Pathak,V.K. and Coffin,J.M. (2009) Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. *PLoS Pathog.*, **5**, e1000367.
3. Suspene,R., Rusniok,C., Vartanian,J.P. and Wain-Hobson,S. (2006) Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucleic Acids Res.*, **34**, 4677–4684.
4. Yu,Q., Konig,R., Pillai,S., Chiles,K., Kearney,M., Palmer,S., Richman,D., Coffin,J.M. and Landau,N.R. (2004) Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.*, **11**, 435–442.

5. Armitage,A.E., Deforche,K., Chang,C.H., Wee,E., Kramer,B., Welch,J.J., Gerstoft,J., Fugger,L., McMichael,A., Rambaut,A. *et al.* (2012) APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. *PLoS Genet.*, **8**, e1002550.
6. Bishop,K.N., Verma,M., Kim,E.Y., Wolinsky,S.M. and Malim,M.H. (2008) APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS Pathog.*, **4**, e1000231.
7. Newman,E.N., Holmes,R.K., Craig,H.M., Klein,K.C., Lingappa,J.R., Malim,M.H. and Sheehy,A.M. (2005) Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.*, **15**, 166–170.
8. Burnett,A. and Spearman,P. (2007) APOBEC3G multimers are recruited to the plasma membrane for packaging into human immunodeficiency virus type 1 virus-like particles in an RNA-dependent process requiring the NC basic linker. *J. Virol.*, **81**, 5000–5013.
9. Huthoff,H. and Malim,M.H. (2007) Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. *J. Virol.*, **81**, 3807–3815.
10. Wissing,S., Galloway,N.L. and Greene,W.C. (2010) HIV-1 Vif versus the APOBEC3 cytidine deaminases: an intracellular duel between pathogen and host restriction factors. *Mol. Aspects Med.*, **31**, 383–397.
11. Chelico,L., Pham,P., Calabrese,P. and Goodman,M.F. (2006) APOBEC3G DNA deaminase acts processively 3′ –> 5′ on single-stranded DNA. *Nat. Struct. Mol. Biol.*, **13**, 392–399.
12. Iwatani,Y., Takeuchi,H., Strebel,K. and Levin,J.G. (2006) Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J. Virol.*, **80**, 5992–6002.
13. Bulliard,Y., Turelli,P., Rohrig,U.F., Zoete,V., Mangeat,B., Michielin,O. and Trono,D. (2009) Functional analysis and structural modeling of human APOBEC3G reveal the role of evolutionarily conserved elements in the inhibition of human immunodeficiency virus type 1 infection and Alu transposition. *J. Virol.*, **83**, 12611–12621.
14. Wedekind,J.E., Gillilan,R., Janda,A., Krucinska,J., Salter,J.D., Bennett,R.P., Raina,J. and Smith,H.C. (2006) Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. *J. Biol. Chem.*, **281**, 38122–38126.
15. Bennett,R.P., Salter,J.D., Liu,X., Wedekind,J.E. and Smith,H.C. (2008) APOBEC3G subunits self-associate via the C-terminal deaminase domain. *J. Biol. Chem.*, **283**, 33329–33336.
16. Salter,J.D., Krucinska,J., Raina,J., Smith,H.C. and Wedekind,J.E. (2009) A hydrodynamic analysis of APOBEC3G reveals a monomer-dimer-tetramer self-association that has implications for anti-HIV function. *Biochemistry*, **48**, 10685–10687.
17. Huthoff,H., Autore,F., Gallois-Montbrun,S., Fraternali,F. and Malim,M.H. (2009) RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog.*, **5**, e1000330.
18. Chiu,Y.L., Witkowska,H.E., Hall,S.C., Santiago,M., Soros,V.B., Esnault,C., Heidmann,T. and Greene,W.C. (2006) High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc. Natl Acad. Sci. USA*, **103**, 15588–15593.
19. Soros,V.B., Yonemoto,W. and Greene,W.C. (2007) Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. *PLoS Pathog.*, **3**, e15.
20. Khan,M.A., Goila-Gaur,R., Opi,S., Miyagi,E., Takeuchi,H., Kao,S. and Strebel,K. (2007) Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. *Retrovirology*, **4**, 48.
21. Kozak,S.L., Marin,M., Rose,K.M., Bystrom,C. and Kabat,D. (2006) The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. *J. Biol. Chem.*, **281**, 29105–29119.
22. Zhang,W., Du,J., Yu,K., Wang,T., Yong,X. and Yu,X.F. (2010) Association of potent human antiviral cytidine deaminases with 7SL RNA and viral RNP in HIV-1 virions. *J. Virol.*, **84**, 12903–12913.
23. Wang,T., Tian,C., Zhang,W., Luo,K., Sarkis,P.T., Yu,L., Liu,B., Yu,Y. and Yu,X.F. (2007) 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. *J. Virol.*, **81**, 13112–13124.
24. Bach,D., Peddi,S., Mangeat,B., Lakkaraju,A., Strub,K. and Trono,D. (2008) Characterization of APOBEC3G binding to 7SL RNA. *Retrovirology*, **5**, 54.
25. Mbisa,J.L., Barr,R., Thomas,J.A., Vandegraaff,N., Dorweiler,I.J., Svarovskaia,E.S., Brown,W.L., Mansky,L.M., Gorelick,R.J., Harris,R.S. *et al.* (2007) 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.
26. Mbisa,J.L., Bu,W. and Pathak,V.K. (2010) APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. *J. Virol.*, **84**, 5250–5259.
27. Browne,E.P., Allers,C. and Landau,N.R. (2009) Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. *Virology*, **387**, 313–321.
28. Luo,K., Wang,T., Liu,B., Tian,C., Xiao,Z., Kappes,J. and Yu,X.F. (2007) Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J. Virol.*, **81**, 7238–7248.
29. Bishop,K.N., Holmes,R.K. and Malim,M.H. (2006) Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J. Virol.*, **80**, 8450–8458.
30. Iwatani,Y., Chan,D.S., Wang,F., Maynard,K.S., Sugiura,W., Gronenborn,A.M., Rouzina,I., Williams,M.C., Musier-Forsyth,K. and Levin,J.G. (2007) Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res.*, **35**, 7096–7108.
31. Li,X.Y., Guo,F., Zhang,L., Kleiman,L. and Cen,S. (2007) APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription. *J. Biol. Chem.*, **282**, 32065–32074.
32. Wang,X., Ao,Z., Chen,L., Kobinger,G., Peng,J. and Yao,X. (2012) The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. *J. Virol.*, **86**, 3777–3786.
33. Sliva,K., Erlwein,O., Bittner,A. and Schnierle,B.S. (2004) Murine leukemia virus (MLV) replication monitored with fluorescent proteins. *Virol. J.*, **1**, 14.
34. Langlois,M.A., Beale,R.C., Conticello,S.G. and Neuberger,M.S. (2005) Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res.*, **33**, 1913–1923.
35. Naldini,L., Blomer,U., Gallay,P., Ory,D., Mulligan,R., Gage,F.H., Verma,I.M. and Trono,D. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*, **272**, 263–267.
36. Zhang,H., Zhou,Y., Alcock,C., Kiefer,T., Monie,D., Siliciano,J., Li,Q., Pham,P., Cofrancesco,J., Persaud,D. *et al.* (2004) Novel single-cell-level phenotypic assay for residual drug susceptibility and reduced replication capacity of drug-resistant human immunodeficiency virus type 1. *J. Virol.*, **78**, 1718–1729.
37. Simon,J.H. and Malim,M.H. (1996) The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes. *J. Virol.*, **70**, 5297–5305.
38. Prochnow,C., Bransteitter,R., Klein,M.G., Goodman,M.F. and Chen,X.S. (2007) The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*, **445**, 447–451.
39. Krzysiak,T.C., Jung,J., Thompson,J., Baker,D. and Gronenborn,A.M. (2012) APOBEC2 is a monomer in solution: implications for APOBEC3G models. *Biochemistry*, **51**, 2008–2017.
40. Sato,Y., Probst,H.C., Tatsumi,R., Ikeuchi,Y., Neuberger,M.S. and Rada,C. (2010) Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J. Biol. Chem.*, **285**, 7111–7118.
41. Conticello,S.G., Langlois,M.A., Yang,Z. and Neuberger,M.S. (2007) DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv. Immunol.*, **94**, 37–73.

42. Langlois,M.A., Kemmerich,K., Rada,C. and Neuberger,M.S. (2009) The AKV murine leukemia virus is restricted and hypermutated by mouse APOBEC3. *J. Virol.*, **83**, 11550–11559.
43. O'Doherty,U., Swiggard,W.J. and Malim,M.H. (2000) Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. *J. Virol.*, **74**, 10074–10080.
44. Butler,S.L., Hansen,M.S. and Bushman,F.D. (2001) A quantitative assay for HIV DNA integration in vivo. *Nat. Med.*, **7**, 631–634.
45. Brussel,A., Delelis,O. and Sonigo,P. (2005) Alu-LTR real-time nested PCR assay for quantifying integrated HIV-1 DNA. *Methods Mol. Biol.*, **304**, 139–154.
46. Wichroski,M.J., Robb,G.B. and Rana,T.M. (2006) Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS Pathog.*, **2**, e41.
47. Stenglein,M.D., Matsuo,H. and Harris,R.S. (2008) Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *J. Virol.*, **82**, 9591–9599.
48. Feng,Y. and Chelico,L. (2011) 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.
49. Harris,R.S., Petersen-Mahrt,S.K. and Neuberger,M.S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell*, **10**, 1247–1253.
50. Mariani,R., Chen,D., Schrofelbauer,B., Navarro,F., Konig,R., Bollman,B., Munk,C., Nymark-McMahon,H. and Landau,N.R. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*, **114**, 21–31.
51. Ao,Z., Yu,Z., Wang,L., Zheng,Y. and Yao,X. (2008) Vpr14-88-Apobec3G fusion protein is efficiently incorporated into Vif-positive HIV-1 particles and inhibits viral infection. *PloS One*, **3**, e1995.
52. Zhang,S., Pointer,D., Singer,G., Feng,Y., Park,K. and Zhao,L.J. (1998) Direct binding to nucleic acids by Vpr of human immunodeficiency virus type 1. *Gene*, **212**, 157–166.
53. Lavens,D., Peelman,F., Van der Heyden,J., Uyttendaele,I., Catteeuw,D., Verhee,A., Van Schoubroeck,B., Kurth,J., Hallenberger,S., Clayton,R. *et al.* (2010) Definition of the interacting interfaces of Apobec3G and HIV-1 Vif using MAPPIT mutagenesis analysis. *Nucleic Acids Res.*, **38**, 1902–1912.
54. Bulliard,Y., Narvaiza,I., Bertero,A., Peddi,S., Rohrig,U.F., Ortiz,M., Zoete,V., Castro-Diaz,N., Turelli,P., Telenti,A. *et al.* (2011) Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.*, **85**, 1765–1776.
55. Chelico,L., Prochnow,C., Erie,D.A., Chen,X.S. and Goodman,M.F. (2010) Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J. Biol. Chem.*, **285**, 16195–16205.
56. Levy,D.N., Aldrovandi,G.M., Kutsch,O. and Shaw,G.M. (2004) Dynamics of HIV-1 recombination in its natural target cells. *Proc. Natl Acad. Sci. USA*, **101**, 4204–4209.
57. Gelderblom,H.C., Vatakis,D.N., Burke,S.A., Lawrie,S.D., Bristol,G.C. and Levy,D.N. (2008) Viral complementation allows HIV-1 replication without integration. *Retrovirology*, **5**, 60.
58. Iwabu,Y., Mizuta,H., Kawase,M., Kameoka,M., Goto,T. and Ikuta,K. (2008) Superinfection of defective human immunodeficiency virus type 1 with different subtypes of wild-type virus efficiently produces infectious variants with the initial viral phenotypes by complementation followed by recombination. *Microbes Infect.*, **10**, 504–513.
59. Charpentier,C., Nora,T., Tenaillon,O., Clavel,F. and Hance,A.J. (2006) Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J. Virol.*, **80**, 2472–2482.
60. Wodarz,D. and Levy,D.N. (2009) Multiple HIV-1 infection of cells and the evolutionary dynamics of cytotoxic T lymphocyte escape mutants. *Evolution*, **63**, 2326–2339.
61. Russell,R.A., Moore,M.D., Hu,W.S. and Pathak,V.K. (2009) 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.
62. Refsland,E.W., Hultquist,J.F. and Harris,R.S. (2012) Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n. *PLoS Pathog*, **8**, e1002800.
63. Sadler,H.A., Stenglein,M.D., Harris,R.S. and Mansky,L.M. (2010) APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.*, **84**, 7396–7404.
64. Weil,A.F., Ghosh,D., Zhou,Y., Seiple,L., McMahon,M.A., Spivak,A.M., Siliciano,R.F. and Stivers,J.T. (2013) Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc. Natl Acad. Sci. USA*, **110**, E448–E457.
65. Shindo,K., Li,M., Gross,P.J., Brown,W.L., Harjes,E., Lu,Y., Matsuo,H. and Harris,R.S. (2012) A comparison of two single-stranded dna binding models by mutational analysis of APOBEC3G. *Biology*, **1**, 260–276.
66. Chen,K.M., Harjes,E., Gross,P.J., Fahmy,A., Lu,Y., Shindo,K., Harris,R.S. and Matsuo,H. (2008) Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature*, **452**, 116–119.
67. Dapp,M.J., Holtz,C.M. and Mansky,L.M. (2012) Concomitant lethal mutagenesis of human immunodeficiency virus type 1. *J. Mol. Biol.*, **419**, 158–170.

Appendix III: Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis.

Kasandra Bélanger, Mathieu Savoie, Halil Aydin, Tyler Milston Renner, Zahra Montazeri and Marc-André Langlois.

Contribution:

I performed most of the experiments presented in the manuscript. Biostatistical analyses were performed by Dr. Zahra Montazeri and 3D-PCR experiments were conducted by Tyler Renner and Mathieu Savoie. I contributed to the assembly of the manuscript, proofread the manuscript, and assisted in the response to reviewers.

Published: *Virology*, October 2013

Virology 448 (2014) 168–175

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis

Kasandra Bélanger [a], Mathieu Savoie [a], Halil Aydin [a], Tyler Milston Renner [a], Zahra Montazeri [b], Marc-André Langlois [a,*]

[a] Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada
[b] Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 22 August 2013
Returned to author for revisions
15 September 2013
Accepted 4 October 2013
Available online 25 October 2013

Keywords:
APOBEC3
Hypermutation
Cytidine deamination
HIV
Moloney murine leukemia virus

ABSTRACT

Enzymatic deamination of cytidines in DNA is an intrinsic component of antibody maturation and retroviral resistance, but can also be a source of HIV drug resistance and cancer-causing mutations. Here, we developed a high-throughput method based on high resolution melt (HRM) analysis called HyperHRM that can screen genomic DNA for rare hypermutated proviral sequences and accurately quantify the number of C-to-T or G-to-A mutations in each sequence. We demonstrate the effectiveness of the approach by profiling in parallel the intensity of the DNA mutator activity of all seven human APOBEC3 proteins on the near full-length sequence of HIV-1 and the Moloney murine leukemia virus. Additionally, HRM was successfully used to identify hypermutated proviral sequences in peripheral blood mononuclear cells from an HIV-1 patient. These results exemplify the effectiveness of HRM-based approaches for hypermutation quantification and for the detection of hypermutated DNA sequences potentially associated with disease or retroviral drug resistance.

Introduction

Deamination of cytosine (C) into uracil (U) in DNA is a potentially mutagenic process that is fundamental to both adaptive and innate immunity (Conticello et al., 2007). Deoxycytidine deaminases such as the activation-induced deaminase (AID) and APOBEC3 proteins (A3) are enzymes that extensively convert deoxycytidines in single-stranded DNA (ssDNA) into deoxyuridines, a process called hypermutation [see (Conticello, 2008) for a review]. Because human DNA polymerases cannot distinguish U from thymidine (T) in a DNA template, adenine (A) is usually incorporated opposite U during DNA replication, resulting in G-to-A/C-to-T mutations. The physiological roles of the mutator activities of AID and A3 proteins are to promote adaptive immunity through antibody affinity maturation and innate immunity by restricting retroviral infection, respectively. However, recent studies have established strong correlations between the enzymatic activity of A3 proteins and antiretroviral drug resistance (Fourati et al., 2012; Hache et al., 2006; Jern et al., 2009; Mulder et al., 2008), and also with certain forms of cancer (Alexandrov et al., 2013; Burns et al., 2013a, 2013b; Nik-Zainal et al., 2012; Pavri and Nussenzweig, 2011; Roberts et al., 2013). For these reasons, a simple, quick and high-throughput tool for detecting and quantifying hypermutation caused by the enzymatic activities of cytidine deaminases in both the laboratory and clinic is of significant interest.

Until now, current methods were mostly designed to either screen for hypermutated sequences or conduct quantitative analyses of mutation frequency, but not both. Direct DNA sequencing of cloned proviral DNA is an approach whereby PCR is first performed on DNA isolated from an infected cell population. Amplicons are then cloned, and a small selection of clones are analyzed by DNA sequencing [see (Bélanger et al., 2013) for an application example]. Although this is a relatively simple approach for quantifying mutations in a sequence and identifying which bases are mutated, the method is not suited for large-scale screening purposes, is labor intensive because it requires cloning and plasmid purification, and does not provide high-throughput capabilities.

Next generation sequencing (NGS) is another approach that is occasionally used to analyze hypermutations in integrated retroviral DNA. It can provide direct sequence information on mutated residues and the frequency at which these mutations occur in a heterogeneous sample. This method, however, is very costly, must usually be outsourced, and the data analysis is labor intensive and requires specialized software tools. Furthermore, NGS is not a suitable approach for hypermutation screening purposes or for routine

* Correspondence to: Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada K1H 8M5. Tel.: +1 613 562 5800 x7110.
E-mail address: langlois@uottawa.ca (M.-A. Langlois).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2013.10.008

hypermutation analyses *in vitro*. Finally, the read lengths of NGS technologies are relatively short and are thereby not amenable for the analysis of long contiguous stretches (> 9 kb) of proviral DNA as required for near full-length single genome analyses.

For hypermutation screening purposes, there exists a very simple and powerful technique called differential DNA denaturing PCR, or commonly called 3D-PCR (Suspene et al., 2005b). 3D-PCR involves performing a first round of PCR on the gDNA extracted from a cell population, followed by a narrow gradient PCR on a short nested region (generally less than 300 bp). Amplification of experimental samples at lower melting temperatures than unmutated controls suggests the presence of A/T-rich DNA sequences that are a hallmark of A3-induced hypermutation. The greatest advantage of 3D-PCR over other methods is that it can amplify very rare hypermutated sequences in a heterogeneous sample (Suspene et al., 2011, 2005a; Vartanian et al., 2008). However, the strength of this method is also its weakness because it is naturally biased towards amplifying only the most heavily hypermutated sequences. It cannot therefore provide quantitative information on the frequency of hypermutated sequences or the mean intensity of the mutations. Another shortcoming of the method is that it is not amenable to high-throughput applications because of the numerous tubes required for each experimental sample at the gradient PCR stage. Lastly, 3D-PCR only performs well on relatively short DNA amplicons containing high numbers of mutations.

High resolution melting (HRM) is a post-PCR method that detects small nucleotide variations in a given DNA sequence in comparison to a reference. Base pair variations are measured as a function of their impact on the melting temperature of the DNA segment (Vossen et al., 2009; Wittwer, 2009). Accordingly, increasing the A/T content of a short DNA sequence, albeit by only a single base pair substitution, will decrease the melting temperature of that DNA fragment. HRM relies on dyes that specifically emit fluorescence only when bound to dsDNA. HRM analysis is performed following a quantitative PCR (qPCR) reaction by slowly melting dye-saturated PCR amplicons. Fluorescence is then measured at regular temperature intervals until the DNA strands are completely denatured and the fluorescence level reaches background. HRM is most commonly used in single nucleotide polymorphism (SNP) genotyping and DNA methylation analyses (Reed and Wittwer, 2004; Wojdacz and Dobrovic, 2007).

Here, we adapted HRM analysis combined with colony PCR and a custom analysis algorithm to identify hypermutated DNA sequences and directly evaluate the number of G-to-A/C-to-T mutations, without the need for plasmid purification or DNA sequencing. We call this approach HyperHRM. We provide application examples for our methods by analyzing the near full-length genomes of both HIV-1 and MoMLV that have been individually hypermutated by each of the seven human A3 family members. This analysis has allowed us to identify very weakly deaminated zones in the genomes of both viruses and to characterize the specific deamination profile of each A3 protein. Finally, we demonstrate the effectiveness of HRM-based approaches for identifying hypermutated proviral sequences bearing hallmarks of A3 mutator activity in the gDNA of peripheral blood mononuclear cells (PBMCs) from an HIV-1 infected patient.

Results and discussion

Here we propose two strategies that employ HRM to detect sequences with lower melting temperatures due to C-to-T/G-to-A transition mutations (Fig. 1A and Note S1). The first strategy is a direct HRM analysis performed using a single round of qPCR on the gDNA extracted from a cell population. This method is useful for detecting hypermutated sequences within a large pool of unmutated sequences. The second strategy, called HyperHRM, offers more sensitivity and resolution because sequences of interest are first cloned and then their HRM fluorescence output is measured to controls with known numbers of mutations using the algorithm defined in the Materials and Methods Section.

To validate the accuracy of mutation quantification by HyperHRM, we first cloned eGFP-derived DNA sequences containing 1 and 8G-to-A mutations in defined segments of 100 bp to 1 kb in length (Fig. 1B and Table S1). We then resolved the mathematical relationship between the fluorescence measurements taken during the HRM step and the number of mutations in each DNA amplicon of increasing length (see Materials and Methods Section). We established that there exists a robust linear correlation between fluorescence and mutation number at the temperature that maximizes the distance between the melting curves of the controls. R-square values above 0.94 were obtained for every DNA amplicon length tested below 1 kb (Fig. 1B). Furthermore, the approach is also very sensitive; a single mutation could be accurately detected in amplicons of up to 648 bp in length, while the 900 bp amplicon had an accurate detection limit of three mutations (Table 1). Considering that hypermutated sequences contain, by definition, several more C-to-T/G-to-A transition mutations than any other type of mutation, the technique is therefore sensitive enough to identify hypermutation in amplicons up to 900 bp in length. We then examined if HRM is sufficiently sensitive to serve as an effective screening tool for detecting rare hypermutated sequences among a majority of unmutated sequences. To investigate this, we performed serial dilutions of plasmid DNA (containing either 1, 3 or 8 mutations) in unmutated plasmid (Fig. 1C). Our experiment shows that fluorescence was detected above background in all conditions for dilutions up to 1/1000 (Fig. 1C and Table S3).

For comparison purposes, we challenged the sensitivity of our HRM-based approached with that of 3D-PCR (Fig. S1) (Suspene et al., 2005b). Here we tested the sensitivity of 3D-PCR on amplicons of 279 bp and 648 bp in length, containing 1, 3 or 8 G-to-A mutations (Fig. S1). Although 3D-PCR was able to detect all mutations in a short 279 bp amplicon, the method failed to provide a positive signal when 1 or 3 mutations were present in the larger 648 bp amplicon.

Having established that HyperHRM is both sensitive and accurate, we next tested its high-throughput capabilities in a standard retroviral hypermutation assay by analyzing the DNA deamination intensity of each of the seven human A3 proteins along the near full-length (~8.3 kb) proviral DNA sequence of MoMLV. We produced MoMLV reporter virus in cells that transiently expressed one of the seven human A3 proteins. For the mutation analysis, we designed an approach in which the genome of MoMLV is amplified in 5 large PCR amplicons (A–E) (Fig. 2A). HRM is then performed on nested regions within each amplicon (R1–15) (Fig. 2B). Relative fluorescence is calculated as the difference between experimental sample DNA and the unmutated control sequence. HRM hypermutation profiles revealed that APOBEC3A (A3A) had very little activity on MoMLV, as expected, and that APOBEC3G (A3G) displayed the highest mutator activity on the virus.

Considering that many cell types simultaneously express more than one A3 protein, we plotted the combined effect of all seven A3 proteins on the virus to provide a hypothetical hypermutation profile if all the A3 proteins were expressed (Fig. 2C). An intriguing result emerged with region 9 (R9), which appears to be much less targeted for deamination by all members of the A3 family. This type of deamination pattern has been observed before for A3G acting on HIV-1, where the dead zones were identified at the central polypurine tract (cPPT) and LTR proximal polypurine tract (3'PPT) of the virus (Hu et al., 2010; Suspene et al., 2006; Wurtzer et al., 2006). Although the 3'PPT is located immediately upstream of the 3'LTR for all retroviruses, a cPPT has never been described for MoMLV or for retroviruses other than lentiviruses. Sequence analysis of MoMLV R9 reveals the presence of a

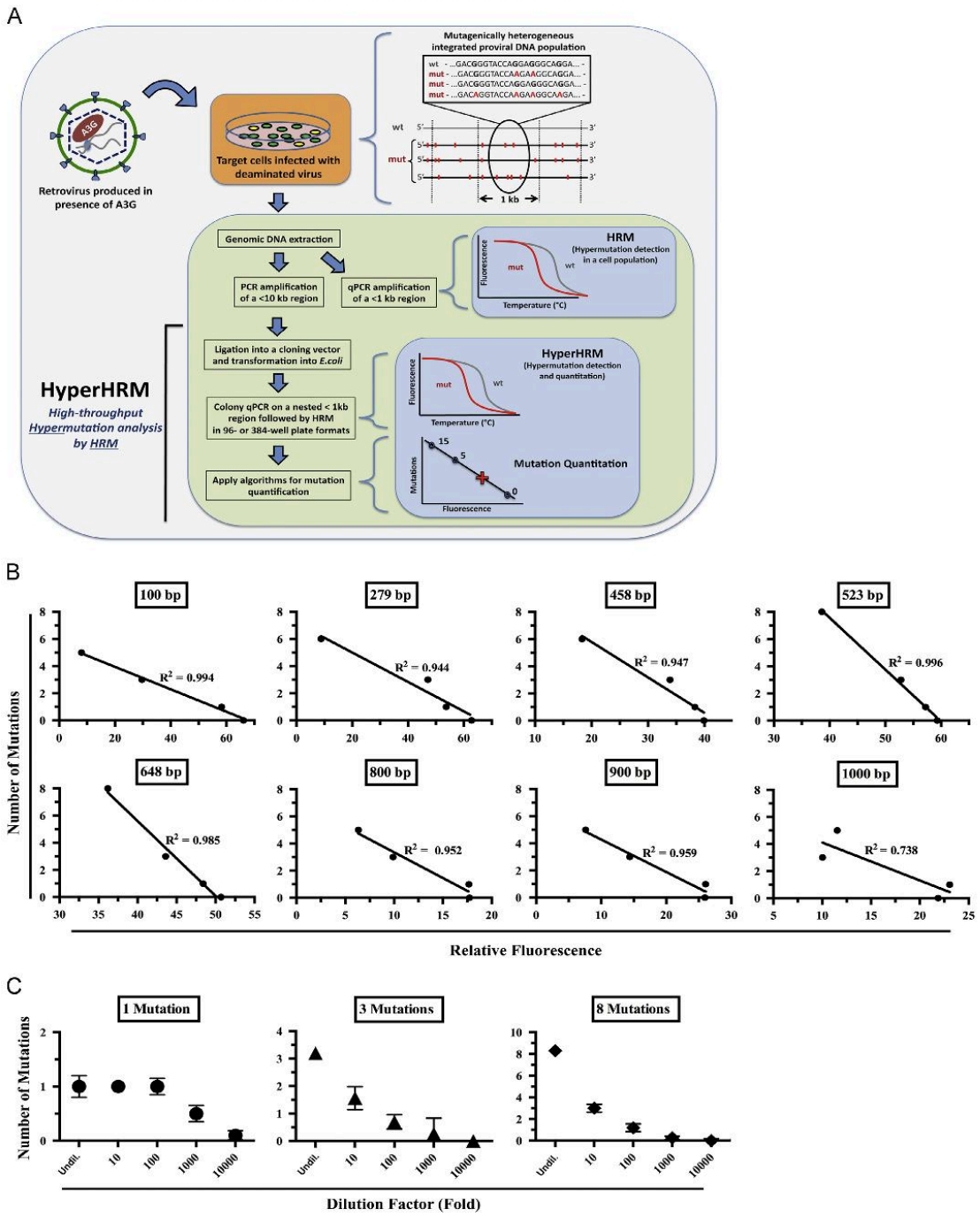


Fig. 1. *Accuracy and sensitivity of HyperHRM*: (A) Flow chart for HyperHRM analysis (see Note S1 for more details). (B) Correlation between PCR amplicon length and the accuracy of mutation number determination by HyperHRM. Plasmids containing a cloned virus-derived sequence of increasing lengths with known numbers of mutations were used for HyperHRM analysis. The analysis was performed twice in quadruplicate. Data points represent average values. (C) Sensitivity of HyperHRM for mutation detection. Plasmids containing a cloned eGFP-derived sequence of 648 bp in length with 1, 3 or 8 mutations were diluted at the indicated ratios with a plasmid containing the parental sequence with no mutations (Undil.: undiluted). The analysis was performed three times in quadruplicate. Data points represent mean ± standard deviation (SD).

Table 1
Correlation between PCR amplicon length and the accuracy of mutation number determination by HyperHRM analysis.

HRM amplicon length (bp)[a]	Mutations in control sequence determined by HyperHRM analysis						Optimal temperature (°C)[b]	R^{2c}
	0	*1*	*3*	*5*	*6*	*8*		
100	0.0	0.8	3.6	5.7	–	–	85.2	0.994
279	0.3	1.3	2.1	–	6.2	–	87.8	0.944
428	0.3	1.0	2.2	–	6.2	–	87.9	0.947
523	0.2	1.0	2.7	–	–	8.1	86.6	0.996
648	0.0	1.0	3.6	–	–	7.7	87.9	0.985
800	0.4	0.4	3.4	4.7	–	–	89.6	0.952
900	0.5	0.4	3.2	4.9	–	–	89.7	0.959
1000	1.0	0.4	6.6	5.9	–	–	89.4	0.738

Results represent the average of quadruplicate values. Values in *italics* indicate the actual number of G-to-A mutations in the control sequence as determined by DNA sequencing.

[a] All amplicons used here are located within regions R11–R13 of MoMLV proviral DNA (Fig. 2A).
[b] Calculated temperature at which fluorescence is maximized between the melting curves of the calibration controls.
[c] R-square value of the standard curve.

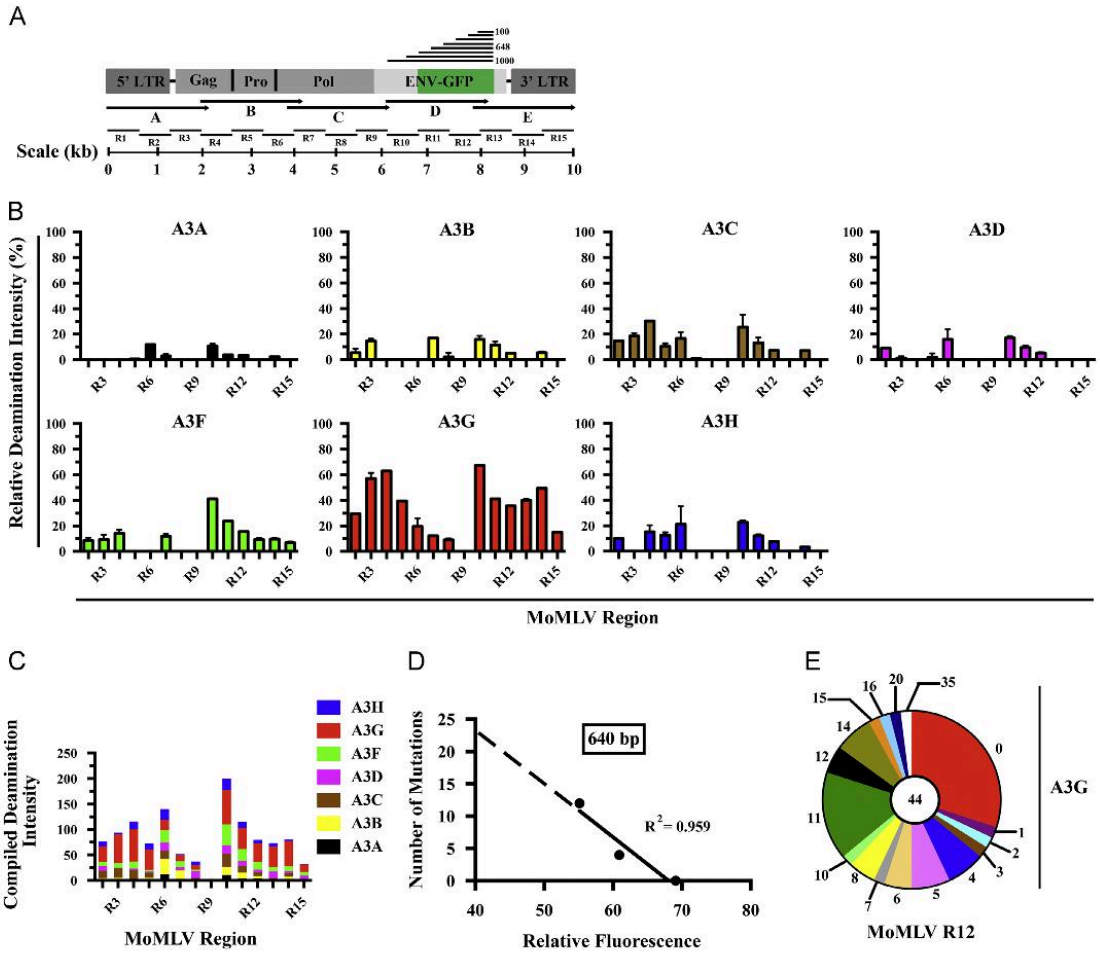


Fig. 2. *Hypermutation analysis of MoMLV proviral DNA deaminated by each of the seven human A3 proteins*: (A) Schematic representation of the regions analyzed in MoMLV proviral DNA. (B) HRM analysis performed on each of the 15 proviral regions. MoMLV was independently produced in presence of each of the seven A3 proteins. (C) Compilation of the results from (B) illustrating the contribution of A3 proteins to the overall deamination of each proviral region. (D) Linear regression plot used to calculate mutation numbers in R12 of MoMLV by HyperHRM. (E) HyperHRM performed on 44 clones of MoMLV region 12 (R12) deaminated by A3G. The size of each sector of the pie chart represents the percentage of all sequences containing the indicated number of mutations.

purine-rich patch that shares a striking resemblance with elements of the cPPT sequences from various lentiviruses (Table S4). Functional assays are required to confirm if R9 contains indeed a *bona fide* cPPT, or if it is simply a region with strong secondary structures that are resistant to deamination.

We next selected a region that was highly susceptible to A3G deamination (in this case R12), and analyzed individual sequences by HyperHRM. The optimal temperature for performing the analysis was identified by applying the algorithm to the melting curve fluorescence data points obtained with various controls (Fig. 2D).

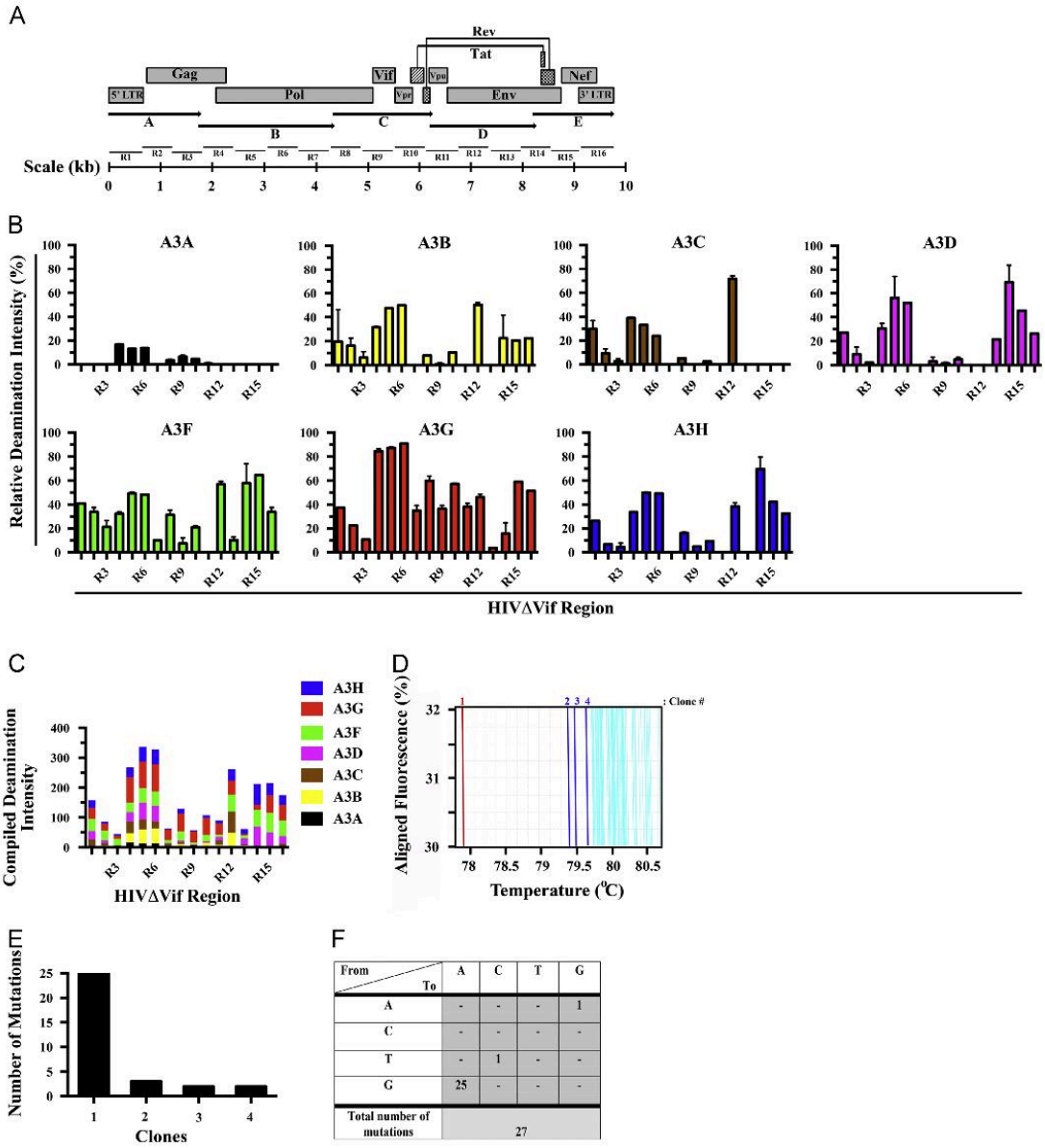


Fig. 3. *Hypermutation analysis of HIV-1 proviral DNA in vitro and in an HIV clinical sample*: (A) Schematic representation of the regions analyzed in HIV-1ΔVif proviral DNA. (B) HRM analysis performed on each of the 16 proviral regions individually deaminated by the seven A3 proteins. (C) Compilation of the results from (B) illustrating the contribution of A3 proteins to the overall deamination of each proviral region. (D) HRM analysis output graph of 96 clones isolated an HIV-1 infected individual. Four sequences with lower melting temperatures than the median value are highlighted in red and dark blue. (E) Number of G-to-A mutations in each of the 4 clones that displayed lower melting temperatures assessed by DNA sequencing. (F) Mutation analysis of hypermutated clone #1 by DNA sequencing.

We then used linear regression to quantify the number of mutations in 44 randomly selected clones (Fig. 2E). We found that most sequences were hypermutated, with some having up to 35 mutations per 640 bp sequence. The overall average mutation rate as measured by HyperHRM was approximately 11 mutations per kb, which is fully consistent with a recently published report from our group using direct DNA sequencing to measure the deamination rate in that same region (Bélanger et al., 2013). To confirm these hypermutation levels, we also analyzed the clones using standard DNA sequencing. Our results indicate that all the mutations were G-to-A transitions (Fig. S2), and present within a dominant 5'-**G**GG-3' local sequence consensus as read on the viral plus-strand DNA (Table S5).

Although hypermutation has a prominent role in maintaining host health and fitness through adaptive and innate immunity, it has been shown that A3G-induced hypermutation of HIV-1 can give rise to drug-resistant viruses (Fourati et al., 2012; Hache et al., 2006; Jern et al., 2009; Mulder et al., 2008). A high-throughput assay that can quickly analyze a large number of viral quasispecies and identify hypermutation in defined viral regions could help predict potential drug resistance in HIV patients. Here we asked whether HRM-based approaches could be used for such a task. As with MoMLV, we partitioned the A3-sensitive HIV-1ΔVif genome (~9.7 kb) in 5 large segments (A–E) and then further into 16 smaller regions (Fig. 3A). We then examined the effect of all seven A3 proteins on integrated proviral DNA using HRM (Fig. 3B). We found that, as previously published, A3D, A3F, A3G and A3H have a very potent mutator effect on HIV-1ΔVif (Hultquist et al., 2011). Combined deamination profiles show that deamination intensity is weakest around regions 9 and 13, which are in close proximity to the central polypurine tract (cPPT) and LTR proximal polypurine tract (3'PPT) respectively (Fig. 3C) (Suspene et al., 2006).

Finally, we used HRM to directly screen the gDNA of PBMCs of an HIV-1 infected patient (Group M, subtype B). Our analysis identified 4 clones displaying significantly lower melting temperatures than the median value of the 96 clones analyzed. DNA sequencing revealed that proviral sequences in dark blue contained 2 or 3 G-to-A mutations (Fig. 3D and E), however these were not in a 5'-**G**G or 5'-**G**A local sequence context typical of A3-mediated deamination (data not shown). These mutations are most likely the result of viral reverse transcriptase error rather than A3-mediated deamination. On the other hand, clone 1 shown in red was heavily hypermutated (Fig. 3D and E), with almost all mutations being G-to-A's with adjacent sequence hallmarks associated with A3 deamination (Fig. 3F). Despite being unable to determine the exact numbers of G-to-A transition mutations in a HIV-1 clinical sample due to the broad genetic diversity of proviral quasispecies, HRM can identify sequences with significantly lower melting temperatures that have a much higher probability of being hypermutated. DNA sequencing can afterwards be performed on a far smaller cohort of samples, thus reducing processing times and costs associated with DNA sequencing.

Conclusions

Overall, HRM and HyperHRM offer important technical, practical and financial advantages over DNA sequencing and 3D-PCR for the detection and quantification of hypermutations in gDNA. We have shown that both these techniques can be used to easily identify hypermutated DNA sequences within a pool of unmutated sequences, as well as DNA containing a single C-to-T/G-to-A substitution. In addition to being a powerful screening tool, HyperHRM can also determine with near perfect accuracy the number of these substitutions in DNA amplicons of up to 900 bp in length.

By using HRM analysis, we identified regions in HIV-1 and MoMLV proviral DNA that are preferentially targeted by each of the seven human A3 family members. Several hundred clones for each virus would have been required to carry out such an experiment using standard DNA sequencing. By taking advantage of the high-throughput efficiency of HRM analysis, the effect of up to six different A3 proteins can be analyzed in parallel on a near full-length retroviral genome in less than 2 h using a single 96-well plate. As such, HRM lead to the identification of a previously undescribed region in the MoMLV genome that is infrequently targeted by A3 proteins. Interestingly, a closer look at this region revealed sequence similarities to the cPPT of lentiviruses. We also show in this report how HRM can be used to identify and isolate hypermutated sequences in an HIV-1 clinical sample. Afterwards, sequences of interest can be individually analyzed for the presence of specific drug-resistance mutations caused by A3 deamination.

In summary, HyperHRM is a quick, inexpensive, accurate and high-throughput method for both the detection and quantification of C-to-T and G-to-A hypermutation in genomic DNA. Standard HRM analysis is a useful tool to quickly screen for the presence of hypermutated sequences in a pool of non-hypermutated sequences. HyperHRM can additionally be used to accurately quantify the number of G-to-A and C-to-T mutations in a given DNA amplicon. The combination of these unique features suggests that HyperHRM has the potential to make important contributions to studies investigating the impact of A3-induced hypermutation on host/virus interactions.

Materials and methods

Cells, viruses and A3 expression vectors

Human embryonic kidney epithelium 293T (293T) and mouse embryonic fibroblast (NIH 3T3) cells were cultured in HyClone DMEM/High Glucose medium (Thermo Fischer Scientific) supplemented with 10% fetal bovine serum (FBS), 100U/ml penicillin and 100 μg/ml streptomycin (Wisent). Human T4-lymphoblastoid cells (CEM-SS) were cultured in RPMI medium (Promega) supplemented with 10% FBS, 100U/ml penicillin and 100 μg/ml streptomycin (Wisent). eGFP-tagged A3 and A2 expression plasmids and replicative MoMLV (pMOV-eGFP) have previously been described (Bélanger et al., 2013). HIV-1ΔVif consists of the molecular clone NL4-3 (CXCR4 tropic; subtype B) with two non-sense mutations introduced in the vif gene by site-directed mutagenesis (Bélanger et al., 2013). Expression plasmid pNL4-3 was kindly provided by the NIH AIDS Research and Reference and Reagent Program (Catalog #114).

Transfections and virus production

Viruses were produced by transfecting 3×10^5 293T cells seeded in a 6-well plate with 800 ng of pMOV-eGFP or 500 ng of pHIV-1ΔVif using the GeneJuice transfection reagent (EMD Millipore) (Bélanger et al., 2013). Standard restriction assays on MoMLV and HIV-1ΔVif were done by co-transfecting 80 ng or 150 ng of expression vectors encoding for A2 or A3 cDNA at a 1:10 or 1:3 ratio in respect to pMOV-eGFP or pHIV-1ΔVif, respectively. Cells were washed with phosphate-buffered saline (PBS) 16 h following transfection and grown for an additional 48 h in complete DMEM media. Virus-containing supernatants were then collected, cleared by centrifugation and filtered through 0.45 μm cartridge filters and treated with the *Dpn I* endonuclease to remove possible plasmid carry-over. Virus production was quantified and normalized by enzyme-linked immunosorbent assay for p30 or p24 (QuickTiter™ ELISA Kit, Cell Biolabs Inc.)

Infection assays and genomic DNA extractions

Target NIH 3T3 or CEM-SS cells were infected with a volume of virus-containing supernatant yielding an MOI of 0.5 when co-transfected with the APOBEC2 control (Bélanger et al., 2013). Infections were carried out by spinoculation at 800 × *g* for 1 h in presence of 8 μg/ml of polybrene. Genomic DNA was extracted from infected cells 24 h (MoMLV) and 48 h (HIV-1ΔVif) after infection using the Wizard genomic DNA purification kit (Promega) according to the manufacturer's specifications.

HIV clinical samples

Genomic DNA extracted from the PBMCs of an HIV-1-infected patient was graciously provided by Dr. Jonathan Angel, Director of the Ottawa Hospital HIV clinic.

Hypermutation analysis by DNA sequencing

Region 12 of MoMLV proviral DNA was amplified from the gDNA of infected NIH 3T3 cells using the PrimeStar HS high fidelity polymerase (Takara Bio Inc.) with the corresponding primers listed in Table S1. The reaction parameters were 92 °C for 5 min, followed by 30 cycles (92 °C for 30 s, 55 °C for 10 s, and 72 °C for 1 min), and finally 10 min at 72 °C. Purified PCR products were then cloned using the TOPO-Blunt cloning kit (Invitrogen) and sequenced using M13 reverse primers (Bélanger et al., 2013).

HyperHRM analysis

HyperHRM sample preparation

Proviral DNA was amplified from the gDNA of virus-infected cells by PCR using primer pairs for viral regions A–E (Tables S1 and S2). Amplicons were then cloned using the TOPO-Blunt cloning kit (Invitrogen). Colonies positive for a cloned insert were diluted in 250 μl of water, of which 8 μl were used for HRM analysis in a 96-well plate format. To generate standard curves for quantification purposes, 8 μl of 12.5 fg/μl stocks of plasmid DNA containing defined numbers of G-to-A mutations in each of the 15 (MoMLV) or 16 (HIV-1ΔVif) viral R-regions were used.

HRM sample preparation

20 ng of gDNA (diluted to 2.5 ng/μl) isolated from virus-infected cell lines were used directly for HRM analysis described below.

HRM sample preparation from an HIV-1 patient

100 ng of gDNA from the PBMCs of an HIV-1 infected patient were amplified using primers HIV-Pol FWD and HIV-Pol REV (Table S2). Amplicons were then cloned using the TOPO-Blunt cloning kit (Invitrogen). Colonies positive for a cloned insert were diluted in 250 μl of water, of which 8 μl were used for HRM analysis in a 96-well plate format.

HRM analysis

Each 20 μl reaction contained 10 μL of the MeltDoctor Master Mix (Life Technologies, Catalog # 4415440), 0.6 μl (0.3 μM final) of each R-region primer, 8 μl of the DNA template and 0.8 μl of water. Primer sequences for the amplification and analysis of mutations in MoMLV and HIV-1 proviral DNA are found in Tables S1 and S2 respectively. The conditions for qPCR were 94 °C for 2 min, followed by 40 cycles (95 °C for 30 s, 55 °C for 45 s, and 72 °C for 1 min). qPCR amplification was immediately followed by a melting curve analysis in which DNA amplicons were gradually heated from 72 °C to 95 °C and fluorescence values were acquired at 0.025 °C intervals. qPCR and HRM were performed on a ViiA™ 7 real-time PCR instrument (Applied Biosystems), and data was analyzed using the ViiA™ 7 system software (Applied Biosystems).

Bioinformatics analysis for mutation quantification (algorithm determination)

By systematically testing the correlation between mutation numbers and fluorescence at different temperatures along the melting curves, we found a strong linear relationship at the temperature at which the relative distance between the melting curves of the controls is maximal. To determine mutation numbers in samples, we imported the HRM melting curve data file to a Microsoft Excel spreadsheet and identified the temperature at which the difference in fluorescence between the controls was maximized using the following formula: $\max_{1 \le j \le n} \sum_{i=1}^{k} (x_{i,j} - x_{i+1,j})$
where:

n is the total number of temperatures of the sample analyzed;
k is number of control samples to generate the standard curve;
$x_{i,j}$ is the amount of fluorescence for curve i at temperature j.

Linear regression was then used to evaluate the number of mutations in experimental samples as a function of the relative fluorescence at the optimal temperature using the formula $y = A + Bx$; where y is the number of mutations (dependent variable), and x is the amount of fluorescence (independent variable). The R Statistical Software (R Development Core Team) was used to conduct correlation analyses and Microsoft Excel was utilized for optimal temperature identification and regression analyses.

Funding

This research was supported by grant #89774 from the Canadian Institutes of Health Research, and an Early Researcher Award from the Ontario Ministry of Research and Innovation to M.-A.L.

Acknowledgments

We are especially grateful to Martin Bourbonnière for technical support with HRM. Special thanks also go to Jonathan Angel and Jason Fernandes for providing HIV-1-infected PBMCs, and to Kristin Kemmerich for helpful discussions and comments on the manuscript. K.B. holds an Ontario Graduate Scholarship and a scholarship from le Fonds de Recherche en Santé du Québec. M.-A.L. holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.10.008.

References

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjord, J.E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinsk, M., Jager, N., Jones, D.T., Jones, D., Knappskog, S., Kool, M., Lakhani, S.R., Lopez-Otin, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J.V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N., Valdes-Mas, R., van Buuren, M.M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Australian Pancreatic Cancer Genome, I., Consortium, I.B.C., Consortium, I.M.-S., PedBrain, I., Zucman-Rossi, J., Andrew Futreal, P., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M.,

Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J., Stratton, M.R., 2013. Signatures of mutational processes in human cancer. Nature (in press).
Bélanger, K., Savoie, M., Rosales Gerpe, M.C., Langlois, M.A., 2013. Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. Nucleic Acids Res. (in press).
Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., Yee, D., Temiz, N.A., Donohue, D.E., McDougle, R.M., Brown, W.L., Law, E.K., Harris, R.S., 2013a. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494, 366–370.
Burns, M.B., Temiz, N.A., Harris, R.S., 2013b. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. (in press).
Conticello, S.G., 2008. The AID/APOBEC family of nucleic acid mutators. Genome. Biol. 9, 229.
Conticello, S.G., Langlois, M.A., Yang, Z., Neuberger, M.S., 2007. DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv. Immunol. 94, 37–73.
Fourati, S., Malet, I., Lambert, S., Soulie, C., Wirden, M., Flandre, P., Fofana, D.B., Sayon, S., Simon, A., Katlama, C., Calvez, V., Marcelin, A.G., 2012. E138K and M184I mutations in HIV-1 reverse transcriptase coemerge as a result of APOBEC3 editing in the absence of drug exposure. AIDS. 26, 1619–1624.
Hache, G., Mansky, L.M., Harris, R.S., 2006. Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. AIDS Rev. 8, 148–157.
Hu, C., Saenz, D.T., Fadel, H.J., Walker, W., Peretz, M., Poeschla, E.M., 2010. The HIV-1 central polypurine tract functions as a second line of defense against APOBEC3G/F. J. Virol. 84, 11981–11993.
Hultquist, J.F., Lengyel, J.A., Refsland, E.W., LaRue, R.S., Lackey, L., Brown, W.L., Harris, R.S., 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J. Virol. 85, 11220–11234.
Jern, P., Russell, R.A., Pathak, V.K., Coffin, J.M., 2009. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS. Pathog. 5, e1000367.
Mulder, L.C., Harari, A., Simon, V., 2008. Cytidine deamination induced HIV-1 drug resistance. Proc. Natl. Acad. Sci. USA 105, 5501–5506.
Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K.W., Mudie, L.J., Varela, I., McBride, D.J., Bignell, G.R., Cooke, S.L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P.S., Davies, H.R., Papaemmanuil, E., Stephens, P.J., McLaren, S., Butler, A.P., Teague, J.W., Jonsson, G., Garber, J.E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerod, A., Tutt, A., Martens, J.W., Aparicio, S.A., Borg, A., Salomon, A.V., Thomas, G., Borresen-Dale, A.L., Richardson, A.L., Neuberger, C., Futreal, P.A., Campbell, P.J., Stratton, M.R., Breast Cancer Working Group of the International Cancer Genome, C., 2012. Mutational processes molding the genomes of 21 breast cancers. Cell 149, 979–993.
Pavri, R., Nussenzweig, M.C., 2011. AID targeting in antibody diversity. Adv. Immunol. 110, 1–26.
Reed, G.H., Wittwer, C.T., 2004. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clin. Chem. 50, 1748–1754.
Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., Harris, S., Shah, R.R., Resnick, M.A., Getz, G., Gordenin, D.A., 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat. Genet. (in press).
Suspene, R., Aynaud, M.M., Guetard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J.P., Wain-Hobson, S., 2011. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proc. Natl. Acad. Sci. USA 108, 4858–4863.
Suspene, R., Guetard, D., Henry, M., Sommer, P., Wain-Hobson, S., Vartanian, J.P., 2005a. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc. Natl. Acad. Sci. USA 102, 8321–8326.
Suspene, R., Henry, M., Guillot, S., Wain-Hobson, S., Vartanian, J.P., 2005b. Recovery of APOBEC3-edited human immunodeficiency virus G- > A hypermutants by differential DNA denaturation PCR. J. Gen. Virol. 86, 125–129.
Suspene, R., Rusniok, C., Vartanian, J.P., Wain-Hobson, S., 2006. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucleic. Acids. Res. 34, 4677–4684.
Vartanian, J.P., Guetard, D., Henry, M., Wain-Hobson, S., 2008. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science 320, 230–233.
Vossen, R.H., Aten, E., Roos, A., den Dunnen, J.T., 2009. High-resolution melting analysis (HRMA): more than just sequence variant screening. Hum. Mutat. 30, 860–866.
Wittwer, C.T., 2009. High-resolution DNA melting analysis: advancements and limitations. Hum. Mutat. 30, 857–859.
Wojdacz, T.K., Dobrovic, A., 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic. Acids. Res. 35, e41.
Wurtzer, S., Goubard, A., Mammano, F., Saragosti, S., Lecossier, D., Hance, A.J., Clavel, F., 2006. 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.

Appendix IV: *N*-Linked Glycosylation Protects Gammaretroviruses against Deamination by APOBEC3 Proteins.

María Carla Rosales Gerpe, Tyler Milston Renner, Kasandra Bélanger, Cindy Lam, Halil Aydin and Marc-André Langlois.

Contribution:

I performed site-directed mutagenesis to generate mutant viruses used in this manuscript and provided intellectual and technical expertise for the achievement of HRM and other experiments. I proofread the manuscript and participated in the review process.

Published: *Journal of Virology*, June 2015

JVI Journals.ASM.org

N-Linked Glycosylation Protects Gammaretroviruses against Deamination by APOBEC3 Proteins

María Carla Rosales Gerpe, Tyler Milston Renner, Kasandra Bélanger, Cindy Lam, Halil Aydin,* Marc-André Langlois

Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, Canada

ABSTRACT

Retroviruses are pathogens with rapid infection cycles that can be a source of disease, genome instability, and tumor development in their hosts. Host intrinsic restriction factors, such as APOBEC3 (A3) proteins, are constitutively expressed and dedicated to interfering with the replication cycle of retroviruses. To survive, propagate, and persist, retroviruses must counteract these restriction factors, often by way of virus genome-encoded accessory proteins. Glycosylated Gag, also called glycosylated Pr80 Gag (gPr80), is a gammaretrovirus genome-encoded protein that inhibits the antiretroviral activity of mouse A3 (mA3). Here we show that gPr80 exerts two distinct inhibitory effects on mA3: one that antagonizes deamination-independent restriction and another one that inhibits its deaminase activity. More specifically, we find that the number of *N*-glycosylated residues in gPr80 inversely correlates with the sensitivity of a gammaretrovirus to deamination by mouse A3 and also, surprisingly, by human A3G. Finally, our work highlights that retroviruses which have successfully integrated into the mouse germ line generally express a gPr80 with fewer glycosylated sites than exogenous retroviruses. This observation supports the suggestion that modulation of A3 deamination intensity could be a desirable attribute for retroviruses to increase genetic diversification and avoid immune detection. Overall, we present here the first description of how gammaretroviruses employ posttranslational modification to antagonize and modulate the activity of a host genome-encoded retroviral restriction factor.

IMPORTANCE

APOBEC3 proteins are host factors that have a major role in protecting humans and other mammals against retroviruses. These enzymes hinder their replication and intensely mutate their DNA, thereby inactivating viral progeny and the spread of infection. Here we describe a newly recognized way in which some retroviruses protect themselves against the mutator activity of APOBEC3 proteins. We show that gammaretroviruses expressing an accessory protein called glycosylated Gag, or gPr80, use the host's posttranslational machinery and, more specifically, *N*-linked glycosylation as a way to modulate their sensitivity to mutations by APOBEC3 proteins. By carefully controlling the amount of mutations caused by APOBEC3 proteins, gammaretroviruses can find a balance that helps them evolve and persist.

Retroviruses are exceptional pathogens in that they permanently modify the genome of their host upon infection. Proviral integration can lead to deleterious insertions in the coding sequence of genes and thereby alters the sequence, stability, splicing, and function of host mRNAs. Additionally, because the genomes of retroviruses encode an active promoter and enhancer sequences, they can also influence the expression of nearby host genes, which can lead to diseases such as cancer (1). In response to this potential threat to their genome, vertebrates express several intrinsic antiretroviral restriction factors that are dedicated to the prevention of infection, replication, release, and spread of retroviruses (2). There are several retroviral restriction factors that operate in parallel in mammals; the most notable are BST-2, TRIM5-α, SAMHD1, and APOBEC3 (A3) proteins, as well as several other factors that are dependent on an interferon response for expression (3). To persist, retroviruses therefore need to concomitantly develop countermeasures to all these restrictions factors, and these are often in the form of accessory proteins or genetic substitutions at sites of interactions with restriction factors (1, 4, 5). HIV-1 and HIV-2 are among the retroviruses that have been the most successful at avoiding restriction by the host, whereby BST-2 is defeated by expression of the viral accessory protein Vpu (6, 7), SAMHD1 is defeated by Vpx (8, 9), A3 (A3F, A3G, A3D, and A3H) is defeated by Vif (10), and TRIM5α is defeated by the virus evolving target sequences that avoid its capsid being recognized by this restriction factor (11–14). HIV is not the only retrovirus that successfully counteracts the effects of restriction factors. Some murine leukemia viruses (MLVs) have also developed ways to avoid detection and restriction by host retroviral restriction factors (15–18).

A3 proteins constitute a family of cytidine deaminases that convert deoxycytidines into deoxyuridines in single-stranded DNA (for a review, see reference 19). These proteins have a major role in intrinsic defenses against foreign naked DNA, some DNA viruses, and retroelements but mostly against retroviruses. Hu-

Received 17 November 2014 Accepted 1 December 2014

Accepted manuscript posted online 10 December 2014

Citation Rosales Gerpe MC, Renner TM, B langer K, Lam C, Aydin H, Langlois M-A. 2015. *N*-Linked glycosylation protects gammaretroviruses against deamination by APOBEC3 proteins. J Virol 89:2342–2357. doi:10.1128/JVI.03330-14.

Editor: S. R. Ross

Address correspondence to Marc-Andr Langlois, langlois@uottawa.ca.

M.C.R.G. and T.M.R. are co-first authors.

* Present address: Halil Aydin, Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

doi:10.1128/JVI.03330-14

mans and primates express seven A3 proteins (A3A, A3B, A3C, A3DE, A3F, A3G and A3H), with A3G being the most potent against retroviruses. Deamination occurs primarily on single-stranded minus-strand viral DNA during reverse transcription, resulting in C-to-U transition mutations. Newly generated uracils then direct the incorporation of adenines on the plus-strand DNA, thereby generating G-to-A mutations. Intense deamination or hypermutation results in inactivating substitutions and premature stop codons in viral genes. A3 proteins can also, under some experimental conditions, inhibit retrovirus infection by multiple deamination-independent mechanisms (20–28).

Mice, in contrast to humans and primates, express APOBEC3 (mouse A3 [mA3]) proteins only from a single gene (29, 30). Although mA3 is able to potently restrict and hypermutate HIV and simian immunodeficiency virus in a Vif-independent manner *in vitro* (29, 31), it displays a variable and often much weaker ability to restrict murine retroviruses *in vivo* and *in vitro* (32–35). This variability is likely due to mA3's prevailing inability to hypermutate most murine retroviruses, and therefore, mA3 does not genetically inactivate circulating viruses. The most striking example is the Moloney MLV (M-MLV) gammaretrovirus, which is modestly restricted by mA3 in the absence of detectable G-to-A hypermutation, despite efficiently packaging mA3 into virions (33, 36–40). However, M-MLV becomes much more sensitive to restriction by mA3 when expression of the glycosylated viral Pr80 Gag (gPr80) protein is suppressed (15, 18). This is reportedly caused by gPr80 blocking access of mA3 to the reverse transcription complex in viral cores (15).

The gPr80 protein of M-MLV appears as a multiband protein with an apparent molecular mass ranging from 80 to 100 kDa on SDS-polyacrylamide gels (41, 42). This glycoprotein is expressed from an alternative in-frame CUG start codon 264 bases upstream of the primary AUG initiation codon of the structural polyprotein $Pr65^{Gag}$, resulting in 88 additional amino acids at the N terminus (43). Glycosylated Pr80 is then further processed by cellular proteases that cleave it into a 40-kDa amino-terminal fragment (containing the gPr80 leader polypeptide and viral matrix [MA] and p12 proteins) and a 55-kDa carboxy-terminal moiety (containing the viral capsid [CA] and nucleocapsid [NC]) (44, 45). Although the N-terminal 40-kDa moiety can associate with the cell membrane as a type II integral membrane protein, it is also present in secreted viral particles (43, 45, 46). While the primary function of gPr80 has not been fully elucidated, its described roles include facilitating late-stage viral release from infected cells, increasing viral core stability and integrity, and improving viral spreading and pathogenesis *in vivo* (15, 38, 47–50).

AKV MLV is an endogenous murine gammaretrovirus that is highly similar in genomic sequence to M-MLV. However, AKV, in contrast to M-MLV, is sensitive to deamination by mA3 (33). In this study, we mapped AKV's sensitivity to deamination and restriction by mA3 and human A3G (hA3G) to the *pr80* gene sequence of the virus. We identified three putative *N*-linked glycosylated sites in gPr80 of M-MLV but only two in that of AKV. Our biochemical and cell-based analyses show that the number of glycosylated sites in gPr80 inversely correlates with the level of resistance to deamination. Abolishing gPr80 expression also resulted in hypermutated virus, indicating that gPr80 and the *N*-linked glycans attached to it are together an integral part of the resistance mechanism. Additionally, we demonstrate that genetically modified mouse gammaretroviruses with fewer glycosylated sites in gPr80 are mutated by endogenous mA3 expressed in murine primary splenocytes. However, to our surprise, these mutated viruses remained infectious and capable of replication. These results highlight the important contribution of the host's posttranslational modification machinery to helping gammaretroviruses modulate their sensitivity to deamination by mA3. This could constitute a novel strategy used by gammaretroviruses to increase their genetic diversity and avoid immune detection.

MATERIALS AND METHODS

Mice. All breeding and manipulations performed on animals were conducted in accordance with the Ontario Animals for Research Act and were approved by the University of Ottawa Animal Ethics Committee (protocol number ME-133). A3-deficient ($mA3^{-/-}$) mice (mA3-knockout [KO] mice) were backcrossed 12 times to a C57BL/6 mouse background. These mice have the A3 gene disrupted by the insertion of a neomycin resistance cassette in exon 3 (51). C57BL/6 (mA3 wild-type [WT]) and mA3-KO mice were maintained in the barrier unit of the University of Ottawa Animal Care Facility.

Cells. Human embryonic kidney epithelium (293T) cells and mouse embryonic fibroblasts (NIH 3T3 cells) were cultured in HyClone Dulbecco modified Eagle medium (high-glucose medium) supplemented with 10% decomplemented fetal bovine serum (FBS), 100 U/ml penicillin, and 100 μg/ml streptomycin and propagated at 37°C in a 5% CO_2 incubator. Mouse splenocytes were prepared by homogenizing the spleens of neonatal mouse pups 3 to 5 days of age by enforced passage through a 70-μm-mesh size nylon cell strainer as previously described (33).

Viruses and APOBEC expression vectors. The pMOV-eGFP expression vector encoding replicative M-MLV and the pAKV-NB-eGFP viral plasmid encoding replicative AKV MLV have been described before (33). Expression vectors for the Flag-tagged C57BL/6 mA3 delta exon 5 allele (referred to throughout as mA3), Flag-tagged human APOBEC expression vectors (hA2 and hA3G), and their respective catalytically inactive mutants (mA3 [E73A] and hA3G [E257A]) have been described before (33, 52). Hybrid viruses were generated by replacing DNA sequences in M-MLV by orthologous sequences from AKV (see Fig. 3A). Viruses with point mutations in the *pr80* gene sequence (for M-MLV, N113Q, N480Q, and N505Q; for AKV, N113D and S507N) were made by using a QuikChange XL site-directed mutagenesis kit (Agilent Technologies) according to the manufacturer's specifications. M-MLV (CTA) and AKV (CTA), which do not express gPr80, were generated by replacing the CTG alternative initiation codon in proviral plasmid DNA by CTA using site-directed mutagenesis.

Viral infection and spreading assays. The procedures used for the viral infection and spreading assays are graphically presented in Fig. 1.

(i) *In vitro* assay. To produce replicative viruses in the *in vivo* assay (Fig. 1A), 3×10^5 293T cells were seeded in 6-well plates and grown for 24 h until they reached 50 to 60% confluence. Before transfection, the medium was replaced. Cotransfections were done with 800 ng of viral expression plasmid and 100 ng of A3 expression plasmids, using the GeneJuice transfection agent (Novagen) according to the manufacturer's instructions. Virus-containing supernatant was harvested at 48 h after transfection. At 24 h prior to infection, 1×10^5 NIH 3T3 target cells were seeded in 12-well plates and incubated for 24 h. At the time of the infection, NIH 3T3 cell medium was replaced by fresh decomplemented medium containing Polybrene (Sigma-Aldrich) at a concentration of 8 μg/ml. Virus-containing supernatants were cleared by centrifugation, and the concentration of viral p30 CA protein was evaluated by enzyme-linked immunosorbent assay (ELISA; QuickTiter ELISA kit; Cell Biolabs Inc.). The amount of p30 CA protein yielding a multiplicity of infection (MOI) of 1 for virus produced in the presence of A2 was calculated by infection titration. Similar amounts of p30 protein were then used for virus produced in the presence of mA3, hA3G, and their catalytically inactive mutants. The cells were spin infected at $800 \times g$ for 1 h. Infected target cells

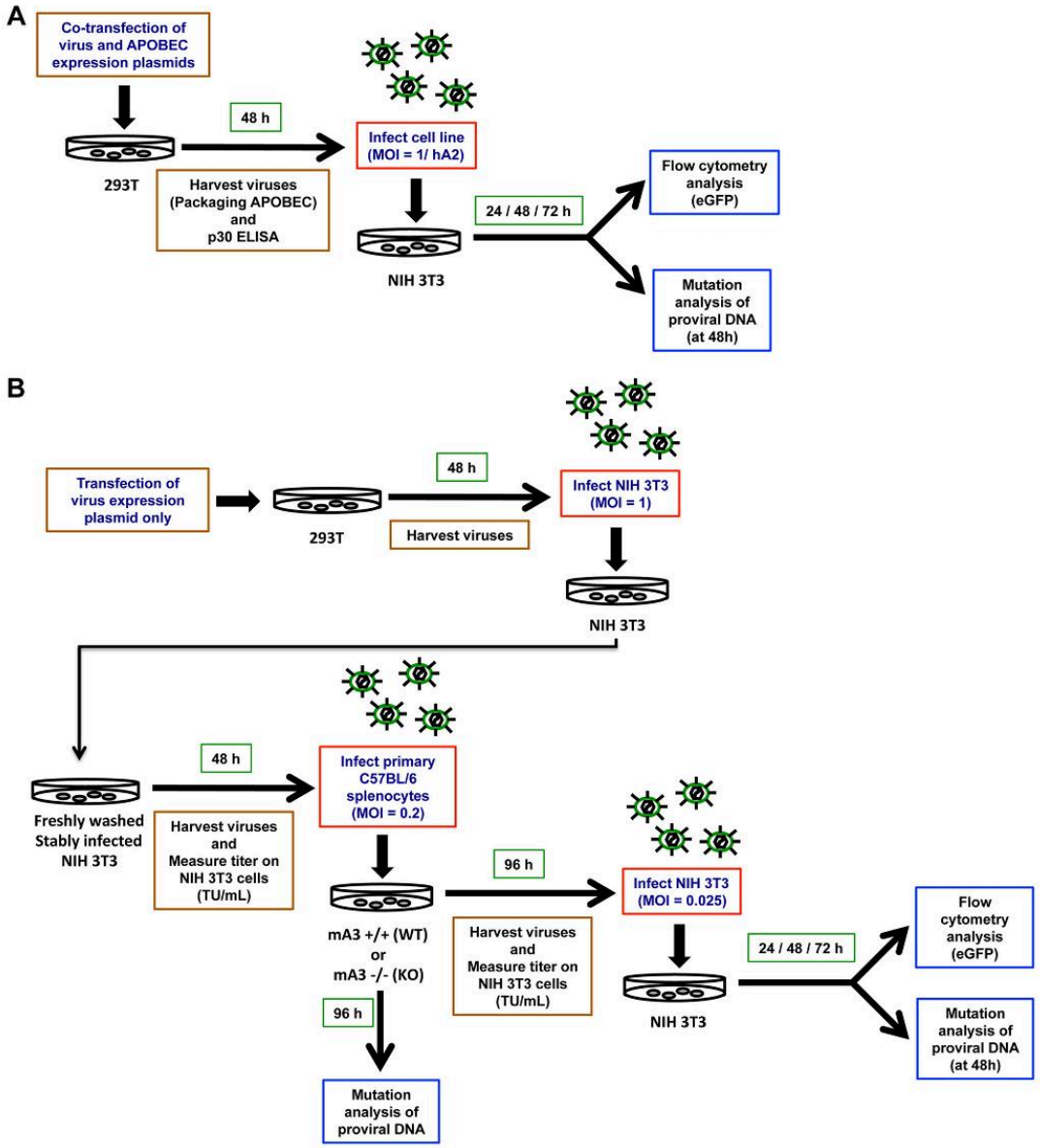


FIG 1 Flow charts describing the infection assays. (A) *In vitro* viral spreading assay. Viruses were produced by plasmid cotransfection in 293T cells and were harvested after 48 h. NIH 3T3 cells were infected with an MOI of 1 in respect to the amount of control viruses produced in the presence of hA2. Infection was measured at 24-h intervals for 72 h. (B) *Ex vivo* viral spreading assay in primary murine splenocytes. Stably infected NIH 3T3 cells were first generated. Viruses released from these cells were then used to infect primary murine splenocytes. Splenocytes from WT and mA3$^{-/-}$ (KO) C57BL/6 mice were infected with an MOI of 0.2. Cells were washed at 24 h after infection, and viruses were harvested 96 h later and used to infect NIH 3T3 cells with an MOI of 0.025. Infection was monitored every 24 h for 72 h.

were grown for 24, 48, and 72 h before being harvested and split into two fractions: one for enhanced green fluorescent protein (eGFP) reporter gene expression using a CyAn ADP flow cytometer (Beckman Coulter) and the other for hypermutation analysis by high-resolution melt (HyperHRM), three-dimensional PCR (3D-PCR), and direct DNA sequencing.

(ii) *Ex vivo* assay. Viruses harvested from stably infected NIH 3T3 cells were used as stock for the *ex vivo* assay (Fig. 1B). At 1 h prior to

TABLE 1 Mutation analysis of proviral DNA by direct sequencing[a]

Virus	3D-PCR band sequenced	No. of clones analyzed	No. of clones with G-to-A mutations	Total no. of mutations[b]		% mutations in a 5′-**Y**C context
				G to A	Other	
Virus produced with transfected mA3						
M-MLV	2	5	1	1	1	100
M-MLV (N113Q)	3	6	6	31	2	71
M-MLV (N505Q)	3	5	5	15	1	75
M-MLV (N113Q/N505Q)	4	5	5	41	0	59
M-MLV (CTA)	4	5	4	26	2	69
Hybrid 1	2	5	1	1	3	100
Hybrid 2	3	5	5	9	4	67
Hybrid 3	3	5	4	12	2	75
AKV	3	5	5	30	0	67
AKV (N113D)	4	5	5	34	0	77
AKV (S507N)	2	8	3	3	2	67
AKV (CTA)	4	5	5	27	0	70
Virus from Spl.						
M-MLV	3	10	7	7	2	29
M-MLV (N113Q/N505Q)	4	8	8	59	3	98
AKV	4	11	10	27	4	74
AKV (S507N)	3	6	2	2	2	50
Virus from 3T3 (Spl. derived)						
M-MLV	3	8	0	0	4	0
M-MLV (N113Q/N505Q)	4	7	6	18	3	78
AKV	4	25	23	62	9	84
AKV (S507N)	3	12	6	6	7	84

[a] The clones contain the first 279 bp of the eGFP reporter gene. Viruses were produced in the presence of transfected mA3 or endogenous mA3 in murine splenocytes. Clones displaying identical deamination profiles were excluded from the calculations. Spl., proviral sequences from infected wild-type C57BL/6 mouse splenocytes; 3T3, proviral sequences from NIH 3T3 cells infected with viruses released from splenocytes.

[b] Mutations compiled on the coding strand of the proviral DNA.

infection, activated splenocytes were counted and seeded in activation medium (33). Splenocytes were spin infected at an MOI of 0.2. At 24 h postinfection, infected cells were washed with 1× phosphate-buffered saline, pH 7.4, resuspended in complete RPMI with 10 μg of lipopolysaccharide, and incubated for 96 h for virus production. Virus-containing supernatants were passed through 0.45-μm-pore-size cartridge filters, and viral titers (in number of transducing units [TU]/ml) were determined on NIH 3T3 cells by limiting dilution using flow cytometry. The secondary infections (viral spreading assays) were performed on NIH 3T3 cells, using an MOI of 0.025. Cells were analyzed by flow cytometry for eGFP expression at 24, 48, and 72 h postinfection. Genomic DNA (gDNA) was extracted from infected splenocytes at 96 h postwash and from the NIH 3T3 cells at each time point.

Viral APOBEC packaging assays. Virus-containing supernatants were filtered and then purified by ultracentrifugation through a sucrose cushion as previously described (22). Viral lysates were then processed for immunoblot analysis.

Western blotting. Details on sample preparation for Western blotting can be found elsewhere (22). Blots were probed with anti-Flag (Sigma-Aldrich), anti-β-tubulin (Abcam), and anti-p30 (clone R187; ATCC). Detection of M-MLV-expressed gPr80 was performed using an anti-$p30^{CA}$ antibody kindly provided by Hung Fan (University of California, Irvine). Detection of recombinant gPr80-V5 was performed using an anti-V5 polyclonal antibody (catalog number V8137; Sigma).

3D-PCR analysis. 3D-PCR was performed in a two-step protocol using PrimeSTAR high-fidelity polymerase (TaKaRa). A first-round PCR was performed to amplify a 717-bp eGFP amplicon using primers GFP-717 FWD and GFP-717 REV (53). A second-round gradient PCR targeting a 279-bp nested fragment within the eGFP sequence was then performed using primers R279-FWD and GFP-REV. PCR cycles were 94°C for 50 s, followed by 30 cycles of a denaturation gradient from 91.1 to 88.0°C for 50 s, annealing at 56°C for 30 s, and an extension at 72°C for 1 min and a final extension of 72°C for 5 min. Samples that amplified at the lowest denaturation temperature were column purified, cloned using TA cloning (Promega), and sequenced (at the Nanuq Sequencing Facility at McGill University and the Genome Quebec Innovation Center). Deamination intensity graphs were generated by illustrating the number of amplicons that could be detected in a 7-well 3D-PCR gradient set. Confirmation of the presence of a hypermutation in an A3G or mA3 context was done by DNA sequencing (Table 1). Our experimental conditions were optimized so that 3D-PCR performed on integrated proviral DNA of virus alone or virus produced in the presence of catalytically inactive mutant A3 proteins consistently yielded amplification only at the two highest denaturing temperatures of the gradient.

HyperHRM analyses. HyperHRM analyses were carried out as previously published (53). Briefly, proviral DNA was amplified from the gDNA of virus-infected cells using primers R648-FWD and GFP-REV to generate a 648-bp amplicon. Amplicons derived from each infection condition were then cloned, and colonies positive for an insert were diluted in 250 μl of water, of which 8 μl was used for high-resolution melting analysis in a 96-well plate format. To generate standard curves for mutation quantification purposes, bacterial clones containing defined numbers of G-to-A mutations were used. Amplification by quantitative PCR was then carried out and immediately followed by a melting curve analysis in which DNA amplicons were gradually heated from 72°C to 95°C and fluorescence values were acquired at 0.025°C intervals. Determination of the number of mutations in each clone was achieved by applying the algorithm previously described (53). Unmutated clones were excluded from the calcula-

TABLE 2 HyperHRM analysis of editing of M-MLV, AKV, and glycosylation mutants by mA3

Virus	No. of clones analyzed	No. of mutated clones	% clones mutated	Total no. of mutations	Predicted G-to-A mutation frequency (no. of mutations/kb)
M-MLV	80	3	4	9	0.17
M-MLV (N505Q)	66	29	44	117	2.74
M-MLV (N113Q/N505Q)	71	30	42	145	3.15
AKV	50	20	40	100	3.09
AKV (N113D)	43	32	74	232	8.33
AKV (S507N)	37	2	5	5	0.16

tion of mutation frequency presented in Tables 1 and 2. DNA sequencing of randomly selected proviral clones was performed as a quality control and to ensure that the results were not biased by clonal amplification of a unique sequence (data not shown).

RESULTS

Murine gammaretroviruses display different sensitivities to restriction and deamination by APOBEC3. M-MLV and AKV are two murine gammaretroviruses with very high sequence identity but with strikingly different sensitivities to restriction and deamination by mA3. Here we performed infection assays with the two viruses being produced in cells expressing human APOBEC3G (hA3G), its catalytically inactive mutant (hA3G [E259Q]), mA3, and its catalytically inactive mutant (mA3 [E73A]). hA2 was used as a negative control because it neither deaminates nor restricts these viruses (22). Viruses were produced by cotransfection in 293T cells, harvested, and normalized for p30 content by ELISA (Fig. 1A). The levels of cell expression and efficiencies of virion packaging of all APOBEC proteins were monitored and comparable (Fig. 2A). Target NIH 3T3 cells, which do not express detectable levels of mA3, were infected at an MOI of 1 in respect to the amount of control viruses produced in the presence of hA2, and infection was assessed by measuring eGFP reporter gene expression over 72 h (Fig. 2B). By using NIH 3T3 cells as targets, viruses that survive the initial restriction by the virion-packaged A3 proteins are then free to replicate and spread over the course of the experiment. Our results showed that all A3 proteins tested, including the catalytically inactive mutants, reduced the level of infection of both viruses at the 24-h time point compared to the level of infection achieved with hA2 (Fig. 2B). This is the result of the well-documented deamination-independent restriction that is especially prominent in tissue culture assays. After 48 h postinfection and likely after a second cycle of viral infection, both viruses remained completely restricted by hA3G but continued to rapidly infect new cells when produced with catalytically inactive hA3G (E259Q) and mA3 (E73A). In contrast, mA3 delayed the ability of AKV to spread, as can be observed by a downward inflection in the growth curve (Fig. 2B) and a nearly 45% reduction in the level of infection compared to that achieved for M-MLV at 48 h (Fig. 2C).

We next looked at the intensity of cytidine deamination in integrated provirus DNA. Here we used 3D-PCR, which is a method used to selectively amplify hypermutated DNA sequences on the basis of the premise of reduced PCR amplicon melting temperatures as a consequence of A/T base enrichment due to the deamination of cytosines into uracils (54). 3D-PCR, however, does not provide information on the proportion of sequences mutated in a population. It is a method that makes a qualitative assessment of how intensely proviral sequences are hypermutated. We optimized our assays so that unmutated proviral DNA (virus alone) amplified only at the two highest melting temperatures selected from a gradient of seven possible temperatures (Fig. 2D). Positive amplification at lower melting temperatures is indicative of hypermutated proviral DNA sequences. Under our conditions, the threshold for reliable mutation detection by 3D-PCR is 3 or more G-to-A/C-to-T mutations per 279-bp sequence (53). For amplicons indicative of hypermutation, the presence of transition mutations in an A3 deamination context (5′-C**C** or 5′-T**C**, where the boldface C represents the deaminated C residue) was confirmed by DNA sequencing (Table 1). 3D-PCR assays were repeated at least 3 times from independent infection assays, and representative gel images are presented. Our results show that hA3G deaminates both M-MLV and AKV at the same intensity, meaning that the most intensely mutated sequences have similar numbers of G-to-A mutations (Fig. 2D). Catalytically inactive A3G (E259Q) was used as a negative control, and virus produced with A3G (E259Q) displayed amplification only under the first two melting conditions, similar to the findings for virus produced alone or with hA2 (data not shown). On the other hand, mA3 weakly hypermutated AKV, as judged by the presence of amplification in the 3rd lane in Fig. 2D, and did not at all hypermutate M-MLV.

Resistance to deamination maps to the gPr80 accessory protein. In order to identify the regions of AKV that are responsible for rendering the virus sensitive to deamination by mA3, we generated three hybrid viruses by progressively replacing the proviral DNA of M-MLV with that of AKV (Fig. 3A). We then performed an infection assay under the same conditions described above. We found that A3 protein encapsidation was similar for all hybrid viruses (data not shown). Viral infection assays showed that hybrid 1, containing an AKV segment spanning from the R region to the N terminus of Gag, spread similarly to wild-type M-MLV (Fig. 3B). Hybrids 2 and 3, however, displayed a delay in spreading at the 48-h time point, similar to that seen for AKV, with, respectively, 18% and 24% reductions in the levels of infection relative to that for M-MLV being seen (Fig. 3C). Catalytically inactive mutants A3G (E259Q) and mA3 (E73A) did not have a significant effect on hybrid virus spread (Fig. 3B and C). 3D-PCR analysis performed on infected target cells revealed that hA3G hypermutated all three hybrids with the same intensity, while mA3 was able to hypermutate only hybrids 2 and 3 to levels similar to those for AKV (Fig. 3D).

The common proviral DNA segment of hybrids 2 and 3 maps to the *gag* gene of AKV, which includes sequences coding for both the gPr80 and Pr65 polyproteins. Because the spread curves and deamination intensities were similar between the two hybrids, this

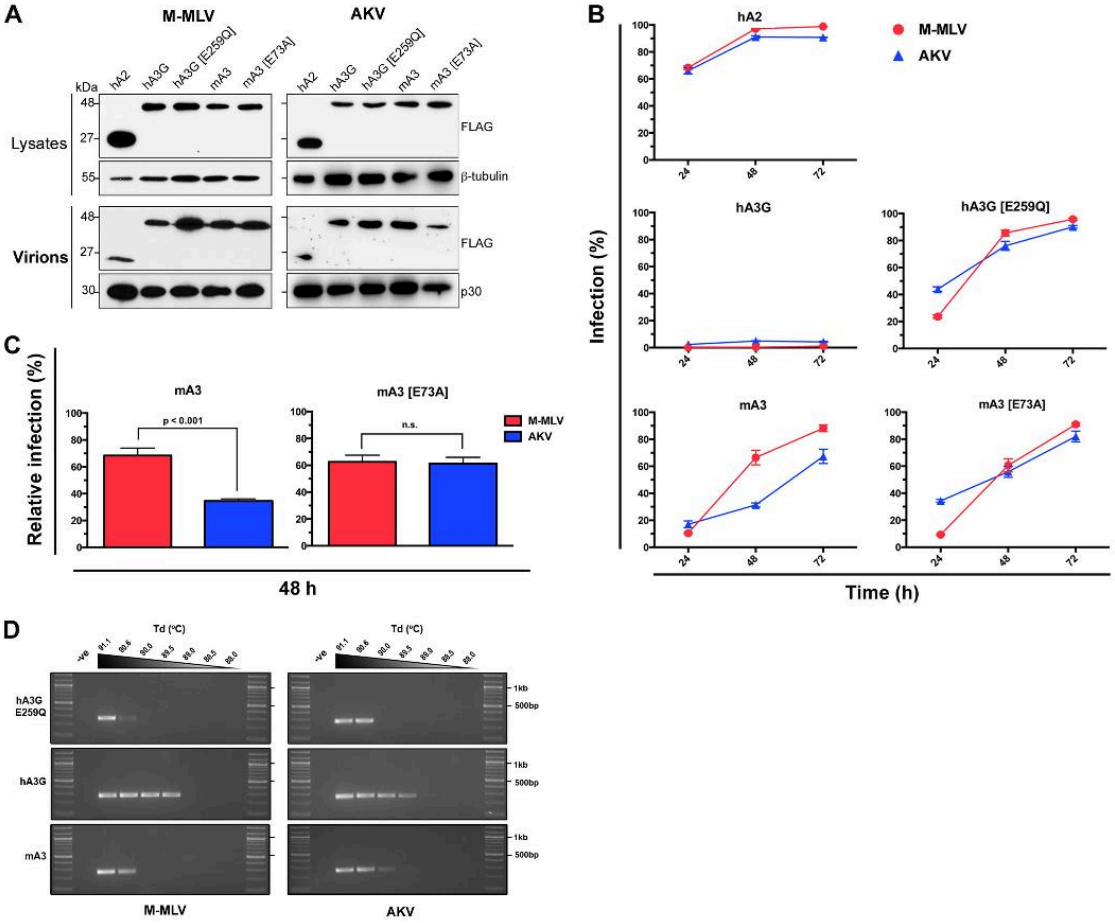


FIG 2 Sensitivity of M-MLV and AKV to deamination and restriction by A3 proteins. (A) (Top) Expression of the various APOBEC proteins in lysates of cotransfected 293T cells used to produce M-MLV and AKV was analyzed; (bottom) virions harvested from the culture supernatant of 293T cells were analyzed for the efficiency of packaging of the various APOBEC proteins. (B) Analysis of the spread of M-MLV and AKV infection produced in the presence of APOBEC proteins, represented as the percentage of cells expressing the eGFP reporter at the various time points after infection. Infection results are from at least three independent transfections with triplicate infection values for each. Results are presented as the mean level of infection ± standard deviation. (C) Comparison of the relative infection of M-MLV and AKV produced in the presence of mA3 or mA3 (E73A) at 48 h postinfection. Values were normalized to those for infection with virus produced in the presence of hA2 at 48 h. Error bars represent standard deviations. Statistical significance was determined using a two-tailed unpaired Student *t* test; *P* values are indicated on the graph. n.s., not significant. (D) 3D-PCR analysis performed on genomic DNA extracted from infected NIH 3T3 cells from the 48-h time point for M-MLV and AKV. Representative gels from 3 independent assays are shown. Td, denaturing temperature; −ve, negative.

led us to conclude that all determinants responsible for the contrasting phenotypes between M-MLV and AKV are located within this region. Two recent studies have highlighted the involvement of the gPr80 Gag protein in resisting restriction by mA3 (15, 18). To distinguish if resistance to deamination was conferred by elements within the Pr65 or gPr80 polyproteins, we generated viral mutants, termed CTA mutants, that do not express gPr80 because their CTG initiation codon was replaced by a CTA trinucleotide (Fig. 4). Packaging of A3 proteins into these mutant viruses was similar to that for their WT counterparts (Fig. 4A). Restriction assays and growth curves of both viruses showed that they had similar profiles that closely resembled the profile for AKV (Fig. 4B and C). However, there was a small but noticeable decrease in infectivity of these mutants compared to that of the WT parental viruses. M-MLV and AKV produced with A2 infected 65 to 70% of the cell population after 24 h (Fig. 2B). With the CTA mutant viruses, this infection was reduced to 40% when the same amount of input virus (as measured by p30 ELISA) was used (Fig. 4B). Although both the CTA and CTG codons code for leucine, the position of this trinucleotide is located in the stem of one of the major stem-loops involved in packaging the viral RNA dimer, which could explain this decrease in infectivity (55). Hypermutation analysis of the M-MLV CTA and AKV CTA mutants revealed an increase in sensitivity to hypermutation by both mA3 and hA3G (Fig. 4D). This therefore allowed us to

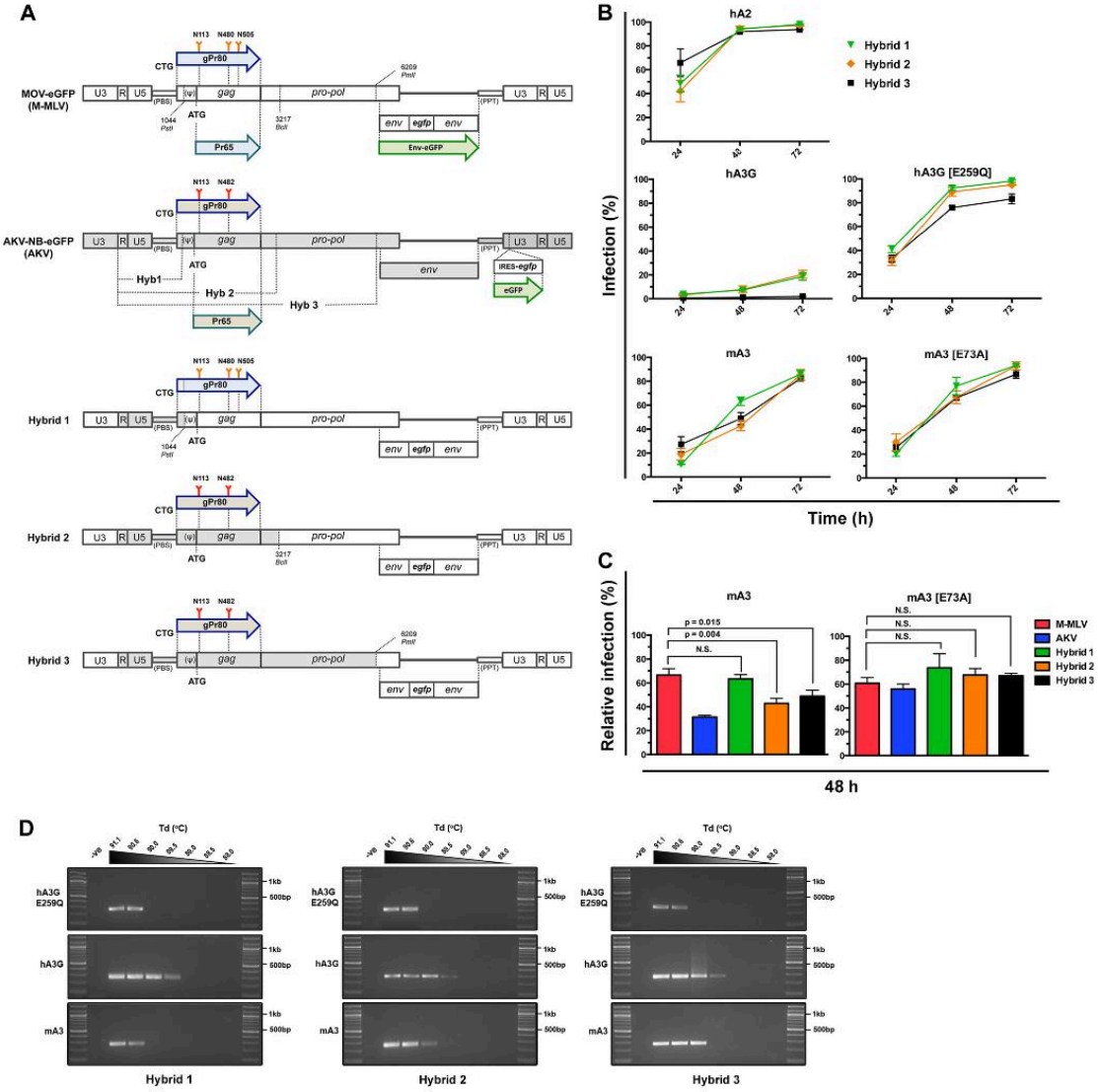


FIG 3 M-MLV–AKV hybrid viruses reveal that deamination resistance maps to the *gag* gene. (A) Schematic illustrations of the M-MLV, AKV, and hybrid proviruses. The segments used to construct the three hybrid viruses are indicated. Hybrid viruses are composed of M-MLV into which orthologous segments of AKV have been inserted. Hyb, hybrid; IRES, internal ribosome entry site; PBS, primer binding site; PPT, polypurine tract. (B) Analysis of spread of hybrid virus infection produced in the presence of the various A3 proteins. Infection results are from at least three independent transfections with triplicate infection values for each. Results are presented as the mean ± standard deviation. (C) Comparison of the level of infection with the hybrid viruses relative to that with M-MLV and AKV produced in the presence of mA3 or mA3 (E73A) at 48 h postinfection. Values were normalized individually to the value for infection with virus produced in the presence of hA2 at 48 h. Error bars represent standard deviations. Statistical significance was determined using a two-tailed unpaired Student *t* test; *P* values are indicated on the graph. N.S., not significant. (D) 3D-PCR analysis performed on genomic DNA extracted from infected NIH 3T3 cells from the 48-h time point. Representative gels from 3 independent assays are shown.

exclude the possibility of any potential roles for elements within Pr65 to be the cause for the resistance to deamination and to focus our attention on gPr80.

Three *N*-linked glycosylated sites in M-MLV gPr80 are required for complete resistance to deamination. Because the amino acid sequences of the gPr80 proteins of M-MLV and AKV are highly similar, we focused our attention on differences in the *N*-linked glycosylation patterns of the proteins (Fig. 3). We carried out an *in silico* analysis of the gPr80 proteins of both viruses using the NetNGlyc server to identify putative glycosylated sites

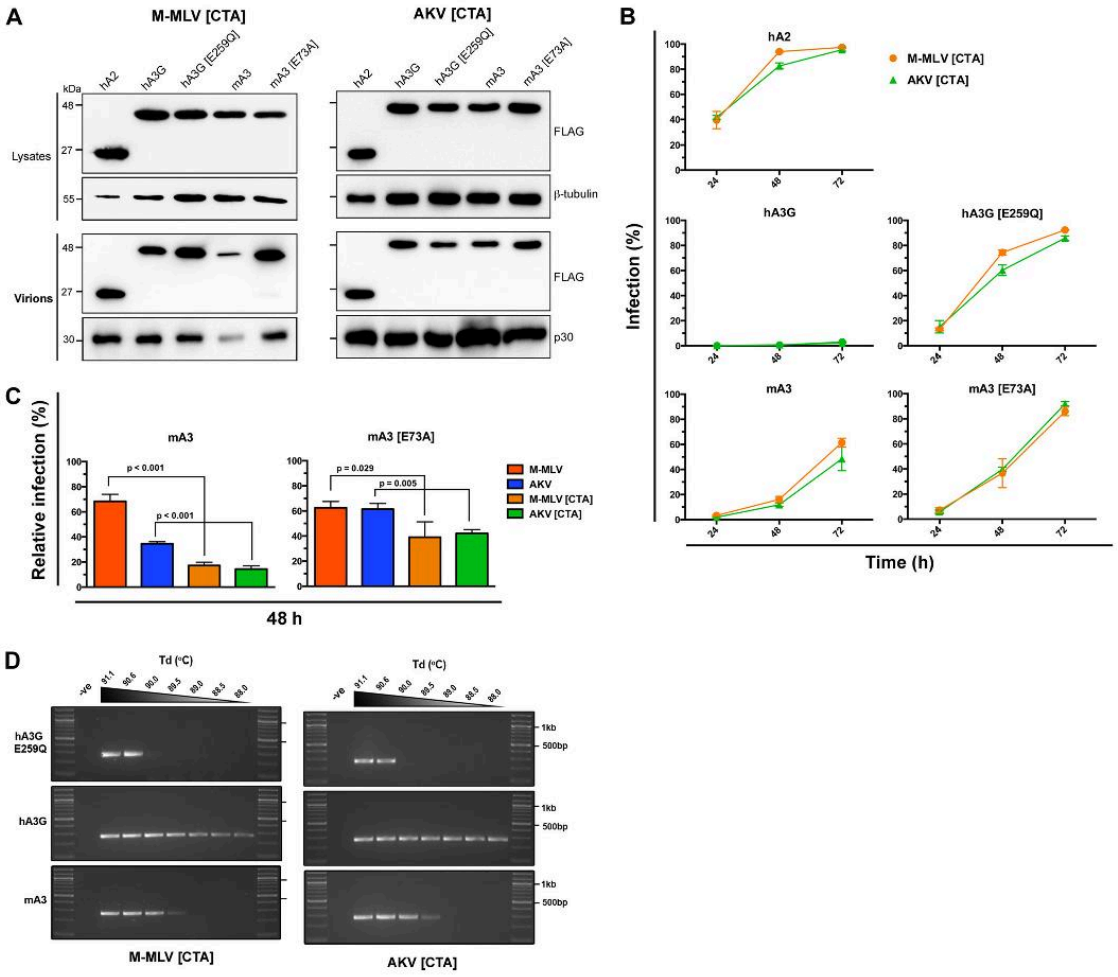


FIG 4 Analysis of M-MLV and AKV mutants that do not express gPr80. (A) (Top) Western blots showing the expression of the various APOBEC proteins in lysates of cotransfected 293T cells used to produce M-MLV (CTA) and AKV (CTA) mutant viruses are shown; (bottom) viruses (virions) harvested from the culture supernatant of 293T cells were analyzed for the efficiency of packaging of the various APOBEC proteins. (B) Analysis of the spread of infection of mutant viruses produced in the presence of the various A3 proteins. Infection results are from at least three independent transfections with triplicate infection values for each. Results are presented as the mean ± standard deviation. (C) Comparison of the level of infection with the CTA mutant viruses relative to that with M-MLV and AKV produced in the presence of mA3 or mA3 (E73A) at 48 h postinfection. Values were normalized individually to the values for infection with virus produced in the presence of hA2 at 48 h. Error bars represent standard deviations. Statistical significance was determined using a two-tailed unpaired Student *t* test; *P* values are indicated on the graph. (D) 3D-PCR analysis performed on genomic DNA extracted from infected NIH 3T3 cells from the 48-h time point for the various viruses. Representative gels from 3 independent assays are shown.

within an Asn-XAA-Ser/Thr (where XAA is not Pro) sequon. We identified three likely glycosylated sites in the gPr80 of M-MLV (N113, N480, and N505) but only two in that of AKV (N113 and N482). N113 is located in the matrix (MA), while N480/N482 and N505 are located in the capsid (CA).

Using site-directed mutagenesis, we generated point mutants for each of the predicted glycosylated amino acids of gPr80 of M-MLV and AKV. We also introduced an additional *N*-linked glycosylated site at position S507 of AKV (AKV [S507N]), which is in a sequon favorable for glycosylation. The asparagine-to-aspartic acid substitutions chosen remove the glycosylated site, while they introduce a chemically conservative mutation. It should also be noted that changes in the gPr80 sequence downstream of amino acid 88 are also reflected in the Pr65 Gag protein. We also generated double and triple glycosylation mutants when appropriate. Western blot analyses of transfected cell extracts clearly show that the apparent molecular mass of gPr80 shifts according to the number of putative glycosylated sites mutated (Fig. 5A to C). Glycosidase treatment of the extracts reveals a gPr80 band lower than that of the gPr80 band of the M-MLV (N113Q/

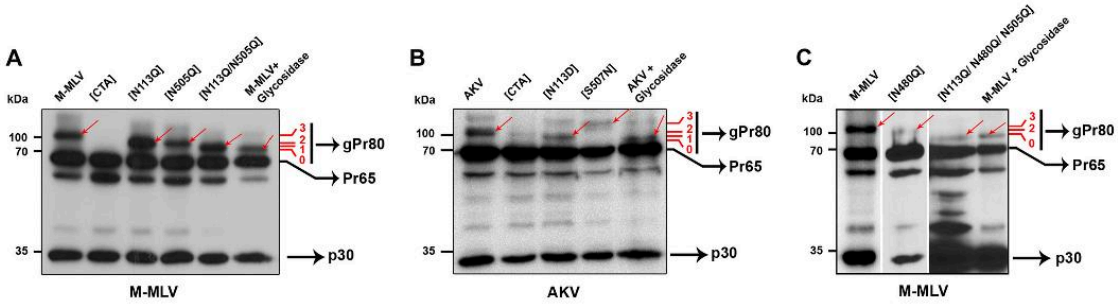


FIG 5 Identification of *N*-linked glycosylated sites in M-MLV and AKV. (A to C) Western blot analysis of transfected 293T cell extracts showing the altered migration patterns caused by *N*-linked glycosylation. The last lane of each blot shows either M-MLV- or AKV-infected cell extracts treated with glycosidase. The gPr80 protein band for each virus is indicated by red arrows. Numbers in red indicate the number of glycosylated residues.

N505Q) double mutant, which indicates that amino acid N480 is also likely glycosylated, as predicted (Fig. 5A and B). The M-MLV (N113Q/N480Q/N505Q) triple glycosylation mutant demonstrated a band for gPr80 at the same height as that in the glycosidase-treated sample (Fig. 5C). Western blot analysis of AKV gPr80 showed that the S507N mutation increases the size of the band, suggesting that *N*-linked glycosylation at this site was indeed restored (Fig. 5B).

Restriction assays with the M-MLV single point mutants and the N113Q/N505Q double mutant clearly showed a delay in infection spread at 48 h with mA3 (Fig. 6A). The N113Q/N505Q double mutant was about twice as sensitive to mA3 restriction as the single point mutants (Fig. 6C). Mutant M-MLVs produced with mA3 (E73A) showed little or no difference in infectivity compared to that of the wild-type virus (Fig. 6C). Restoring glycosylation at position S507 of AKV increased the spreading kinetics to resemble that of mA3-resistant M-MLV (Fig. 6B); however, the N113D substitution did not render the virus more sensitive to mA3 restriction (Fig. 6D). Viral mutants M-MLV (N480Q) and AKV (N482D) displayed very poor infectivity (less than 2%), despite being efficiently released from the cells, as judged by p30 ELISA (data not shown). For this reason, these mutants were not analyzed further.

Although the sequential mutation of glycosylated residues in gPr80 had an overall modest effect on the sensitivity to restriction by mA3, a more pronounced impact on hypermutation could clearly be observed. Both mA3 and hA3G mutated the M-MLV and AKV mutants with an intensity higher than that for their WT counterparts (Fig. 6E and F). All mutants with N-to-Q point mutations of M-MLV gPr80 became more sensitive to deamination; this effect was slightly increased in the M-MLV (N113Q/N505Q) double mutant and the AKV (N113D) mutant. AKV (S507N), which had three glycosylated sites, became resistant to deamination (Fig. 6F).

To further characterize the intensity and frequency of deamination in proviral sequences, we performed mutation analyses on individual clones isolated from infected cells. We used HyperHRM analysis, which is a high-throughput method that we developed to quantify the number of A3-induced mutations in a PCR amplicon (53). Confirming the 3D-PCR data, we found that there was an inverse correlation between the number of glycosylated residues in gPr80 and the proportion of hypermutated sequences in a specific virus (Fig. 7A and B and Table 2). Our data also show that the intensity of hypermutation increases when fewer glycosylated sites are present in either virus.

gPr80 antagonizes both arms of mA3 restriction. Having identified M-MLV mutants that are sensitive to deamination, we next wanted to evaluate how gPr80 glycosylation affects deamination-independent restriction by mA3. Here we normalized the infection data at the 48-h time point to those for the hA2 control independently for each virus set: M-MLV (Fig. 7C) and AKV (Fig. 7D). We found that viruses were more restricted by catalytically inactive mA3 (E73A) when the gPr80 protein had fewer glycosylated sites or was not expressed altogether. The increased sensitivity to restriction of the CTA mutant viruses by mA3 (E73A) could reflect the protective effect of N480 glycosylation. Altogether, these results clearly show that *N*-linked glycosylation of gPr80 prevents restriction by deamination and also by deamination-independent mechanisms.

Endogenous mA3 inhibits MLV infection but is also a source of genetic diversification. Two reports have already established that the antiretroviral activity of mA3 is inhibited by gPr80 *in vivo* (15, 18). These studies were carried out by comparing the infectivity of deamination-resistant MLV strains with coisogenic strains unable to either express or produce a full-length gPr80 protein. Here we asked whether endogenous levels of mA3 are sufficient to hypermutate deamination-sensitive MLVs to affect viral fitness and the replication of released viruses. We performed our experiments only with MLVs that expressed gPr80 in order to measure the direct contribution of *N*-linked glycosylation to the resistance.

We isolated splenocytes from neonatal C57BL/6 mice with WT mA3 or mA3-deficient (mA3-KO) neonatal C57BL/6 mice. The splenocytes were cultured, activated, and then infected with a similar MOI with viruses that were produced in the absence of mA3 in NIH 3T3 cells (Fig. 1B). The infectious titer of the viruses released by the splenocytes, measured as the number of transducing units per ml of supernatant harvested, was assessed 96 h later (Table 3). M-MLV was generally more efficient at replicating in these splenocytes than AKV. Differences in titers were also presented as relative infectivity to better illustrate that endogenous mA3 had a similarly potent effect at reducing the infectious titers of all viruses, including those resistant to deamination (Fig. 8A). This

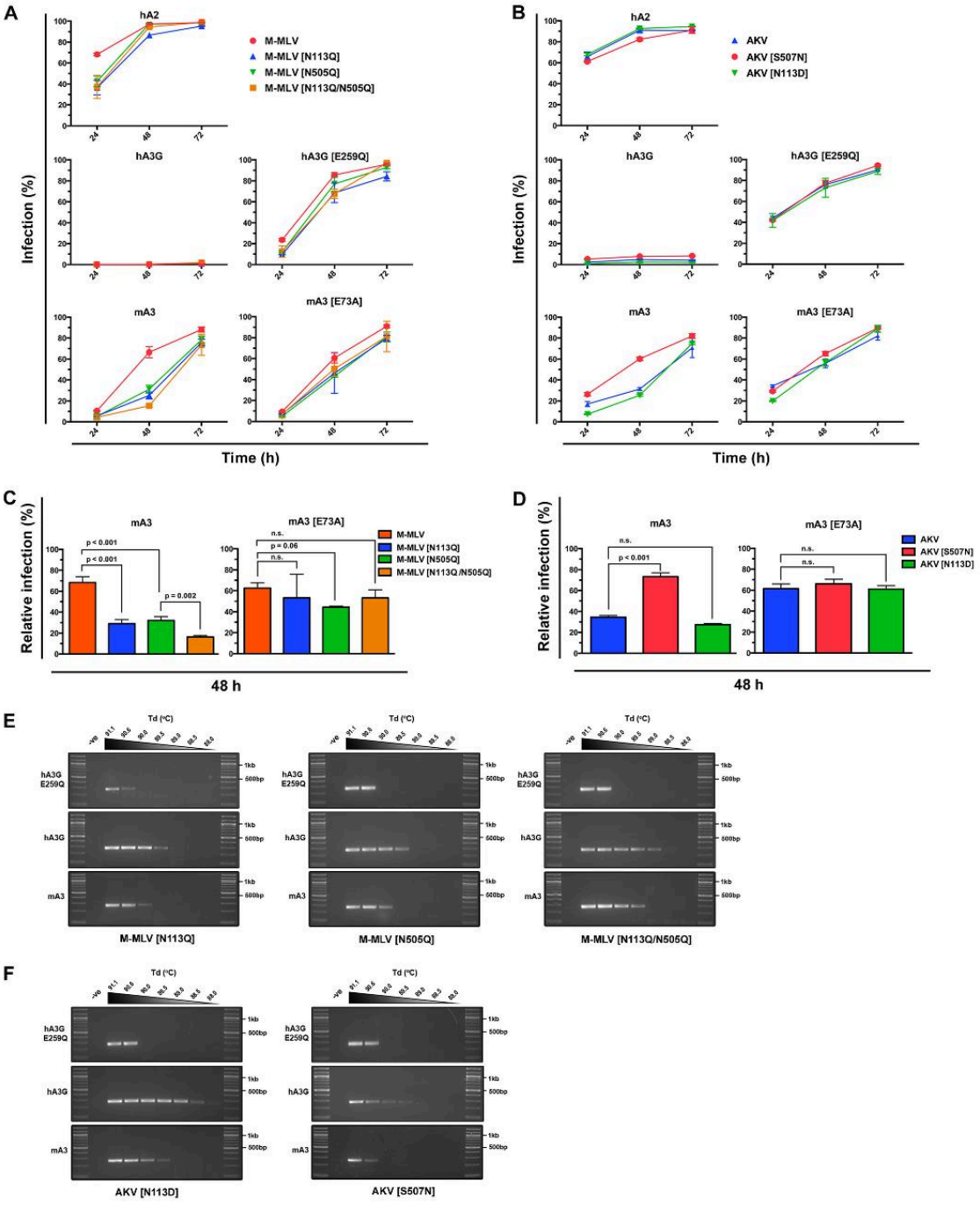


FIG 6 Intensity of gPr80 glycosylation correlates with sensitivity to deamination by mA3. (A and B) Analysis of spread of infection of glycosylation mutants of M-MLV (A) and AKV (B) coproduced with various A3 proteins. (C and D) Comparison of the infection with the mutant viruses relative to that with M-MLV (C) and AKV (D) produced in the presence of mA3 or mA3 (E73A) at 48 h postinfection. Values were normalized individually to those for infection with virus produced in the presence of hA2 at 48 h. Error bars represent standard deviations. Statistical significance was determined using a two-tailed unpaired Student *t* test; *P* values are indicated on the graphs. n.s., not significant. (E and F) 3D-PCR analysis performed on genomic DNA extracted from infected NIH 3T3 cells from the 48-h time point for glycosylation mutants of M-MLV (E) and AKV (F).

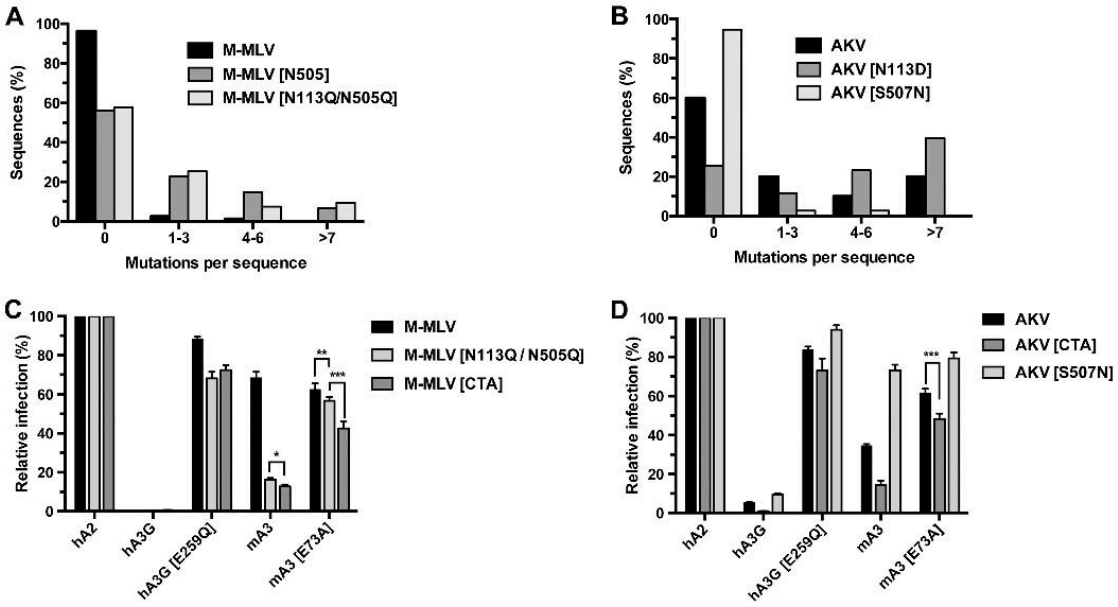


FIG 7 *N*-linked glycosylation inhibits deamination and deamination-independent restriction. (A and B) Analysis of mutation intensities in proviral DNA by HyperHRM analysis. The histograms depict the proportion of total sequences containing the indicated number of mutations. The results of clone analysis are presented in Table 2. (C and D) Relative infection depicting the effect of gPr80 mutants on virus sensitivity to restriction at 48 h postinfection. Infection levels from experiments whose results are presented in Fig. 2B, 4B, and 6A and B were normalized to those for hA2 for each virus. Values presented are the means ± SEMs from three independent transfection experiments with triplicate infection samples. The Student unpaired *t* test was performed to assess statistical significance: *, $P = 0.05$; **, $P = 0.19$; ***, $P = 0.02$.

indicates that deamination by endogenous mA3 has little effect on viral restriction under these conditions.

To ensure that hypermutation of proviral DNA did occur in the WT splenocytes, 3D-PCR was carried out on proviral DNA from splenocytes at 96 h postinfection (Fig. 8B). Hypermutated viral sequences were detected only in WT splenocytes at various intensities for all viruses. Surprisingly, M-MLV and AKV (S507N) also showed evidence of hypermutation caused by mA3, but the intensity and frequency of the mutations were very low (Table 1). These data were also supported by HyperHRM analysis of individual clones, showing that they contained several transition mutations (Fig. 8C and Table 4).

TABLE 3 Titers of gammaretroviruses released from splenocytes[a]

Virus	Source	Virus titer (no. of TU/ml [10^5])[a]
M-MLV	WT	0.047 ± 0.006
	KO	5.099 ± 1.964
M-MLV (N113Q/N505Q)	WT	0.037 ± 0.007
	KO	3.896 ± 1.740
AKV	WT	0.025 ± 0.005
	KO	1.796 ± 0.615
AKV (S507N)	WT	0.012 ± 0.002
	KO	0.302 ± 0.053

[a] TU, transducing units; WT, wild type C57BL/6 mouse splenocytes; KO, mA3$^{-/-}$ C57BL/6 mouse splenocytes. Values represent the means of triplicate infection values ± standard deviation (for KO mice, $n = 5$; for WT mice, $n = 6$).

To address whether the replicative fitness of MLVs sensitive to deamination was affected by being produced in splenocytes expressing mA3, we then used these viruses to infect NIH 3T3 cells with an MOI of 0.025. We monitored their spread in culture every 24 h for 72 h. These assays revealed that the pool of infectious viruses released from WT splenocytes did not exhibit a proliferation efficiency different from that of viruses released from mA3-KO cells (Fig. 8D). These results were confirmed by statistical analyses indicating that the slopes of the viral spreading curves were similar for viruses produced in WT or KO splenocytes (data not shown). We also analyzed proviral DNA sequences amplified from infected NIH 3T3 cells. We found by 3D-PCR clear evidence of hypermutation in viruses originating from WT splenocytes (Fig. 8E and Table 1). HyperHRM analysis revealed that the intensity and frequency of the transition mutations were, however, lower (Fig. 8F and Table 4). The mutation frequency for AKV dropped from 2 mutations per kb in splenocytes to 1.19 mutations per kb in NIH 3T3 cells (Table 4). This indicates that more heavily mutated viruses were selected against and were not represented in NIH 3T3 cells but a significant proportion of sublethally mutated viruses still continued to replicate.

DISCUSSION

The aim of this study was to identify what makes AKV different from M-MLV in respect to its sensitivity to deamination by mA3. Our results have clearly revealed that the number of *N*-linked

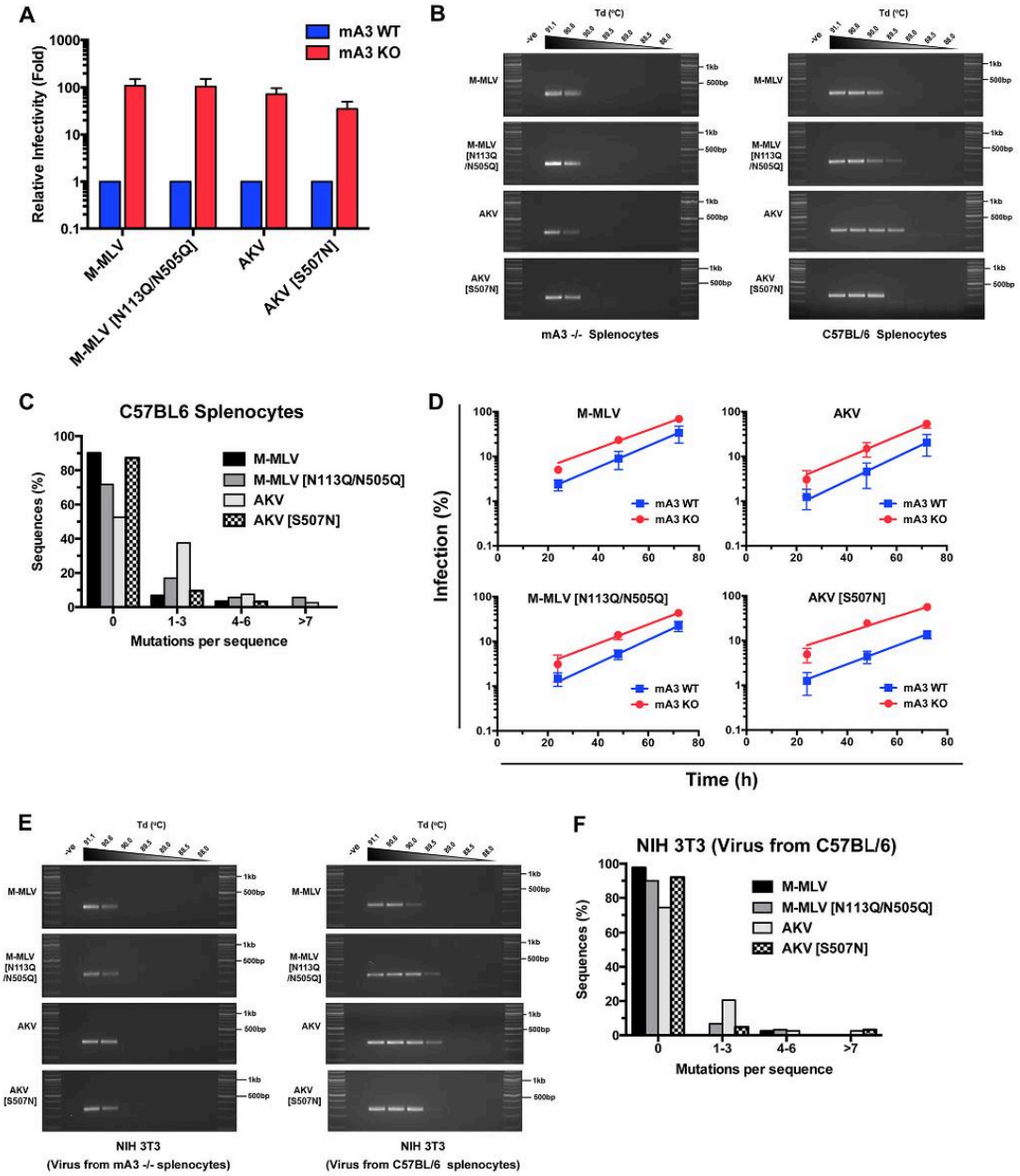


FIG 8 gPr80 glycosylation levels influence the intensity of deamination by endogenous mA3 expressed in mouse splenocytes. (A) Relative infectivity of viruses released from infected splenocytes. The viral titers used to generate the figure are presented in Table 3. (B) 3D-PCR analysis performed on genomic DNA extracted from infected KO (left) or WT (right) C57BL/6 splenocytes. Representative gels from 3 independent assays are shown. (C) Analysis by HyperHRM of mutation intensities in proviral DNA isolated from infected C57BL/6 splenocytes. The histograms depict the proportion of total sequences containing the indicated number of mutations. The results of clone analysis are presented in Table 2. (D) Infection spread in NIH 3T3 cells. Cells were infected with an MOI of 0.2 with viruses released from the splenocytes for which the results are depicted in panel A. Infection was measured every 24 h for 72 h. The graphs represent nonlinear regression curves of the infection. Results are presented as the mean level of infection ± standard deviation. (E) 3D-PCR analysis performed on genomic DNA extracted from infected NIH 3T3 cells. Representative gels from 3 independent assays are shown. (F) Analysis by HyperHRM of mutation intensities in proviral DNA isolated from infected NIH 3T3 cells.

TABLE 4 HyperHRM analysis of editing of M-MLV, AKV, and glycosylation mutants by endogenous mA3 expressed in C57BL/6 splenocytes[a]

Infected cell and virus	No. of clones analyzed	No. of mutated clones	% clones mutated	Total no. of mutations	Predicted G-to-A mutation frequency (no. of mutations/kb)
WT splenocytes					
M-MLV	81	12	15	38	0.73
M-MLV (N113Q/N505Q)	53	15	28	59	1.72
AKV	40	19	48	52	2.00
AKV (S507N)	63	8	13	21	0.51
NIH 3T3 cells					
M-MLV	42	1	3	4	0.15
M-MLV (N113Q/N505Q)	60	6	10	16	0.41
AKV	39	10	26	30	1.19
AKV (S507N)	63	5	8	21	0.51

[a] HyperHRM analysis was performed on gDNA from infected WT splenocytes 96 h after infection. NIH 3T3 cells were infected with viruses released from WT splenocytes. HyperHRM analysis was performed 48 h later.

glycosylated sites in the gammaretrovirus-specific gPr80 accessory protein inversely correlates with the intensity of mA3-induced deamination. This therefore supports the concept that gammaretroviruses employ the host's protein glycosylation machinery to protect themselves against innate restriction by mA3. Considering that the gPr80 protein of MLVs also reduces the potency of the mutator activity of human A3G, *N*-linked glycosylation may therefore be part of a broader strategy allowing gammaretroviruses to increase the success of zoonotic transmission between hosts of different species.

M-MLV is not the only murine retrovirus that is resistant to deamination by mA3. In fact, most mouse retroviruses, including the prevalent Friend MLV, which has the same number of putative gPr80 glycosylated sites as M-MLV, are resistant (Fig. 9A) (34). The xenotropic MLV-related virus (XMRV) is perhaps the only other currently known exception, along with AKV, of a murine

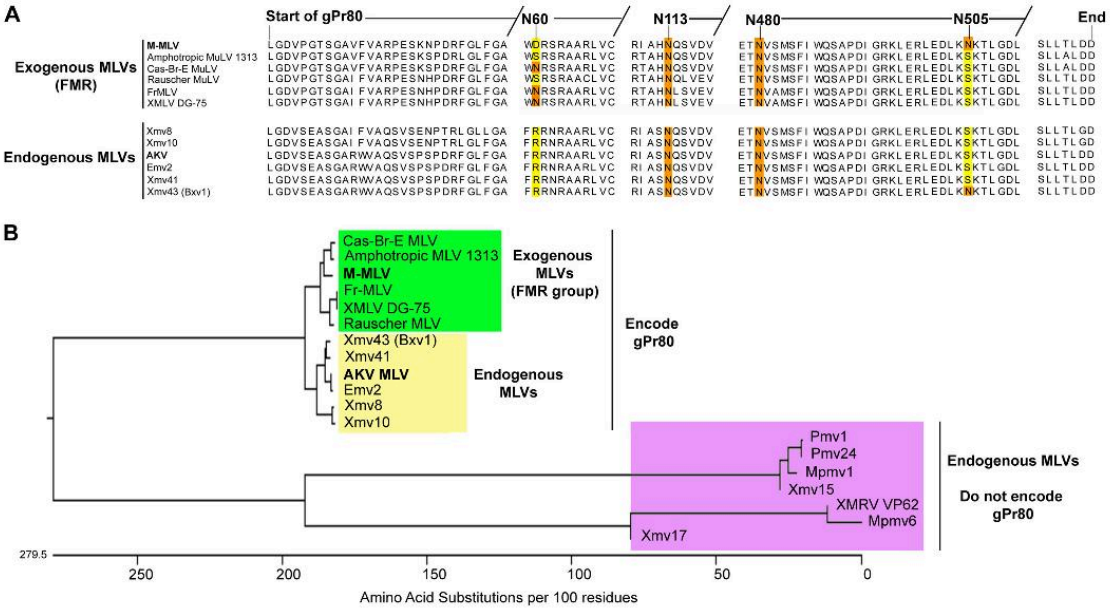


FIG 9 Phylogenetic analysis of the gPr80 amino acid sequence of various MLVs. (A) Detailed alignment of MLVs able to produce gPr80. Residues at positions homologous to putative glycosylated sites for M-MLV are highlighted in orange. Amino acids different from asparagine at these positions are highlighted in yellow. N60 is a putative glycosylated site for several MLVs of the Friend-Moloney-Rauscher (FMR) group. (B) The amino acid sequences of exogenous and endogenous MLVs were aligned by use of the DNAStar Lasergene (version 8) MegAlign program (Clustal W method), and a phylogenetic tree was generated. Green shading, exogenous retroviruses belonging to the Friend-Moloney-Rauscher group; yellow shading, endogenous MLVs that produce a full-length gPr80; purple shading, endogenous MLVs unable to produce gPr80 because of deletions in the sequence. Endogenous MLV subgroups were xenotropic (Xmv), ecotropic (Emv), polytropic (Pmv), and modified polytropic (Mpmv). FrMLV and Fr-MLV, Friend MLV.

retrovirus that is sensitive to deamination by mA3 (56). However, XMRV, contrary to AKV, is not an endogenous murine retrovirus with a long history of coevolution with its host but, rather, a recently emerged laboratory virus that was inadvertently created through provirus (PreXMRV-1 and PreXMRV-2) recombination during human xenograft passages in mice (57). What is further interesting about XMRV is that it has a deletion in the leader sequence of the *pr80* gene, as do some other murine endogenous retroviruses (ERVs), and therefore does not express gPr80 or any alternate form of a glycosylated Gag-like protein (58). The mouse mammary tumor virus (MMTV) is a betaretrovirus that is restricted by mA3 but is resistant to deamination (35). Although the MMTV genome does not encode a glycosylated Gag protein, it does encode a superantigen (SAg) with multiple *N*-linked glycosylated residues that could be packaged into virions and act as a functional surrogate to gPr80 (59). It would also be interesting to investigate whether *N*-linked glycosylation of virion-packaged proteins other than gPr80 can also inhibit A3 deaminase activity.

The exact role of the gPr80 protein of MLVs has long remained enigmatic. Early studies investigated various functional aspects of the gPr80 protein by comparing viral release, infectivity, and virulence in the presence or absence of the protein or by exchanging *gag* gene segments between different MLV strains (40, 49, 50, 58, 60, 61). The general conclusion of these studies was that the absence of a functional and full-length gPr80 decreases virus release and infectivity *in vitro*, but especially *in vivo*. However, recent studies have revealed new and important roles of this accessory protein in helping the virus evade innate immune defenses. Two studies have shown that gPr80 inhibits the antiviral activity of mA3 (15, 18). Stavrou et al. took these findings further by showing that gPr80 acts by preventing mA3 from accessing the reverse transcription complex in viral cores and also by hiding replication intermediates from cytosolic sensors (15). Our work here adds to current knowledge by showing that the *N*-linked glycans attached to gPr80 are, in fact, essential for its activity against host innate defenses. MLVs that express few *N*-linked glycosylated sites in gPr80 are more susceptible to being hypermutated and restricted by mA3.

In contrast to hA3G, which stringently restricts MLV infection and spread in human cells, mA3 does not inactivate mouse retroviruses with the same potency. All MLVs that we tested, including those with fewer glycosylated sites in gPr80, and that were produced in the presence of mA3 spread to nearly all target cells 72 h after infection. Differences in the intensity of the deamination cannot alone explain this striking difference. In a previous study, we reported that hA3G mutated AKV at an average rate of 8 mutations per kb of proviral DNA; here we have shown that AKV (N113D) is also mutated to that level by mA3 (Table 2) (33). It is therefore likely that a factor in addition to deamination intensity, such as the consensus DNA sequence being deaminated, is also required for viral genetic inactivation. While mA3 deaminates cytidines preferentially in a 5′-T**C** context, hA3G prefers 5′-C**C** (26, 33, 62). With the tryptophan codon being TGG (5′-C**C**A when it is read on the minus-strand DNA), hA3G is much better suited than mA3 at generating all of the three termination codons. Mutations caused by mA3 therefore appear to have subtler and less deleterious effects on viral replicative fitness than hA3G-induced mutations, which are almost always lethal for the virus.

An interesting result that emerged from the *ex vivo* assays is that infectious viruses with low levels of mutations have replicative fitness similar to that of nonmutated viruses (Fig. 8D). In a previous study, we showed how AKV released from WT splenocytes expressing endogenous mA3 was restricted when analyzed 48 h after infection in NIH 3T3 cells (33). Here we took these results further by looking at the impact of mutations on virus spread over time. By normalizing input infectious viral titers, we compared the infection efficiency of a pool of mA3-mutated viruses to that of nonmutated viruses. Viruses that infected target NIH 3T3 cells had, as would be expected, fewer mutations than those recovered from the splenocytes, but nevertheless, a relatively large proportion (26% for AKV) had between 1 and 3 mutations per sequence analyzed (Fig. 8C and F). It therefore appears that viruses containing low levels of mutations can be as infectious as unmutated viruses under these experimental conditions. In the context of a natural infection, such mutations could be beneficial and help the virus evade immune defenses.

These results now raise the question as to whether gammaretroviruses that are permissive to low levels of deamination, by way of less intensely glycosylated gPr80, have a greater success at infecting and persisting *in vivo*. In support of this concept, a prior study in which the authors performed a phylogenetic analysis of the nucleotide sequence of various MLV *pr80* genes revealed that clade A, B, and C murine ERVs (xenotropic, polytropic, and modified polytropic viruses) are descendants of the same ancestral progenitor, while viruses of the Friend-Moloney-Rauscher (FMR) group come from a different progenitor (58). If we now look instead at the amino acid sequence of their respective gPr80 proteins, we find that clade B and C endogenous retroviruses (containing AKV) cluster together and that their genomes generally encode a gPr80 with fewer putative glycosylated sites than the gPr80 proteins of the FMR group (Fig. 9A and B). The only exceptions are Rauscher MLV and amphotropic MLV, which contain only two putative glycosylated sites. However, it remains to be experimentally determined whether these viruses are sensitive or resistant to deamination by mA3 and also whether they in fact contain only two glycosylated sites. Additionally, these two viruses have a serine at position 60, whereas Friend MLV has an asparagine in a perfect *N*-glycosylation sequon. It would therefore be interesting to investigate the role of this variable residue, especially if Rauscher MLV and amphotropic MLV were, in fact, shown to be sensitive to deamination. Overall, these observations indicate that moderate sensitivity to mA3 deamination may have provided a subset of exogenous gammaretroviruses with an additional strategy to increase their genetic diversity. As such, permitting sublethal mA3 mutations could help these gammaretroviruses persist in their murine host and thereby increase their chances of becoming endogenized in the mouse germ line. This concept that sublethal levels of A3-induced deamination could be a driving force behind retroviral/HIV evolution, drug resistance, and immune escape has been raised several times before (63–68).

How the *N*-linked glycans of gPr80 prevent A3-induced deamination is still unknown. Although gPr80 is cleaved by cellular proteases, with its N-terminal extremity (containing N113) being packaged into virions (46), it is intriguing that the secreted C-terminal fragment containing glycosylated residues N480 and N505 also prevents deamination. It remains to be established whether gPr80 and mA3 interact. If mA3 were to facilitate the packaging of the full-length gPr80 protein or even that of each of its cleaved polypeptides, then a case could be made that gPr80 glycans act collectively to physically block A3 proteins from ac-

cessing the viral cDNA substrate in the reverse transcription complex. Further studies are required to fully understand how *N*-linked glycans on gPr80 impede the deamination of gammaretroviruses by A3 proteins. Such efforts could potentially lead to the identification of a new class of A3 inhibitors.

ACKNOWLEDGMENTS

We are grateful to Tara Read and Kristin Kemmerich for technical and logistical support to the project and Daniel Figeys and Rui Chen for advice with analyzing protein glycosylation. We also specially acknowledge Hung Fan and thank him for helpful discussions, comments on the manuscript, and the anti-p30^{CA} antibody.

M.-A.L. holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity.

This research was supported by grant 89774 from the Canadian Institutes of Health Research and an early researcher award from the Ontario Ministry of Research and Innovation to M.-A.L.

REFERENCES

1. **Stoye JP.** 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat Rev Microbiol **10:**395–406. http://dx.doi.org/10.1038/nrmicro2783.
2. **Duggal NK, Emerman M.** 2012. Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol **12:**687–695. http://dx.doi.org/10.1038/nri3295.
3. **Rehwinkel J.** 2014. Mouse knockout models for HIV-1 restriction factors. Cell Mol Life Sci **71:**3749–3766. http://dx.doi.org/10.1007/s00018-014-1646-8.
4. **Compton AA, Hirsch VM, Emerman M.** 2012. The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. Cell Host Microbe **11:**91–98. http://dx.doi.org/10.1016/j.chom.2011.11.010.
5. **Meyerson NR, Sawyer SL.** 2011. Two-stepping through time: mammals and viruses. Trends Microbiol **19:**286–294. http://dx.doi.org/10.1016/j.tim.2011.03.006.
6. **Neil SJ, Zang T, Bieniasz PD.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451:**425–430. http://dx.doi.org/10.1038/nature06553.
7. **Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, Stephens EB, Guatelli J.** 2008. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell Host Microbe **3:**245–252. http://dx.doi.org/10.1016/j.chom.2008.03.001.
8. **Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J.** 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature **474:**658–661. http://dx.doi.org/10.1038/nature10195.
9. **Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, Yatim A, Emiliani S, Schwartz O, Benkirane M.** 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature **474:**654–657. http://dx.doi.org/10.1038/nature10117.
10. **Sheehy AM, Gaddis NC, Choi JD, Malim MH.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418:**646–650. http://dx.doi.org/10.1038/nature00939.
11. **Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J.** 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature **427:**848–853. http://dx.doi.org/10.1038/nature02343.
12. **Li Y, Li X, Stremlau M, Lee M, Sodroski J.** 2006. Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. J Virol **80:**6738–6744. http://dx.doi.org/10.1128/JVI.00270-06.
13. **Sokolskaja E, Berthoux L, Luban J.** 2006. Cyclophilin A and TRIM5alpha independently regulate human immunodeficiency virus type 1 infectivity in human cells. J Virol **80:**2855–2862. http://dx.doi.org/10.1128/JVI.80.6.2855-2862.2006.
14. **Yap MW, Nisole S, Lynch C, Stoye JP.** 2004. Trim5alpha protein restricts both HIV-1 and murine leukemia virus. Proc Natl Acad Sci U S A. **101:**10786–10791. http://dx.doi.org/10.1073/pnas.0402876101.
15. **Stavrou S, Nitta T, Kotla S, Ha D, Nagashima K, Rein AR, Fan H, Ross SR.** 2013. Murine leukemia virus glycosylated Gag blocks apolipoprotein B editing complex 3 and cytosolic sensor access to the reverse transcription complex. Proc Natl Acad Sci U S A **110:**9078–9083. http://dx.doi.org/10.1073/pnas.1217399110.
16. **Kozak CA, Chakraborti A.** 1996. Single amino acid changes in the murine leukemia virus capsid protein gene define the target of Fv1 resistance. Virology **225:**300–305. http://dx.doi.org/10.1006/viro.1996.0604.
17. **Ohkura S, Goldstone DC, Yap MW, Holden-Dye K, Taylor IA, Stoye JP.** 2011. Novel escape mutants suggest an extensive TRIM5alpha binding site spanning the entire outer surface of the murine leukemia virus capsid protein. PLoS Pathog **7:**e1002011. http://dx.doi.org/10.1371/journal.ppat.1002011.
18. **Kolokithas A, Rosenke K, Malik F, Hendrick D, Swanson L, Santiago ML, Portis JL, Hasenkrug KJ, Evans LH.** 2010. The glycosylated Gag protein of a murine leukemia virus inhibits the antiretroviral function of APOBEC3. J Virol **84:**10933–10936. http://dx.doi.org/10.1128/JVI.01023-10.
19. **Refsland EW, Harris RS.** 2013. The APOBEC3 family of retroelement restriction factors. Curr Top Microbiol Immunol **371:**1–27. http://dx.doi.org/10.1007/978-3-642-37765-5_1.
20. **Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM.** 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol **15:**166–170. http://dx.doi.org/10.1016/j.cub.2004.12.068.
21. **Bishop KN, Holmes RK, Malim MH.** 2006. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J Virol **80:**8450–8458. http://dx.doi.org/10.1128/JVI.00839-06.
22. **Bélanger K, Savoie M, Rosales Gerpe MC, Couture JF, Langlois MA.** 2013. Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. Nucleic Acids Res **41:**7438–7452. http://dx.doi.org/10.1093/nar/gkt527.
23. **Luo K, Wang T, Liu B, Tian C, Xiao Z, Kappes J, Yu XF.** 2007. Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. J Virol **81:**7238–7248. http://dx.doi.org/10.1128/JVI.02584-06.
24. **Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG.** 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res **35:**7096–7108. http://dx.doi.org/10.1093/nar/gkm750.
25. **Wang X, Ao Z, Chen L, Kobinger G, Peng J, Yao X.** 2012. The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. J Virol **86:**3777–3786. http://dx.doi.org/10.1128/JVI.06594-11.
26. **MacMillan AL, Kohli RM, Ross SR.** 2013. APOBEC3 inhibition of mouse mammary tumor virus infection: the role of cytidine deamination versus inhibition of reverse transcription. J Virol **87:**4808–4817. http://dx.doi.org/10.1128/JVI.00112-13.
27. **Nowarski R, Prabhu P, Kenig E, Smith Y, Britan-Rosich E, Kotler M.** 2014. APOBEC3G inhibits HIV-1 RNA elongation by inactivating the viral trans-activation response element. J Mol Biol **426:**2840–2853. http://dx.doi.org/10.1016/j.jmb.2014.05.012.
28. **Guo F, Cen S, Niu M, Yang Y, Gorelick RJ, Kleiman L.** 2007. The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA. J Virol **81:**11322–11331. http://dx.doi.org/10.1128/JVI.00162-07.
29. **Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, Munk C, Nymark-McMahon H, Landau NR.** 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell **114:**21–31. http://dx.doi.org/10.1016/S0092-8674(03)00515-4.
30. **Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS.** 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol **22:**367–377. http://dx.doi.org/10.1093/molbev/msi026.
31. **Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH.** 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol **14:**1392–1396. http://dx.doi.org/10.1016/j.cub.2004.06.057.
32. **Santiago ML, Montano M, Benitez R, Messer RJ, Yonemoto W, Chesebro B, Hasenkrug KJ, Greene WC.** 2008. Apobec3 encodes Rfv3, a gene influencing neutralizing antibody control of retrovirus infection. Science **321:**1343–1346. http://dx.doi.org/10.1126/science.1161121.

33. **Langlois M-A, Kemmerich K, Rada C, Neuberger MS.** 2009. The AKV murine leukemia virus is restricted and hypermutated by mouse APOBEC3. J Virol **83:**11550–11559. http://dx.doi.org/10.1128/JVI.01430-09.
34. **Takeda E, Tsuji-Kawahara S, Sakamoto M, Langlois M-A, Neuberger MS, Rada C, Miyazawa M.** 2008. Mouse APOBEC3 restricts Friend leukemia virus infection and pathogenesis in vivo. J Virol **82:**10998–11008. http://dx.doi.org/10.1128/JVI.01311-08.
35. **Okeoma CM, Lovsin N, Peterlin BM, Ross SR.** 2007. APOBEC3 inhibits mouse mammary tumour virus replication in vivo. Nature **445:**927–930. http://dx.doi.org/10.1038/nature05540.
36. **Rulli SJ, Jr, Mirro J, Hill SA, Lloyd P, Gorelick RJ, Coffin JM, Derse D, Rein A.** 2008. Interactions of murine APOBEC3 and human APOBEC3G with murine leukemia viruses. J Virol **82:**6566–6575. http://dx.doi.org/10.1128/JVI.01357-07.
37. **Okeoma CM, Low A, Bailis W, Fan HY, Peterlin BM, Ross SR.** 2009. Induction of APOBEC3 in vivo causes increased restriction of retrovirus infection. J Virol **83:**3486–3495. http://dx.doi.org/10.1128/JVI.02347-08.
38. **Low A, Okeoma CM, Lovsin N, de las Heras M, Taylor TH, Peterlin BM, Ross SR, Fan H.** 2009. Enhanced replication and pathogenesis of Moloney murine leukemia virus in mice defective in the murine APOBEC3 gene. Virology **385:**455–463. http://dx.doi.org/10.1016/j.virol.2008.11.051.
39. **Browne EP, Littman DR.** 2008. Species-specific restriction of Apobec3-mediated hypermutation. J Virol **82:**1305–1313. http://dx.doi.org/10.1128/JVI.01371-07.
40. **Sanchez-Martinez S, Aloia AL, Harvin D, Mirro J, Gorelick RJ, Jern P, Coffin JM, Rein A.** 2012. Studies on the restriction of murine leukemia viruses by mouse APOBEC3. PLoS One **7:**e38190. http://dx.doi.org/10.1371/journal.pone.0038190.
41. **Edwards SA, Fan H.** 1979. gag-Related polyproteins of Moloney murine leukemia virus: evidence for independent synthesis of glycosylated and unglycosylated forms. J Virol **30:**551–563.
42. **Evans LH, Dresler S, Kabat D.** 1977. Synthesis and glycosylation of polyprotein precursors to the internal core proteins of Friend murine leukemia virus. J Virol **24:**865–874.
43. **Prats AC, De Billy G, Wang P, Darlix JL.** 1989. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. J Mol Biol **205:**363–372. http://dx.doi.org/10.1016/0022-2836(89)90347-1.
44. **Edwards SA, Fan H.** 1980. Sequence relationship of glycosylated and unglycosylated Gag polyproteins of Moloney murine leukemia virus. J Virol **35:**41–51.
45. **Fujisawa R, McAtee FJ, Zirbel JH, Portis JL.** 1997. Characterization of glycosylated Gag expressed by a neurovirulent murine leukemia virus: identification of differences in processing in vitro and in vivo. J Virol **71:**5355–5360.
46. **Fujisawa R, McAtee FJ, Favara C, Hayes SF, Portis JL.** 2001. N-terminal cleavage fragment of glycosylated Gag is incorporated into murine oncornavirus particles. J Virol **75:**11239–11243. http://dx.doi.org/10.1128/JVI.75.22.11239-11243.2001.
47. **Corbin A, Prats AC, Darlix JL, Sitbon M.** 1994. A nonstructural gag-encoded glycoprotein precursor is necessary for efficient spreading and pathogenesis of murine leukemia viruses. J Virol **68:**3857–3867.
48. **Low A, Datta S, Kuznetsov Y, Jahid S, Kothari N, McPherson A, Fan H.** 2007. Mutation in the glycosylated Gag protein of murine leukemia virus results in reduced in vivo infectivity and a novel defect in viral budding or release. J Virol **81:**3685–3692. http://dx.doi.org/10.1128/JVI.01538-06.
49. **Nitta T, Kuznetsov Y, McPherson A, Fan H.** 2010. Murine leukemia virus glycosylated Gag (gPr80gag) facilitates interferon-sensitive virus release through lipid rafts. Proc Natl Acad Sci U S A **107:**1190–1195. http://dx.doi.org/10.1073/pnas.0908660107.
50. **Portis JL, Fujisawa R, McAtee FJ.** 1996. The glycosylated Gag protein of MuLV is a determinant of neuroinvasiveness: analysis of second site revertants of a mutant MuLV lacking expression of this protein. Virology **226:**384–392. http://dx.doi.org/10.1006/viro.1996.0666.
51. **Mikl MC, Watt IN, Lu M, Reik W, Davies SL, Neuberger MS, Rada C.** 2005. Mice deficient in APOBEC2 and APOBEC3. Mol Cell Biol **25:**7270–7277. http://dx.doi.org/10.1128/MCB.25.16.7270-7277.2005.
52. **Langlois M, Beale R, Conticello S, Neuberger M.** 2005. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. Nucleic Acids Res **33:**1913–1923. http://dx.doi.org/10.1093/nar/gki343.
53. **Bélanger K, Savoie M, Aydin H, Renner TM, Montazeri Z, Langlois MA.** 2014. Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis. Virology **448:**168–175. http://dx.doi.org/10.1016/j.virol.2013.10.008.
54. **Suspene R, Henry M, Guillot S, Wain-Hobson S, Vartanian JP.** 2005. Recovery of APOBEC3-edited human immunodeficiency virus G → A hypermutants by differential DNA denaturation PCR. J Gen Virol **86:**125–129. http://dx.doi.org/10.1099/vir.0.80426-0.
55. **Gherghe C, Weeks KM.** 2006. The SL1-SL2 (stem-loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. J Biol Chem **281:**37952–37961. http://dx.doi.org/10.1074/jbc.M607380200.
56. **Paprotka T, Venkatachari NJ, Chaipan C, Burdick R, Delviks-Frankenberry KA, Hu WS, Pathak VK.** 2010. Inhibition of xenotropic murine leukemia virus-related virus by APOBEC3 proteins and antiviral drugs. J Virol **84:**5719–5729. http://dx.doi.org/10.1128/JVI.00134-10.
57. **Paprotka T, Delviks-Frankenberry KA, Cingoz O, Martinez A, Kung HJ, Tepper CG, Hu WS, Fivash MJ, Jr, Coffin JM, Pathak VK.** 2011. Recombinant origin of the retrovirus XMRV. Science **333:**97–101. http://dx.doi.org/10.1126/science.1205292.
58. **Nitta T, Lee S, Ha D, Arias M, Kozak CA, Fan H.** 2012. Moloney murine leukemia virus glyco-Gag facilitates xenotropic murine leukemia virus-related virus replication through human APOBEC3-independent mechanisms. Retrovirology **9:**58. http://dx.doi.org/10.1186/1742-4690-9-58.
59. **Wirth S, Vessaz A, Krummenacher C, Baribaud F, Acha-Orbea H, Diggelmann H.** 2002. Regions of mouse mammary tumor virus superantigen involved in interaction with the major histocompatibility complex class II I-A molecule. J Virol **76:**11172–11175. http://dx.doi.org/10.1128/JVI.76.21.11172-11175.2002.
60. **Nitta T, Tam R, Kim JW, Fan H.** 2011. The cellular protein La functions in enhancement of virus release through lipid rafts facilitated by murine leukemia virus glycosylated Gag. mBio **2**(1):e00341-10. http://dx.doi.org/10.1128/mBio.00341-10.
61. **Fan H, Chute H, Chao E, Feuerman M.** 1983. Construction and characterization of Moloney murine leukemia virus mutants unable to synthesize glycosylated Gag polyprotein. Proc Natl Acad Sci U S A **80:**5965–5969. http://dx.doi.org/10.1073/pnas.80.19.5965.
62. **Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, Richman D, Coffin JM, Landau NR.** 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol **11:**435–442. http://dx.doi.org/10.1038/nsmb758.
63. **Simon V, Zennou V, Murray D, Huang Y, Ho DD, Bieniasz PD.** 2005. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS Pathog **1:**e6. http://dx.doi.org/10.1371/journal.ppat.0010006.
64. **Sadler HA, Stenglein MD, Harris RS, Mansky LM.** 2010. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. J Virol **84:**7396–7404. http://dx.doi.org/10.1128/JVI.00056-10.
65. **Haché G, Mansky LM, Harris RS.** 2006. Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. AIDS Rev **8:**148–157. http://www.aidsreviews.com/files/2006_08_3_148-157.pdf.
66. **Monajemi M, Woodworth CF, Benkaroun J, Grant M, Larijani M.** 2012. Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment. Retrovirology **9:**35. http://dx.doi.org/10.1186/1742-4690-9-35.
67. **Jern P, Russell RA, Pathak VK, Coffin JM.** 2009. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS Pathog **5:**e1000367. http://dx.doi.org/10.1371/journal.ppat.1000367.
68. **Kim EY, Bhattacharya T, Kunstman K, Swantek P, Koning FA, Malim MH, Wolinsky SM.** 2010. Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. J Virol **84:**10402–10405. http://dx.doi.org/10.1128/JVI.01223-10.

Appendix V: RNA-binding residues in the N-terminus of APOBEC3G influence its DNA sequence specificity and retrovirus restriction efficiency.

Kasandra Bélanger and Marc-André Langlois

Contribution:

Assisted by Dr. Marc-André Langlois, I planned, executed, and analyzed all of the experiments presented in this manuscript. Finally, I wrote the manuscript with input from Dr. Marc-André Langlois and participated in the response to reviewers.

Published: *Virology*, April 2015

Virology 483 (2015) 141–148

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

RNA-binding residues in the N-terminus of APOBEC3G influence its DNA sequence specificity and retrovirus restriction efficiency

Kasandra Bélanger, Marc-André Langlois *

Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 13 March 2015
Returned to author for revisions
1 April 2015
Accepted 15 April 2015

Keywords:
APOBEC3G
Restriction factor
HIV-1
DNA deamination

ABSTRACT

APOBEC3G (A3G) is a host-expressed protein that inactivates retroviruses through the mutagenic deamination of cytosines (C) to uracils (U) in single-stranded DNA (ssDNA) replication products. A3G prefers to deaminate cytosines preceded by a cytosine (5′CC), whereas all other A3 proteins target cytosines in a 5′TC motifs. Structural and mutational studies have mapped the dinucleotide deamination preference of A3G to residues in loop 7 of the catalytic C-terminal domain of the protein. Here we report that A3G with a double W94A/W127A substitution in its N-terminus, designed to abolish RNA binding and protein oligomerization, alters the DNA deamination specificity of the enzyme from 5′CC to 5′TC on proviral DNA. We also show that the double substitution severely impairs its deaminase and antiretroviral activities on Vif-deficient HIV-1. Our results highlight that the N-terminal domain of the full length A3G protein has an important influence on its DNA sequence specificity and mutator activity.

Introduction

Retroviruses, such as HIV-1, can insert their genetic material into the genome of their host within only a few hours after entering a cell. To avoid damage or modifications to their genomic DNA, mammals have evolved several defense mechanisms against these unique pathogens. Members of the APOBEC3 (A3) family are deoxycytidine deaminases that present an important barrier to retroviral replication and spread in mammals (Reviewed in Desimmie et al., 2014; Harris and Dudley, 2015) There are seven A3 in humans, A3A to A3H, with each enzyme being characterized by the presence of one (A3A and A3C) or two (A3B, A3D, A3F and A3G) zinc (Z)-coordinating domains (Jarmuz et al., 2002; LaRue et al., 2009). There are three distinct types of Z domains in humans and primates, each containing a conserved $Hx_1Ex_{24-28}PCx_{2-4}C$ motif that constitutes a putative catalytic site in which the histidine and cysteines coordinate a zinc ion, whereas the glutamic acid acts as a proton shuttle. It is the hydrolytic attack of a zinc-activated water molecule on the C4 amine of the substrate deoxycytidine that results in the generation of a déoxyuridine (Neuberger et al., 2003). A3 proteins deaminate exclusively single-stranded DNA (ssDNA) (Harris et al., 2003; Suspene et al., 2004; Yu et al., 2004). The minus strand viral cDNA that is synthesized during the reverse transcription of retroviral genomic RNA is targeted and uracilated by A3 proteins. Uracils then template the insertion of adenosines (A) during the synthesis of the plus-strand proviral DNA thereby giving rise to G-to-A mutations. A high density of mutations, also called hypermutation, result in the incorporation of premature stop codons and protein defects that inactivate viral functions (Lecossier et al., 2003; Mariani et al., 2003).

The deaminase activity of A3 proteins is highly sequence-specific. A3G is the only enzyme of the family that strongly prefers to target cytosines (underlined) immediately preceded by a cytosine (5′CC), whereas all other A3 deaminate cytosines in a 5′TC motif (Beale et al., 2004; Dang et al., 2006; Doehle et al., 2005; Harris et al., 2003; Langlois et al., 2005; Liddament et al., 2004; OhAinle et al., 2006). Because the crystal structure of the full-length double-domain A3G protein has not yet been solved, there is an incomplete understanding of the global residues that influence target DNA specificity. Structurally, all A3 proteins share several conserved features including at least one Z domain consisting of five β-sheets surrounded by six α-helices (Bohn et al., 2013; Byeon et al., 2013; Chen et al., 2008; Kitamura et al., 2012; Siu et al., 2013). A recent study in which A3A and A3G residues were interchanged found that replacing D317 of the CTD of A3G, situated in loop 7 between β4 and α4, with the homologous tyrosine residue of A3A (Y132) was sufficient to alter the local dinucleotide preference for deamination from 5′CC to 5′TC (Rathore et al., 2013). Those results are supported by previous reports also implicating loop 7 residues in the dinucleotide deamination preference of several A3 proteins and also that of

* Correspondence to: Department of Biochemistry, Microbiology and Immunology Faculty of Medicine, University of Ottawa. 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5. Tel.: +1 613 562 5800x7110.
E-mail address: langlois@uottawa.ca (M.-A. Langlois).

http://dx.doi.org/10.1016/j.virol.2015.04.019
0042-6822/

the activation-induced deaminase (AID) (Carpenter et al., 2010; Kohli et al., 2009, 2010; Langlois et al., 2005; Wang et al., 2010).

Because of its potent antiretroviral activity against a wide range of retroviruses, A3G is the best-characterized member of the A3 family. A3G has two Z domains (Z2-Z1) with each one containing a seemingly intact catalytic site, however only its C-terminal Z1 domain (CTD) is catalytically active (Hache et al., 2005; Navarro et al., 2005). The N-terminal domain (NTD) of the enzyme has been ascribed various other functions such as binding to HIV-1 Vif and virion encapsidation (Huthoff et al., 2009; Huthoff and Malim, 2007; Lavens et al., 2010; Mangeat et al., 2004; Schrofelbauer et al., 2004). Residues located in the NTD have also been implicated in nucleic acid binding, especially RNA (Bach et al., 2008; Bélanger et al., 2013; Bulliard et al., 2009; Huthoff et al., 2009; Iwatani et al., 2006; Khan et al., 2007; Shindo et al., 2012).

In addition to its cytosine deaminase activity, it is well established that A3G also has the ability, in certain experimental conditions, to hinder retroviral infection by means that are independent of deamination (Bélanger et al., 2013; Bishop et al., 2006, 2008; Guo et al., 2007; Iwatani et al., 2007; Li et al., 2007; Lu et al., 2015; Luo et al., 2007; Mariani et al., 2003; Mbisa et al., 2010; Newman et al., 2005; Wang et al., 2012). In particular, several groups have demonstrated defects in $tRNA_3^{Lys}$primer annealing and strand transfers during replication, which consequently lead to decreased levels of early and late reverse viral transcript accumulation as well as reduced proviral integration. Overall, these restriction mechanisms are thought to occur as a result of the direct interaction of A3G with the viral reverse transcriptase (RT) and integrase (IN), and possibly also through an alteration in the processivity of the reverse transcriptase caused by the clamping of A3G onto its ssDNA substrate during reverse transcription (Bishop et al., 2008; Luo et al., 2007; Wang et al., 2012). Recent work from our group has demonstrated that RNA binding by A3G is required for deamination-independent restriction (Bélanger et al., 2013). We mapped the RNA-binding ability of A3G to two tryptophans, W94 and W127, in the non-catalytic NTD of the protein. By individually substituting these two residues to alanine, A3G lost most deamination-independent functions including the inhibitions of late reverse transcript accumulation and proviral integration (Bélanger et al., 2013).

In this report we show that an RNA-binding defective mutant of A3G with two substitutions in its N-terminus, W94A/W127A, displays altered DNA substrate specificity. The loss of RNA-binding did not prevent the mutant enzyme from being incorporated into retrovirus virions, however its antiretroviral activity against a Vif-deficient HIV-1 pseudovirus was completely lost. Our observations support that the ability to bind RNA by full length A3G is required for retrovirus restriction and for high frequency provirus hypermutation. Surprisingly, the few hypermutated proviral sequences that were recovered from infected cells displayed a shift in the preferred target motif for deamination from 5′CC to 5′TC. This shift in target DNA specificity of the enzyme was also independently confirmed in a bacterial mutator assay where the double mutant was found to be as potent as the wild type A3G protein. This represents the first report to date showing the involvement of the NTD in the target specificity and regulation of A3G's mutator activity on a retrovirus.

Results and discussion

In a previous study we demonstrated that A3G [W94A] and A3G [W127A] mutants each had a significantly reduced ability to bind RNA, and more specifically, to 7SL, Alu, hY1 and hY3 RNAs (Bélanger et al., 2013). However, weak RNA binding could still be detected with these mutants. Here we were first interested to determine whether the double W94A/W127A substitution could account for all detectable RNA binding by A3G. Lysates from 293T cells transfected with Flag-tagged A3G or mutants were used for Co-IP using anti-Flag conjugated agarose beads followed by RNA isolation. Samples were then subjected to qPCR to measure the relative binding of each APOBEC to RNA. In these conditions, [W94A/W127A] did not bind to any of the RNAs tested above the background level set by the APOBEC2 (A2) negative control (Fig. 1A). A3G has a well-established ability to assemble into large oligomeric or high molecular mass (HMM) ribonucleic complexes (Chiu et al., 2005). In a previous report we showed that the W94A and W127A point mutants displayed a partial ability to assemble into oligomeric structure compared to wild type A3G (Bélanger et al., 2013). Here we analyzed the double mutant in presence and absence of RNase treatment. We could not detect any evidence of assembly into complexes that are RNA-dependent (Fig. 1B). Surprisingly, although RNA binding was lost, cytoplasmic foci believed to be P-bodies, stress granules and other RNA processing bodies that normally associate with A3G were still visible when similar levels of protein expression were compared (Fig. 1C and D) (Gallois-Montbrun et al., 2007; Kozak et al., 2006; Wichroski et al., 2006). This observation thereby indicates that association of A3G with these structures is not dependent on cellular RNA binding, nor is it dependent on protein aggregation or self-assembly (Fig. 1E).

We next investigated the impact of the double mutation on antiretroviral activity. Here we produced HIV [p8.9] pseudoviruses by co-transfecting viral expression plasmids along with A3G and various A3G mutants, including the catalytically inactive E259Q mutant and single point mutants W94A and W127A into 293T cells. We then confirmed by Western blot analysis that all the mutants, including [W94A/W127A], were expressed in 293T cells and were also efficiently packaged into virions, as judged by p24 capsid expression used for normalization (Fig. 2A). For reasons that are yet unclear, A3G mutants containing W94A or W127A point substitutions or the double substitution were all efficiently packaged into HIV [p8.9] pseudoviruses (Fig. 2A), but not into HIV-1 Δ*V* if viruses as would be expected (data not shown) (Bélanger et al., 2013). Consistent with our previous findings showing that the W94A and W127A point mutants displayed reduced antiretroviral activities (Bélanger et al., 2013), [W94A/W127A] was completely unable to restrict the infection of the virus (Fig. 2B).

Loss of the antiretroviral activity of the double mutants could be the result of protein misfolding causing defective deaminase activity. To evaluate the mutator activity of A3G [W94A/W127A] we conducted a bacterial mutator assay in *E. coli*. An increased frequency of rifampicin-resistant (Rif^R) colonies arises when A3G mutates the *rpoB* gene of *E.coli* rendering it resistant to rifampicin (Harris et al., 2002). Our results show that the full-length double mutant is well expressed in *E. coli* and generates similar frequencies of Rif^R colonies than wild type A3G (Fig. 2C and D) and (Table 1). Having confirmed that the double mutant's mutator activity is unaffected, we next investigated whether it could deaminate proviral DNA. Genomic DNA from 293T cells infected with viruses produced in presence of A3G and mutant proteins was isolated after 24 h. The eGFP reporter gene contained within the provirus was amplified by PCR and 120 clones were sequenced. None of the clones displayed G-to-A hypermutation in presence of the double mutant (data not shown). To determine whether there was a very low frequency of hypermutated sequences we next used 3D-PCR and HRM analysis (Bélanger et al., 2014; Suspene et al., 2005). These methods promote the selective PCR amplification (3D-PCR) or detection (HRM) of hypermutated sequences and provide qualitative information as to the intensity of the mutations. Results of these assays revealed the presence of mutations at intensities much lower than wild type A3G but similar to that of the single mutants (Fig. 2E and F) and (Table 1). No mutations were detected in the integrated proviral DNA of viruses produced in presence of A3G E259Q or A2. These results therefore indicate that the double mutant is capable of mutating provirus DNA, but at very low frequencies.

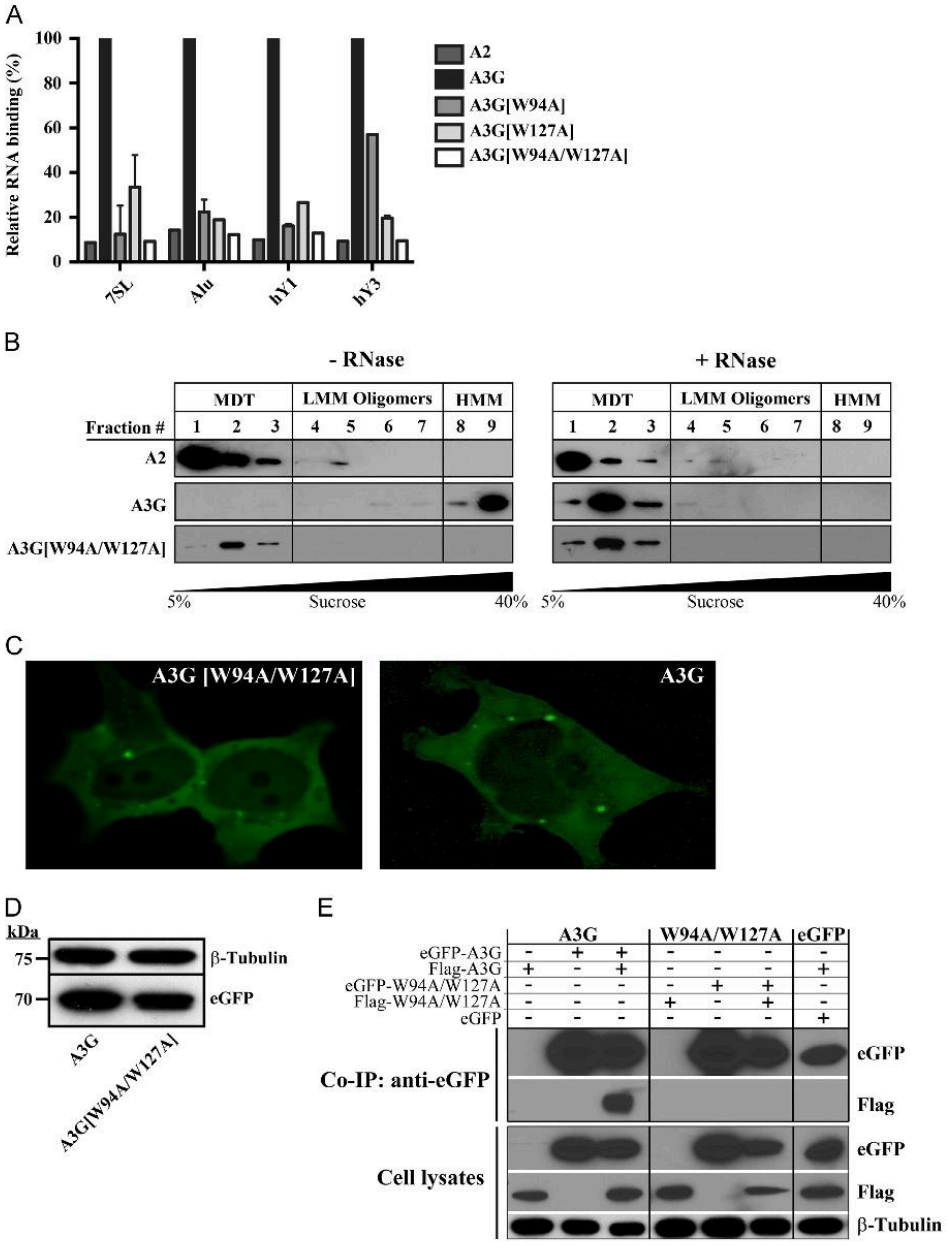


Fig. 1. Binding of A3G [W94A/W127A] to RNA and visualization of its subcellular localization. (**A**) Flag-A2, Flag-A3G or RNA-binding mutants of A3G were transfected in 293T cells. 48 h after transfection, cells were harvested, lysed in NP40 lysis buffer and subjected to Co-IP using anti-Flag-conjugated agarose beads. RNA was then isolated from immunoprecipitates and the binding of the different APOBECs to 7SL, Alu, hY1 and hY3 RNAs was determined by qPCR. Relative binding to A3G is depicted. Results represent the mean ± SD of triplicate values from three independent transfection experiments. (B) Lysates of transfected 293T cells, treated (right panels) or untreated with RNase A (left panels), were resolved by non-denaturing 5–40% sucrose gradient velocity ultra-centrifugation and analyzed by Western blot. Protein detection was performed with an anti-Flag antibody. MDT, monomers, dimers and tetramer; LMM, low molecular mass; HMM, high molecular mass. (C) eGFP-A3G and eGFP-A3G [W94A/W127A] expression plasmids were transfected in 293T cells and cultured in DMEM without phenol red for 48 h. Fluorescent images of the subcellular localization of the proteins were acquired with a 63X objective. (D) Cells transfected with eGFP-A3G or eGFP-A3G [W94A/W127A] were lysed 48 h after transfection and analyzed by Western blot using an anti-eGFP antibody; β-Tubulin was used as a loading control. (E) 293T cells were singly transfected (lanes 1, 2, 4 and 5) or co-transfected (lanes 3, 6 and7) with Flag- or eGFP-tagged APOBEC3 proteins. Pull-downs were carried out with anti-GFP magnetic beads. Homodimers were detected with an anti-Flag antibody (lanes 7–8).

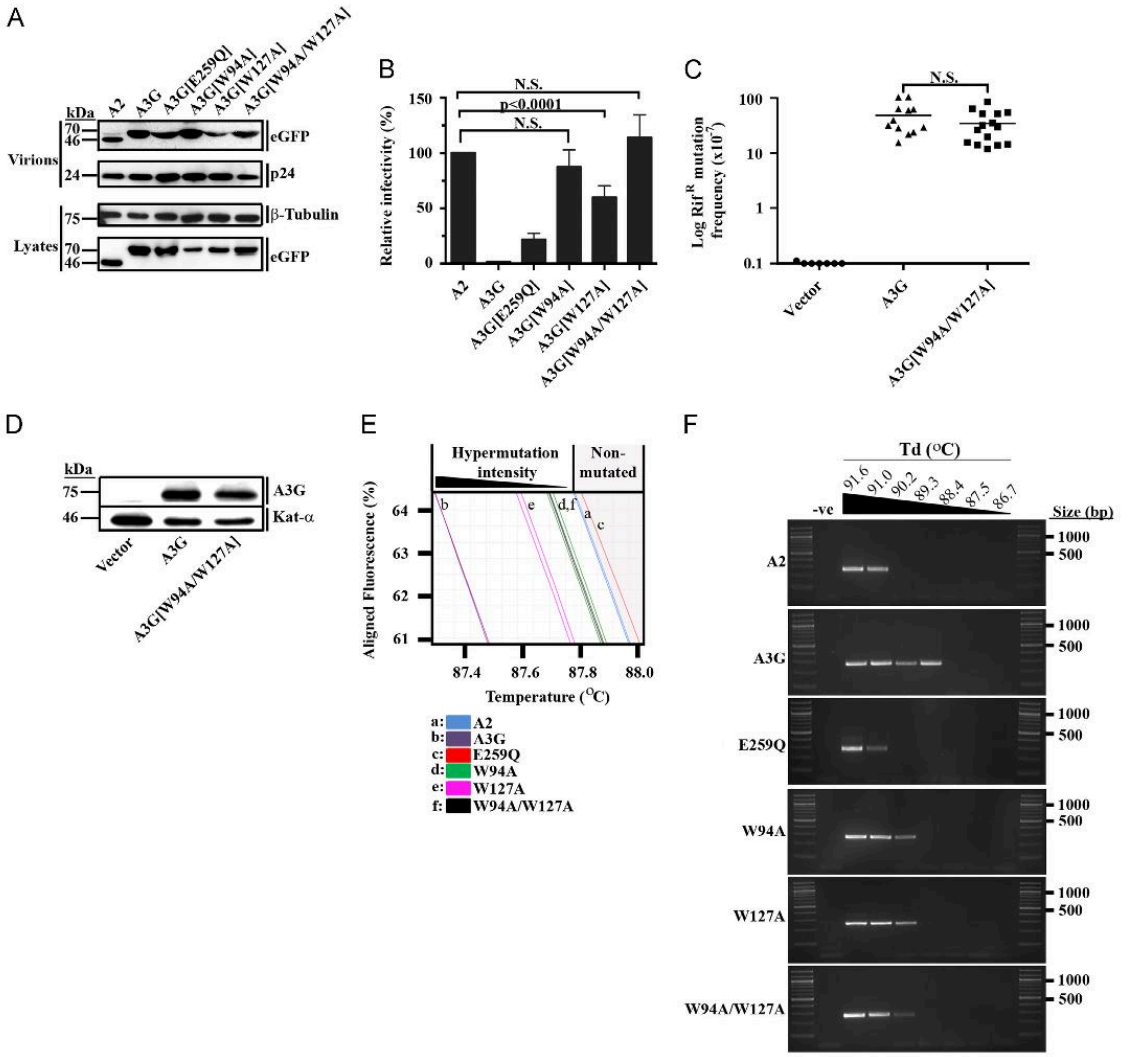


Fig. 2. A3G [W94A/W127A] hypermutates proviral DNA at a low frequency. (**A**) HIV [p8.9] viral particles produced in the presence of A3G or mutants of A3G were purified from cell supernatants, lysed and assayed for protein levels by Western blot analysis. Lysates of virus-producer cells are also presented. (B) Restriction of HIV [p8.9] infection by A3G and mutants as measured by flow cytometry in 293T target cells 24 h after infection. Results represent the mean ± SD of triplicate values from three independent transfection experiments. N.S., not significant. (C) Comparison of the mutator activities of A3G and A3G [W94A/W127A] in *E. coli*. Each point represents in Log the mutation frequency (Rif^R mutants per 10^7 viable cells) of an independent culture; the median values are indicated by a bar. Statistical significance was calculated using a two-tailed Student's *t*-test. (D) Western blot analysis performed on lysates of IPTG-induced bacterial cells used in the mutator assay described above. The expression of Kat-α was used as a loading control. (E and F) HIV [p8.9] viral particles produced in the presence of APOBEC proteins were used to infect 293T target cells. gDNA was extracted from infected cells 24 h after infection and used for (E) HRM analysis or (F) 3D-PCR analysis. (E) Duplicate melting curves for one representative experiment of three is shown. A2-treated samples (blue) delineate to the right a shaded region containing non-hypermutated sequences. PCR amplicons enriched in A/T content due to DNA deamination induce a shift in their melt curves towards lower temperatures. The linear segment of the melt-point threshold of the curve is presented for clarity. (F) 3D-PCR analyses. One representative experiment of three is shown. Amplification at lower temperatures is indicative of increased G-to-A hypermutation intensity.

We next analyzed the local DNA sequence context of the mutations in hypermutated proviral clones obtained by 3D-PCR. Our analysis revealed that the preferred dinucleotide DNA target site for A3G [W94A/W127A] had changed from 5′CC to a very firm 5′TC context (Table 2). We also observed a change in DNA deamination specificity at position -1 in a segment of the *rpoB* gene isolated from Rif^R *E. coli* colonies from a dominant 5′CC to 5′KC (where K is A or G) (Table 2 and Fig. 3) (Beale et al., 2004; Harris et al., 2003). Such a shift in deamination specificity to favor 5′TC is likely to further contribute to the low viral gene-inactivating potential of the mutant enzyme. An important part of the gene inactivation potency of A3G relies on the generation of TAG termination codons caused by deamination opposite to TGG tryptophan codons (Armitage et al., 2012). Weak mutational activity and frequency of mutated proviral DNA sequences combined to a 5′TC deaminase specificity would altogether highly compromise the antiretroviral activity of the A3G [W94A/W127A] protein.

Table 1
Mutation analysis in the eGFP gene of HIV [p8.9] and in the *rpoB* gene of *E. coli*.

APOBEC3	Sequences analyzed		Base pairs analysed[a]		G-to-A mutations		C-to-T mutations	
	eGFP	*rpoB*	eGFP	*rpoB*	eGFP	*rpoB*	eGFP	*rpoB*
A3G	5	33	1350	6534	29	2	0	28
A3G [W94A/W127A]	10	53	2700	10494	88	11	0	32

[a] The fragment analyzed in eGFP was 279 bp long whereas the *rpoB* fragment length was 198 bp.

Table 2
Local sequence preference for DNA deamination.

APOBEC3	HIV [p8.9]			*rpo*B		
	−2	−1	0	−2	−1	0
A3G						
A	24	–	–	–	3	–
C	38	76	100	87	67	100
T	24	24	–	3	23	–
G	14	–	–	10	7	–
A3G [W94A/W127A]						
A	24	2	–	3	16	–
C	26	6	100	67	12	100
T	26	92	–	16	42	–
G	24	–	–	14	30	–

Values represent the percentage of occurrence of each base at positions −1 and −2 respective to the deaminated cytidine (position 0). Sequence analysis with the A3G [E259Q] mutant did not yield any mutated sequences. The number of G-to-A/C-to-T mutations analyzed are as follows: HIV [p8.9]+A3G, $n=29$; HIV [p8.9]+A3G [W94A/W127A], $n=88$; *rpoB*+A3G, $n=30$; *rpoB*+A3G [W94A/W127A], $n=43$.

Overall, our results suggest that tryptophan to alanine substitutions at positions 94 and 127 of the NTD may affect the overall DNA substrate binding efficiency of the full length A3G protein. In fact, several published reports have speculated that W94 and W127 may interact with ssDNA substrates; residue W94 has been recently proposed to contact the DNA substrate in a full-length model of A3G (Lu et al., 2015), and moreover, the residue homologous to W94 in A3A (W98) has been shown to be directly involved in DNA binding (Logue et al., 2014). Furthermore, Chen et al. (2008) previously identified W285, which is homologous to W94 in the CTD of A3G, as an important contributor to DNA-binding. Substrate interactions with residues within the NTD could contribute to its positioning and stabilization for efficient deamination by the active site of the CTD. Finally, the W127A substitution has been shown to alter the scanning mechanism of the protein on ssDNA causing a decreased mutagenic potential, presumably because of the loss of protein homodimerization (Ara et al., 2014). Our experimental data therefore supports these studies by implicating both W94 and W127 in not only RNA binding, but also in physical interactions with the ssDNA substrate.

To date, the DNA dinucleotide preference of A3G has been attributed mainly to amino acids comprising loop 7 of the CTD of the protein. More specifically, Rathore et al. identified that residue D317 is responsible for α-helix capping resulting in restrained mobility of loop 7 and the generation of a binding pocket more favorable for the inclusion of cytosine than any other base (Rathore et al., 2013). In addition, according to the "brim model" for DNA binding, the initial engagement of the substrate has been proposed to occur via the NTD of A3G followed by a prolonged binding by the CTD that then carries out the deamination (Shindo et al., 2012). Because W127 is also located in loop 7 of the NTD of A3G, it is possible that mutations at positions 94 and 127 could destabilize the initial DNA substrate

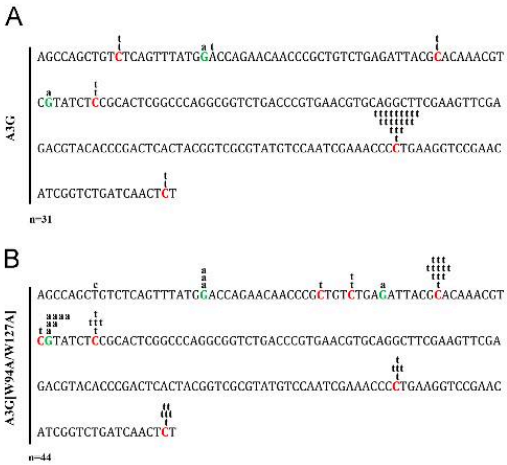


Fig. 3. Distribution of mutations in the *rpoB* gene of Rif^R *E. coli* expressing A3G or A3G [W94A/W127A]. Rif^R colonies generated by transforming *E.coli* with (**A**) GST-A3G or (B) GST-A3G [W94A/W127A] expression plasmids were randomly selected and their *rpoB* gene was amplified by PCR. Amplicons were then cloned and sent for sequencing. Mutated cytosines on the coding strand DNA are indicated in red and in green are mutated guanines; n indicates the total number of mutations in each library. Additional sequence information is presented in Table 1.

binding resulting in a less stringent substrate selection by the catalytic CTD. The crystal structure of full-length A3G bound to its ssDNA substrate would certainly help better understand this process.

Another factor that may also contribute to the loss of antiretroviral activity of the [W94A/W127A] mutant is the lack of deamination-independent restriction. We previously reported that both the W94A and W127A point mutants did not exhibit the typical features of deamination-independent restriction such as reduced late-reverse transcript accumulation and proviral integration (Bélanger et al., 2013). Although the same loss-of-function is expected of the double mutant, an antiretroviral feature that was not investigated is binding to the viral IN and RT. Interactions of A3G with these two viral proteins have previously been linked to reduced retroviral infection (Luo et al., 2007; Wang et al., 2012).

In order to address this question, HIV [p8.9] pseudoviruses were produced in the presence or absence of Flag-tagged APOBEC proteins followed by co-immunoprecipitation (co-IP) using anti-Flag-conjugated agarose beads. Viral proteins were detected in immunoprecipitates by Western blotting. Our results show an interaction between all the A3G mutants, including the double mutant, with both the viral RT and IN suggesting that the binding to the viral enzymes does not require RNA (Fig. 4A). Although [W94A/W127A] was not expressed as well as the point mutants or the wild type protein in presence of virus, the binding ratios to either RT or IN were comparable for all A3G proteins (Fig. 4B). We therefore find no evidence that the binding to IN or RT is sufficient to impede the efficiency of viral infection. However, these results may also indicate that interactions with IN and RT need to occur in conjunction to RNA or ssDNA binding by the NTD in order for A3G to exhibit deamination-independent restriction of retroviral infection.

Materials and methods

Antibodies and cells

Human embryonic kidney epithelium (HEK) 293T cells were cultured in HyClone DMEM/High Glucose medium (Thermo Fischer

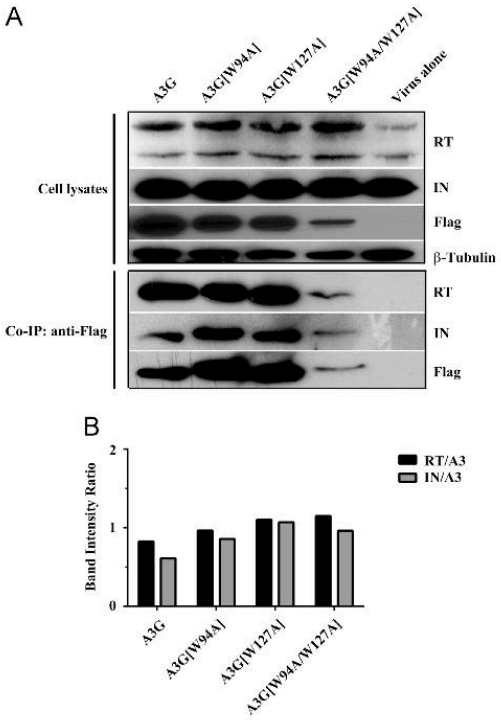


Fig. 4. Interaction of A3G and mutants with HIV-1 RT and IN. (A) HIV [p8.9] particles were produced in the presence of Flag-A3G or mutants of A3G by co-transfection in 293T cells. Cells were harvested 48 h after transfection and lysed for 30 min. Lysates were subjected to Co-IP analysis using anti-Flag-conjugated agarose beads. Following high-pH elution of the proteins from the agarose beads, precipitates were analyzed by Western blot and probed with anti-Flag, anti-IN and anti-RT antibodies. The amount of input cell lysate was normalized using β-tubulin expression. One representative Co-IP experiment of two is depicted. (B) Quantification of Western blot band intensities. Data is presented as the intensity ratio of the viral proteins (RT or IN) to the specific A3 proteins in each experimental condition. Ratios were calculated from the average band intensities from two independent Co-IP experiments.

Scientific) supplemented with 10% fetal bovine serum (FBS), 100U/ml penicillin and 100 μg/ml streptomycin (Multicell). The following antibodies were used for this study: HRP-conjugated anti-Flag (A8592, Sigma), anti-eGFP (no.632381, Clontech), anti-p24 (ab9069, Abcam) and HRP-conjugated anti-β-tubulin (ab21058,Abcam). The anti-Katα was provided by Dr. Stinzi from the University of Ottawa. The anti-HIV-1 integrase polyclonal and anti-HIV-1 reverse transcriptase polyclonal antibodies were obtained from the NIH AIDS Research Reference and Reagent Program (#757 and #6195 respectively) (Grandgenett and Goodarzi, 1994; Szilvay et al., 1992).

Plasmids

Flag-tagged and eGFP-tagged A3G and A2 expression plasmids were constructed previously (Langlois et al., 2005). Mutants of A3G were generated from the corresponding plasmids by PCR-based site directed mutagenesis as previously described (Bélanger et al., 2013).

The single-cycle HIV[p8.9] reporter pseudovirus has been described in previous reports (Bélanger et al., 2013; Langlois et al., 2005). Particles are produced by co-transfecting a Vif-deficient HIV-1 Gag-Pol expression plasmid (p8.9), a packageable HIV-1 LTR backbone containing the eGFP reporter gene under the control of an internal SFFV promoter (pCSGW), and a VSV-G expression vector (pMDG). For the bacterial pGST-expression plasmids, the coding sequences of A3G and A3G [W94A/W127A] were amplified by PCR from pFlag-APOBEC plasmids using the following primers; Xho1-FWD: 5′-GAATTCCCTCGAGGCCACCATGAAGCCTCACTTCAGAAAC-3′ and Pst1-REV: 5′-CGTCGACTGCAGTCAGTTTTCCTGATTCTGGA-3′. The PCR products were then digested and inserted in the pGST-2 vector (Addgene) using the Xho1 and Pst1 restriction sites.

Transfections and viral infectivity assays

Transfections and infections were conducted as described elsewhere (Bélanger et al., 2013). The packaging of APOBEC proteins inside viral particles was assayed by producing HIV [p8.9] viruses in the presence of eGFP-tagged APOBECs for 96 h. Virus-containing supernatants were then purified by ultra-centrifugation and pellets were resuspended in RIPA lysis buffer for Western blot analysis.

Velocity sedimentation

Velocity sedimentation analyses were conducted following the methods described in a previous report (Bélanger et al., 2013). Briefly, 293T cells expressing Flag-A2, Flag-A3G or Flag-A3G [W94A/W127A] were lysed for 30 min on ice with NP40 lysis buffer supplemented with a protease inhibitor cocktail (Roche). Half of each sample was then treated for 15 min at room temperature with 1 μg/mL of RNase A, and both treated and untreated samples were then loaded onto a 5–40% sucrose step gradient and resolved by ultra-centrifugation.

Fluorescence microscopy

Detailed procedures for this part can be found in (Bélanger et al., 2013). 293T cells were seeded in 35 mm glass bottom dishes (MatTech) and transfected with plasmids expressing eGFP-A3G or eGFP-A3G [W94A/W127A]. Images of live cells were captured using a Zeiss Axio Observer.Z1 inverted fluorescent microscope and a Plan-Apochromat 63x/1.4 oil immersion objective. Acquisition and deconvolution were performed using the AxioVision software (Zeiss).

Rif^R bacterial mutator assay

Bacterial mutator assays were performed as previously described (Bélanger et al., 2013). In brief, the *E. coli* uracil excision-defective strain BW310 was transformed with an empty pGST or an APOBEC-expressing pGST plasmid and plated overnight on ampicillin-containing plates. Single transformants were picked and used to inoculate 6 independent liquid media cultures supplemented with 1 mM IPTG and 100 μg/mL ampicillin. To obtain rifampicin resistant (Rif^R) mutants, 300 μL of each culture was spread onto low salt LB media containing 100 μg/mL ampicillin and 50 ug/mL rifampicin and incubated at 37 °C overnight.

For sequencing analysis, Rif^R colonies produced in presence of A3G or A3G [W94A/W127A] were diluted in 200 μl of water and 2 μl of the dilution was used in a PCR reaction mix with primers against a 198 bp region of the RNA polymerase B (*rpoB*) gene. PCR amplification was carried out by the PrimeStar HS high fidelity polymerase (Takara Bio Inc.). The sequence of the primers used for the amplifications are as follows: *rpoB*-FWD: 5′-TTGGCGAAATGGCGGAAAACC-3′; *rpoB*-REV: 5′-CACCGACGGATACCACCTGCTG-3′. Cycling conditions were: 94 °C for 2 min, 94 °C for 30 sec, 65 °C for 1 min and 72 °C for 6 min. Steps 2 to 4 were repeated 8 times with a decrease of 1 °C in the annealing temperature (step 3) at every

successive cycle. That was followed by 21 cycles at 94 °C for 30 sec, 56 °C for 1 min and 72 °C for 5 min. PCR products were purified and sent for Sanger sequencing.

Immunoprecipitation

Methods for immunoprecipitation have been described in (Bélanger et al., 2013). Briefly, lysates of 293T cells transfected or co-transfected with Flag-APOBEC and HIV [p8.9] expression vectors were incubated for 3 h at 4 °C with anti-Flag conjugated agarose beads pre-treated with salmon sperm DNA (500 μg/ml). Bound complexes were then eluted from the beads with a high-pH buffer prior to RNA isolation or Western blot analysis.

RNA binding assay

Methods for RNA amplification and quantification were performed as described in previous reports (Bach et al., 2008; Bélanger et al., 2013). Briefly, RNA was purified from normalized quantities of Flag-tagged APOBEC immunoprecipitates using Trizol (Sigma). After DNAse treatment of the samples, RNA was reverse-transcribed using random hexamers and the ImProm-II™ reverse transcriptase according to the manufacturer's recommendations (Promega). The resulting cDNA was used for qPCR using GoTaq-Green Master Mix (Promega) and specific primers for 7SL, Alu, hY1, hY3 were used as previously described (Bach et al., 2008; Bélanger et al., 2013).

3D-PCR

This procedure is detailed in (Suspene et al. (2005)). Infected 293T cells were collected 24 h after infection for gDNA extraction and 3D-PCR was performed in a two-step protocol using the PrimeStar HS high fidelity polymerase (Takara Bio Inc.). A first-round PCR was performed to amplify a 717 bp eGFP amplicon using primers eGFP-FWD: 5′-GTGAGCAAGGGCGAGGAGCTGTTC-3′ and eGFP-REV:5′CTTGTACAGCTCGTCCATGCCGAGA-3′. A second-round 7-point gradient PCR targeting a 279 bp nested fragment within the eGFP sequence was then performed using primers R279-FWD: 5′–ACAACAGCCACAACGTCTATATCAT-3′ and 279-REV: 5′–CGTCCATGCCGAGAGTGAT-3′. The band of the gradient amplifying at the lowest temperature was then purified and cloned in the pBlueScript vector (Agilent Technologies). Independent clones were sequenced using M13 reverse primers and mutations computed on the plus-strand DNA.

HRM analysis

This method was detailed in a previous report by our group (Bélanger et al., 2014). The protocol was performed on 20 ng of gDNA isolated from virus-infected cells. A 717 bp eGFP amplicon was produced using the MeltDoctor Master Mix (Life Technologies) and the following primers: eGFP-FWD and eGFP-REV. qPCR and HRM were carried out on a ViiA™ 7 real-time PCR instrument (Applied Biosystems), and data was analyzed using the ViiA™ 7 system software (Applied Biosystems).

Protein self-association analysis

Pull-downs for eGFP epitope tags were performed according to the manufacturer's specifications detailed in the uMACS™ Epitope Tag Protein Isolation kits (Miltenyi Biotec).

Statistical analysis and software

All statistical analyses were performed using Student's paired *t*-test using GraphPad Prism software as previously described (Bélanger et al., 2013). Western blot band intensity analysis was performed using ImageJ.

Conclusions

Here we have demonstrated that A3G residues W94 and W127 together account for all detectable RNA binding by the protein. The dual substitution of these tryptophans to alanines resulted in a complete loss of retroviral restriction and in a very low frequency of hypermutated retroviral sequences recovered from infected target cells. Integrated proviral DNA sequences that were deaminated by the W94A/W127A mutant displayed a shift in preferred target sites from 5′CC to 5′TC. Reduced frequencies of hypermutated proviral sequences are not a reflection of reduced mutator activity by the A3G double mutant. Mutation assays in bacteria revealed that A3G [W94A/W127A] was as efficient as the wild type A3G protein, however with a sequence specificity that had also changed, but this time from 5′CC to 5′KC. These results highlight the essential role played by these RNA-binding residues in the efficient restriction and hypermutation of retroviruses. Detailed structural studies have only focused on the crystal of the CTD of A3G. As a consequence, there is still a very limited understanding of how the NTD interacts with both the ssDNA substrate and with the CTD active site of the enzyme. This study is the first to highlight the involvement of the NTD of A3G in influencing both the enzyme's substrate specificity and mutator activity on retroviral DNA.

Acknowledgments

The authors thank Tyler Renner for helpful discussions, and Kristin Kemmerich for comments on the manuscript. We are grateful to the NIH AIDS Research and Reference Reagent Program for reagents. K.B. holds an Ontario Graduate Scholarship and a scholarship from le Fonds de Recherche en Santé du Québec. M.-A. L. holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity. This research was supported by grant #89774 from the Canadian Institutes of Health Research, and an Early Researcher Award from the Ontario Ministry of Research and Innovation to M.-A.L.

References

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

Armitage, A.E., Deforche, K., Chang, C.H., Wee, E., Kramer, B., Welch, J.J., Gerstoft, J., Fugger, L., McMichael, A., Rambaut, A., Iversen, A.K., 2012. APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. PLoS Genet. 8, e1002550.

Bach, D., Peddi, S., Mangeat, B., Lakkaraju, A., Strub, K., Trono, D., 2008. Characterization of APOBEC3G binding to 7SL RNA. Retrovirology 5, 54.

Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., Neuberger, M.S., 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J. Mol. Biol. 337, 585–596.

Bélanger, K., Savoie, M., Aydin, H., Renner, T.M., Montazeri, Z., Langlois, M.A., 2014. Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis. Virology 448, 168–175.

Bélanger, K., Savoie, M., Rosales Gerpe, M.C., Couture, J.F., Langlois, M.A., 2013. Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. Nucleic Acids Res. 41, 7438–7452.

Bishop, K.N., Holmes, R.K., Malim, M.H., 2006. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J. Virol. 80, 8450–8458.

Bishop, K.N., Verma, M., Kim, E.Y., Wolinsky, S.M., Malim, M.H., 2008. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS Pathog. 4, e1000231.
Bohn, M.F., Shandilya, S.M., Albin, J.S., Kouno, T., Anderson, B.D., McDougle, R.M., Carpenter, M.A., Rathore, A., Evans, L., Davis, A.N., Zhang, J., Lu, Y., Somasundaran, M., Matsuo, H., Harris, R.S., Schiffer, C.A., 2013. Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. Structure 21, 1042–1050.
Bulliard, Y., Turelli, P., Rohrig, U.F., Zoete, V., Mangeat, B., Michielin, O., Trono, D., 2009. Functional analysis and structural modeling of human APOBEC3G reveal the role of evolutionarily conserved elements in the inhibition of human immunodeficiency virus type 1 infection and Alu transposition. J. Virol. 83, 12611–12621.
Byeon, I.J., Ahn, J., Mitra, M., Byeon, C.H., Hercik, K., Hritz, J., Charlton, L.M., Levin, J.G., Gronenborn, A.M., 2013. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. Nat. Commun. 4, 1890.
Carpenter, M.A., Rajagurubandara, E., Wijesinghe, P., Bhagwat, A.S., 2010. Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair9, 579–587.
Chen, K.M., Harjes, E., Gross, P.J., Fahmy, A., Lu, Y., Shindo, K., Harris, R.S., Matsuo, H., 2008. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature 452, 116–119.
Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W., Greene, W.C., 2005. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+T cells. Nature 435, 108–114.
Dang, Y., Wang, X., Esselman, W.J., Zheng, Y.H., 2006. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. J. Virol. 80, 10522–10533.
Desimmie, B.A., Delviks-Frankenberrry, K.A., Burdick, R.C., Qi, D., Izumi, T., Pathak, V.K., 2014. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. J. Mol. Biol.426, 1220–1245.
Doehle, B.P., Schafer, A., Cullen, B.R., 2005. Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. Virology 339, 281–288.
Gallois-Montbrun, S., Kramer, B., Swanson, C.M., Byers, H., Lynham, S., Ward, M., Malim, M.H., 2007. Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. J. Virol. 81, 2165–2178.
Grandgenett, D.P., Goodarzi, G., 1994. Folding of the multidomain human immunodeficiency virus type-I integrase. Protein Sci.: Publ. Protein Soc. 3, 888–897.
Guo, F., Cen, S., Niu, M., Yang, Y., Gorelick, R.J., Kleiman, L., 2007. The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA. J. Virol. 81, 11322–11331.
Hache, G., Liddament, M.T., Harris, R.S., 2005. The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. J. Biol. Chem. 280, 10920–10924.
Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., Malim, M.H., 2003. DNA deamination mediates innate immunity to retroviral infection. Cell 113, 803–809.
Harris, R.S., Dudley, J.P., 2015. APOBECs and virus restriction. Virology 479–480, 131–145.
Harris, R.S., Petersen-Mahrt, S.K., Neuberger, M.S., 2002. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol. Cell. 10, 1247–1253.
Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F., Malim, M.H., 2009. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. PLoS Pathog. 5, e1000330.
Huthoff, H., Malim, M.H., 2007. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. J. Virol. 81, 3807–3815.
Iwatani, Y., Chan, D.S., Wang, F., Maynard, K.S., Sugiura, W., Gronenborn, A.M., Rouzina, I., Williams, M.C., Musier-Forsyth, K., Levin, J.G., 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res. 35, 7096–7108.
Iwatani, Y., Takeuchi, H., Strebel, K., Levin, J.G., 2006. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. J. Virol. 80, 5992–6002.
Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., Navaratnam, N., 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics 79, 285–296.
Khan, M.A., Goila-Gaur, R., Opi, S., Miyagi, E., Takeuchi, H., Kao, S., Strebel, K., 2007. Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. Retrovirology 4, 48.
Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A., Sugiura, W., Iwatani, Y., 2012. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. Nat. Struct. Mol. Biol. 19, 1005–1010.
Kohli, R.M., Abrams, S.R., Gajula, K.S., Maul, R.W., Gearhart, P.J., Stivers, J.T., 2009. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J. Biol. Chem. 284, 22898–22904.
Kohli, R.M., Maul, R.W., Guminski, A.F., McClure, R.L., Gajula, K.S., Saribasak, H., McMahon, M.A., Siliciano, R.F., Gearhart, P.J., Stivers, J.T., 2010. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. J. Biol. Chem. 285, 40956–40964.
Kozak, S.L., Marin, M., Rose, K.M., Bystrom, C., Kabat, D., 2006. The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. J. Biol. Chem. 281, 29105–29119.
Langlois, M., Beale, R., Conticello, S., Neuberger, M., 2005. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. Nucleic Acids Res. 33, 1913–1923.
LaRue, R.S., Andresdottir, V., Blanchard, Y., Conticello, S.G., Derse, D., Emerman, M., Greene, W.C., Jonsson, S.R., Landau, N.R., Lochelt, M., Malik, H.S., Malim, M.H., Munk, C., O'Brien, S.J., Pathak, V.K., Strebel, K., Wain-Hobson, S., Yu, X.F., Yuhki, N., Harris, R. S., 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. J. Virol. 83, 494–497.
Lavens, D., Peelman, F., Van der Heyden, J., Uyttendaele, I., Catteeuw, D., Verhee, A., Van Schoubroeck, B., Kurth, J., Hallenberger, S., Clayton, R., Tavernier, J., 2010. Definition of the interacting interfaces of Apobec3G and HIV-1 Vif using MAPPIT mutagenesis analysis. Nucleic Acids Res. 38, 1902–1912.
Lecossier, D., Bouchonnet, F., Clavel, F., Hance, A.J., 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 300, 1112.
Li, X.Y., Guo, F., Zhang, L., Kleiman, L., Cen, S., 2007. APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription. J. Biol. Chem. 282, 32065–32074.
Liddament, M.T., Brown, W.L., Schumacher, A.J., Harris, R.S., 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr. Biol.: CB 14, 1385–1391.
Logue, E.C., Bloch, N., Dhuey, E., Zhang, R., Cao, P., Herate, C., Chauveau, L., Hubbard, S.R., Landau, N.R., 2014. A DNA sequence recognition loop on APOBEC3A controls substrate specificity. PloS One 9, e97062.
Lu, X., Zhang, T., Xu, Z., Liu, S., Zhao, B., Lan, W., Wang, C., Ding, J., Cao, C., 2015. Crystal structure of DNA cytidine deaminase abobec3g catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. J. Biol. Chem. 290, 4010–4021.
Luo, K., Wang, T., Liu, B., Tian, C., Xiao, Z., Kappes, J., Yu, X.F., 2007. Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. J. Virol. 81, 7238–7248.
Mangeat, B., Turelli, P., Liao, S., Trono, D., 2004. A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J. Biol. Chem. 279, 14481–14483.
Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., Landau, N.R., 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell 114, 21–31.
Mbisa, J.L., Bu, W., Pathak, V.K., 2010. APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. J. Virol. 84, 5250–5259.
Navarro, F., Bollman, B., Chen, H., Konig, R., Yu, Q., Chiles, K., Landau, N.R., 2005. Complementary function of the two catalytic domains of APOBEC3G. Virology 333, 374–386.
Neuberger, M.S., Harris, R.S., Di Noia, J., Petersen-Mahrt, S.K., 2003. Immunity through DNA deamination. Trends Biochem. Sci. 28, 305–312.
Newman, E.N., Holmes, R.K., Craig, H.M., Klein, K.C., Lingappa, J.R., Malim, M.H., Sheehy, A.M., 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr. Biol.: CB 15, 166–170.
OhAinle, M., Kerns, J.A., Malik, H.S., Emerman, M., 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J. Virol. 80, 3853–3862.
Rathore, A., Carpenter, M.A., Demir, O., Ikeda, T., Li, M., Shaban, N.M., Law, E.K., Anokhin, D., Brown, W.L., Amaro, R.E., Harris, R.S., 2013. The local dinucleotide preference of APOBEC3G can be altered from 5′-CC to 5′-TC by a single amino acid substitution. J. Mol. Biol. 425, 4442–4454.
Schrofelbauer, B., Chen, D., Landau, N.R., 2004. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc. Natl. Acad. Sci. USA 101, 3927–3932.
Shindo, K., Li, M., Gross, P.J., Brown, W.L., Harjes, E., Lu, Y., Matsuo, H., Harris, R.S., 2012. A comparison of two single-stranded DNA binding models by mutational analysis of APOBEC3G. Biology 1, 260–276.
Siu, K.K., Sultana, A., Azimi, F.C., Lee, J.E., 2013. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. Nat. Commun. 4, 2593.
Suspene, R., Henry, M., Guillot, S., Wain-Hobson, S., Vartanian, J.P., 2005. Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. J. Gen. Virol. 86, 125–129.
Suspene, R., Sommer, P., Henry, M., Ferris, S., Guetard, D., Pochet, S., Chester, A., Navaratnam, N., Wain-Hobson, S., Vartanian, J.P., 2004. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. Nucleic Acids Res. 32, 2421–2429.
Szilvay, A.M., Nornes, S., Haugan, I.R., Olsen, L., Prasad, V.R., Endresen, C., Goff, S.P., Helland, D.E., 1992. Epitope mapping of HIV-1 reverse transcriptase with monoclonal antibodies that inhibit polymerase and RNase H activities. J. Acquir. Immune Defic. Syndr. 5, 647–657.
Wang, M., Rada, C., Neuberger, M.S., 2010. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. J. Exp. Med. 207, 141–153.
Wang, X., Ao, Z., Chen, L., Kobinger, G., Peng, J., Yao, X., 2012. The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. J. Virol. 86, 3777–3786.
Wichroski, M.J., Robb, G.B., Rana, T.M., 2006. Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. PLoS Pathog. 2, e41.
Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., Landau, N.R., 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat. Struct. Mol. Biol. 11, 435–442.

Appendix VI: Comparative Analysis of the Gene-Inactivating Potential of Human Retroviral Restriction Factors APOBEC3F and APOBEC3G.

Kasandra Bélanger and Marc-André Langlois.

Contribution:

I executed and analyzed all the experiments presented in the manuscript with input from Dr. Marc-André Langlois. I wrote the first draft of the manuscript and participated in the review process.

Published: *Journal of General Virology*, 2015

Journal of General Virology

Comparative Analysis of the Gene-Inactivating Potential of Retroviral Restriction Factors APOBEC3F and APOBEC3G

--Manuscript Draft--

Manuscript Number:	VIR-D-15-00342R1
Full Title:	Comparative Analysis of the Gene-Inactivating Potential of Retroviral Restriction Factors APOBEC3F and APOBEC3G
Short Title:	A3F promotes retroviral genetic diversification when the mutational burden is low
Article Type:	Standard
Section/Category:	Animal - Retroviruses
Corresponding Author:	Marc-André Langlois, Ph.D. University of Ottawa Ottawa, Ontario CANADA
First Author:	Kasandra Bélanger
Order of Authors:	Kasandra Bélanger
	Marc-André Langlois, Ph.D.
Abstract:	APOBEC3 (A3) proteins are host-encoded restriction factors that inhibit retrovirus infection by mutagenic deamination of cytosines in minus-strand DNA replication intermediates. APOBEC3F (A3F) and APOBEC3G (A3G) are two of the most potent A3 enzymes in humans with each having a different target DNA specificity. A3G prefers to deaminate cytosines preceded by a cytosine (5'CC), whereas A3F preferentially targets cytosines preceded by a thymine (5'TC). Here we performed a detailed comparative analysis of retrovirus-encoded gene sequences edited by A3F and A3G with the aim of correlating the context and intensity of the mutations with their effects on gene function. Our results reveal that, when there are few (TGG) tryptophan codons in the sequence, both enzymes alter gene function with a similar efficiency when given equal opportunities to deaminate in their preferred target DNA context. In contrast, tryptophan-rich genes are efficiently inactivated in presence of a low mutational burden through termination codon generation by A3G but not A3F. Overall our results clearly demonstrate that the target DNA specificity of an A3 enzyme along with the intensity of the mutational burden and the tryptophan content of the gene being targeted are the factors that have the most forceful influence on whether A3-induced mutations will favor either terminal inactivation or genetic diversification of a retrovirus.

Comparative Analysis of the Gene-Inactivating Potential of Retroviral Restriction Factors APOBEC3F and APOBEC3G

Kasandra Bélanger and Marc-André Langlois*

Department of Biochemistry, Microbiology and Immunology,
Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

*Corresponding author. Mailing address: Department of Biochemistry, Microbiology and Immunology
Faculty of Medicine, University of Ottawa. 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5.
Phone: +1 (613) 562-5800, x7110. E-mail: langlois@uottawa.ca

SUMMARY

APOBEC3 (A3) proteins are host-encoded restriction factors that inhibit retrovirus infection by mutagenic deamination of cytosines in minus-strand DNA replication intermediates. APOBEC3F (A3F) and APOBEC3G (A3G) are two of the most potent A3 enzymes in humans with each having a different target DNA specificity. A3G prefers to deaminate cytosines preceded by a cytosine (5'CC), whereas A3F preferentially targets cytosines preceded by a thymine (5'TC). Here we performed a detailed comparative analysis of retrovirus-encoded gene sequences edited by A3F and A3G with the aim of correlating the context and intensity of the mutations with their effects on gene function. Our results reveal that, when there are few (TGG) tryptophan codons in the sequence, both enzymes alter gene function with a similar efficiency when given equal opportunities to deaminate in their preferred target DNA context. In contrast, tryptophan-rich genes are efficiently inactivated in presence of a low mutational burden through termination codon generation by A3G but not A3F. Overall our results clearly demonstrate that the target DNA specificity of an A3 enzyme along with the intensity of the mutational burden and the tryptophan content of the gene being targeted are the factors that have the most forceful influence on whether A3-induced mutations will favor either terminal inactivation or genetic diversification of a retrovirus.

Key Words: APOBEC3F; APOBEC3G; hypermutation; cytidine deamination; HIV; retroviral restriction factors.

INTRODUCTION

The human APOBEC3 (A3) family is composed of 7 proteins (A3A, A3B, A3C, A3D, A3F, A3G and A3H) with DNA cytosine deaminase activity. Together they play an important role in preventing the replication and spread of a wide range of viruses, retroviruses and retroelements (For reviews see (Desimmie *et al.*, 2014; Harris & Dudley, 2015; Koito & Ikeda, 2013)).

A3F and A3G are potent and well-characterized members of the A3 family that are mostly recognized for their roles in mutating and restricting the infection of HIV-1. However, this virus has evolved effective strategies to dodge the effects of these host intrinsic proteins, one of them is through the expression of the viral infectivity factor (Vif) (Desimmie *et al.*, 2014). The main role of Vif is to protect HIV-1 from the deleterious actions of A3G and A3F (and some other members of the A3 family) by linking these proteins to the E3 ubiquitin ligase complex, which then targets them for degradation in the proteasome (Kobayashi *et al.*, 2005; Liu *et al.*, 2004; Mehle *et al.*, 2004; Yu *et al.*, 2003). In absence of Vif early in the infection or when Vif levels are low, A3 proteins are readily packaged into nascent viral particles that are released from virus-infected cells. Upon virus entry into new susceptible target cells, stowaway A3 proteins then deaminate cytosines (C) into uracils (U) in the minus-strand viral cDNA during reverse transcription (Desimmie *et al.*, 2014; Harris & Dudley, 2015; Koito & Ikeda, 2013). This results in the accumulation of abundant G-to-A mutations, also called hypermutation, on the viral plus-strand that can lead to gene inactivation in progeny viruses. Some A3 proteins display important differences in their selection of the cytosines that they prefer to deaminate. The bases immediately upstream to the deaminated C (underlined) influence this selection. A3G has a strong preference to deaminate a C preceded by a C (5'-C<u>C</u>), while all other human A3 proteins, including A3F, prefer to deaminate C preceded by a thymine (5'-T<u>C</u>) (Bogerd *et al.*, 2007; Chelico *et al.*, 2006; Hultquist *et al.*, 2011; Langlois *et al.*, 2005; Love *et al.*, 2012; Rausch *et al.*, 2009). Considering that A3-induced deamination events occur predominantly on the minus-strand viral DNA, deamination in a 5'C<u>C</u> context opposite to a tryptophan (Trp) codon (T<u>G</u>G) during reverse transcription produces a T<u>A</u>G termination codon, and also possibly TG<u>A</u> and T<u>AA</u> non-sense codons if the TGG codon is followed by a G (i.e. T<u>GG</u> G) (Ara *et al.*, 2014; Yu *et al.*, 2004; Zennou & Bieniasz, 2006). Equally, deamination in a 5'T<u>C</u> context can also generate a TGA termination codon, but only when a Trp codon is immediately followed by an A (i.e. TG<u>G</u> A).

A3-induced mutations do not always lead to gene inactivation and can potentially be beneficial for a retrovirus given the right conditions. Low levels of non-deleterious G-to-A mutations are believed to help retroviruses evolve and evade immune defenses (Ara *et al.*, 2014; Armitage *et al.*, 2014; Kim *et al.*, 2014; Monajemi *et al.*, 2012; Monajemi *et al.*, 2014). This phenomenon may be especially important soon after a new cell infection. At this time, Vif levels are increasing but have not yet reached a threshold capable of completely preventing the packaging of A3 proteins into HIV-1 virions (Strebel, 2013). Importantly, sublethal mutations have been associated with the emergence of antiretroviral drug resistance and reduced cytotoxic T lymphocyte recognition (Fourati *et al.*, 2012a; Fourati *et al.*, 2012b; Hache *et al.*, 2006; Jern *et al.*, 2009; Monajemi *et al.*, 2014; Mulder *et al.*, 2008; Neogi *et al.*, 2013; Sadler *et al.*, 2010; Simon *et al.*, 2005). It is therefore of significant interest to better understand the conditions that lead A3-induced mutations towards either viral gene inactivation or towards sequence evolution that ultimately benefits the virus.

Tangentially, reporter cells such as CEM-GFP or TZM-bl, that express a reporter protein under the control of an integrated HIV-1 promoter are tools commonly used to measure the impact of A3 deamination on HIV infection. These cells express a reporter gene, often luciferase, β-galactosidase or eGFP, upon the translation of a functional Tat and/or Rev protein contributed by an infecting HIV-1 provirus (Derdeyn *et al.*, 2000; Gervaix *et al.*, 1997; Platt *et al.*, 1998). Similarly, indicator retroviruses that express an eGFP reporter protein once integrated in a target cell are also frequently used to assess the retroviral restriction potency of the various A3 proteins. In light that A3 proteins can display different target DNA specificities, a bias could be introduced in these reporter assays that is related to the composition of the specific gene sequences being mutated and their differential susceptibility to be deaminated by any one A3 protein. Furthermore, the mutation burden intensity required to altered or inactivate the function of these proteins has not been formally assessed experimentally.

To address some of these issues, we performed a detailed analysis of how A3-induced mutations and amino acid substitutions affect the fluorescence intensity of eGFP expressed in two types of reporter retroviruses: HIV-1 ΔVif and the Moloney murine leukemia virus (M-MLV). We investigated the efficiency of A3F and A3G to reduce the fluorescence of either the eGFP reporter protein alone, or eGFP as a fusion protein with a viral envelope glycoprotein (Env). Our results show that there is an incongruent relationship between mutation intensity and gene inactivation by A3F and A3G, which is dependent on the nature of the sequence being mutated. When target sequences have no or few Trp

codons and nearly equal numbers of 5'TC and 5'CC deamination sites, A3F and A3G display similar abilities to mutate and diminish reporter gene fluorescence, with A3G being slightly more efficient at completely inactivating gene function. However, A3G potently inactivates gene function by high-frequency stop codon generation in target sequences containing numerous Trp codons. In contrast, A3F mainly inactivates through intense hypermutation, mostly in absence of stop codons. Additionally, our study reveals that mutations located outside the eGFP region of the Env-eGFP fusion protein can also alter fluorescence intensity. This indicates that A3 mutations could have an impact on various other parameters affecting protein function and gene expression such as mRNA and protein stability.

RESULTS

A3F deamination contributes to sequence diversification and rarely creates termination codons.

While eGFP has but a single Trp codon (TGG), the envelope glycoprotein gene of M-MLV (Env-eGFP) has 13 additional Trp codons in the N-terminal region of the protein (N-ter.), and 5 in its C-terminus (Table 1). Of these, there are 4 Trp codons followed by and adenine (A) (i.e. TGG A), which can result in a TGA termination codon when targeted by A3F in a 5'TC context on the minus strand. The number of 5'TC and 5'CC deamination target sites were nearly identical in the sequences coding for eGFP and for the C-terminus of Env, but there were 72 5'CC A3G target sites in the N-terminal sequence of Env compared to 45 5'TC target sites for A3F (Table 1).

To quantify the gene inactivating potential of A3 proteins, we produced reporter viruses expressing either eGFP alone under the control of an internal Tat/Rev-independent promoter (HIV-1 ΔVif), or eGFP embedded in-frame within Env-eGFP of M-MLV in cells co-transfected with either A3F or A3G (Fig.1). The M-MLV and HIV-1 ΔVif viruses produced were then used to infect target cells from which integrated proviral DNA was amplified by PCR. Flow cytometry analysis was used to monitor virus restriction by the A3 proteins. Both A3F and A3G potently restricted the viruses, with A3G causing near complete restriction (Fig. 1a and b). Human APOBEC2 (A2) was used as a negative control. A3 protein expression in the virus producer cells was similar in all cases (Fig 1c and d).

To assess the effect of A3-induced mutations on reporter protein fluorescence, PCR amplicons containing the complete coding sequence for eGFP or Env-eGFP were digested and ligated into a CMV expression plasmid and then transformed into *E.coli* to establish a library of A3F and A3G mutated sequences for each virus (Fig. 2a). Assays were performed on 20 randomly selected mutated clones and one randomly selected unmutated clone from each experimental condition (Table 2). Amongst mutated clones of eGFP, A3F displayed an average mutation rate of 8.2 G-to-A mutations per kb and A3G 12.4 mutations per kb. These results show that A3G has greater mutagenic potential than A3F when provided with equal numbers of optimal consensus sites and when expressed at similar levels in the virus producing cells (Fig. 1c and d). The rationale for using two different reporter viruses was to see if the profiles, positions and intensities of the mutations in eGFP differed, but we found no clear evidence of this (data not shown).

To visualize the distribution of putative mutations in the Env-eGFP gene, all 5'TC and 5'CC dinucleotides were first located and their positions were then mapped to the corresponding codon that would be affected by the deamination of the targeted cytosine (Fig. 2b). In total, there are 163 codons that can be affected by at least one deamination event occurring in a 5'TC context, and 167 in a 5'CC context; some codons are affected by mutations in both contexts (e.g., GAG GNN). The sequence of the adjacent codons was also therefore taken into consideration for these calculations. We next pooled the mutations observed in all eGFP and Env-eGFP clones and analyzed them separately for A3F (Fig. 2c) and A3G (Fig. 2d). We then plotted every codon containing a mutation occurring in either a 5'TC or 5'CC context (indicated as Mutated Total). We also plotted the location of the codons that were not deaminated (indicated as Non-mutated). Analysis of the local DNA sequence context flanking these non-mutated codons did not reveal a common motif (data not shown).

Analysis of the clones derived from mutated M-MLV and HIV-1 ΔVif proviral sequences revealed that 144 different codons were targeted by A3F: 63% of mutations occurred in a 5'TC context, 13% were in 5'CC, and the remaining 24% were in a 5'RC context, where R is either A or G (Fig. 2c). Only 4 of these 40 clones contained at least one stop codon indicated by red lines in Fig. 2(c) (Tables 3, S1 and S2). In contrast to A3F, A3G deaminated 187 different codons, 65% of which in a 5'CC context, 19% in 5'TC, and 16% in a 5'RC context (Fig. 2d). The most striking difference between A3F and A3G is the large number of stop codons that were generated by A3G. A total of 152 stop codons in 40 clones were generated by A3G that converted 17 out of the possible 19 Trp codons into termination codons (Tables 3, S3 and S4). Of interest, only one of the 80 clones analyzed (Env-A3G-20) contained a stop codon in the eGFP coding sequence (Fig. 2d and Table S8). In terms of generating non-synonymous (NS) amino acid substitutions in the eGFP gene, A3G was only slightly more efficient than A3F (45% compared to 32%) (Table 4 and Figs. S5 and S6). As for the full-length Env-eGFP gene sequence, A3F produced overall 6% more NS substitutions than A3G (71% compared to 65%) (Table 4 and Figs. S7 and S8).

Disparity between reporter fluorescence intensity and mutation burden.

We next wanted to correlate A3-induced mutation intensity with alterations in reporter gene fluorescence. The main question we wanted to address is whether there is a minimal mutational threshold that would reliably predict gene inactivation. Mutated eGFP and Env-eGFP clones were individually transfected into 293T cells in three independent experiments and analyzed for eGFP

expression by flow cytometry (Tables S1 to S4). The relative mean fluorescence, the total number of G-to-A mutations and the number of NS substitutions in eGFP are graphically presented for eGFP in Fig. 3(a) and the N-ter. of Env-eGFP clones in Fig. 3(b). Clones containing at least one termination codon are indicated by a red hexagon with the total number of stop codons contained within. Our results with A3F-mutated HIV ΔVif-derived eGFP clones show that total G-to-A mutations per sequence ranged from 1 to 21, with NS substitutions ranging from 0 to 6 per sequence (Fig. 3a, Table S1 and S5). A3G-mutated eGFP clones contained between 2 and 26 G-to-A mutations, and between 1 and 10 NS substitutions (Fig. 3a, Table S2 and S6). The mutation burden could not reliably predict a reduction in fluorescence intensity as some heavily mutated sequences still exhibited high levels of fluorescence (Fig 3a). However, no detectable fluorescence was observed in clones containing 11 or more G-to-A mutations caused by either A3F or A3G.

Interestingly, several eGFP clones displayed very low fluorescence when few mutations were present such as clones A3F-3 and A3F-5. Both these clones share a common substitution, E214K, but the defect in fluorescence appears to be partially rescued in presence of mutation S209N, as seen in clone A3F-15 that contains a dual mutation (Fig. 3a and Table S5). This specifically illustrates the epistatic-like effect that mutations can have on gene function. Conversely, several intensely mutated A3F clones remained strongly fluorescent (e.g., A3F-18 and A3F-19). This is in contrast to the more heavily mutated A3G-derived clones that all exhibit reduced fluorescence, but these generally also contained slightly higher numbers of NS substitutions than their equivalent A3F counterparts.

A3G inactivates reporter function by termination codon generation when Trp codons are abundant.

M-MLV encoding the Env-eGFP fusion offers 13 additional Trp codons upstream of eGFP that can be converted into stop codons by deamination in a 5'CC context (Table 1). Of these, 4 Trp codons were followed by an A (i.e., TGG A), and can therefore also be potentially turned into TGA stop codons by deamination in a 5'TC context. All Env-eGFP clones mutated by A3G contained at least two stop codons that completely abrogated eGFP fluorescence (Fig. 3b and Table S4). A3F, on the other hand, inhibited mainly through NS substitutions that resulted in reduced reporter fluorescence, although 4 clones contained a stop codon at the same position in the N-terminal region of Env (Fig. 3b and Table S2).

DISCUSSION

It is now well documented that human A3F and A3G are potent antiretroviral restriction factors that inhibit HIV-1 infection when the Vif protein is absent or expressed at low levels. It is also clear that when A3 proteins are expressed at near-physiological levels the main mechanism for the restriction is their ability to lethally mutate retroviral DNA replication intermediates (Albin *et al.*, 2014; Browne *et al.*, 2009; Kobayashi *et al.*, 2014; Miyagi *et al.*, 2007). The mutator activities of A3F and A3G can therefore be involved in shaping part of the early repertoire of HIV-1 viruses produced during a natural infection in humans when Vif levels are low enough to allow A3 virion incorporation (Feng *et al.*, 2014). In addition, A3 mutation intensity is strongly influenced by the position of a gene within a retroviral genome, as twin 3' to 5' mutational gradients have been described for HIV-1 and M-MLV (Belanger *et al.*, 2014; Suspene *et al.*, 2006; Yu *et al.*, 2004). Here we were interested in comparing the gene-inactivating potential of A3F and A3G, especially in light that these two enzymes have different DNA target specificities and biochemical processes for deaminating cytidines in single-stranded DNA (Ara *et al.*, 2014). We focused our interest on using the eGFP reporter protein as a model because it offers an equivalent distribution of 5'TC and 5'CC putative deamination targets. This thereby provided us with the ability to directly compare how DNA target specificity impacts gene inactivation and retroviral sequence evolution. This is an issue of particular importance in light that A3F, more than A3G, has been shown to promote HIV-1 diversification and evolution *in vivo* (Ara *et al.*, 2014; Armitage *et al.*, 2012; Armitage *et al.*, 2014; Kim *et al.*, 2014; Sato *et al.*, 2014; Wood *et al.*, 2009). However, the extent to which this diversification correlates with alterations in gene function had not been assessed until now.

Biochemical analyses have previously demonstrated that A3G is a more potent cytosine deaminase than A3F, in that purified A3G is able to deaminate more cytidines in a single-stranded DNA substrate within a defined unit of time (Ara *et al.*, 2014). These assays, however, were conducted using DNA oligos with containing the preferred consensus target site for each enzyme. Here we revisited this issue using a retroviral restriction assay and analyzing a sequence offering nearly identical 5'TC and 5'CC target sites. Our results also support that A3G is a more potent deaminase by introducing approximately 34% more G-to-A mutations in the eGFP gene than A3F (178 mutations for A3G compared to 118 for A3F) (Table 2). Despite the ability of both deaminases to reduce eGFP fluorescence was remarkably similar, A3G generated 20% more sequences that did not encode for a fluorescent protein (Fig. 3 and

Table 3). This result is in line with our observation that A3G produced larger numbers and a greater proportion of NS substitutions in the gene coding for eGFP than A3F (Table 4).

Our analysis of mutations in the full-length Env-eGFP gene yielded quite contrasting results. We found that A3G was prominently more effective in mutating and inactivating reporter fluorescence than A3F (Fig. 3 and Table 2). This was primarily due to a combination of factors, notably a larger number of overall 5'CC dinucleotides than 5'TC, and also, more importantly, a large number of Trp codons in the sequence. In fact, every A3G-mutated Env-eGFP clone analyzed contained at least 2 stop codons upstream of the eGFP coding sequence (Fig. 3 and Table 3). Only four A3F clones contained a termination codon in the N-terminus of the Env gene, all of which at position W48 (Fig. 2c and Table S2). The termination codon generated in this case is TGA, thereby indicating that it is the result of A3F deaminating in the less favorable 5'CC context. This shows that although A3G is much more efficient at generating stop codons, A3F can also target the 5'CC dinucleotide opposite to the TGG of the Trp codon, but at much lower frequencies. Our assays also revealed that A3F targets 5'CC in 13% of the compiled mutations (Fig. 2c). Although A3F was less efficient than A3G at inactivating eGFP fluorescence, it did however introduce similar NS substitutions in the Env-eGFP gene (Table 4). With a lower gene inactivation potential and a high rate of NS substitution generation, our results support that A3F is much more effective than A3G at promoting viral sequence evolution. This is also consistent with previous findings that employed other analysis methods and systems (Ara *et al.*, 2014; Armitage *et al.*, 2012; Armitage *et al.*, 2014; Kim *et al.*, 2014; Sato *et al.*, 2014; Wood *et al.*, 2009).

Through this study we also identified a number of gene-inactivating mutations in the eGFP gene caused by DNA deamination. The most potent mutations that completely abolished fluorescence are E133K and E223K; the latter has been found to define the charged state on the β-barrel chromophore structure of eGFP and is critical for fluorescence (Arpino *et al.*, 2012). Both these substitutions were caused by the deamination of a glutamate codon (GAA) into a lysine (AAA) in a 5'TC context by both A3F and A3G. Other residues such as E214K (5'TC context) and M219I (5'CC context) also significantly reduced fluorescence, but were influenced by the presence of other mutations. For example, the E214K mutation reduced fluorescence to 1.5% when present alone (clone A3F-5), but fluorescence was restored to 74% when accompanied with a S209N mutation (clone A3F-15). This emphasizes the epistatic-like effect of mutations when A3 deaminates proviral DNA. But it also underlines a possible shortcoming when using eGFP fluorescence intensity to measure infection by a reporter virus. Another

interesting observation is that some mutations outside the eGFP coding sequence also influenced fluorescence. Clones Env-A3F-2, Env-A3F-4 and Env-A3F-7 do not contain NS substitutions in the eGFP sequence but display reduced fluorescence (Fig. 3b and Tables 5, S2 and S7). These examples suggest that A3 deamination events may have effects not only on protein function but also on various other parameters such as mRNA translation efficiency, and protein and mRNA stability. Determining the exact cause of reduced fluorescence for each affected clone is beyond the scope of this study.

TZM-bl and CEM-GFP indicator cell lines are laboratory workhorses frequently used to monitor HIV-1 infection and the restriction potential of various A3 proteins. Such indicator cells require Tat and or Rev protein expression from the infecting provirus for cell-encoded HIV-1 LTR-mediated reporter gene transcription and expression. Interestingly, all HIV-1 genes contain several Trp residues, however, early genes that code for Tat and Rev only contain one (Table S9). As such, these HIV reporter cells rely on the expression of these early genes, which may not be inactivated through stop codon generation if the intensity of hypermutation is low and especially if the investigated A3 proteins deaminate with a preferred 5'TC consensus. This could result in viral restriction being largely underestimated. Because most human A3 proteins other than A3G target preferentially 5'TC dinucleotides, it would be more advisable to measure inhibition of proviral DNA integration or p24 particle release from infected cells to compare the restriction efficiency of various A3 proteins.

Here we have functionally demonstrated that A3G is more efficient at mutating retroviral DNA than A3F when given equal numbers of preferred putative deamination targets. But more importantly, we have also shown that A3G will potently inactivate genes containing Trp codons through stop codon generation even when the mutation burden is very low, which was not the case for A3F. Therefore, A3F, like all other A3 proteins that prefer 5'TC deamination targets, is more amenable to contribute to retroviral sequence evolution than restriction, especially when the mutation burden is low. The ability to diversify the genetic code of retroviruses without inactivating viral genes further strengthens the possible implications of some A3 proteins in immune evasion, but also in retroviral drug resistance. This may become a special concern for the new HIV-1 therapies comprised of highly neutralizing antibodies that bind to viral epitopes that could be putative A3 deamination targets.

METHODS

Cells. HEK 293T Human embryonic kidney epithelium (293T) and NIH mouse embryonic fibroblasts (NIH 3T3) were cultured in HyClone DMEM/High Glucose medium supplemented with 10% fetal bovine serum, 100U/ml penicillin and 100µg/ml streptomycin.

Plasmids. Flag-tagged A2, A3F and A3G expression plasmids were constructed previously (Langlois *et al.*, 2005). The single-cycle HIV[p8.9] pseudovirus expressing eGFP and the replicative Moloney murine leukemia virus (M-MLV) expressing eGFP as an Env-eGFP fusion have been previously described (Bélanger *et al.*, 2013; Langlois *et al.*, 2009). HIV[p8.9] (referred to as HIV-1 ΔVif pseudovirus in the text) expresses eGFP from an internal spleen focus-forming virus promoter (SFFV) (Langlois *et al.*, 2005). M-MLV expresses eGFP from the viral LTR promoter as a nested fusion protein with Env where it is inserted in the proline-rich region of the protein (Sliva *et al.*, 2004).

Viral mutator assays, reporter gene cloning and sequencing. Infections using HIV-1 ΔVif pseudoviruses and M-MLV produced in presence of A3F or A3G were carried out as previously described (Bélanger *et al.*, 2013). A3F and A3G expression in virus producing cells were monitored by Western blot and were consistently equal in each transfection assay (Fig. 1). Viral titers were normalized by p24 or p30 ELISA. Target cells were infected with the same p24 or p30 amounts that yield a MOI of 0.5 as when the viruses are produced in presence of the A2 negative control (Bélanger *et al.*, 2013). The restriction profile of each A3 was monitored 24h after infection by measuring eGFP fluorescence in target cells by flow cytometry analysis.

For gene inactivation analysis, the gDNA of 293T cells infected with HIV-1 ΔVif or NIH 3T3 cells infected with M-MLV was extracted and purified. PCR was used to amplify eGFP (HIV-1 ΔVif) or Env-eGFP (M-MLV) from integrated proviral sequences. The following primers were used for the PCR; eGFP-FWD-Nhe1: 5'-GCTAGCTATCCACCGGTCGCCACCAT-'3, eGFP-REV-HindIII: 5'-AAGCTTGCAGGTCGACTCTAGAGT-'3, ENV-eGFP-FWD-Nhe1: 5'-GCTAGCATGGCGCGTTCAACG-'3, ENV-eGFP-REV-HindIII: 5'-AAGCTTTATGGCTCGTACTCTATAGG-'3. PCR products were then digested and cloned into the pcDNA3.1 expression vector (Invitrogen) downstream of the CMV promoter using the NheI and HindIII restriction sites. In total, 42 eGFP clones derived from HIV-1 ΔVif (20 mutated by A3F and 20 by A3G; and one unmutated clone for each); and 42 Env-eGFP clones derived from M-

MLV (20 mutated by A3F and 20 by A3G; and one unmutated clone for each) were randomly selected from a small library of clones that were sequenced. We ensured that each mutated clone selected displayed a unique mutational profile.

eGFP reporter assay. eGFP and Env-eGFP clones were co-transfected along with a transfection control plasmid into 293T cells in triplicate in three independent experiments. Cells were harvested by trypsinization (0.25% (w/v) Trypsin, 1 mM EDTA) 48h post transfection, washed twice in PBS and analyzed by flow cytometry on a Beckman Coulter CyAN ADP flow cytometer. Transfection efficiency for all clones was consistently between 75% and 90% as determined by the reporter control. Mean eGFP fluorescence was assessed on the gated transfected cell population delineated by the reporter control. Mutated clones with fewer than 5% of gated cells expressing mean eGFP fluorescence above background were registered as having a relative mean fluorescence of 0. Data is presented as the relative mean eGFP fluorescence of the mutated clones compared to the wild type eGFP or Env-eGFP reporter for three independent experiments ± standard deviation (SD).

ACKNOWLEDGEMENTS

The authors thank Laura Goodwin for technical assistance, Maria Rosales Gerpe and Tyler Renner for helpful discussions, and Kristin Kemmerich for comments on the manuscript. We are grateful to the NIH AIDS Research and Reference Reagent Program for reagents. K.B. holds an Ontario Graduate Scholarship and a scholarship from le Fonds de Recherche en Santé du Québec. M.-A.L. holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity. This research was supported by grant #89774 from the Canadian Institutes of Health Research, and an Early Researcher Award from the Ontario Ministry of Research and Innovation to M.-A.L.

REFERENCES

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

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

Armitage, A. E., Deforche, K., Chang, C. H., Wee, E., Kramer, B., Welch, J. J., Gerstoft, J., Fugger, L., McMichael, A., Rambaut, A. & Iversen, A. K. (2012). APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. *PLoS genetics* **8**, e1002550.

Armitage, A. E., Deforche, K., Welch, J. J., Van Laethem, K., Camacho, R., Rambaut, A. & Iversen, A. K. (2014). Possible Footprints of APOBEC3F and/or Other APOBEC3 Deaminases, but Not APOBEC3G, on HIV-1 from Patients with Acute/Early and Chronic Infections. *Journal of virology* **88**, 12882-12894.

Arpino, J. A., Rizkallah, P. J. & Jones, D. D. (2012). Crystal structure of enhanced green fluorescent protein to 1.35 A resolution reveals alternative conformations for Glu222. *PloS one* **7**, e47132.

Belanger, K., Savoie, M., Aydin, H., Renner, T. M., Montazeri, Z. & Langlois, M. A. (2014). Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis. *Virology* **448**, 168-175.

Bélanger, K., Savoie, M., Rosales Gerpe, M. C., Couture, J. F. & Langlois, M. A. (2013). Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. *Nucleic acids research* **41**, 7438-7452.

Bogerd, H. P., Wiegand, H. L., Doehle, B. P. & Cullen, B. R. (2007). The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology* **364**, 486-493.

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

Chelico, L., Pham, P., Calabrese, P. & Goodman, M. F. (2006). APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nature structural & molecular biology* **13**, 392-399.

Derdeyn, C. A., Decker, J. M., Sfakianos, J. N., Wu, X., O'Brien, W. A., Ratner, L., Kappes, J. C., Shaw, G. M. & Hunter, E. (2000). Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *Journal of virology* **74**, 8358-8367.

Desimmie, B. A., Delviks-Frankenberrry, K. A., Burdick, R. C., Qi, D., Izumi, T. & Pathak, V. K. (2014). Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *Journal of molecular biology* **426**, 1220-1245.

Feng, Y., Baig, T. T., Love, R. P. & Chelico, L. (2014). Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. *Frontiers in microbiology* **5**, 450.

Fourati, S., Lambert-Niclot, S., Soulie, C., Malet, I., Valantin, M. A., Descours, B., Ait-Arkoub, Z., Mory, B., Carcelain, G., Katlama, C., Calvez, V. & Marcelin, A. G. (2012a). HIV-1 genome is often defective in PBMCs and rectal tissues after long-term HAART as a result of APOBEC3 editing and correlates with the size of reservoirs. *The Journal of antimicrobial chemotherapy* **67**, 2323-2326.

Fourati, S., Malet, I., Lambert, S., Soulie, C., Wirden, M., Flandre, P., Fofana, D. B., Sayon, S., Simon, A., Katlama, C., Calvez, V. & Marcelin, A. G. (2012b). E138K and M184I mutations in HIV-1 reverse transcriptase coemerge as a result of APOBEC3 editing in the absence of drug exposure. *Aids* **26**, 1619-1624.

Gervaix, A., West, D., Leoni, L. M., Richman, D. D., Wong-Staal, F. & Corbeil, J. (1997). A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4653-4658.

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

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

Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L. & Harris, R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *Journal of virology* **85**, 11220-11234.

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

Kim, E. Y., Lorenzo-Redondo, R., Little, S. J., Chung, Y. S., Phalora, P. K., Maljkovic Berry, I., Archer, J., Penugonda, S., Fischer, W., Richman, D. D., Bhattacharya, T., Malim, M. H. & Wolinsky, S. M. (2014). Human APOBEC3 Induced Mutation of Human Immunodeficiency Virus Type-1 Contributes to Adaptation and Evolution in Natural Infection. *PLoS pathogens* **10**, e1004281.

Kobayashi, M., Takaori-Kondo, A., Miyauchi, Y., Iwai, K. & Uchiyama, T. (2005). Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. *The Journal of biological chemistry* **280**, 18573-18578.

Kobayashi, T., Koizumi, Y., Takeuchi, J. S., Misawa, N., Kimura, Y., Morita, S., Aihara, K., Koyanagi, Y., Iwami, S. & Sato, K. (2014). 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.

Koito, A. & Ikeda, T. (2013). Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Frontiers in microbiology* **4**, 28.

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

Langlois, M.-A., Kemmerich, K., Rada, C. & Neuberger, M. S. (2009). The AKV Murine Leukemia Virus Is Restricted and Hypermutated by Mouse APOBEC3. *Journal of virology* **83**, 11550-11559.

Liu, B., Yu, X., Luo, K., Yu, Y. & Yu, X. F. (2004). Influence of primate lentiviral Vif and proteasome inhibitors on human immunodeficiency virus type 1 virion packaging of APOBEC3G. *Journal of virology* **78**, 2072-2081.

Love, R. P., Xu, H. & Chelico, L. (2012). Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. *The Journal of biological chemistry* **287**, 30812-30822.

Mehle, A., Strack, B., Ancuta, P., Zhang, C., McPike, M. & Gabuzda, D. (2004). Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *The Journal of biological chemistry* **279**, 7792-7798.

Miyagi, E., Opi, S., Takeuchi, H., Khan, M., Goila-Gaur, R., Kao, S. & Strebel, K. (2007). Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1. *Journal of virology* **81**, 13346-13353.

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

Monajemi, M., Woodworth, C. F., Zipperlen, K., Gallant, M., Grant, M. D. & Larijani, M. (2014). Positioning of APOBEC3G/F mutational hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells. *PloS one* **9**, e93428.

Mulder, L. C., Harari, A. & Simon, V. (2008). Cytidine deamination induced HIV-1 drug resistance. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5501-5506.

Neogi, U., Shet, A., Sahoo, P. N., Bontell, I., Ekstrand, M. L., Banerjea, A. C. & Sonnerborg, A. (2013). Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals. *Journal of the International AIDS Society* **16**, 18472.

Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. (1998). Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *Journal of virology* **72**, 2855-2864.

Rausch, J. W., Chelico, L., Goodman, M. F. & Le Grice, S. F. (2009). Dissecting APOBEC3G substrate specificity by nucleoside analog interference. *The Journal of biological chemistry* **284**, 7047-7058.

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

Sato, K., Takeuchi, J. S., Misawa, N., Izumi, T., Kobayashi, T., Kimura, Y., Iwami, S., Takaori-Kondo, A., Hu, W. S., Aihara, K., Ito, M., An, D. S., Pathak, V. K. & Koyanagi, Y. (2014). APOBEC3D and APOBEC3F potently promote HIV-1 diversification and evolution in humanized mouse model. *PLoS pathogens* **10**, e1004453.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D. D. & Bieniasz, P. D. (2005). Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS pathogens* **1**, e6.

Sliva, K., Erlwein, O., Bittner, A. & Schnierle, B. S. (2004). Murine leukemia virus (MLV) replication monitored with fluorescent proteins. *Virology journal* **1**, 14.

Strebel, K. (2013). HIV accessory proteins versus host restriction factors. *Current opinion in virology* **3**, 692-699.

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

Wood, N., Bhattacharya, T., Keele, B. F., Giorgi, E., Liu, M., Gaschen, B., Daniels, M., Ferrari, G., Haynes, B. F., McMichael, A., Shaw, G. M., Hahn, B. H., Korber, B. & Seoighe, C. (2009). HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS pathogens* **5**, e1000414.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J. M. & Landau, N. R. (2004). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nature structural & molecular biology* **11**, 435-442.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P. & Yu, X. F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-1060.

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

TABLES

Table 1. Total number of putative A3 deamination target sites in the genes coding for eGFP and Env.

Sequence motifs	eGFP[b]	N-terminal Env[b]	C-terminal Env[b]	Env-eGFP (Complete sequence)[b]
5'-**TC**[a]	52	45	70	167
5'-**CC**[a]	51	72	70	193
TGG (*Trp*) codons	1	13	5	19
Trp followed by a G (TGG.G)	0	6	1	7
Trp followed by an A (TGG.A)	0	4	0	4

[a] Cytosines in bold are putative deamination.
[b] Target motifs tabulated on the retroviral minus DNA strand.

Table 2. Comparative analysis of the mutations induced by A3F and A3G.

A3	Deaminated sequence	Clones analyzed	Mutated clones	Base pairs analyzed in mutated clones	Total number of G-to-A mutations	Mutation rate (mutations/ Kb)
A3F	eGFP[a]	21	20	14400	118	8.2
A3G	eGFP	21	20	14400	178	12.4
A3F	Env-eGFP[b]	21	20	54060	293	5.4
A3G	Env-eGFP	21	20	54060	970	18.0

[a] The eGFP sequence is 720 bp in length. [b]Env-eGFP is 2703 bp.

Table 3. Gene inactivating potential of A3F and A3G

A3	Deaminated sequence	Mutated clones	Mutated clones with reduced fluorescence[a]	Mean relative fluorescence[b]	Mutated clones with no fluorescence[c]	Clones with stop codons	Total number of stop codons
A3F	eGFP	20	7 (35%)	47	6 (30%)	0	0
A3G	eGFP	20	6 (30%)	45	10 (50%)	0	0
A3F	Env-eGFP	20	8 (40%)	55	11 (55%)	4 (20%)	6
A3G	Env-eGFP	20	0	-	20 (100%)	20 (100%)	152

[a] Clones with reduced fluorescence are defined as clones displaying less than 85% of the mean fluorescence of the unmutated control but more than background. [b] Mean relative fluorescence was calculated for mutated clones with reduced fluorescence compared to the control. [c] Clones were designated as having no fluorescence (background) if their mean fluorescence was below 3% of the control.

Table 4. Genetic diversification caused by A3F and A3G

A3	Deaminated sequence	Mutated clones	Total codons deaminated	NS in GFP	NS total	NS (%)
A3F	eGFP	20	118	38	38	32
A3G	eGFP	20	184	83	83	45
A3F	Env-eGFP	20	283	24	201	71
A3G	Env-eGFP	20	870	60	567	65

NS: Non-synonymous substitutions.

Table 5. Amino acid substitutions outside of the eGFP sequence that influence fluorescence

			Substitution type and position in Env-eGFP						
			N-terminal Env		C-terminal Env				
Clones with reduced eGFP fluorescence but no mutations in egfp	Total number of G-to-A mutations	Mean relative fluorescence	E119K	D211N	G722E	E755K	L814L	M858I	C865Y
Env-A3F-2	1	74						X	
Env-A3F-4	2	44	X			X			
Env-A3F-7	4	66		X	X		X		X

FIGURE LEGENDS

Fig. 1. Restriction of M-MLV and HIV-1 ΔVif by A3F and A3G. (A) M-MLV or (B) HIV-1 ΔVif viral particles were produced in 293T cells in the presence of Flag-tagged APOBEC proteins and used to infect NIH 3T3 or 293T target cells respectively. Retroviral restriction was measured by eGFP expression in infected cells by flow cytometry 24 h after infection. Results represent the mean ±SD of triplicate values from three independent transfection experiments. (C) and (D), Western blots showing the expression of Flag-tagged A3F and A3G, respectively, in virus producer cells from one representative experiment. The expression of β-tubulin was used as a loading control.

Fig. 2. Codons in the gene coding for Env-eGFP that are affected by A3F and A3G deamination. (A) Schematic representation of the eGFP or ENV-eGFP genes cloned into the pcDNA3.1 expression vector. The green triangle represents the start codon (ATG), and the red square represents termination codon (TAG). (B) Graphical representation of all the codons in Env-eGFP affected by a deamination in a 5'TC (*top*) or 5'CC (*bottom*) dinucleotide context. The total number (n) of affected codons is represented to the right of the graph. (C) Positions of all the codons targeted by A3F deamination (*top*) and codons in a 5'TC context that were not mutated (*bottom*). Compiled mutations for all A3F and Env-A3F clones are represented. Red bars indicate a deamination event causing a termination codon. The total number (n) of mutated and non-mutated codons is indicated to the right. The proportion (%) of codons mutated in a 5'TC or 5'CC dinucleotide context is also indicated. (D) Similar analysis than in (C) but for viruses produced in presence of A3G.

Fig. 3. Fluorescence of mutated eGFP and Env-eGFP reporter proteins. (A) Relative mean fluorescence intensity (MFI) of the eGFP reporter protein mutated by A3F (*left*) and A3G (*right*). Relative MFI (green bars) and the total number of G-to-A mutations (black bars) are indicated for each clone. Light blue bars indicate the number of NS amino acid substitutions in the eGFP sequence of the clone. Red hexagons indicate the presence of one or more stop codons in a clone (the exact number is specified inside the hexagon); Asterisks (*) indicate clones with determined eGFP-inactivating mutations other than stop codons. (B) Relative MFI of the Env-eGFP reporter protein mutated by A3F (*left*) and A3G (*right*). Only mutations in N-terminal Env and eGFP are displayed (N-Ter.). Fluorescence results represent MFI ± SD normalized to a non-mutated plasmid control and were compiled from at least three independent transfections.

Figure 1
Click here to download Figure: Figure 1_JGV.tif

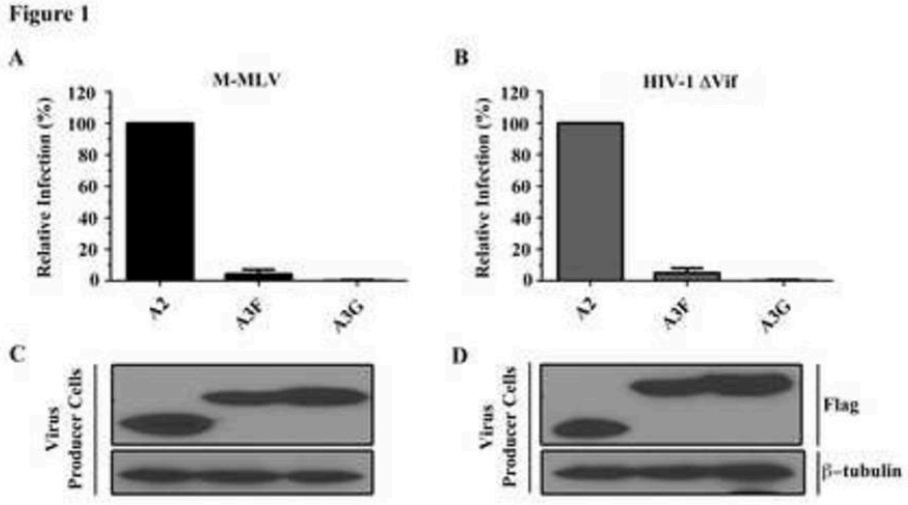


Figure 2
Click here to download Figure: Figure 2_JGV.tif

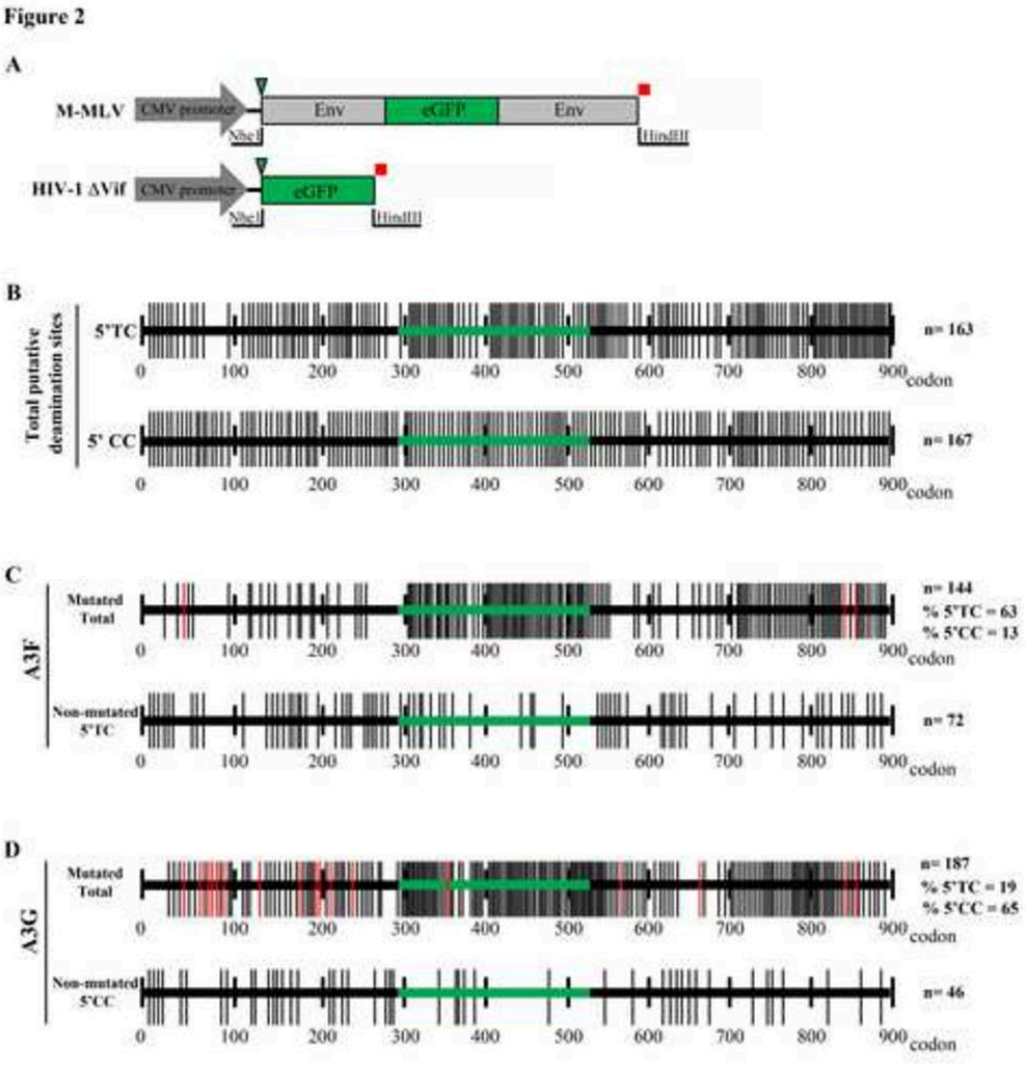
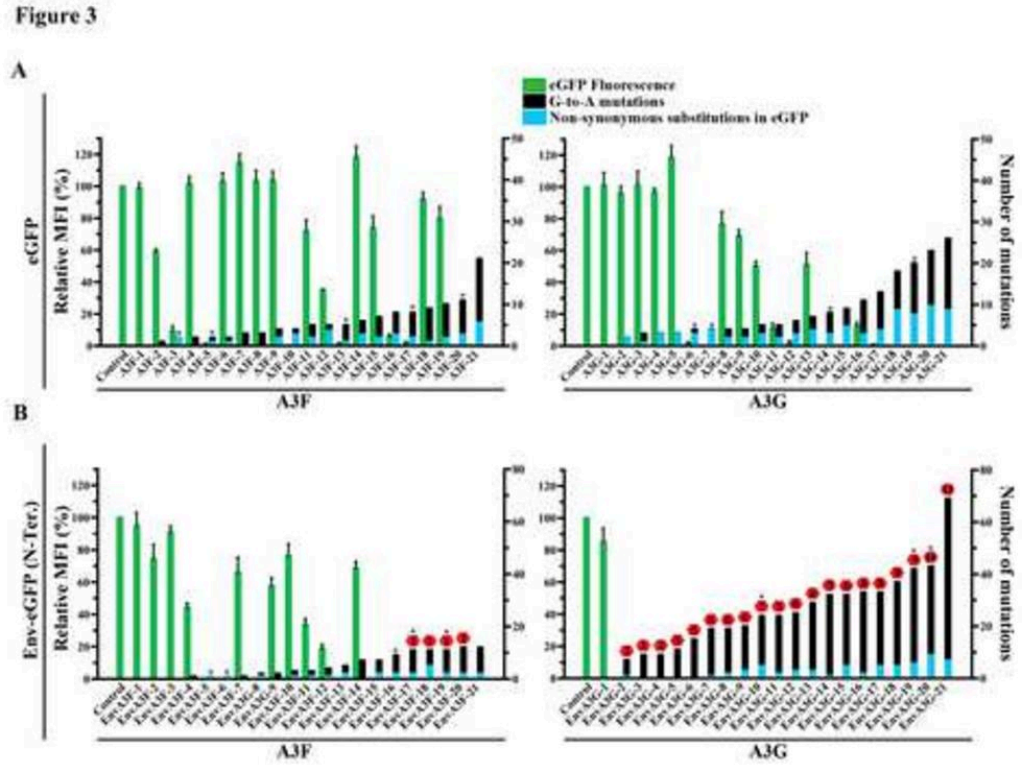


Figure 3
Click here to download Figure: Figure 3_JGV.tif



Appendix VII: Kasandra Bélanger, Curriculum Vitae

Kasandra Bélanger

Education

Doctor of Philosophy (**Ph.D.),** Biochemistry, Microbiology and Immunology
Under the supervision of Dr. Marc-André Langlois
University of Ottawa – Ottawa, ON January 2011- present
Thesis submitted: November 2015. Defense: February 2016.
Thesis title: Investigation of the various modes of retroviral and endogenous retroelements restriction by APOBEC3 proteins.

Master's of Science (**M.Sc.),** Biochemistry, Microbiology and Immunology ***(Transfer to Ph.D.)***
Under the supervision of Dr. Marc-André Langlois
University of Ottawa – Ottawa, ON Sept 2009 – January 2011

Bachelor of Science (**B.Sc.),** Honours in Cellular and Molecular Biology
Under the supervision of Dr. Charles-Antoine Darveau
University of Ottawa – Ottawa, ON January 2006 – May 2009
Thesis title: Intraspecific variation in metabolic rate of *Bombus impatiens*.

Expertise

- Molecular and Cellular biology
- Virology (human, murine and avian retroviruses)
- Immunology (innate immunity, host restriction factors)

Awards and Achievements

1. Award of Excellence in Graduate Studies for the Ph.D. in Microbiology and Immunology program, (500$), December 2015.
2. Poster Presentation Award, Canadian Society of Immunology (CSI), (100$), Winnipeg, June 2015.
3. Travel Award, Canadian Society of Immunology (CSI), (1000$), Winnipeg, June 2015.
4. Best poster presentation by a Ph.D. student, Biochemistry, Microbiology and Immunology (BMI) Annual Poster Day (75$), 2014.
5. Fonds de Recherche en Santé du Québec (FRSQ), (20 000$/yr, 3 years), April 2013-present.
6. Young Investigator Award, Conference on Retroviruses and Opportunistic Infections (CROI), Atlanta, USA, March 2013.

7. Ontario Graduate Scholarship (OGS) (15 000$), April 2013-April 2014.
8. University of Ottawa Excellence Scholarship (7,500$), 2012-2013.
9. CIHR Travel Award, Canadian Student Health Research Forum, (1,000$), 2011.
10. Best presentation by a Ph.D. student, Biochemistry, Microbiology and Immunology (BMI) Annual Symposium, (100$), 2011.
11. Ontario Graduate Scholarship in Science and Technology (OGSST), (12,000$), 2009-2010.
12. Dean's Honor's list, University of Ottawa, 2007-2009.

Summary of Published Work

Type of Publication	Number of publication
First author manuscript	4
Co-author manuscript	2
Posters	15
Oral presentations	6

Key research skills

Molecular biology

- Development of a novel high throughput method (96 and 384 well plate) to identify mutations in genomic DNA. This application has potential clinical relevance for personalized treatment plan of HIV patients.
- PCR (2 and 3-step PCR, RT-PCR, real-time PCR, digital droplet PCR and high resolution melt).
- DNA, RNA and protein extraction from tissues.
- Molecular cloning (primer design, digestion with restriction enzymes, ligation and transformation).
- Western and Southern blotting analyses.
- Co-immunoprecipitation.
- ELISA (p24 and p30 viral capsid proteins).
- Velocity sedimentation (sucrose and optiprep gradients).

Cellular biology

- Level 2 and 3 Biocontainment work.
- HEK 293T, NIH 3T3, CEM, CEM-SS, Jurkat and DF-1 cell culture.
- Isolation and culture of immune cell population (T lymphocytes) from human blood.
- Fluorescence microscopy.
- Transfection (lipogenic reagent and $CaCl_2$).
- Virus production and infection with replicative or self-inactivating viruses.

Flow cytometry

- Monitoring of transfection and infection with viral vectors expressing an eGFP reporter gene.
- Multi-color flow cytometry.

Leadership

- Design of experiments for research projects in molecular and cellular biology (review of the literature, selection of the most appropriate methodologies, optimization).
- Supervision of undergraduate students (design of experiments for their project, analysis of their results, trouble-shooting, revision of their scientific reports and presentations).
- Writing and redaction of manuscripts for publication in peer-reviewed journals.
- Lab demonstrator for the Biochemistry undergraduate course (demonstration of laboratory techniques, explanation of biochemistry concepts, grading of students' reports).

List of Publications

Manuscripts Published-Peer Reviewed

1. **Kasandra Bélanger** and Marc-André Langlois. **(2015)** Comparative Analysis of the Gene-Inactivating Potential of Human Retroviral Restriction Factors APOBEC3F and APOBEC3G. ***Journal of General Virology***, 96, 2878-2887.

2. **Kasandra Bélanger** and Marc-André Langlois. **(2015)** RNA-binding residues in the N-terminus of APOBEC3G influence its DNA sequence specificity and retrovirus restriction efficiency. ***Virology***, 483, 141-148. [*citations: 1]
   **This article was selected by the editor to be one of the highlighted manuscripts in volume 483 (September issue) of the journal.

3. María Carla Rosales Gerpe, Tyler Renner, **Kasandra Bélanger**, Cindy Lam, Halil Aydin and Marc-André Langlois. **(2015)** N-linked glycosylation protects gammaretroviruses against deamination by APOBEC3 proteins. **Journal of Virology**, 89, 2342-2357. [*citations: 2]

4. **Kasandra Bélanger**, Mathieu Savoie, Maria Rosales Gerpe, Jean-François Couture and Marc-André Langlois. **(2013)** Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. ***Nucleic Acid Research***, 41, 7438-7452. [*citations: 22]
5. **Kasandra Bélanger**, Mathieu Savoie, Halil Aydin, Tyler Renner, Zahra Montazeri and Marc-André Langlois. **(2013)** Deamination intensity profiling of human APOBEC3 protein activity along the near full-length genomes of HIV-1 and MoMLV by HyperHRM analysis. ***Virology***, 448, 168-175. [*citations: 5]

6. Charles-Antoine Darveau, Fannie Billardon and **Kasandra Bélanger**. **(2013)** Intraspecific variation in flight metabolic rate in the bumblebee Bombus impatiens: repeatability and

functional determinants in workers and drones. ***Journal of Experimental Biology***.10, 1242. [*citations: 6]

**citations according to Google Scholar; December 2015*

Manuscripts in Preparation to be Submitted (< 2 months)

1. **Kasandra Bélanger**, Tyler Renner, Cindy Lam, María Carla Rosales Gerpe and Marc-André Langlois. The glycosylated gag protein of Moloney murine leukemia virus inhibits the antiretroviral activity of A3 proteins by inferring a physical blockade prohibiting their deamination activity. (To be submitted to the *Journal of Virology*, December 2015).

2. **Kasandra Bélanger**, Laura Rose Goodwin, Mark Campbell and Marc-André Langlois. RNA binding is required for the inhibition of LINE-1 retrotransposition by the APOBEC3 proteins. (To be submitted to *Plos One*, January 2016).

Posters *(presenting author underlined)*

National and International Conferences:

1. **Kasandra Bélanger** and Marc-André Langlois. DNA Target Specificity for deamination by the APOBEC3 Proteins influence their potency for HIV-1 gene inactivation. Canadian Society of Immunology (CSI), Winnipeg, June 2015.

2. **Kasandra Bélanger** and Marc-André Langlois. DNA Target Specificity for deamination by the APOBEC3 Proteins influence their potency for HIV-1 gene inactivation. Canadian Association of HIV Research (CAHR), Toronto, April 2015.

3. **Kasandra Bélanger** and Marc-André Langlois. Does Sublethal Mutagenesis Caused by APOBEC3 Proteins Contribute to the Establishment of Latent HIV Reservoirs? CanCure meeting, Montréal, November 2014.

4. **Kasandra Bélanger**, Mathieu Savoie, Halil Aydin, Zahra Montazeri and Marc-André Langlois. Comparative profiling of deamination rates and hotspots on vif-deficient HIV-1 and MoMLV proviral DNA by all seven human apobec3 family members. Cold Spring Harbor Laboratory – Retrovirus meeting, New York, May 2013.

5. **Kasandra Bélanger**, Mathieu Savoie, Jean-François Couture and Marc-André Langlois. RNA-Binding properties of APOBEC3G control deamination-independent restriction of retroviruses. Cold Spring Harbor Laboratory – Retrovirus meeting, New York, May 2013.

6. **Kasandra Bélanger**, Mathieu Savoie, Halil Aydin, Zahra Montazeri and Marc-André Langlois. Comparative profiling of deamination rates and hotspots on vif-deficient HIV-1 and MoMLV proviral DNA by all seven human apobec3 family members. Conference on retroviruses and opportunistic infections (CROI), Atlanta, USA, March, 2013.

7. **Kasandra Bélanger**, Mathieu Savoie, Jean-François Couture and Marc-André Langlois. RNA-Binding properties of APOBEC3G control deamination-independent restriction of retroviruses. Conference on retroviruses and opportunistic infections (CROI), Atlanta, USA, March, 2013.

8. **Kasandra Bélanger** and Marc-André Langlois. Regulation of APOBEC3G-mediated intrinsic immunity to HIV infection. Canadian Student Health Research Forum, Winnipeg, Manitoba, June 2011.

9. Dimitri A. Skandalis, **Kasandra Bélanger**, Taehyoung Lee and Charles-A Darveau. Maturation of flight in the adult bumblebee (Bombus impatiens). Canadian Society of Zoologists, Toronto, ON, May 2009.

University of Ottawa:

1. Mark Campbell, **Kasandra Bélanger** and Marc-André Langlois. Structural determinants of LINE-1 and Alu retrotransposon restriction by the family of APOBEC3 proteins. Poster, Undergraduate Students Poster Day, ON, April 2014.

2. **Kasandra Bélanger**, Mathieu Savoie and Marc-André Langlois. Deamination is not sufficient for APOBEC3G-mediated retroviral restriction. Pathology Research Day, University of Ottawa, ON, April 2012.

3. Mathieu Savoie, **Kasandra Bélanger** and Marc-André Langlois. Intrinsic immunity to retroviral infection: Mouse APOBEC3 as a model ortholog for human APOBEC3G. Poster, Undergraduate Students Poster Day, ON, April 2012.

4. Maria Rosales Gerpe, Halil Aydin, **Kasandra Bélanger** and Marc- André Langlois. Developing an in vitro detection method for Xenotropic Murine Leukemia Virus-Related Virus (XMRV). Poster, Undergraduate Students Poster Day, University of Ottawa, ON, April 2011.

5. **Kasandra Bélanger** and Marc-André Langlois. Regulation of APOBEC3G-mediated intrinsic immunity to HIV infection. Pathology Research Day, University of Ottawa, ON, April 2011.

6. **Kasandra Bélanger** and Marc-André Langlois. Mechanisms for APOBEC3G-mediated intrinsic immunity to HIV infection. Pathology Research Day, University of Ottawa, ON, April 2009.

Oral presentations *(presenting author underlined)*

University of Ottawa:

1. **Kasandra Bélanger**, Mathieu Savoie and Marc-André Langlois. Mechanisms for A3G-mediated intrinsic immunity to HIV-1 retroviral infection. Poster. Biochemistry, Microbiology and Immunology (BMI), Annual Symposium, University of Ottawa, ON, May 2014.

2. **Kasandra Bélanger**, Mathieu Savoie and Marc-André Langlois. Mechanisms for A3G-mediated intrinsic immunity to HIV-1 retroviral infection. Biochemistry, Microbiology and Immunology (BMI), Annual Symposium, University of Ottawa, ON, March 2013.

3. **Kasandra Bélanger**, Mathieu Savoie and Marc-André Langlois. Deamination is not sufficient for APOBEC3G-mediated retroviral restriction. Biochemistry, Microbiology and Immunology (BMI), Annual Symposium, University of Ottawa, ON, May 2012.

4. Shabnam Rahimi-Khameneh, **Kasandra Bélanger**, Olga Agah and Marc-André Langlois. Identification of a potential retroviral component associated with prion disease. Biochemistry, Microbiology and Immunology (BMI), Annual Symposium, University or Ottawa, ON, May 2011.

5. **Kasandra Bélanger** and Marc-André Langlois. Mechanisms for APOBEC3G-mediated intrinsic immunity to HIV infection. Biochemistry, Microbiology and Immunology (BMI), Annual Symposium, University of Ottawa, ON, May 2009.

6. Halil Haydin, **Kasandra Bélanger** and Marc-André Langlois. Assessment and Analysis of the Restriction of Retroviral Infection by the Murine APOBEC3 Protein. Biochemistry, Microbiology and Immunology (BMI), Annual Symposium, University of Ottawa, ON, May 2009.