

Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates

A Thesis Submitted to the College of
Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in the Department of Microbiology and Immunology
University of Saskatchewan
Saskatoon, Saskatchewan

By Madison B. Adolph

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirement for a Postgraduate degree from the University of Saskatchewan, I agree that the libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done.

DISCLAIMER

References in this thesis to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes. Requests for permission to copy or to make other use of the material in this thesis, in whole or in part should be addressed to:

Head of the Department of Microbiology and Immunology
107 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E5
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116-110 Science Place
Saskatoon, Saskatchewan S7N 5CN
Canada

ABSTRACT

Apolipoprotein B mRNA-editing enzyme-catalytic, polypeptide-like 3 (APOBEC3) enzymes are a family of single-stranded (ss)DNA cytosine deaminases that serve as a host restriction factors for retrotransposons and viruses that contain a ssDNA intermediate. While the APOBEC3 family was originally expanded to restrict endogenous retroelements, the majority of these targets are now inactivated and the family has been found to act on different targets, such as viral ssDNA as a mechanism of viral defense. Some of the family members also catalyze “off-target” deaminations in human ssDNA and have been implicated in somatic mutagenesis that can lead to cancer.

My PhD thesis research investigated the biochemical mechanisms underlying both the benefits and the risks of the A3 family of enzymes. The central hypothesis of this work is that A3 enzymes have evolved distinct biochemical mechanisms to deaminate ssDNA that differentiate A3 restriction of retroelements and retroviruses from A3-induced somatic mutagenesis. Therefore, we investigated the biochemical mechanisms the A3 enzymes utilize during restriction of Human Immunodeficiency Virus (HIV) in both deamination-dependent and deamination-independent modes, as well as the unique mechanisms employed during mutation of genomic DNA. There are seven APOBEC3 members (A-H, excluding E) and of these, four members (A3D, A3F, A3G, A3H) have been identified to restrict HIV replication in CD4 T⁺ cells, and currently three members (A3A, A3B and A3H haplotype I) have been implicated in somatic mutagenesis.

The APOBEC3 enzymes that restrict HIV replication function optimally in the absence of the HIV viral infectivity factor (Vif). Vif targets APOBEC3 for degradation via the proteasome in HIV infected cells. For APOBEC3s that are able to fortuitously escape Vif mediated degradation and become encapsidated, the enzymes can deaminate cytosines to form uracils in viral (-)DNA. Replication of (-)DNA to (+)DNA causes HIV-1 reverse transcriptase to incorporate adenines opposite uracils which creates C/G→T/A transition mutations. While restriction of HIV most commonly occurs through this deamination-dependent mechanism, APOBEC3s have also been identified to interfere with HIV reverse transcriptase processes in a deamination-independent manner, although this mechanism is secondary to the deamination-dependent mode. In order to restrict retroelements and viruses such as HIV, the A3 enzymes

require an efficient processive ssDNA scanning mechanism that allows them to search for, and locate cytosines on ssDNA in the finite amount of time the substrate is available during formation of the double-stranded DNA provirus. To understand if this requirement was lost in A3 enzymes unable to restrict HIV, we studied A3C. Human A3C is not able to restrict HIV, in comparison to the more active gorilla and chimpanzee A3C orthologs that can restrict HIV, however an explanation for this difference was lacking. A polymorphism, S188I, in human A3C, which is found in less than 3% of the global population leads to increases enzymatic activity and HIV restriction. However, chimp and gorilla also possess the S188 allele suggesting a secondary residue responsible for increasing the activity of the A3C enzyme, in addition to residue 188. We determined that both I188 as well as N115 increases the activity of A3C by promoting dimerization of A3C. The dimerization of A3C increases its processivity on ssDNA and correlates with higher levels of mutagenesis during HIV reverse transcription, as is known for other A3 members studied. Since these amino acid changes also lie within the known interface for A3C interaction with Vif, we examined whether the different oligomerization states changed the ability of Vif to degrade A3C and found that Vif was able to induce A3C degradation regardless of its oligomerization state. Nevertheless, we determined that dimerization of A3C was able to predict the activity of the enzyme.

Similarly, it is known that A3 enzymes exhibiting reduced mutagenesis due to inefficient cytosine deamination may compensate by having an increased deamination-independent antiviral activity. A3 mediated inhibition of reverse transcriptase movement along the template may have an effect on processes such as HIV reverse transcriptase (RT) template switching and insertion fidelity, all of which would inhibit HIV replication. By examining these HIV RT processes in the presence of the A3 enzymes we were able to determine that A3F, which binds RNA and DNA with high affinity, promoted HIV RT template switching by blocking the progression of RT and forcing the RT to switch templates. Other A3s such as A3G and A3C, which bind the template with lower affinity, did not affect template switching. Interestingly, A3G decreased the fidelity of RT, causing misincorporation of nucleotides. Both of these outcomes, an increase in template switching and a decrease in polymerase fidelity, may promote virus evolution and emphasizes the importance of viral inactivation through the deamination-dependent mode of restriction.

If APOBEC3s are expressed in the wrong cell at the wrong time, they can facilitate mutagenesis of human genomic DNA and contribute to the C/G→T/A mutations evident in

multiple human cancers. A3 enzymes induce mutations during processes that generate ssDNA, such as transcription and replication, and we sought to determine the biochemical features of A3 enzymes that cause somatic mutations. Interestingly, instead of enzyme processivity, which is required for efficient restriction of HIV, the determining characteristic for mutagenesis of genomic DNA was the ability of A3 enzymes to cycle rapidly between ssDNA substrates. For example, A3A, a non-processive enzyme, readily cycled between substrates and induced higher levels of mutations than A3G, a processive enzyme that is unable to cycle. In addition to substrate cycling, enzyme oligomerization also influences its ability to induce mutations in cells due to a size limitation imposed by the dynamic transcription bubble. Deamination during transcription by A3B tetramers was poor, but A3A monomers, and A3H dimers were efficient. Therefore, the biochemical properties of the enzymes, in addition to availability of ssDNA, determine whether A3s will be able to induce mutagenesis in cells.

Taken together, these data have allowed us to better understand the biochemical mechanisms behind A3 enzyme evolution that has influenced their ability to restrict HIV in hominids, their ability to manipulate retroviral polymerases, and their capacity to induce somatic mutagenesis in human genomes.

ACKNOWLEDGEMENT

First and foremost, the most important acknowledgement belongs to my supervisor Dr. Linda Chelico. She took a chance on an undergraduate student with no research experience, and allowed me to pursue my passion for research. Her mentorship over the years has been invaluable and she has provided me with a level of training that I would not have found elsewhere. Her passion and love for science has amazed me everyday, and I strive to be one fraction as successful as she has become. To the members of the lab past and present, particularly Robin Love and Yuqing Feng, thank you for pushing me to become a better researcher and most importantly for your friendship over the years.

I would like to thank the members of my graduate committee, Dr. Wei Xiao, Dr. Kerri Kobryn, Dr. Joyce Wilson and Dr. Erique Lukong for their constant advice and support. Their presence on my committee has allowed my research and scientific knowledge to grow to new levels. I would also like to thank both the Department of Microbiology and Immunology and the Department of Biomedical Sciences here at the U of S. I have been apart of both departments for many years (more than I'd like to admit), and will always view this place as home.

A very large thank you goes out to my family, and my parents Bev and Dave Adolph. They have always supported this unconventional career path (both emotionally and financially), and have nurtured all of my aspirations. Last, but certainly not least, I would like to thank my best friend and partner in crime, Lauren Hopper. She has always been my biggest cheerleader and has had more faith in me than I have ever had in myself. She has been there through all ups and downs, and I cannot imagine my life without her friendship. To the numerous others who have been a part of this journey, thank you for your support.

DEDICATION

**I dedicate this these to my mentor Dr. Linda Chelico.
This thesis would contain very little substance if it weren't for her
constant support and guidance.**

TABLE OF CONTENTS

PERMISSION TO USE	i
ACKNOWLEDGEMENT.....	v
TABLE OF CONTENTS	vii
LIST OF FIGURES.....	x
LIST OF ABBREVIATIONS	xii
1.0. Introduction and Literature Review	1
1.1. APOBEC family of enzymes.....	1
1.2. Evolution of APOBEC3 activity	4
1.3. General HIV biology	6
1.4. Restriction of HIV by APOBEC3	10
1.4.1. Deamination-dependent restriction of HIV by APOBEC3.....	13
1.4.1.1. <i>Deamination-dependent restriction of HIV by A3G</i>	13
1.4.1.2. <i>Deamination-dependent restriction of HIV by A3F</i>	14
1.4.1.3. <i>Deamination-dependent restriction of HIV by A3D</i>	18
1.4.1.4. <i>Deamination-dependent restriction of HIV by A3H</i>	18
1.4.1.5. <i>Deamination-dependent restriction of HIV by A3C</i>	19
1.4.2. Interaction of APOBEC3 with HIV Vif	19
1.4.3. Deamination-independent restriction of HIV by APOBEC3	22
1.5. Restriction of endogenous retroelements by APOBEC3.....	22
1.6. Role of APOBEC family in somatic mutagenesis.....	24
2.0. Hypothesis and Objectives	29
2.1. Rationale and Hypothesis	29
2.2. Objectives	29
3.0. A single nucleotide polymorphism in human APOBEC3C enhances restriction of lentiviruses	30
3.1. Abstract.....	31
3.2. Introduction	31
3.3. Results	33
3.3.1. <i>APOBEC3C has evolved under positive selection in primates, suggesting an ancient role in protection from pathogens</i>	33
3.3.2. <i>Human APOBEC3C SNP I188 increases antiviral activity</i>	35
3.3.3. <i>Isoleucine at position 188 of APOBEC3C enhances enzymatic activity in vitro</i>	40
3.3.4. <i>Dimerization correlates with enhanced antiviral activity of human APOBEC3C</i>	41
3.3.5. <i>APOBEC3C variants are targeted by Vif</i>	46
3.3.6. <i>APOBEC3C I188 is an ancient human polymorphism that is not found in other hominids</i>	48
3.4. Discussion.....	50
3.5. Materials and Methods	54
3.5.1. <i>APOBEC3C sequences</i>	54

3.5.2.	<i>Sequence analysis</i>	55
3.5.3.	<i>APOBEC3, provirus and LINE-1 plasmids</i>	55
3.5.4.	<i>Infectivity assays</i>	55
3.5.5.	<i>LINE-1 assays</i>	56
3.5.6.	<i>Intracellular protein expression and packaging</i>	56
3.5.7.	<i>Recombinant protein expression and purification</i>	57
3.5.8.	<i>Size exclusion chromatography</i>	57
3.5.9.	<i>Protein crosslinking by BS3</i>	57
3.5.10.	<i>In vitro deamination assay</i>	58
3.5.11.	<i>In vitro reverse transcription assay</i>	58
3.5.12.	<i>SNP analysis</i>	59
4.0.	Examination of A3C orthologs for differences in enzymatic activity	60
5.0.	Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization	61
5.1.	Abstract	62
5.2.	Introduction	62
5.3.	Results	64
5.3.1.	<i>Human A3C S188I is a processive enzyme</i>	64
5.3.2.	<i>Chimpanzee and gorilla A3C is distinct from human A3C processivity</i>	65
5.3.3.	<i>Dimerization of hominid A3C is mediated through α helix 6 or β- strand 4</i>	68
5.3.4.	<i>Dimerization correlates with efficient ssDNA scanning</i>	75
5.3.5.	<i>A monomer-dimer equilibrium is required for scanning ssDNA by intersegmental transfer</i>	76
5.3.6.	<i>Ability to processively scan ssDNA correlates with mutagenesis ability</i>	78
5.4.	Discussion	86
5.5.	Materials and Methods	94
5.5.1.	<i>Cloning and site-directed mutagenesis</i>	94
5.5.2.	<i>Protein expression and purification</i>	94
5.5.3.	<i>In vitro deamination assay</i>	95
5.5.4.	<i>In vitro reverse transcription assay</i>	95
5.5.5.	<i>Size exclusion chromatography</i>	96
5.5.6.	<i>Protein crosslinking</i>	96
5.5.7.	<i>Steady-state rotational anisotropy</i>	96
5.5.8.	<i>Co-immunoprecipitation</i>	97
5.5.9.	<i>Single-cycle infectivity assay</i>	97
5.5.10.	<i>Quantitative immunoblotting</i>	97
5.5.11.	<i>Sequencing of integrated proviral DNA</i>	98
5.5.12.	<i>Structural models of A3C</i>	98
6.0.	Differences in Vif degradation efficiency among the A3C orthologs	100
7.0.	Investigation of the deamination independent restriction of HIV by APOBEC3s	107
8.0.	Determining the mechanism of inhibition of HIV reverse transcriptase by APOBEC3s	108
8.1.	Abstract	108

8.2.	Introduction	108
8.3.	Results	111
8.3.1.	<i>Modulation of HIV-1 RT template switching by APOBEC3 enzymes</i>	111
8.3.2.	<i>RNA template binding by A3F promotes HIV RT template switching</i>	113
8.3.3.	<i>Alteration of HIV RT insertion fidelity by APOBEC3</i>	117
8.4.	Discussion.....	125
8.5.	Materials and Methods	127
8.5.1.	<i>Synthesis of RNA templates</i>	127
8.5.2.	<i>Protein expression and purification</i>	127
8.5.3.	<i>In vitro template switching assay</i>	128
8.5.4.	<i>Steady state rotational anisotropy</i>	129
8.5.5.	<i>In vitro polymerase insertion fidelity assay</i>	129
9.0.	Biochemical determinants of deamination during transcription and replication by APOBEC3s	132
10.0.	Enzyme cycling contributes to the efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B.	134
10.1.	Abstract.....	135
10.2.	Introduction	135
10.3.	Results	137
10.3.1.	<i>A3B is a processive enzyme that forms higher order oligomers</i>	137
10.3.2.	<i>Processivity is not required for deamination during transcription</i> ...	146
10.3.3.	<i>Enzyme cycling is required for efficient deamination during DNA replication</i>	153
10.3.4.	<i>Enzyme cycling enables APOBEC3 enzymes to compete with RPA for ssDNA</i>	158
10.4.	Discussion.....	159
10.5.	Materials and Methods	167
10.5.1.	<i>Cloning and site directed mutagenesis</i>	167
10.5.2.	<i>Protein expression and purification</i>	167
10.5.3.	<i>In vitro deamination assays</i>	168
10.5.4.	<i>Electrophoretic mobility shift assay</i>	171
10.5.5.	<i>Size exclusion chromatography</i>	171
10.5.6.	<i>In vitro transcription assay</i>	171
10.5.7.	<i>In vitro DNA replication assay</i>	172
10.5.8.	<i>Steady state rotational anisotropy</i>	172
11.0.	Discussion and Future Directions.....	174
12.0.	References	179

LIST OF FIGURES

<i>Figure 1.1. Overview of human APOBEC cytidine deaminase functions</i>	<i>2</i>
<i>Figure 1.2. Schematic of the fate of the uracil after cytosine deamination in genomic DNA</i>	<i>3</i>
<i>Figure 1.3. Evolution and domain structure of the APOBEC family.....</i>	<i>5</i>
<i>Figure 1.4. Schematic overview of HIV-1 genome architecture and life cycle.....</i>	<i>7</i>
<i>Figure 1.5. Schematic of reverse transcriptase and replication of HIV</i>	<i>9</i>
<i>Figure 1.6. APOBEC3 restriction of HIV</i>	<i>11</i>
<i>Figure 1.7. Consequence of deaminations at 5'CC and 5'TC motifs.</i>	<i>15</i>
<i>Figure 1.8. DNA scanning by facilitated diffusion.....</i>	<i>17</i>
<i>Figure 1.9. Vif interaction interface with APOBEC3</i>	<i>20</i>
<i>Figure 1.10. Simplified view of a replication fork</i>	<i>27</i>
 <i>Figure 3.1. APOBEC3C is rapidly evolving in primates</i>	 <i>34</i>
<i>Figure 3.2. I188 is a SNP in APOBEC3C and is the conserved residue in the other six human APOBEC3 paralogs.</i>	<i>36</i>
<i>Figure 3.3. APOBEC3C SNP Isoleucine 188 confers increased antiviral activity.....</i>	<i>38</i>
<i>Figure 3.4. Inhibition of Line-1 replication by APOBEC3A and S188 and I188 APOBEC3C.....</i>	<i>39</i>
<i>Figure 3.5. In vitro characterization of APOBEC3C S188 and I188</i>	<i>41</i>
<i>Figure 3.6. In vitro reverse transcription assay.....</i>	<i>42</i>
<i>Figure 3.7. Purified APOBEC3C I188 forms dimers in solution.....</i>	<i>44</i>
<i>Figure 3.8. A synthetic dimer of APOBEC3C has increased antiviral activity</i>	<i>45</i>
<i>Figure 3.9. APOBEC3C is targeted by Vif.....</i>	<i>47</i>
<i>Figure 3.10. Isoleucine 188 changes to serine in some hominoids but was maintained or reverted back to isoleucine for some human populations.....</i>	<i>49</i>
 <i>Figure 5.1. Processivity of A3G and A3A</i>	 <i>66</i>
<i>Figure 5.2. Analysis of A3C processivity on ssDNA oligonucleotides.....</i>	<i>67</i>
<i>Figure 5.3. Sequence alignment and structural analysis of A3C.....</i>	<i>70</i>
<i>Figure 5.4. A3C dimerization is mediated through α-helix 6 or β-strand 4</i>	<i>71</i>
<i>Figure 5.5. SDS-PAGE of size exclusion chromatography resolved fractions of A3C enzymes.....</i>	<i>73</i>
<i>Figure 5.6. Dimerization influences processive ssDNA scanning</i>	<i>77</i>
<i>Figure 5.7. Intersegmental transfer ability of A3C is determined by the oligomeric state</i>	<i>79</i>
<i>Figure 5.8. Monomeric A3C induces lower levels of mutagenesis than dimeric A3C in vitro</i>	<i>81</i>
<i>Figure 5.9. Ability to dimerize correlates with increased mutagenesis in vitro.</i>	<i>83</i>
<i>Figure 5.10. Ability to dimerize correlates with increased mutagenesis in integrated proviral HIV Δvif DNA.....</i>	<i>87</i>
<i>Figure 5.11. Models of A3C dimerization.....</i>	<i>92</i>
 <i>Figure 6.1. Residues of importance in A3C</i>	 <i>102</i>
<i>Figure 6.2. Vif can disrupt A3C oligomerization.....</i>	<i>103</i>
<i>Figure 6.3. HIV Vif is able to promote the degradation of human and chimp A3C</i>	<i>105</i>

<i>Figure 6.4 A3C does not inhibit the initiation of primer extension.</i>	106
<i>Figure 8.1 Current models and schematic of template switching protocol.</i>	112
<i>Figure 8.2. Effect of A3s on the template switching efficiency of RT.</i>	114
<i>Figure 8.3. Interaction of A3 with the RNA template.</i>	115
<i>Figure 8.4. Interaction of A3G NPM and A3F NPM with p/t and the effect on template switching.</i>	118
<i>Figure 8.5. Steady state standing start kinetics of RT insertion fidelity</i>	120
<i>Figure 8.6. Steady state running start kinetics of RT insertion fidelity.</i>	122
<i>Figure 10.1. A3B NTD mediates enzyme activity and oligomerization.</i>	139
<i>Figure 10.2. Time course of deamination activity for fl A3B and CTD mutants</i>	141
<i>Figure 10.3. Interaction of A3 enzymes with nucleic acids.</i>	142
<i>Figure 10.4. SDS-PAGE of size exclusion chromatography fractions.</i>	143
<i>Figure 10.5. A3B is processive enzyme that both slides and moves 3-dimensionally on ssDNA.</i>	145
<i>Figure 10.6. Comparison of processivity for fl A3B and A3B 193 CTD.</i>	147
<i>Figure 10.7. Deamination during transcription is mediated by size and not by processivity.</i>	148
<i>Figure 10.8. Deamination activity of A3 enzymes on dsDNA undergoing transcription.</i>	151
<i>Figure 10.9. Deamination activity of A3 enzymes on R-loop substrates</i>	152
<i>Figure 10.10. A3-mediated mutagenesis during DNA replication requires enzyme cycling</i>	155
<i>Figure 10.11. Deamination activity of A3 enzymes in the absence or presence of unlabeled DNA</i>	156
<i>Figure 10.12. Electrophoretic mobility shift assay of RPA bound to ssDNA.</i>	160
<i>Figure 10.13. A3 enzymes can compete with RPA to deaminate ssDNA.</i>	161
<i>Figure 10.14. SDS-PAGE of purified A3 enzymes</i>	169
<i>Figure 11.1. Model for APOBEC3 activity on disperse ssDNA substrates</i>	175

LIST OF ABBREVIATIONS

-nt- Nucleotide(s)
(-) DNA-minus strand/template strand
(+) DNA-positive strand/coding strand
(fl)- Full length
(R)- Repeat region
3'UTR- 3' untranslated region
A3- APOBEC3
A3H hap I- APOBEC3H haplotype I
AID-Activation Induced cytosine Deaminase
AIDS- Acquired immunodeficiency syndrome
APE-apurinic/apryidiminic endonuclease
APOBEC1- apolipoprotein B mRNA editing, catalytic-enzyme polypeptide like 1
APOBEC3-apolipoprotein B mRNA editing, catalytic-enzyme polypeptide like 3
BER- base excision repair
bp- Base pair
BS3- bis(sulfosuccinimidyl)suberate
CA- Capsid
cA3C- chimpanzee APOBEC3C
CBF β - Core binding factor β
Co-IP- Coimmunoprecipitation
CRCL5- Culling RING ubiquitin ligase 5
CTD- C-terminal domain
Cul5- Cullin 5
dN- Non-synonymous change
dNTP- deoxyribonucleoside triphosphate
dS- Synonymous change
DTT- Dithiothreitol
E.coli- *Escherichia coli*
EloB- Elongin B
EloC- Elongin C
Env- Envelope
FAM- Fluorescein
gA3C- gorilla APOBEC3C
Gag- Group specific antigen
hA3C- human APOBEC3C
HERV-K- Human endogenous retrovirus K
HIV- Human Immunodeficiency Virus
IN- Integrase
K_d- Dissociation constant
LINE-1- Long interspersed nuclear element 1
LTR-Long terminal repeat
MA-Matrix
mC- Methyl-cytosine
MBP- Maltose binding protein
mRNA- Messenger RNA

NC- Nucleocapsid
 NTD- N-terminal domain
 p/t- Primer/template
 PBMC- Peripheral blood mononuclear cell
 PBS-Primer binding site
 PISA- Protein interfaces, surfaces and assemblies
 Pol- polymerase
 PPT- Poly-purine tract
 PR- Protease
 Prot- Protease
 RBX2- Ring box protein 2
 RNase A- Ribonuclease A domain enzyme
 RNase H- Ribonuclease H domain enzyme
 RNP- Ribonucleoprotein
 RPA- Replication protein A
 RT- Reverse transcriptase
 SDM- Site directed mutagenesis
 SDS-PAGE- Sodium dodecyl sulfate polyacrylamide gel electrophoresis
 SEC- Size exclusion chromatography
Sf9- Spodoptera frugiperda 9
 SINE- Short interspersed nuclear element
 SIV- Simian Immunodeficiency Virus
 SIV_{cpz}- Simian immunodeficiency virus from chimpanzees
 SNP- Single nucleotide polymorphism
 SOCS2- Suppressor of cytokine signaling 2
 ssDNA- single-stranded DNA
 Tat- Transcriptional transactivator
 UDG/UNG-uracil DNA glycosylase
 VCBC- Vif/ CBF β /EloB/C complex
 Vif -Viral Infectivity Factor
 Z1- Zinc binding domain 1
 Z2- Zinc binding domain 2
 Z3- Zinc binding domain 3

1.0. Introduction and Literature Review

Portions of this chapter were reprinted with permission from Adolph MB., *et al.* (2018) [1].

1.1 APOBEC Family of Enzymes

The APOBEC family of enzymes are single-stranded polynucleotide cytidine deaminases and are named after the function of their namesake member, Apolipoprotein B mRNA editing complex, APOBEC1 (A1) which was the first member to be identified [2]. The AID/APOBEC family share structural and functional similarities to other zinc-dependent deaminases, and the APOBEC locus evolved function from the ancestral member, activation induced cytosine deaminase (AID) [3]. AID was genetically discovered prior to the APOBEC3 family members and has maintained a nonconventional name compared to the other enzymes in this family [4]. APOBEC and AID enzymes modify cytidine in either mRNA or single-stranded (ss) DNA to form uracil [5-7]. Other members have also been identified, including APOBEC2 and APOBEC4, though they do not have demonstrated catalytic activity [3, 8, 9]. The primary substrate of family members other than A1 is ssDNA (Figure 1.1), though the RNA editing role of A1 may not have been its ancestral role [10]. In RNA, where uracil has a coding function, deamination is an RNA editing tool. In DNA, the formation of uracil by APOBEC/AID is a promutagenic lesion which in some cases can be detrimental and cause cancer, but in other cases enables evolution of antibody genes and inactivation of viruses, retroelement, or other cytoplasmic inflammation-inducing DNA [5, 11, 12].

The multiple functions of the APOBEC family of proteins are possible because uracil has multiple fates in ssDNA. Most commonly, uracil is recognized as a lesion by the base excision repair (BER) enzyme uracil-DNA glycosylase (UNG) and excised. Subsequent processing by AP endonuclease (APE) and polymerase gap filling is required to complete the repair [13]. If UNG removes uracils during replication, DNA damage tolerance mechanisms such as template switching can allow for error-free lesion bypass [14]. If left unrepaired, the uracil will template addition of adenine at a site that should have incorporated a guanine, thus creating a C/G→T/A transition mutation (Figure 1.2A) [13]. In B cells, the uracils formed by AID in antibody genes act as nucleating factors for error prone DNA repair to evolve and mature antibody genes or create double-stranded DNA breaks to facilitate immunoglobulin class switching [11]. In T cells, macrophages, and germ cells, various APOBEC3 (A3) members (A3A, A3B, A3C, A3D, A3F,

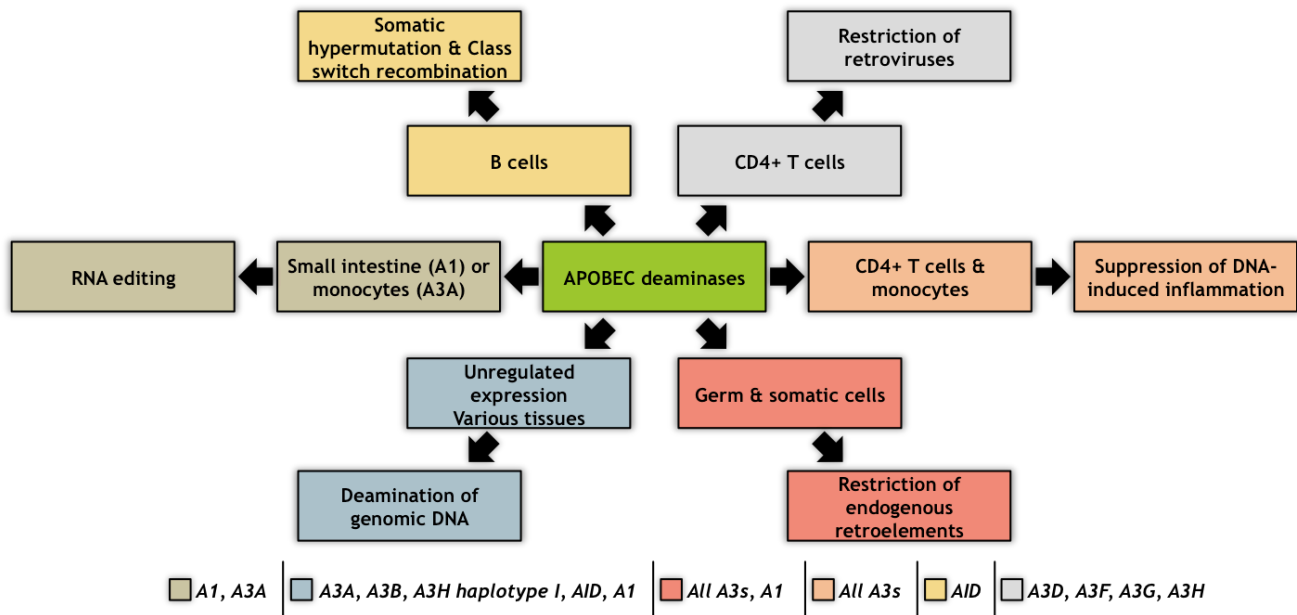


Figure 1.1. Overview of human APOBEC cytidine deaminase functions. During normal function, the APOBEC family members are each expressed in unique cell types. RNA editing by APOBEC enzymes can occur in the small intestine (A1) or in monocytes (A3A). In B-cells, AID deaminations initiate antibody diversification pathways for somatic hypermutation and class-switch recombination. In CD4+ T cells, A3 deamination of retroviral ssDNA intermediates restricts their replication. In CD4+ T cells and monocytes, A3s can restrict foreign DNA in order to reduce the DNA-induced inflammatory response. In germ cells and somatic cells, APOBEC enzymes may restrict retrotransposons through RNA binding or deamination of reverse transcripts. In various tissues, unregulated APOBEC expression can lead to unregulated deaminations in genomic DNA that can initiate genomic instability or cell transformation, which ultimately may lead to cancer. The APOBECs involved in each process are depicted below the figure. Reprinted with permission from Adolph MB., Love RP., Chelico, L. (2018) Biochemical basis of APOBEC3 deoxycytidine deaminase activity on diverse DNA substrates. Reprinted with permission from *Adolph et al.*, 2018 [1].

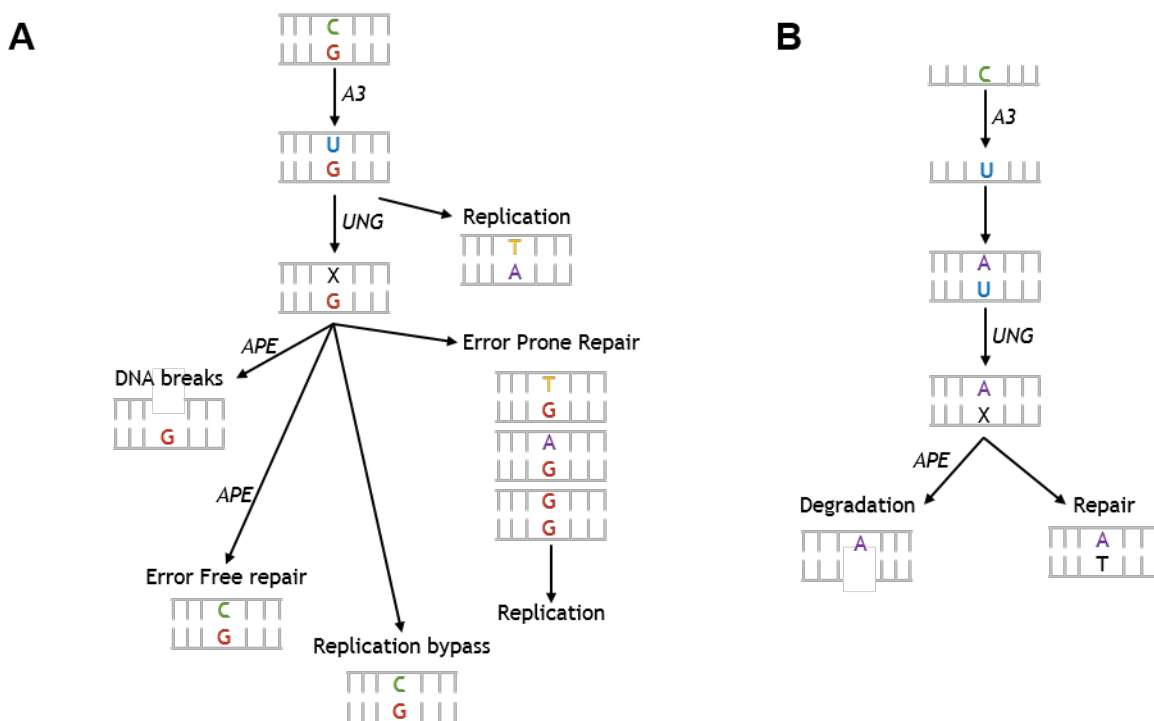


Figure 1.2. Schematic of the fate of the uracil after cytosine deamination in genomic DNA. (A) Without repair, replication of the U results in a T/A base pair instead of a C/G base pair. Alternatively, uracils created by cytosine deamination are removed by uracil DNA glycosylase (UNG) to create an abasic site (X). This abasic site can be processed by AP endonuclease (APE) that cleaves the DNA backbone at the abasic site to enable subsequent repair by Base Excision Repair (BER) polymerases, the Mismatch Repair (MMR) pathway, or depending on the number of uracils in the immediate area this can lead to DNA breaks. Alternatively, this lesion can be bypassed through error free mechanisms or gap-filling DNA synthesis can occur by high fidelity polymerases in an error free manner. Depending on cellular conditions, translesion synthesis polymerases may also synthesize DNA within the gap, resulting in error prone repair. (B) A3 enzymes deaminate single-stranded (-) DNA intermediates of retroelements, HIV-1, or endogenous retroviruses. The uracil formed by cytosine deamination in the (-) DNA strand serves as a template for the reverse transcriptase during (+) DNA synthesis, which leads to concomitant G→A mutation in the (+) DNA due to the absence of an unmodified template. In the host nucleus, the uracil can be excised by UNG/APE and depending on the number of uracils this can lead to degradation of the template or repair of uracil and integration of the retroelement or proviral DNA. Reprinted with permission from *Adolph et al.*, 2018 [1].

A3G, and A3H) can deaminate cytidines in ssDNA intermediates of retroviruses, retrotransposons, and other foreign DNA to cause mutagenesis. This mutagenesis occurs to the point of destroying gene function or inducing DNA repair mediated DNA degradation to block replication of foreign elements in these cells (Figure 1.2B) [2, 15]. The APOBEC enzymes must remain tightly regulated to avoid mutagenesis of the cell's own genomic DNA. Regulation failure leads to APOBEC-induced somatic mutagenesis and the C/G→T/A mutation signature is evident in multiple cancers [16, 17].

1.2 Evolution of APOBEC3 activity

Despite the risk of harboring these promutagenic enzymes, the APOBEC family is ancient and has been present since the emergence of jawed vertebrates, with AID being the ancestral member (Figure 1.3B) [3]. AID is selectively expressed in B cells, and is essential for the regulation of B cell diversification [4]. Cytidine deamination by AID on ssDNA during transcription of variable regions in the immunoglobulin locus is required for the initiation of both class-switch recombination and somatic hypermutation [18, 19]. Interestingly, the activity of AID is regulated at the transcription bubble through the interaction with transcriptional co factors that assist with AID catalytic activity [20, 21]. The expression of AID in B cells is essential as AID deficiency leads to Hyper-IgM syndrome and the inability to produce antibodies other than IgM M2 [22]. Despite APOBEC2 (A2) being the second APOBEC deaminase that evolved, catalytic activity has not been demonstrated for this enzyme, although it is required for proper muscle development in mice [3, 8, 9, 23]. As more complex organisms evolved, duplications in the APOBEC family of genes occurred to create A1, likely from AID, and in placental mammals, the A3 enzymes appeared (Figure 1.3B) [3, 23]. In spite of A1's role in RNA editing it evolved from AID, a DNA targeting enzyme. However, there is evidence that A1 from the green anole lizard cannot deaminate RNA, suggesting that its ancestral function was as a deoxycytidine deaminase [10]. AID was also initially thought to be a RNA editing enzyme, due to homology with other RNA deaminases, however an RNA substrate has not been identified [24]. The A3 enzymes have been amplified from one member in mice to seven members in primates, and all possess a similar Zinc-binding domain structure (Figure 1.3C) [3, 23]. This massive expansion of the A3 locus correlates with the increased number of primate retrotransposons and endogenous retroviruses, and is thought to have been required to suppress excessive damage to the genome by these elements [23, 25]. With the majority of retrotransposons and endogenous retroviruses in humans

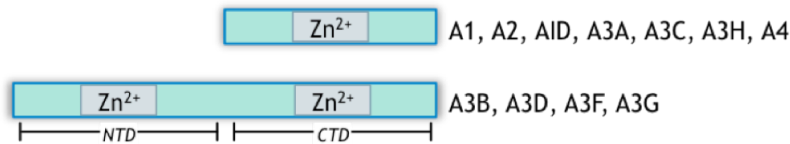
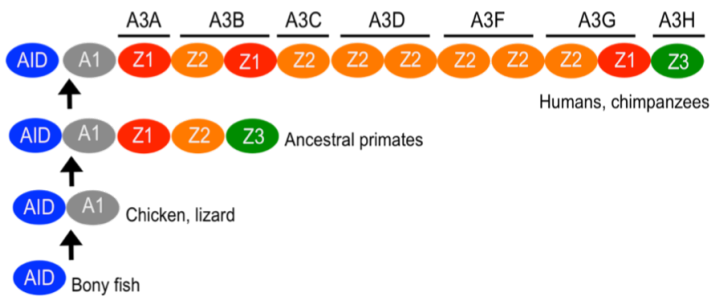
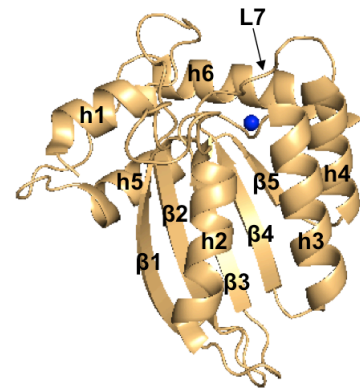
A**B****C**

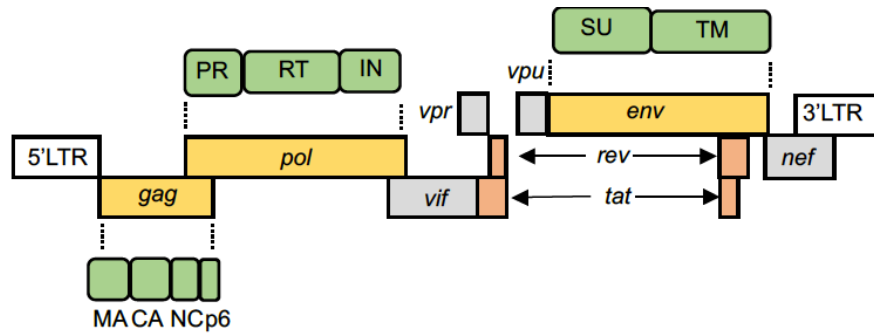
Figure 1.3. Evolution and domain structure of the APOBEC family. (A) The APOBEC family in humans is composed of eleven members having either a single or double Zn²⁺-coordinating domain per polypeptide chain. Of the double domain enzymes, the N-terminal domain (NTD) is not catalytically active, but often serves as a processivity domain. The C-terminal domain (CTD) confers the catalytic activity. (B) Expansion of the primate APOBEC3 locus with 7 APOBEC3 members in humans with corresponding Zinc-binding domain (Z) classification based on the phylogenetic cluster of the domain (Z1-Z3) (modified from Lackey et al. 2012). (C) A3 enzymes have a basic structure in each Z-type domain that is composed of a five-stranded β -sheet core surrounded by six α -helices. Numerical assignments to β -strands and α -helices are superimposed in (A). Zinc atom is shown as a blue sphere.

inactivated, it is thought that the risk of A3 enzymes may now outweigh the benefits of their protection [25]. Evidence of this comes from population stratifications, where there are deletions in the human A3 locus on chromosome 22 that removes A3B or sequence variations in A3 genes that decrease deaminase activity, such as some forms of A3H, A3D, and A3C [26-28]. Despite AID posing a risk to genomic integrity, this enzyme is indispensable to proper immunity and must be maintained [19].

1.3 General HIV biology

Human Immunodeficiency Virus type 1 (HIV-1, hereafter HIV) currently infects 37 million people worldwide, with the majority of those infected residing in Sub-Saharan Africa [29]. HIV, and the related though less infectious virus HIV-2, arose through zoonotic transfer from primates infected with simian immunodeficiency virus (SIV). HIV is a positive-sense, single-stranded (ss)RNA virus that belongs to the retrovirus family [30]. It consists of a 9.7 kb genome that encodes numerous proteins that exist within three main groups: structural proteins, regulatory proteins, and viral accessory proteins (Figure 1.4A) [31]. The HIV contains two dimerized RNA genomes, although only one proviral DNA is produced during infection [32]. Group-specific antigen (*gag*), polymerase (*pol*) and envelope protein (*env*) are structural proteins essential for HIV replication [31]. Of the structural proteins, *gag* encodes the precursor gag protein, which is cleaved by a virally encoded protease into smaller structural subunits, namely matrix (MA), capsid (CA), and nucleocapsid (NC); *pol*, which arises from a Gag-pol precursor, encodes reverse transcriptase (RT), protease (PR) and integrase (IN) enzymes required for the virus to undergo reverse transcription (RT), integration (IN) and cleavage of the Gag precursor; *env* is the viral envelope precursor protein that is cleaved into gp120 (also called SU) and gp41 (also called TM) and facilitates the attachment to and infection of host cells (Figure 1.4A) [33]. The HIV regulatory proteins are transcriptional transactivator (Tat), and regulator of viral gene expression (Rev), and are required for elongation and regulation of transcription products, and nuclear export of viral mRNA, respectively [34]. HIV also encodes accessory proteins, which aid the virus in avoiding the host immune mechanisms and are crucial virulence factors [35-37]. The accessory proteins (Vpr, Vpu, Nef and Vif) enhance viral fitness by hijacking host protein degradation pathways in order to overcome the host's viral restriction factors or change immune signaling [35-37]. Vpr is incorporated into viral particles and aids the infection of host cells by regulating the localization of the proviral DNA pre-integration complex (PIC) [36, 37]. Vpr also hijacks a

A



B

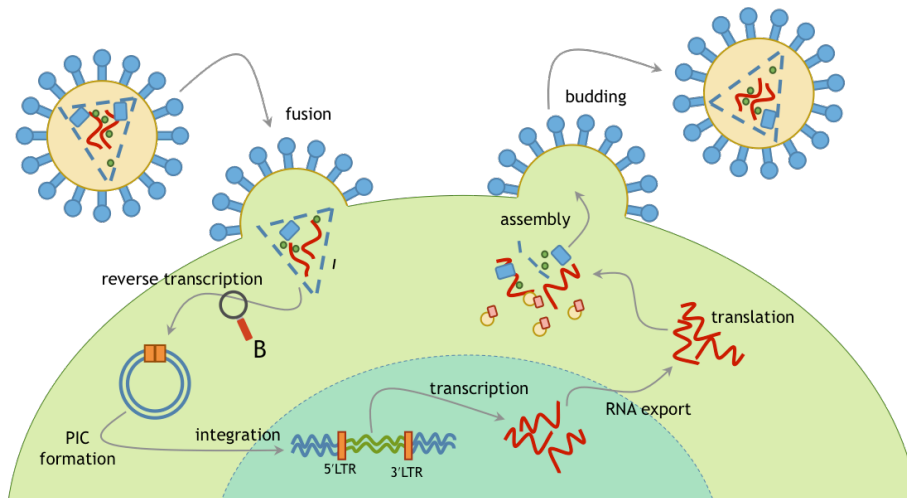


Figure 1.4. Schematic overview of HIV-1 genome architecture and life cycle. (A) HIV encodes three structural genes (yellow), gag pol and env, which are processed into mature proteins (green). The HIV genome contains 15 distinct proteins: the Gag and Env structural proteins MA (matrix), CA (capsid), NC (nucleocapsid), p6, SU (surface, gp120), and TM (transmembrane, gp41); the Pol proteins: PR (protease), RT (reverse transcriptase), and IN (integrase); two regulatory proteins Tat and Rev (orange); four accessory proteins Nef, Vif, Vpr, and Vpu (wheat) (B) Major steps involved in the HIV-1 replication cycle. HIV-1 envelope glycoprotein binds to the cell surface receptor to trigger viral fusion. Uncoating of the viral capsid, coupled with reverse transcription, leads to the formation of a double-stranded DNA and a pre-integration complex (PIC). The PIC-associated viral integrase (vertical box) mediates integration into the host genome. Proviral DNA (red) transcription is mediated by the host RNA polymerase II and HIV-1 Tat, producing viral mRNA transcripts (red) for viral protein production and viral genomic RNA. Protease-mediated maturation occurs after budding. Reprinted with permission from Adolph et al., 2018 [1].

host E3 ubiquitin ligase complex and causes G2 cell cycle arrest [38]. Vpu enhances virion release by triggering ubiquitin-mediated degradation of CD4 glycoprotein as well as CD4 down-modulation in complex with Env [39, 40]. Nef prevents immune targeting of virally infected cells by down regulating the expression of MHC-1 at the surface of the infected cells (Figure 1.4A) [35-37]. Viral infectivity factor (Vif), targets for degradation certain APOBEC3 enzymes that target and restrict viral replication [41]. The nature of this interaction will be expanded in detail in section 1.4.3.

In the HIV mature virion, the RNA genomes and enzymes required for proviral DNA synthesis are contained in a protein capsid. The capsid is surrounded a protein coat termed the matrix which also interacts with the virus lipid envelope. Within the lipid envelope the HIV surface (SU, gp120) and transmembrane proteins (TU, gp41) bind to a receptor on the host CD4 T cell receptor in concert with the chemokine co-receptor CCR5 or CXCR4 to enable fusion of the viral and cellular membranes. This allows for insertion of the HIV core into the host cell (Figure 1.4B) [42]. It is thought that the capsid remains intact until the viral complexes reach the nuclear membrane, however the precise timing and location of viral uncoating are poorly understood [43]. Maintaining an intact capsid allows the stoichiometry of the viral components to remain high and promotes efficient reverse transcription. The HIV capsid becomes permeable to small macromolecules and the high level of host deoxynucleotide triphosphates in the cytoplasm allows diffusion through the capsid pore to initiate the reverse transcription process (Figure 1.4B) [44]. The HIV pre-integration complex (PIC) is formed upon virus uncoating and once viral double-stranded (ds)DNA is formed from the RNA genome, the PIC integrates into the chromosome of the host cell. With the aid of the viral IN protein, this forms the provirus, which is flanked by long terminal repeats (LTRs) [33, 42]. The proviral DNA is transcribed by RNA polymerase II to generate viral RNA that is subsequently translated to form viral proteins or packaged into newly assembled virions. These progeny virions undergo viral maturation and budding, which produces infectious viral particles (Figure 1.4B) [45].

Reverse transcription utilizes a virally encoded RT to covert the RNA genome into DNA (Figure 1.5) [46]. This process can be exploited by the APOBEC3 enzymes, and will be a central component of the presented work. HIV RT is a heterodimer that contains p66 and p51 subunits that are 560 and 440 amino acids in length, respectively [47]. The p51 subunit is thought to play a structural role, while p66 contains an RNA- and DNA-dependent DNA polymerase and an

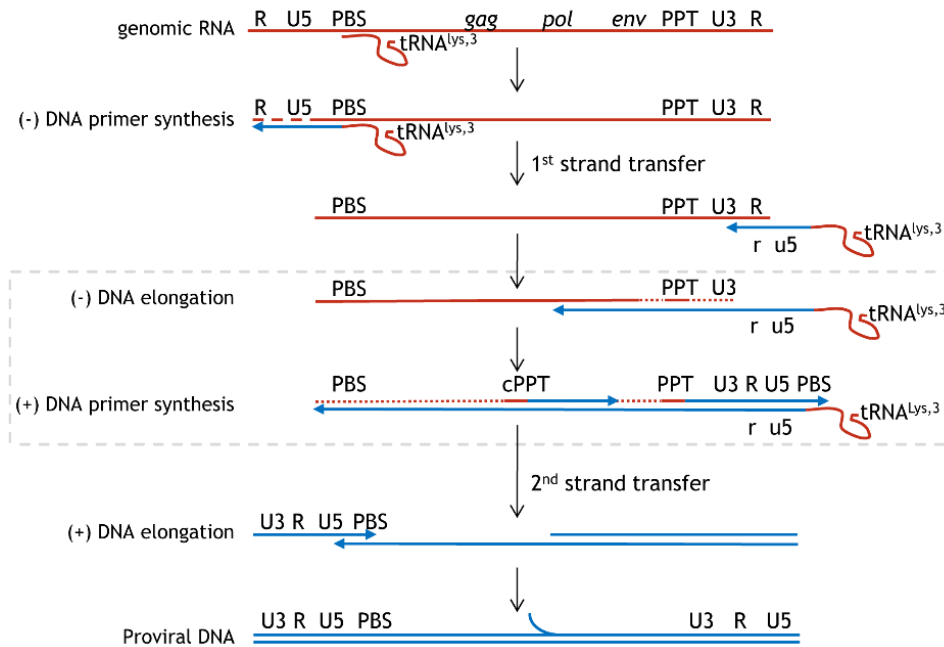


Figure 1.5. Schematic of reverse transcriptase and replication of HIV. The genomic RNA (red line) is primed for replication using a host tRNA^{lys3} primer. Synthesis of the minus strand (-) cDNA (blue line) generates an RNA/DNA hybrid that is a substrate for RNase H. RNase H degrades the genomic RNA strand (hatched red line), leaving the nascent (-) cDNA single-stranded. First strand transfer occurs when the (-) DNA hybridizes to the repeated (R) sequence at the 3' end of the viral RNA. The (-) DNA synthesis resumes and the reverse transcriptase associated RNaseH domain degrades the RNA strand (hatched line). Two purine-rich sequences called the polypurine tracts (PPT) are relatively resistant to RNaseH cleavage and serve as the primer for (+) DNA synthesis. HIV-1 has a central and 3' end PPT (cPPT and PPT). The (+) DNA synthesis enables removal of the tRNA primer. In a process called the second strand transfer, both the (+) and (-) DNA are extended until the entire DNA becomes double-stranded, eventually creating a dsDNA that has the same sequences (U3-R-U5) at both ends (called the long terminal repeats, LTRs). The nucleocapsid protein facilitates the strand transfer. Figure is adapted from Coffin et al.[30]. Reprinted with permission from Adolph et al., 2018 [1].

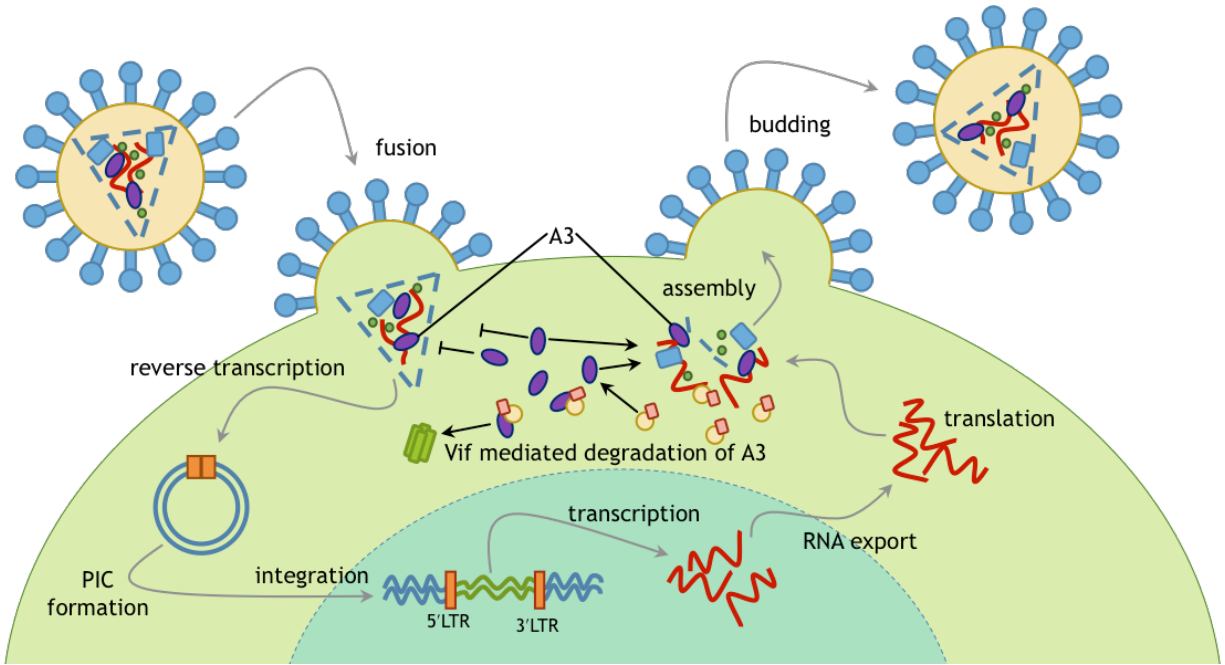
RNaseH domain that cleaves RNA in RNA/DNA hybrids. To initiate reverse transcription, there is a primer binding site (PBS) that is 18 nucleotides long near the 5' end of the genome. The PBS is primed by a host tRNA_{lys3}. The RT synthesizes the minus strand (-) complementary DNA (-)DNA as the RNaseH degrades the RNA. First strand transfer of the (-)DNA is facilitated by HIV nucleocapsid (NC) and transfers the (-)DNA to the 3' repeated (R) sequence. The plus-strand (+)DNA synthesis is primed by polypurine tracts (PPT) within the RNA genome that are resistant to RNaseH activity [30]. Completion of reverse transcription results in a double stranded DNA segment, with each end containing the same U3-R-U5 sequence. These sequences represent the LTRs, which contain the viral promoter for transcription of viral mRNAs, including the full length genomic RNA synthesis.

1.4 Restriction of HIV by APOBEC3

When the HIV-1 fuses with a target CD4+ T cell, the capsid is released into the cytoplasm (Figure 1.6A). The deoxynucleotide triphosphates (dNTP) from the host cell diffuse into the nucleocapsid and reverse transcription begins [44]. Due to the impermeability of the capsid to cellular proteins at this early stage, if A3 enzymes are to access the HIV-1 (-) DNA synthesized by reverse transcriptase, they must be present in the viral capsid (Figure 1.6A). Thus, the model for HIV-1 restriction by A3 enzymes in CD4+ T cells requires that they become encapsidated into newly formed virus particles in the producer cell [48]. To mediate encapsidation the A3 enzymes bind viral or cellular RNAs with the same sequence affinity as the NC domain of Gag that packages the HIV-1 RNA [49, 50]. Encapsidation is also facilitated by A3 oligomerization on RNA although the mechanism is not entirely understood [51]. It could simply be that oligomerization is a way of A3s hedging bets for encapsidation through multiple mechanisms, since there is no guarantee of an A3 in each virion.

Usually, the encapsidation of A3 enzymes is blocked by the HIV-1 protein Vif (virus infectivity factor), which causes polyubiquitination and degradation of A3 enzymes through the 26S proteasome [41]. Vif can specifically interact with A3 enzymes and acts as a substrate receptor by interacting with components of a Cullin RING ligase 5 E3 ubiquitin ligase complex, and results in a complex capable of binding and polyubiquitinating an A3 (Figure 1.6A) [41, 52-55]. In HIV-1 infected people, the suppression of A3 enzymes by Vif is not complete since proviral genomes have been found to contain footprints of A3 deaminase activity [56, 57]. However, the inhibition is clearly sufficient enough to enable ongoing viral replication. From cell

A



B

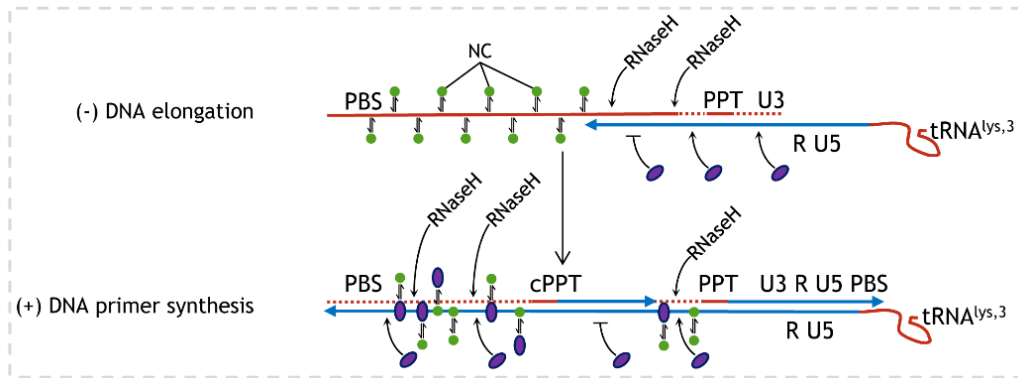


Figure 1.6. APOBEC3 restriction of HIV. (A) Cellular A3s (purple) are unable to interact with viral RNA undergoing reverse transcription due to the presence of the capsid (hatched line) and must be encapsidated in the virus producer cell to deaminate proviral DNA. The action of A3 enzymes is thwarted by HIV-1 Vif that induces their degradation through the 26S proteasome. Vif (red square) interacts with CBF β , Elongin C and Cullin 5, which enables recruitment of E3 ligase components (shown as a circle for simplicity) and degradation of A3 enzymes (purple). (B) Nucleocapsid (NC) (green) coats the DNA during (-) DNA elongation and (+) DNA synthesis, however the A3 (purple) is able to displace the transiently bound NC in order to access the ssDNA (dotted line) created through RNaseH degradation of the template strand. Reprinted with permission from Adolph *et al.*, 2018 [1].

culture studies with Δ Vif HIV-1, only A3D, A3F, A3G, and A3H are encapsidated and restrict HIV-1 replication [58]. Although all four of the A3s have variants in the human population, A3H is the most notable since it is the most variable and has seven major haplotypes (hap), but only hap II, hap V, and hap VII are able to restrict HIV-1 replication [59, 60].

The encapsidated A3 enzymes are in a competitive situation where they must deaminate cytosines in a ssDNA substrate that is not available initially and only available for a finite time after synthesis (Figure 1.6B). To form the (-) DNA, the HIV-1 RNA genome is primed by a tRNA^{lys,3} that is unfolded and annealed to the PBS by the HIV-1 nucleic acid chaperone, NC [61]. After reverse transcription and degradation of the RNA by the RNaseH domain associated with the HIV-1 reverse transcriptase, NC facilitates a strand transfer to another complementary repeat (R) region and full synthesis of the (-) DNA can begin. During synthesis of the (-) DNA, the RNaseH activity of reverse transcriptase will cleave the RNA at internal sites until the RNA cleavage leads to dissociation [30]. The NC is abundant in the capsid and protects the HIV-1 genetic material by binding the RNA or ssDNA at approximately one nucleocapsid every seven nucleotides [62]. NC facilitates reverse transcriptase by melting secondary structure in the template [63]. Although not fully understood, the mechanism of nucleic acid chaperoning by nucleocapsid involves the protein rapidly binding and unbinding the RNA or DNA [63]. Thus, A3 enzymes can compete with NC for access to the (-) DNA (Figure 1.6B). However, there are still physical obstacles the enzymes must overcome such as annealed RNA fragments, the reverse transcriptase, and constraints such as finding the deamination motif.

Each A3 deaminates cytosine within a preferred sequence. A trinucleotide motif is the preferred substrate, although a dinucleotide motif can also be deaminated. For A3G the preferred sequence is 5'CCC (underlined C deaminated) and for other A3 enzymes it is 5'TC, with each enzyme having a specific preference, e.g., 5'TTC for A3F and 5'CTC for A3H [64-66]. The C→U deaminations catalyzed on the (-) DNA strand become G→A mutations when reverse transcriptase is forced to use uracil as a template during (+) DNA strand synthesis (Figure 1.2B) [67-70]. Throughout this process, the A3 enzymes also directly disrupt proviral DNA synthesis by oligomerizing on the template DNA resulting in a “roadblock” for the reverse transcriptase [71-74]. In addition, A3G can interact with reverse transcriptase directly and inhibit polymerase activity [71, 75]. The double-stranded proviral genome is imported into the nucleus and either the uracils cause DNA repair mediated cleavage of the provirus or the provirus is integrated through

the action of HIV-1 integrase and host protein interactions [75] (Figure 1.2B, Figure 1.4B). Depending on the number and type of mutations induced by the A3 enzymes, the integrated provirus may be functionally inactivated. Most commonly labs have studied A3-induced mutagenesis of a Δ Vif HIV-1 virus. In these situations, the A3 “hypermutation” arises where enzymes such as A3G can induce 5 or more mutations per kilobase [67, 69, 70]. Although hypermutated and integrated proviruses have been isolated from HIV-1 infected patients, the number of replication cycles that the provirus underwent before becoming hypermutated and nonfunctional is unknown [76, 77]. In the presence of Vif, the A3 enzymes are depleted from cells to such an extent that viruses can replicate, although a low level of mutations can persist. Measurements of the number of A3-induced mutations in a single round of infection are lacking and need to be determined in order to identify whether A3 deaminations in a wild type virus infection induces hypermutation and viral inactivation, a low level of mutation that leads to virus evolution, or has a neutral effect. All of these possibilities have been characterized in different labs using different experimental systems [57, 78-82]. The variability may be due to which A3 enzyme studied and the amount that was encapsidated. The number of A3-induced mutations can be influenced by encapsidation levels, which are indirectly related with the ability of Vif-to mediate A3 degradation, but can also be affected by the inherent ability of the A3 enzymes to bind and oligomerize on cellular and viral RNA.

1.4.1 Deamination-dependent restriction of HIV by APOBEC3

1.4.1.1 Deamination-dependent restriction of HIV by APOBEC3G.

APOBEC3G was the first APOBEC3 family member to be identified and is the most widely studied, mainly due to the enzyme having the highest levels of HIV restriction in comparison to the other APOBEC3s that are relevant to HIV restriction [48]. In order for A3G to be an efficient restriction factor, it must bind RNA and become encapsidated into the budding virion. Newly synthesized A3G that has not been sequestered to stress granules and P-bodies through host RNA binding [83-85], bind viral or cellular RNA that is subsequently bound by the HIV Gag protein during viral encapsidation [86-97]. Encapsidation of A3G is facilitated by RNA binding and oligomerization mediated through its the N-terminal domain (NTD) [98]. The NTD, specifically residues found on loop 7, of A3G are known to be responsible for A3G virion encapsidation [98-101]. A3G exists as monomers and dimers in solution, however it is able to further oligomerize once bound on DNA and RNA [102, 103]. Further oligomerization is

mediated by the C-terminal domain (CTD) through residues found on loop 7. Once encapsidated, A3G is able to access single-stranded (ss)DNA during (-)DNA synthesis. However, the ability of A3 to access the viral DNA being reverse transcribed is not sufficient for restriction, and A3G must possess efficient search mechanisms in order to locate target ssDNA.

For DNA modification enzymes, efficiency generally means that the enzyme is processive and can deaminate multiple cytosines in a single enzyme-substrate encounter. For enzymes that do not use an energy source to move on DNA, they can search the DNA for a specific target motif using facilitated diffusion [104, 105]. Facilitated diffusion describes how enzymes can use the electrostatic interaction with DNA to facilitate their Brownian motion driven diffusion [104, 105]. The diffusion can be described by different terms depending on whether the enzyme diffuses along the phosphate backbone (sliding), diffuses along the DNA without always making direct contact with the phosphate backbone (jumping or hopping), or moving through a doubly-bound state (intersegmental transfer) (Figure 1.7) [2, 104-106]. Sliding and hopping movements are efficient for searching < 20 nt of DNA, and this movement is essential for finding the deamination motif for A3 enzymes that slide [2]. Jumping and intersegmental transfer facilitates movements on DNA > 20 nt, but there is no local search at each place the enzyme lands unless the enzyme can also slide [106-108]. Thus, if an enzyme only has sliding or only has a three-dimensional search mechanism such as hopping, jumping, or intersegmental transfer, the search will be inefficient [2, 107-110]. Numerous studies have demonstrated that A3G is a processive enzyme that is able to use both one-dimensional and three-dimensional motions [107, 111]. This observation correlates with data from integrated proviral genomes, which demonstrates clusters of A3G-induced deaminations due to the processive one-dimensional movements of the enzyme. These findings are further confirmed when examining mutants of A3G that are unable to undergo sliding such as A3G H186R and A3G 193NPMins that have decreased mutagenesis in both *in vitro* studies and HIV proviral genomes [107, 108]. Similarly, an A3G mutant that is unable to undergo three-dimensional jumping movements, A3G F126A/W127A, exhibited decreased mutagenesis during *in vitro* studies [107]. Therefore, these data demonstrate that A3 enzymes must possess multiple ssDNA search mechanisms in order to efficiently induce mutations during the dynamic process of reverse transcription.

1.4.1.2 Deamination-dependent restriction of HIV by APOBEC3F.

A3F has also been shown to restrict HIV, although studies from most labs show that it is

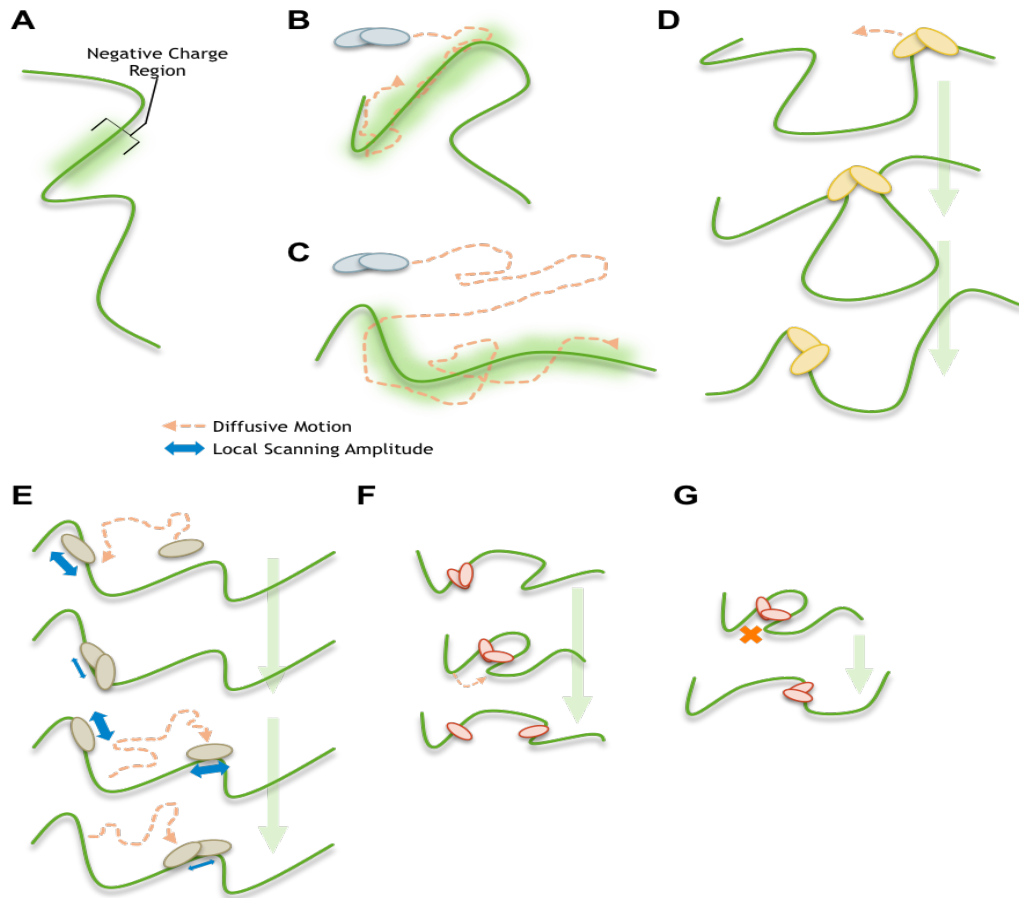


Figure 1.7. DNA scanning by facilitated diffusion. (A) Sketch of the DNA showing the negative charge region, which is important for the electrostatic interaction of APOBEC3 enzymes with DNA and enables facilitated diffusion. (B) Sketch depicting the one-dimensional scanning by sliding of the enzyme (shown as a dimer). Sliding is important for a localized search of the substrate. Dotted line represents the diffusion motion of the enzyme. (C) Sketch representing the three-dimensional scanning by jumping. Jumping enables larger translocations, but lacks a local search processes. During jumping, the enzyme microscopically dissociates from the DNA without diffusing into the bulk solution. (D) Sketch demonstrating three-dimensional scanning by intersegmental transfer, which allows for larger translocations that are mediated by an enzyme having a doubly-bound state. An enzyme with two binding domains binds two regions of DNA simultaneously before dissociating from one region before moving to another. (E) Dynamics of A3G jumping as revealed by single molecule experiments. A3G monomers bind to ssDNA and can slide rapidly. Over time the monomers oligomerize and the mobility of A3G on ssDNA decreases. The oligomers can spontaneously dissociate enabling each monomer subunit to jump to a distal place on the same ssDNA. The local search process that ensues is again slowed by oligomer formation. (F-G) Dynamics of intersegmental transfer of A3C as revealed by studies on A3C dimerization. (F) An intersegmental transfer can occur when the doubly bound state mediated by a dimer is broken through A3C dimer dissociation. (G) Stable dimers of A3C are unable to undergo intersegmental transfer. Reprinted with permission from *Adolph et al.*, 2018 [1].

less efficient than A3G [58, 108, 112]. It was observed that HIV proviral genomes had mutational signatures in the 5'TC context, which implicated an A3 enzyme other than A3G that deaminates within a 5'CC context. It was then demonstrated that this mutational signature could be attributed to A3F [113-115]. Similar to A3G, A3F also interacts with RNA in order to be encapsidated into the virion, and packages more efficiently than A3G into the core of HIV particle [97, 116, 117]. This more efficient packaging of A3F can be attributed to the greater binding affinity of A3F for both DNA and RNA substrates in comparison to A3G. Despite the more efficient packaging of A3F, it induces mutations at lower levels than A3G in both *in vitro* assays and in sequenced proviral DNA [108]. Similar to mutants of A3G that lost their ability to use one-dimensional or three-dimensional motions, wild-type A3F can only use three-dimensional jumping and lacks the ability to slide on ssDNA (Figure 1.6) [108]. This led to fewer clustered mutations in sequenced genomes as A3F lacked the local search mechanism of sliding. This lack of sliding movements was due to a 190NPM192 motif in the connecting domain between the A3F NTD and CTD, and mutation of this motif to an 190NGM192 imparted A3F with the ability to slide [108]. Therefore, the more rigid proline in the wild-type motif of A3F blocked the sliding ability. However, despite adding the one-dimensional search mechanism to the repertoire of A3F, it was still less efficient than A3G because A3F was still less able to jump on DNA. This further emphasizes the requirement for multiple processive scanning mechanisms, as A3F is 4-fold less effective at restricting HIV than A3G in single-cycle infectivity assays[108].

The mutational outcome also differs between A3G and A3F. An early observation with A3G was that it was able to induce stop codons [64]. The ability to induce stop codons arises from the single codon used to code for tryptophan (Figure 1.8A). In the (-) DNA orientation the tryptophan antisense codon contains a deamination motif for A3G (Figure 1.8A). Deaminations of cytosine within the antisense codon can result in two possible stop codons (Figure 1.8A). Depending on the surrounding sequence context, A3F may also be able to induce a stop codon (Figure 1.8A). However, in proviral DNA sequences, the A3G-induced tryptophan mutation to a stop codon is recovered more often than the A3F-induced tryptophan mutation to a stop codon [74, 108] (Figure 1.8B). A3F is more likely to induce a variety of missense (amino acid changing) mutations than A3G, which primarily causes mutations at glycine codons (Figure 1.7B)[74, 108]. The A3G-induced mutations at glycine codons tend to be nonconservative and inactivating mutations whereas the A3F-induced missense mutations have less inactivation ability

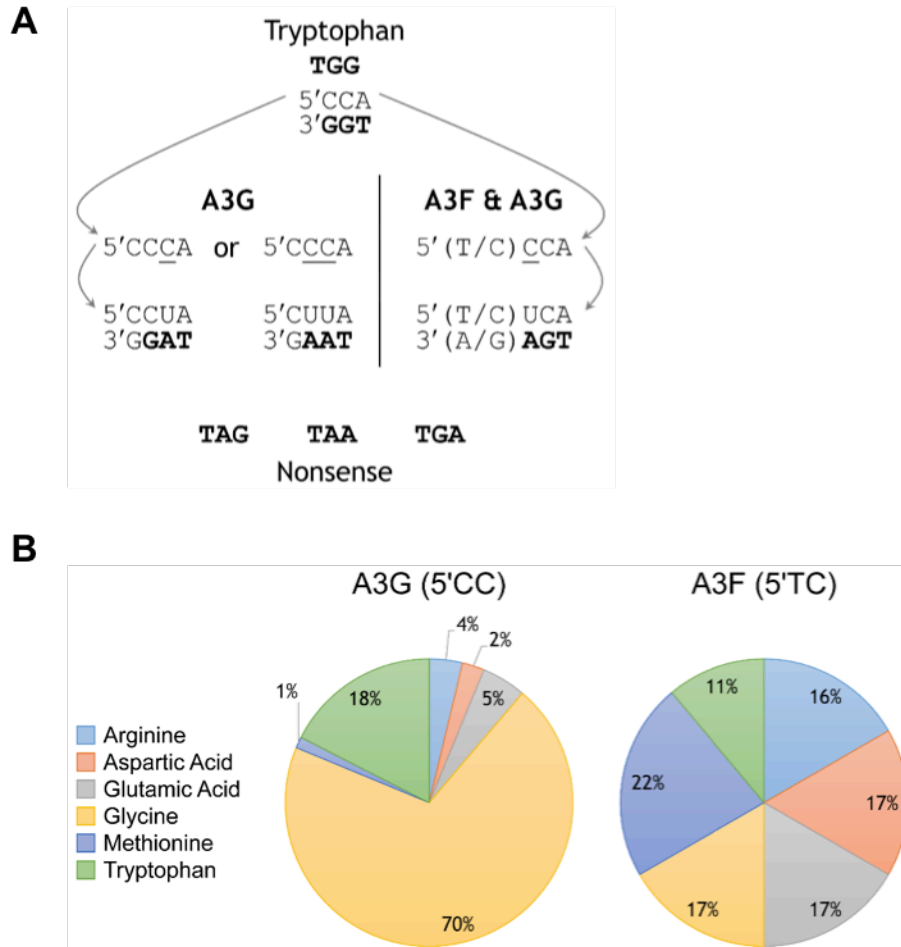


Figure 1.8. Consequence of deaminations at 5'CC and 5'TC motifs. (A) The path to a stop codon through deamination of cytosine in the antisense codon for tryptophan in (-) DNA (5'CCA). (B) Deaminations by A3G at 5'CC motifs in the (-) DNA lead to codon changes in the (+) DNA primarily at glycine. Glycine can be changed to serine, glutamic acid, arginine, or lysine by A3G-induced mutations. The second highest number of mutations occur at tryptophan, which leads to a stop codon. The A3F deaminations at 5'TC motifs in the (-) DNA result in much more variable amino acid changes in the (+) DNA. Sequences of the protease gene of Δ Vif HIV-1 exposed to A3G or A3F were analyzed from Ara et al.[74, 108]. Reprinted with permission from Adolph et al., 2018 [1].

[74, 108, 117, 118]. Similar data has been found using *in vitro* substrates for A3A that also deaminate at 5'TTC[119]. Accordingly, there is evidence that 5'TC modifying A3 enzymes may contribute more than A3G to HIV-1 evolution in the form of immune escape or drug resistance, rather than inducing HIV-1 inactivation [79, 118, 120-124].

1.4.1.3 Deamination-dependent restriction of HIV by A3D

Despite limited *in vitro* data demonstrating the mutation potential of A3D during HIV reverse transcription; A3D has been implicated in the restriction of HIV, as A3D is able to become encapsidated into the virion and is suppressed by HIV Vif. A3D was also demonstrated to restrict HIV in single-cycle infectivity assays with the corresponding mutational signature existing in 5'GC and 5'TC contexts [125]. Despite these findings, A3D is less restrictive than both A3G and A3F [58, 126]. Interestingly, A3D, similar to A3C, has been shown to be more active in other primates. For example, chimpanzee A3D is more restrictive than human A3D despite being encapsidated at similar levels [127]. A study found that a residue within loop 7 reduced the human A3D activity and a C320Y substitution increased the activity more than 20-fold [126]. This suggests that the activity of A3D was lost during evolution from chimpanzees to humans, or the human ortholog has evolved to restrict unique targets not yet identified.

1.4.1.4 Deamination-dependent restriction of HIV by A3H.

A3H is unique to the A3 family as it is the only member that exists as multiple different haplotypes I-VII [59, 128]. These sequence polymorphisms are unique in having varying antiviral activities. Only A3H haplotype II, V, and VII are relevant to HIV restriction while the others are thermodynamically unstable or localized to different cellular compartments [27, 129]. Interestingly, A3H haplotype I, the most common haplotype found in human population, is an unstable variant, suggesting the majority of the population possess an inactive A3H [27, 129]. Also unique to A3H is the Zinc (Z) -domain family to which it belongs. A3 members have been characterized as having three distinct Z-type domains (Z1, Z2, Z3) based on the phylogenetic cluster to which it belongs [23]. Each Zinc-binding domain is required for catalytic activity and is characterized by a conserved Hx₁Ex₂₃₋₂₈Cx₂₋₄C (x is any amino acid) [23]. A3H is the only member that has a Z3-domain; A3D and A3F have two Z2-domains, A3G and A3B have a Z1- (CTD) and Z2- (NTD) domain [23]. A3H is also the only single domain enzyme to oligomerize in solution and similar to A3G and A3F, A3H hap II is able to interact with cellular RNA and become encapsidated into virions [51]. As both virion encapsidation and the biochemical

properties of the enzymes are important for the efficient restriction of HIV, biochemical characterization of A3H has demonstrated that it is a processive enzyme. A3H Hap II and A3H Hap V preferentially deaminate 5' TC motifs and can utilize both one-dimensional sliding and three-dimensional translocations such as jumping and intersegmental transfer [110]. This allows for efficient induction of mutations during reverse transcription. The activity of A3H is improved through its ability to oligomerize, dimerization is uniquely mediated through an RNA intermediate [130-132]. Interestingly, two studies recently identified that A3H may be able to dimerize through an RNA mediated mechanism that involves loops 1 and 7 as well as helix 6 and that the presence of bound RNA alters the activity of the A3H [131, 133]. These findings warrant future work on the exact nature of A3H oligomerization and activity.

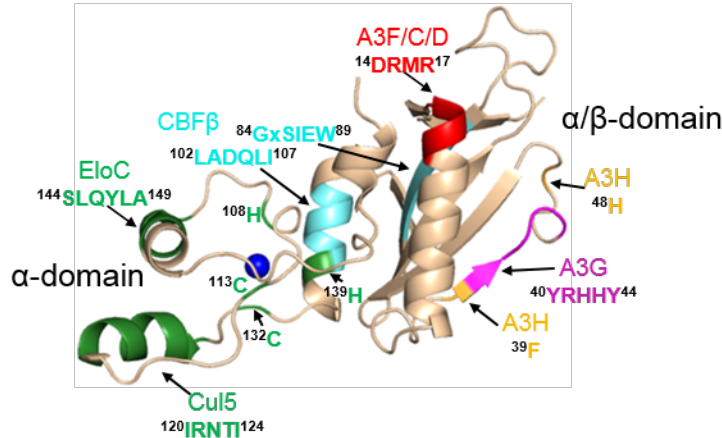
1.4.1.5 Deamination-dependent restriction of HIV by A3C

A3C has been identified to have a limited ability to restrict both viruses and endogenous retroelements. Despite this weak antiviral activity, A3C is expressed in CD4⁺ T cells and effectively degraded through its interaction with Vif and demonstrates evidence of positive selection (recent evolution) indicating that it must have played some role in the restriction of viruses. Previous work has identified that an A3C S188I polymorphism exists in ~10% of people of African descent [27]. Analysis of the HIV restriction abilities of the common (S188) and variant A3C (I188) is the focus of Aim 1 of the thesis research.

1.4.2 Interaction of APOBEC3 with HIV Vif

HIV viral infectivity factor, or Vif, was identified to be essential for HIV replication however the function of Vif was not understood until a human protein CEM15 (later identified as A3) was shown to be suppressed in the presence of Vif [134, 135]. Vif promotes A3 degradation by inducing polyubiquitination of the enzyme. Vif must interact with the host transcription cofactor CBF β to maintain conformational stability [53, 136]. Vif also interacts with host protein Elongin C, which forms an obligate heterodimer with Elongin B (EloB/C) as well as Cullin 5 (Cul 5) (Figure 1.9A) [52, 137, 138]. This interaction with EloB/C occurs through a SLQ motif known as the EloB/C box [52, 64]. Vif interacts with Cul5 through two cysteine residues at position 114 and 133, which are part of a HCCH Zinc binding motif [139]. Binding of Vif/CBF β /EloB/C then recruits the binding of Cul5 and forms the CRL5 ligase complex with subsequent recruitment of Rbx2 [140, 141]. Vif mimics the human protein suppressor of cytosine signaling-2 (SOCS2) that is normally the substrate receptor of the CRL5 ligase in order to

A



B

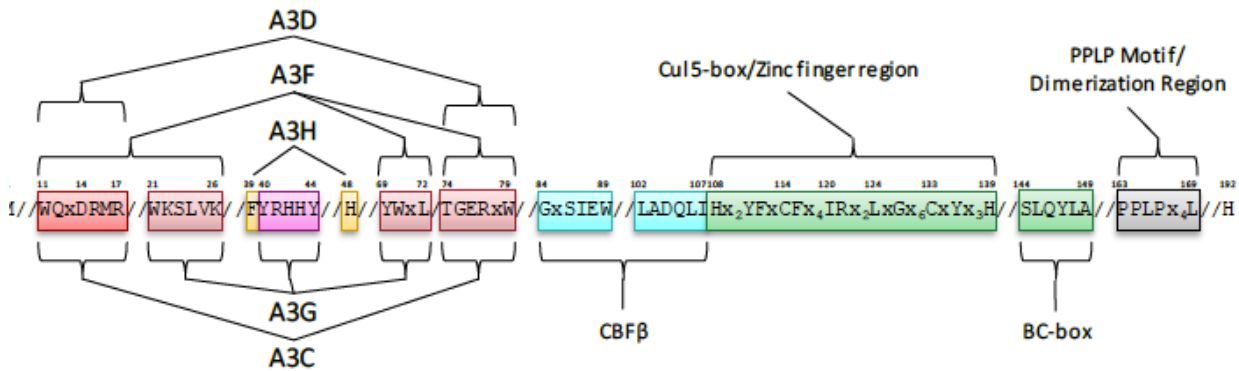


Figure 1.9. Vif interaction interface with APOBEC3. (A) The crystal structure of Vif (wheat) shows that it has two domains on either side of a bound Zinc (blue). The α/β -domain contains the binding interface for CBF β and A3 enzymes. The $^{11}\text{WQxDRMR}_{17}$ motif (red) is used to interact with A3F, A3C, and A3D, the $^{40}\text{YRHHY}_{44}$ motif (magenta) is used to interact with A3G. The α -domain contains two alpha helices that mediate two separate interactions with ELoC (green) and Cul5 (green) (B) Domain organization of Vif. Vif uses specific motifs to interact with A3G (magenta), A3F/A3C/A3D (red), and A3H (orange). In conjunction with these specific motifs, there are shared interaction motifs for A3F and A3G with Vif (pink). CBF β interacts with Vif through two adjacent motifs (cyan). The Zinc finger region (green, amino acids 108-139) coordinates the Zinc through an $^{108}\text{H}_{114}\text{C}_{133}\text{C}_{139}\text{H}$ motif and stabilizes Vif structure, which indirectly enables an interaction with Cullin 5 (Cul5). The BC box mediates an interaction with Elongin C (green). Vif oligomerizes through a PPLP motif (grey). Figure reprinted from Feng et al., 2014.

interact with this ligase [140, 141]. This complex then polyubiquitinates an A3 that is able to interact with Vif. Vif has independently evolved an interaction site for each A3 for which it can promote degradation via the proteasome [142]. The A3 then becomes the substrate of the Cul5 E3 ubiquitin ligase complex, and A3s are targeted for degradation through the proteasome pathway by K48 linked polyubiquitination.

Vif interacts with A3 enzymes on both unique and overlapping sites. Two positively charged regions on Vif comprise the domains that interact with A3G, namely K26, 40YRHHY44, and W70 on one side and residues 14-17 and 21WxSLVK26 on the opposite side [143-145]. The Vif domains that interact with A3F are similarly located, with the main interface being 11WQxDRMR17 and 74TGERxW79 being the secondary interface [144, 146]. Vif also interacts with both A3F and A3G via a 69YWxL72 motif [144]. These interfaces were originally identified through alanine scanning, but have been recently refined through structural modeling and molecular docking studies. Studies have identified that the 14DRMR17 motif that interacts with A3F is also the interface for the interaction with A3C and A3D [144, 146]. Most recently, studies have shown that A3H interacts with Vif through another unique interface composed of residues on the β - sheet of Vif, (residues 40-44) [147]. A3H was previously thought to have a unique interface to A3G however these residues identified overlap with the residues known for A3G. A3H can also interact with Vif on unique residues at position 63 and 90 [147]. These multiple interfaces ensure that Vif is able to counteract distinct A3s effectively.

Conversely, while there are specific interfaces on Vif that are important for the interaction with A3s, each A3 has a specific set of amino acids that interact with Vif. Vif interacts with A3s through three interfaces on the A3s that are grouped as an A3G, A3F/C/D or A3H interface (Figure 1.9B) [143, 146-148]. Although the A3F/C/D interaction site is identified as one, there are small differences in the amino acid interactions for each A3, but overall the interface is largely the same [146, 149]. For A3G, the amino acid D128 is important for the interaction, and mutation of this residue is sufficient to prevent the interaction [148, 150]. Similar to A3G, residue D121 on A3H has been found to be essential for the interaction while E289 on A3F is important for sensitivity to Vif [151-153]. Interestingly, while the interfaces are essential for maintaining an interaction, there is also species specificity to this interaction as SIV Vif is not able to effectively antagonize human A3G, and HIV Vif was shown to not effectively degrade primate A3s. This species specificity demonstrates that the A3s may be a cross species barrier to infection by

viruses spread from related primates, and that Vifs have evolved to antagonize the A3s from the species in which they reside.

1.4.3 Deamination-independent restriction of HIV by APOBEC3

While the primary mechanism of restriction of HIV by the APOBEC3 enzymes is deamination-dependent, studies have observed that A3G and other A3 enzymes also possess deamination-independent modes of restriction. While the majority of viral restriction occurs through deamination of cytosines during reverse transcription, A3 enzymes can also inhibit processes such as tRNA primer binding, initiation and elongation of reverse transcription, plus-strand DNA synthesis as well as inhibition of integration [71, 73, 154-159]. These processes were demonstrated to be due to deamination-independent mechanisms as mutations that affected the catalytic activity of the enzyme (E259Q for A3G and E251Q for A3F) did not change the observed antiviral activity [160, 161]. One mechanism for deamination-independent restriction suggests that A3G or A3F binding to the primer template may cause a “road-block” to RT polymerization, which slows down the extension of the template [71, 74]. A3F is thought to have a stronger deamination-independent mode of restriction than A3G, since it binds substrates with a higher affinity and may impair polymerization to a greater extent. A3F was also shown to inhibit the production of full-length products of RT more efficiently than A3G, and can inhibit the accumulation of late reverse transcriptase products [74]. Interestingly, the deamination-independent mode represents less than 1% of the restriction activity for A3G, whereas for A3F it represents approximately 30% [162]. Therefore, these studies on the deamination-independent mechanisms have highlighted the contribution of these mechanisms to the restriction of HIV, but this mode of restriction is notably less potent than the restriction caused by cytosine deamination.

1.5 Restriction of endogenous retroelements by APOBEC3

Although some A3 enzymes have evolved the biochemical characteristics necessary to restrict HIV-1 replication, their role in protecting the host genome integrity predates immunodeficiency viruses. More of the A3 family members are able to restrict retrotransposons than HIV-1 and retrotransposon pressure on primate genomes is thought to have caused the expansion of the A3 locus (Figure 1.3B) [23]. Retrotransposons replicate through an RNA intermediate and are divided into two groups, long terminal repeat (LTR) retrotransposons (also known as endogenous retroviruses) and non-LTR retrotransposons [15].

LTR retrotransposons (endogenous retroviruses) constitute approximately 10% of the human genome [15], but owing to the accumulation of mutations, endogenous retroviruses, such as HERV-K, have become largely inactive in humans. Reconstruction of endogenous retrovirus genomes has led to the observation that they accumulated many C/G→T/A mutations and were inactivated by A3 enzyme activity [163]. When active, endogenous retroviruses were single-strand (+) RNA viruses that infected germ cells and thus were transmitted vertically, but otherwise replicated similarly to HIV-1 and underwent reverse transcription in the cytoplasm [15].

Non-LTR retrotransposons constitute approximately 30% of the human genome [15]. Three types of non-LTR retrotransposons are still active in humans, the long interspersed element-1 (LINE-1), the short interspersed element (SINE) Alu, and the composite retrotransposon SINE-VNTR-Alu (SVR) [15, 164]. Since Alu and SVR depend on LINE-1 activity for their retrotransposition, the LINE-1 is the model used to study A3-mediated restriction of retrotransposon activity. All A3s except A3G can restrict LINE-1 retrotransposition to various degrees [15, 165]. The LINE-1 RNA is reverse transcribed in the nucleus, and a cellular RNaseH degrades the LINE-1 RNA and exposes the (-) single-stranded DNA, enabling A3 deamination of cytosines [15, 166]. Although multiple labs have demonstrated that A3 enzymes can restrict LINE-1 retrotransposition, it was not known until recently whether this was due to a deamination-dependent or deamination-independent mechanism [15, 166].

There are two potential mechanisms that allow A3 to restrict the activity of LINE-1. First, since A3 enzymes bind RNA with high affinity, it was thought that the A3 enzymes would act as a roadblock to reverse transcription. Similarly, A3s could bind the LINE-1 RNA in the cytoplasm and cause it to accumulate in cytoplasmic RNA processing bodies (P-bodies), preventing nuclear import [167, 168]. DNA sequencing of integrated LINE-1 genomes had little evidence of cytosine deamination, which supports these mechanisms mediated by RNA binding. In addition, many A3 enzymes bind RNA in cells and can be purified from cells in a Ribonucleoprotein (RNP) mass [15, 83]. Second, if A3 deaminations occurred in the nucleus, the uracils formed on the exposed single-stranded (-) DNA would be recognized by the BER enzyme UNG and excised. If the uracils were numerous enough, this would result in DNA breaks and degradation of the LINE-1 (-) DNA [13] (Figure 1C). This mechanisms of DNA degradation may have more time to occur since LINE-1 relies on cellular polymerases and DNA repair enzymes to copy the (-

) DNA to form a dsDNA LINE-1 [166]. This mechanism was supported by experiments using an UNG inhibitor, and demonstrated that A3A mediated inhibition of LINE-1 involved deamination of cytosines in LINE-1 single-stranded (-) DNA. Therefore, the primary fate of the LINE-1 was degradation, not integration of a mutated retrotransposons [166, 169]. However, it does not seem to be a common mechanism for other A3s since A3H uses a deamination independent mechanism to inhibit LINE-1[169]. Thus, although there are some parallels with HIV-1 restriction in that reverse transcription is needed to generate the ssDNA substrate, these restriction mechanisms need to be addressed to determine if A3s other than A3A can deaminate LINE-1 (-) DNA and identify the mechanism responsible for the deamination-independent restriction of LINE-1 retrotransposition. Interestingly, the fate of deaminated LINE-1 (-) DNA is different than for HIV-1 or endogenous retroviruses. Perhaps due to the longer time that the LINE-1 (-) DNA may remain single-stranded owing to a dependence on cellular rather than self-encoded mechanism of (+) DNA synthesis, the A3-catalyzed uracils result in LINE-1 degradation more often than for HIV-1 or endogenous retroviruses.

1.6 Role of APOBEC in somatic mutagenesis

Despite these benefits of A3 enzymes for restriction of retroelements and viruses, there is evidence that there is a cost to this defense system caused by off-target A3-catalyzed mutations to our genomes during our lifetime. This happens when the expression of A3 enzymes occurs in the wrong cell, at the wrong time, and detrimental effects can be caused if deamination occurs in the a gene such as a tumor suppressor [170]. Usually, redundant DNA repair mechanisms can remove uracils and negate most of these promutagenic lesions [13]. However, with the development of Next Generation Sequencing technology and the availability of large sequence datasets, it has become clear that mutations in many cancers have a C/G→T/A bias, and most cancer cells or tumors also overexpress of A3B or A3H hap I mRNA, suggesting that A3s are inducing somatic mutations [16, 17, 171-175]. The uracils created by A3 mutations also can lead to C→G or C→A mutations depending on the repair pathway initiated by the abasic site after APE-mediated removal [176, 177] (Figure 1.2A). Several lines of evidence demonstrate that A3 enzymes provide tumor cells with a “just right” rate of mutagenesis, which provides the genetic diversity for the tumor to adapt to selection pressures in the tumor environment [170]. For example, one report found that women diagnosed with Estrogen Receptor positive breast cancer that also have high *A3B* mRNA expression are associated with poor survival [178]. The resulting

increase in mutation rate has also been shown to lead to tamoxifen therapy resistance [179]. A3B is the best studied A3 in cancer mutagenesis and was identified to contribute to not only breast cancer progression but also head/neck, lung, bladder, cervical, and ovarian cancer [17, 175, 180, 181]. Since A3 enzymes each deaminate cytosine in a specific nucleotide sequence context, A3B activity was identified by retroactive sequence analysis in combination with *in vitro* analysis of the preferred nucleotide sequence context [17, 175]. The A3B “footprint” is mutations at 5'ATCA (underlined base becomes mutated) [175, 182]. Additionally, studies have found the A3A mutational footprints of 5'TTCA or 5'CTCA in tumors, but no corresponding A3A expression, which suggested that A3A may be upregulated early in tumor development, but inactivated later. This may be due to A3A being the most active deaminase, which could cause cell death through its activity over time, rather than a selective advantage [182-184]. In addition or alternatively, A3H hap I is able to induce mutations in breast and lung cancer cells and there is a high association of the A3H hap I allele in an A3B^{-/-} genetic background [174]. The A3H hap I footprint is 5'CTCA and overlaps with A3A [110, 174]. Thus more tumor based studies are needed to differentiate the activity of A3A and A3H hap I. Interestingly, A3A has been demonstrated to deaminate methyl-cytosines (mC) with high efficiency, unlike A3H, and is able to deaminate mC in CpG islands (mCpG) [130, 185]. Differentiating mutational contexts from mCpG and CpG may allow further identification of the A3 involved in addition to A3B.

Importantly, the A3 enzymes that induce somatic mutagenesis must localize to the nucleus and deaminate transiently available ssDNA created during the processes of transcription or replication. For a related family member AID, which normally deaminates the immunoglobulin genes to enable antibody maturation and class switching [11], targeting of AID to the correct genomic region and maximizing its access to both strands of the genomic DNA requires many interacting partner [186]. Although AID can access other regions of the genome, it is thought that interactions with transcription machinery occur to facilitate its specific role in deaminating cytosine in immunoglobulin genes [20, 21]. These protein-protein interactions may not only assist in targeting immunoglobulin genes during B-cell activation, but may be required for AID to catalyze deaminations since its catalytic rate is slow [187]. The slow catalytic rate of AID may be a protective mechanism against AID-catalyzed “off-target” activity [187, 188]. In contrast, *in vitro* studies indicate that the A3 enzymes relevant to cancer mutagenesis are able to deaminate cytosines during transcription or replication without the requirement for interacting partners

[189]. However, while no interacting partners have yet been identified, there is the potential that they may also facilitate A3 mutagenesis. In contrast to AID, studies in yeast and analysis of TCGA sequences have identified that A3 enzymes primarily deaminate ssDNA on the lagging strand during DNA synthesis, deaminate less on the leading strand, and do not show significant amounts of deamination during transcription [190-194]. These studies are based on the analysis of strand biases that would place deaminations either on the lagging strand or nontranscribed strand of DNA during transcription.

Interestingly, in an understudied area of A3 biology, it has been proposed that cytoplasmic A3 can deaminate incoming foreign DNA or stress associated cytoplasmic dsDNA released from mitochondria [12, 195, 196], and this may play a role in modulating dsDNA induced inflammation. Cytoplasmic dsDNA can activate RIG-I inflammatory pathways if it is transcribed into dsRNA by RNA polymerase III [12]. A3 enzymes presumably inhibits this proinflammatory process by deaminating the ssDNA generated during transcription, which leads to DNA degradation by UNG, or prevents reannealing of the DNA strands [12]. This transcription-associated deamination also appears lack any targeting mechanisms.

To avoid degradation by cellular enzymes like A3s, and to promote DNA processing the ssDNA generated during DNA replication, recombination, and transcription is protected by ssDNA binding proteins Replication Protein A (RPA) or RAD51 [197, 198]. A3 enzymes must be able to breach this protective barrier to access the ssDNA. Under normal replication conditions, The helicase moves along the leading strand immediately in front of the polymerase, However, if the polymerase is blocked by a lesion in the DNA or imbalances in dNTP pools, then the helicase can become uncoupled and will unwind DNA without concomitant synthesis, generating long stretches of ssDNA [199, 200]. The ssDNA is then bound by the ssDNA binding protein RPA to protect from nucleases or other chemical damage. A3s must compete with RPA to access to the ssDNA substrate [199, 200]. Thus, in a normal cell, A3 must access ssDNA between the polymerase and helicase, which are usually traveling along the DNA rapidly [200, 201] (Figure 1.10). Although there are ssDNA gaps on the lagging strand, which is probably why this strand is favored by A3 enzymes, access to the ssDNA on the lagging strand also requires the A3 to displace RPA (Figure 1.10).

Replication stress is a hallmark of cell transformation and this stress causes the uncoupling of the polymerase, which facilitates A3 deamination by generating ssDNA targets [200, 202, 203].

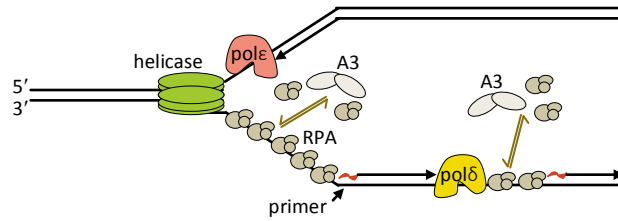


Figure 1.10. Simplified view of a replication fork. At the replication fork, single stranded DNA in both the leading and lagging strand is protected by bound RPA. There is more RPA bound ssDNA in the lagging strand than the leading strand due to discontinuous synthesis of the lagging strand. In order for A3 enzymes to access the ssDNA formed in the lagging strand, they must be able to compete with RPA. Reprinted with permission from *Adolph et al.*, 2018 [1].

The generation of increased amounts of ssDNA promotes the activity of A3 and increases the mutational frequency. Thus, it is notable that A3B is not problematic until a cellular alteration occurs that increases A3B expression and increases replication stress [177, 192, 203]. In contrast, the ssDNA generated on the nontranscribed strand during transcription is accessible at multiple times during the cell cycle of a normal or transformed cell, especially during RNA polymerase stalling [20, 21]. Although A3 enzymes would still need to access transiently available nontranscribed ssDNA that interacts with the RNA polymerase, the ssDNA is solvent accessible [204]. Despite this, the A3 enzymes do not use ssDNA generated during transcription as a major substrate, in contrast to AID. This may be due to the lack of protein-protein interactions that recruit A3 enzymes to sites of transcription or a lack of temporal resolution in cellular studies.

While the APOBEC3 enzymes were discovered to be antiviral enzymes over 15 years ago, there still remains many questions on the biochemical mechanisms the APOBEC3 enzymes use in order to find their target ssDNA. The focus of the presented work is to determine how these enzymes find the ssDNA substrates in vastly different biological contexts, both as an innate immune factor that restricts HIV replication and as a threat to genomic integrity by deaminating “off-target” genomic DNA. Despite the apparent differences in these two systems, the APOBEC3 enzymes still must be able to have an efficient biochemical search mechanism to find and deaminate ssDNA that is transiently available during reverse transcription, replication and transcription. Through careful biochemical analyses of A3s on substrates that model those encountered in cells, the biological functions of the APOBEC3 enzymes can be better understood.

2.0 Hypothesis and Objectives

2.1 Rationale and Hypothesis

APOBEC3 enzymes are a family of host restriction factors originally expanded by gene duplication to restrict endogenous retroelements. While many of these original targets are now inactivated, their presence served to provide the evolutionary pressure to maintain the APOBEC3 gene locus, and adapt the family to act on target single-stranded DNA for which they were not originally intended. The APOBEC3 family can be considered a “double-edged” sword. They act as essential components to our innate immune response, but if unregulated can cause damage to our genomic DNA. However, how the APOBEC3 enzymes can act on such diverse DNA substrates, both during HIV replication and cellular transcription and replication, remains unknown. Characterization of the HIV restriction capacity of A3 primate orthologs, further determining if the ability of A3 to manipulate retroelement polymerases in a deamination independent manner is a secondary restriction mechanism, and determining the biochemical mechanisms of A3 genomic mutation, will aid in defining the biochemical functions for these disparate roles of A3s. This understanding will aid in determining their suitability for HIV-1 therapy and predict if the A3 enzymes are too detrimental for the host to maintain.

My PhD thesis aims to determine the biochemical restriction mechanisms of primate A3 enzymes in the presence and absence of Vif, the ability of the A3 enzymes to interfere with reverse transcriptase processes, and for A3 enzymes to catalyze mutations during transcription and replication. The central hypothesis to this work is that A3 enzymes utilize unique biochemical mechanisms to access ssDNA during viral replication and ssDNA generated during transcription and replication. Which mechanisms the A3s use is dependent on the biochemical characteristics of the enzyme.

2.2 Objective

The objectives of my PhD work are as follows:

1. Characterize the HIV restriction capacity of the APOBEC3C enzyme through examination of the interaction of APOBEC3C with ssDNA substrates *in vitro* and in cells.
2. Characterize the deamination-independent mechanism of APOBEC3 restriction of HIV through modulation of reverse transcriptase processes.
3. Characterize the ability of APOBEC3B to mutate genomic DNA through biochemical assays of transcription and replication.

3.0 A single nucleotide polymorphism in APOBEC3C enhances restriction of lentiviruses.

Cristina J. Wittkopp^{1,2,3}, Madison B. Adolph⁴, Lily I. Wu^{2,3}, Linda Chelico⁴, Michael Emerman^{2,3}.

¹Department of Microbiology, University of Washington, Seattle, Washington, United States of America. ²Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America. ³Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America. ⁴Department of Microbiology and Immunology, College of Medicine, University of Saskatchewan, Saskatoon, Canada.

The information in this chapter was previously published:

Wittkopp, C.J., Adolph, M.B., Wu, L.I., Chelico, L., Emerman, M. (2016). A single nucleotide polymorphism in APOBEC3C enhances restriction of lentiviruses. *PLoS Pathogens*, 12(10): e1005865.

The PLOS journals state that the published materials are licensed by the respective authors of such articles for use and distribution by them subject to citation of the original source in accordance with the Creative Commons Attribution (CC BY) license.

All experiments in this chapter were performed by C.J.W. except Figures 3.5, 3.6, and 3.7 that resulted from experiments that were performed by M.B.A. and Figure 3.8B that resulted from an experiment performed by L.I.W. C.J.W., M.B.A., L.C., M.E. conceived and designed the experiments. C.J.W., M.B.A. analyzed the data. C.J.W., M.B.A., L.C., M.E. wrote the article. C.J.W., M.B.A., L.I.W., L.C., M.E. revised the article.

3.1 Abstract

Humans express seven human APOBEC3 proteins, which can inhibit viruses and endogenous retroelements through cytidine deaminase activity. The seven paralogs differ in the potency of their antiviral effects, as well as in their antiviral targets. One APOBEC3, APOBEC3C, is exceptional as it has been found to only weakly block viruses and endogenous retroelements compared to other APOBEC3s. However, our positive selection analyses suggest that APOBEC3C has played a role in pathogen defense during primate evolution. Here, we describe a single nucleotide polymorphism in human APOBEC3C, a change from serine to isoleucine at position 188 (I188) that confers potent antiviral activity against HIV-1. The gain-of-function APOBEC3C SNP results in increased enzymatic activity and hypermutation of target sequences when tested in vitro, and correlates with increased dimerization of the protein. The I188 is widely distributed in human African populations, and is the ancestral primate allele, but is not found in chimpanzees or gorillas. Thus, while other hominids have lost activity of this antiviral gene, it has been maintained, or re-acquired, as a more active antiviral gene in a subset of humans. Taken together, our results suggest that APOBEC3C is in fact involved in protecting hosts from lentiviruses.

3.2 Introduction

The *APOBEC3* locus encodes seven cytidine deaminase proteins that inhibit endogenous retroelements, lentiviruses such as HIV-1, and other viruses [205]. The *APOBEC3* locus arose through duplication events on chromosome 22[206] of cytidine deaminase domains, resulting in single domain *APOBEC3*s (*APOBEC3A*, *APOBEC3C*, and *APOBEC3H*) and double-domain *APOBEC3* genes (*APOBEC3B*, *APOBEC3D*, *APOBEC3F*, and *APOBEC3G*). In order for APOBEC3 proteins to restrict lentiviruses such as HIV-1, they are packaged into virions, brought to a target cell, and deaminate cytidines on ssDNA during reverse transcription, resulting in cytidine to uracil mutations in the viral genome. APOBEC3 proteins exert selective pressure on primate lentiviruses, which have evolved to encode a protein, Vif, which targets APOBEC3 proteins for proteasomal degradation.

Vif-mediated antagonism leads to non-synonymous mutations in *APOBEC3* that allow for escape from Vif but maintenance of antiviral activity [207]. Lentiviruses, in turn, select for Vif alleles that target these APOBEC3 variants, leading to further adaptive evolution of *APOBEC3* genes through non-synonymous mutations. As such, enrichment of the rate of nonsynonymous

mutations (dN) compared to the rate of synonymous mutations (dS) is a common signature of antiviral genes [207]. This enrichment, or $dN/dS > 1$, is referred to as positive selection.

APOBEC3 genes involved in blocking viral replication are expected to exhibit signatures of positive selection. Specifically, APOBEC3s involved in lentiviral restriction should have signatures of positive selection at the Vif:APOBEC3 interface[208].

There is considerable variation in the antiviral activity of each of the seven human APOBEC3 paralogs. APOBEC3G potently inhibits *vif*-deleted-HIV-1 (Δvif) [112]. Human APOBEC3D, APOBEC3F, and APOBEC3H also inhibit HIV-1 (Δvif), but to a lesser extent than APOBEC3G [58, 112, 115, 126]. In contrast, APOBEC3A and APOBEC3B do not potently block HIV infection of T cells [58, 112, 115, 209], which are the primary target of HIV (although a target-cell effect has been reported in monocytes for APOBEC3A) [6]. Instead, APOBEC3A and APOBEC3B drastically inhibit replication of endogenous retroelements and some DNA viruses [210-215]. In studies that compare the ability of the seven human APOBEC3s to restrict lentiviruses and endogenous retroelements, the only APOBEC3 that has weak activity against both lentiviruses and endogenous retroelements is APOBEC3C [58, 112-114, 211, 215-219]. For another *APOBEC3* gene, *APOBEC3H*, the most common human variant does not block HIV infection although other haplotypes exist that potently restrict lentivirus replication [59]. In fact, one haplotype of APOBEC3H restricts HIV-1(Δvif) as potently as APOBEC3G [59] and has been shown to impact clinical outcomes in HIV-1+ patients[66, 220, 221]. Thus, we considered the possibility that while the common human haplotype of *APOBEC3C* encodes a protein with little antiviral activity, other variants of *APOBEC3C* may in fact encode more potent anti-lentiviral proteins. Compellingly, the Vif protein of HIV-1 targets human APOBEC3C for proteosomal degradation [222]. Moreover, APOBEC3C mRNA is highly expressed in the major HIV-1 target cells, activated T cells [223]. Thus, the high expression of APOBEC3C in HIV target cells and the antagonism of APOBEC3C by HIV-1 Vif are consistent with the hypothesis that APOBEC3C may have an overlooked role in combating lentivirus infection.

In this study, we found that *APOBEC3C* has evolved under positive selection in primates in a manner that suggests that APOBEC3C has played a role in blocking primate lentiviruses. This provided motivation to determine if there are naturally occurring variants of APOBEC3C that potently block lentivirus replication. In humans, only one APOBEC3C coding variant is present at a frequency above 1% and this is a serine to isoleucine change at position 188, here

called APOBEC3C I188 [27]. Here, we show that the polymorphism APOBEC3C I188 is present at about 10% frequency in diverse populations throughout Africa, and thus did not recently arise in a particular subpopulation of humans, but is an ancient allele that has likely been circulating in humans for much of human history. Moreover, we show that the APOBEC3C I188 single nucleotide polymorphism (SNP) has about 10-fold more potent anti-lentiviral activity than the common human APOBEC3C variant and has greater *in vitro* cytidine deaminase specific activity. The greater activity of APOBEC3C I188 in turn correlates with its ability to dimerize. Moreover, construction of a forced dimer of APOBEC3C S188 also gains enhanced antiviral activity to a level comparable to APOBEC3G. We show that the APOBEC3C I188 allele is likely the ancestral state since all sequenced Old World monkeys and some great apes carry isoleucine at position 188. However, gorillas, chimpanzees and most humans carry the S188, the apparent loss of function allele. Taken together, our results suggest that APOBEC3C is involved in protecting hosts from lentiviruses, and we speculate that some humans may be afforded some level of additional protection from lentiviruses by a more active antiviral version of this protein.

3.3 Results

3.3.1 *APOBEC3C has evolved under positive selection in primates suggesting an ancient role in protection from pathogens.* In studies that compare the antiviral activity of the seven APOBEC3 paralogs, APOBEC3C consistently has poorer restriction activity than the other paralogs [58, 112, 113, 217, 218]. However, we reasoned that if APOBEC3C is in fact a *bona-fide* restriction factor then we would expect that the gene has an evolutionary signature of positive selection [207]. We performed positive selection analyses of twenty-two *APOBEC3C* sequences derived from eighteen primate species with sequences representing diverse clades of catarrhines, a subdivision of primates including old world monkeys and apes (Figure 3.1A). Among these, multiple sequences were obtained from African green monkeys, because we chose to include three subspecies (vervet, tantalus, and sabeus). The sequences were aligned and tests for positive selection were conducted using maximum likelihood ratio tests comparing M8 (a model that allows positive selection across the gene) to M8a (a model that disallows positive selection). Our results indicate *APOBEC3C* shows a gene-wide signature of positive selection ($p < 0.0008$) (Figure 3.1B).

We next analyzed individual lineages to determine which branches of the *APOBEC3C* tree have signatures of positive selection. Branch analysis identified two branches with

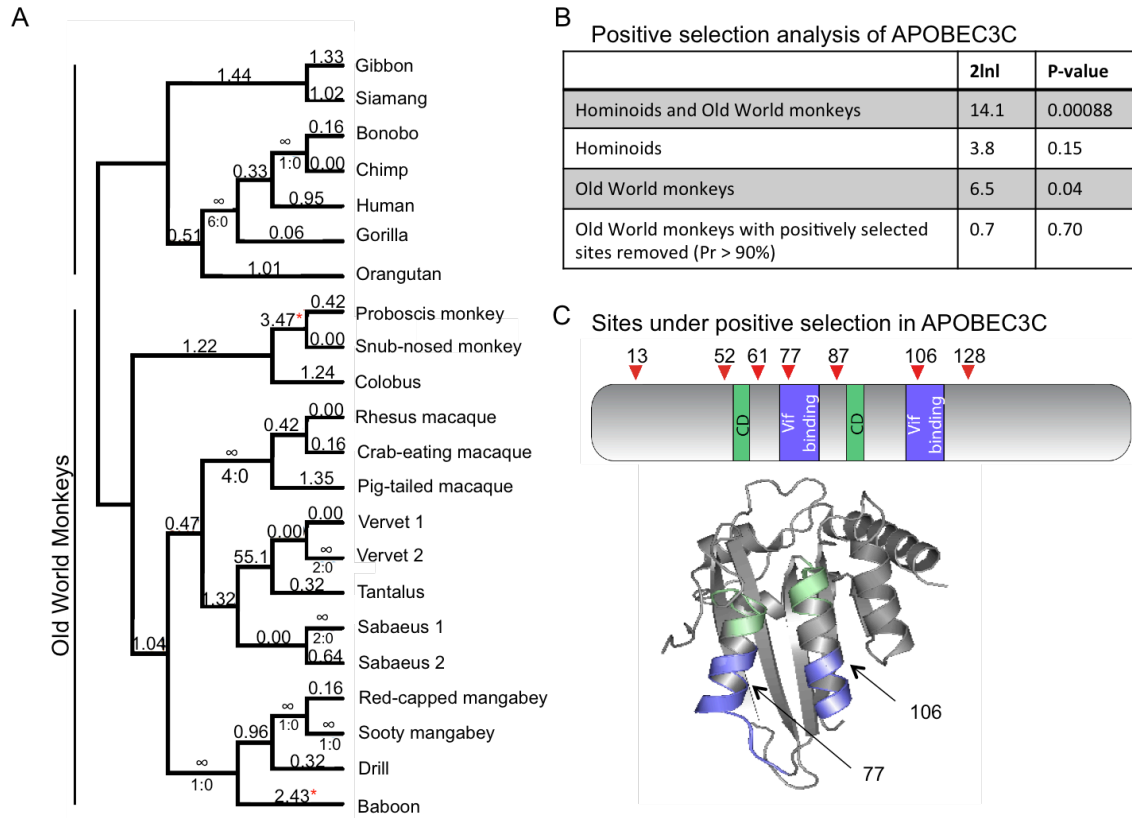


Figure 3.1. APOBEC3C is rapidly evolving in primates. (A) Twenty-two primate APOBEC3C coding sequences were obtained by PCR or from the NCBI sequence database. A phylogeny of APOBEC3C, indicating the branch analysis results of the positive selection tests. The ratio of rate of nonsynonymous changes (dN) and the rate of synonymous changes (dS) that occurred along each branch are shown above each branch. For dN/dS values = ∞ , the total number of non-synonymous changes (N) and synonymous changes (S) are shown (N:S) below the branch. The red asterisks mark the branches where the dN/dS is significantly greater than 1 across the entire gene. (B) Maximum likelihood tests for positive selection, with 2lnl values indicating twice the log difference between the model that allows for positive selection (M8) and the model that does not allow for positive selection (M8a), as well as a P-value to indicate whether the M8 model better fits the data than the M8a model. (C) Sites under positive selection in APOBEC3C are shown in a cartoon diagram, comparing these sites to the Vif binding domain and the cytidine deaminase enzymatic domain (CD). Red triangles depict sites with a posterior probability >0.99 (red triangle). The structure of APOBEC3C[149] is represented, with the Vif binding domain[149] shown in blue. Two of the seven positively selected sites (PP > 0.99) overlap with this domain, are shown with arrows. The cytidine deaminase domain is shown in green.

statistically significant signatures of positive selection, both in Old World monkeys (Figure 3.1A), and while most were not statistically significant, many branches had a dN/dS >1 (Figure 3.1A). Furthermore, we performed M8 vs M8a analysis of the hominoid and Old World monkey clades of the tree separately, and found that the Old World monkey clade has a statistically significant signature of positive selection ($p < 0.05$) (Figure 3.1B). We did not see a statistically significant signature of positive selection in the hominoid-only branch ($p = 0.15$), although this could be due to a smaller sample size ($n = 7$).

For antiviral genes, sites under positive selection often correlate with sites of interaction with a viral antagonist [224]. APOBEC3C is antagonized by the lentiviral protein Vif and the interface of Vif binding has been extensively mapped [149]. If APOBEC3C is in fact an anti-lentiviral gene, the Vif binding interface may be evolving under positive selection. Therefore, we performed a site-analysis to determine which amino acids are under positive selection across the tree. Our analysis indicated seven sites under positive selection (posterior probability > 99%) (Figure 3.1C). Next, we mapped the positively selected sites onto the structure of human APOBEC3C and compared these to the Vif interface of APOBEC3C. Of the seven positively selected sites, two of these, residues 106 and 77, are located within the two helices that are targeted by Vif (Figure 3.1C). Strikingly, residue 106 has been identified as the most important for Vif binding and this interaction has been documented in two separate studies [149, 222]. Thus, *APOBEC3C* has evolved under selection, gene-wide, as well as at the Vif-binding interface. These results suggest that although the common human APOBEC3C variant does not potentially block lentivirus replication, primate *APOBEC3C* may have evolved as an anti-lentiviral protein.

3.3.2 Human APOBEC3C SNP I188 increases antiviral activity. Because the positive selection analyses suggested an ancient or ongoing role of APOBEC3C in lentiviral restriction (Figure 3.1), we re-evaluated human polymorphisms in *APOBEC3C* for potential variants with increased activity. There is only one SNP in *APOBEC3C* above 1% frequency globally, and this is a serine to isoleucine change at position 188 [27]. To evaluate the potential significance of this SNP, we aligned this region of *APOBEC3C* to other human *APOBEC3* genes. Strikingly, we found that in contrast to *APOBEC3C*, the other ten *APOBEC3* deaminase domains all encode a conserved isoleucine at the position homologous to APOBEC3C 188 (Figure 3.2). Thus, the human I188 polymorphism in APOBEC3C actually encodes an amino acid that is highly

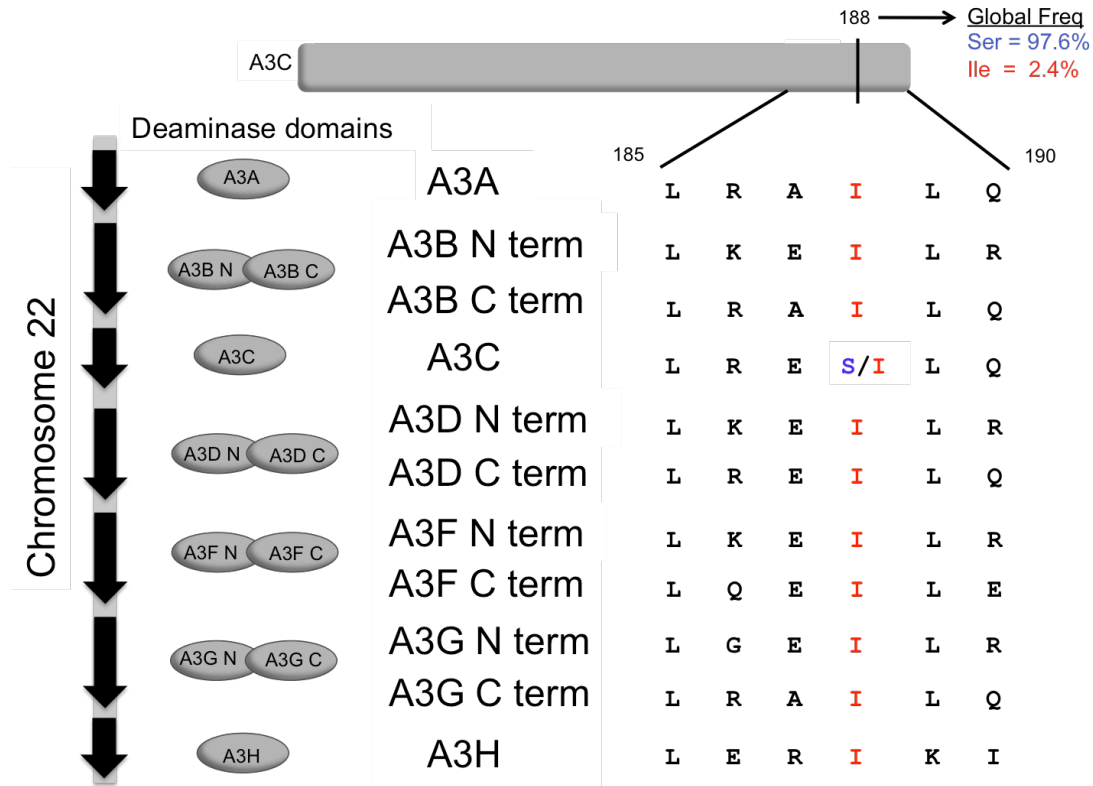


Figure 3.2. I188 is a SNP in APOBEC3C and is the conserved residue in the other six human APOBEC3 paralogs. Alignment of the seven APOBEC3 proteins, using both deaminase domains (N and C terminal) of the double-domain APOBEC3 proteins (11 domains total). The residue homologous to APOBEC3C in the other ten deaminase domains is conserved as an isoleucine, whereas APOBEC3C is the only domain with a serine at that position. However, human APOBEC3C is polymorphic at that position, with an isoleucine at an allele frequency of 2.4% globally[27].

conserved at this position across human APOBEC3s, while the more common APOBEC3C in the human population has a different amino acid at position 188.

Since conserved sequences are often important for function and comparative studies indicate that human APOBEC3C (S188) has weak antiviral/anti-retroelement activity compared to the other human APOBEC3s, we posited that the serine change may contribute to the weak restriction activity of the common variant of APOBEC3C. Therefore, we directly compared APOBEC3C S188 and APOBEC3C I188 for their ability to restrict HIV-1. We transfected the two APOBEC3C variants, S188 and I188, along with VSV-G and an *env-vif*-deleted luciferase-expressing HIV-1 provirus (Δenv , Δvif). Normalized amounts of virus were subsequently used to infect SupT1 cells and infectivity of the viruses was compared by measuring virus-encoded luciferase. Viral infectivity in the presence of no APOBEC3 is set to 100%. APOBEC3G was used as a positive control because it potently inhibits HIV-1 (Δvif). We found that APOBEC3C I188 restricts infectivity of HIV-1(Δvif) to a level approximately ten-fold greater than the common APOBEC3C, S188, (approx. 30% infectivity versus 3%, respectively) (Figure 3.3A) even though both proteins are expressed at similar levels. Furthermore, infectivity assays were conducted as a dose-response in the presence of decreasing concentrations of APOBEC3, and the 188 isoleucine variant restricts HIV-1(Δvif) more potently for all conditions (Figure 3.3B) at similar protein expression levels.

To determine if the APOBEC3C I188 variant has increased potency against another lentivirus, we evaluated its activity against SIVagm, which is a simian immunodeficiency virus that infects African green monkeys. As shown by others, the S188 variant of APOBEC3C restricted infectivity of SIVagm to a greater extent than HIV-1 [209]. However, the APOBEC3C I188 restricted SIVagm infectivity ten-fold more than the restriction caused by APOBEC3C S188 (10 % versus 1% infectivity, respectively, $p < 0.05$) (Figure 3.3C). Some APOBEC3s also restrict endogenous retroelements, such as LINE-1s [213, 214]. However, the APOBEC3C I188 variant does not confer increased restriction of LINE-1 as we have previously published [27] and have repeated for this study (Figure 3.4). Therefore, the human polymorphism in APOBEC3C at position 188 enhances restriction of at least two primate lentiviruses. Thus, we conclude that a SNP in human APOBEC3C has increased anti-lentiviral activity relative to the APOBEC3C encoded by most humans.

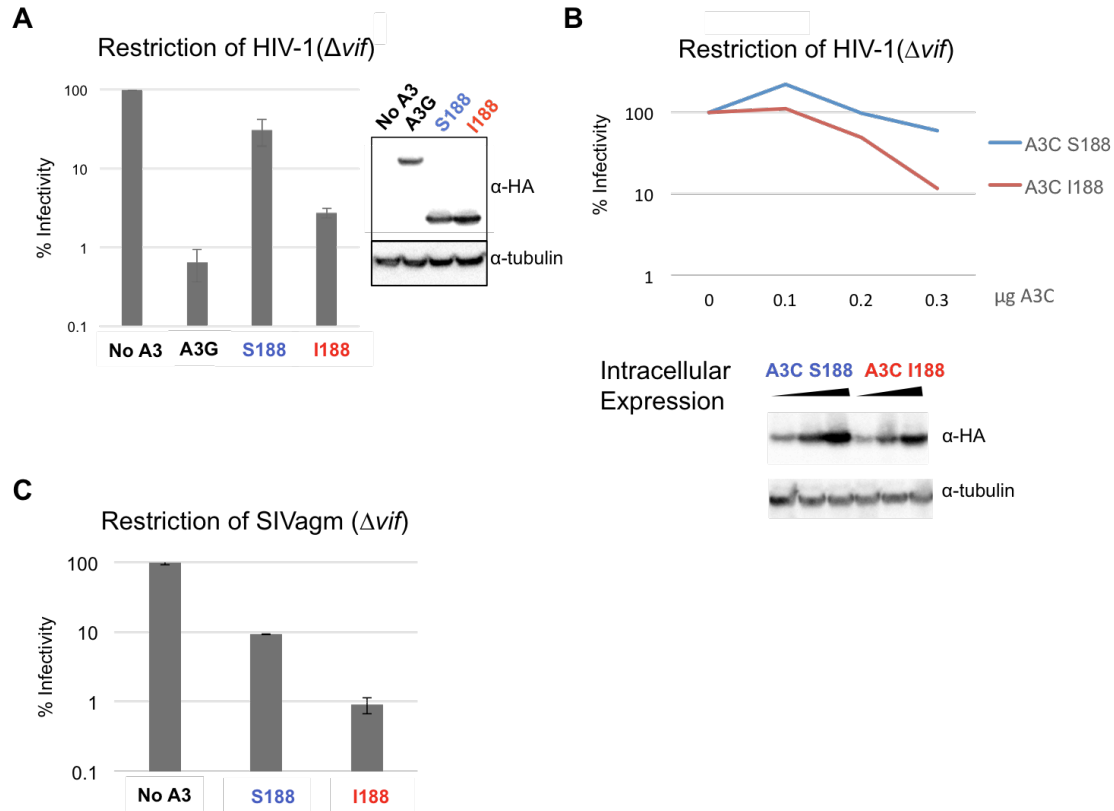


Figure 3.3. APOBEC3C SNP Isoleucine 188 confers increased antiviral activity. (A) Infectivity of HIV-1 Δvif in the absence of APOBEC3 (No A3), APOBEC3G (A3G), APOBEC3C S188 (A3C S188) and APOBEC3C I188 (A3C I188). 0.3 μg of HA-tagged APOBEC3 was expressed in virus-producing cells, and viruses were collected and used for infection. Infectivity in the absence of APOBEC3 is set to 100%. Error bars indicate the standard deviation of triplicate transfections and infections, and this experiment was repeated four times with similar results. Intracellular expression of APOBEC3 was measured by Western Blot using an anti-HA antibody. A section of the blot was probed with an anti-tubulin antibody as a loading control. (B) Dose-response analysis showing restriction of HIV Δvif in the presence of two-fold dilutions of transfected APOBEC3C S188, or APOBEC3C plasmids I188 along with Western blot analysis of APOBEC3C S188, and APOBEC3C I188 protein expression during virus production. This experiment was performed three times, and a representative result is shown. (C) Infectivity of Simian Immunodeficiency virus SIVagm Δvif , in the presence of APOBEC3C S188, APOBEC3C I188. Infectivity is set to 100% for infection with No APOBEC3 present. Error bars indicate the standard deviation of three independent experiments.

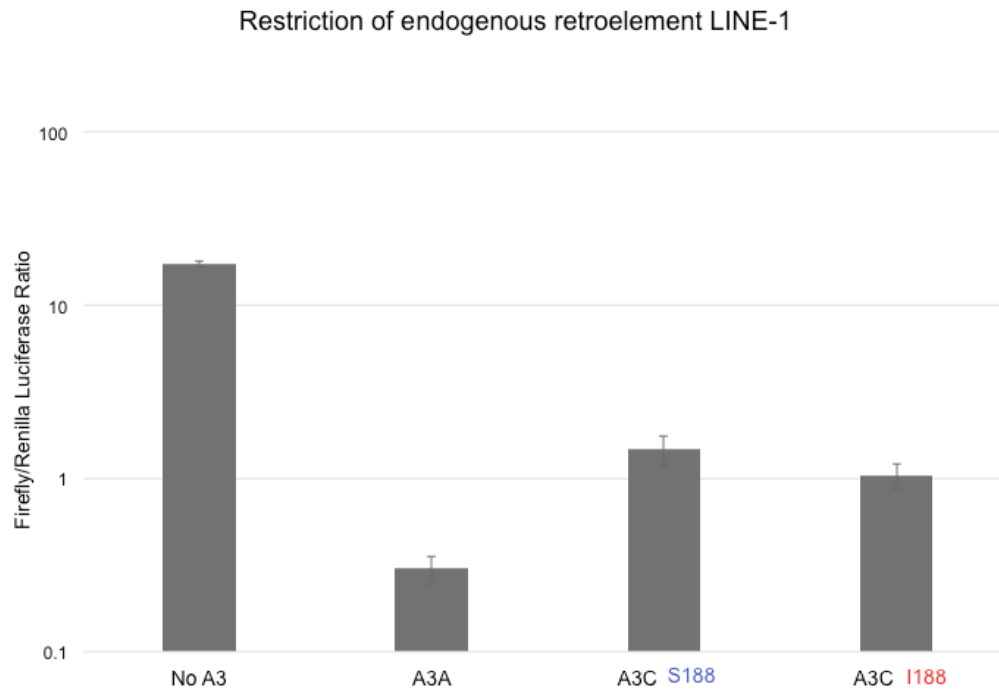


Figure 3.4. Inhibition of LINE-1 replication by APOBEC3A and S188 and I188 APOBEC3C. Ten times more APOBEC3C was used in this assay than APOBEC3A. The LINE-1 plasmid constitutively expresses renilla luciferase, and only expresses firefly luciferase upon retrotransposition. Values are shown as the ration of firefly luciferase expressed over renilla luciferase expression. Averages of three replicates are shown and this experiment was repeated twice.

3.3.3 *Isoleucine at position 188 enhances enzymatic activity in vitro.* We wished to investigate whether or not the more potent antiviral activity of APOBEC3C I188 compared to APOBEC3C S188 could be explained by differences in their inherent enzymatic activity. Thus, each protein was produced by expression in a recombinant baculovirus system, purified as described in the Materials and Methods, and tested for its ability to cause cytidine deamination. We examined APOBEC3C S188 and I188 activity using a ssDNA substrate containing two deamination target motifs (Figure 3.5, top sketch). 5' TTC deamination motifs were used because APOBEC3C preferentially targets this motif [218]. Reactions were carried out as a time-course over 60 minutes and next the substrates were incubated with uracil DNA glycosylase, which modifies uracil-containing DNA and makes it sensitive to cleavage at high pH. Cytidine to uracil mutations leading to DNA cleavage were detected based on a fluorescein label placed between the two deamination motifs. Substrate usage was calculated from integrated gel band intensity of cleaved product at either deamination motif relative to the uncleaved substrate. We found that at all time points substrate usage of APOBEC3 I188 was higher than S188, and by 60 minutes I188 had led to twice as many cleavage events as S188 (Figure 3.5, top). The specific activity of APOBEC3C was determined by calculating the picomoles of substrate used (or deamination events) per microgram of enzyme per minute on a 118 nt ssDNA. The specific activity values were calculated using initial reaction times where the substrate usage was in the linear range (Figure 3.5, bottom left). We found that APOBEC3C S188 had a specific activity approximately 10-fold lower than I188 (0.010 pmol/μg/min vs 0.130 pmol/μg/min) (Figure 3.5, bottom right). Therefore, the I188 APOBEC3C more rapidly deaminated cytosines *in vitro* than S188. Since APOBEC3C I188 has greater cytidine deaminase activity *in vitro* than APOBEC3C S188 (Figure 3.5), we predicted that it would also have a higher mutational frequency than the APOBEC3C S188. To test this prediction, we used a model *in vitro* system that reconstitutes reverse transcription of RNA to DNA, and observed the ability of APOBEC3 enzymes to induce mutagenesis. The template includes the gene *lacZα*, and blue/white screening was performed to identify mutated reverse transcription products. White colonies, representing templates that were mutated, were then sequencing and the number of mutations induced by each APOBEC3 were quantified. We found that addition of APOBEC3C I188 induced two-fold higher clonal mutation frequency compared to APOBEC3C S188 (Figure 3.6, 0.33×10^{-2} mutations/bp versus 0.15×10^{-2}

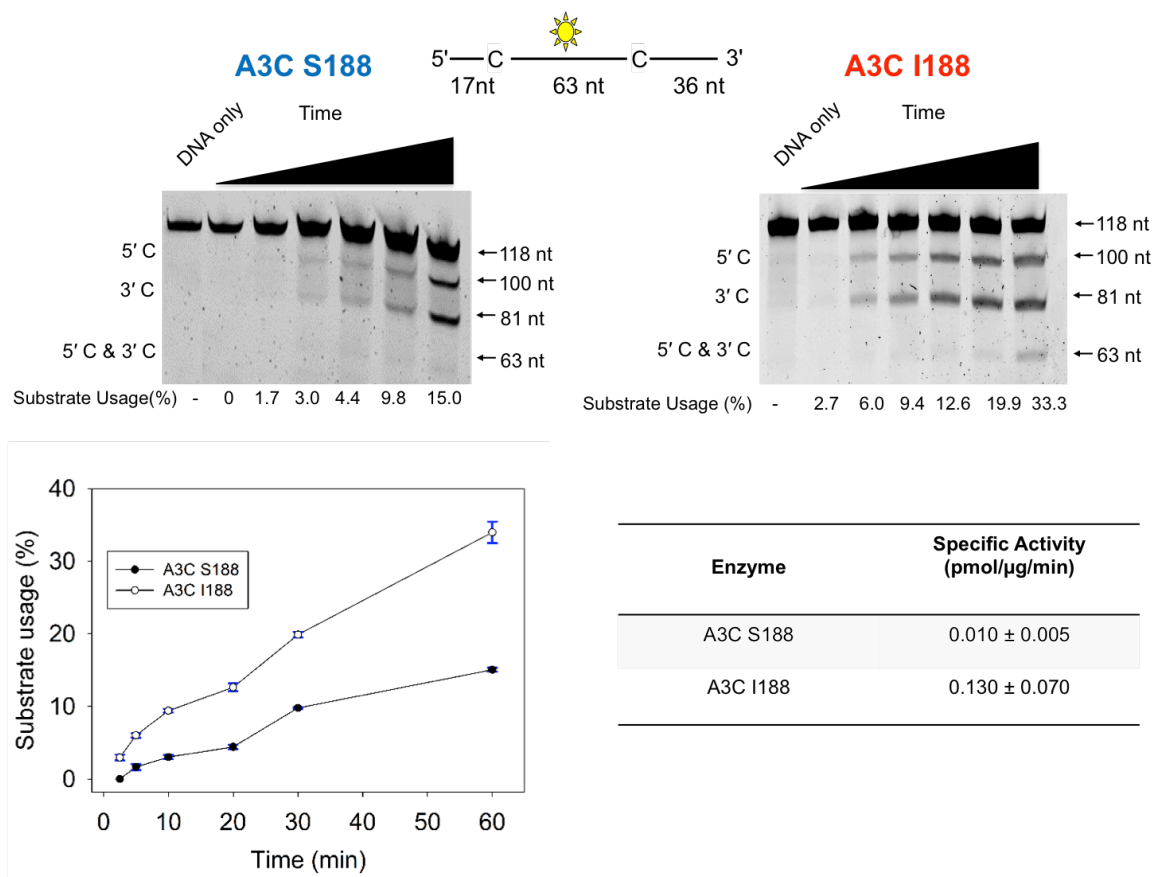


Figure 3.5. In vitro characterization of APOBEC3C S188 and I188. (Top) The specific activity of APOBEC3C S188 and I188 was determined by incubating the enzyme with a 118 nt ssDNA substrate with an internal fluorescein label (yellow star) and 2 possible sites for cytidine deamination (marked as “C”). Single deaminations of the 5'C and 3'C are detected as the appearance of fluorescently labeled 100 nt and 81 nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. Substrate usage is quantified for below each lane of the gels. (Bottom) The substrate usage during a 60 min time course was plotted from three independent experiments (bottom left) and used to calculate the specific activity of the enzymes (bottom right).

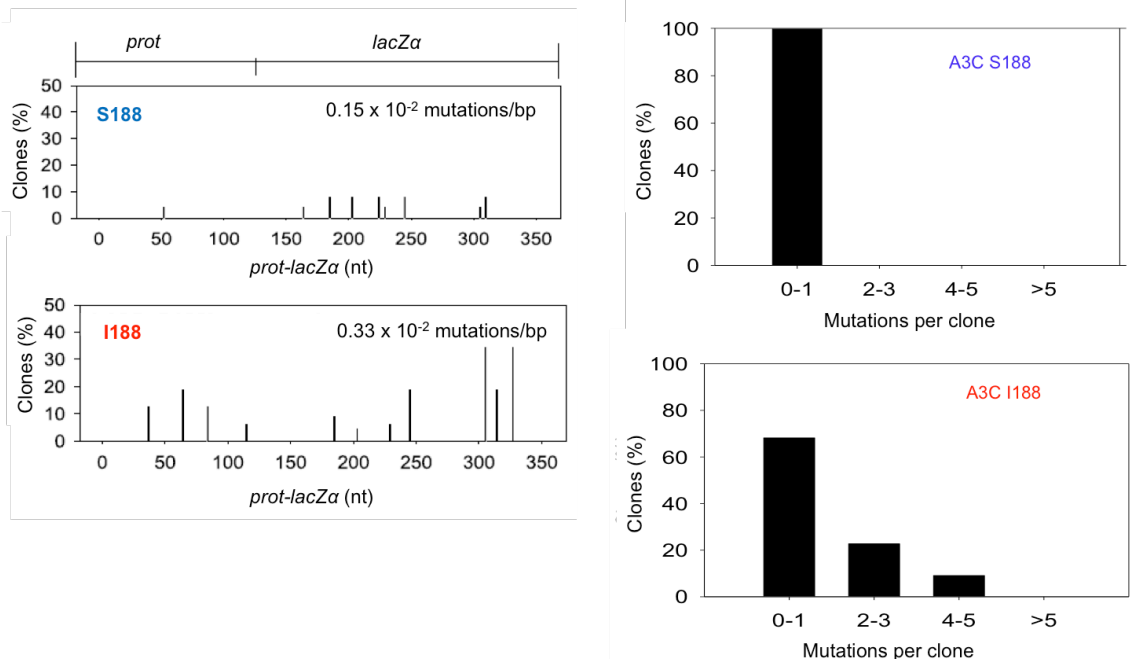


Figure 3.6. In vitro reverse transcription assay. An *in vitro* HIV replication assay was utilized to determine the APOBEC3C enzyme ability to catalyze deaminations during proviral DNA synthesis. This system reconstitutes reverse transcription of (–)DNA and synthesis of (+)DNA by using a substrate which contains a polypurine tract (PPT), 120-nt of the protease gene (*prot*) of HIV, and a *lacZα* reporter. (Left) G→A mutations are scored for each clone and mutational spectra are plotted as the percentage of clones containing a mutation at a particular location (nt) in the 368 nt *prot-lacZα* construct. The number of mutations per base pair for each APOBEC3C is indicated above the spectra. (Right) Histograms depicting the population distribution of mutations per *prot-lacZα* for the APOBEC3C enzymes.

² mutations/bp, respectively). For reactions containing APOBEC3C S188, 100% of clones had zero to one G→A mutation. In contrast, the presence of APOBEC3C I188 caused a noticeable shift in the number of G→A mutations with 32% of clones having more than one mutation and up to four to five mutations in some individual clones (Figure 3.6). Overall, isoleucine at position 188 increased the APOBEC3C-induced mutagenesis of ssDNA *in vitro*.

3.3.4 Dimerization correlates with enhanced antiviral activity of human APOBEC3C.

Previous studies have reported that the S188 variant of APOBEC3C is a monomeric protein, both in solution [149] and in cells [51]. Indeed, by size exclusion chromatography we also found that baculovirus/Sf9-produced APOBEC3C S188 (the common variant) is monomeric (Figure 3.7A, apparent molecular weight 17 kDa). However, the baculovirus-produced APOBEC3C I188 was in equilibrium between monomer and dimer forms (Figure 3.7A, apparent molecular weight 21 kDa and 42 kDa, respectively). We confirmed this result using an alternative method of cross-linking the proteins in solution followed by SDS-PAGE and Western blotting. Baculovirus/Sf9-produced APOBEC3C A3C S188 or I188 were incubated in the absence or presence of 10 μM bis(sulfosuccinimidyl)suberate (BS3), an amine-amine chemical crosslinker, and then visualized through SDS-PAGE and Western blotting (Figure 3.7B). A3C S188 remained monomeric in the presence of crosslinker, whereas A3C I188 was dimeric in the presence of the crosslinker. The observation that the isoleucine residue at position 188 was able to shift the oligomeric profile of APOBEC3C suggests that residue 188 is important for dimerization. Dimerization has been previously correlated with improved APOBEC3 catalytic activity because it enables efficient scanning of ssDNA to find cytosine targets for deamination.[107]. This provides a potential explanation for the increased *in vitro* enzymatic activity of A3C I188. In order to further test the effects of dimerization of A3C on antiviral activity, we constructed an artificial dimer that consists of two tandem S188 APOBEC3Cs (Figure 3.8A) and tested the anti-lentiviral activity of this protein. We used the linker that naturally exists between the N- and C-terminal domains of the two double-domain APOBEC3s, APOBEC3D and APOBEC3F, which are the APOBEC3 proteins with the highest sequence identity shared with APOBEC3C. This linker consists of amino acids Arg-Asn-Pro followed by the second APOBEC3 domain starting at Met12 (labeled Met12' here—see schematic at top of Figure 3.8A). Western blot analysis shows that this artificial double domain APOBEC3C is expressed in cells and runs at about the same molecular

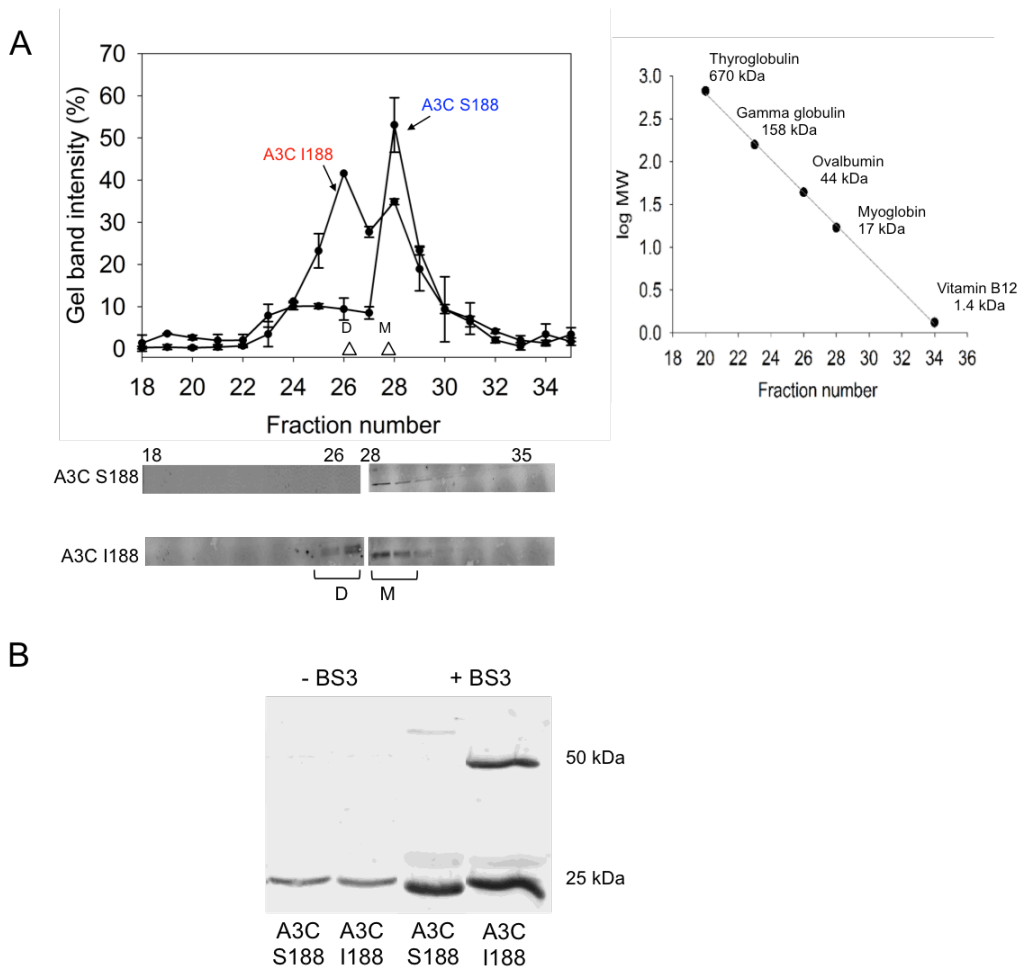


Figure 3.7. Purified APOBEC3C I188 forms dimers in solution. (A) Size exclusion chromatography profiles of the APOBEC3C S188 and I188 from a 10mL G200 Superdex column were used to calculate the oligomerization states of the enzymes. Molecular weights were calculated by comparing to a calibration curve (see inset on right). When APOBEC3C S188 and APOBEC3C I188 were loaded onto the column, APOBEC3C S188 was a monomer in solution (apparent MW 21 kDa in peak fraction), whereas APOBEC3C I188 could form dimers (apparent MW 42 kDa in peak fraction) in addition to monomers (apparent MW 21 kDa in peak fraction). Chromatograms were made using the integrated gel band intensities from three independent experiments of each protein fraction after resolution by SDS-PAGE. A representative Western blot of the size exclusion chromatography fractions is shown. (B) A3C S188 or I188 were incubated in the absence or presence of 10 μ M bis(sulfosuccinimidyl)suberate (BS3) crosslinker (indicated as -BS3 or +BS3) and then visualized through SDS-PAGE and Western blotting. The Western blot demonstrates that A3C S188 remained monomeric in the presence of crosslinker, whereas A3C I188 was both monomeric and dimeric in the presence of the crosslinker. Molecular weight standards are indicated.

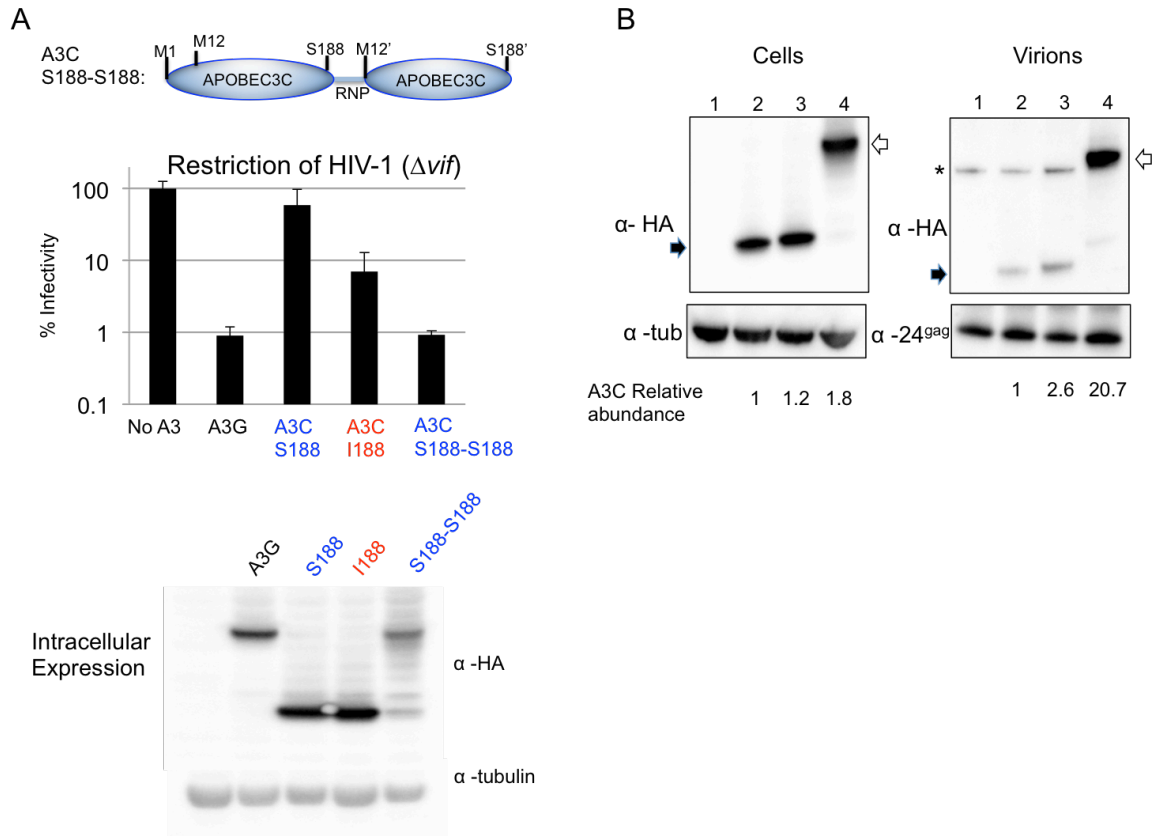


Figure 3.8. A synthetic dimer of APOBEC3C has increased antiviral activity (A) Cartoon schematic showing the sequence of the double-domain APOBEC3C. Restriction of HIV-1 Δvif by APOBEC3G(A3G), APOBEC3C S188 (A3C S188), APOBEC3C I188 (A3C I188) and the double-domain APOBEC3C (S188-S188). 0.3 μ g APOBEC3 was used in this assay. Error bars represent standard deviation of four independent transfections and infections. A Western blot for expression of the APOBEC3 proteins is shown, and is representative of three experiments. (B) Packaging of APOBEC3C variants. Left side: Intracellular expression. Right side: Proteins in the pelleted virions. The intracellular blot was probed with antibody to the HA tag and with antibody to tubulin. The virion pellet blot was probed with antibody to the HA tag and with an antibody to p24gag. A background band in the virion blot with the HA antibody is marked with an asterisk. The single domain APOBEC3C proteins are marked with a solid arrow, while the synthetic double domain APOBEC3C is marked with an open arrow. Relative quantitation of the amounts of APOBEC3C is shown under the panels. An HIV provirus is transfected in each condition; Lanes 1, no APOBEC3; lanes 2, APOBEC3C I188; lanes 3, APOBEC3C S188; lanes 4; the double-domain APOBEC3C (S188-S188).

weight as the natural double domain APOBEC3 protein, APOBEC3G (Figure 3.8A). We examined the antiviral activity of the synthetic dimer *APOBEC3C* gene (with S188 in both domains, called S188-S188) compared to APOBEC3C S188 and APOBEC3C I188 (Figure 3.8A) against HIV-1 Δ *vif*). Again, APOBEC3G was used as a positive control. While the APOBEC3C I188 restricted 5-10 fold better than APOBEC3C S188 (Figure 3.8A: 30% infectivity compared to 8 % infectivity), strikingly, APOBEC3C S188-S188 dimer restricted infection as efficiently as APOBEC3G (Figure 3.7B: approximately 1% infectivity for both conditions). Importantly, the APOBEC3C S188-S188 synthetic dimer restricts infection far greater than two-fold more than the APOBEC3C S188 monomer (Figure 3.8A: 30% infectivity relative to 1% infectivity), suggesting that the increased antiviral activity is not simply the result of having twice as many active sites. Thus, these results indicate that forced dimerization is sufficient to induce anti-HIV activity of APOBEC3C regardless of the isoleucine at position 188. In a separate series of experiments, we also examined the ability of each of the A3C variants to be packaged into virions. We found that A3C I188 was not packaged to a greater extent than A3C S188 (compare Fig 3.8B lanes 2 (A3C I188) to Figure 3.8B lanes 3 (A3C S188)). Thus the greater activity of A3C I188 correlates better with its increased enzymatic activity than with virion packaging. On the other hand, the synthetic dimer of A3C S188-S188 is packaged into virions 10-20 fold better than the single domain versions of A3C (Figure 3.8B). This increased packaging could additionally explain the enhanced antiviral activity of the synthetic dimer. This suggests that while natural dimers of APOBEC3C have increased enzymatic activity, a synthetic dimer of an APOBEC3 protein can be created with improved antiviral activity due to increased packaging into virions.

3.3.5 *APOBEC3C* variants are targeted by *Vif*. In the absence of direct clinical or cohort data, we next sought to further evaluate the relevance of APOBEC3C to HIV infection. Previous studies had found that APOBEC3C mRNA is well expressed in primary T cells [223]. We reasoned that if APOBEC3C is indeed a restriction factor relevant to HIV, then one would expect it to be antagonized by the viral *Vif* protein. To test this, we produced HIV-1 (either lacking *vif*, or expressing either HIV-1 or HIV-2 *vif*) in the presence of APOBEC3C. When we express APOBEC3C I188 during HIV production, the infectivity of the virus is reduced by more ten-fold. However, in the presence of APOBEC3 S188 or I188, both HIV-1 *Vif* and HIV-2 *Vif*, restored viral infectivity (Figure 3.9A). We also conducted western blots analysis to probe for APOBEC3C (S188 and I188) expression in the presence of HIV-1 and HIV-2 *Vif* proteins

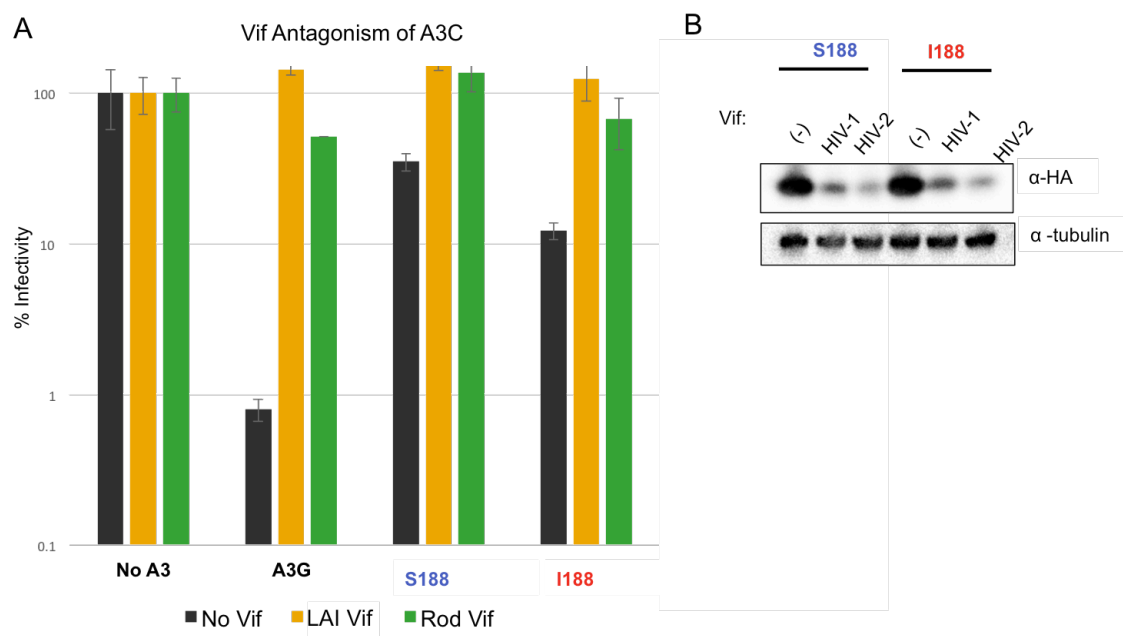


Figure 3.9. APOBEC3C is targeted by Vif. (A) Restriction of HIV-1 Δvif by APOBEC3C S188 (A3C S188), APOBEC3C I188 (A3C I188), and full recovery of infectivity by the presence of HIV-1 (LAI strain) and HIV-2 (ROD strains) *vif*. 0.4 μ g of APOBEC3 plasmid was used for each condition, and 0.6 μ g of each provirus was used. Infectivity of each virus is set to 100% for infection with No APOBEC3 (No A3) present. Error bars represent standard deviation of three independent transfections and infections. Black bars indicate no Vif, orange bars indicate HIV-1 (LAI) Vif, and green bars indicate HIV-2 (ROD) Vif (B) Vif degradation of APOBEC3C S188 and I188 was detected by Western blot analysis. 0.4 μ g of APOBEC3C and 0.6 μ g of an HIV-1 provirus (either Δvif) or containing HIV-1 *vif* or containing HIV-2 *vif* were used to transfect 293T cells and lysates were probed for APOBEC3C expression. Tubulin was used as a loading control.

(Figure 3.9B). Consistent with other reports, A3C S188 protein levels are significantly decreased in the presence of HIV-1 Vif [149, 222] as well as HIV-2 Vif [225]. Likewise, the expression of the A3C I188 variant also dramatically decreased in the presence of HIV-1 and HIV-2 Vif. Thus, APOBEC3C I188 effectively antagonized by HIV-1 and HIV-2 Vif which suggests that even the more active form of APOBEC3C in its partial dimer form can still be targeted by both human lentiviral pathogens. These results suggest that APOBEC3C is relevant to HIV infections since Vif has evolved to induce its degradation and APOBEC3C is expressed in HIV target cells.

3.3.6 *APOBEC3C I188 is an ancient human polymorphism that is not found in other hominids.* The APOBEC3C I188 variant is present at frequency of 2.4% in the 1000 Genomes Project [27]. Therefore, a relatively small proportion of humans carry a variant of APOBEC3C that is more enzymatically active against lentiviruses. To determine which allele is ancestral at position 188, we constructed a phylogeny of primate *APOBEC3C* sequences. All old world monkeys (N=15) analyzed encode an isoleucine at position 188 (Figure 3.10A). Moreover, orangutans, siamangs, and gibbons also encode isoleucine, but the serine change at amino acid 188 occurred in the lineage leading to gorillas, chimpanzees, and humans (Figure 3.9A). Thus, isoleucine at position 188 is likely the ancestral state, and changed during the evolution of hominids. There are two possible explanations for the existence of the I188 in humans: 1) a reversion back to isoleucine may have occurred in a subpopulation or 2) a polymorphism has been maintained at this site for millions of years, since humans split from their ancestor with gorillas and chimpanzees. If a serine to isoleucine reversion mutation occurred in recent human evolution, we would expect it to be present only in a limited subset of humans. The frequency of the allele in the 1000 Genomes Project data is 8.9% in populations of African descent, less than 1% frequency in the Americas, and not present in Asia and Europe [27] (Figure 3.10B). Humans are dramatically more genetically diverse in Africa than on any other continent, so the presence of the allele almost exclusively in Africa does not provide strong evidence that the isoleucine reverted only in African populations. Therefore, we sought to determine if the APOBEC3C I188 allele is distributed across divergent populations in Africa, or if it is present in only a particular subpopulation. The APOBEC3C I188 allele is present in all six African subpopulations analyzed by the 1000 Genomes project, with a frequency ranging between 5.6% and 13% (Figure 3.10B). However, many of the sub-populations included in the 1000 Genomes Project live in

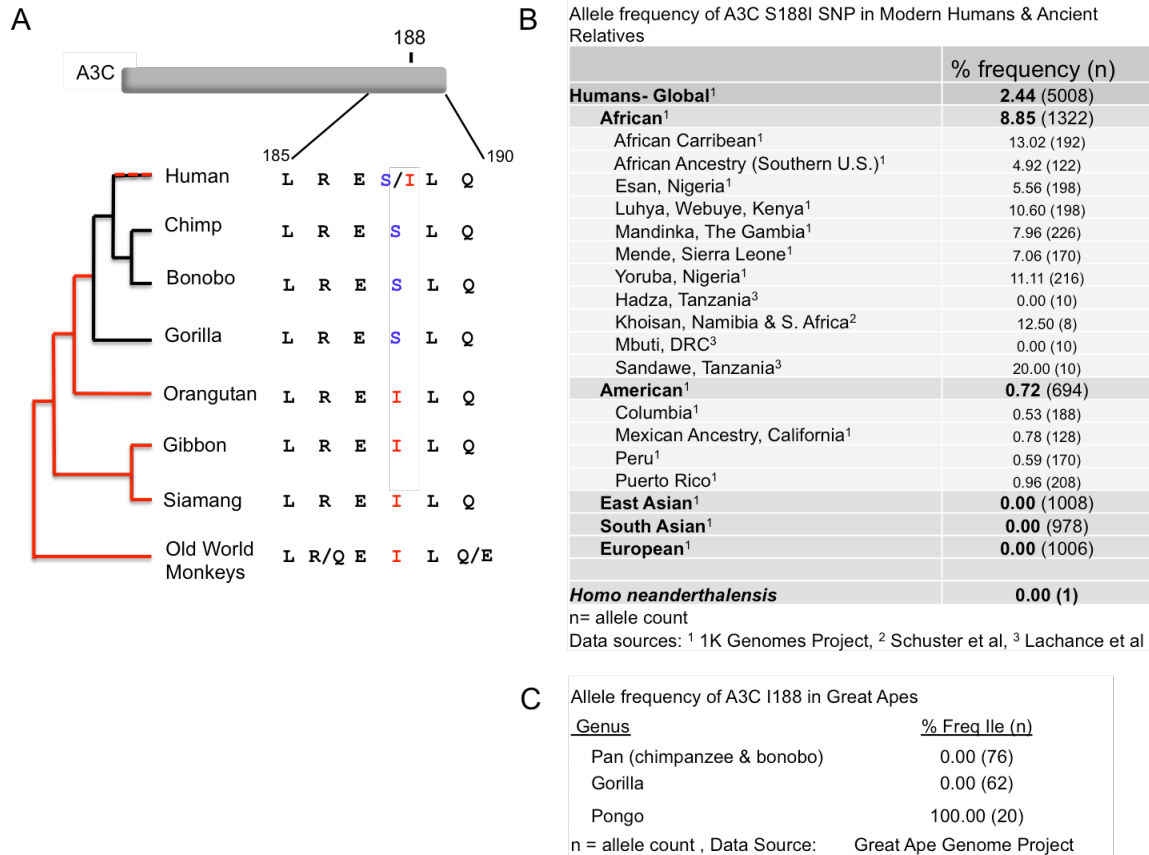


Figure 3.10. Isoleucine 188 changes to serine in some hominoids but was maintained or reverted back to isoleucine for some human populations. (A) A phylogram of catarrhines, along with an alignment of the C-terminus of APOBEC3C. All old world monkey sequences contained isoleucine at 188 (15 sequences total). (B) Allele frequency of Ile SNP in global human populations, as well as *Homo neanderthalensis*. (C) Allele frequency of Ile 188 in chimpanzees, bonobos, gorillas, and orangutans (n=79). Sequences were derived from the Great Ape Genome Project[226].

regions affected by the Bantu Expansion, a migration event when Bantu-speaking tribes swept across the continent approximately 3,000 years ago [227, 228]. To determine if the isoleucine allele is present in more diverse African genomes, we determined the *APOBEC3C* sequence from individuals from four hunter-gatherer groups (Hadza, Sandawe, Mbuti, and Khoe-San)[229, 230]. We found that one of the four Khoe-San individuals was heterozygous for the I188 allele, and two out of five Sandawe individuals were heterozygous for the I188 allele (Figure 3.10B). In conclusion, I188 seems to be a widely distributed SNP in African populations suggesting that the more active allele is very ancient, and may have even been circulating in humans since the birth of the species. Presence of the I188 in the ancient human relative *Homo neanderthalensis* would have provided evidence that the allele has been present in the Homo lineage for at least 600,000 years but we failed to find the I188 SNP in the published Neanderthal genomes. To determine if other hominoids also possess variation at position 188 we probed the *APOBEC3C* sequences from the Great Ape Genome project[226], and found that none of the great apes included in the study (n=79) were polymorphic at position 188 (Figure 3.10C). Ten orangutans were included in the study, and all encoded isoleucine at position 188. In contrast, all gorillas (n= 31), and chimpanzees and bonobos (n=38), encoded serine at position 188. Humans, gorillas, and chimpanzees diverged from their most recent common ancestor approximately 10 to 20 million years ago [231, 232], and in this ancestral lineage the more active isoleucine allele was lost. However, since some humans express the I188 allele, it is possible S188 never rose to fixation and I188 was maintained as a minor allele for a long period of the evolutionary history of hominoids. Alternatively, it is possible that serine became fixed in the ancestor to gorillas, chimpanzees and humans, but more recently the serine reverted to isoleucine in a subpopulation of humans. Nonetheless, we find that the *APOBEC3C* I188 is relatively ancient to humans, but is not present to an appreciable extent in out-of-Africa human populations, nor have we found it in other hominids.

3.4 Discussion

APOBEC3C stood out among the seven human *APOBEC3* paralogs as it has no known function. We observed that the six *APOBEC3*s with known functions possess a conserved isoleucine at the residue homologous to *APOBEC3C* position 188, whereas *APOBEC3C* encodes a serine at this position. However, human *APOBEC3C* is in fact polymorphic at this site, and some humans encode an isoleucine, the residue that correlates with *APOBEC3* antiviral/anti-

retroelement function. This led us to hypothesize that APOBEC3C may have an as yet overlooked role as a restriction factor, and that the I188 variant may have enhanced antiviral activity compared to the more common variant, S188. *APOBEC3C* has evolved under positive selection in primates and within the interface of binding by the viral protein Vif, suggesting that this gene may have played a role in restriction of lentiviruses over primate evolution. Furthermore, we found that *APOBEC3C* I188 encodes a protein with increased antiviral activity, increased enzymatic activity, and the ability to dimerize in solution. Consistent with this conclusion, an artificial forced dimer of APOBEC3C S188 has vastly increased antiviral activity. We find that the isoleucine at position 188 was lost during hominid evolution but was either reacquired by some humans since humans split with our most recent common ancestor with chimpanzees, or alternatively, has never been lost as an allele and has been maintained as a polymorphism through several million years of hominoid evolution.

Positive selection of APOBEC3C in primates suggests an ancient role in antiviral defense

Previous studies have shown that APOBEC3C binds to HIV-1 Vif and that E106 is important for Vif binding since mutation to lysine at position 106 completely abrogated HIV-1 Vif binding to APOBEC3C [149, 222]. We found that this residue within the Vif binding interface is evolving under positive selection, and another residue in the Vif-binding region, 77, is also under positive selection. Residue 77 is within the α -2 helix of APOBEC3C, which has also been shown to be important for HIV-1 Vif binding [149]. Additionally, it is possible that Vifs from other lentiviruses target APOBEC3C at different motifs, driving the positive selection in other regions of the protein. For example, APOBEC3C is under positive selection at residues 128 and 130. While these residues are not in the known APOBEC3C:HIV-1 Vif binding interface, the homologous residues of APOBEC3G are involved in HIV-1 Vif binding[148, 233]. Therefore, it is possible that other Vif proteins from other lentiviruses target APOBEC3C at positions 128 and 130, or that ancient lentiviruses have targeted these residues in the past. In summary, rapid evolution of APOBEC3C at the known APOBEC3C:Vif binding interface suggests that APOBEC3C has evolved to block lentiviruses in primates.

Mechanism of increased activity of APOBEC3C I188 relative to APOBEC3C S188

Our results indicate that the difference in the anti-HIV activity of the APOBEC3C variants S188 and I188 lies in the enzymatic efficiency of the two APOBEC3C proteins. We found that I188 more rapidly deaminates ssDNA *in vitro*. Furthermore, in an *in vitro* RT model system, the

presence of APOBEC3C cause a higher mutation frequency than APOBEC3C S188. A previous study correlated multimerization of APOBEC3s with the capacity to restrict lentiviruses [51], and our finding that the monomeric variant (S188) was less antivirally active than the dimer-forming, more active variant (I188), is consistent with this conclusion. Therefore, our model is that isoleucine at position 188 of APOBEC3C enhances lentiviral restriction by improving dimerization and in turn, the enzymatic activity of the protein. One possible reason dimerization is important for APOBEC3C activity, is that it could improve the protein's ability to scan DNA substrates for cytidine deamination motifs. In fact, I188 lies within α -helix 6, which has implicated as important for DNA scanning of another APOBEC3, APOBEC3G [108]. Ongoing work suggests that the APOBEC3C I188 is more processive (able to consecutively deaminate cytidine motifs without dissociating from substrate), than APOBEC3C S188 (Adolph, et al, manuscript in preparation). Taken together, these results bolster the model that the APOBEC3C I188 protein has greater antiviral activity than the more common APOBEC3C protein due to better enzymatic activity that correlates with increased dimerization. This model that dimerization is a key determinant of APOBEC3C activity is further supported by the fact that a synthetic dimer formed by linking two tandem S188 APOBEC3Cs drastically enhances antiviral activity. In fact, activity is improved even in comparison to I188, the more active variant. I188 only partially dimerizes, and compared to S188 and S188-S188, has an intermediate ability to restrict HIV. Interestingly, the mechanism of increased antiviral activity of A3C S188-S188 is likely due to its increased ability to be packaged into virions. These results suggest that artificial forms of human APOBEC3C proteins can have enhanced antiviral properties that may have therapeutic uses in controlling viral infection.

Population genetics of human APOBEC3C suggests that the I188 polymorphism is ancient.

The isoleucine at position 188 of APOBEC3C is present at approximately 10% frequency across diverse African populations, but almost absent from all other global populations. All human populations outside of Africa are thought to have descended from one or a few migration events out of Africa[234]. As such, humans from non-African populations may lack the APOBEC3C I188 allele because it was excluded in a population bottleneck during the migrations. Or, the allele may have been lost in non-African populations due to drift or a lack of selective pressure, and this possibility is consistent with the fact that no primate lentiviruses are known to originate outside of Africa. Alternatively, it is possible that loss of the allele was

selected for non-African populations. Expression of another APOBEC3, APOBEC3B, has been associated with increased risk of cancer [17, 175]. Therefore, the antiviral function of APOBEC3s may come at an evolutionary trade-off. In fact, this may have driven the maintenance of the less enzymatically active S188 allele for millions of years in humans and ancient human ancestors. Our phylogenetic analysis shows that APOBEC3C I188 is ancestral in primates, but changed to serine in the clade of apes including gorillas, chimpanzee, and humans. The fact that humans have a polymorphism that corresponds with the ancestral residue could be due to a reversion back to the amino acid present in other primates, but not in gorillas nor chimpanzees. If a reversion occurred it must have happened long ago in human history, since the allele is present in such deeply divergent populations across Africa. However, the allele was likely lost due to a bottleneck in the out-of-Africa populations because it is almost completely missing from non-African populations. Alternatively, it is possible that the isoleucine allele has continued in the human lineage through incomplete lineage sorting (the maintenance of a polymorphism after the divergence of species), since before humans split with their most recent common ancestor with gorillas more than 10 million years ago. Notably, the isoleucine codon, ATT, at position 188 is the same in the human SNP as in all other primates with an Ile at this position in APOBEC3C. While we did not find support for incomplete lineage sorting since we did not find any other hominids that were polymorphic at position 188, the limited number of great ape sequences were included does not allow us to completely rule out this second possibility. Nonetheless, given the increased antiviral activity of APOBEC3C I188 and its fixation in primates other than hominids argues that the gain (or maintenance) of this allele in humans has been driven by a function for protection against pathogens.

Potential impact on human health

We discovered that an APOBEC3C single nucleotide polymorphism (SNP) that is common in Africa enhances anti-lentiviral activity. This polymorphism may impact human susceptibility to cross-species transmissions of lentiviruses because Vifs from other lentiviruses may not antagonize human APOBEC3C. HIV-1 and HIV-2 Vif are able to antagonize both variants of APOBEC3C so the I188 SNP may not block HIV transmission, so Vif may effectively counteract I188 activity during infection. However, the fact that APOBEC3C is antagonized by Vif does suggest that APOBEC3C is an important barrier that must be countered by the virus during natural infections. Alternatively, it is possible that APOBEC3C antagonism by Vif is an

unintended consequence due to Vif binding to another APOBEC3, APOBEC3F, as APOBEC3C has a Vif binding pocket that is nearly identical to the Vif binding pocket of APOBEC3F [149, 151, 222]. Despite the ability of Vif to antagonize APOBEC3C, it is possible that APOBEC3C I188 influences HIV susceptibility. In infected individuals possessing the whole APOBEC3 repertoire, Vif has to adapt to counteract multiple antiviral proteins and this may constrain Vif and weaken its activity. In fact, viral genomes sequenced from HIV-1-infected patient cells are extensively mutated by APOBEC3s despite the presence of Vif [235, 236] and the extent of APOBEC3-induced mutagenesis negatively correlates with disease progression rate [237]. As such, it is possible that APOBEC3C I188 may provide some level of protection from HIV transmission or pathogenesis.

3.5 Materials and Methods

3.5.1 APOBEC3C sequence. APOBEC3C was amplified by RT-PCR from total RNA extracted from chimpanzee, gorilla, orangutan, white-cheeked gibbon, siamang, baboon, sooty mangabey, and red-capped mangabey, and proboscis monkey cells (either fibroblast or lymphoid) obtained from Coriell Repository as well as from the vervet monkey cell line Vero, the tantalus monkey cell line CV-1, and the saeius cell line V038 provided by the Nonhuman Primate Research Resource (NPRR). Primers were designed to amplify from the 3' and 5' UTRs of APOBEC3C mRNA transcripts (5'UTR: CTAAGAGGCTGAACATGAATC'3, 3'UTR: 5'GGCTAGAGGAGACAGACCATGA'3). The APOBEC3C amplicons were cloned into pGEM vectors, and then sequenced. The S188-188 forced dimer was designed to mimic the linker between the two domains of the double-domain APOBEC3F. The N-terminal subunit consists of APOBEC3C residues 1–189 (residue 190 is removed), followed by the residues RNP, which serve as a linker. The C-terminal APOBEC3C begins at the second start codon, M12. The dimer S188-S188 APOBEC3C was constructed by overlap extension PCR. Two separate PCRs were performed for the N terminal and C terminal APOBEC3C subunits (1st domain, For: TTCAGGATCCATGAATCCAGAGATC, 1st domain, Rev: GCCTCCATTGGGTCCCGGAGACTCTCCCGTAGCCTTCTTT, 2nd domain, For: TCCAGGATCCATGAATCCACAGATC, 2nd Rev: GCCCTCTAGATTAGGCGTAGTCAGG), and these amplicons were annealed in a third PCR reaction using the 1st domain For and the 2nd domain Rev primers.

3.5.2 Sequence analysis. APOBEC3C genes were aligned using Geneious software. To test for positive selection, maximum likelihood tests were performed using the PAML statistical software suite [238]. The APOBEC3C genes were subjected to tests that allowed for positive selection (M8 model), or disallowed positive selection (M8a model). The analyses were performed with the F3X4 codon model, and multiple starting omega values were used, ranging between 0.5 and 1.4. Specific residues with signatures of positive selection with a posterior probability of 99% or greater were identified by Bayes Empirical Bayes analysis. Ancestral APOBEC3C sequences were reconstructed by the likelihood/Empirical Bayes approach using the codeml program in PAML. Branch analysis to identify particular primate branches with signatures of positive selection in APOBEC3C were performed in two ways. Overall dN/dS values were calculated with PAML, using the free ratio model. Additionally, a branch-site test to identify statistically significant signatures of episodic selection was performed using the Branch-site REL method in the HyPhy software suite [239] .

3.5.3 APOBEC3, provirus, and LINE-1 plasmids. APOBEC3Cs were cloned into the BamHI and XhoI sites of pCDNA3.1 by PCR addition of restriction sites (BamHI and XhoI) to the N and C termini of APOBEC3C. The human APOBEC3C plasmid we previously obtained from the AIDS Repository contained the SNP rs11551111, which is not common (no reported frequency according to dbSNP). Therefore, we used site-directed mutagenesis to change the asparagine at position 31 to aspartic acid (For: GCCAACGATCGGGACGAAACTTGGC, Rev: GCCAAGTTTCGTCCCGATCGTTGGC). A hemagglutinin tag was inserted into the XhoI and XbaI sites of pCDNA3.1, at the C-terminus of each APOBEC3C sequence. APOBEC3G and APOBEC3A were also in a pCDNA3.1 backbone, with a Kozak sequence, as well as a hemagglutinin tag at the N-terminus. HIV Δ env, Δ vif, HIV Δ vif + HIV-1 vif, HIV Δ vif +HIV-2 vif have been described elsewhere [153]. SIVagm Δ env, Δ vif was kindly provided by Nathaniel Landau.

3.5.4 Infectivity assays. Single round HIV-1 and SIVagm infectivity assays were performed as previously described [240]. 293T cells (American Type Culture Collection) were plated at a density of 5×10^3 cells per well of a 24-well plate. The next day, the cells were transfected with 0.3 μ g provirus encoding luciferase as a marker gene 0.1 μ g pL-VSV-G, and 0.3 μ g pCDNA3.1.APOBEC3.HA or empty pCDNA3.1 plasmid. For the dose response infectivity assay, either 0.1 μ g, 0.2 μ g, or 0.3 μ g APOBEC3 plasmid was used. For experiments

involving Vif expression, 0.2µg of APOBEC3 was used. Forty-eight hours after transfection, virions were harvested. For SIVagm infectivity assays, SupT1 cells were infected with 10µl of each virus and treated with 20µg/ml DEAE/dextran. For HIV infectivity assays, ELISA was performed to quantify p24, and virus equivalent to 2ng p24 was used for infections. For all infectivity assays, 5×10^4 were infected in a 96 well dish. Seventy-two hours later, infected cells were lysed in luciferase lysis reagent (Brightglo, Promega) and luciferase expression was measured on a luminometer (LUMISTAR Omega, BMG). Infectivity of each virus was compared by setting infectivity of the “No APOBEC3” control to 100%. All HIV-1 constructs are based on the LAI strain.

3.5.5 *LINE-1 assays.* To assay for restriction of LINE-1 retrotransposition 293T cells were transfected with 200ng LINE-1 plasmids pYX016 and pYX015[241], along with 100ng of APOBEC3C S188 or I188, APOBEC3C, 10ng APOBEC3A, or empty pCDNA3.1 plasmid. The next day, the cells were treated with 2.5 µg/ul puromycin to select for transformants. Three days later, expression of renilla and firefly luciferase were assayed using a luminometer. The LINE-1 plasmids encode firefly luciferase disrupted by a splice site, so expression only occurs after retrotransposition, whereas renilla luciferase expression is not dependent upon retrotransposition. Percent retrotransposition is reported by setting retrotransposition (firefly luciferase values divided by renilla luciferase values) in the absence of APOBEC3 to 100%.

3.5.6 *Intracellular protein packaging and expression.* Intracellular expression of the APOBEC3 proteins during virion production was evaluated by lysis of the virion-producing 293T cells with Radio Immunoprecipitation Assay buffer (RIPA), with protease inhibitor (50mM Tris, 150mM sodium chloride, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40, protease inhibitor cocktail cOmplete by Roche). Lysates were resolved on an SDS-PAGE gel in MES buffer, and transferred to a PVDF membrane for Western blot analysis, using anti-HA (BioLegend) antibody and anti-tubulin (Sigma-Aldrich) antibody. Endogenous levels of APOBEC3C were measured by Western blotting with antibody purchased from Fisher (product # PA5- 27629). HRP-conjugated secondary antibodies (Santa Cruz) were used to detect primary antibodies.

Packaging of APOBEC3 into virions was evaluated by co-transfection of 100 ng of each APOBEC3 expression plasmid with 500 ng of an HIV proviral clone (LAI) containing a deletion in vif in each well of a 12-well plate. Three days after transfection, 1 ml of supernatant was collected, filtered through a 0.2 micron filter, and concentrated by pelleting in a microcentrifuge

at 13K rpm for 60 minutes and resuspended in 80 μ l. The amount of p24gag was determined by ELISA (Advanced Bioscience Laboratories). Equal quantities of p24gag were lysed and run on an SDS-PAGE gel. The Western blots were probed with an anti-HA antibody for A3C protein and with a p24gag antibody for virus production and HRP-conjugated secondary antibodies were used to detect primary antibodies. Cells were lysed as described above. The chemiluminescent signals from the Western blots were imaged using a ChemiDoc MP Imaging System (Bio-Rad) and quantified in the linear detection range.

3.5.7 Recombinant protein expression and purification. Recombinant baculovirus production for APOBEC3C S188 was carried out in the pACG2T transfer vector (BD Biosciences), as described previously [111]. Recombinant baculovirus production for APOBEC3C I188 was carried out in the pFastbac1-GST-APOBEC3C vector according to the Bac-to-Bac expression system (Life Technologies) and as described previously [110]. Recombinant virus was then used to infect Sf9 cells. Cells were harvested 72 hours after infection, lysed, treated with RNaseA, and clarified cell lysates were incubated with glutathione-sepharose 4B resin (GE Healthcare) at 4°C and subjected to a series of salt washes, as described previously[102]. The APOBEC3C S188, APOBEC3C I188 enzymes were eluted from the glutathione-sepharose resin (GE Healthcare) with the GST tag, as previously described [102]. The samples were then treated with thrombin (GE Healthcare) for 6 hr at 21°C to cleave the GST tag.

3.5.8 Size exclusion chromatography. The oligomerization states of the APOBEC3C enzymes were determined by loading 10 μ g of purified enzyme on a 10 mL Superdex 200 (GE Healthcare) size exclusion column. The column was prepared by pouring the resin bed in a column with 16-cm height and 0.5-cm diameter. The running buffer contained 50 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT. The Bio-Rad standard set was used to generate a standard curve from which molecular masses and oligomerization states of the enzymes were determined.

3.5.9 Protein crosslinking by BS3. A3C S188 and A3C I188 (0.5 μ M) were incubated with 10 μ M BS3 in 20 mM Hepes (pH 7.5), 150 mM NaCl and 1 mM DTT for 1 hour at 21°C. Crosslinked proteins were resolved on a 12% SDS-PAGE gel, transferred to a nitrocellulose membrane for Western Blot analysis and visualized using primary antibody for native APOBEC3C (GeneTex) and secondary IRdye labeled goat anti-rabbit antibody compatible with the LI-COR/Odyssey system.

3.5.10 In vitro deamination assay. All ssDNA substrates were obtained from Tri-Link Biotechnologies as previously published [108]. Reactions were carried out under single-hit conditions (i.e. <15% substrate usage) to ensure that a single enzyme carried out the deaminations on the ssDNA[242]. A ssDNA substrate containing two 5'-TTC motifs (100 nM) was incubated with 350 nM of APOBEC3C I188 or 700 nM of APOBEC3C S188 for 5 to 30 min at 37 °C in RT buffer (50 mM Tris, pH 7.5, 40 mM KCl, 10 mM MgCl₂, and 1 mM DTT). The reaction time was varied on each ssDNA according to the specific activity of the enzymes to ensure <15% substrate usage. Reactions were started by the addition of the ssDNA substrate. APOBEC3C-catalyzed deaminations were detected by treating the ssDNA with uracil DNA glycosylase (New England Biolabs) and heating under alkaline conditions before resolving the fluorescein-labeled ssDNA on 10 or 20% (v/v) denaturing polyacrylamide gels, depending on the sizes of the ssDNA fragments. Gel photos were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare), and integrated gel band intensities were analyzed using ImageQuant (GE Healthcare). The specific activity was calculated from single-hit condition reactions by determining the picomoles of substrate used per minute for a microgram of enzyme.

3.5.11 In vitro reverse transcription assay. Mutagenesis of ssDNA by A3 enzymes during reverse transcription of an RNA template was assessed using an in vitro assay, which models reverse transcription of an RNA template and second-strand synthesis. The method is described in detail in Feng and Chelico 2011 [107]. This system uses an in vitro synthesized RNA, which contains a polypurine tract (PPT), a protease gene (prot) of HIV, and a lacZ α reporter for blue/white screening. The RNA is reverse transcribed to (-)DNA by reverse transcriptase (RT) by annealing a DNA primer and after the RNaseH domain of RT removes the RNA, the PPT enables second-strand (+)DNA synthesis by acting as a primer. A 368-nt RNA template (50 nM) is annealed to a DNA primer (24-nt) and incubated with 1.5 μ M of nucleocapsid (NC), 1.2 μ M of reverse transcriptase (RT) and 500 μ M of dNTPs in RT buffer in the presence or absence of 350 nM of each APOBEC3C enzyme. The RNA template contained an HIV-1 PPT, nucleotides (nt 2282–2401) from the HIV-1 clone 93th253.3 (accession number U51189), and lacZ α . The resulting dsDNA that is synthesized from this in vitro system was PCR amplified using Pfu C_x Turbo Hotstart (Agilent Technologies) that can use uracils as a template with high fidelity. These amplicons were then cloned into a pET-Blue vector backbone that

allows for blue-white screening of the synthesized lacZ α . At least twenty-five mutated clones for each condition were tested.

3.5.12 SNP analysis. 1000 Genomes Project data was mined for the presence of SNPs at position 188 of APOBEC3C (SNP ID rs112120857). To further elucidate the frequency of the APOBEC3C I188 SNP across Africa, we analyzed the genomes reported by Schuster et al.[230] and Lachance et al.[229] for the presence of the I188 allele. To assay for the presence of SNP at position 188 in other hominoids, we mined the Great Ape Genome Project [233] (accession number SRP018689) sequences in the NCBI short read archive.

4.0 Examination of A3C orthologs for differences in enzymatic activity

In Chapter 3, it was observed that the common A3C allele, S188, had weak HIV restriction ability. We found that the reason for this observed inactivity was due to the enzyme not being able to form dimers. This is in contrast to a single-nucleotide polymorphism (SNP) at amino acid position 188, I188 that formed dimers in solution. Strikingly, the requirement for dimerization was clearly demonstrated through the use of an artificial dimer of A3C. A3C that was dimerized through an engineered peptide linker had an increased restriction capacity over the I188 SNP and a reduction in virus infectivity similar to that of A3G. It is unique to A3C that dimerization seems to be the most important factor for improving the catalytic activity of the enzyme.

We then were interested to examine whether this residue at position 188 was solely responsible for the HIV restriction activity for A3C. The human A3C shares a high level of sequence similarity with the chimpanzee and gorilla A3C orthologs and chimpanzee and gorilla A3C both possess a S188. This would suggest that the S188 chimpanzee and gorilla A3C were also not active against HIV (or SIV). We sought to determine the restriction activity of chimpanzee and gorilla A3C. We were also interested in determining whether the requirement for dimerization for increased human A3C activity would also extend to being essential for chimpanzee and gorilla A3C. The A3C oligomerization interface has not been previously determined. Altogether, the initial study on A3C activity presented in Chapter 3 has provided a framework for a further biochemical characterization of the enzyme.

5.0 Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization.

Madison B. Adolph¹, Anjuman Ara¹, Yuqing Feng¹, Cristina J. Wittkopp^{2,3,4}, Michael Emerman^{3,4}, James S. Fraser⁵, Linda Chelico¹.

¹Department of Microbiology and Immunology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

²Department of Microbiology, University of Washington, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

³Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

⁴Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

⁵Department of Bioengineering and Therapeutic Science and California Institute for Quantitative Biology, University of California, San Francisco, San Francisco, CA, USA

The information in this chapter was previously published:

Adolph, M.B., Ara, A., Feng, Y., Wittkopp, C.J., Emerman, M., Fraser, J.S., Chelico, L. (2017). Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. *Nucleic Acids Research*, 45(6): 3378-3394.

Copyright © The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

All experiments in this chapter were performed by M.B.A. Figures 4.3B and 4.11 were modeled by J.S.F.. A.A. contributed to preliminary experiments. Y.F., C.J.W., M.E. provided reagents. M.B.A. and L.C. conceived and designed the experiments and analyzed the data. M.B.A., L.C., M.E., J.S.F. wrote the article. M.B.A., A.A., Y.F., M.E., J.S.F. revised the article.

5.1 Abstract

The seven APOBEC3 (A3) enzymes in primates restrict HIV/SIV replication to differing degrees by deaminating cytosine in viral (-)DNA, which forms promutagenic uracils that inactivate the virus. A polymorphism in human APOBEC3C (A3C) that encodes an S188I mutation increases the enzymatic activity of the protein and its ability to restrict HIV-1, and correlates with increased propensity to form dimers. However, other hominid A3C proteins only have an S188, suggesting they should be less active like the common form of human A3C. Nonetheless, here we demonstrate that chimpanzee and gorilla A3C have approximately equivalent activity to human A3C I188 and that chimpanzee and gorilla A3C form dimers at the same interface as human A3C S188I, but through different amino acids. For each of these hominid A3C enzymes, dimerization enables processivity on single-stranded DNA and results in higher levels of mutagenesis during reverse transcription in *in vitro* and in cells. For increased mutagenic activity, formation of a dimer was more important than specific amino acids and the dimer interface is unique from other A3 enzymes. We propose that dimerization is a predictor of A3C enzyme activity.

5.2 Introduction

The human APOBEC3 (A3) family of single-stranded (ss) DNA cytidine deaminases has seven members that act as host restriction factors against retroelements, retroviruses, and other DNA viruses that contain ssDNA intermediates [2]. For A3 enzymes to restrict HIV-1 in CD4+ T cells, they must first be encapsidated into the budding virion in order to facilitate cytosine to uracil deaminations on the (-)DNA synthesized by reverse transcriptase [243]. When the (-) DNA is copied to form the (+) DNA the uracils template the addition of adenine, which results in C/G→T/A transition mutations that reduce the infectivity of HIV-1 [67-69]. The A3 enzymes A3D, A3F, A3G, and A3H (haplotypes II, V, and VII) are able to restrict HIV-1 infection in this manner to varying degrees [58, 59, 65, 66, 126, 244, 245]. However, lentiviruses such as HIV-1 have evolved a protein, Vif, which antagonizes the A3 proteins by inducing their degradation [52, 66, 114, 126, 246-251]. Vif physically interacts with A3 enzymes and functions as a substrate receptor for a Cullin 5 E3 ubiquitin ligase complex inducing the polyubiquitination of the A3 proteins followed by degradation in the proteasome [52]. Vif is stabilized in host cells by interacting with the transcriptional cofactor CBF β and Elongin C [54, 55, 252].

For A3 enzymes that fortuitously escape Vif mediated degradation or in the absence of Vif, the encapsidated A3s must be able to induce enough mutations to inactivate the proviral DNA. To achieve high levels of mutations, the enzymes must efficiently find cytosines for deamination within their preferred target motif within a limited amount of time [2]. HIV-1 replication is a dynamic process with (-)DNA synthesis, RNA degradation, and (+)DNA synthesis occurring at the same time. A3 enzymes A3F, A3G, and A3H have been characterized to locate these targets on ssDNA by facilitated diffusion [2]. Facilitated diffusion is Brownian motion driven diffusion of enzymes on DNA that occurs in the absence of an energy source to drive the enzyme's motion [104, 105]. Facilitated diffusion can involve one-dimensional sliding of the enzyme along the DNA phosphate backbone and three-dimensional translocations that are described as jumping or intersegmental transfer [104, 105]. Jumping is used to describe the movement of the enzyme as it diffuses within the charged domain of the DNA without directly interacting with the DNA phosphate backbone [104, 105]. Intersegmental transfer involves a doubly-bound state where the enzyme leaves the charged domain of the DNA and enters into the bulk solution to bind another DNA segment before releasing the first bound DNA [104, 105]. Sliding allows for deamination of cytosines that are closely spaced (< 20 nt) whereas the jumping or intersegmental transfer movements allow for deamination of more distantly spaced targets [106, 107, 253]. The A3 enzymes that most efficiently induce mutagenesis in HIV-1 proviral DNA, such as A3G and A3H, use a combination of both one-dimensional short-range sliding and three-dimensional long-range scanning movements which enables a rapid sampling of DNA for the preferred target motif [107, 110, 111, 254]. In contrast, A3F that is limited to using only long-range movements induces less mutagenesis than A3G and A3H [108]. However, A3G, A3H, and A3F are all processive enzymes, meaning they can deaminate multiple cytosines in a single enzyme-substrate encounter, but their level of processivity differs as a result of their scanning movements and this influences their mutagenic efficiency.

The processive mechanism of A3C has not been characterized previous to this study, however, A3C has been found in the majority of studies to be weakly restrictive or not restrictive for HIV-1 replication, yet it is still highly expressed in CD4+ lymphocytes and can be encapsidated [28, 209, 218, 255-258]. Recently, a human (h) A3C polymorphism, S188I, which exists in approximately 10% of people of African descent was found to enable hA3C S188I to restrict HIV-1 replication 5- to 10-fold more than the common hA3C [27, 28]. The hA3C S188I was able

to dimerize *in vitro*, unlike the common hA3C [28, 149]. While this is usually required for encapsidation for other A3s, for hA3C both the common and hA3C S188I were able to encapsidate relatively equally [28]. These data suggested that the higher restriction levels induced by hA3C S188I in comparison to the common hA3C were due to differences in the enzymes inherent biochemical characteristics [28].

Here, we investigated the mechanism by which hA3C S188I has greater activity than the common hA3C. Our biochemical analysis demonstrates that the hA3C S188I is more processive and more mutagenic during reverse transcription than the common hA3C. Since we had previously shown that the hA3C S188I mutation correlates with dimerization of the protein [28], we also analyzed the closely related chimpanzee (cA3C) and gorilla A3C (gA3C) enzymes since these encode an S188, like to the common hA3C version. Surprisingly, we found that cA3C and gA3C were efficient at inducing mutagenesis during reverse transcription despite the presence of an S188, suggesting another determinant for activity. We found that cA3C and gA3C have achieved this increased activity through dimerization that is mediated by a unique amino acid. These observations are consolidated by the finding that dimerization is required for A3C processivity and although the hominid A3Cs share a common dimer interface, they use different amino acids to form the dimer. Identifying these determinants enables prediction of A3C enzymatic activity.

5.3 Results

5.3.1 Human A3C S188I is a processive enzyme. We were interested in understanding why the hA3C S188I polymorphism encodes a more active enzyme than the more common hA3C S188. Since the ssDNA scanning mechanisms used by an A3 enzyme contributes to the efficiency of HIV-1 restriction we first examined the mechanisms by which hA3C scanned ssDNA [2]. By characterizing the ssDNA scanning mechanisms we are able to measure enzyme processivity, the ability of the enzyme to deaminate more than one cytosine in a single enzyme-substrate encounter. An *in vitro* deamination assay was used in which A3C processivity was tested on different synthetic ssDNA substrates that contained two 5'TTC deamination motifs spaced varying distances apart. Processive deaminations are measured under single-hit conditions (<15% substrate usage) to ensure that each ssDNA substrate was acted upon by only one enzyme during the course of the reaction [242]. Under these conditions a processivity factor is calculated as the ratio of the processive deaminations occurring during the experiment, i.e., deamination of

both 5'TTC motifs in a single enzyme-substrate encounter, in comparison to the calculated theoretical number of deaminations that would occur independently at both 5'TTC motifs if the enzyme were non-processive (see Materials and Methods) [111]. As a reference for the range of processivity, a highly processive deaminase, such as A3G, has processivity factors ranging from 4 to 8, depending on the ssDNA substrate (Figure 5.1) [107, 108]. An enzyme that is non-processive, such as A3A, would have a processivity factor of 1.0 or an undetectable level of deamination of both motifs under the single-hit reaction kinetics of the experiment (Figure 5.1) [119].

To measure processive sliding movements we used an ssDNA substrate with the deamination targets spaced 5 nt apart. For the common hA3C (S188, referred to as hA3C), a processive deamination band (5' & 3') was not detected under single-hit reaction kinetics (Figure 5.2A). Thus, hA3C was unable to processively slide to deaminate multiple cytosines. In contrast, on this substrate the hA3C S188I had a processivity factor of 1.4 (Figure 5.2A). This processivity factor meant that hA3C S188I was 1.4-fold more likely to catalyze processive deaminations of both the cytosine motifs in a single enzyme-ssDNA encounter than to catalyze both deaminations in separate enzyme-ssDNA encounters. To investigate if the A3Cs had differences in their ability to scan ssDNA by three-dimensional translocations, we utilized ssDNA substrates with more distantly spaced deamination targets (63 nt apart). The processivity factors indicated that hA3C S188I was 2-fold more processive than the more common form of hA3C (Figure 5.2B). Thus differences in enzyme processivity could explain why hA3C S188I is able to restrict HIV-1 replication more than hA3C [28].

5.3.2 Chimpanzee and gorilla A3C processivity is distinct from human A3C processivity. We previously showed the cA3C and gA3C encode a serine at position 188 [28], suggesting that they should be less active than hA3C S188I if the S188I mutation is the only mutation able to confer increased processivity for A3C. Surprisingly, however, on the substrate with deamination targets spaced 5 nt apart, cA3C and gA3C had processivity factors of 1.9 and 2.1, respectively (Figure 5.2C). Similarly, on ssDNA substrates with deamination targets spaced 63 nt apart, the processivity factors of cA3C and gA3C demonstrated that these enzymes were approximately 3-fold more likely to catalyze processive deaminations, but hA3C had a processivity factor of 1.4 (Figure 5.2D). To extend these results we tested the processivity of

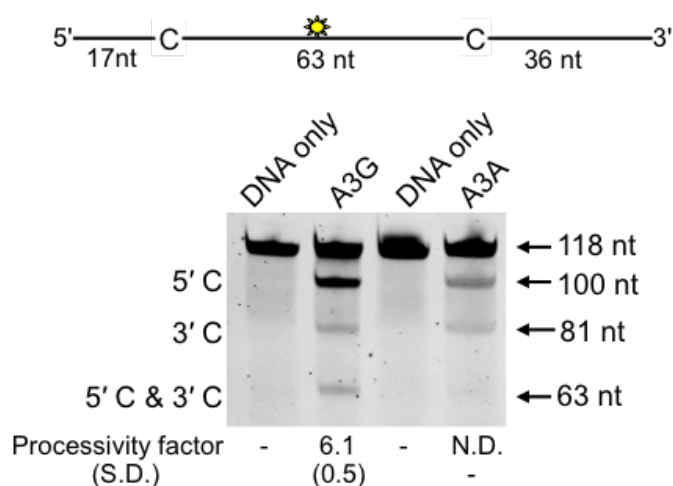


Figure 5.1. Processivity of A3G and A3A. Processivity of A3 enzymes was tested on an ssDNA substrate that contained fluorescein-labeled deoxythymidine (yellow star) between two 5'CCC (A3G) or 5'TTC (A3A) deamination motifs. Deamination of a 118 nt ssDNA substrate with deamination targets spaced 63 nt apart. Single deaminations of the 5' C & 3' C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. A measurement using the integrated gel band intensities determines a processivity factor that indicates the likelihood of the enzyme to deaminate both motifs in a single enzyme-substrate encounter (see Materials & Methods). If no 5'C & 3'C band was detected, the processivity was denoted with N.D. (not detected) and means that the enzyme is not processive. The measurements of enzyme processivity (processivity factor) and the S.D. are shown below the gels. All values are calculated from at least three independent experiments.

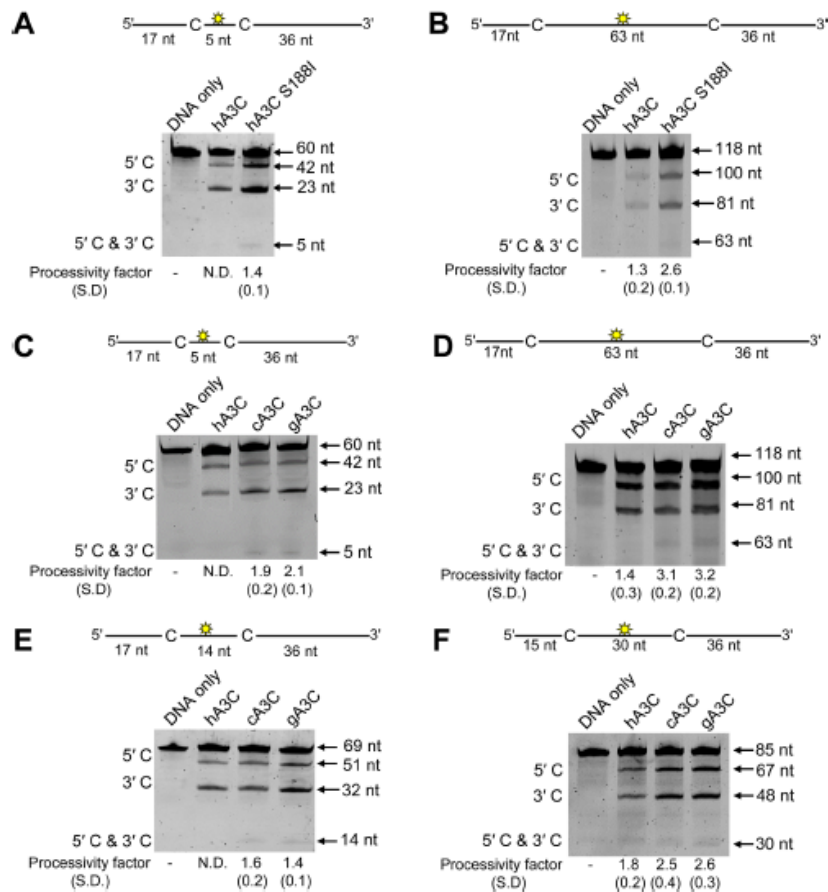


Figure 5.2. Analysis of A3C processivity on ssDNA oligonucleotides. Processivity of A3C was tested on ssDNA substrates that contain a fluorescein-labeled deoxythymidine (yellow star) between two 5'TTC deamination motifs separated by different distances. (A-B) hA3C S188I is more processive than hA3C. (A) Deamination of a 60 nt ssDNA substrate with deamination targets spaced 5 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment. (B) Deamination of a 118 nt ssDNA substrate with deaminated cytosines spaced 63 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. (C-F) cA3C and gA3C are more processive than hA3C. (C) Deamination of a 60 nt ssDNA substrate as for panel (A). (D) Deamination of a 118 nt ssDNA as for panel (B). (E) Deamination of a 69 nt ssDNA substrate with deamination targets spaced 14 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 14 nt labeled fragment. (F) Deamination of an 85 nt ssDNA substrate with deaminated cytosines spaced 30 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 30 nt labeled fragment. If no 5'C & 3'C band was detected, the processivity was denoted with N.D. (not detected). The measurements of enzyme processivity (processivity factor) and the S.D. are shown below the gels. All values are calculated from at least three independent experiments.

hA3C, cA3C, and gA3C on substrates with deamination targets spaced 14- and 30- nt apart. Consistently, cA3C and gA3C were more processive than hA3C (Figure 5.2E-F).

Due to the higher processivity factors of all A3Cs on the ssDNAs with distantly spaced deamination targets, these data suggest that the A3C enzymes were primarily using three-dimensional translocations in order to locate the target cytosines. However, hA3C S188I, cA3C, and gA3C were more processive than hA3C. We also determined that the specific activity of cA3C and gA3C were similar to hA3C S188I and 5- to 8- fold higher than hA3C (Table 5.1). Altogether the data suggested that for cA3C and gA3C, increases in deamination activity did not require an I188, as for hA3C.

5.3.3 *Dimerization of hominid A3C is mediated through α -helix 6 or β -strand 4.* We reasoned that by examining differences in hA3C, cA3C, and gA3C sequences we could identify amino acids that were unique to cA3C and gA3C and investigate if these amino acids enabled increased processivity and specific activity. From this analysis, only the amino acids at positions 85 (β -strand 3), 99 (α -helix 3) and 115 (β -strand 4) that were different in both cA3C and gA3C from hA3C were candidates (Figure 5.3A). To aid in making mutations we examined the existing crystal structure of hA3C (Figure 5.3B) [149]. Although hA3C is a monomer in solution, the enzyme crystallized as a dimer [149]. Since dimerization was important for increased hA3C S188I activity [28], we hypothesized that cA3C and gA3C were more active than hA3C due the ability to dimerize. We thus began by mutating amino acids closest to the predicted dimer interface. The amino acid at position 115 was closer to the predicted dimer interface in the hA3C crystal structure than amino acid 85 or 99 (Figure 5.3B, β -strand 4). As a result, we converted the cA3C and gA3C amino acid at position 115 to the hA3C amino acid at that position, making a K115N mutant (Figure 5.3A).

Using size exclusion chromatography (SEC) and a calibration curve, we determined the oligomerization states of cA3C, gA3C, and their K115N mutants. Consistent with increased processivity and specific activity, we found that cA3C and gA3C were able to form dimers, similar to hA3C S188I, but in contrast to hA3C (Figure 5.4A-C and Figure 5.5, apparent molecular weights 45 kDa (dimer) and 19 kDa (monomer)). The cA3C and gA3C SEC showed the presence of both monomer and dimer peaks indicating that the dimer and monomer forms were in equilibrium (Figure 5.4A-B). Interestingly, the hA3C crystal structure correctly predicted the key dimerization amino acid for cA3C and gA3C [149]. The cA3C K115N existed solely as a

Table 5.1. Specific activity of A3 enzymes on ssDNA.

Enzyme	Specific Activity (pmol/μg/min)
hA3C	0.020 ± 0.005
cA3C	0.100 ± 0.010
gA3C	0.150 ± 0.010
hA3C S188I	0.160 ± 0.050
hA3C N115K	0.085 ± 0.020
hA3C S188I/N115K	0.220 ± 0.020
cA3C S188I	0.210 ± 0.030
cA3C K115N	0.050 ± 0.007
gA3C S188I	0.280 ± 0.030
gA3C K115N	0.080 ± 0.002

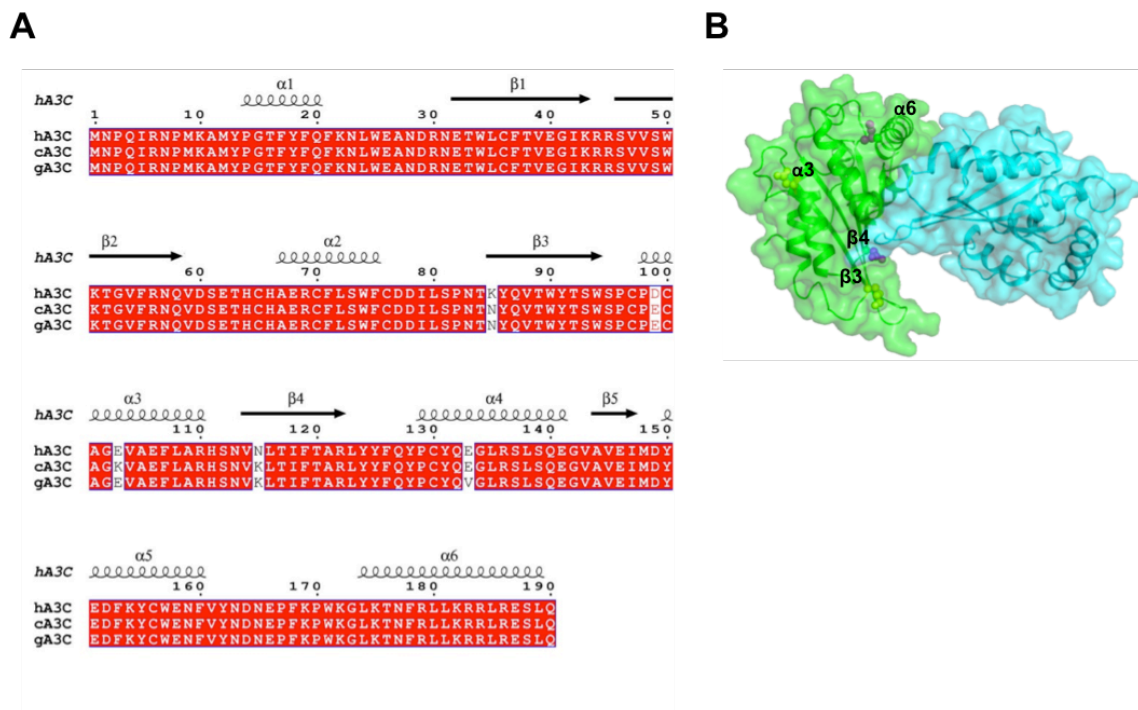


Figure 5.3. Sequence alignment and structural analysis of A3C. (A) Sequence alignment of hA3C, cA3C, and gA3C with amino acid differences shown in white. The sequence alignment was performed by a Clustal Omega multiple sequence alignment [259] and plotted using the program ESPrpt [260]. (B) Surface representation of a hA3C dimer from the crystal structure (PDB: 3VOW). Amino acids unique to hA3C that are potentially involved in the dimer interface are shown in purple (α -helix 6, S188; β -strand 4, N115) and other amino acids unique to hA3C are shown in yellow (β -strand 3, K85; α -helix 3, D99).

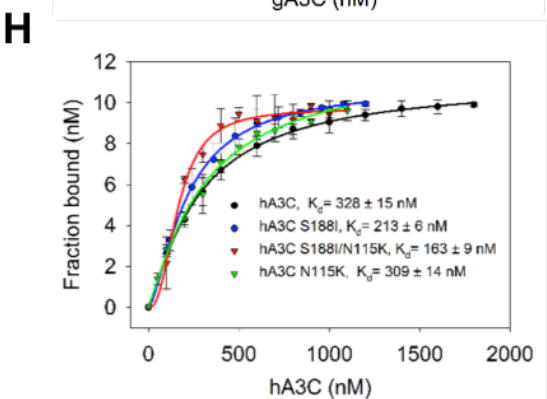
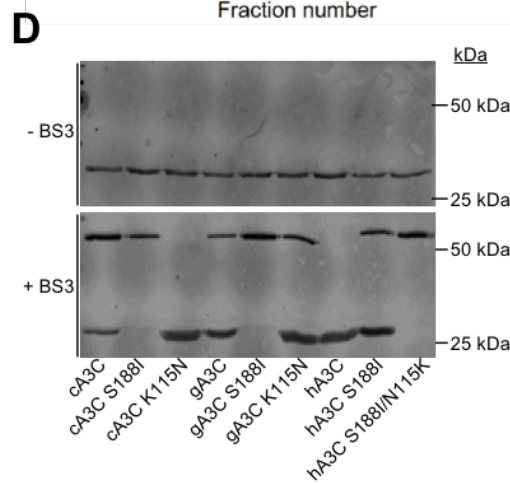
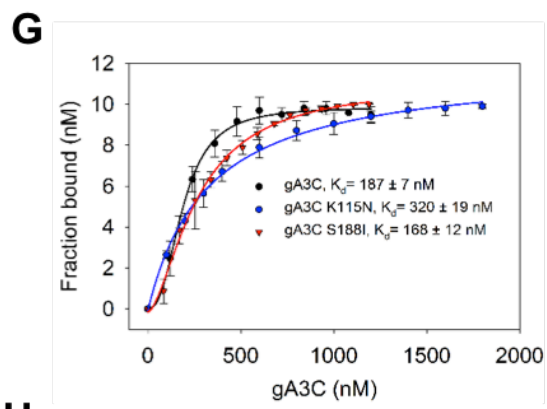
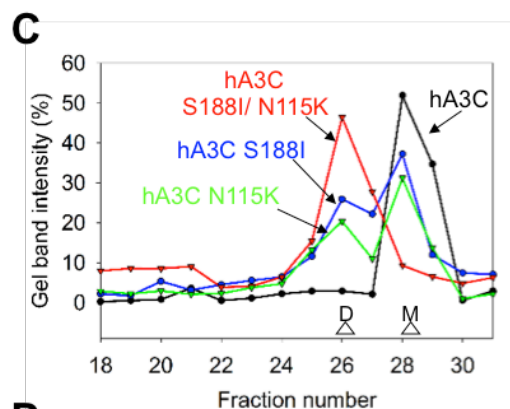
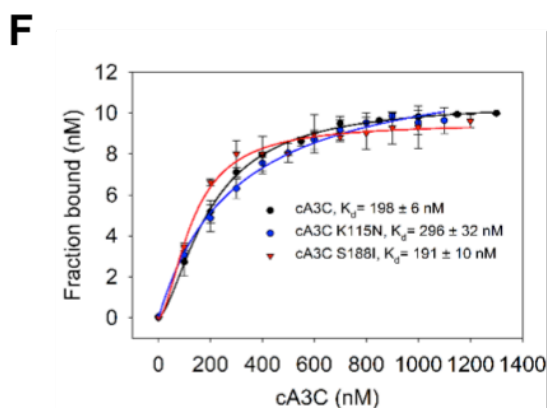
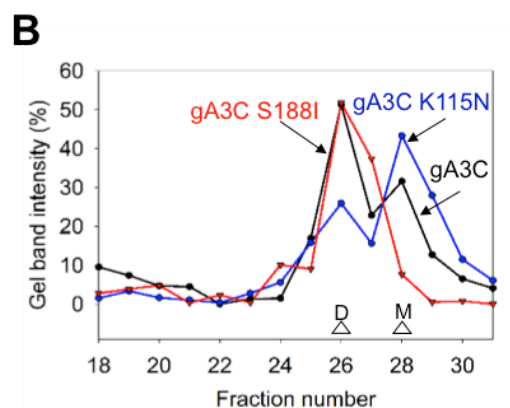
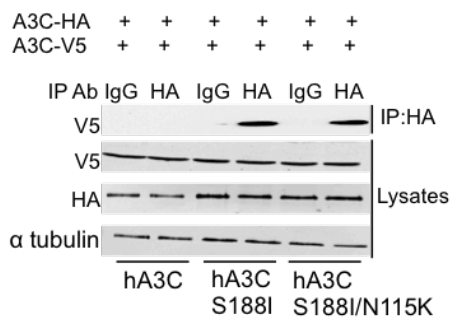
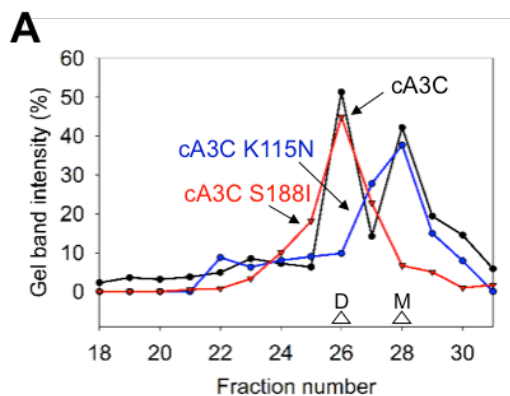


Figure 5.4. A3C dimerization is mediated through α -helix 6 or β -strand 4. (A-C) SEC profile for 10 μ g of (A) cA3C, cA3C K115N, and cA3C S188I; (B) gA3C, gA3C K115N, and gA3C S188I; and (C) hA3C, hA3C S188I, hA3C N115K, and hA3C S188I/N115K from a 10 mL Superdex 200 column was used to calculate the oligomerization state of the enzyme from a standard calibration curve. An M denotes a monomer fraction and a D denotes a dimer fraction. (A) cA3C formed monomers and dimers (apparent molecular weights 19 kDa and 45 kDa, respectively), cA3C S188I formed a stable dimer (apparent molecular weight 45 kDa), and cA3C K115N formed monomers (apparent molecular weight 19 kDa). (B) The gA3C SEC profiles were similar to cA3C, except for gA3C K115N that was mainly monomers (apparent molecular weight 19 kDa), but also retained a small proportion of dimers (apparent molecular weight 45 kDa). (C) hA3C formed monomers in solution (apparent molecular weight 19 kDa), hA3C S188I and hA3C N115K were an equilibrium of monomers and dimers (apparent molecular weights 17 kDa and 45 kDa, respectively) and hA3C S188I/N115K was a stable dimer (apparent molecular weight 45 kDa). The chromatograms were constructed by analyzing the integrated gel-band intensities of each protein in each fraction after resolution by SDS-PAGE (Supplementary Figure S3). (D) A3C enzymes were incubated in the absence or presence of 20 μ M BS3 crosslinker and subsequently visualized with SDS-PAGE and immunoblotting. Monomeric A3C enzymes remained as monomers in the presence of crosslinker (cA3C K115N, hA3C). A3C enzymes that were able to form dimers according to SEC, were also present as monomers/dimers (cA3C, gA3C, gA3C K115N, hA3C S188I) or as dimers (cA3C S188I, gA3C S188I, hA3C S188I/N115K) in the presence of the crosslinker. Molecular weight standards are indicated. (E) Coimmunoprecipitation of A3C-V5 with A3C-HA. The A3C-HA and A3C-V5 were transfected in combination and the immunoprecipitation of cell lysates used either anti-HA antibody or Rabbit IgG (mock) and was immunoblotted with antibodies against α -tubulin, HA, and V5. Cell lysates show the expression of α -tubulin, HA, and V5. (F-H) The apparent K_d of A3C enzymes from a 118 nt ssDNA was analyzed by steady-state rotational anisotropy for (E) cA3C, cA3C S188I, and cA3C K115N; (F) gA3C, gA3C S188I, and gA3C K115N; and (G) hA3C, hA3C S188I, hA3C N115K, and hA3C S188I/N115K. Apparent K_d values are shown in the figure. Hill coefficients for cooperative binding curves are (E) cA3C, 1.6; cA3C S188I, 1.7; (F) gA3C, 1.8; gA3C S188I, 2.1; (G) hA3C S188I, 1.6; hA3C N115K, 1.5; hA3C S188I/N115K, 1.9. Error bars represent the S.D. from three independent experiments.

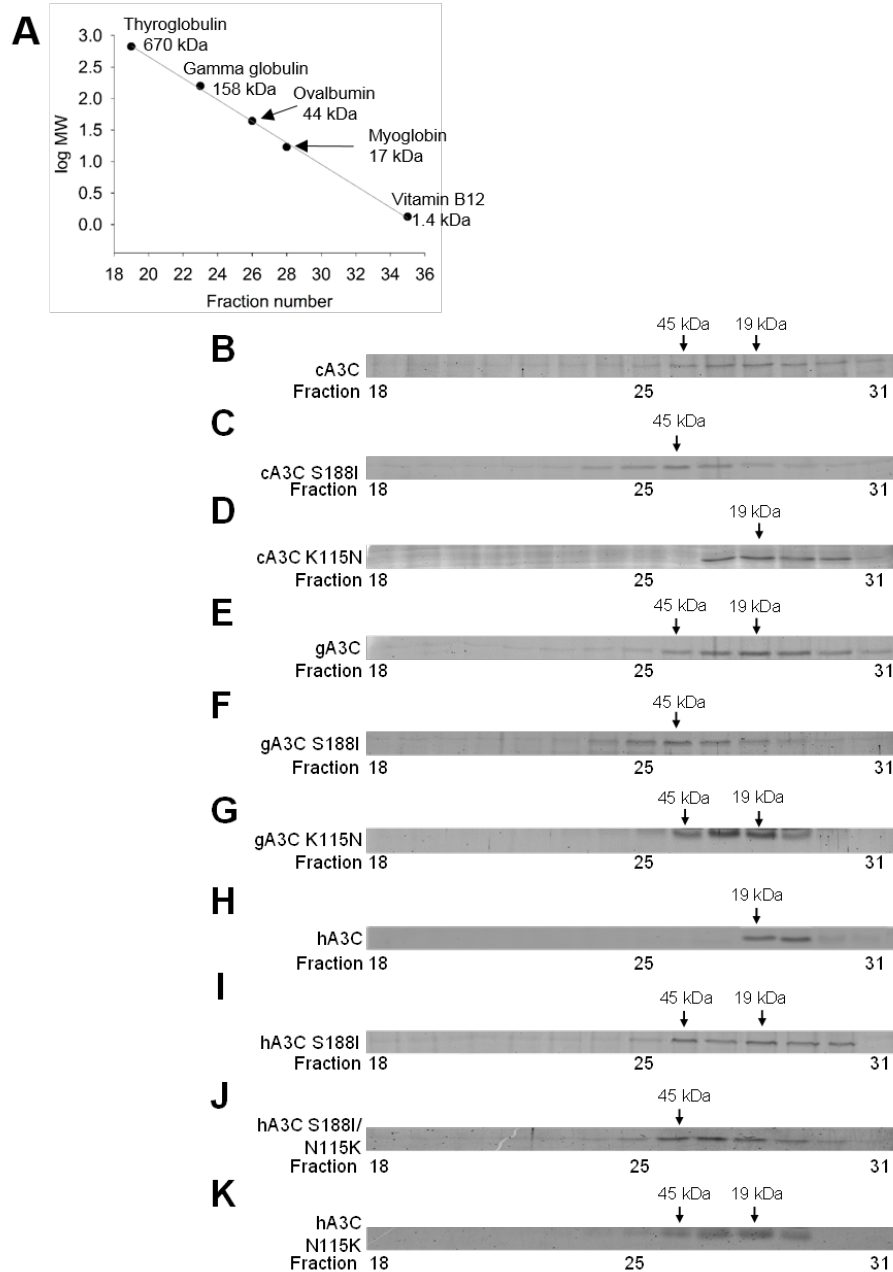


Figure 5.5. SDS-PAGE of size exclusion chromatography resolved fractions of A3C enzymes. (A) The standard curve obtained from the 10 ml Superdex 200 column from which molecular weight and oligomerization states were calculated (see Figure 3). (B-K) The chromatograms from the 10 ml Superdex 200 column were constructed by analyzing the integrated gel band intensities of the protein in each fraction after resolution by SDS-PAGE (see Figure 3). The gels for each panel were resolved, stained with Oriole stain, and scanned in parallel. The gels show the size exclusion chromatography fractions resolved by SDS-PAGE for each A3C, as labeled on each panel. The molecular weight calculated from the standard curve are shown for the peak fractions.

monomer population, similar to hA3C (Figure 5.4A, C). The gA3C K115N formed a majority of monomers, however there were still a small proportion of dimers in solution (Figure 5.4B). This suggested that an amino acid unique to gA3C further stabilized dimerization. Only V133 is unique to gA3C, suggesting that the dimer interface in gA3C may directly or indirectly involve α -helix 4 (Figure 5.3A). Further, in hA3C a K115 alone can mediate dimerization since a hA3C N115K mutant formed an equilibrium of monomers and dimers (Figure 5.4C). That the wild type cA3C and gA3C dimerization was similar to hA3C S188I and hA3C N115K (Figure 5.4A-C) was consistent with the conclusion that dimerization was enabling A3C to be more processive and have a higher specific activity (Figure 5.2 and Table 5.1).

In order to more rigorously test this conclusion using deamination assays, we first created a panel of mutants at positions 115 and 188 in each of the three hominid A3C enzymes. We introduced the hA3C S188I mutation on α -helix 6 in cA3C and gA3C. Both cA3C S188I and gA3C S188I shifted from an equilibrium of monomer and dimer populations to a stable dimer (Figure 5.4A-B). We also mutated hA3C to form the hA3C S188I/N115K double mutant. The hA3C S188I/N115K formed a stable dimer (Figure 5.4C). Analysis of these mutants suggested that the hA3C, cA3C, and gA3C dimerization interfaces were the same, but were stabilized by different amino acids.

To confirm the differences in oligomerization as determined by SEC, we used chemical crosslinking and co-IP. For crosslinking, the BS3 amine to amine chemical crosslinker enabled the resolution of A3C complexes by SDS-PAGE, which were then visualized by immunoblotting. Although the amount of protein added to each crosslinking reaction was the same (Figure 5.4D, -BS3), the total intensities of crosslinked proteins as determined by immunoblotting appeared to be unequal, which may have been due to amino acids that reacted with the BS3 preventing the antibody from binding (Figure 5.4D, +BS3). Nonetheless, the crosslinked A3C data were consistent with the SEC and identified the same monomeric (cA3C K115N, hA3C), monomeric/dimeric (cA3C, gA3C, gA3C K115N, hA3C S188I), and dimeric (cA3C S188I, gA3C S188I, hA3C S188I/N115K) forms (Figure 5.4D). We also confirmed that the dimerization was physiological by demonstrating that A3C-HA could co-IP A3C-V5. The co-IP experiment was conducted in the presence of RNaseA to ensure that we were detecting protein-protein interactions. This was tested for hA3C, hA3C S188I, and hA3C S188I/N115K. We found that

hA3C S188I and hA3C S188I/N115K, but not hA3C, could co-IP from the lysates of 293T cells, consistent with the SEC and crosslinking data (Figure 5.4E).

To determine if these differences in dimerization also occurred on ssDNA we used fluorescence depolarization to measure the rotational anisotropy of fluorescently labeled ssDNA when A3C was titrated into the solution. The binding of A3C to the fluorescently labeled ssDNA will result in a change in rotation speed (anisotropy) until the fluorescently labeled ssDNA is saturated with A3C. From these data an apparent dissociation constant (K_d) was determined. We observed that each A3C was able to bind ssDNA in the nanomolar range, with cA3C, gA3C, and hA3C S188I having less than a 2-fold higher affinity for ssDNA than hA3C (Figure 5.4F-H). This indicated that the minimal processivity observed for hA3C was not due to the enzyme having a weak interaction with the ssDNA (Figure 5.2). However, we observed that the nature of the interaction of the A3C enzymes with ssDNA was different. The cA3C, gA3C, and hA3C S188I bound ssDNA cooperatively (Figure 5.4F-H, Hill coefficients of 1.5 to 2.1), while hA3C bound ssDNA non-cooperatively (Figure 5.4H, data fit a rectangular hyperbola). This indicated that the cA3C, gA3C, and hA3C S188I monomers in solution were able to oligomerize by binding ssDNA and the dimers in solution were stabilized or oligomerized further upon binding ssDNA. This was not observed for hA3C that bound ssDNA non-cooperatively and indicated that hA3C remained a monomer even when bound to ssDNA (Figure 5.4H). With the mutants, we found that the A3C enzymes that were able to form dimers in solution bound to ssDNA cooperatively and the A3C enzymes that lacked the ability to dimerize bound ssDNA non-cooperatively with saturation curves that best fit a rectangular hyperbola by least squares regression analysis (Figure 5.4F-H). The exception was gA3C K115N that had a minor dimer population, but bound ssDNA non-cooperatively (Figure 5.4B and 5.4G). These data demonstrate that the V133 amino acid is not sufficient to promote dimerization on ssDNA and either I188 or K115 are required for this function. Altogether, the SEC and binding data indicated that the oligomerization state and not the binding affinity for ssDNA differed between hA3C and hA3C S188I, cA3C, and gA3C. Thus, the higher processivity of hA3C S188I, cA3C, and gA3C correlated with their ability to form dimers.

5.3.4 *Dimerization correlates with efficient ssDNA scanning.* Our analysis of A3C orthologs indicated that dimerization was required for processive ssDNA scanning. However, due to other amino acid differences between the A3C orthologs (Figure 5.3A) we analyzed the

monomer and dimer forms of each A3C ortholog individually to test that improved dimerization would lead to improved processivity. To test for short-range processivity due to sliding we used the substrate with deamination targets separated by 5 nt. We observed that for hA3C, dimerization through the S188I mutation enabled processive sliding and this increased further when the hA3C S188I dimer was stabilized by the N115K mutation (Figure 5.6A, short-range). For the cA3C S188I stable dimer we observed improved sliding compared to the wild type enzyme (Figure 5.6B, short-range). In contrast, the monomeric cA3C K115N was not processive (Figure 5.6B, short-range). The results with the gA3C mutants were similar to cA3C with the gA3C S188I having improved sliding from the wild type enzyme and the gA3C K115N having approximately 3-fold reduced sliding from the wild type enzyme (Figure 5.6C, short-range).

To test for long-range processivity due to jumping or intersegmental transfer we used the substrate with deamination targets separated by 63 nt. For hA3C, the S188I mutation improved the processivity of the enzyme approximately 3-fold (Figure 5.6A, long-range). The formation of a stable dimer did not further improve processivity (Figure 5.6A-C, long range). However, the cA3C K115N and gA3C K115N mutants were approximately 1.5- to 2- fold less processive than the wild type enzymes (Figure 5.6B-C, long-range). The changes in oligomerization also resulted in changes to the specific activity of the enzymes. With the introduction of the hA3C S188I or hA3C S188I/N115K mutation, the hA3C specific activity was increased 8- to 11- fold from the wild type (Table 5.1). Stabilization of the cA3C or gA3C dimer with the S188I mutation resulted in a 2-fold increase in specific activity from the wild type enzymes (Table 5.1). The K115N mutation in cA3C and gA3C resulted in an approximately 2-fold lower specific activity than the wild type enzymes (Table 5.1). Since the amino acid differences were not in the active site, the data suggested that differences in catalytic activity were due to the different efficiencies in the ssDNA scanning mechanisms employed.

5.3.5 *A monomer-dimer equilibrium is required for scanning ssDNA by intersegmental transfer.* The long-range processivity movements can be of two types, either jumping or intersegmental transfer. We determined whether the A3C orthologs and their mutants were scanning ssDNA using jumping, intersegmental transfer, or both in order to determine if the stability of the dimer correlated with either of these types of movements.

For A3C that has a single DNA binding domain per polypeptide chain, intersegmental transfer movements would require dimerization to achieve a doubly bound state. We conducted an

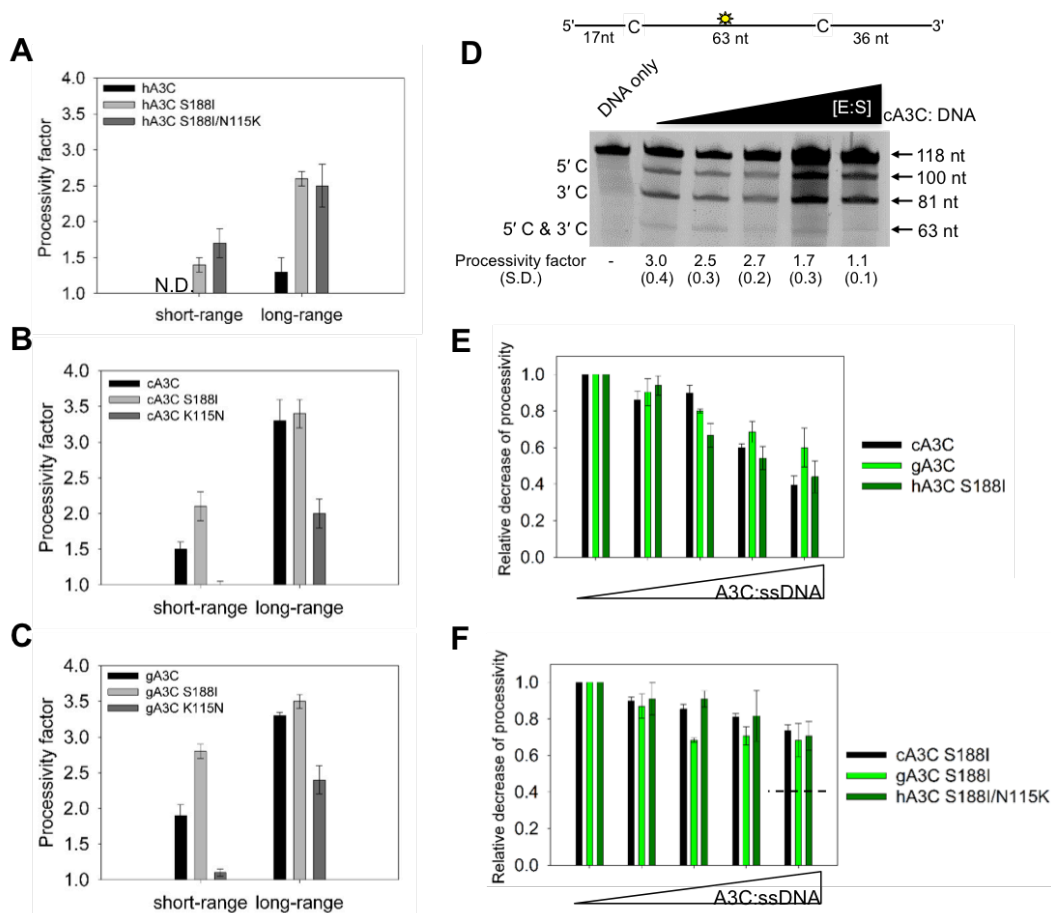


Figure 5.6. Dimerization influences processive ssDNA scanning. Processivity of A3C mutants was tested on ssDNA substrates and compared to the wild type enzyme. (A-C) Processivity factor values are shown for short-range movements based on deamination of a 60 nt ssDNA substrate with deamination targets spaced 5 nt apart and long-range movements based on deamination of a 118 nt substrate with deaminated cytosines spaced 63 nt apart for (A) hA3C, hA3C S188I, and hA3C S188I/N115K, (B) cA3C, cA3C S188I, and cA3C K115N, and (C) gA3C, gA3C S188I, and gA3C K115N. See Supplementary Figure S4 for a representative gel. (D) Intersegmental transfer ability of cA3C was determined by keeping an A3C/ssDNA ratio of 7:1 constant, but increasing the total reaction components. If the enzyme is able to undergo intersegmental transfer, the assay will result in an apparent decrease in the processivity factor with increasing concentrations of reaction components. The ssDNA substrate contained a fluorescein-labeled deoxythymidine (yellow star) between two deamination targets separated by 63 nt. The measurements of enzyme processivity (processivity factor) and the S.D. are shown below the gel. (E-F) Summary of intersegmental transfer assays shown in Supplementary Figure S5. (E) The monomer/dimer forms of A3C (cA3C, gA3C, hA3C S188I) are better able to undergo intersegmental transfer than the (F) stable dimer forms of A3C (cA3C S188I, gA3C S188I, hA3C S188I/N115K). For comparison, the hatched line in (F) denotes the decrease in processivity observed for monomer/dimer forms of A3C in (E). All values are calculated from at least three independent experiments.

intersegmental transfer assay by increasing the concentration of the A3C and ssDNA substrate while keeping the ratio of the components the same. In doing this, the reaction environment becomes crowded which makes the enzyme more likely to translocate to a different ssDNA than to translocate within the same ssDNA substrate [110, 261]. If intersegmental transfer is occurring this results in an apparent decrease in the enzyme's processivity as the enzyme cycles between substrates [110, 261]. For example, for cA3C we observed that as the concentration of the reaction components increased, the processivity decreased (Figure 5.6D, cA3C processivity of 3.0 decreased to 1.1). For cA3C the processivity decreased to 1.1, which is essentially non-processive. This meant that the majority of the three-dimensional translocations for cA3C were due to intersegmental transfer and not jumping (Figure 5.2D and F). Since jumping occurs within the localized charged domain of the DNA it is insensitive to crowding of the reaction [2, 110, 261]. To compare the A3Cs we summarized the intersegmental transfer assay data by calculating the relative decrease of processivity with increasing A3C and ssDNA concentration for the A3Cs that are monomer/dimer (Figure 5.6E) or dimer (Figure 5.6F). For the A3Cs with a monomer/dimer equilibrium, all could undergo intersegmental transfer relatively equally as demonstrated by the characteristic decrease of the processivity factor with increasing reaction components (Figure 5.6E). This suggested that for A3C, dimerization and not specific amino acid motifs was the main determinant for this ssDNA scanning mechanism. However, the stable dimer forms were not able to undergo intersegmental transfer as well (Figure 5.6F). The processivity factors did not decrease throughout the titration and plateaued (Figure 5.6F). However, the stable dimer forms of A3C did remain processive and had processivity factors that were not significantly different than the corresponding monomer/dimer A3C form at the initial A3C: ssDNA concentration used in the experiments (gA3C, compare Figure 5.7 panels A and C; hA3C, Figure 5.7 panels B and E; cA3C, compare Figure 5.6D and Figure 5.7D). These data indicate that jumping was used instead of intersegmental transfer (Figure 5.7). Thus, these forms of ssDNA scanning for A3C were interchangeable and did not depend on specific amino acid motifs, but were mediated by dimerization.

5.3.6 Ability to processively scan ssDNA correlates with mutagenesis ability. Our results showing that cA3C and gA3C were more processive in comparison to hA3C would predict that they are also able to induce a higher frequency of mutations than hA3C [107, 108]. To investigate this we used an in vitro assay to test the mutagenic potential of the A3C enzymes

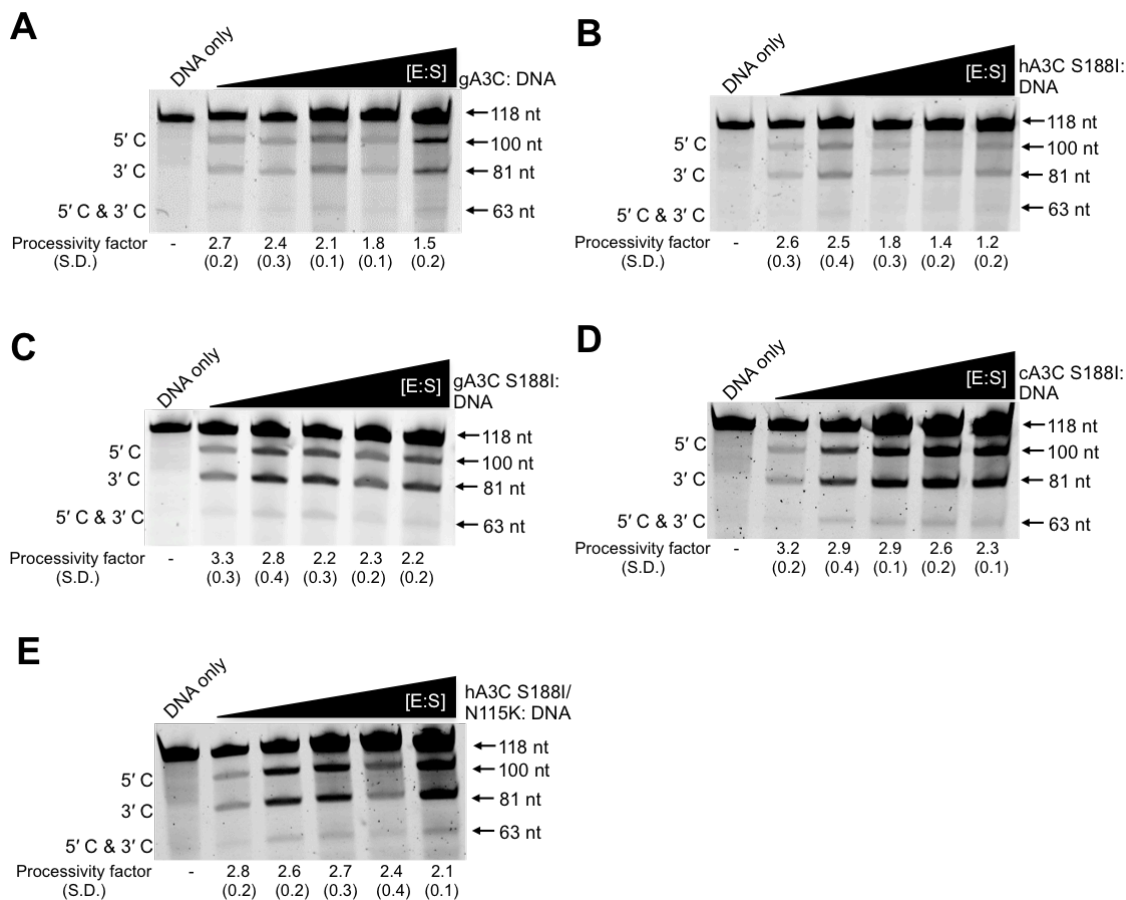


Figure 5.7. Intersegmental transfer ability of A3C is determined by the oligomeric state. The ssDNA substrate contained a fluorescein-labeled deoxythymidine (yellow star) between two 5'TTC deamination motifs. The deamination targets were separated by 63 nt. Intersegmental transfer ability of (A) gA3C (monomer/dimer), (B) hA3C S188I (monomer/dimer), (C) gA3C S188I (dimer), (D) cA3C S188I (dimer), and (E) hA3C S188I/N115K (dimer) was determined by keeping an A3C/ssDNA ratio of 3:1 constant, but increasing the total reaction components. If the enzyme is able to undergo intersegmental transfer, the assay will result in an apparent decrease in the processivity factor with increasing concentrations of reaction components. The measurements of enzyme processivity (processivity factor) and the S.D. are shown below the gels. All values are calculated from at least three independent experiments.

in which we could add the same molar amounts of each enzyme. This assay recapitulates reverse transcription of (-) DNA and synthesis of (+) DNA in vitro (see Materials and Methods) [107, 108]. The mutational spectra showed that hA3C induced at least 2.5-fold less mutations than cA3C and gA3C (Figure 5.8A-C, hA3C 0.13×10^{-2} mutations/bp; cA3C 0.33×10^{-2} mutations/bp; gA3C 0.40×10^{-2} mutations/bp). We also examined the difference in the mutational load per clone sequenced for each of the enzymes. For hA3C, all of the clones sequenced had at most one G→A mutation per clone (Figure 5.8D). Since clones were chosen for sequencing based on mutagenesis of the lacZ α region (blue/white screening), when hA3C clones contained zero G→A mutations, they were recovered due to other mutations induced by reverse transcriptase. This was in contrast to cA3C and gA3C that were able to cause a greater number of G→A mutations per clone (Figure 5.8E-F). We next tested the A3C mutants to determine if their mutagenic efficiency would correlate with processivity and dimerization. We summarized the mutation frequency of the A3C enzymes according to their oligomeric state (Figure 5.8G, monomer, monomer/dimer, dimer). In contrast to other A3C orthologs that formed dimers, the gA3C K115N could not oligomerize on ssDNA, despite forming a small amount of dimers in solution (Figure 5.4B and 5.4G). As a result, for the experiment, we considered this A3C to be a monomer. The results show that the in vitro mutation frequency increases with dimer formation. In comparison to the monomeric A3C forms, the increases were ~2-fold for monomer/dimer and 3- to 4- fold for dimer and were independent of the A3C ortholog (Figure 5.8G). The improvement in overall mutations induced was also observed on a per clone basis (Figure 5.9).

We also confirmed that these increases in mutagenic efficiency were relevant to virus restriction by conducting single-cycle infectivity assays with virus derived from a LAI Δ vif construct. In this assay, an increase in A3-induced mutagenesis results in a decrease in virus infectivity. Consistent with previous studies using HIV-1 Δ vif, hA3C does not restrict the virus, A3G highly restricts the virus, and hA3C S188I restricts the virus more than hA3C (~3-fold more) (Figure 5.8H) [28]. The single-cycle infectivity assays were in agreement with the in vitro mutagenesis assay and demonstrated that monomer/dimer and dimer A3C forms were more able decrease virus infectivity than monomeric A3C, regardless of the A3C ortholog. We confirmed that decreases in viral infectivity were the result of A3-induced mutations by PCR amplifying and sequencing a 351 nt portion of the protease gene from integrated proviral DNA (Table 5.2). It should be noted that the mutation frequencies recovered from the proviral DNA are not directly

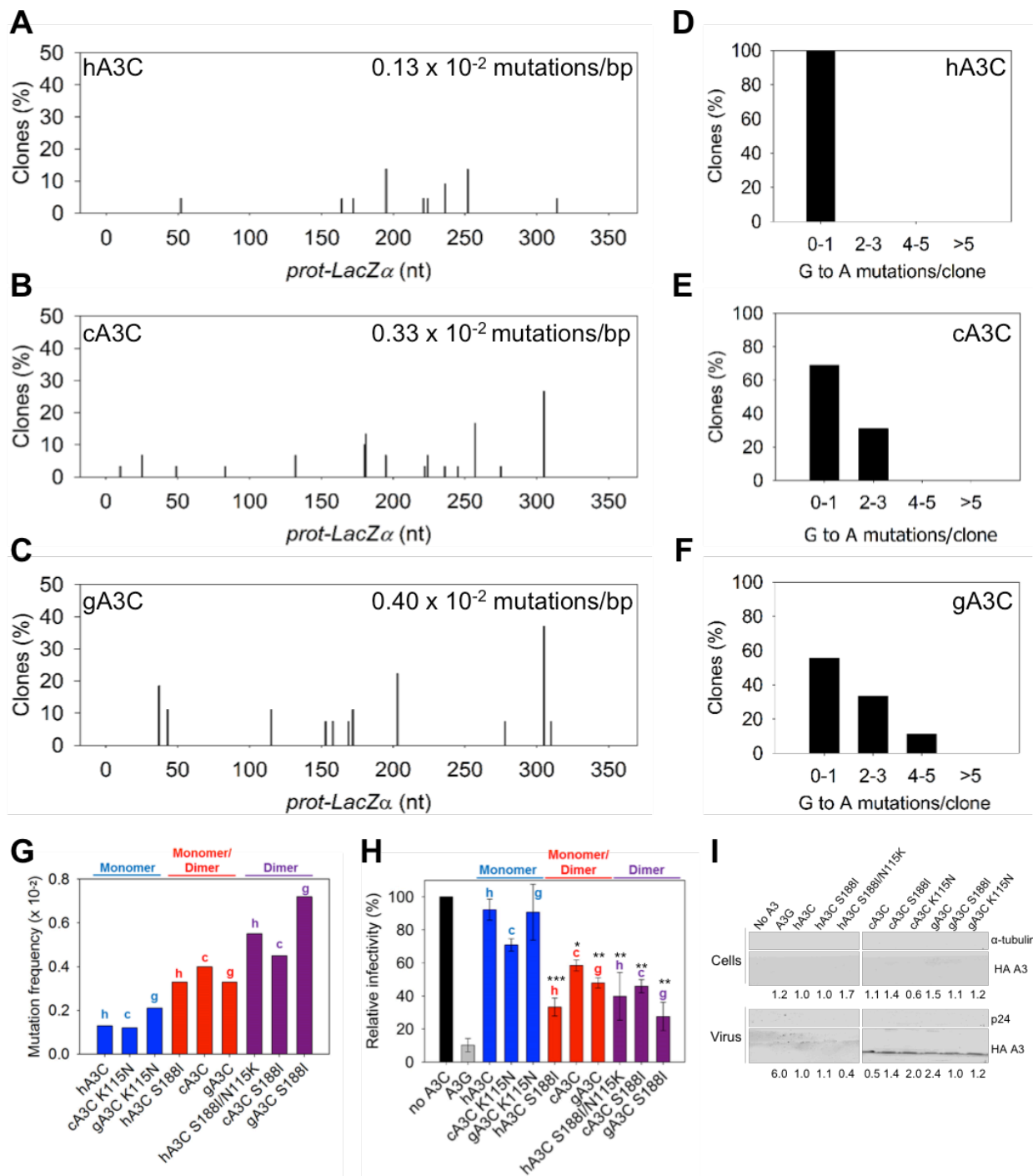


Figure 5.8. Monomeric A3C induces lower levels of mutagenesis than dimeric A3C in vitro. An *in vitro* HIV replication assay was utilized to determine the enzymes abilities to catalyze deaminations during proviral DNA synthesis. (A-C) Spectra of mutations are plotted as the percentage of clones containing a G→A mutation at a particular location (nt) in the 368 nt *prot-lacZα* construct for (A) hA3C, (B) cA3C or (C) gA3C. (D-F) Histograms illustrate the number of mutations that can be induced by (D) hA3C, (E) cA3C or (F) gA3C within individual clones. (G) Summarized G→A mutation frequency for A3C monomers (hA3C, cA3C K115N, gA3C K115N), monomers/dimers (hA3C S188I, cA3C, gA3C), and dimers (hA3C S188I/N115K, cA3C S188I, gA3C S188I). The graph denotes whether the A3C is from human (h), chimpanzee (c), or gorilla (g). Individual spectra and clonal mutation frequencies not included in Figure 5 are in Supplementary Figure S6. (H) HIV Δvif infectivity was measured by β-galactosidase expression driven by the HIV-1 5'LTR from HeLa CD4+ HIV-1 LTR-β-gal cells infected with HIV Δvif that was produced in the absence or presence of A3G or A3C orthologs. Relative decrease in virus infectivity is shown for A3G, A3C monomers (hA3C, cA3C K115N, gA3C K115N), A3C monomers/dimers (hA3C S188I, cA3C, gA3C), and A3C dimers (hA3C S188I/N115K, cA3C S188I, gA3C S188I). The graph denotes whether the A3C is from human (h), chimpanzee (c), or gorilla (g). Results normalized to the no A3 condition are shown with the Standard Deviation of the mean calculated from at least three independent experiments. Statistical significance of HIV Δvif restriction for each A3C ortholog was determined in comparison to the monomer form (hA3C, cA3C K115N, or gA3C K115N). Designations for significant difference of values were $p \leq 0.001$ (***), $p \leq 0.01$ (**), or $p \leq 0.05$ (*). (I) Immunoblotting for the HA tag was used to detect A3 enzymes expressed in cells and encapsidated into HIV Δvif virions. The cell lysate and virion loading controls were α-tubulin and p24, respectively. Quantification of the relative amount of A3 was normalized to hA3C.

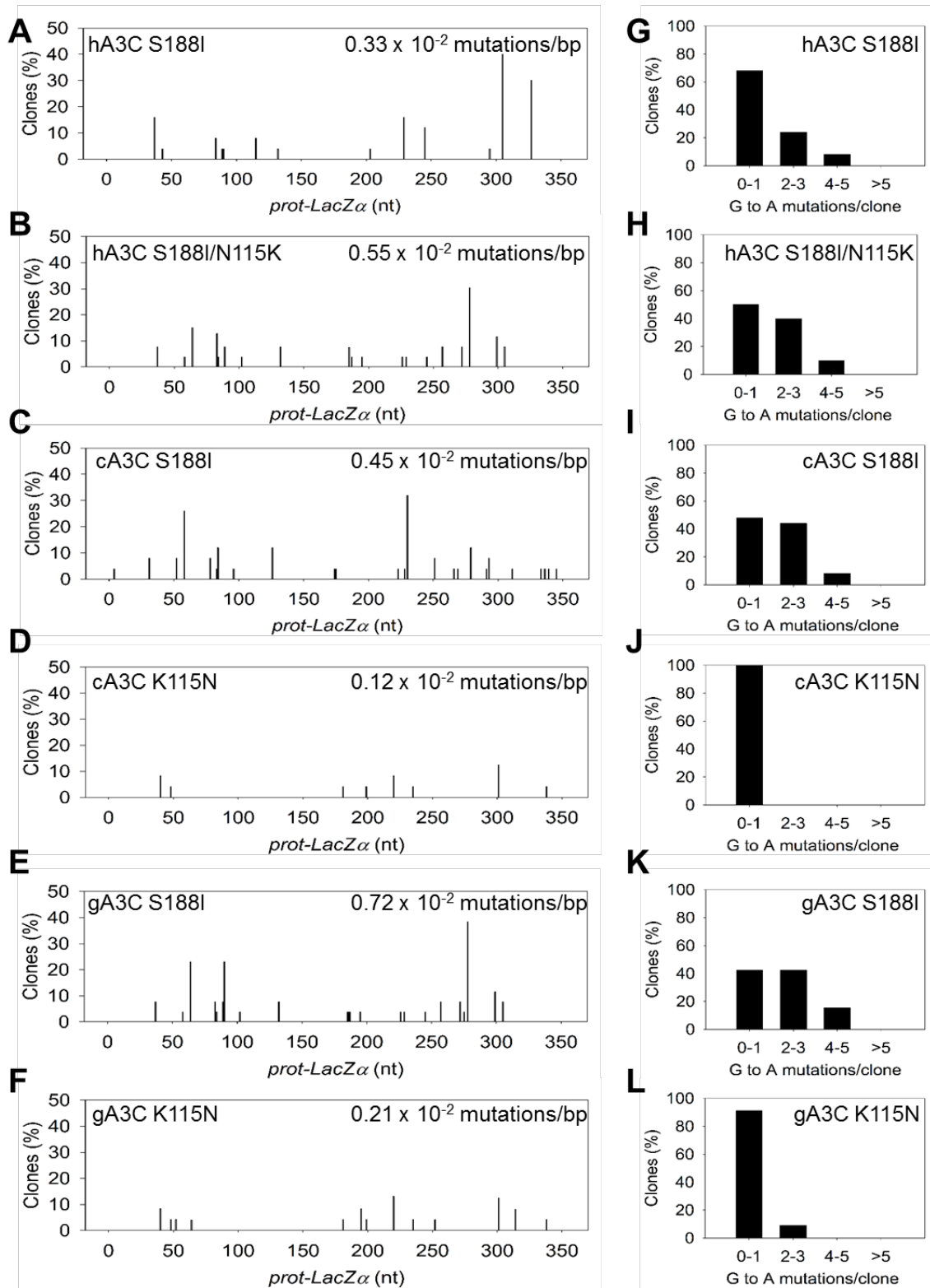


Figure 5.9. Ability to dimerize correlates with increased mutagenesis in vitro. An *in vitro* HIV reverse transcription assay was utilized to determine the mutant A3C enzymes abilities to catalyze deaminations during proviral DNA synthesis. (A–F) Spectra of mutations are plotted as the percentage of clones containing a mutation at a particular location (nt) in the 368 nt *prot-lacZα* construct for (A) hA3C S188I, (B) hA3C S188I/N115K, (C) cA3C S188I, (D) cA3C K115N, (E) gA3C S188I, (F) gA3C K115N. (G–L) Histograms illustrate the number of mutations per clone that can be induced by (G) hA3C S188I, (H) hA3C S188I/N115K, (I) cA3C S188I, (J) cA3C K115N, (K) gA3C S188I, (L) gA3C K115N.

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

A3C	Base pairs sequenced	Total G→A mutations	Mutations per kb
hA3C	9477	3	0.03
cA3C K115N	9477	2	0.02
gA3C K115N	10179	4	0.03
hA3C S188I	10179	8	0.07
cA3C	8073	3	0.04
gA3C	9126	6	0.06
hA3C S188I/N115K	9126	15	0.16
cA3C S188I	9126	11	0.13
gA3C S188I	10179	17	0.17

comparable to the in vitro assay since the in vitro assay uses a selection process to recover mutated clones and the proviral DNA is sequenced without selection, thus, the latter mutation frequencies are lower. Nonetheless, consistent with the in vitro data (Figure 5.8A-F and Figure 5.9), with increased dimerization we observed an increase in the total number of mutations (Table 5.2) and an increase in the number of mutations per clone (Figure 5.10). Specifically, formation of a monomer/dimer resulted in a 2-fold increase in mutation frequency, stable dimers resulted in a 3- to 5-fold increase in mutation frequency, and formation of a monomer resulted in a 2-fold decrease in mutation frequency (Table 5.2). The only exception is that the cA3C mutation frequency is not as high as the other monomer/dimer forms (hA3C S188I and gA3C) (Table 5.2). This is also reflected in the similar decreases of infectivity induced by the monomer (cA3C K115N) and monomer/dimer (cA3C) forms (Figure 5.8H). This is likely due to cA3C encapsidating 2- to 4-fold less than other monomer/dimer A3C forms (hA3C S188I and gA3C) (Figure 5.8I). Notably, A3G encapsidation is at least 2.5-fold more efficient than A3C, which provides reasoning for why dimer forms of A3C are still not as effective as A3G in decreasing viral infectivity, although A3G is also more processive than A3C (Figure 5.8I and Figure 5.1). Despite the level of encapsidation being an additional determining factor to processivity in virus restriction between A3G and A3C, within the A3C orthologs, the decreases in infectivity in the majority of conditions correlates better with the processivity rather than encapsidation. For example, gA3C S188I is encapsidated 2-fold less than gA3C, but is able to restrict virus infectivity 2-fold more (Figure 5.8H), is the more processive A3C (Figure 5.6C), is able to induce more mutations (Table 5.2), and more mutations per clone (Figure 5.10). Similarly, hA3C S188I/N115K is encapsidated 2-fold less than hA3C, but is able to restrict virus infectivity ~2.5-fold more (Figure 5.8H), is the more processive A3C (Figure 5.6A), is able to induce more mutations (Table 5.2), and more mutations per clone (Figure 5.10). Altogether, these data demonstrate the importance of an enzyme's processive ssDNA scanning mechanisms for inducing mutagenesis in a dynamic system where there is a limited amount of time that the ssDNA is available.

5.4 Discussion

Our biochemical analysis has established that A3C dimerization correlates with processive DNA scanning. The hA3C S188I variant dimerizes, in contrast to the common hA3C (Figure 5.4). The cA3C and gA3C were able to dimerize despite containing an S188 by using an

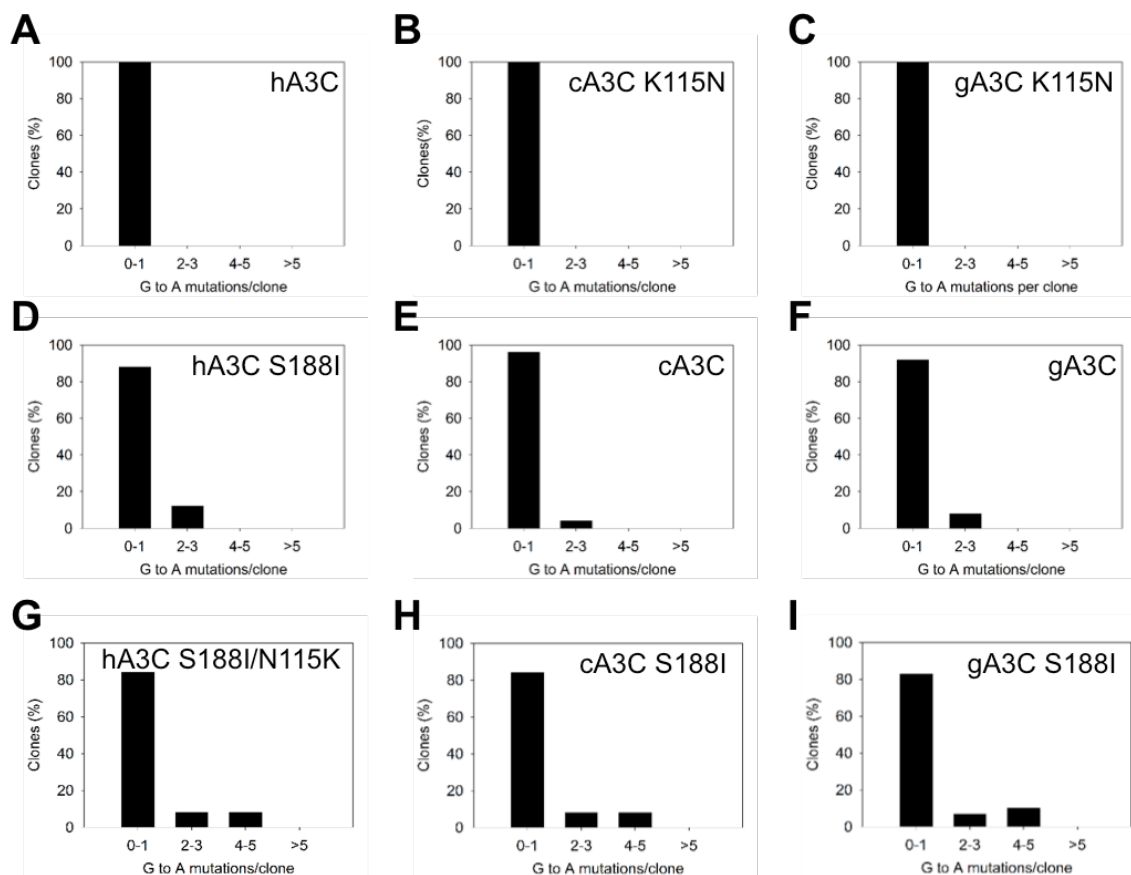


Figure 5.10. Ability to dimerize correlates with increased mutagenesis in integrated proviral HIV Δ vif DNA. A single-cycle infectivity assay was utilized to determine the ability of A3C enzymes to catalyze deaminations during proviral DNA synthesis. Histograms illustrate the number of mutations per integrated proviral DNA clone that can be induced by (A) hA3C, (B) cA3C K115N, (C) gA3C K115N, (D) hA3C S188I, (E) cA3C (F) gA3C, (G) hA3C S188I/N115K, (H) cA3C S188I, and (I) gA3C S188I.

amino acid unique to cA3C and gA3C, K115 (Figure 5.4). These data support the model that dimerization of A3C is primarily mediated through α -helix 6 (human S188I) or β -strand 4 (chimpanzee and gorilla). The importance of these biochemical characteristics is that higher levels of A3C-mediated mutagenesis during reverse transcription requires a dimeric and processive enzyme. Altogether, these data provide a biochemical explanation for why hA3C S188I is able to restrict HIV-1 more than hA3C [28].

Modulation of catalytic activity in the AID/APOBEC family

The A3C data on processivity and specific activity demonstrate that differences in catalytic activity were due to the different efficiencies in the ssDNA scanning mechanisms employed and that this was dependent on the oligomerization state (Figure 5.2 and Table 5.1). In the AID/APOBEC family there have been several common observations that amino acids outside of the active site can mediate catalytic activity. A3F specific activity can be increased by introducing sliding activity through mutations in α -helix 6, a structure that is located away from the active site [108, 262, 263]. A3G and A3A deamination activity can be enhanced by a secondary Zinc ion that binds loop 3, outside of the active site [264]. The Zinc ion is not used directly for catalysis, but stabilizes loop 3 in a conformation that promotes the correct orientation of substrate binding for catalysis [264]. A3A activity is also enhanced by cooperative dimerization [265]. Further, the related family member Activation Induced cytidine Deaminase (AID) that deaminates within specific regions of immunoglobulin genes to promote somatic hypermutation and class switching has been characterized as being a catalytically inefficient enzyme [188, 266, 267]. The low efficiency of catalysis is despite AID's high processivity and ability to remain bound to ssDNA for an average of 5 min [268]. Data indicates that the low efficiency is due to an inaccessibility of the catalytic pocket to DNA [266]. It was proposed that this causes a high propensity of catalytically unfavorable ssDNA-AID binding conformations, which may afford some protection against off-target mutations in genomic DNA [188, 266]. This may also be why the majority of humans carry an A3H allele for a thermodynamically unstable enzyme (haplotypes I, III, IV, or VI) although this does not completely protect from the ability of A3H to contribute to mutations that arise during cancer [59, 60, 174]. The relative inactivity of A3C in the majority of humans due the loss of oligomerization may also be a mechanism to decrease off-target mutations, despite its ability to access genomic DNA in cells [28, 269].

Processivity of A3C

We observed that the primary mechanism used by the A3C enzymes to locate target cytosines on ssDNA was long-range translocations involving jumping or intersegmental transfer. Even though hA3C S188I, cA3C, and gA3C were able to slide on ssDNA, in contrast to hA3C, the processivity factors indicated that they were only 1.5- to 2- fold more likely to catalyze processive deaminations than non-processive deaminations on closely spaced cytosines (Figure 5.2). Thus, the sliding was a minor component of their processive scanning mechanism. The long-range movements, mediated by jumping or intersegmental transfer, were more significant and enabled hA3C S188I, cA3C, and gA3C to undergo processive long-range movements ~3-fold more than non-processive interactions with ssDNA (Figure 5.2). Notably, all the A3C enzymes we tested were 2- to 4- fold less processive and 1.7- to 20- fold less mutagenic during *in vitro* reverse transcription than other A3 enzymes that restrict HIV-1 characterized to date (Figure 5.6 and Figure 5.8) [108, 110, 111]. This may explain why although A3C mutations during proviral DNA synthesis increased with dimerization, there was still a portion of proviral clones that only had one mutation even in the presence of a processive and dimeric A3C (Table 5.2 and Figure 5.10). In agreement with another study by Byeon *et al.*, we found A3C to have less specific activity than other A3s [108, 110, 119, 254, 270-272]. However, the A3C in the study from Byeon *et al.* was produced from *E. coli* and is less active than our A3C produced from Sf9 cells, similar to what has been found for AID produced from Sf9 and *E. coli* cells [270, 273]. Based on these data, we expect that hA3C S188I would contribute to mutagenesis along with other A3 enzymes rather than be able to fully suppress the virus independently of other A3s [28].

The ability to oligomerize has been correlated with improved processivity and specific activity for A3G and A3H, although monomers of the enzymes are still catalytically active [99, 107, 110, 254, 274]. Oligomerization of A3 enzymes allows for multiple binding and interaction domains, which imparts a larger selection of ssDNA scanning mechanisms to efficiently locate the target cytosine. Consistent with previous literature, the current data with A3C enzymes also found that those enzymes that were able to form dimers had higher processivity factors than the enzymes that were unable to dimerize [107, 110]. Further, the high processivity factors were due to the enzyme not only improving in processivity, but also gaining a processive mechanism, e.g., hA3C S188I improved long-range processivity, but also gained the ability to slide (Figure 5.6A). The A3C enzymes that formed stable dimers had the most improvement in their ability to

catalyze processive deaminations (Figure 5A-C, hA3C S188I/N115K, cA3C S188I, gA3C S188I).

For A3C, a single Zinc coordinating domain (Z-domain) enzyme, dimerization is required for intersegmental transfer to occur. We initially expected that the stable dimer forms of A3C would have enhanced intersegmental transfer since the dissociation or reassociation of dimers on ssDNA would not be rate limiting. Rather, the data supports the conclusion that for A3C, intersegmental transfer required an unstable dimer (Figure 5.6D-F). The data suggest that the stable dimer is in a conformation where it is not favorable to bind two different ssDNA segments simultaneously. That stable dimer forms of A3C were unable to undergo intersegmental transfer may also explain why enzymes like A3G and A3F that have two DNA binding domains within a single polypeptide chain are unable to utilize intersegmental transfer, under the same reaction conditions tested with A3C [108], although under low salt buffer reaction conditions A3G may be able to undergo intersegmental transfer [254]. The improved processivity of the stable A3C dimer forms was through sliding and long-range movements, but the long-range movements were largely comprised of jumping (Figure 5.6 and Figure 5.7). We found A3C to use long-range movements similarly to A3H, in which jumping and intersegmental transfer are redundant processive mechanisms and for A3C either one can improve the mutagenic efficiency of the enzyme [110].

Our study with A3C raises the possibility that the structures that mediate processivity are different for single and double Z-domain A3 enzymes. In contrast to A3G and A3F that contain two Z-domains, α -helix 6 was not directly involved in mediating A3C sliding [107, 108]. The hA3C, cA3C, and gA3C all had identical amino acids in the α -helix 6 despite differing abilities to slide along ssDNA (Figure 5.2). In A3C, the ability of α -helix 6 to mediate dimerization or be part of the dimer interface was the only requirement for sliding movements on ssDNA. In contrast, a monomer of A3G can still retain sliding ability and an oligomer of A3F is unable to slide unless specific mutations are made to the loop region extending from α -helix 6 [107, 108]. In a study of a single Z-domain enzyme, A3H, specific amino acid changes on α -helix 6 could influence sliding of the dimeric A3H [110]. Nonetheless, a β 2-strand dimerization mutant had compromised sliding ability, despite a wild type α -helix 6 [110]. These results suggest that the dimerization interface in single Z-domain A3s creates a groove along the enzyme that promotes an interaction with ssDNA and facilitates sliding [275]. The double Z-domain enzymes A3G and

A3F may inherently have an extended interaction interface with ssDNA due to possessing two ssDNA binding domains in a single polypeptide chain. This appears to facilitate extensive ssDNA interactions that may also involve bending of the ssDNA around the enzyme [102, 253, 276].

Dimerization of A3C

Consistent with previous studies *in vitro* and in cells, we determined that hA3C was a monomer [28, 51, 149]. This is in contrast to Stauch *et al.* that determined hA3C to be a dimer by co-IP [255]. Although Stauch *et al.* and our study used a similar co-IP strategy, the transfected plasmid amounts were different and Stauch *et al.* did not use RNaseA in their procedure. These factors can lead to nonspecific interactions in the co-IP [277, 278]. A surprising feature of our results was that the biochemical data we obtained support a model in which the hA3C dimer formed in the crystal structure did form on a physiologically relevant interface for hA3C S188I, cA3C, and gA3C (Figure 5.3B) [149]. However, it should be noted that multiple additional interfaces did form when hA3C was crystallized and these do not appear to represent the dimer interface in hA3C S188I, cA3C, or gA3C [149].

Although the biochemical and co-IP data definitively show that amino acids 115 and 188 mediate A3C dimerization, there is a caveat in our model that relies on the hA3C crystal structure. Namely, visual inspection of amino acids 115 and 188 on the structure of hA3C suggest that amino acid 115, but not amino acid 188, is directly involved in the dimer interface (Figure 5.3B). For hA3C the N115 is 4.7 Å away from the backbone of R44. In contrast, a model of cA3C K115 positions the backbone of R44 only 3.0 Å away (Figure 5.11A-B). Thus, the mechanism by which the K115 in cA3C promotes dimerization appears to be due to enabling a new hydrogen bond with R44 on the other subunit (Figure 5.11A-B). The interface between subunits is already surface complementary and this complementarity is unlikely to be disrupted by the change in the 115 side chain identity. Therefore, the new hydrogen bond formed by the K115 side chain and R44 backbone would mainly serve to strengthen the interface with minimal tradeoffs.

Although the amino acid at position 188 mutation does not appear to be directly involved in the dimer interface by PISA analysis or in the structural model [149, 279], PISA analysis does suggest that amino acids surrounding position 188 on α -helix 6 are involved in dimerization. The most likely explanation for the role of the S188I mutation in altering the dimer affinity is that it

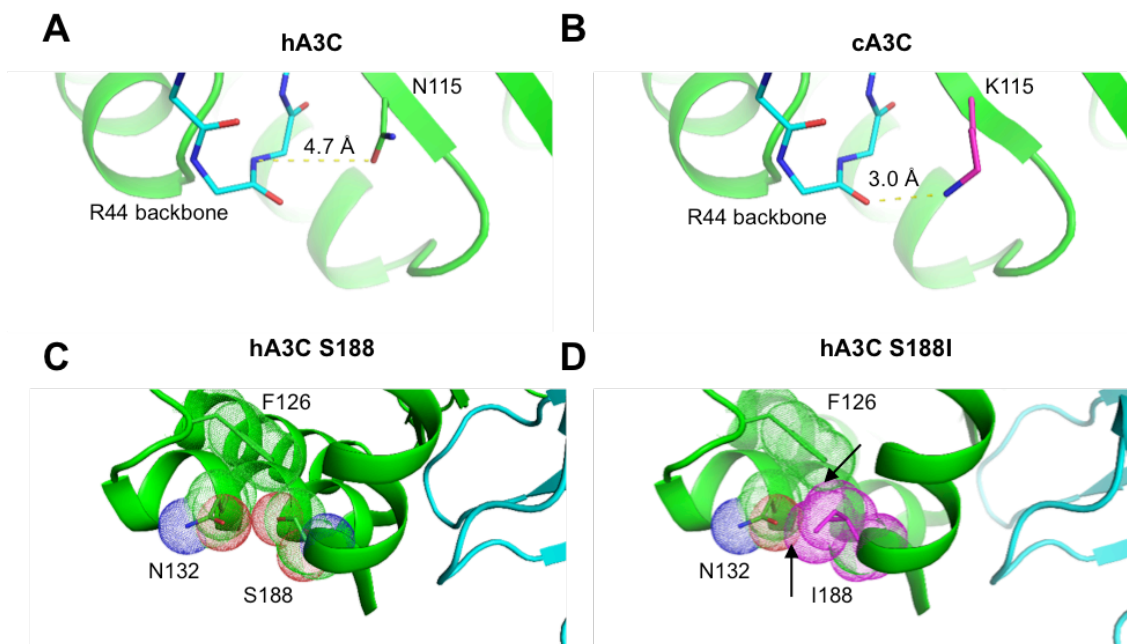


Figure 5.11. Models of A3C dimerization. (A) For hA3C, N115 (chain A) is 4.7 Å away from the backbone of R44 (chain B). (B) In contrast, a model of cA3C K115 (chain A) positions the backbone of R44 (chain B) only 3.0 Å away. This distance could indicate a new hydrogen bond with potential to stabilize the dimer between chains A and B. (C) In hA3C, S188, shown with van der Waal space filling dots, packs closely to F126 and N132, but does not clash with either. (D) In contrast, a model of hA3C shows how I188 would clash with F126 and N132 (arrows indicating overlap of van der Waal space filling dots). Conformational changes, potentially including a repositioning of the helix to enable formation of an A-B dimer, would be needed to accommodate this amino acid variant.

causes steric clash with other buried residues that are accommodated by shifting the position of helix 6. This repositioning of the dimer interface could alter the surface complementarity and other interactions, leading to the change in affinity (Figure 5.11C-D). Based on our studies, the hA3C crystal structure gives a good approximation of the dimer interface, but more mutagenesis studies or crystallization of a dimer form of A3C is required to identify the interface in more detail [149].

A3C forms a dimer interface unique from other A3 dimer interfaces determined thus far [98, 99, 110, 262, 263, 265, 280, 281]. Notably, the other dimer interfaces characterized have been from the A3G N-terminal domain in the full length enzyme (Z2-Z1-type Z-domain organization), rhesus macaque A3G N-terminal domain alone (Z2-Z1-type Z-domain organization), A3F C-terminal domain alone (Z2-Z2-type Z-domain organization), A3A (a Z1-type Z-domain), and A3H (a Z3-type Z-domain) [98, 99, 110, 262, 263, 265, 280, 281]. A3C is the first Z2-type single Z-domain to have the dimer interface characterized. However, the other Z2-type Z-domains that have been characterized structurally, A3F C-terminal domain and rhesus macaque A3G N-terminal domain further support the involvement of α -helix 6 [262, 280]. In the rhesus macaque A3G N-terminal domain structure, α -helix 6 and loop 7 are involved in the dimer interface [280]. There have been several A3F C-terminal domain (CTD) crystal structures solved and each with a different predicted dimer interface based on crystal contacts [262, 263, 281]. In one A3F crystal structure, the amino acid equivalent to A3C N/K115 in the A3F CTD, N298 is involved in a dimer interface [262]. It is not clear why the A3F Z2-type Z-domain CTD structures are not consistent with respect to the dimer contacts made in each of the crystal structures, however, this may be due to different constructs being used between labs to promote crystallization, that the A3F CTD is mutated to improve solubilization, and that it is not the full length A3F enzyme [262, 263, 281].

Conclusions

Our results with these three hominid A3C orthologs demonstrates that there is more than one mechanism to maintain enzyme processivity, although the common endpoint requires dimerization of the enzyme. Since A3C is a Z2-type domain, the dimer model may represent how other Z2-type domains such as A3F and A3D dimerize (both have Z2-Z2-type Z-domain organization). That the mechanism for achieving processivity in A3C is based primarily on dimerization rather than specific amino acid motifs that interact with the ssDNA substrate is

novel from other A3s and indicates that specific adaptations to maintain activity have been different for individual A3 enzymes. Although we have shown that the dimerization of hA3C through the S188I mutation increases processivity and mutagenic ability, thus providing a reason for why it is able to restrict HIV-1, in contrast to hA3C [28], it remains to be determined if the increased activity naturally found in cA3C and gA3C results in these enzymes restricting SIV. Despite their higher enzymatic activity it has not been demonstrated if cA3C and gA3C are able to encapsidate into SIV, which would be required for them to act as restriction factors. Alternatively, the naturally higher activity of cA3C and gA3C may be used to restrict viruses other than SIV or retroelements. Overall, the data support the model that A3 enzymes must be able to deaminate multiple cytosines processively to be efficient at mutagenesis during reverse transcription and for A3C this requires dimerization.

5.5 Materials and Methods

5.5.1 Cloning and site directed mutagenesis. The sequence for hA3C, cA3C, and gA3C have been previously described [27, 28]. The A3C sequences were cloned into a baculovirus transfer vector (pAcG2T or pFAST-bac1) containing an N-terminal GST tag or pcDNA3.1 with an N-terminal HA or V5 tag. Mutants were made by site-directed mutagenesis of the wild type sequences to create hA3C S188I, hA3C S188I/N115K hA3C N115K, cA3C S188I, cA3C K115N, gA3C S188I, and gA3C K115N. All constructed plasmids were verified by DNA sequencing.

5.5.2 Protein expression and purification. Recombinant baculovirus production for expression of wild type and mutant hA3C, cA3C and gA3C in *Sf9* cells was carried out using the pACG2T or pFast-bac1 transfer vector as previously described [110, 282]. *Sf9* cells were infected with recombinant GST-A3C virus at an MOI of 1 (hA3C, cA3C, gA3C, cA3C S188I, hA3C S188I/N115K, gA3C S188I, and gA3C K115N) or an MOI of 5 (hA3C S188I, hA3C N115K and cA3C K115N). Recombinant baculovirus infected *Sf9* cells were harvested after 72 hours of infection. Cells lysates treated with RNaseA were incubated with glutathione-Sepharose 4B resin (GE Healthcare) at 4°C and were subjected to a series of salt washes, as previously described [99]. The enzymes were eluted with the GST tag in elution buffer (100 mM Tris, pH 8.8, 150 mM NaCl, 10% (v/v) glycerol, and 50 mM reduced glutathione). The A3C enzymes were then cleaved from the GST tag in solution at 21°C for 6 hours with Thrombin (GE Healthcare) before being dialyzed against 100 mM Tris pH 7.5, 250 mM NaCl, 10% (v/v) glycerol, and 1 mM DTT.

To purify A3C from the free GST and thrombin, the enzyme stock was diluted to achieve a solution of 50 mM Tris pH 7.5, 50 mM NaCl, 10% (v/v) glycerol and 1 mM DTT for loading onto a DEAE FF column (GE Healthcare). A3C was eluted with a linear gradient of NaCl. The A3C eluted at approximately 350 mM NaCl and was >95% pure, as determined by SDS-PAGE

5.5.3 *In vitro* deamination assay. All ssDNA substrates were obtained from Tri-Link Biotechnologies and have been published previously [108]. Reactions were carried out under single-hit conditions (*i.e.* <15% substrate usage) to ensure that each ssDNA reacted with at most a single enzyme [242]. Under these conditions, a processivity factor can be determined by comparing the quantified total amount of deaminations occurring at two sites on the same ssDNA with a calculated theoretical value of deaminations at these two sites if the deamination events were uncorrelated (not processive) [111]. A ssDNA substrate containing two 5'TTC motifs (100 nM) was incubated with 350 nM (gA3C, hA3C S188I, gA3C S188I, cA3C S188I, hA3C S188I/N115K, and hA3C N115K) or 700 nM (hA3C, cA3C, cA3C K115N, gA3C K115N) of A3C for 2.5 to 30 min at 37 °C in RT buffer (50 mM Tris, pH 7.5, 40 mM KCl, 10 mM MgCl₂, and 1 mM DTT). The reaction time was varied on each ssDNA according to the specific activity of the enzymes to ensure ~10% substrate usage. For intersegmental transfer assays, the A3C:ssDNA ratio (3:1 or 7:1) was kept the same, but increasing concentrations of enzyme and substrate was titrated in (ssDNA: 100-500 nM, A3C: 350-1750 nM or 700-3500 nM). Reactions were started by the addition of the ssDNA substrate. A3C-catalyzed deaminations were detected by treating the ssDNA with uracil DNA glycosylase (New England Biolabs) and heating under alkaline conditions before resolving the fluorescein-labeled ssDNA on 10 or 20% (v/v) denaturing polyacrylamide gels, depending on the sizes of the ssDNA fragments. Gel photos were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare) and integrated gel band intensities were analyzed using ImageQuant (GE Healthcare). The specific activity was calculated from single-hit condition reactions by determining the picomoles of substrate used per minute for a microgram of enzyme.

5.5.4 *In vitro* reverse transcription assay. Mutagenesis of ssDNA by A3 enzymes during reverse transcription of an RNA template was assessed using a previously established *in vitro* assay, which models reverse transcription of an RNA template and second-strand synthesis [107]. In brief, a synthetic (+)RNA that contains the poly-purine tract (PPT), the catalytic region of the protease gene (120 nt), and lacZ α (248 nt) is used. In this construct the PPT is used as a

primer for (+)DNA synthesis and enables synthesis of dsDNA. The lacZ α serves as a reporter gene for mutations by blue/white screening. To start the assay, the 368 nt RNA template (50 nM) is annealed to a DNA primer (24 nt) and incubated with 1.5 μ M of nucleocapsid, 1.2 μ M of reverse transcriptase and 500 μ M of dNTPs in RT buffer in the presence or absence of 350 nM of each A3C enzyme. The resulting dsDNA that was synthesized from this in vitro system was PCR amplified using Pfu C $_x$ Turbo Hotstart (Agilent Technologies) that can use uracils as a template with high fidelity. These amplicons were then cloned into a pET-Blue vector backbone that allows for blue-white screening of the synthesized lacZ α . At least twenty-five mutated clones for each condition were sequenced.

5.5.5 Size exclusion chromatography. The oligomerization states of the A3C enzymes were determined by loading 10 μ g of purified enzyme on a 10 mL Superdex 200 (GE Healthcare) size exclusion column. The column was prepared by pouring the resin bed in a column with 16 cm height and 0.5 cm diameter. The running buffer contained 50 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT. The chromatograms from the 10 ml Superdex 200 column were constructed by analyzing the integrated gel band intensities of the protein in each fraction after resolution by SDS-PAGE. The gels for each panel were resolved, stained with Oriole fluorescent gel stain, and scanned in parallel. The Bio-Rad standard set was used to generate a standard curve from which molecular weight and oligomerization states of the enzymes were determined.

5.5.6 Protein crosslinking. A3C enzymes (0.5 μ M) were incubated with 20 μ M BS3 (bis (sulfosuccinimidyl) suberate), an amine to amine crosslinker, in 20 mM HEPES (pH 7.5), 150 mM NaCl and 1 mM DTT for 1 h at 21°C. Crosslinked proteins were resolved by SDS-PAGE, transferred to a nitrocellulose membrane for immunoblotting, and visualized using primary antibody for native A3C (GeneTex) and secondary IRdye labeled goat anti-rabbit antibody compatible with the LI-COR/Odyssey system.

5.5.7 Steady state rotational anisotropy. Steady state fluorescence depolarization (rotational anisotropy) was used to measure the binding affinity of A3C-ssDNA and a 118 nt fluorescein-labeled ssDNA. Reactions were 60 μ L and contained fluorescein-labeled ssDNA (10 nM) in RT buffer. The enzyme was titrated into the solution until saturation. For saturation the required titrations used were: hA3C (0-1600 nM); hA3C S188I, cA3C S188I, gA3C S188I, cA3C, and gA3C (0-1200 nM); hA3C S188I/N115K, hA3C N115K, and cA3C N115K (0-1100 nM); and gA3C K115N (0-1800 nM). A QuantaMaster QM-4 spectrofluorometer (Photon

Technology International) with a dual emission channel was used to collect data and calculate anisotropy. Samples were excited with vertically polarized light at 495 nm (6 nm band pass) and vertical and horizontal emissions were measured at 520 nm (6 nm band pass). Apparent dissociation constants (K_d) were obtained by fitting to a rectangular hyperbola or sigmoidal curve using Sigma Plot 11.2 software.

5.5.8 Co-immunoprecipitation. Co-immunoprecipitation (Co-IP) assays were conducted as described previously [283]. Briefly, 293T cells (2.5×10^6 cells per 75 cm² flask) were transfected with 1 µg total DNA. Equal amounts of each plasmid for hA3C-HA and hA3C-V5, hA3C S188I-HA and hA3C S188I-V5, or hA3C S188I/N115K-HA and hA3C S188I/N115K-V5 were used to transfect the cells. Genejuice transfection reagent (EMD Millipore) was used according to the manufacturer's instructions. At 72 h post transfection, the cells were washed with PBS and lysed in IP buffer (50 mM Tris-Cl pH 7.4, 1% Nonidet-P40, 0.1% sodium deoxycholate, 10% glycerol, 150 mM NaCl) supplemented with RNaseA (20 µg/mL; Roche) and EDTA-free protease inhibitor (Roche). One half of the precleared supernatant was then incubated with Protein A-agarose conjugated polyclonal rabbit anti-HA antibody (2 µg, Santa Cruz Biotechnology) and the other half (mock) was incubated with Protein A-agarose-conjugated normal rabbit IgG (2 µg, Santa Cruz Biotechnology) at 4°C for 2 h. Resin was washed and the samples were then resolved by SDS-PAGE and transferred to a nitrocellulose membrane. For detection of A3C-HA and A3C-V5 in cell lysates, the membrane was probed with polyclonal Rabbit HA (1:1000, Sigma) and polyclonal Rabbit V5 (1:200, Santa Cruz Biotechnology), respectively. For the loading control, monoclonal mouse anti-α-tubulin (1:1000, Sigma) was used. HA- and rabbit IgG- immunoprecipitated lysates were probed with anti-V5 mouse monoclonal antibodies. After incubation with Horse Radish Peroxidase (HRP) conjugated secondary antibodies, the blots were visualized with X-ray film using Super Signal West Pico chemiluminescence substrate (Thermo-Scientific).

5.5.9 Single cycle infectivity assays. The 293T cells were plated at a density of 5×10^4 cells per well of a 12 well plate. The next day, the cells were transfected with 0.50 µg pLAI HIV Δvif, 0.15 µg pVSV-G, and 0.30 µg pCDNA3.1 A3C-HA or 0.10 µg pCDNA3.1 A3G-HA. Different levels of transfected DNA were used to achieve the same steady state expression levels in cells for A3C-HA and A3G-HA. All transfection amounts of DNA were equalized with empty pCDNA3.1 plasmid. Forty-eight hours after transfection, culture supernatants containing virus

were collected, filtered, and used to infect HeLa CD4+ HIV-1 LTR- β -gal cells (MAGI) cells [284]. To detect A3 by immunoblotting, harvested supernatants were centrifuged at 16 000 x g for 60 min to recover virions and cells from the same well were washed with PBS and lysed with Laemmli sample buffer. Forty-eight hours after infection of HeLa CD4+ HIV-1 LTR- β -gal cells, Pierce β -galactosidase assay reagent was used to lyse cells and detect β -galactosidase activity through colorimetric detection using a spectrophotometer. Infectivity of each virus was compared by setting infectivity of the “No A3” to 100%. Statistical significance of results was determined using an unpaired t-test.

5.5.10 Quantitative immunoblotting. The 293T cells expressing A3C-HA and A3G-HA from the single-cycle infectivity assays were detected using anti-HA (1:1000; Sigma) in cell lysates (30 μ g total protein) and virions (20 μ l of concentrated virus). Loading controls for cell lysates (α -tubulin, Sigma) and virions (p24, Cat #3537, NIH AIDS Reagent Program) were detected using mouse monoclonal antibodies. Odyssey software was used to quantitate the amount of p24 in each lane and ensure that similar amounts of virus lysate were loaded. Proteins of interest and loading controls were detected in parallel by using the Licor/Odyssey system (IRDye 680-labeled goat anti-rabbit secondary and IRDye 800-labeled goat anti-mouse secondary antibody). Using Odyssey software, the relative amount of A3 on the blot was calculated by first normalizing each sample lane to the corresponding control. Normalized values were then converted to relative amounts of A3 by setting the hA3C band at 1.0 and calculating the relative amounts of A3 in other lanes.

5.5.11 Sequencing of integrated proviral DNA. Total DNA from infected 293T cells was extracted 48 h after infection using DNAzol reagent (Life Technologies). DNA was treated with DpnI (New England Biolabs) for 1 h at 37°C to remove possible contaminating plasmid DNA and the *protease* (nt 2280-2631) sequences were amplified by PCR using Q5 polymerase (New England Biolabs). Primers have been published previously [108]. The PCR products were purified and cloned with the CloneJET PCR cloning kit (Thermo Fisher). DNA was sequenced with kit-specific primers and carried out at the National Research Council of Canada (Saskatoon, Canada).

5.5.12 Structural models of A3C. The dimer model used was present in the asymmetric unit of the crystal structure (PDB: 3VOW). The model mutations created were done by manual

mutagenesis in Coot (Crystallographic Object-Oriented Toolkit), selecting rotamers that minimized clashes to the surrounding residues.

6.0 Comparison of A3C orthologs to other A3 enzymes that restrict HIV.

In chapter 3, A3C residues within the Vif interface were demonstrated to be undergoing positive selection. Specifically, residue 77, 128, and 130 all within the predicted A3/Vif interface were identified (Figure 6.1). Therefore, A3C is efficiently targeted for degradation by Vif, satisfying one of the major requirements to be considered a bona-fide restriction factor [285].

As observed in chapter 3 and 5, dimerization of A3C is required for effective restriction of HIV. If Vif were able to disrupt this dimerization, the activity of these enzymes could be reduced. For A3G that is able to restrict HIV and is targeted by Vif, Vif can disrupt A3G oligomerization and reduce A3 activity, as a secondary mechanism of inhibition [283]. Conversely, Vif cannot fully disrupt oligomerization of A3H and A3H is partially resistant to Vif-mediated degradation [283]. Therefore, we aimed to assess the impact of Vif on A3C oligomerization to determine if either of these features characterized for other A3s were relevant to A3C. The Vif interaction interface on A3C includes α -helices 2, 3, 4, and β -strand 3 and 4. Residue 188, which is involved in hA3C oligomerization interface, is located on helix 6 (Figure 6.1). We hypothesized that an interaction between hA3C and Vif would not disrupt A3C oligomerization. We further hypothesized that this may promote resistance to Vif mediated degradation [142]. This was based on speculation that resistance to Vif might be the reason why human A3C has retained or regained the sequence I188. Conversely, residue 115 on β -strand 4 that is involved in chimpanzee A3C (cA3C) and gorilla A3C (gA3C) dimerization is within the Vif interaction interface. Therefore, we hypothesized that Vif would disrupt cA3C and gA3C oligomerization, as was observed for A3G [283].

To determine if Vif could disrupt oligomerization interface of A3C that overlapped with the Vif interaction interface, we compared the SEC profile of A3C alone, a Vif_{HXB2} heterotetramer complex (Vif/CBF β /Elongin B/Elongin C, VCBC), and A3C and VCBC in a equimolar mix to determine if there was a corresponding change in the A3C SEC profile (Figure 6.2). If Vif were able to change the oligomerization status of A3C we would expect the apparent molecular weights calculated from the SEC profile to show a 1:1 A3C:VCBC complex. If Vif does not change the oligomerization status of A3C, then we would expect the apparent molecular weights calculated from the SEC profile to show a 2:1 A3C:VCBC complex. As expected, the control profile of VCBC resolved as a 1:1:1:1 complex (molecular weight 66 kDa), comprised of

Vif/CBF β /EloB/EloC [283]. If VCBC interacts with a dimer of A3C the combined expected molecular weight of the complex would be 111 kDa, however if VCBC interacts with a monomer of A3C the combined expected molecular weight of the complex would be 89 kDa. Using this information, we examined the SEC profiles of different A3Cs in complex with VCBC.

When VCBC was resolved with human A3C (hA3C) S188I, which is an equilibrium of monomer and dimer populations, the VCBC-A3C complex resolved in fractions with molecular weights ranging from 111-52 kDa (Figure 6.2A), indicating that VCBC could interact with both monomer and dimer forms of hA3C S188I, and that Vif did not disrupt the dimerization of hA3C S188I. This is expected as residue 188 lies outside the Vif interface and would not interfere Vif binding. The interaction of VCBC with the stable dimer of hA3C, hA3C S188I/N115K, the A3C-VCBC complex resulted in a 111 kDa species, indicating that VCBC interacted primarily with a hA3C S188I/N115K dimer (Figure 6.2B). However, there was a smaller molecular weight fraction at 73 kDa, which demonstrates that there is some ability of VCBC to disrupt the dimerization. As observed with hA3C S188I, when the dimer is formed through residue 188 it is insensitive to Vif binding and hA3C remains as a dimer in solution.

To further investigate this interaction, we examined the interaction of VCBC and chimpanzee (cA3C), which dimerizes primarily through residue 115 located within the Vif interaction interface. HIV Vif is able to be used in studies with cA3C, instead of the corresponding SIV Vif, as it has been shown that cA3C is able to interact with an HIV derived Vif, unlike gorilla (gA3C) which is not sensitive to HIV Vif [246]. When we examined the SEC profile of cA3C in complex with VCBC we observed a molecular weight range of 78 to 37 kDa, meaning that VCBC is able to disrupt the dimerization of cA3C (Figure 6.2C) and interacts with a monomer form of cA3C. Therefore, cA3C dimerization is sensitive to Vif binding due to the location of residue 115 on β -strand 4, and the dimerization is subsequently disrupted. This is in contrast to the dimerization through residue 188 on helix 6, which is not contained within the interaction interface, and is unable to be disrupted by Vif. We confirmed this observation with a cA3C S188I mutant that forms a stable dimer through the helix 6 residue. Vif was unable to disrupt the dimer portion of this mutant (Figure 6.2D), similar to hA3C S188I (Figure 6.2A).

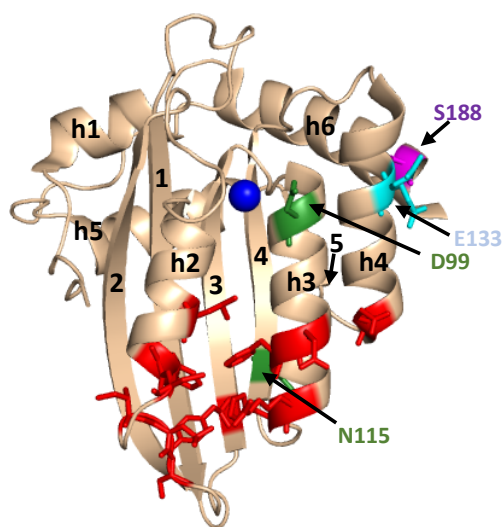


Figure 6.1. Residues of importance to Vif-mediated degradation in hA3C. A3C has the basic APOBEC3 structure that is comprised of a five stranded β -sheet core surrounded by six α -helices. Numerical assignments to β -strands and α -helices (h) are superimposed on the hA3C structure (PDB: 3VOW). Zinc atom shown as a blue sphere. Amino acids unique to hA3C from cA3C and gA3C are K85, D99, and N115 (green). Amino acid 133 is an amino acid unique to gA3C from hA3C and cA3C (cyan). The S188 amino acid has previously been identified to be the location of a SNP in hA3C (magenta, I188). Amino acids where Vif interacts with A3C are shown in red.

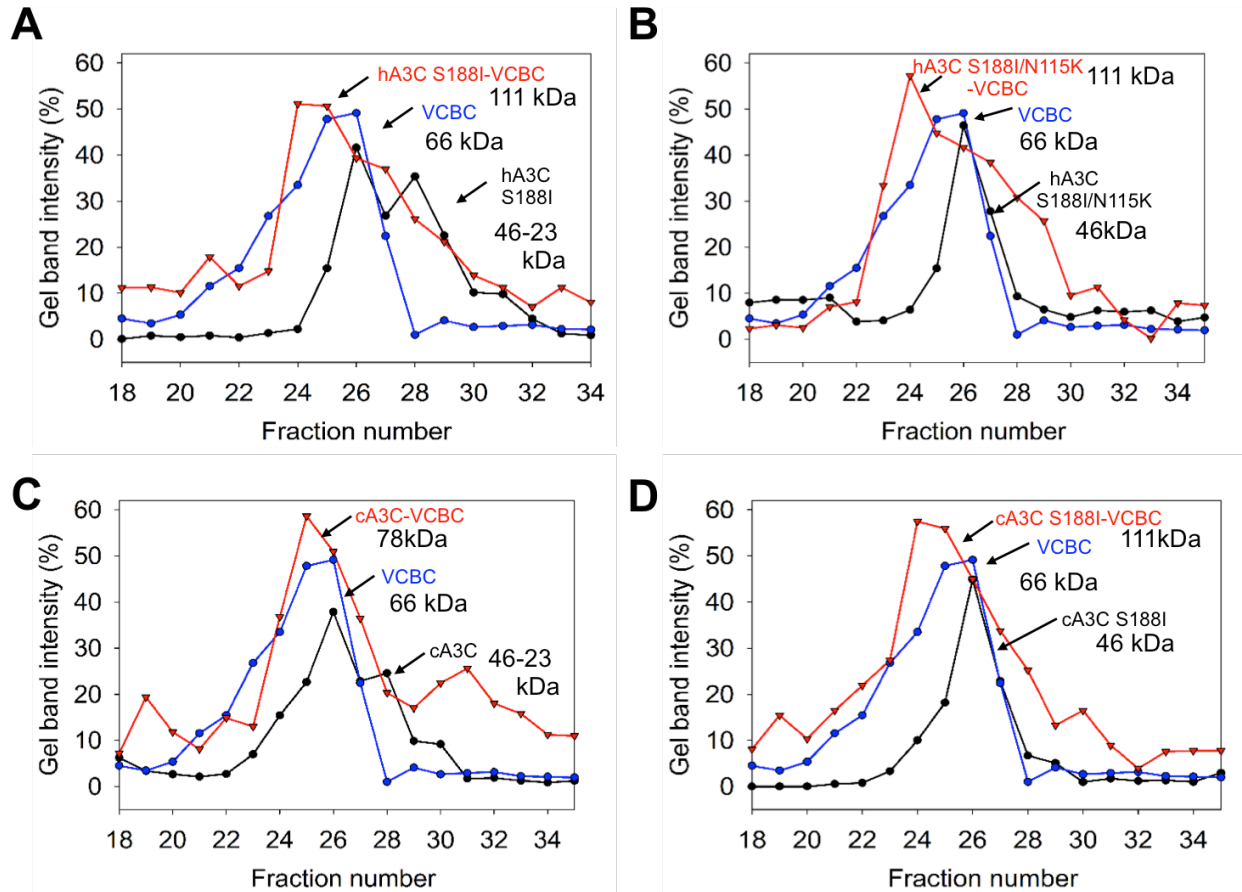


Figure 6.2. *Vif* can disrupt A3C oligomerization. (A-D) The size-exclusion chromatography profile of 10 μ g of (A) hA3C S188I, VCBC, and hA3C S188I with VCBC; (B) hA3C S188I/N115K, VCBC, and hA3C S188I/N115K with VCBC; (C) cA3C, VCBC, and cA3C with VCBC; and (D) cA3C S188I, VCBC and cA3C S188I/VCBC from a 10 mL Superdex 200 column was used to calculate the oligomerization state of the enzyme from a standard calibration curve. (A) *Vif* interacted with both a monomer and dimer form of hA3C S188I (apparent molecular weight 112-58 kDa). (B) *Vif* interacted with both a monomer and dimer form of hA3C S188I/N115K (apparent molecular weight 108-73 kDa). (C) *Vif* interacted with a monomer of cA3C (apparent molecular weight 78-37 kDa). (D) *Vif* interacted with both a monomer and dimer form of cA3C S188I (apparent molecular weight 109-50 kDa). The chromatograms were constructed by analyzing the integrated gel-band intensities of each protein (A3C, VCBC or A3C with VCBC) in each fraction after resolution by SDS-PAGE and quantitative western blotting.

Despite the inability of VCBC to disrupt dimers mediated through helix 6, Vif was able to degrade all A3Cs tested regardless of oligomerization status (Figure 6.3). We had initially hypothesized that when dimerization occurred through residues outside of the Vif interaction interface, Vif would be unable to disrupt these dimers and this would lead to resistance to degradation. However, this was not observed as A3C mutants were degraded in cell based Vif degradation assays regardless of which residues promoted dimerization. These data suggest that the A3C I188 was not retained or regained in humans to impart resistance to Vif-mediated degradation as hypothesized. Altogether, A3C dimerization is sensitive to Vif binding if the interfaces overlap between oligomerization and Vif interaction. However, dimerization of A3C outside this Vif interface does not impart resistance to Vif mediated degradation as seen for A3H, and therefore the positive selection and SNP observed for A3 was not to facilitate Vif resistance.

If A3C S188I is truly an active version of A3C, it should possess common features with the other A3s known to restrict HIV replication efficiently. The ssDNA binding data showed that hA3C S188I could oligomerize on ssDNA, in contrast to hA3C S188 (Figure 5.4H). As a result, we investigated if the hA3C S188I could also restrict HIV by inhibiting RT activity, similar to what has been shown for A3G oligomers, and will be expanded on in Chapter 8 [283]. This may be an additional mechanism used by A3C to restrict HIV and would explain why A3C has undergone positive selection. Although, the hA3C S188I did not inhibit primer initiation or synthesis of DNA by RT (Figure 6.4), we reasoned that the slowing down of RT through oligomerization may alter the ability of RT to undergo template switching or affect its insertion fidelity. This was investigated in comparison to A3G, A3F, and A3H that are known to have restriction activities against HIV that is both deamination -dependent and -independent.

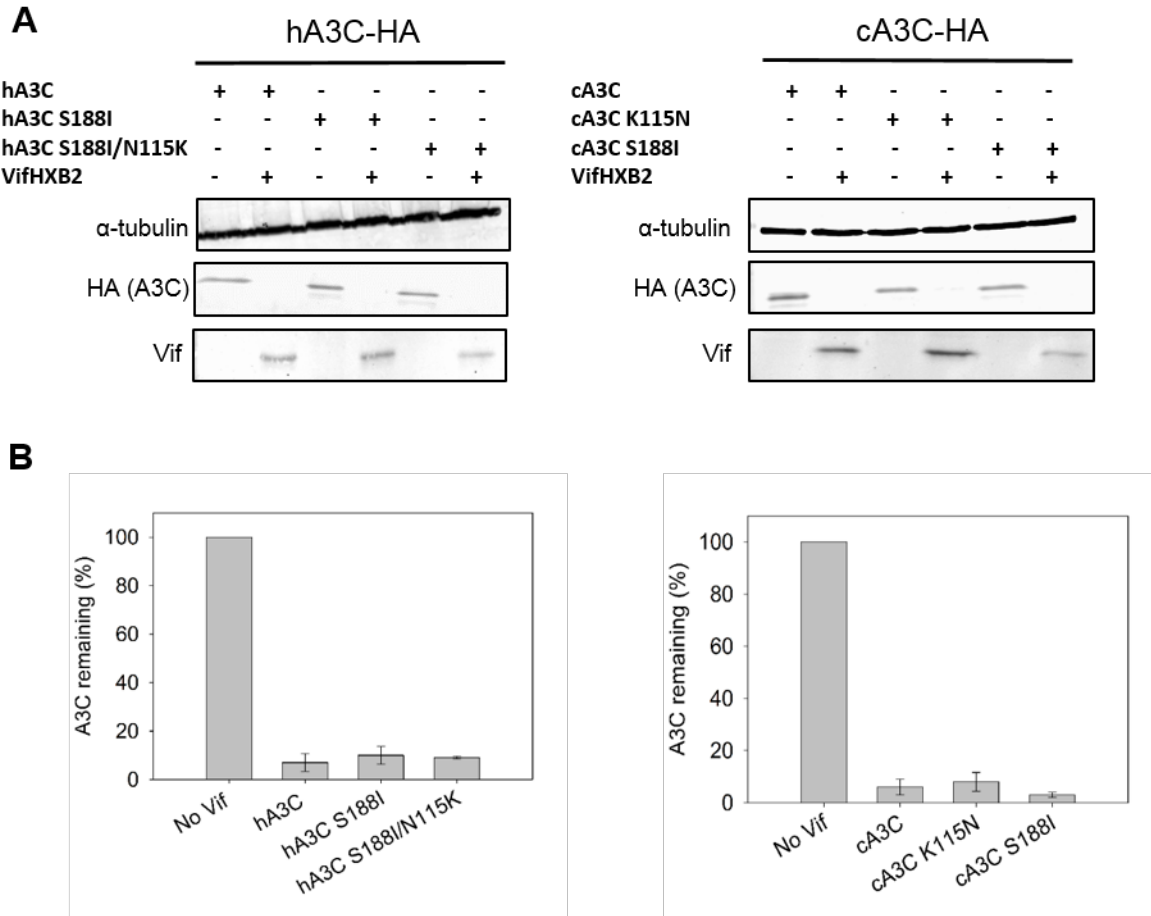


Figure 6.3. *HIV Vif is able to promote the degradation of human and chimp A3C.* (A) Degradation of hA3C-HA and cA3C-HA by Vif_{HXB2}. 293T cells were transiently transfected with A3C-HA expression plasmid with or without cotransfection of a Vif_{HXB2} expression plasmid. Cell lysates prepared in Laemmli sample buffer were analyzed by immunoblotting with antibodies against α -tubulin, HA, and Vif. (B) A bar graph depicting the % A3C remaining in the presence of Vif. Values were obtained from the integrated gel band intensities in the presence and absence of HIV Vif. Average and standard deviation were calculated from three independent experiments.

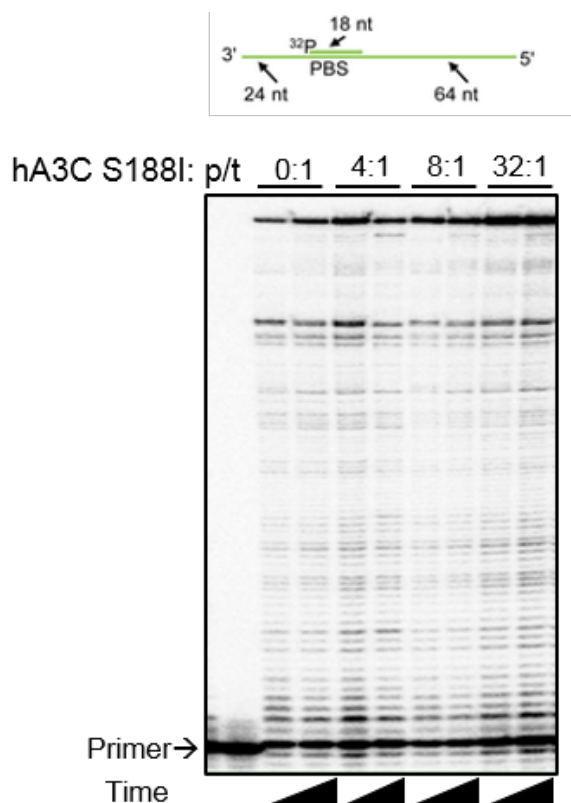


Figure 6.4 A3C does not inhibit the initiation of primer extension. An 18 nt ^{32}P -labeled RNA primer containing a sequence complementary to the HIV-1 PBS was annealed to a 106 nt RNA containing the PBS (sketch). Complete extension of the primer results in a product of 82 nt (sketch). The p/t was used at a concentration of 10 nM. (A) Primer extension by RT (400 nM) in the absence (0:1) or presence (4:1, 8:1, 32:1,) of increasing amounts of hA3C S188I relative to the p/t concentration. Reactions were sampled at 10 and 60 min.

7.0 Investigation of the deamination independent restriction of HIV by APOBEC3s

As discussed in section 1.4.3, A3 enzymes that possess multiple processive mechanisms are able to induce higher levels of mutagenesis to ensure inactivation of the virus. However, for those A3s that induce low levels of mutations and restrict viruses to a lesser degree, having secondary mechanisms of restriction can facilitate complete virus inactivation. The secondary mechanisms occur through deamination-independent modes of restriction. While A3G has weak deamination-independent activity, other A3 enzymes, such as A3F and A3H, have potent deamination-independent activity [74, 154, 169, 286]. A3C restricts LINE-1 retroelements by deamination-independent mechanisms [167, 214]. Therefore A3C, and other A3s, may utilize deamination-independent mechanisms in order to supplement the observed weak restriction activity (e.g., Figure 5.8). Deamination mediated mutagenesis, as well as deamination-independent restriction likely act in concert to increase the chances of complete inactivation of HIV [74].

Despite many years of studying deamination –dependent and –independent mechanisms of A3 enzymes, the field is still divided on their contributions. Some studies have shown that this deamination-independent mode of restriction may be able to decrease the accumulation of reverse transcripts by up to 90%, while others suggest that catalytic activity of A3 enzymes cannot be removed from its restriction capacity [161, 287]. Our lab has recently shown that A3G and A3F are able to reduce the formation of late reverse transcript products (LRTs) independent of deamination activity [74]. While the evidence suggests that a deamination-independent mode of restriction exists, the exact mechanism of this restriction has yet to be conclusively determined. There is evidence that this non-catalytic mode of restriction occurs through direct or RNA-mediated interactions that inhibit nucleocapsid mediated annealing or inhibit reverse transcriptase mediated DNA synthesis [71, 73, 155, 158]. Therefore, further investigation into the mechanism of deamination-independent restriction is warranted for hA3C S188I and other A3 enzymes.

8.0 Determining the mechanism of inhibition of HIV reverse transcriptase by APOBEC3s

8.1 Abstract

The APOBEC3 (A3) family is well studied for their ability to restrict HIV, in the absence of HIV Vif, through mutagenesis of the virus by cytosine deamination of single-stranded viral DNA. However, the APOBEC3s that are relevant to HIV restriction, A3G, A3F, A3H hap II and the recently characterized A3C S188I, may also possess deamination-independent modes of restriction in order to ensure complete viral inactivation. Previous observations indicate that A3G and A3F may competitively bind the RNA template and act as a roadblock to reverse transcriptase (RT) polymerization. Here we demonstrate that the roadblock imposed by A3s delays the progression of RT along the template and forces RT to undergo higher levels of template switching. We examined HIV RT template switching using an *in vitro* primer extension assay in which template switching results in a longer DNA product than without template switching. A3F bound the RNA template with higher affinity than the other A3s tested and promoted RT to form almost entirely longer template products, indicating higher levels of template switching, in our *in vitro* assays. When we created a mutant of A3G, A3G NPM, that increased the affinity of A3G for the template, we were able to observe an increase in longer product formation compared to wild type A3G, confirming that the roadblock model was promoting the template switch. Interestingly, we also observed that the A3s, other than A3F, were able to decrease the fidelity of RT *in vitro*, which would increase the mutational load in the virion. Altogether, these additional modes of deamination independent restriction of A3s may work in concert with the deamination dependent mechanisms to ensure complete inactivation of the virus. Increased template switching may cause higher frequencies of deletions or insertions, and decreased polymerase fidelity would lead to higher mutation rates. Both of these activities would promote increased likelihood of producing an inactive HIV provirus.

8.2 Introduction

The APOBEC3 (A3) deoxycytidine deaminases are a family of host restriction factors for retroviruses, retrotransposons, and endogenous retroviruses. The A3 enzymes are most well characterized for restricting replication of the human immunodeficiency virus-1 (HIV-1, referred to as HIV). For A3-mediated HIV restriction to occur, the A3 enzymes must become virion

encapsidated. The HIV-1 protein Vif induces the polyubiquitination and degradation of A3 enzymes to prevent their encapsidation, but degradation of A3 enzymes is not complete and some A3s can become encapsidated at a low amount into virions in the presence of Vif. HIV restriction is primarily mediated by the deamination of cytosine (C) to uracil (U) on single-stranded (ss) DNA during synthesis of (-)HIV-1 DNA after virus entry into a target cell [288, 289]. These deaminations lead to guanine (G) to adenine (A) hypermutation of the virus genome when uracil is used as a template during HIV-1 reverse transcription [67-69]. The G to A hypermutation ultimately leads to inactivation of the provirus [67-69]. Of the seven human A3 deaminases, A3G, A3F, A3D, and A3H haplotype II (hap II) are able to restrict replication of HIV to varying degrees [2]. In addition, a rare human polymorphism of A3C (S188I) has HIV-1 restriction activity, while the more common S188 polymorphism of A3C does not.

Although viral inactivation mediated by the A3 enzymes primarily occurs through cytosine deamination, deamination-independent mechanisms have been characterized. During reverse transcription, the RNaseH domain of reverse transcriptase degrades the RNA template, which leaves regions of single-stranded (ss) DNA available for deamination by A3 [289, 290]. As these ssDNA regions are only available for a limited amount of time, A3 enzymes must possess efficient ssDNA scanning mechanisms, and scan the template processively in order to locate target cytosines [111, 291]. The processive action is mediated through facilitated diffusion, which includes both one-dimensional sliding and three-dimensional translocations such as jumping and intersegmental transfer [104, 105]. A3 enzymes that scan processively and are able to use a combination of one-dimensional and three-dimensional movements and induce higher mutational loads than non-processive A3s [108-110]. However, certain A3s are able to additionally inhibit HIV through deamination-independent mechanisms that act in concert with the deamination activity [71-74, 155, 156, 159, 292]. A3 enzymes, such as A3G, A3F, A3H and A3C are able to restrict both HIV-1 and the endogenous retroelement LINE-1 through deamination-independent modes [72, 73, 155, 156, 159, 169, 214, 286, 293-295]. A3G and A3F delay the initiation of primer extension by HIV-1 reverse transcriptase (RT) to varying degrees by binding to the RNA template and acting as a physical block to the scanning mechanism of RT. A3G also binds directly with RT and disrupts DNA synthesis [71, 74].

HIV reverse transcription requires both RT and nucleocapsid (NC). NC is a nucleic acid chaperone that promotes annealing of the host tRNA_{Lys3} to the primer-binding site (PBS), which

primes RT initiation of the minus strand (-) strong-stop DNA [61, 63]. NC is also required to facilitate the first-strand transfer of the minus strand (-) strong-stop DNA to the 3' end of the genome in order for (-)DNA synthesis to proceed further [62, 296]. During reverse transcription, RT undergoes two separate template-switching events that are necessary to complete synthesis of (+) DNA [62, 296, 297]. It has been suggested that the low template affinity and low processivity of the RT is advantageous because it allows for these template-switching events to occur that are essential to producing the proviral DNA [297]. However, RT can also switch templates at other random locations in a homology directed fashion and this can create diverse retroviral populations if the switch occurs between distinct templates (intermolecular switch) [298-301] or lead to deletions of genomic regions if the switch occurs on the same RNA molecule (intramolecular switch) [302, 303].

These inter- and intra- molecular template switches have been described to occur through either a dynamic copy choice or dynamic forced copy choice model. Both HIV and Murine Leukemia Virus (MLV) RTs utilize a dynamic copy choice model of recombination which is dependent on the equilibrium between the polymerase and RNaseH activity of RT [304-306]. A slow polymerase undergoes higher rates of template switching, whereas a slow RNase H domain causes lower rates of template switching compared to the wild-type equilibrium [304]. RT can also undergo what is known as a dynamic forced copy choice model, in which a roadblock to polymerization, or a nick in the template strand forces the RT to switch templates. Both models are supported by *in vitro* and cellular studies mutations and particularly strong evidence for the dynamic copy choice model comes from RT mutants that have disrupted dynamics between polymerization and RNaseH degradation that can promote RT to switch templates to varying degrees [304].

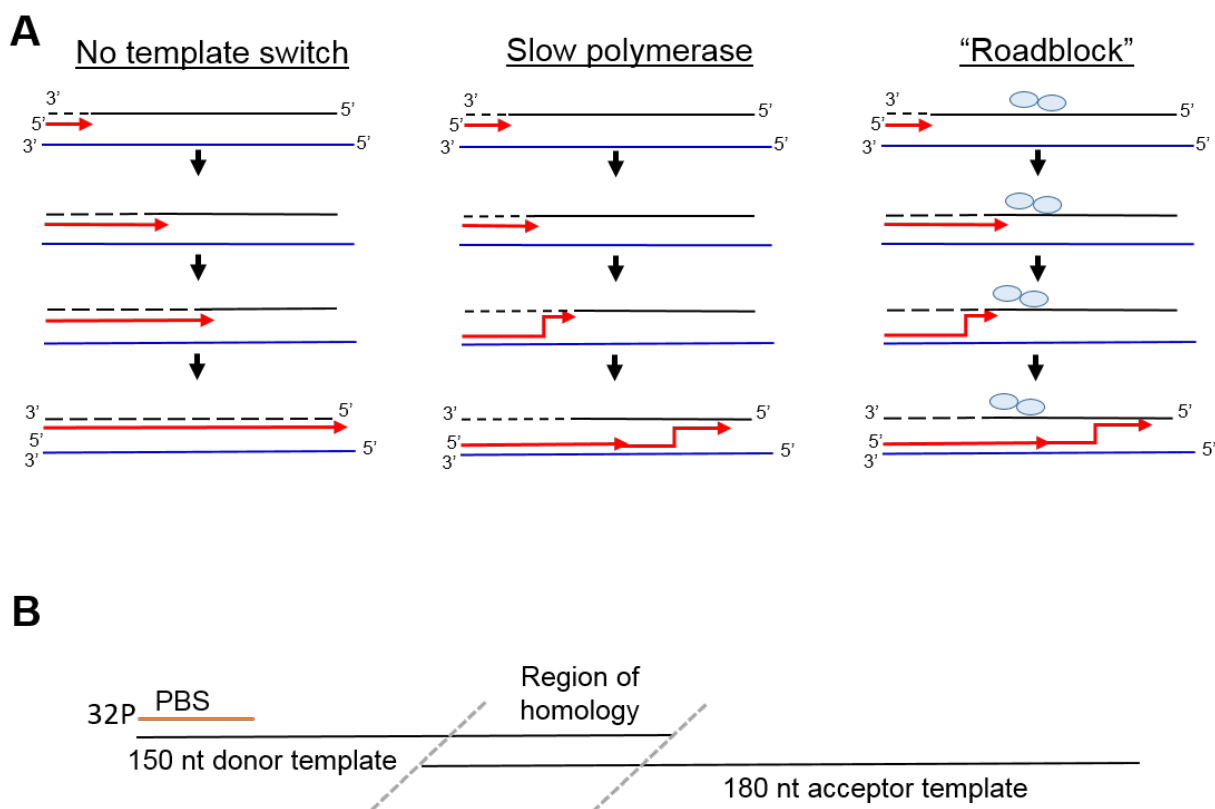
RT does not have the ability to proofread its insertions due to a lack of 3' to 5' exonuclease activity [290]. Thus, HIV RT, like most other viral polymerases has a higher misincorporation frequency than many other non-viral replicative polymerases [307, 308]. Misincorporation of bases by RT leads to rapid virus evolution, escape from the host immune system, and emergence of viruses with resistance to HIV therapeutics [309]. Several studies have characterized the insertion fidelity of HIV RT, using both *in vitro* and cellular complementation assays, with the majority examining the role of nucleoside analogs on replication fidelity [47, 310-313]. However, the fidelity of the enzyme *in vitro* is significantly lower than that observed in cell based assays

[310], likely due to reaction components and experimental conditions. Therefore, these experimental variables leave numerous questions about what affects RT insertion fidelity. Similarly, how restriction factors such as A3 enzymes may affect RT insertion fidelity has not been examined.

We hypothesized that if an A3 inhibited RT-mediated DNA polymerization either through a “roadblock” or direct interaction, that it may also facilitate the dissociation of RT from the template and disrupt the equilibrium between polymerization and RNaseH degradation, thus affecting template switching. Although not known why, the frequency of template switching can influence the insertion fidelity of RT. To test this hypothesis we used A3s with the strongest HIV restriction ability, A3G, A3F, A3H hap II, and a recently characterized A3C S188I variant. Our data demonstrate that the deamination-independent functions of A3 enzymes can decrease the insertion fidelity of RT, and if they bind the template with high enough affinity, can promote template switching. Both of these events could increase the probability of generating a mutated and nonfunctional virus independent of cytosine deamination.

8.3 Results

8.3.1 Modulation of HIV-1 RT template switching by APOBEC3 enzymes. We hypothesized that a roadblock imposed by A3s could slow the polymerization down and alter the equilibrium between polymerization and RNaseH activity, or like a nick in the template, force RT to switch templates (Figure 8.1A). The proposed dynamic copy choice model of template switching suggests that the equilibrium between the polymerization of the template and the RNaseH mediated degradation of the RNA behind the RT affects the rate of template switching observed [304, 306]. In order to test this hypothesis, we established an *in vitro* RT template-switching model, similar to previously established experimental systems [305, 314-316]. We used the PBS region to establish a “donor region” that resembled the PBS region of the HIV genome and extends 150 nucleotides to the trans-activation response (TAR) element. (Figure 8.1B). This region is sensitive to delays in initiation by A3 due to the inefficient initiation of DNA synthesis of RT from an RNA primer [317, 318]. We also used an “acceptor” template that does not contain the PBS, but possesses a 100 nucleotide region of homology with the donor template as well as unique bases that allow for a distinct acceptor template (Figure 8.1B). If template switching occurs between the donor and the acceptor then the resulting product should be up to 30 nt longer than that of the donor product alone.



*Figure 8.1 Current models and schematic of template switching protocol. (A) Dynamic copy choice model (slow polymerase) and forced copy choice model (roadblock) adapted from Hwang *et al.* 2001 [304] and Anderson *et al.* 1998 [319]. The black and blue lines represent RNA templates whereas the red arrow indicates the newly synthesized DNA template. The hashed line behind the polymerase indicates RNase H degradation of the RNA, and the efficiency of this degradation is depicted in the size of the dashes (short is fast RNaseH degradation and long is slow RNaseH degradation). In the presence of a slow polymerase or a road block, the newly synthesized template can base pair with the second template (blue) and promote RT to template switch. (B) Schematic of template switching assay. A radiolabeled primer is annealed to the donor RNA template, which creates the primer/template. An acceptor template is included in the reaction that lacks the primer-binding site but has a region of homology with the donor template. Template switching from donor to acceptor template will generate a longer acceptor product after resolution on by denaturing PAGE.*

For our template switching assays we aimed to use physiologically relevant concentrations of each component as they would be found in a virion. Thus, based on evidence of approximately 100 molecules of RT in a virion and 2 RNA genomes in a virion, we used 400 nM of RT, 10 nM of donor template, and saturating amounts of nucleocapsid (NC) (1NC/7nt) in the *in vitro* assays [71, 74]. In addition, we utilized a 40:1 acceptor to donor ratio to promote template switching, similar to previous reports that showed an excess of acceptor template is required to observe template switching *in vitro* [315, 316, 320]. We then examined the ability of RT to undergo template switching in the presence of varying amounts of A3 and primer/template (p/t). The A3:p/t ratios used were 0:1, 4:1, 8:1, 32:1. Interestingly, we found that in the presence of increasing amounts of A3G, there was a statistically insignificant but reproducible reduction in the amount of acceptor template, indicating less template switching, compared to the absence of A3 (Figure 8.2A, compare 0:1 and 32:1 lanes) with the amount of acceptor template decreasing from 21% to 16% (Figure 8.2E). However, in the presence of increasing amounts of A3F we observed the opposite trend. The presence of high amounts of A3F promoted the formation of the acceptor product only, indicating that the template switching efficiency was increased from 27% to 90% (A3F:p/t of 32:1, Figure 8.2B and Figure 8.2E). We also examined two other A3s known to exert a deamination-independent mode of restriction, A3H and A3C I188. There was no observable effect on template switching by these two enzymes, indicating that they are not able to modulate this function of RT (Figure 8.2C-D, and Figure 8.2E).

8.3.2 RNA template binding by A3F promotes HIV RT template switching. To determine why A3F was able to promote template switching and the other A3s were not, we characterized how the A3 enzymes interacted with the primer/template complex (p/t). Using rotational anisotropy we determined the strength of the A3-p/t interaction. A3F bound the p/t in the nanomolar range (K_d of 0.06 μ M, Figure 8.3B, Table 8.1). A3G also bound the p/t less well than previously observed for ssDNA and 13-fold less well than A3F binding the same substrate (K_d of 0.82 μ M, Figure 8.3A, Table 8.1) [108]. This suggests that the high affinity of A3F with the p/t induces more of a roadblock to RT than A3G, and promotes increased template switching due to

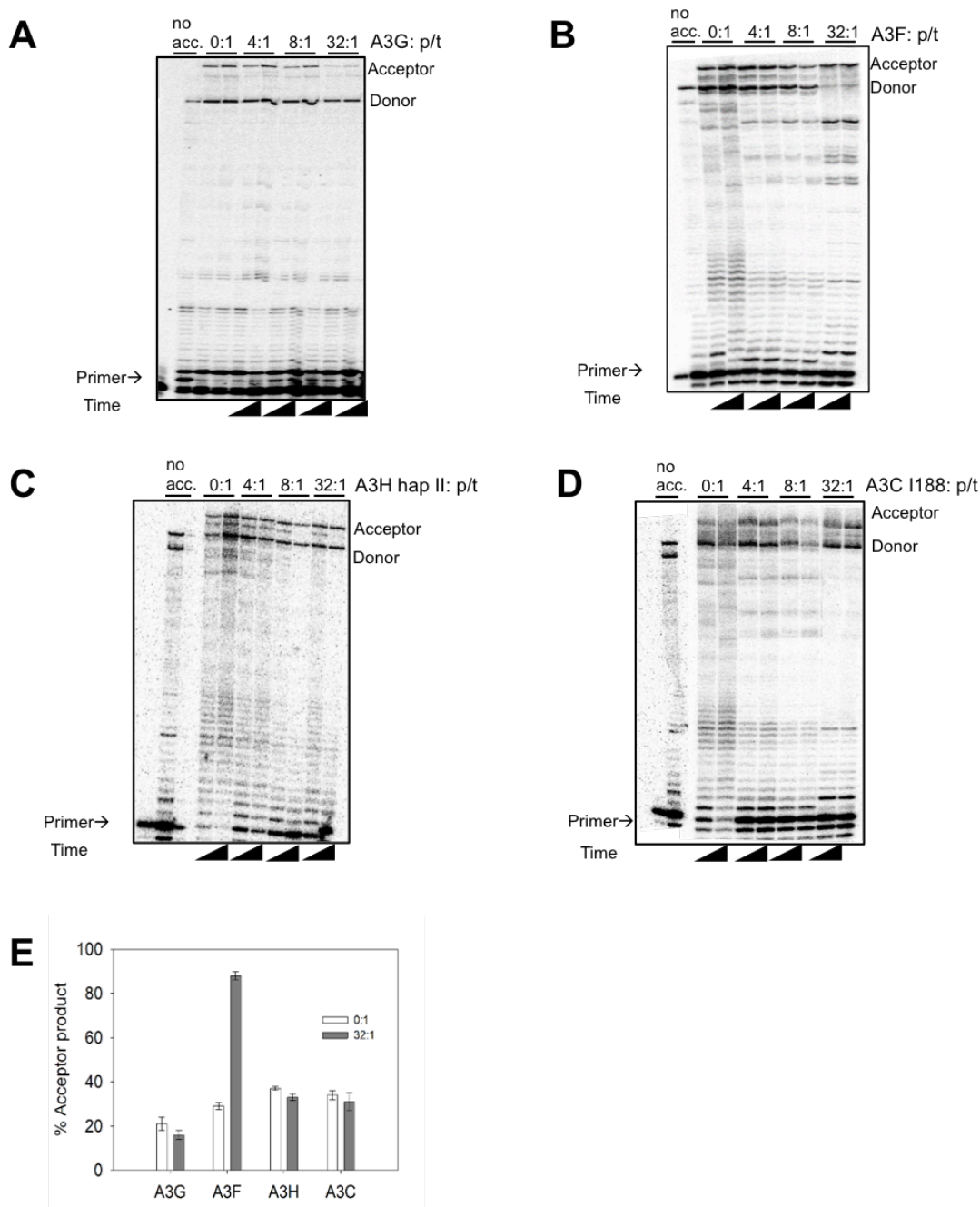


Figure 8.2. Effect of A3s on the template switching efficiency of RT. Sketch of the annealed 18-nt DNA primer to the 150 nt RNA template (p/t) in Figure 8.1B. Complete extension of the p/t (10 nM) results in a 150 nt donor product, whereas in the presence of acceptor, template switching results in the production of a longer acceptor product. (A-D) Extension of the p/t by 400 nM of RT in the absence (0:1) or presence (4:1, 8:1, 32:1) of (A) A3G, (B) A3F, (C) A3H hap II, or (D) A3C I188. Reactions were sampled at 60 or 90 minutes. (E) Bar graph showing the quantified percentage of acceptor template at 0:1 and 32:1 A3 to p/t.

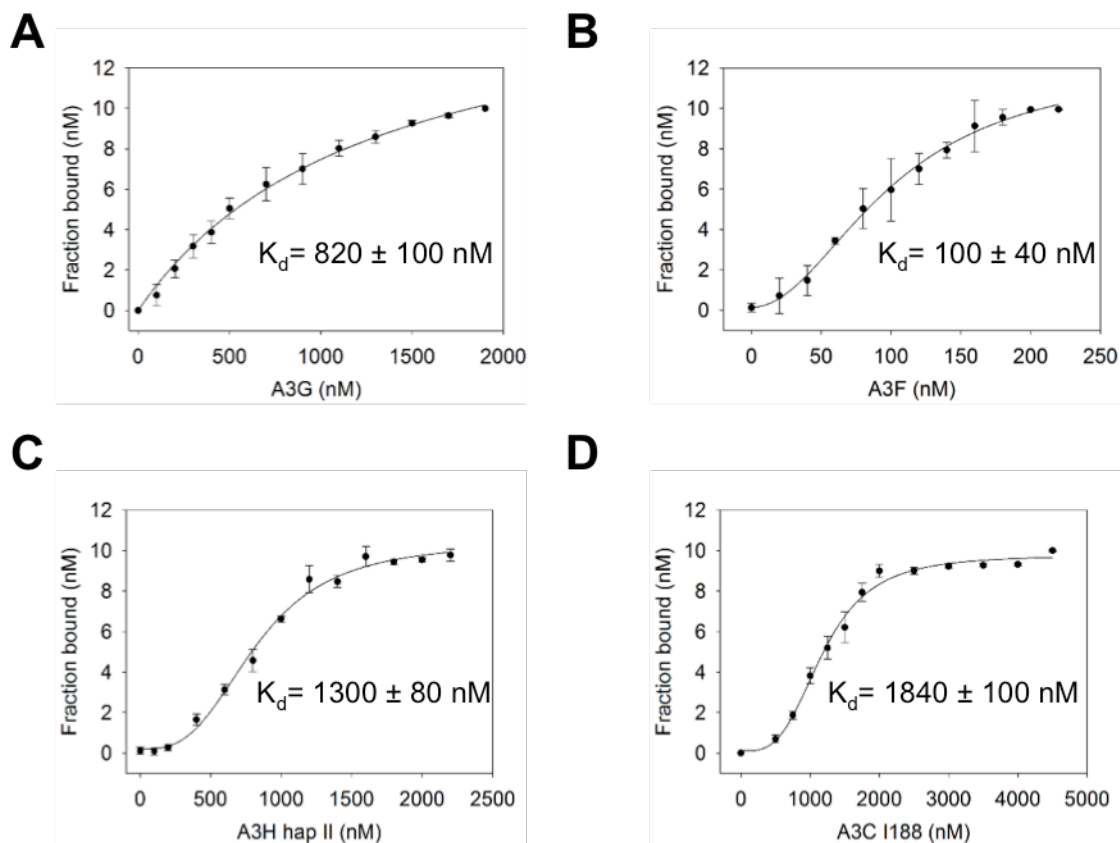


Figure 8.3. Interaction of A3 with the DNA/RNA primer/template. The apparent dissociation constant (K_d) was determined by steady-state rotational anisotropy by titrating increasing amounts of (A) A3G, (B) A3F, (C) A3H hap II or (D) A3C I188 into a buffered solution containing 10 nM of fluorescently labeled p/t. The K_d is shown on the graph.

Table 8.1 Apparent dissociation constants for A3s binding the p/t.

Enzyme	Binding primer/template (μM)
A3G	0.82 ± 0.10
A3F	0.10 ± 0.04
A3G NPM	0.20 ± 0.02
A3F NAM	0.23 ± 0.05
A3C I188	1.84 ± 0.10
A3H hap II	1.30 ± 0.08
HIV RT	0.50 ± 0.06

a slow polymerase. Consistent with this finding A3H and A3C both bound the p/t less well than A3G, and significantly less well than A3F, suggesting that the weaker interaction of the enzymes with the substrate does not allow for them to induce template-switching (Figure 8.3C-D, Table 8.1). To further test this rationale, we repeated the binding assays using a previously described mutant of A3G having an enhanced ssDNA interaction and a converse mutant in A3F, that interacts weakly with ssDNA. The strongly ssDNA binding A3G (A3G NPM) mutant has an 195NPM197 insertion in the connection loop between the NTD and CTD [108] that was derived from the amino acid sequence of A3F, and has been shown to impair A3F sliding movements on ssDNA, due to the enhanced ssDNA binding. The A3F mutant A3F NAM has a mutation at the rigid proline in the linker region, 190NAM192, and binds ssDNA with less affinity than wild-type A3F [108]. We confirmed that these mutations changed the binding affinity as expected. A3GNPM bound the p/t 4-fold more tightly than wild-type A3G (K_d of 0.20 μ M, Figure 8.4A, Table 8.1) and A3F NAM bound the p/t 2-fold less tightly than wild-type A3F (K_d of 0.23 μ M, Figure 8.4B and Table 8.1). To test whether that strengthening the association of A3G with the p/t affects RT template switching we assessed them in our *in vitro* template switching assay. In support of our hypothesis, A3G NPM increased the production of template switched acceptor products compared to wild-type A3G, but 1.6-fold less than efficiently than A3F (Figure 8.4C and Figure 8.2A). This can be attributed to the weaker binding to the p/t of A3G NPM compared to A3F (Table 8.1). Similarly, we observed 1.4-fold less acceptor template extension in the presence of A3F NAM compared to A3F, as the binding affinity for the p/t is 2-fold less than wild-type A3F (Figure 8.4B and Figure 8.4D). However, we observed similar levels of template switching promotion in the presence of both A3G NPM and A3F NAM, which correlates with the similar affinity to the p/t by the mutants. Therefore, the higher affinity for ssDNA correlates with greater ability to promote template switching.

8.3.3 Alteration of HIV RT insertion fidelity by APOBEC3. We also determined if the presence of the A3s on the RNA template could alter the insertion fidelity of the RT. It has been demonstrated previously that polymerase mutants that exhibit a higher frequency of template switching also alter the misincorporation frequencies of the polymerase, though it remains unclear as to how template switching may affect the fidelity of the polymerase [321]. Therefore, the data obtained from the template switching assays would predict that A3 enzymes that bind the

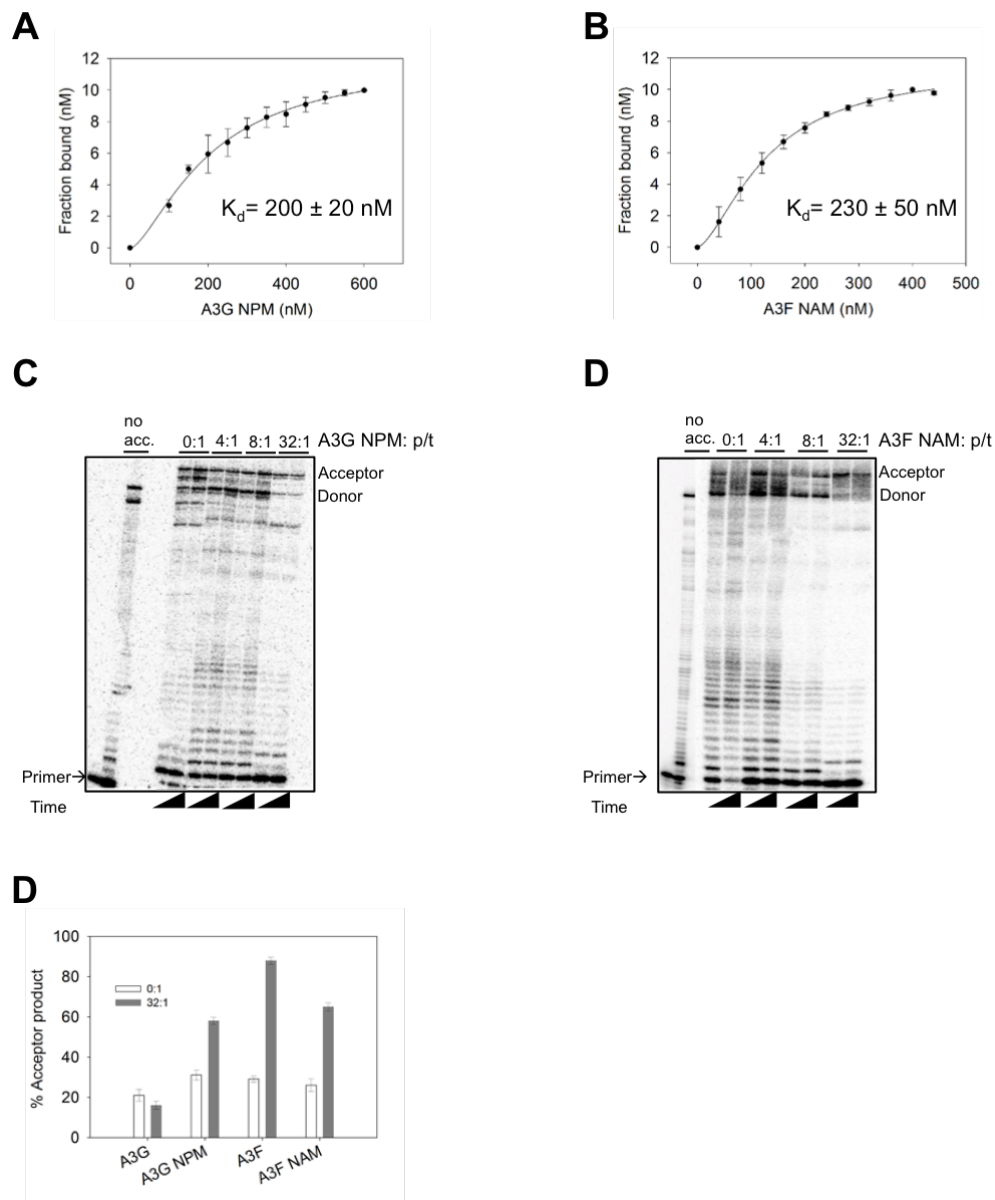


Figure 8.4. Interaction of A3G NPM and A3F NPM with p/t and the effect on template switching. (A-B) The apparent dissociation constant (K_d) was determined by steady-state rotational anisotropy by titrating increasing amounts of (A) A3G NPM or (B) A3F NAM with 10 nM of fluorescently labeled p/t. The K_d is shown on the graph. (C-D) Complete extension of the 10 nM p/t results in a 150 nt donor product, whereas in the presence of acceptor, template switching results in the production of a longer acceptor product. Extension of the p/t by 400 nM of RT in the absence (0:1) or presence (4:1, 8:1, 32:1) of (A) A3G NPM, or (B) A3F NAM. Reactions were sampled at 60 and 90 minutes. (E) Bar graph showing the quantified percent acceptor template at 0:1 and 32:1 ratios of A3 to p/t. Hatched lines show percent acceptor for A3G and A3F at 32:1 for reference.

RNA template with higher affinity and promote increased RT template switching may also exhibit increased insertion fidelity of RT. We utilized steady-state polymerase insertion kinetic analyses to examine the incorporation of either a right dNTP or a wrong dNTP into the template. We assayed RT insertion fidelity in the context of a standing start (mismatched base immediately follows primer) and a running start (mismatched base is 1 nt downstream of primer) and used a radiolabeled template in order to detect the incorporation of bases [322-324]. The templates were resolved on 16% PAGE gels and the integrated intensities of the bands quantified. For standing start experiments the initial velocity (v) is calculated as the percent p/t extended per minute. For running start experiments the initial velocity is calculated as the insertion of base 2 (mismatch) over the insertion of base 1 (correct base) and is represented by I_2/I_1 , where I is the percent insertion [322]. Data were analyzed to determine Michaelis-Menten kinetics and we used a rectangular hyperbola to be fit to the curve (initial velocity versus dNTP concentration) in order to determine the V_{max} and K_m of both right (R) and wrong (W) base incorporation [322]. The insertion fidelity is calculated as follows:

$$F_{ins} = (V_{max}/K_m)W / (V_{max}/K_m)R$$

The schematics of the fidelity assays are included in Figure 8.5 and Figure 8.6. The *in vitro* insertion fidelity of HIV RT is 1.0×10^{-4} [325, 326]. Similar insertion RT fidelity kinetics of 4.4×10^{-4} and 2.6×10^{-4} , were measured for standing and running start experiments respectively (data not shown). Notably, RT fidelity values obtained from *in vitro* studies are typically 3- to 4- fold higher than cell based assays. As a result, the data presented here is a relative measure of fidelity in the absence or presence of an A3. The *in vitro* fidelity assays were performed in the presence of a heparin trap, which does not allow for RT to re-associate with the p/t once dissociated, to ensure that only single enzyme and p/t encounters were considered when calculating the insertion fidelity.

In order to determine the effect of an A3 on the fidelity of RT, we examined the incorporation of a dNTP in the presence of different A3 enzymes. We observed that in the presence of weakly p/t associating A3s, namely A3G, A3H and A3C I188, the insertion fidelity of RT was decreased. For example, A3G decreased the fidelity in standing start experiments 6-fold (Figure 8.5A, $f_{ins} = 3.0 \times 10^{-3}$ and Table 8.2) and 30-fold in running start experiments (Figure 8.6A, $f_{ins} = 7.5 \times 10^{-3}$ and Table 8.2). Similarly, A3H and A3C I188 induced decreases in the insertion fidelity in both standing and running start experiments. A3H decreased the insertion fidelity to 3.2×10^{-3} (7-fold)

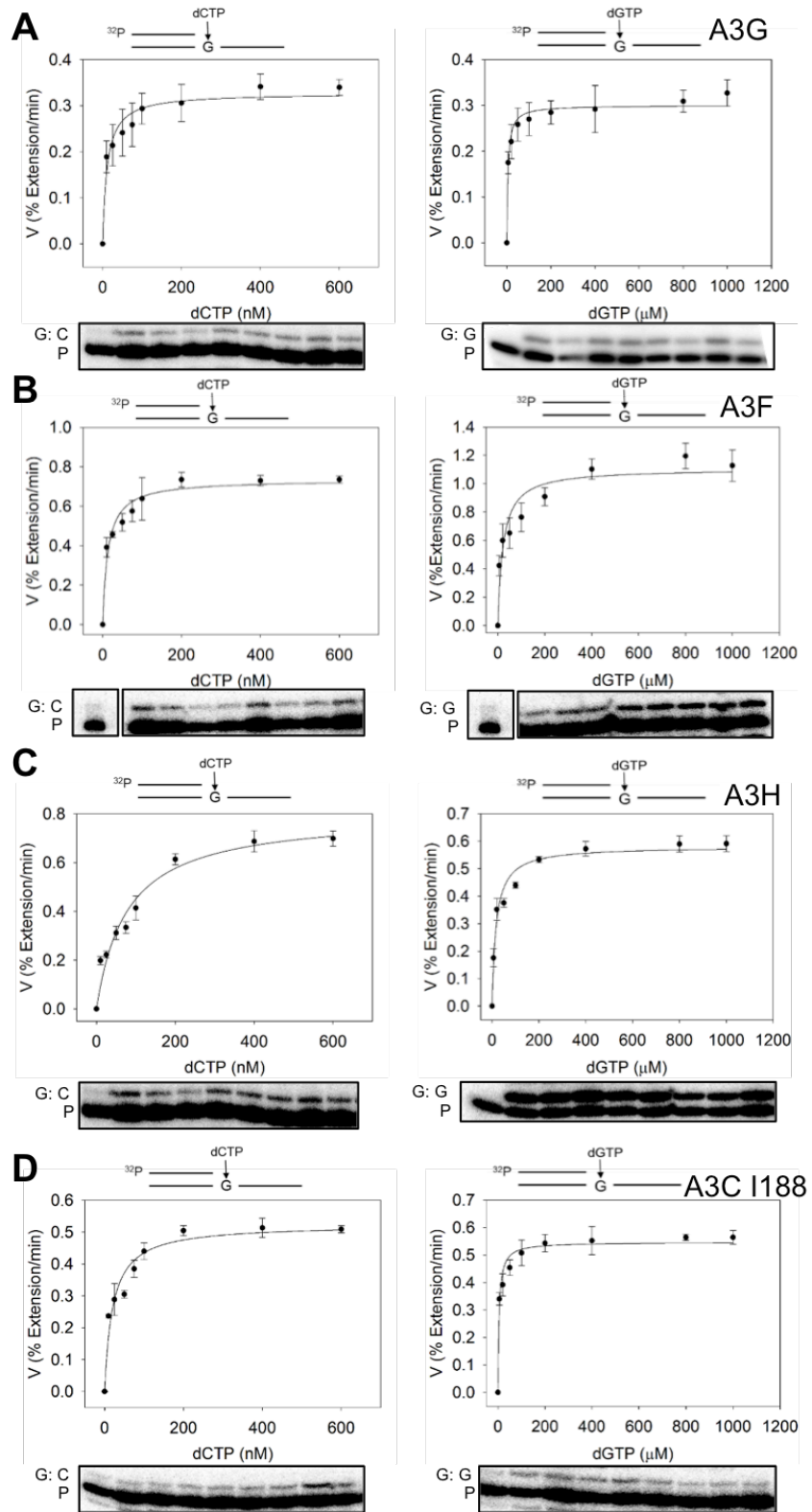


Figure 8.5. Steady state standing start kinetics of RT insertion fidelity. (A-D) Standing-start velocities for extension of primer (P) by incorporation of a “right” C (G:C, left panel) or “wrong” G (G:G, right panel) in the presence of (A) A3G, (B) A3F, (C) A3H and (D) A3C I188 are shown as a function of corresponding dNTP concentration. A rectangular hyperbola fit to the data is used to determine V_{\max}/K_m for right and wrong incorporations. The standing-start misinsertion frequencies, $f_{\text{ins}} = (V_{\max}/K_m)_W/(V_{\max}/K_m)_R$ are shown in Table 7.2.

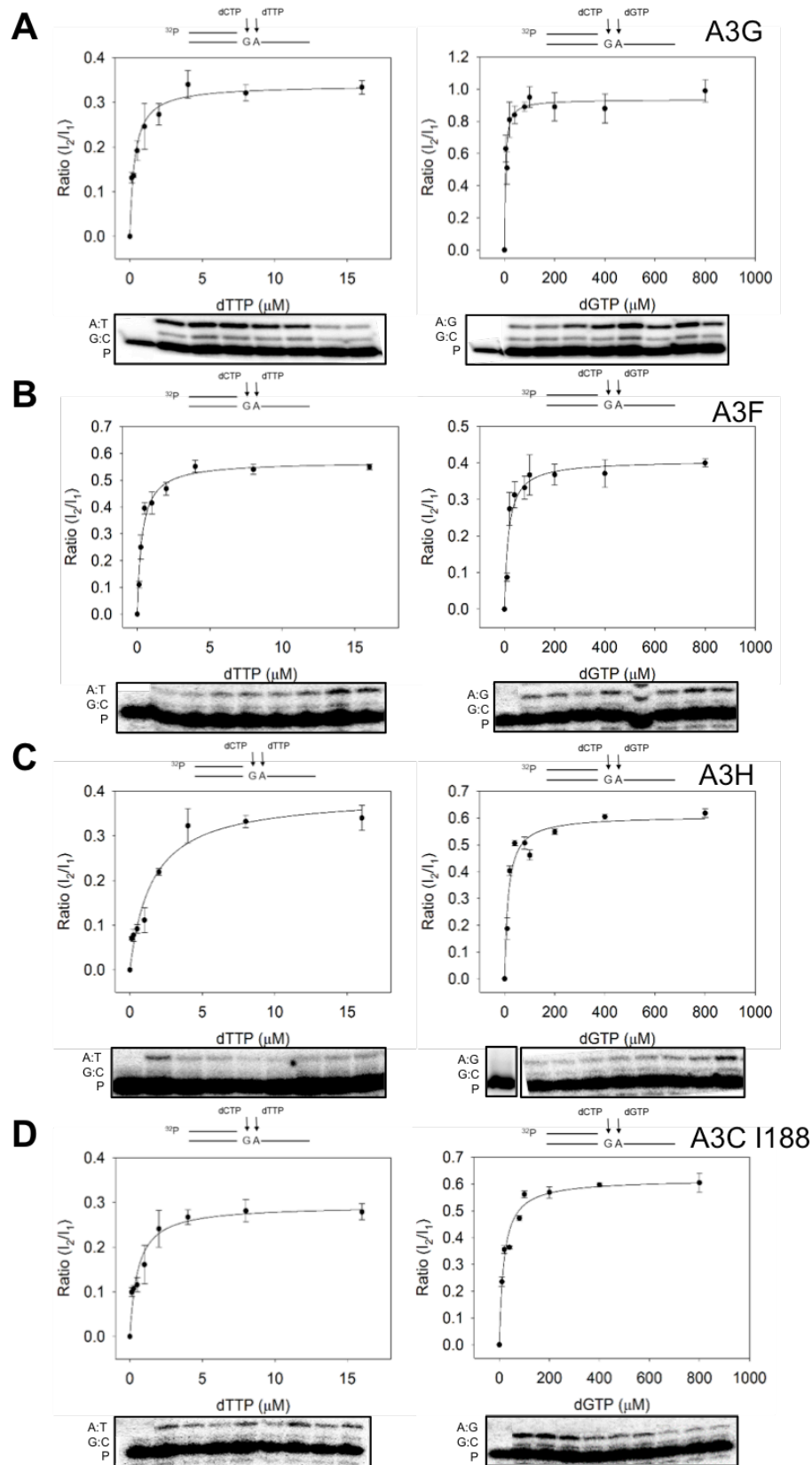


Figure 8.6. Steady state running start kinetics of RT insertion fidelity. (A-D) Running start kinetics are plotted as the Ratio (I_2/I_1) as a function of dNTP concentration in the presence of 200 nM of (A) A3G, (B) A3F, (C) A3H and (D) A3C I188. The ratio I_2/I_1 is the velocity of extended primer where I_2 is the band intensity at site 2 (A:T or A:G) and I_1 is the band intensity at site 1 (G:C). The running-start dCTP is kept at a constant concentration (50 μ M) when measuring right incorporation of T opposite template A (A:T, left panel) and when measuring wrong incorporation of G opposite template A (A:G, right panel). A rectangular hyperbola fit to the data is used to determine V_{\max}/K_m for right and wrong incorporations (see text for details). The running-start misinsertion frequencies, $f_{\text{ins}} = (V_{\max}/K_m)_W/(V_{\max}/K_m)_R$ are shown in Table 7.2

Table 8.2. HIV RT insertion fidelity in the presence of A3s.

Enzyme	HIV RT fidelity (standing start)	HIV RT fidelity (running start)
No A3	4.4×10^{-4}	2.6×10^{-4}
A3G	3.0×10^{-3}	7.5×10^{-3}
A3F	4.1×10^{-4}	3.9×10^{-4}
A3C I188	4.1×10^{-3}	5.1×10^{-3}
A3H hap II	3.2×10^{-3}	9.9×10^{-3}

and 9.9×10^{-3} (38-fold) and A3C I188 decreased the insertion fidelity to 4.1×10^{-3} (9-fold) and 5.1×10^{-3} (20-fold) on standing and running start templates, respectively (Figure 8.5C-D, Figure 8.6B-D, and Table 8.2). Interestingly, in the presence of A3F which binds the template with high affinity, the insertion fidelity remained essentially unchanged for standing and running start p/t unlike the other A3s (Figure 8.5B, $f_{\text{ins}}=4.1 \times 10^{-4}$, Figure 8.6B, $f_{\text{ins}}=3.9 \times 10^{-4}$ and Table 8.2). Since A3F promoted almost complete template switching in our *in vitro* assay (Figure 8.2B and Figure 8.2F), it can be reasoned that this observed increase in template switching frequency in the presence of A3F prevented any effect on the incorporation frequency of RT, as shown in previous reports. This is in contrast to the effects of A3G, A3H, and A3C I188. Therefore, the A3 enzymes that have lower affinity for the RNA template decreased the fidelity of RT. In contrast, A3F that binds the template with high affinity forced the RT to switch templates to avoid the roadblock and had no effect the insertion fidelity.

8.4 Discussion

A3 enzymes are able to restrict viral replication through deamination-independent mechanisms in addition to the well-characterized deamination-dependent mechanism [73, 155, 156, 169, 214, 286, 294, 295]. A3G and A3F delay the initiation of primer extension, which leads to production of less full-length extension products [71, 74]. A3G and A3F were also able to inhibit the formation of late reverse transcriptase products (LRTs) through delay of extension from the primer [74]. A3G has been found to interact with RT directly, which also leads to less full-length extension products being formed [71]. We sought to determine if the disruption of RT polymerase activity affects template switching and polymerase insertion fidelity.

A3G and A3F have been found to act as a scanning block and delay RT polymerization [317]. We further extended the downstream effect of this roadblock model by examining the effect on RT template switching. Interestingly, A3F uniquely promoted RT template switching to 90% (Figure 8.2B) and all but A3F, could decrease the insertion fidelity of RT (Figure 8.5, Figure 8.6, and Table 8.2). The data indicated that the high affinity binding of A3F to the p/t promoted template switching. For the other A3s tested, their inability to promote template switching was consistent with an ability to decrease the insertion fidelity.

The data obtained here are in agreement with previous reports studying RT template switching, but in the absence of A3s. Studies from the Pathak lab have examined the dynamics of template switching, including mutants of RT. They observed that when the polymerase activity

was slowed through either single or double mutations within the polymerase domain, that the frequency of template switching was increased [304]. This observed increase is due to increased RNaseH cleavage behind the slow polymerase, which allows for increased base pairing with the acceptor template (Figure 8.1A, slow polymerase). Conversely, if there were mutations that affected the RNase H domain by decreasing RNaseH cleavage, then less template was available for template switching and the rate of switching was reduced [304]. Alternatively, the forced copy choice model suggests that a roadblock to polymerization, such as a nick in the RNA template or a competitive binder, stalls the progression of the polymerase and forces template switching (Figure 8.1A) [319]. We observed that the presence of the A3 bound to the template is able to copy the effect of a slow polymerase mutant as well as a roadblock on the template. This forces RT to switch templates, which it is able to do at high levels due to the increased RNase H degradation behind the stalled polymerase.

A previous report identified that increased template switching, correlates with reduced misincorporation of bases by RT [321]. The exact mechanism by which template switching reduces the misincorporation of bases remains to be elucidated. However, this model fits with our results, as all A3s except A3F, decreased template switching and decrease the insertion fidelity of RT, and A3F increased template switching and reduced misincorporation. Thus, A3s can modulate RT fidelity. To avoid the roadblock imposed by these A3s, RT must be able to circumvent the bound A3s on the RNA template either through dissociation and reassociation downstream or through template switching if a region of homology is present. Although our *in vitro* experiments used an excess of acceptor template to promote template switching, in virions, the two different viral genomes encapsidated in the confines of the virus particle, are more likely to recombine through template switching despite their equimolar ratio. This could lead to one of two outcomes: if intermolecular template switching occurs, this may promote recombination and virus evolution which would be detrimental to the host, or intramolecular switching may create regions of insertions and deletions that could inactivate the virus [302-304]. The presence of A3s on the RNA template may also increase the misincorporation frequency of RT. The increase in error rate may be a mechanism to ensure virus inactivation in the presence of Vif where the number of A3s encapsidated is low and the deamination induced mutations alone may not be enough to inactivate the virus. It is interesting that A3G was the only A3 that could decrease

template switching. Although not tested for other A3s, A3G can interact with RT directly [327], and this may enable A3G to uniquely decrease RT template switching.

Overall, this study has found that A3 enzymes are not only able to decrease the initiation of primer extension, but are able to modulate RT processes of template switching and affect insertion fidelity. While these deamination-independent mechanisms alone are not likely to cause similar levels of viral inhibition observed for cytosine deamination by A3s, and therefore not as potent at restricting HIV as the deamination-dependent mechanisms, the additive effect of inhibiting RT processes and hypermutating viral DNA may ensure complete inactivation of HIV. The delay in polymerization is thought to allow for a longer time period for A3s to access and deaminate the viral DNA, and this roadblock also forces RT to compensate for the block in polymerization through template switching. Similarly, the competitive binding of A3s prevents proper sampling of the dNTP pool, increasing the frequency of errors. These effects on RT, in addition to deaminations of the viral DNA would cause higher levels of mutation, perhaps to better ensure viral inactivation.

8.5 Materials and Methods

8.5.1 *Synthesis of RNA templates.* For the PBS RNA donor template, a 150 nucleotide (nt) segment near the 5'-end of the HIV-1 genome (nt 521–676) encompassing the PBS and upstream region to the TAR site was PCR amplified. For the PBS RNA acceptor template, a 180 nt segment of the HIV-1 genome (451-635 nt) was PCR amplified. The PCR amplicons were cloned into the BamHI and EcoRI sites of the pSP72 vector (Promega). Sequences were amplified from the HIV-1 clone 93th253.3 (GenBank accession number U51189) obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health; p93th253.3 was from Dr. Feng Gao and Dr. Beatrice Hahn [328]. The sequences were verified and then used for synthesis of RNA template by linearizing the vector at the BamHI site and using it as a T7 RNA polymerase substrate according to manufacturer's instructions for the Ambion Megascript kit. All primers and templates used are listed in Table 7.3

8.5.2 *Protein expression and purification.* Recombinant baculovirus for expression of GST-tagged NC, A3G, A3F, A3H, A3C I188, A3G NPM and A3F NAM was constructed as described previously [99, 107-111]. Sf9 cells were infected with recombinant A3 or NC at an MOI 1 for A3G and NC, MOI of 2 for A3F, MOI of 5 for A3C I188 and an MOI of 20 for A3H

hap II and were harvested after 72 h. Cells were lysed in the presence of RNaseA and purified as described previously to obtain proteins cleaved from the GST tag. NC, A3G, A3G NPM and A3H hap II were subjected to on-column cleavage from the GST tag with thrombin (GE Healthcare) at 21 °C for 18 h in thrombin digestion buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM DTT). A3F, A3C I188 and A3F NAM were eluted with the GST in elution buffer (100 mM Tris, pH 8.8, 150 mM NaCl, 10% (v/v) glycerol, and 50 mM reduced glutathione) digested with thrombin to cleave GST in solution for 6 h at 21 °C and then dialyzed (100 mM Tris, pH 7.5, 250 mM NaCl, 10% glycerol, and 1 mM DTT) overnight at 4 °C. To purify A3 enzymes from the free GST and thrombin, the enzyme stock was diluted to achieve a solution of 50 mM Tris pH 7.5, 50 mM NaCl, 10% (v/v) glycerol and 1 mM DTT for loading onto a DEAE FF column (GE Healthcare). Enzymes were eluted with a linear gradient of NaCl. The *Escherichia coli* strain containing the plasmid to express HIV reverse transcriptase was provided by Stuart Le Grice (National Cancer Institute). Expression of HIV reverse transcriptase was carried out as previously described [329]. In brief, cell lysates produced using sonication were clarified by centrifugation and then purified using a HisTrap FF crude column (GE Healthcare) and HiTrap heparin HP (GE Healthcare) as described previously [329]. Protein fractions were stored at –80°C. All proteins used were ~95% pure.

8.5.3 *In vitro* template switching assays. The 150 nt donor template RNA containing the PBS (nt 571–674) was heat annealed to an 18 nt ³²P-labeled RNA primer to mimic tRNA^{Lys,3} primer binding. The p/t (10 nM) was then used in reactions containing RT buffer (50 M Tris, pH 7.5, 40 mM KCl, 10 mM MgCl₂, 1 mM DTT), 500 μM dNTPs, 175 nM NC, 400 nM RT, and 400 nM acceptor RNA template for a 40:1 acceptor:donor ratio in the absence or presence of A3 (40, 80, or 320 nM). Reactions were preincubated at 37°C for 1 min before the addition of dNTPs which were used to start the reaction. A negative control was used which contained all reaction components except reverse transcriptase to ensure there was no contaminating polymerase activity. A second negative control was used that contained all the reaction components except for the acceptor RNA template to demonstrate the band pattern of the donor template alone. Reactions were stopped by adding a 5-fold excess of 20 mM EDTA and 95% formamide. Template switching was visualized by resolving samples on a 16% denaturing 8 M urea polyacrylamide gel. Gel band intensities were measured by phosphorimaging with a Typhoon Trio multipurpose scanner (GE Healthcare). The integrated gel band intensities of all bands in a

lane were calculated with ImageQuant software (GE Healthcare) and used to determine the relative amounts of donor versus acceptor products.

8.5.4 Steady state rotational anisotropy. Protein-p/t binding measurements were made by monitoring changes in steady-state fluorescence depolarization (rotational anisotropy). For measuring binding to the p/t, the 5'-end fluorescein (F)-labeled primer corresponding to the PBS primer used in template switching assay was heat annealed to the corresponding RNA template (PBS RNA) to form the binding substrate. The rotational anisotropy experiments (60 μ L) were incubated at 21°C and contained RT buffer, p/t (10 nM) and increasing concentrations of RT or A3. Protein concentrations used ranged from 0–5 μ M for measuring p/t binding affinities. Rotational anisotropy was measured with a QuantaMaster QM-4 fluorometer (Photon Technology International). Samples were excited with vertically polarized light at 495 nm (6 nm band pass), and both vertical and horizontal emissions were monitored at 520 nm (6 nm band pass). Apparent dissociation constants (K_d) were determined through regression analysis using Sigma Plot 11.2 software.

8.5.5 In vitro polymerase fidelity assays. For standing start assays, the PBS donor template used above was heat annealed to a standing start primer (Table 5.3) to generate a standing start p/t [322, 324]. 10 nM of this template was then used in reactions containing RT buffer (50 M Tris, pH 7.5, 40 mM KCl, 10 mM MgCl₂, 1 mM DTT), 175 nM NC, 400 nM RT, and 200 nM of the various A3s. To ensure the extension products contained only the insertion bands expected, reactions were preincubated at 37°C for 1 min before the addition of increasing amounts of dNTP and heparin sodium mixture (1 mg/mL) equal in volume to the preheated reaction components. This ensured that once RT dissociated there was no further association with the template. The reactions were allowed to proceed for 10 min before being stopped by the addition of a 5-fold excess of 20 mM EDTA and 95% formamide.

For running start assays, the PBS donor template was heat annealed to a running start primer (Table 5.3) to generate a running start p/t. 10 nM of this template was then used in reactions containing RT buffer (50 M Tris, pH 7.5, 40 mM KCl, 10 mM MgCl₂, 1 mM DTT), 175 nM NC, 400 nM RT, and 200 nM of the various A3s as well as 50 μ M of the running start base dCTP. To ensure the extension products contained only the insertion bands expected, reactions were preincubated at 37°C for 1 min before the addition of increasing amounts of dNTP and heparin sodium mixture (1mg/mL) equal in volume to the preheated reaction components. This ensured

that once RT dissociated there was no further association with the template. The reactions were allowed to proceed for 10 min before being stopped by the addition of a 5-fold excess of 20 mM EDTA and 95% formamide.

Both standing and running start samples were resolved on a 16% denaturing 8 M urea polyacrylamide gel. Gel band intensities were measured by phosphorimaging with a Typhoon Trio multipurpose scanner (GE Healthcare). The integrated gel band intensities of all bands in a lane were calculated with ImageQuant software (GE Healthcare), and either initial velocity (standing start) or ratio of I_2 over I_1 (running start) was plotted against concentration of dNTP.

Table 8.3. List of primers and templates.

Name	Sequence
PBS donor (forward)	AGC CTC AAT AAA GCT TGC CTT GA
PBS donor (reverse)	GTC CCT ATT AAC TTT CGC TTT CA
PBS acceptor (forward)	ACT GGG TCT CTC TTG TTA GGC
PBS acceptor (reverse)	CTG CTA GAG ATT TTT ACT CAG TC
Protease template (forward)	ATA GGA GGA CAA CTG AAA GAA GC
Protease template (reverse)	GAT AAA ACC TCC AAT TCC CCC TAT C
PBS primer	GUC CCU GUU CGG GCG CCA
Standing start primer	GAT AAA ACC TCC AAT TCC <u>C</u>
Running start primer	GAT AAA ACC TCC AAT TCC <u>CC</u>
FAM-PBS primer	Fam-GUC CCU GUU CGG GCG CCA
PBS donor template	AGC CUC AAU AAA GCU UGC CUU GAG UGC UUA AAG UGG UGU GUG CCC AUC UGU GUU AGG ACU CUG GUA ACU AGA GAU CCC UCA GAU CAC UCU AGA CUG AGU AAA AAU CUC UAG CAG UGG CGC CCG AAC AGG GAC UUG AAA GCG AAA GUU AAU AGG GAC
PBS acceptor template	ACU GGG UCU CUC UUG UUA GGC CAG GUC GAG CCC GGG AGC UCU CUG GCU AGC AGG GGA ACC CAC UGC UUA AAG CCU CAA UAA AGC UUG CCU UGA GUG CUU AAA GUG GUG UGU GCC CAU CUG UGU UAG GAC UCU GGU AAC UAG AGA UCC CUC AGA UCA CUC UAG ACU GAG UAA AAA UCU CUA GCA G
Protease template	AUA GGA GGA CAA CUG AAA GAA GCU CUA UUA GAU ACA GGA GCA GAU GAU ACA GUA UUA GAA GAU AUA AAU UUG CCA GGG AAA UGG AAA CCA AAA AUG AUA GGG GGA AUU GGA GGU UUU AUC

9.0 Biochemical determinants of deamination during transcription and replication by APOBEC3s.

Since 2013, there has been increasing research into the role of the A3 family's ability to mutate genomic DNA. If these enzymes are expressed at the wrong time or in the wrong cell type and have access to genomic DNA in the nucleus, there is the potential for deamination of genomic cytosines that can lead to mutations and cellular transformation [170, 330, 331]. A3B, A3A and A3H hap I have been demonstrated to contribute to cancer mutagenesis [17, 174, 175, 184, 196]. Among these three enzymes, A3B has emerged as the leading contributor to mutagenesis [180, 332-336]. Typically, the expression of A3B is restricted to the germ cells and the lymphoid lineages, however inappropriate expression of A3B can lead to genomic instability [170]. A3B recognizes 5'TC motifs that are preferably preceded by a G or A, and this sequence context of mutations has been found in cancer genomes [17, 175, 182]. It is thought that during DNA replication, the cytosines in the exposed ssDNA on the lagging strand are susceptible to deamination by A3B [190-194]. If left unrepaired, these uracils will template addition of the incorrect base leading to C/G to T/A transition mutations [170]. This aberrant cytosine deamination activity of A3B contributes to mutations in multiple cancers, such as breast, cervical, lung (adeno- and squamous cell), bladder, and head and neck [17].

Despite a large amount of cellular based data on the contribution of A3B to cancer mutagenesis, there is limited data on the biochemical mechanism of the wild-type enzyme, as well as the biochemical mechanisms used by other implicated A3s. For HIV restriction by A3 enzymes, encapsidation into the virion is not sufficient for restriction, and the inherent biochemical properties of the enzyme determine the frequency of deaminations in viral DNA. [74, 108-110, 119] Therefore, it is reasonable to assume that availability of the ssDNA in the nucleus is not sufficient for deamination during transcription and replication, and that the biochemical mechanisms of these enzymes may impart this observed activity. It is also interesting to note that the enzymes known to restrict HIV, A3G, A3F, A3H (hap II, V, and VII), and A3D, are not the same enzymes implicated in genomic damage. Therefore, we hypothesize that there will be a mechanism unique to these genomic mutators, and different than the enzymes important for HIV restriction that allows for access to ssDNA during dynamic cellular processes. One clue that these mechanisms may be distinct is that A3A, previously characterized as a non-processive enzyme [119], is an A3 relevant to genomic mutagenesis. Processivity is known to be

a determinant for efficient HIV restriction by A3s, and therefore the processivity of the enzymes may not be as important for deamination during transcription and replication. Further characterization of these mechanisms for A3B, A3H hap I and A3A on ssDNA will be outlined in Chapter 9, and will determine the ability to deaminate “off-target” substrates.

10.0 Enzyme cycling contributes to the efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B.

Madison B. Adolph¹, Robin P. Love¹, Yuqing Feng¹, Linda Chelico¹.

¹Department of Microbiology and Immunology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

The information in this chapter was previously published:

Adolph, M.B., Love, R.P., Feng, Y., Chelico, L. (2017). Enzyme cycling contributes to the efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B. *Nucleic Acids Research*, 45(20): 11925-11940.

Copyright © The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

All experiments in this chapter were performed by M.B.A. except Figure 9.7C and Figure 9.8 that resulted from experiments that were performed by R.P.L. Y.F. provided preliminary data for Figure 9.7B. M.B.A., R.P.L., L.C. conceived and designed the experiments, analyzed the data, and wrote the article. M.B.A., R.P.L., Y.F., L.C. revised the article.

10.1 Abstract

The single-stranded DNA cytidine deaminases APOBEC3B, APOBEC3H haplotype I, and APOBEC3A can contribute to cancer through deamination of cytosine to form promutagenic uracil in genomic DNA. The enzymes must access single-stranded DNA during the dynamic processes of DNA replication or transcription, but the enzymatic mechanisms enabling this activity are not known. To study this, we developed a method to purify full length APOBEC3B and characterized it in comparison to APOBEC3A and APOBEC3H on substrates relevant to cancer mutagenesis. We found that the ability of an APOBEC3 to cycle between DNA substrates determined whether it was able to efficiently deaminate single-stranded DNA produced by replication and single-stranded DNA bound by replication protein A (RPA). APOBEC3 deaminase activity during transcription had a size limitation that inhibited APOBEC3B tetramers, but not APOBEC3A monomers or APOBEC3H dimers. Altogether, the data support a model in which the availability of single-stranded DNA is necessary, but alone not sufficient for APOBEC3-induced mutagenesis in cells because there is also a dependence on the inherent biochemical properties of the enzymes. The biochemical properties identified in this study can be used to measure the mutagenic potential of other APOBEC enzymes in the genome.

10.2 Introduction

The APOBEC family of enzymes in humans has 12 members and is comprised of RNA and DNA cytidine deaminases [5]. The enzymes are named after APOBEC1, the first family member discovered, that edits the mRNA of apolipoprotein B in the intestine by converting cytosine to uracil and forming a stop codon [5]. Other family members with demonstrated deaminase activity have roles in immunity. Activation induced cytidine deaminase (AID) converts cytosine to uracil in single-stranded (ss) DNA created during transcription of immunoglobulin genes in activated B cells [11]. These uracils promote antibody maturation and class switching [11]. The APOBEC3 enzyme family has seven members (A-H, excluding E) that are expressed in germ cells, CD4⁺ T cells, or myeloid derived cells and are able to convert cytosine to uracil in ssDNA of invading viruses, endogenous retroviruses, and retrotransposons [2]. These uracils result in mutagenesis and functional inactivation of the virus [2]. APOBEC3A can also edit mRNAs in monocytes and macrophages [6]. In the APOBEC3 family some of the enzymes have only one Zn-coordinating deaminase domain (A, C, H) and others have two (B, D, F, G). There are also other members such as APOBEC2, APOBEC4, and APOBEC5 that have no

in vitro catalytic activity, although APOBEC2 appears to be involved in muscle development [8-10, 337]. Despite the diverse beneficial functions of this family, there is one unifying characteristic; if the enzymes are expressed at the wrong time or in the wrong cell type and have access to the nucleus they can deaminate cytosines in genomic DNA and this can contribute to cellular transformation, cancer, and the ongoing mutagenic processes in tumors [170, 330, 331]. Thus far, three APOBEC3 (A3) enzyme family members have been characterized to contribute to cancer mutagenesis, APOBEC3A (A3A), APOBEC3B (A3B), and APOBEC3H haplotype I (A3H hap I) [17, 174, 175, 184, 196].

A3B serves as a good example for the role of A3s in cancer since it has been most extensively studied in this context. A3B is localized to the nucleus and has cytidine deaminase activity on ssDNA [209, 338]. When A3B expression occurs in the wrong context, A3B promotes genomic instability at the nucleotide level that results in contributions to the mutations that arise in breast cancer and lead to resistance to the chemotherapeutic drug tamoxifen [175, 179]. A3B can also contribute to mutations that arise in a number of other cancers [17, 180, 332-336, 339, 340]. A3B recognizes 5'TC motifs that are preferably preceded by a G or A, and this sequence context of mutations has been found in cancer genomes [17, 175, 341]. It is thought that during DNA replication, the cytosines in the exposed ssDNA on the lagging strand are susceptible to deamination by A3B [190-194]. If left unrepaired, these uracils will template addition of the incorrect base leading to C/G to T/A transition mutations [170]. The uracils can also induce the formation of double-strand (ds) DNA breaks or be repaired in an error-free or error-prone manner [170].

Despite this understanding of A3B activity at a cellular level, biochemical characterizations have centered on using truncated forms of A3B that contain only the catalytically active C-terminal domain (CTD) [17, 175, 270, 341-346]. A biochemical analysis of full-length (fl) wild-type A3B on substrates relevant to cancer mutagenesis is lacking. Owing to poor solubility of A3B in heterologous *E. coli* expression systems and its mutagenic activity, purification of fl A3B has posed difficulties [58, 343, 345, 347]. Although the N-terminal domain (NTD) of A3B is not catalytically active, it does contribute to activity and the CTD alone is 10-fold less active than fl A3B as measured by a mutator assay where A3B is expressed in *E. coli* or the *in vitro* activity of maltose binding protein tagged A3B [343, 346]. We have little knowledge of why the full length enzyme is more active than the CTD, how fl A3B can access ssDNA in the

genome, if it requires cofactors, or how it competes with other ssDNA binding proteins such as replication protein A (RPA) [270, 342, 344, 345]. Although A3A and A3H have been previously characterized biochemically, there is similarly little information on how at a biochemical level these enzymes would access ssDNA in the genome [110, 119, 174, 272, 348].

To better understand the biochemical mechanisms underlying the ability of A3B, A3H hap I, and A3A to induce genomic mutagenesis we undertook an analysis of these enzymes on ssDNA and used *in vitro* phage based systems of transcription and replication to determine if these enzymes alone can deaminate dynamically produced ssDNA. Importantly, to complete this research, we purified fl A3B and for the first time provide a characterization of the enzyme, which shows that it is different than the commonly used A3B CTD, suggesting that the A3B CTD is not a functional substitute. We also find that deamination activity during DNA synthesis and when ssDNA is bound by RPA requires an enzyme that can rapidly cycle between DNA substrates. For deamination during transcription, we found that the larger oligomeric state of A3B inhibited its activity. These biochemical studies provide evidence that the intrinsic characteristics of the enzymes have a contribution to the mutagenic process, rather than the availability of ssDNA alone. The inherent biochemical characteristics common to A3A, A3B, and A3H hap I can be used to measure the mutagenic potential of other APOBEC enzymes in the genome.

10.3 Results

10.3.1 A3B is a processive enzyme that forms higher order oligomers. The purification of A3B has posed difficulties because even the basal expression levels of A3B from eukaryotic vectors when being amplified in *E. coli* results in the purification of mutated A3B vector sequences [58, 345, 347]. Some labs have overcome this by inserting an intron in the A3B [58, 345]. However, since A3 enzymes are most soluble and active when expressed from recombinant baculovirus infected Sf9 cells [349], and this system is not amenable to splicing of intron containing A3B, we were unable to overcome expression difficulties in this manner (data not shown) [350]. Instead, we made a GST-tagged version of A3B that inactivated the enzyme and amplified this GST-A3B clone in a baculovirus transfer vector for subsequent generation of recombinant baculovirus. We confirmed that the GST-A3B becomes inserted in the baculovirus genome with no mutations (data not shown). The GST-A3B is then expressed and the GST tag is cleaved during purification to impart A3B activity.

The purified fl A3B has a specific activity on ssDNA of 1.60 pmol/μg/min, which is similar to A3G (Figure 10.1A and Table 10.1). In comparison to the literature, the Sf9 insect cell produced fl A3B is 32-fold more active than the Maltose Binding Protein tagged fl A3B produced from *E. coli* [346]. We also produced in Sf9 cells the two commonly used A3B CTD enzymes that contain amino acids 187-392 (187 CTD) or 193-392 (193 CTD) of A3B [342, 343]. The fl A3B is still 40-fold more active than the A3B 193 CTD that was reported to have increased activity over the more commonly used A3B 187 CTD (Figure 10.1A, Table 10.1, and Figure 10.2) [342, 343]. We did not detect in vitro catalytic activity for the A3B 187 CTD (Figure 10.1A, Table 10.1, and Figure 10.2). Thus, despite the A3B NTD not having catalytic activity, it appears to be able to facilitate enzyme activity. This is common to other A3 enzymes that contain two Zn²⁺ coordinating domains, such as A3G and A3F [107, 108]. The NTD in these enzymes is a processivity domain and mediates oligomerization [98, 107, 108]. The fl A3B may also oligomerize since it bound to ssDNA cooperatively as demonstrated by a best fit of the binding data to a sigmoidal binding curve (Table 10.1 and Figure 10.3A). To determine if fl A3B oligomerization occurred in solution and if it was different between fl A3B and the A3B CTDs, we used size exclusion chromatography (SEC). The SEC showed that fl A3B (46 kDa) formed predominantly tetramers (184 kDa) and a small proportion existed as dimers (92 kDa) (Figure 10.1B and Figure 10.4A-B). In contrast, the A3B 187 CTD was monomeric and the A3B 193 CTD was primarily monomeric with a small proportion of dimers, consistent with an earlier report (Figure 10.1C-D, Figure 10.4C-D) [343]. However, both A3B CTDs bound ssDNA non-cooperatively with an apparent dissociation constant (K_d) that was ~4-fold higher than the fl A3B (Table 10.1, and Figure 10.3B, best fit to a rectangular hyperbola). This suggests that the CTD is mainly a catalytic center and that the NTD promotes ssDNA interaction and self-interaction. Altogether, the data strongly support that the A3B CTD is not a suitable replacement for the full-length enzyme and we continued our study with fl A3B only (hereafter referred to as A3B).

Since A3B is an ssDNA binding enzyme that deaminates only in a specific recognition motif, it must first find the motifs by searching the ssDNA non-specifically through a DNA scanning process called facilitated diffusion [104-106, 351]. In facilitated diffusion, electrostatic interactions of the enzyme with the DNA facilitate a search by enabling the enzyme to remain bound longer to the DNA than the time it spends in the bulk solution not bound to a substrate. The ssDNA scanning mechanism enables the enzyme to be processive and deaminate more than

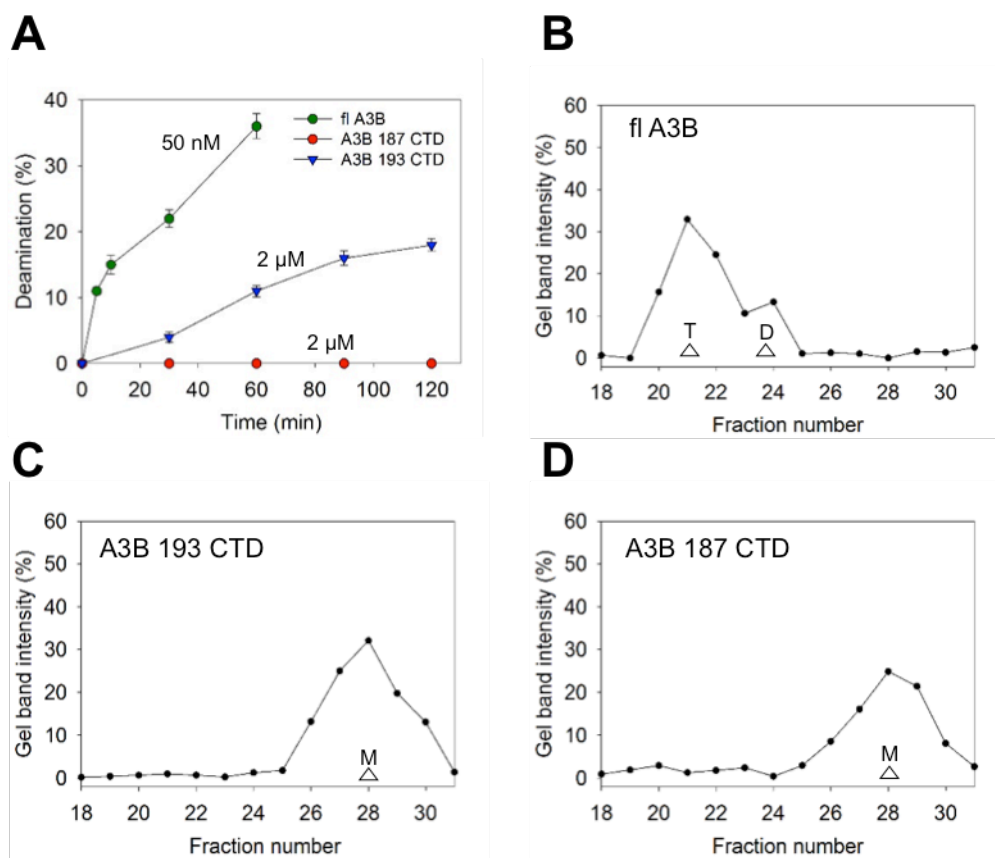


Figure 10.1. A3B NTD mediates enzyme activity and oligomerization. (A) Time course of A3B, A3B 193 CTD, and A3B 187 CTD on a 118 nt fluorescently labeled ssDNA with two 5'ATC deamination motifs spaced 63 nt apart. Reactions were performed with 100 nM substrate DNA and 50 nM (fl A3B) or 2 µM (187 CTD, 193 CTD) for the indicated amount of time (5-60 min fl A3B or 30-120 min CTD mutants). (B-D) Size exclusion chromatography profiles of 10 µg (B) fl A3B, (C) 193 CTD, and (D) 187 CTD from a 10 mL Superdex 200 column was used to calculate the oligomerization state of the enzyme from a standard calibration curve. An “M” denotes a monomer fraction, a “D” denotes a dimer fraction, and a “T” indicates a tetramer fraction. (A) fl A3B formed tetramers (apparent molecular weight 184 kDa) and dimers (apparent molecular weight 92 kDa). (C-D) 193 CTD and 187 CTD resolved as monomers (apparent molecular weight 23 kDa). The chromatograms were constructed by analyzing the integrated gel band intensities of each protein in each fraction after resolution by SDS-PAGE (Figure 10.4).

Table 10.1. Deamination activity and binding of APOBEC3 enzymes on oligonucleotide substrates.

Enzyme	Specific activity (pmol/μg/min) ¹	Apparent K_d, ssDNA (Hill Coefficient)	Specific activity, R-loop (pmol/μg/min)	Apparent K_d, R-loop	Apparent K_d, DNA/RNA (Hill Coefficient)
fl A3B	1.60 ± 0.05	0.35 ± 0.02 μM (2.8)	0.54 ± 0.09	0.90 ± 0.05 μM	0.31 ± 0.03 μM
A3B 187 CTD	None detected	1.51 ± 0.50 μM	Not determined	Not determined	Not determined
A3B 193 CTD	0.04 ± 0.01	1.30 ± 0.34 μM	Not determined	Not determined	Not determined
A3A	1.10 ± 0.10	9.10 ± 2.50 μM (1.7)	0.25 ± 0.07	> 16 μM ²	> 10 μM ²
A3H hap I*	0.60 ± 0.02	0.29 ± 0.01 μM (1.6)	0.35 ± 0.08	> 7 μM ²	1.20 ± 0.16 μM
A3G	2.50 ± 0.15	0.09 ± 0.02 μM (1.9)	0.48 ± 0.10	1.30 ± 0.15 μM	0.16 ± 0.01 μM (2.5)

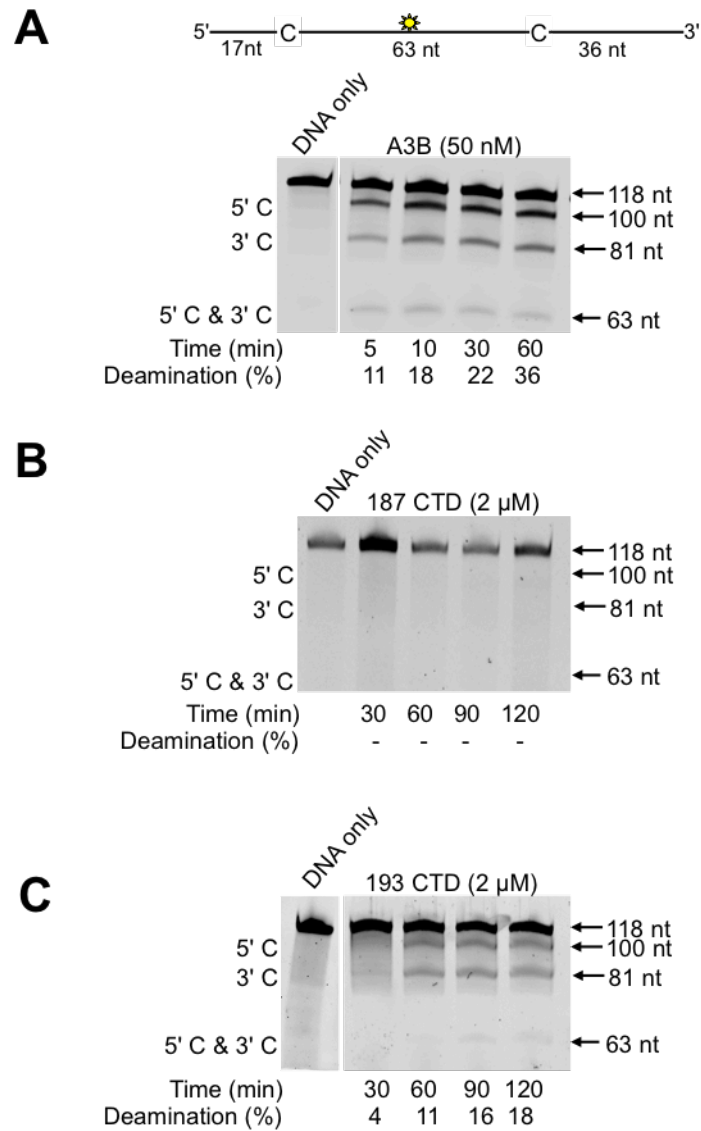


Figure 10.2. Time course of deamination activity for fl A3B and CTD mutants. To determine the activity over time shown in Figure 1A, 100 nM of a 118 nt fluorescently labeled ssDNA with two 5'ATC motifs separated by 63 nt was incubated with (A) 50 nM (fl A3B) or (B-C) 2 μM (187 CTD, 193 CTD) for the indicated amount of time (5-60 min fl A3B or 30-120 min CTD mutants). The percent deamination at each time point is shown below the gel.

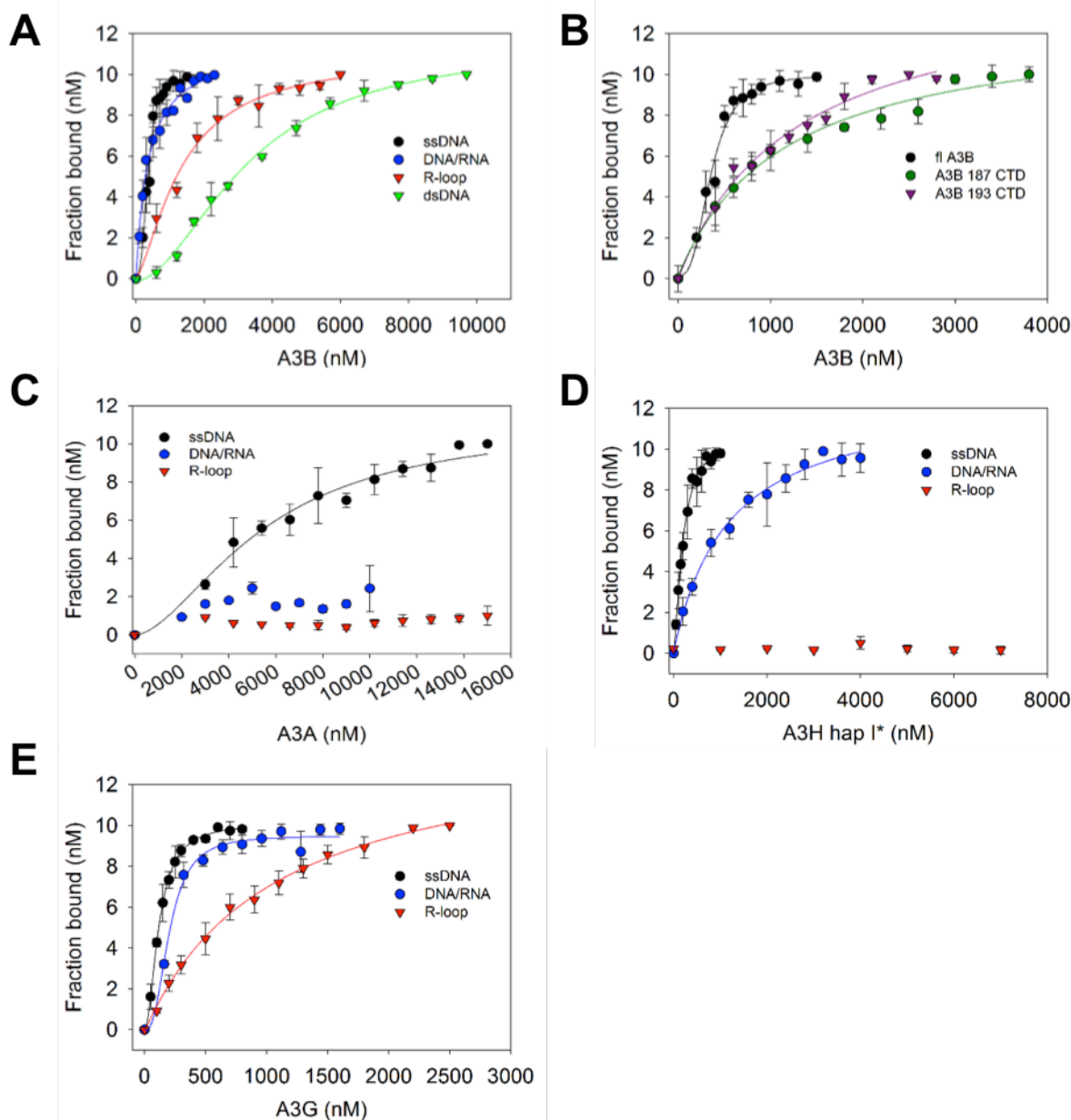


Figure 10.3. Interaction of A3 enzymes with nucleic acids. The apparent dissociation constant (K_d) of A3 enzymes on nucleic acids was determined by steady state rotational anisotropy. Steady state binding to ssDNA, DNA/RNA hybrids, R-loops, or dsDNA was determined. Panels show binding curves for (A) A3B, (B) A3B in comparison to A3B CTDs, (C) A3A, (D) A3H hap I*, and (E) A3G. The apparent K_d for these binding data are summarized in Table 9.1. Error bars represent the standard deviation from three independent experiments.

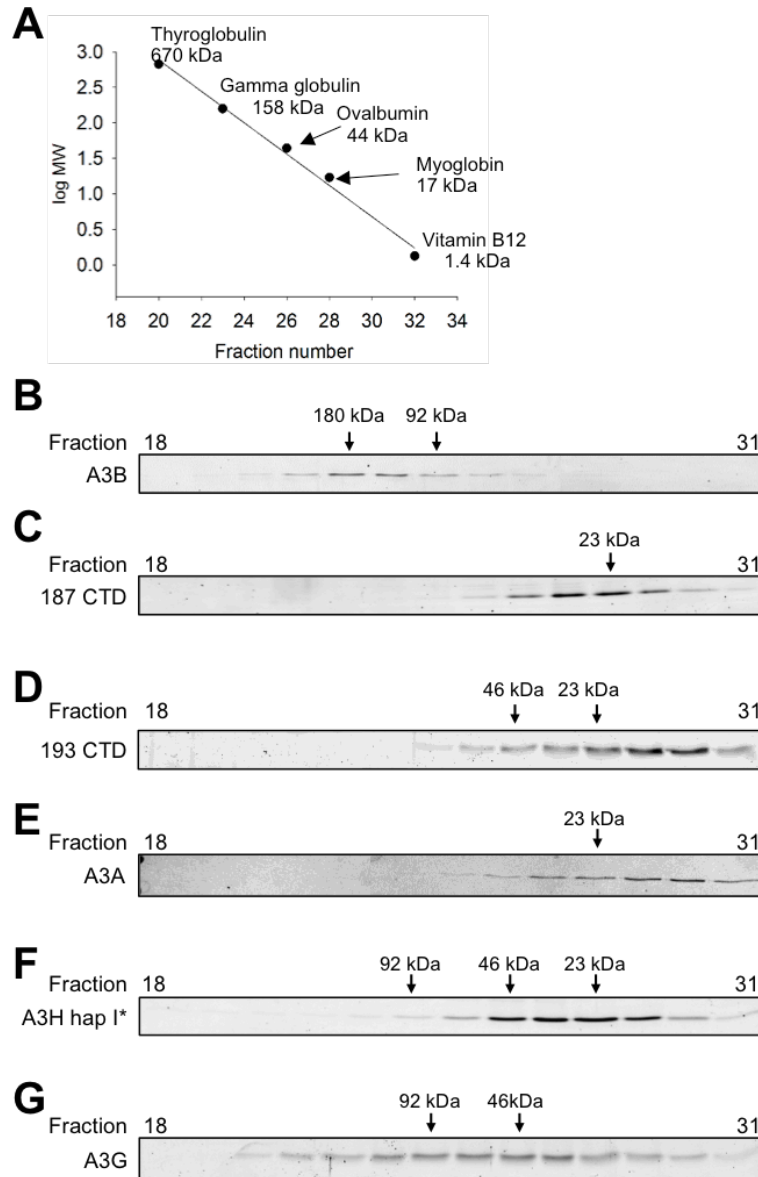


Figure 10.4. SDS-PAGE of size exclusion chromatography fractions. (A) Standard calibration curve used to calculate the molecular weights. (B-G) The chromatograms from the 10 ml Superdex 200 column in Figure 1b, c, d and Figure 3 e, f, g were constructed by analyzing the integrated gel band intensities of the protein in each fraction after resolution by SDS-PAGE. The gels for each panel were resolved, stained with Oriole stain, and scanned in parallel. The gels show the size exclusion chromatography fractions resolved by SDS-PAGE for each A3, as labeled on each panel. The molecular weight calculated from the standard curve are shown for the peak fractions.

one cytosine in a single enzyme-substrate encounter. Other A3 enzymes have been found to use facilitated diffusion [108-111]. Facilitated diffusion is also used by restriction enzymes and for DNA repair proteins, such as Uracil DNA glycosylase [351, 352].

Using an in vitro deamination assay on a synthetic ssDNA substrate containing two A3B 5'ATC deamination motifs we can measure processive deaminations that occurred by facilitated diffusion [111]. Essential to this assay is that the reactions are carried out under single-hit conditions (<15% substrate usage) to ensure that each ssDNA substrate was acted upon by only one enzyme during the course of the reaction [242]. If under single-hit conditions we observe deamination of both 5'ATC motifs we can conclude that the enzyme is processive and deaminated both motifs in a single enzyme-substrate encounter. To decrease the chance that cooperative binding of different A3B molecules at different times on the ssDNA resulted in an apparent processivity, the reactions took place with an excess of ssDNA to promote interaction of single molecules of A3B with the ssDNA, which may be either dimer or tetramer molecules from solution (Figure 10.1B). We also calculated a processivity factor which is a ratio of the observed deaminations at both 5'ATC sites to the calculated theoretical number of deaminations that would occur independently at both 5'ATC motifs if the enzyme were non-processive (see Materials and Methods) [111]. Facilitated diffusion encompasses a range of movements termed sliding, jumping or hopping, and intersegmental transfer [2, 104, 351, 353]. The sliding occurs along the phosphate backbone and for A3 enzymes moving on ssDNA is limited to a scanning length of 20 nt or less [108, 253]. The jumping, hopping, and intersegmental transfer are long range movements where the enzyme can use a 3-dimensional search to sample distal DNA regions either by microscopic dissociations and reassociations along the DNA (jumping or hopping) or through a doubly bound state (intersegmental transfer) [2, 104, 351, 353].

To determine if A3B could slide along the ssDNA phosphate backbone, we conducted the deamination experiment with 5'ATC motifs that were separated by only 5- or 14- nt. A3B is able to slide and had a processivity factor of 4 to 6, depending on the substrate. The processivity factor indicates that A3B was at least 4-fold more likely to processively deaminate both 5'ATC motifs than to make a non-processive deamination of only one 5'ATC motif (Figure 10.5A-B). To determine if A3B could also move by 3-dimensional diffusion we tested deamination on a substrate where the 5'ATC motifs were separated by 63 nt. The A3B had a processivity factor of 4, similar to the processivity exhibited on the substrates with closer spaced motifs (Figure 10.5C).

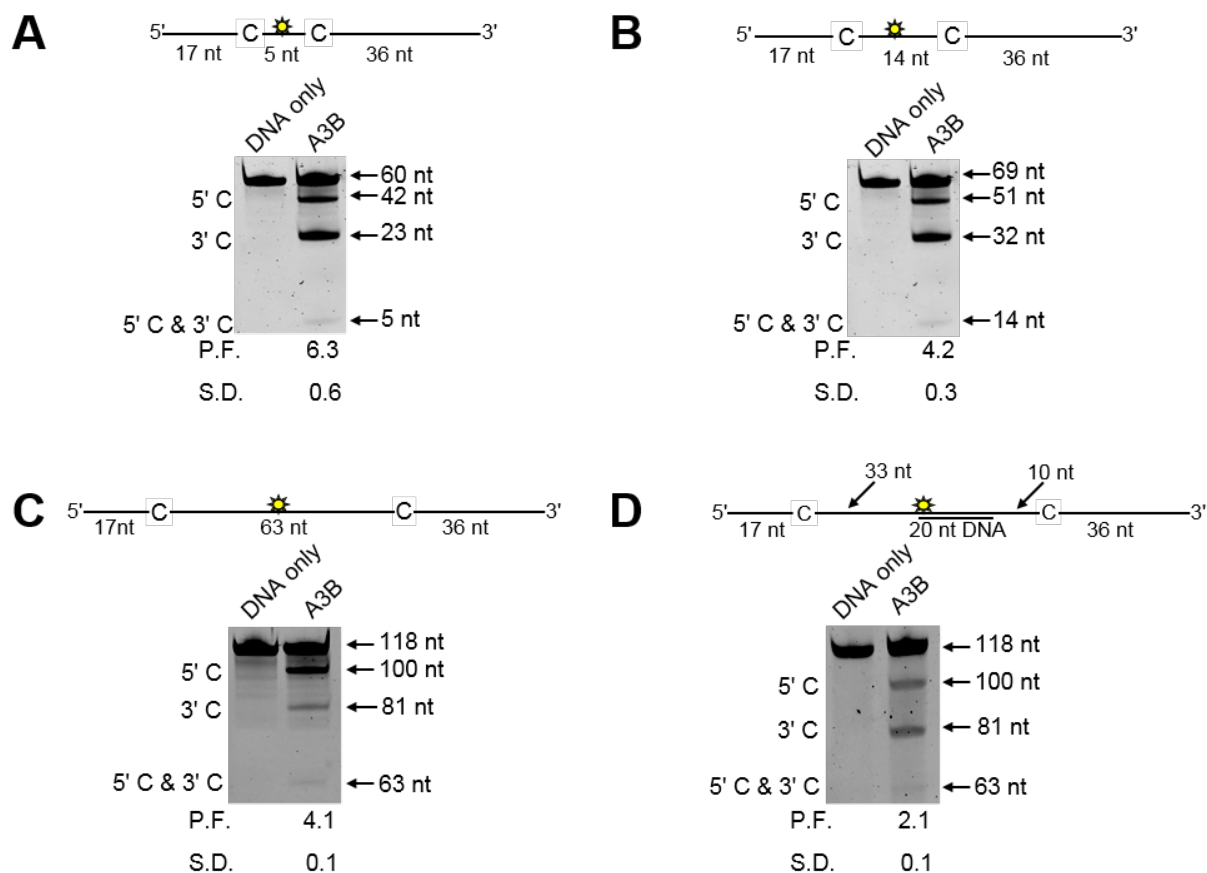


Figure 10.5. A3B is processive enzyme that both slides and moves 3-dimensionally on ssDNA. Processivity of A3B was tested on ssDNA substrates that contained a fluorescein-labeled deoxythymidine (yellow star) between two 5'ATC deamination motifs separated by different distances. (A) Deamination of a 60 nt ssDNA substrate with deamination motifs spaced 5 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment. (B) Deamination of a 69 nt ssDNA substrate with deamination motifs spaced 14 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 14 nt labeled fragment. (C) Deamination of a 118 nt ssDNA substrate with deaminated cytosines spaced 63 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. (D) Deamination of a 118 nt ssDNA substrate as in (C) but with a 30 nt complementary DNA annealed between the deamination motifs. The measurements of processivity factor (P.F.) and the standard deviation (S.D.) from three independent experiments are shown below the gel.

Due to the distance between the motifs alone the data suggest that A3B can move 3-dimensionally by jumping or intersegmental transfer. However, to confirm that A3B could move 3-dimensionally, we annealed a double-stranded (ds) DNA between the two 5'ATC motifs. Since A3B binds dsDNA 13-fold less well than ssDNA (Figure 10.3A), the dsDNA acts to block A3B sliding in between the two 5'ATC motifs. Thus, to deaminate both A3B would have to transverse the dsDNA by 3-dimensional diffusion. Consistent with A3B not being able to slide across the dsDNA, we find that the processivity decreases 2-fold in the presence of the complementary dsDNA, but is not diminished completely (Figure 10.5D). The 2-fold decrease results from failed attempts to slide over the dsDNA, which results in enzyme dissociation [108]. This supports that A3B can move over the dsDNA to complete processive deaminations by either jumping or intersegmental transfer. In contrast, the A3B CTD is not processive on any ssDNA substrate tested (Figure 10.6).

Having characterized the basic biochemical features of A3B we wanted to make a comparison to the other deaminases implicated in cancer mutagenesis to determine if they shared specific biochemical abilities. A3A has been previously characterized by multiple labs and consistent with these past studies we report that A3A is not processive and binds ssDNA in the high micromolar range ($\sim 9 \mu\text{M}$) [119, 272, 348] (Figure 10.7A, Table 10.1, and Figure 10.3C). A3H hap I is not amenable for biochemical studies due to it being thermodynamically unstable [59, 174]. However, it is known that this is due to a Gly at position 105 and mutagenesis of this position to form a G105R mutant stabilizes the protein [59]. This A3H form is found in humans as haplotype VII [60]. For the purposes of our study the A3H hap I G105R (haplotype VII) was used as an A3H hap I proxy since the two enzymes are otherwise identical in their amino acid sequences. We refer to it as A3H hap I*. In contrast to A3A, the A3H hap I* is processive and binds ssDNA in the nanomolar range, similar to other A3H haplotypes previously characterized and A3B (Figure 10.7B, Table 10.1, and Figure 10.3D) [110].

10.3.2 Processivity is not required for deamination during transcription. Deamination of genomic DNA undergoing transcription has not been identified as a major mechanism by which these enzymes access ssDNA in the nucleus [190-194]. We wanted to test whether this was due to inherent characteristics of the enzymes. To determine if A3B, A3A, and A3H hap I* could deaminate during active transcription, we used an in vitro transcription system driven by the phage T7 RNA polymerase. This is a minimal system where transcription requires only the RNA

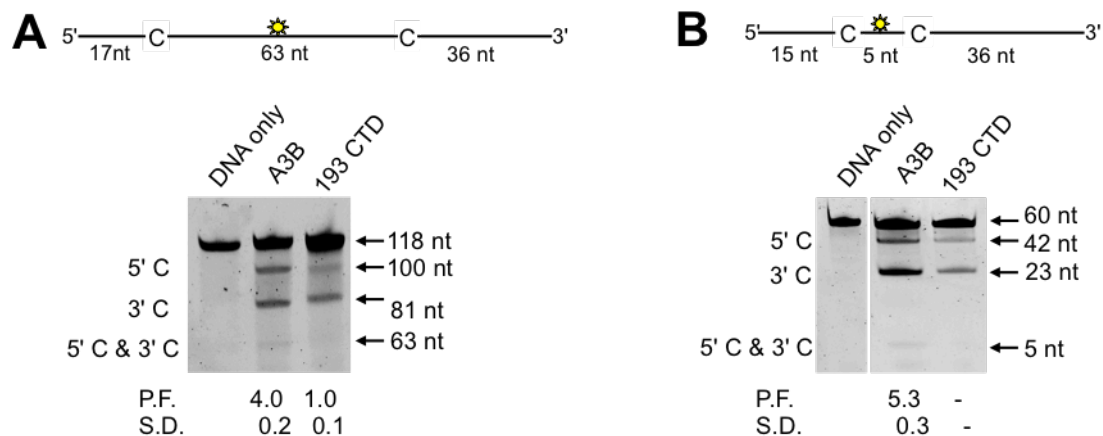


Figure 10.6. Comparison of processivity for fl A3B and A3B 193 CTD. (A) Deamination of a 118 nt ssDNA substrate with deaminated cytosines spaced 63 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. (B) Deamination of a 60 nt ssDNA substrate with deamination targets spaced 5 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 42- and 23- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 5 nt labeled fragment. A processive enzyme has a processivity factor >1 (see Materials and Methods). The measurements of processivity factor (P.F.) and the standard deviation (S.D.) from three independent experiments are shown below the gel.

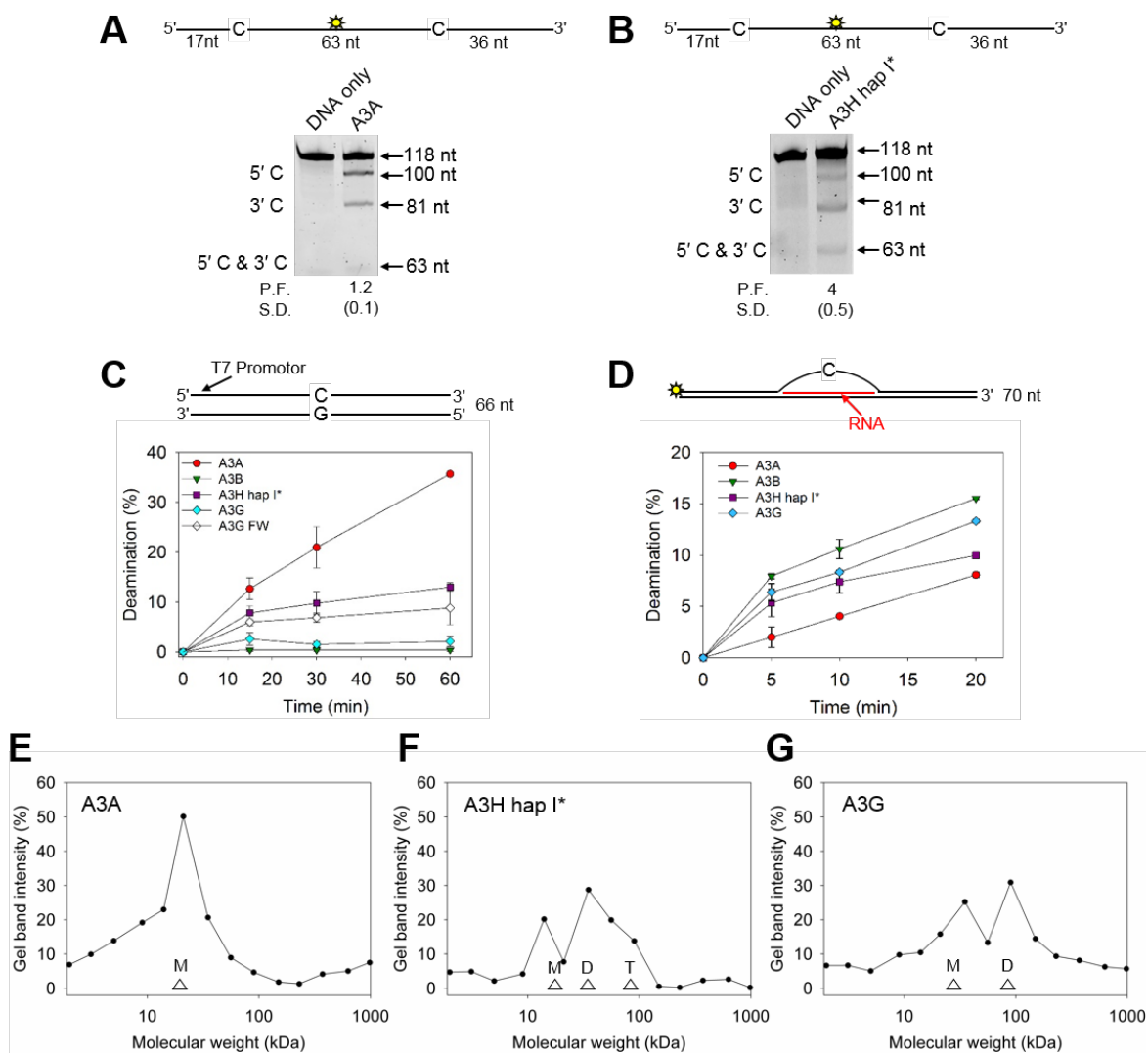


Figure 10.7. Deamination during transcription is mediated by size and not by processivity. (A-B) Processivity of (A) A3A and (B) A3H hap I* was tested on a 118 nt ssDNA substrate with deamination motifs (5'TTC, A3A; 5'CTC, A3H hap I*) spaced 63 nt apart. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 100- and 81- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. A processive enzyme has a processivity factor >1 (see Materials and Methods). (C) Deamination activity of A3B, A3A, A3H hap I*, A3G, and A3G F126A/W127A (FW) during T7 RNA polymerase transcription of a 66 nt dsDNA. The nontranscribed strand contained a single deamination motif for A3A (5'TTC), A3B (5'TTC), A3H hap I* (5'CTC), or A3G (5'CCC). Percent deamination was calculated from the integrated intensity of the product band resolved by PAGE (Figure 10.6). (D) Deamination activity of A3B, A3A, A3H hap I*, and A3G on a 70 nt R-loop substrate with a fluorescein label at the 5' end and a single deamination motif within the R-loop region. The motifs were 5'TTC (A3A, A3B, A3H hap I*) and 5'CCC (A3G). Each A3 enzyme (100 nM) was incubated with 100 nM of R-loop substrate for 5-20 min. Percent deamination was calculated from the integrated intensity of the product band resolved by PAGE (Figure 10.9). (E-G) Size exclusion chromatography profiles of 10 µg (E) A3A, (F) A3H hap I* and (G) A3G from a 10 mL Superdex 200 column was used to calculate the oligomerization state of the enzyme from a standard calibration curve. An "M" denotes a monomer fraction, a "D" denotes a dimer fraction, and a "T" indicates a tetramer fraction. (E) A3A resolved as a monomer (apparent molecular weight 23 kDa). (F) A3H hap I* formed a combination of tetramers (apparent molecular weight 92 kDa), dimers (apparent molecular weight 46 kDa), and monomers (apparent molecular weight 23 kDa). (G) A3G formed both dimers (apparent molecular weight 46 kDa) and monomers (apparent molecular weight 23 kDa). The chromatograms were constructed by analyzing the integrated gel-band intensities of each protein in each fraction after resolution by SDS-PAGE (Figure 10.4).

polymerase, dsDNA substrate and rNTPs. In order to quantify enzyme activity as deamination over time, we used an oligonucleotide dsDNA substrate, rather than a plasmid based T7 system. Each dsDNA substrate used the same transcribed strand, but had a unique non-transcribed strand with a single deamination motif on the non-transcribed strand specific for each A3. We let the reaction proceed in the presence of an A3 and T7 RNA polymerase and then the non-transcribed strand was then sequenced using a poisoned primer assay to detect deaminations [354]. Although not absolute, deamination of the non-transcribed strand would be consistent with what has been observed for APOBEC enzymes in yeast model systems [20, 21, 190], but for the purpose of this experiment we sought to quantify deamination rather than test the propensity for an A3 to deaminate a specific strand in the transcription bubble. The data show that A3B is unable to deaminate within the transcription bubble, but A3A and A3H hap I* are able (Figure 10.3C and Figure 10.8). To further investigate why A3B was unable to deaminate during transcription, we used an R-loop substrate to determine if A3B could deaminate in a “static” transcription bubble. A3B could deaminate this substrate, but the activity was 3-fold less than on ssDNA (Figure 10.7D, Table 10.1, and Figure 10.9A). Even A3A, which could deaminate the ssDNA created during transcription, was 4-fold less active on the R-loop substrate compared to fully ssDNA (Figure 10.7D, Table 10.1, and Figure 10.9B). The A3H hap I* had 1.7-fold less deamination activity on the R-loop substrate than ssDNA (Figure 10.7D, Table 10.1, and Figure 10.9C). Altogether, A3B did not appear to be deficient in deamination of an in vitro R-loop in comparison to A3A and A3H hap I*, suggesting that A3B activity was not limited by the small segment of ssDNA available (~8-20 nt) (Figure 10.7C-D). Although A3B could bind a DNA/RNA hybrid, the apparent K_d value was similar to that for ssDNA suggesting that A3B was not being sequestered away from the ssDNA through preferential binding to a DNA/RNA hybrid (Table 10.1). Notably, we could not reach saturation of the R-loop substrate in steady state binding experiments with A3A or A3H hap I*, despite demonstrated deamination activity, suggesting that the enzymes bound the R-loop with short residence times (Figure 10.7C-D, Table 10.1, and Figure 10.3C-D). This was also observed for A3A with the DNA/RNA hybrid (Table 10.1 and Figure 10.3C).

To understand why A3B did not efficiently deaminate during the dynamic transcription assay in contrast to A3A and A3H hap I* we first conducted a comparative analysis. A3B is made up of four individual enzyme units (Figure 10.1B). In contrast, A3A is a monomer and

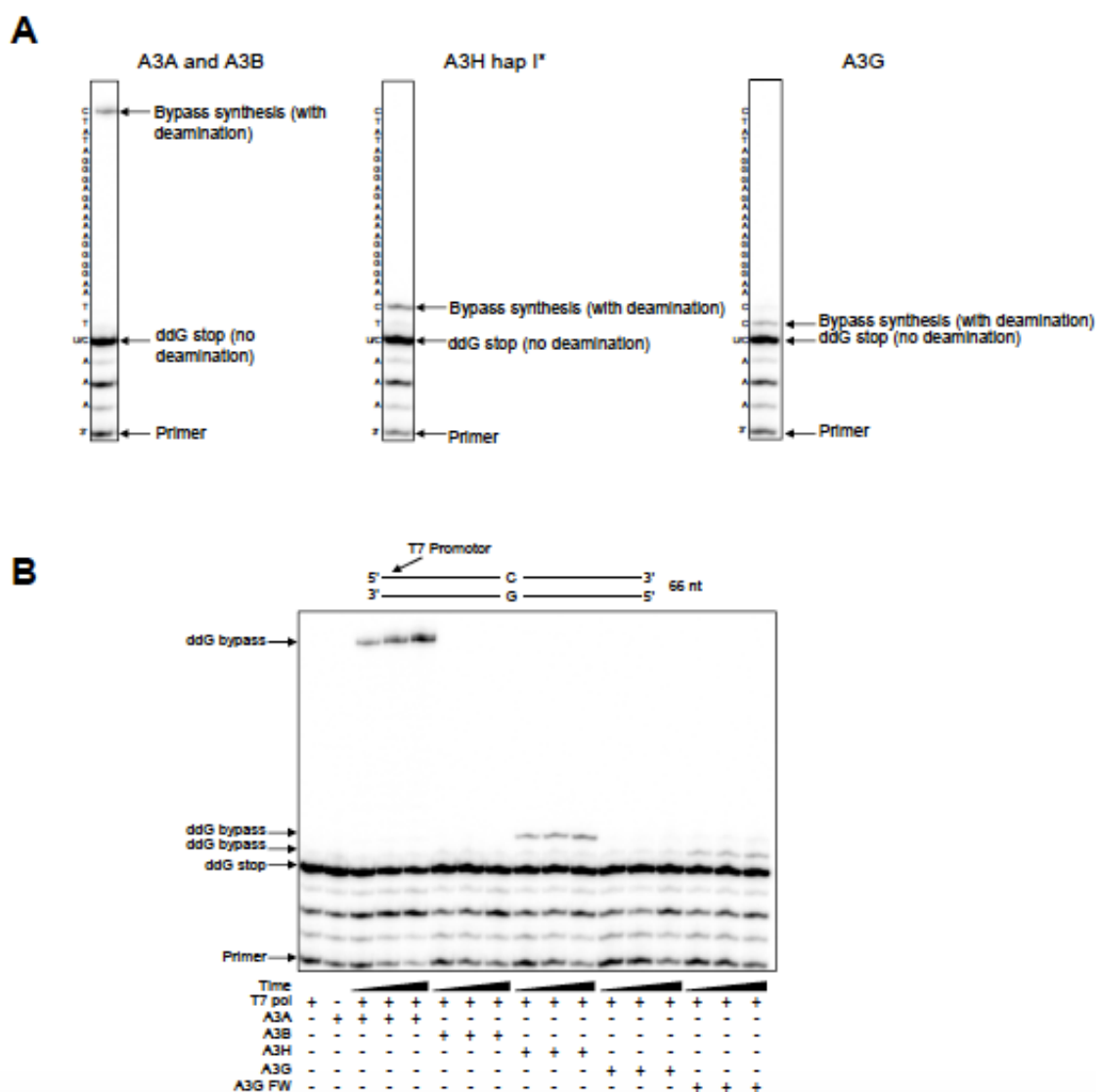


Figure 10.8. Deamination activity of A3 enzymes on dsDNA undergoing transcription. The non-transcribed strand of a dsDNA substrate that underwent T7 RNA polymerase-mediated transcription in the presence of A3A, A3B, A3H hap I*, A3G, or A3G FW was sequenced using a poisoned primer assay. (A) Interpretation of sequencing data. The U/C label indicates the cytosine embedded within the preferred deamination motif for the A3 indicated. When the C is not deaminated, a stop band will be present at that position due to the reaction being poisoned by the inclusion of ddG in place of dG during the sequencing reaction. If deamination occurs, the presence of a U allows synthesis to bypass the U/C site, and the next C in the sequence will have a stop band with an intensity that is proportional to the amount of C deamination. (B) A representative sequencing gel that was used to calculate the deamination over time as presented in Figure 10.7C.

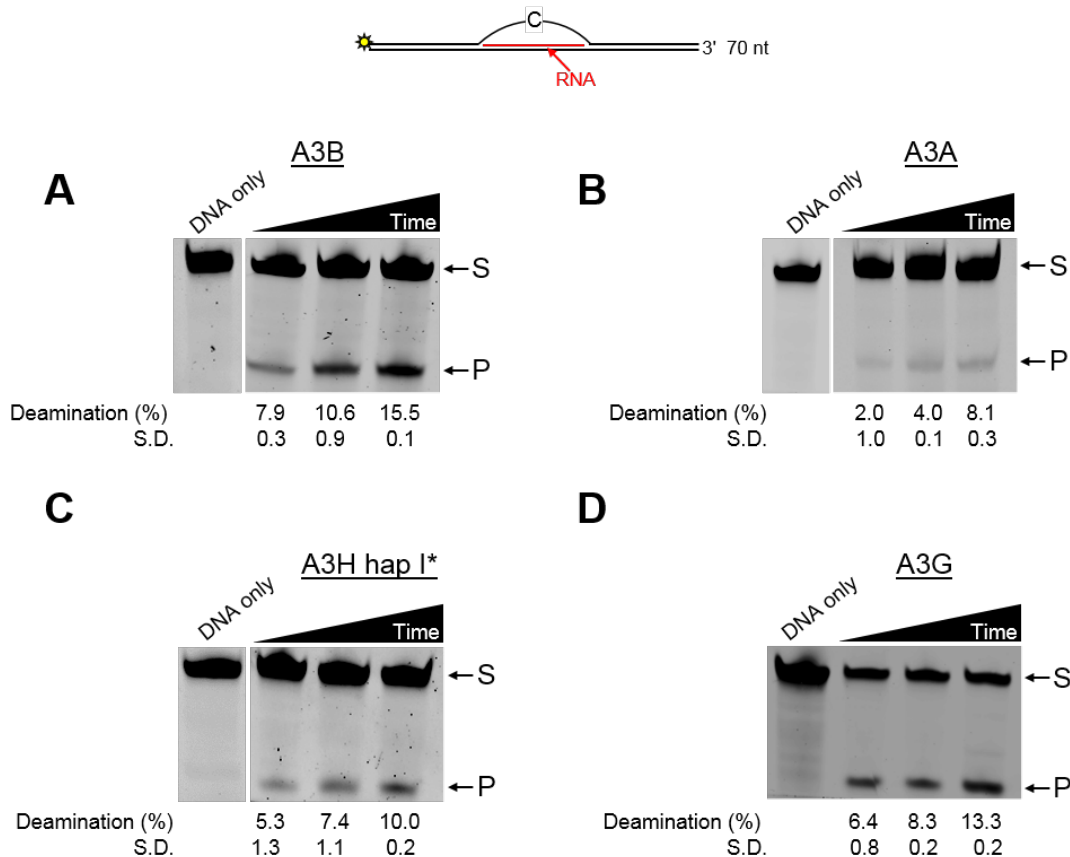


Figure 10.9. Deamination activity of A3 enzymes on R-loop substrates. To determine the activity of A3 enzymes on R-loops, 100 nM of A3 enzyme was incubated with 100 nM of a 70 nt R-loop substrate for 5-20 minutes. The substrate strand was fluorescently labeled at the 5' end (yellow star). An “S” indicates the 70 nt substrate band and a “P” indicates the 33 nt product resulting from deamination. Deamination was determined for (A) A3B, (B) A3A, (C) A3H hap I*, and (D) A3G. Percent deamination and standard deviation (S.D.) from three independent experiments are shown below the gel.

A3H hap I* is primarily a dimer (Figure 10.7E-F). Thus, the ability of A3B to stabilize itself on ssDNA may be diminished when the transcription bubble is moving since it has to account for eight nucleic acid binding domains, whereas A3A or A3H hap I* have only one or two to four, respectively (Figure 10.1B and Figure 10.7E-F). Although A3B bound ssDNA cooperatively, it did not bind the R-loop cooperatively suggesting that A3B bound the ssDNA during transcription at the oligomeric state observed in solution (Table 10.1 and Figure 10.3A). Nevertheless, the tetramer of A3B may become more readily destabilized by a moving substrate than an enzyme with less ssDNA binding sites.

To test this line of reasoning we used another processive and double deaminase domain A3 that has been extensively studied with regards to processivity and oligomerization [98, 99, 355, 356]. A3G forms monomers and dimers in solution and further oligomerizes on ssDNA by binding cooperatively (Figure 10.7G, Table 10.1, and Figure 10.3E) [72, 99, 102]. Similar to A3B, A3G was unable to deaminate ssDNA undergoing transcription, but was able bind and deaminate an in vitro R-loop (Figure 10.7C-D, Table 10.1, and Figure 10.8, and Figure 10.9D). A3G also bound ssDNA with more affinity than the DNA/RNA hybrid (Table 10.1 and Figure 10.3E). We then used a previously characterized monomeric form of A3G, an A3G F126A/W127A mutant (A3G FW) to determine if an A3G with less binding domains could deaminate during transcription [99]. The A3G FW was able to deaminate during transcription at an equivalent level to A3H hap I* (Figure 10.7C and Figure 10.8). The in vitro data support the conclusion that in a dynamic transcription system, the size of the enzyme is a determinant in activity and not substrate binding affinity or processivity. However, since R-loops or ssDNA generated during transcription have not been identified as a major source of mutations from whole genome sequencing data, there are other cellular factors not considered here that must protect the ssDNA at a genome level [190-194].

10.3.3 Enzyme cycling is required for efficient deamination during DNA replication. The deamination of the lagging strand of genomic DNA undergoing replication has been identified as a major source of A3 catalyzed cytosine deaminations [190-194]. This is especially true in cancer cells with a high amount of replication stress where ssDNA can accumulate from incomplete lagging strand synthesis, stalled replication forks, or excessive replication fork firing [192, 332]. As a result, there is a view that the ssDNA is simply left available for A3 enzymes to catalyze

deamination of cytosine [176, 341, 357]. To test this experimentally we set up a minimal phage replication system where the DNA polymerase of phage $\Phi 29$ can initiate rolling circle replication and strand displacement synthesis on a plasmid after random hexamer primers are annealed [358]. This continually generates ssDNA and due to the presence of a large excess of random hexamer primers, the ssDNA being created is constantly being primed, creating a substrate analogous to a lagging strand template, but in the absence of replisome complexes and ssDNA binding proteins. We could use this system to identify whether there is an inherent biochemical characteristic of the enzymes that promotes deamination of ssDNA generated during replication or if simply the availability of the ssDNA promoted deamination activity. We found that A3B, A3A, and A3H hap I* could all induce mutations on the replicated DNA approximately equally, suggesting that the availability of ssDNA promoted deamination (Figure 10.10A). To test this idea further we also included A3G in this assay system. Surprisingly, the mutation frequency induced by A3G was 4-fold lower than A3B and suggested that the A3s able to induce mutations during cancer did share a similar biochemical characteristic that was not common to A3G (Figure 10.10A).

To identify the difference between A3G and the other A3s tested, we used a more defined in vitro system with oligonucleotide ssDNA substrates. We first focused on A3A and A3G since we have previously conducted a side-by-side characterization of these enzymes [119]. The major differences between the two enzymes are that A3A is not processive and A3G is processive, A3A has a K_d in the micromolar range and A3G has a K_d in the nanomolar range, and A3A cycles through substrates more frequently than A3G. The processivity did not appear to be a differentiating factor, since A3B and A3H hap I* are processive and are similar to A3A in this assay (Figure 10.5, Figure, 10.8B, and Figure 10.10A). However, A3G has a long half-life on ssDNA, ranging from 3 to 5 min in contrast to A3A that cycles on and off ssDNA rapidly [72, 99, 102, 119, 265, 359]. This difference between A3A and A3G suggested that cycling on and off ssDNA is a determinant in the efficiency of deamination during DNA replication. To test this we interrogated the ability of the enzymes to cycle through different amounts of unlabeled ssDNA substrate to find and deaminate motifs on a labeled ssDNA substrate (Figure 10.10B). First the enzymes were incubated with 5x or 50x excess unlabeled DNA substrate or no unlabeled DNA substrate (0x), in comparison to the amount of labeled substrate that would be added to the reaction. After an incubation period, the labeled substrate was added to monitor deamination

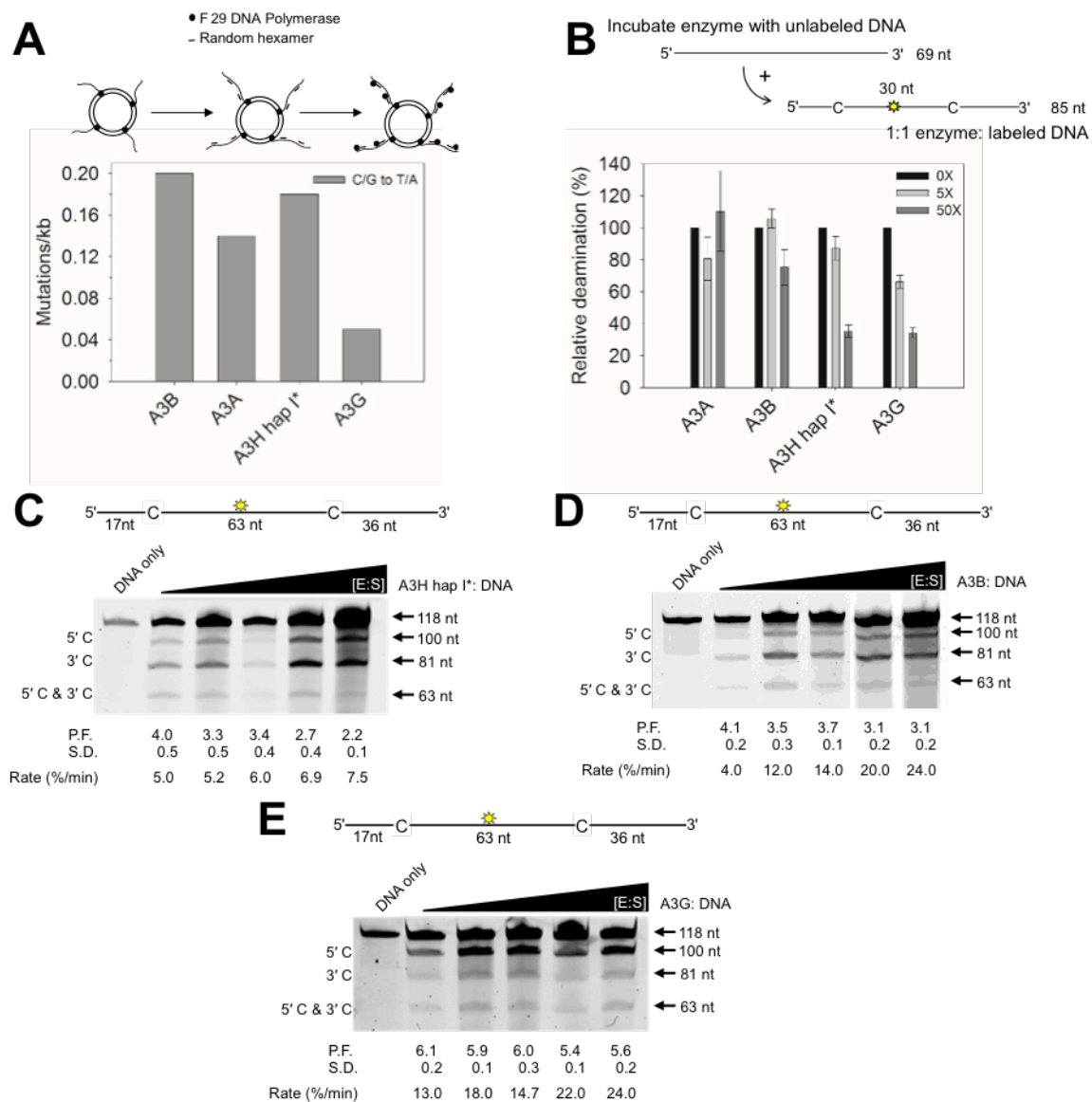


Figure 10.10. A3-mediated mutagenesis during DNA replication requires enzyme cycling. (A) A minimal phage replication system was used where the DNA polymerase of phage ϕ 29 can initiate rolling circle replication and strand displacement synthesis in the presence of random hexamer primers (sketch). The bar graph depicts the number of mutations induced by each A3 per kb sequenced. (B) The ability of an A3 enzyme to cycle through different ssDNA substrates was measured by adding increasing amounts of unlabeled ssDNA relative to the labeled ssDNA (0x, 5x, or 50x). The labeled ssDNA had two deamination motifs separated by 30 nt for A3A (5'TTC), A3B (5'ATC), A3H hap I* (5'CTC), or A3G (5'CCC). (C-E) Intersegmental transfer ability of A3H hap I*, A3B, and A3G were determined by keeping an A3:ssDNA ratio constant, but increasing the total reaction components. The ssDNA substrate contained a fluorescein-labeled deoxythymidine (yellow star) between two deamination motifs for A3A (5'TTC), A3B (5'ATC), A3H hap I* (5'CTC), or A3G (5'CCC) separated by 63 nt. The measurements of processivity factor (P.F.), standard deviation (S.D.), and reaction rate (%) from three independent experiments are shown below the gel.

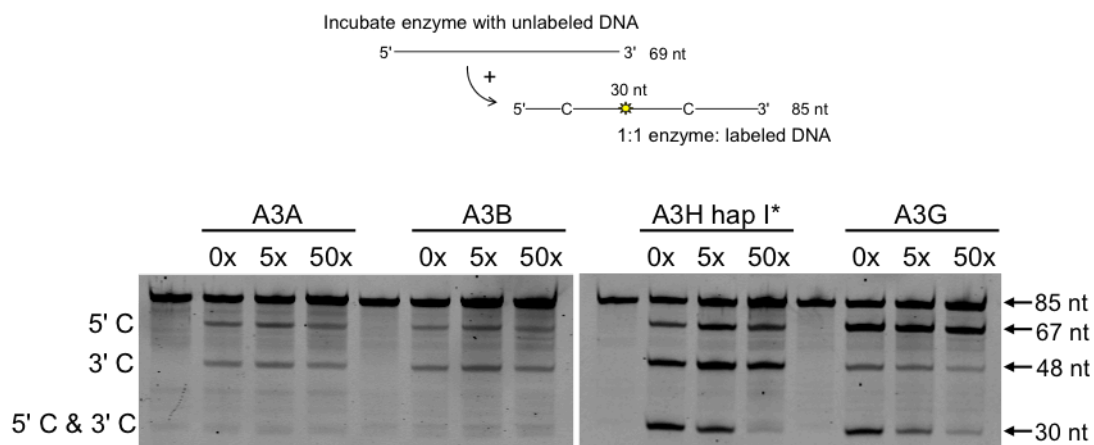


Figure 10.11. Deamination activity of A3 enzymes in the absence or presence of unlabeled DNA. The enzymes were preincubated with 0x, 5x, or 50x unlabeled DNA before the addition of the labeled ssDNA substrate. Deamination reactions were allowed to proceed for 10 min. Representative gels from the experiments that were used to calculate the deamination as presented in Figure 10.10B. The labeled ssDNA had two cytosine motifs separated by 30 nt with 5'TTC for A3A, 5'ATC for A3B, 5'CTC for A3H hap I* and 5'CCC for A3G. On the 85 nt labeled substrate, single deaminations of the 5'C & 3'C are detected as the appearance of labeled 67- and 48- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 30 nt labeled fragment.

activity. Consistent with A3A cycling frequently, the 5x or 50x of unlabeled ssDNA did not diminish the deamination activity (Figure 10.10B and Figure 10.11). Consistent with increased concentrations of ssDNA making molecular collisions more favorable, the A3A activity increased up to ~1.5-fold at 50x concentration of unlabelled substrate DNA, although this value was highly variable (Figure 10.10B). In contrast, A3G deamination activity was increasingly inhibited by the A3G getting “trapped” in the pool of unlabeled ssDNA substrate and not cycling often enough to find and deaminate the labeled ssDNA substrate before the end of the reaction (Figure 10.10B and Figure 10.11). In this cycling assay A3B was similar to A3A and only had a 20% decrease in relative deamination activity even in the presence of 50x unlabeled ssDNA (Figure 10.10B and Figure 10.11). A3H hap I* maintained its activity at a 5x level of unlabeled ssDNA, but could not cycle often enough to deaminate the labeled ssDNA efficiently in the presence of 50x unlabeled ssDNA (Figure 10.10B and Figure 10.11).

These data support the hypothesis that frequent enzyme cycling improves deamination activity during DNA replication. However, it seemed counter intuitive that both A3B and A3H that were processive enzymes, were cycling frequently, since this could preclude the ability to search effectively on each ssDNA. Further, these enzymes had steady state binding constants for ssDNA in the nanomolar range that were similar to A3G and in contrast to A3A (Table 10.1). To reconcile these observations we conducted an assay to further dissect the processive mechanisms of A3B and A3H hap I*. Both A3B and A3H underwent 3-dimensional processive scanning movements, but the mechanism could be jumping or intersegmental transfer (Figure 10.5C and Figure 10.7B). Intersegmental transfer ability enables an enzyme to cycle between ssDNA substrates but through doubly bound intermediates, rather than fully dissociating at each transfer. This type of movement would resolve our disparate observations of the enzymes being processive and able to cycle frequently. In the intersegmental transfer assays, the total concentration of enzyme and ssDNA is increased while keeping their ratio constant. Since intersegmental transfer makes the enzyme sensitive to the local DNA concentration, the increased concentration of reaction components will increase the molecular collisions [261, 351, 360]. This can result in a decrease in the apparent processivity or an increase in the apparent reaction rate as the concentration of reaction components is increased [2, 109, 110].

In our intersegmental transfer assay, we increased the enzyme:ssDNA concentration 5-fold using incremental steps. Under these conditions, A3H hap I* showed the distinct properties of

intersegmental transfer where there was a 2-fold decrease in the processivity factor and a ~2-fold increase in the reaction rate at the end of the titration (Figure 10.10C). These data suggest that A3H hap I* cycles from ssDNA to ssDNA rather than completely dissociating from the pool of ssDNA and then finding another ssDNA, like A3A. As a result, at the 50x concentration of unlabelled trap, despite A3H hap I* moving from ssDNA to ssDNA, the processivity on each ssDNA it encountered was low, causing a decrease in overall deamination activity (Figure 10.10B). For A3B, the enzyme did not have an apparent decrease in processivity, but it appeared that A3B still underwent intersegmental transfer to cycle from one ssDNA to another because the reaction rate increased 6-fold (Figure 10.10D). The activities of A3H hap I* and A3B in this assay were different than A3G that did not have a decrease in processivity and had only a ~2-fold increase in the reaction rate (Figure 10.10E). It has been previously shown biochemically, by single-molecule FRET and by atomic force microscopy that A3G does not undergo intersegmental transfer, but jumps and slides on ssDNA [103, 111, 253]. Based on these data, A3B appears to be very efficient in that it can scan a single ssDNA and processively deaminate cytosines while also cycling from ssDNA to ssDNA using intersegmental transfer. The A3H hap I* data is similar to other A3H haplotypes where a decrease in processivity was observed but the increase in reaction rate is less pronounced [110]. These data indicating that A3B and A3H hap I* undergo intersegmental transfer also suggest that the cooperative binding of A3B and A3H observed with rotational anisotropy could be a combination of individual enzyme molecules binding ssDNA and then promoting the binding of additional enzyme molecules on the same DNA or promoting multiple ssDNA molecules to bind each enzyme subunit in the oligomer (Table 10.1). If the latter occurs, this would be in contrast to A3G that binds ssDNA cooperatively by recruiting additional A3G subunits on the same ssDNA [72]. In conclusion, although the mode of cycling is different for A3A, A3B, and A3H hap I*, probably based in part on their different binding affinities for ssDNA (Table 10.1), this appears to be a distinct feature required for efficient mutagenesis of ssDNA created during DNA replication.

10.3.4 Enzyme cycling enables APOBEC3 enzymes to compete with RPA for ssDNA. In cells, the ssDNA exposed during replication stress would be coated with RPA [361]. It has been assumed that A3 enzymes cannot displace RPA and that they are limited to deaminating ssDNA at gaps in the RPA coating [357]. However, RPA is displaced to enable other recombination proteins to access the ssDNA, such as Rad51 [362, 363]. To determine if RPA would influence

A3 deamination activity we added RPA to an oligonucleotide ssDNA substrate at saturating concentrations and then tested A3 deamination activity and processivity. It has been shown experimentally that RPA can be competed off ssDNA by other proteins in excess through a mechanism known as facilitated dissociation, which can be caused by Rad 51 or excess RPA itself [362, 363]. Notably, the facilitated dissociation occurs when RPA is in excess and it is not displaced as readily when amounts are limited. Thus, we used an in vitro system where the amount of oligonucleotide ssDNA was constant, rather than the Φ 29 replication system where the DNA concentration would be continually increasing and precluding achieving RPA saturation on ssDNA. For this experiment with oligonucleotides, RPA was pre-incubated with ssDNA and saturation was confirmed by an electrophoretic mobility shift assay (EMSA), in which the ssDNA was completely bound by RPA (Figure 10.12). An A3 enzyme was then added to determine if RPA could be displaced. Displacement of RPA was interpreted based on the A3 deamination activity and processivity in comparison to naked ssDNA. We found that each A3 enzyme tested was able to displace RPA from the ssDNA in order to catalyze deaminations on the substrate. However, both the specific activity and processivity of the enzymes was decreased in the presence of RPA. For the A3B, A3H hap I*, and A3G, the processivity decreased 2-fold (Figure 10.13A-C). Interestingly, this is similar to when a dsDNA segment is present between the deamination motifs (Figure 10.5D and Refs [107, 110]) indicating that RPA may be acting as a roadblock to the 1-dimensional sliding motion of the enzymes. The A3A remained nonprocessive, but active (Figure 10.13D). Notably, we determined that there was a correlation between the cycling ability of the A3 enzyme and the specific activity in the presence of RPA. A3A, which cycles the most efficiently, had no change in specific activity in the absence and presence of RPA (Figure 10.10B and Table 10.2). The presence of RPA decreased the specific activities of A3B and A3H hap I* 2- to 3- fold (Table 10.2). However, A3G, an enzyme that does not efficiently cycle between substrates, due to a long residence time on ssDNA, had a 10-fold decrease in specific activity in the presence of RPA (Table 10.2) [72, 99, 102]. These data indicate that the ability to sample multiple substrates through enzyme cycling is important in the presence of the ssDNA binding protein RPA.

10.4 Discussion

The deoxycytidine deaminase activity of A3B, A3A, and A3H hap I has been implicated in having a driving role in cancer and tumor evolution by providing the cells with a diverse pool

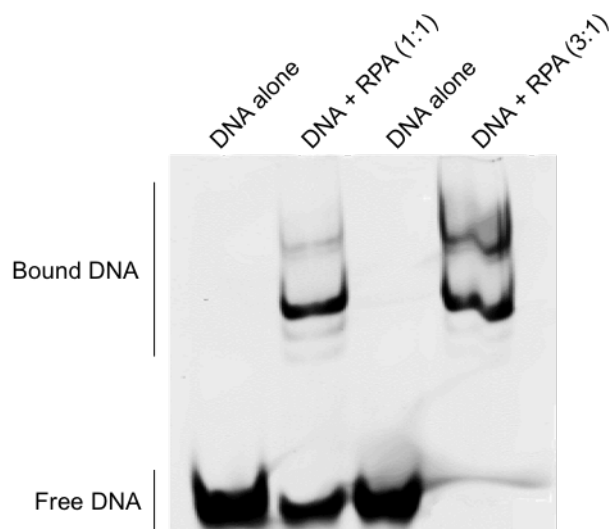


Figure 10.12. Electrophoretic mobility shift assay of RPA bound to ssDNA. RPA at a 1:1 or 3:1 (saturating) ratio to ssDNA was incubated for 5 min at 21°C before resolving the ssDNA-RPA complex on a 5% native acrylamide gel to determine the amount of RPA necessary to saturate the substrates used in Figure 10.13. The assays in Figure 10.13 used a 3:1 RPA:DNA ratio and were completely bound by RPA before the addition of the A3.

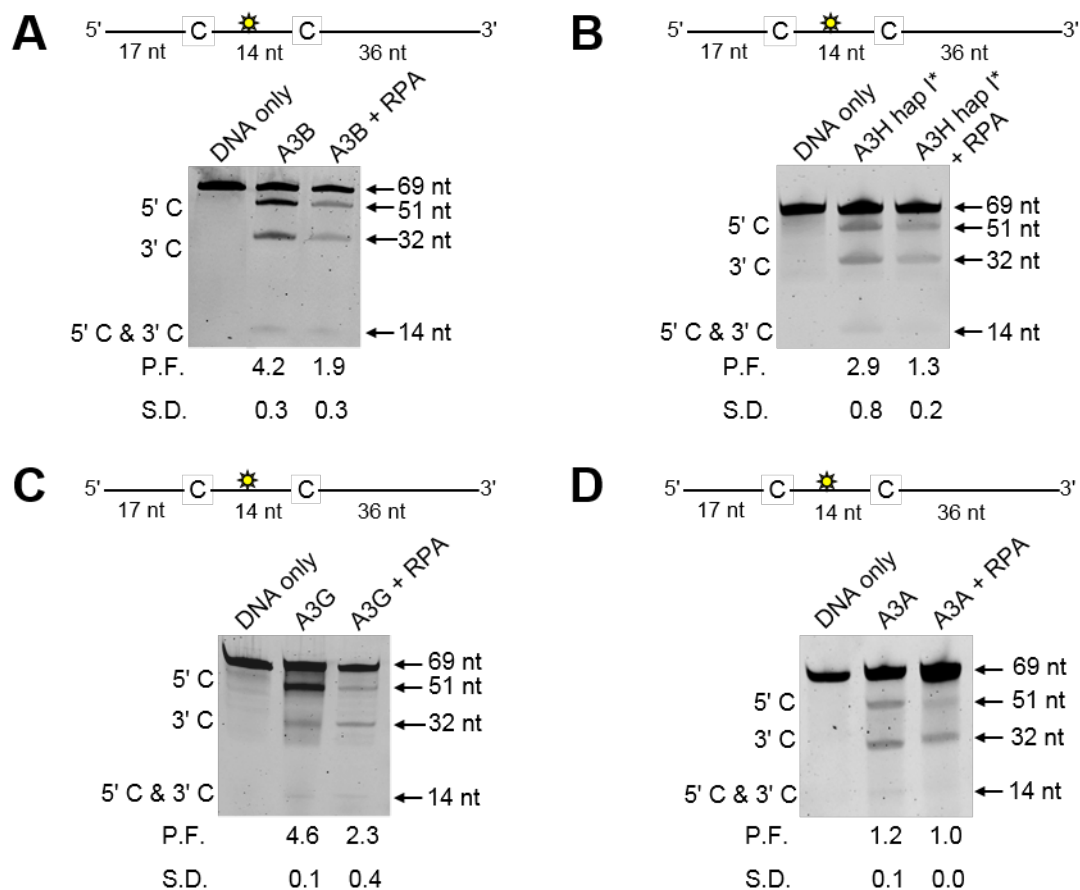


Figure 10.13. A3 enzymes can compete with RPA to deaminate ssDNA. The ability of A3 enzymes to deaminate cytosines on ssDNA in the presence of saturating amounts of RPA was examined on a fluorescently labeled (yellow star) 69 nt ssDNA with deamination motifs for A3A (5'TTC), A3B (5'ATC), A3H hap I* (5'CTC), or A3G (5'CCC) separated by 15 nt. To prepare the substrate, 100 nM of ssDNA was preincubated with 300 nM RPA (for a 3:1 ratio) and 50 nM of (A) A3B, (B) A3A, (C) A3H hap I*, or (D) A3G was added to initiate the reaction. Single deaminations of the 5'C & 3'C are detected as the appearance of labeled 51- and 32- nt fragments, respectively; double deamination of both C residues on the same molecule results in a 14 nt labeled fragment. The measurements of processivity factor (P.F.) and standard deviation (S.D.) from three independent experiments are shown below the gel.

Table 10.2. Specific activity of A3 enzymes on ssDNA in the absence or presence of RPA.

Enzyme	Specific Activity, ssDNA (pmol/μg/min)	Specific Activity, ssDNA with RPA (pmol/μg/min)
A3B	1.6 ± 0.7	0.9 ± 0.2
A3A	1.1 ± 0.1	1.3 ± 0.2
A3H hap I*	1.7 ± 0.2	0.7 ± 0.1
A3G	2.5 ± 0.2	0.3 ± 0.1

of mutations [170, 357]. Beyond analysis of mutational signatures and mRNA levels, there has been a paucity of biochemical data characterizing how these enzymes act on substrates relevant to deamination of genomic DNA [119]. Here we provide the first biochemical characterization of A3B and compare properties of A3B, A3A, and A3H hap I* on substrates relevant to catalyzing cytosine deaminations in genomic DNA. Specific to A3B, the data demonstrate that using an A3B CTD is not a suitable replacement for the fl A3B (Figure 10.1). In regards to how A3 enzymes access ssDNA in the genome during transcription or replication, we find that deamination during transcription is more selective than during replication (Figure 10.7C and Figure 10.10A). A3B was unable to deaminate during transcription and by comparison to other A3 enzymes we demonstrated that this is due to its propensity to form tetramers. Although all enzymes tested could deaminate during DNA replication or on ssDNA bound by RPA, the amount of deamination correlated with the enzyme's ability to cycle between ssDNA substrates (Figure 10.10B, Figure 10.13, and Table 10.2). These data support a model in which the inherent properties of the A3 enzyme determine the amount of A3-induced mutagenesis, not solely the availability of ssDNA.

A3 enzymes have been most extensively studied for their role in HIV-1 restriction where their cytosine deamination activity can inactivate HIV-1 proviral DNA through lethal mutagenesis [2]. In this scenario, the relevant A3 enzymes, A3D, A3F, A3G, and A3H (hap II, V, or VII), which are notably different than the ones that contribute to cancer mutagenesis, contend with a transient ssDNA substrate that is undergoing replication during the time when the enzymes must catalyze deaminations [2]. Studies from our lab have shown that a primary determinant in the mutagenic efficiency of the enzyme is their ability to scan the ssDNA efficiently for the preferred deamination motif and to processively deaminate these motifs [107-109, 150]. Although both the HIV-1 replication and genomic DNA replication and transcription processes are dynamic, they did not require the same biochemical characteristics. Processivity was not detrimental to deamination during transcription, replication, or ssDNA bound by RPA, but it was also not required as exemplified by A3A (Figure 10.7A, C, Figure 10.10A, Figure 10.13, and Table 10.2). This difference may be due to the HIV-1 reverse transcriptase replicating the viral DNA at a slower rate than the eukaryotic polymerases [364, 365]. During HIV-1 replication, an A3 enzyme can use processive ssDNA scanning to compete for ssDNA with the less processive reverse transcriptase and the rapidly cycling HIV-1 nucleic acid chaperone, nucleocapsid [63,

364]. However, in genomic DNA the RPA binds much more stably to ssDNA than nucleocapsid and the polymerases synthesize DNA faster and can use PCNA to increase processivity [361-363, 365]. As a result, the ability to more stochastically interact with ssDNA by cycling to different substrates or regions of the same substrate could provide an advantage. Processively tracking along the ssDNA with the polymerase is not possible since A3 enzymes move on ssDNA without an energy source by facilitated diffusion [2]. Rather, repetitive sampling of ssDNA regions in 3-dimensions by macroscopic cycling or intersegmental transfers may enable better access to the available ssDNA pool that would be constantly changing. There is evidence that this type of cycling can approximate the benefits of processivity for A3A [119]. Nevertheless, this conclusion was unexpected considering the number of processive DNA repair and replication proteins [353]. However, these other proteins that are normally functional in the genome usually bind and scan dsDNA, which is a key difference with A3 enzymes that scan ssDNA or, for A3H, scan ssDNA and DNA/RNA hybrids [2, 110].

Previous work has shown that A3 enzymes primarily deaminate genomic DNA during replication and favor the lagging strand of genomic DNA during replication due to the greater abundance of ssDNA from discontinuous synthesis [190-194]. However, in order to access this ssDNA, A3 enzymes would still have to displace RPA [361]. In the APOBEC family there were two opposing studies on how these enzymes could deaminate ssDNA bound by RPA. For AID, a physical interaction with RPA occurred and promoted AID-catalyzed deaminations in immunoglobulin variable genes and switch regions [366, 367]. However, A3G was characterized to have decreased specific activity and processivity in the presence of RPA [368]. The study with A3G agrees with our data (Figure 10.13C and Table 10.2) and we provide a mechanism for why A3G activity decreases in the presence of RPA. The long lifetime of the enzyme on ssDNA that is beneficial for deamination of HIV-1 proviral genomes is not conducive to competition for ssDNA with RPA (Figure 10.10B, Figure 10.13 and Table 10.3) [99, 107]. Rather the ability to cycle on and off or between substrates is required (Figure 10.10). A3 enzymes are not the only proteins that must compete with RPA for ssDNA. RPA mainly serves to protect ssDNA from nucleases and signal for subsequent processing of ssDNA if replication cannot be completed [361-363]. Thus, other proteins such as Rad51 must also be able to displace RPA. Since RPA facilitated dissociation occurs when RPA is in excess to the ssDNA [362, 363], during replication stress, where there is an increased amount of ssDNA bound by RPA, the A3A, A3B, or A3H hap

I should be able to compete with RPA for ssDNA as demonstrated in our assay (Figure 10.13 and Table 10.2). In support of this, in a study using 293T cells stably expressing A3B, the A3B-induced mutations were shown to significantly increase after DNA replication stress, suggesting that A3B can effectively compete with RPA [203].

Although A3A could deaminate ssDNA generated during active transcription and in R-loops (Figure 10.7C-D), this mechanism of accessing ssDNA in the genome is not a major source of mutations from genome wide sequencing analyses that considered the strand bias of mutations [190-194]. A3A was only found in one study to be capable of deaminating during transcription, but at a low level, which led the authors to conclude that instances where A3A can access ssDNA generated from transcription are limited [190]. A3H hap I induced mutations in cells have not been analyzed in this manner. Altogether, these data indicate that the cellular conditions prevent excessive deamination during transcription. This may be because ssDNA generated during transcription is most accessible only during extremely high levels of transcription, RNA polymerase pausing, R-loop formation, or if specific targeting mechanisms exist as they do for related family member, AID [20, 21, 186, 369]. In support of this, our study confirms that deamination during transcription and transcription related structures such as R-loops is more selective than during replication (Figure 10.7C-D and Figure 10.10). For example, a recent study in yeast found that although A3B is unable to deaminate during transcription of mRNA genes, it can deaminate during transcription of tRNA genes [370]. The ssDNA generated during tRNA gene transcription is thought to be more prone to form R-loops, which consistent with our *in vitro* data, would be required for A3B to exhibit deaminase activity (Figure 10.7C-D) [371]. However, the resulting uracils in tRNA regions are thought to be repaired faithfully and to not contribute to cancer mutagenesis [370].

Although A3B is similar in oligomeric state and processivity to the other double deaminase domain enzymes A3G and A3F, it is uniquely able to cycle through ssDNA substrates in contrast to A3G and A3F that have long lifetimes on single ssDNA substrates [99, 108] (Figure 10.10B, D, E). A3A is still unique among other A3 enzymes because despite not being processive, it has one of the highest specific activities and the highest genotoxicity when ectopically expressed in cell culture [119, 184, 196, 271]. Our studies demonstrate that this is due the ability of A3A to cycle through different substrates so that it is able to sample a larger amount

of transient ssDNA during replication or transcription (Figure 10.10B). The A3H hap I* appears to have features of both A3B and A3A that enable it to also contribute to genomic mutagenesis.

Another aspect of A3 activity not directly addressed in our study is the active site accessibility and the molecular conformation of ssDNA substrate binding. A3A activity is not only high because of its ability to cycle through different substrates, but also because of its highly accessible catalytic pocket that is normally in an open conformation [344, 372]. This is in contrast to A3B CTD that has a closed catalytic pocket [342, 344, 372]. While it has been proposed by others that the closed catalytic pocket state serves as protective measure to reduce off-target mutations by APOBEC family enzymes [188, 372], these conclusions for A3B CTD may need to be reconsidered since the fl A3B is 40-fold more active (Figure 10.1A). It remains to be determined whether increased fl A3B activity is due to different catalytic pocket accessibility in the fl A3B than A3B CTD, if the NTD of A3B is better able to guide the ssDNA substrate into the catalytic pocket, if increased processivity enables a more efficient search for potential deamination motifs, or a combination of all of these factors. Recently, it has been shown that A3A and A3B CTD bend the ssDNA substrate in the active site [344, 373, 374]. How this relates to processivity and if the enzymes can induce the ssDNA bending of larger substrates, such as dynamic ssDNA exposed during replication or transcription remains to be studied. However, it is interesting to speculate on this based on single-molecule data from a study with A3G. It was found that A3G scanned ssDNA very rapidly and bi-directionally, but when A3G encountered a deamination motif, the enzyme movement slowed down and A3G “hovered” over the deamination motif [253]. Perhaps DNA bending for A3G, which has been observed in single-molecule and bulk studies, serves to slow down the diffusion to enable deaminations to take place [102, 253]. If this is true for all processive A3 enzymes, including A3B, then there may be different ssDNA interactions that occur for either processivity (fast diffusion) or deamination (slow diffusion and DNA bending). For A3A, perhaps the DNA bending creates a more stable interaction to offset the low affinity interaction with ssDNA. These molecular features of A3H are understudied compared to A3A, A3B, and A3G and there are many aspects of its interaction with ssDNA that still require further study.

Altogether the data presented here demonstrate not only a system for purifying sufficient quantities of fl A3B for biochemical studies, but provide a number of biochemical systems for mechanistic study of A3 enzymes on DNA substrates relevant to mutagenesis of genomic DNA.

The studies presented here further our understanding of the dynamics of how A3 enzymes can so efficiently induce deaminations that are actually “off-target” events and not part of their normal cellular functions.

10.5 Materials and Methods

10.5.1 Cloning and site directed mutagenesis. The fl A3B (a.a. 1-382), A3B 187 CTD (a.a. 187-382), and A3B 193 CTD (a.a. 193-382) sequences, as well as A3G and A3H haplotype I* were cloned into a baculovirus transfer vector (pAcG2T or pFAST-bac1) containing an N-terminal GST tag as described previously[110, 375]. All constructed plasmids were verified by DNA sequencing.

10.5.2 Protein expression and purification. Recombinant baculovirus production for expression of proteins in *Sf9* cells was carried out using the pACG2T or pFast-bac1 transfer vector as previously described[110, 375]. *Sf9* cells were infected with recombinant GST-A3 virus at an MOI of 1 (A3B 187 CTD and A3B 193 CTD) or an MOI of 10 (fl A3B). The infection conditions for A3A, A3G, and A3H have been previously described[99, 110, 119]. Recombinant baculovirus infected *Sf9* cells were harvested after 72 hours of infection. Cells lysates treated with RNaseA were incubated with glutathione-Sepharose 4B resin (GE Healthcare) at 4°C and were subjected to a series of salt washes, as previously described[99]. For all the enzymes, except the full-length A3B, on-column cleavage from the GST tag with thrombin (GE Healthcare) was performed at 21 °C for 18 h in thrombin digestion buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM DTT). Due to the lower yield of full-length A3B, the enzyme was eluted with the GST tag in elution buffer (100 mM Tris, pH 8.8, 150 mM NaCl, 10% (v/v) glycerol, and 50 mM reduced glutathione). A3B was then cleaved from the GST tag in solution at 21°C for 4 hours with Thrombin (GE Healthcare) before being dialyzed against 100 mM Tris pH 7.5, 250 mM NaCl, 10% (v/v) glycerol, and 1 mM DTT. To purify A3B from the free GST and thrombin, the enzyme stock was diluted to achieve a solution of 50 mM Tris pH 7.5, 50 mM NaCl, 10% (v/v) glycerol and 1 mM DTT for loading onto a DEAE FF column (GE Healthcare). A3B was eluted with a linear gradient of NaCl. The A3B eluted at approximately 300 mM NaCl. Enzymes were assessed as >90% pure by SDS-PAGE (Figure 9.14). The RPA was generously provided by Dr. John Turchi (Indiana University School of Medicine).

10.5.3 In vitro deamination assays. All ssDNA substrates were obtained from Tri-Link Biotechnologies and have been previously described or are in Table 9.3 [108, 110]. Substrates with complementary oligonucleotides annealed were formed by mixing the oligonucleotides in buffer containing 50 mM Tris, pH 7.5 and 100 mM NaCl before heating at 95 °C for 10 min and then slowly cooling (1 °C per minute). For all deamination assays, except the R-loop, the substrates contained specific motifs for each A3 enzyme which were: 5'TTC (A3A), 5'ATC (A3B), 5'CTC (A3H hap I*), or 5' CCC (A3G). The R-loop substrates used only 5' TTC (A3A, A3B, and A3H hap I*) or 5' CCC (A3G). Unless otherwise indicated, reactions were conducted at 37 °C in RT buffer (50 mM Tris, pH 7.5, 40 mM KCl, 10 mM MgCl₂, and 1 mM DTT).

Processivity reactions were carried out under single-hit conditions (i.e. <15% substrate usage) to ensure a single enzyme-substrate encounter. Under these conditions, a processivity factor can be determined by comparing the quantified total amount of deaminations occurring at two sites on the same ssDNA with a calculated theoretical value of deaminations at these two sites if the deamination events were uncorrelated (not processive). Since the processivity factor is a ratio, a value of 1.0 means that the enzyme is not processive. Alternatively, a non-processive enzyme may not have a visible amount of deamination at two sites under the single-hit conditions of the reaction. A ssDNA substrate containing two deamination motifs (100 nM) was incubated with 50 nM of enzyme for 2.5 to 20 min. The reaction time was varied on each ssDNA according to the specific activity of the enzymes to ensure ~10% substrate usage. Reactions were started by the addition of the ssDNA substrate. The specific activity was calculated from these single-hit condition reactions by determining the picomoles of substrate used per minute for a microgram of enzyme.

For the time course of A3B in comparison to the CTD mutants, 100 nM of a 118 nt ssDNA was incubated with 50 nM (A3B) or 2 µM of the CTD mutants for 5-60 min (A3B) or 30-120 min (CTD mutants) in order to compare their activities. Reactions were started by the addition of the ssDNA substrate.

For intersegmental transfer assays, the A3:ssDNA ratio was kept the same, but increasing concentrations of enzyme and substrate was titrated in (118 nt ssDNA: 100-500 nM, A3: 50-250 nM). Reactions were started by the addition of the ssDNA substrate.

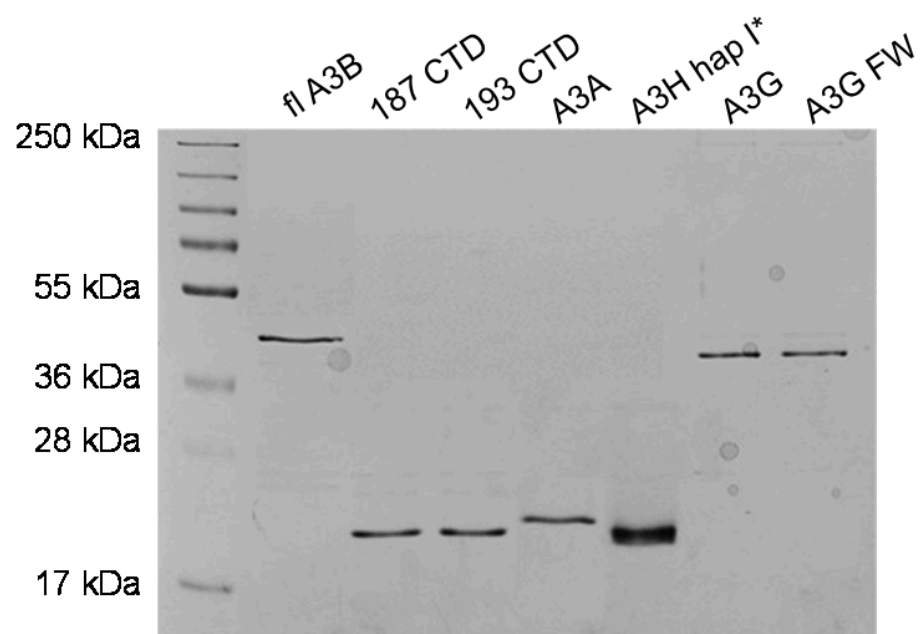


Figure 9.14. SDS-PAGE of purified A3 enzymes. A3 enzymes were purified as detailed in the Materials and Methods and resolved on a SDS-PAGE gel to estimate purity.

Table 9.3. DNA substrates

Name	Sequence
60 nt ATC with deamination motifs 5 nt apart	5' AAA GAG AAA AGT ATA ATC A{dT-FAM}A ATC ATA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
69 nt ATC with deamination motifs 14 nt apart	5' AAA GAG TTA GGG TGA ATC AAA ATT {dT-FAM}AA AGA ATC AAA TGT TAG ATA TGT TAA TGT GTG TGA TGA TGT TGA
85 nt ATC with deamination motifs 30 nt apart	5' AAA GAG AAA GAG TAA ATC AAA GAG TAA AGT {dT-FAM} AAG TAG AGA GAT TAT ATC AAA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
118 nt ATC with deamination motifs 63 nt apart	5' GAA TAT ATG AGT TGA ATC AAA GTA ATG AGA GAG AAT {dT- FAM} TAG ATG AGT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA ATC AAA GAG TAA AGT TGT TAA TGT GTG TAG ATA TGT TAA
5' TTC R-loop top strand	5' {FAM} AAA GAG AAA GTA ATA AGG AAA GAG TAT TAA ATA TTC AAA ATT TCA ATC ATT CTA CAC ATT CAT ACA ATT T
5' CCC R-loop top strand	5' {FAM} AAA GAG AAA GTA ATA AGG AAA GAG TAT TAA ATA CCC AAA ATT TCA ATC ATT CTA CAC ATT CAT ACA ATT T
Bottom strand (R-loop)	5' A AAT TGT ATG AAT GTG TAG AAT GAT TGT TTA AAA TGA ATA ACT TTA CTC TTT CCT TAT TAC TTT CTC TTT
Complementary RNA (R-loop)	5' rUrUrArUrUrCrArUrUrUrU
69 nt unlabeled ssDNA for cycling assay	5' AAA GAG AAA GTA ATA AGG AAA GAG TAA AGT ATA ATC AAA TAA ACA ATC ATT CTA CAC ATT CAT ACA ATT

For the detection of deaminations on R-loop substrates 100 nM of an R-loop substrate was incubated with 100 nM A3 enzyme for 5-20 min. Reactions were started by the addition of the ssDNA substrate.

For cycling assays, 100 nM of A3A, A3B, A3H hap I*, or A3G was incubated for 3 min at 21 °C with 100 nM of an unlabeled 69 nt ssDNA. To start the reaction, 100 nM of a labeled 118 nt ssDNA was added and the reaction was allowed to proceed at 37 °C for 10 min.

For deamination in the presence of RPA, saturating amounts of RPA (300 nM) was preincubated with 100 nM of ssDNA for 5 min before the addition of 50 nM A3 enzyme to initiate the reaction. Reactions proceeded for 10-30 min.

All A3-catalyzed deaminations were stopped using a Phenol: chloroform extraction and cleaned using two additional chloroform extractions. The deaminations were detected by treating the substrates with uracil DNA glycosylase (New England Biolabs) and heating under alkaline conditions before resolving the fluorescein-labeled ssDNA on 10, 16 or 20% (v/v) denaturing polyacrylamide gels, depending on the sizes of the ssDNA fragments. Gel photos were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare) and integrated gel band intensities were analyzed using ImageQuant (GE Healthcare).

10.5.4 Electrophoretic mobility shift assay. A fluorescently labeled 70 nt ssDNA (100 nM) was incubated with 300 nM of RPA for 5 min at 21°C in the presence of RT buffer and 40% glycerol to facilitate gel loading. The samples were run on a 5% (v/v) Native-PAGE gel at 4°C. Gel images were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare).

10.5.5 Size exclusion chromatography. The oligomerization states of the enzymes were determined by loading 10 µg of purified enzyme on a 10 mL Superdex 200 (GE Healthcare) size exclusion column. The column was prepared by pouring the resin bed in a column with 16 cm height and 0.5 cm diameter. The running buffer contained 50 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT. The Bio-Rad standard set was used to generate a standard curve from which molecular weight and oligomerization states of the enzymes were determined.

10.5.6 In vitro transcription assay. Transcription-dependent deamination reactions were performed as described previously[354, 375]. The dsDNA substrates contain a T7 RNA polymerase promoter and either a single 5'TTC (A3A, A3B), 5'CTC (A3H hap I*), or 5'CCC (A3G, A3G FW) motif on the non-transcribed strand. Substrates were previously reported, except

for the 5'CTC substrate, but the DNA sequences are identical other than the deamination motif[119]. The dsDNA substrates (30 nM) were reacted with 90 nM A3A, A3B, A3H hap I*, A3G, or A3G FW in the presence of ribonucleotide triphosphates (500 μ M), T7 polymerase (1 unit, Promega), and DNase-free RNase A (5 ng/ μ l, Roche Applied Science) in transcription buffer (50 mM Tris, pH 7.4, 10 mM MgCl₂, 1 mM DTT) at 37 °C. Control reactions either contained no T7 Polymerase or no APOBEC3 enzyme to ensure that deamination was transcription dependent. Sequencing of the non-transcribed strand for detection of deamination was performed using Thermo Sequenase (Affymetrix) as described previously[354].

10.5.7 In vitro DNA replication assay. Deamination of ssDNA by A3 enzymes during DNA synthesis was assessed using a circular template (pUC19). The pUC19 was heated and slowly cooled to facilitate annealing of random hexamer primers (New England Biolabs). The primed pUC19 was then incubated with Φ 29 DNA polymerase that initiates rolling circle replication and strand displacement synthesis on subsequent DNA synthesis rounds. Specifically, the circular template was mixed with 4 μ g BSA, 200 μ M dNTPs, 200 nM random hexamer primers, and Φ 29 buffer (Thermo Fisher) for DNA synthesis in the absence or presence of 200 nM A3 enzyme. The reaction was incubated at 30°C for 6 hr, and inactivated at 65°C for 10 min before treatment with DpnI (New England Biolabs) for 1 hr at 37°C to remove contaminating plasmid DNA. A 345 nt region of the plasmid (nucleotides 473-818) was PCR amplified using Pfu C_x Turbo Hotstart (Agilent Technologies) that can use uracils as a template with high fidelity. These amplicons were then cloned with the CloneJET PCR cloning kit (Thermo Fisher). At least 25 clones were sequenced with primers specific to pUC19 at the National Research Council of Canada (Saskatoon, Canada).

10.5.8 Steady state rotational anisotropy. Steady state fluorescence depolarization (rotational anisotropy) was used to measure the binding affinity of the enzymes to fluorescein-labeled ssDNA, DNA/RNA, R-loop, and dsDNA. Reactions were 60 μ L and contained fluorescein-labeled DNA substrate (10 nM) in RT buffer and the enzyme was titrated into the solution until saturation. A QuantaMaster QM-4 spectrofluorometer (Photon Technology International) with a dual emission channel was used to collect data and calculate anisotropy. Samples were excited with vertically polarized light at 495 nm (6 nm band pass) and vertical and horizontal emissions were measured at 520 nm (6 nm band pass). Apparent dissociation constants

(K_d) were obtained by fitting to a rectangular hyperbola or sigmoidal curve using Sigma Plot 11.2 software.

11.0 General Discussion and Future Directions

The main focus of this presented work was to characterize mechanistically how A3 enzymes find their ssDNA substrate in different biological contexts such as during HIV proviral DNA synthesis and the “off-target” genomic DNA substrate. We also examined how weakly restrictive A3s may compensate for their reduced activity by modulating RT processes. The major conclusion of this work is that these enzymes must be able to bind to and efficiently deaminate transiently available single-stranded DNA during reverse transcription, replication, or transcription. A summary of these processes is highlighted in Figure 11.1. Specific biochemical characteristics promote binding and deamination in each situation to increase enzyme efficiency through processivity, rapid enzyme cycling between substrates, or oligomerization state. Interestingly, what makes the A3 enzymes able to effectively find these deamination targets differs between the ssDNA available during HIV replication and the ssDNA available in the nucleus. The biochemical differences in A3 enzymes involved in viral restriction versus genomic damage may allow for a prediction tool for the role of the A3 being studied.

For HIV restriction, we examined the ability of A3C to utilize facilitated diffusion to efficiently search for ssDNA substrates [104, 105]. Examining the facilitated diffusion mechanisms of the enzymes in our deamination assays allows for determination of the processive nature of the A3s. We examined whether the enzymes can use one-dimensional (sliding) or three-dimensional (jumping and intersegmental transfer) movements. Prior examination of these mechanisms has provided insight on how A3s access ssDNA during reverse transcription, and enables prediction of restriction capacity. Enzymes that possess multiple ssDNA scanning mechanisms (i.e., both one-dimensional and three-dimensional processes) are able to more effectively search for ssDNA in the finite amount of time that it remains single stranded [107-111, 119].

Work in our lab has previously shown that processivity of the A3s is mediated by specific residues on either loop 7 or helix 6 [107, 108, 110]. Mutating residues on or near these regions allowed for enzymes like A3F, which is unable to slide, to gain this processive mechanism [108]. Similarly, mutating residues on these regions on processive A3s like A3G and A3H decreased the processive action of these enzymes [107, 110]. However, when we applied this knowledge to the study of A3C, we found this to not hold true. Human A3C is a weakly processive enzyme and is not able to restrict HIV in infectivity assays to a high level. In collaboration with the Emerman

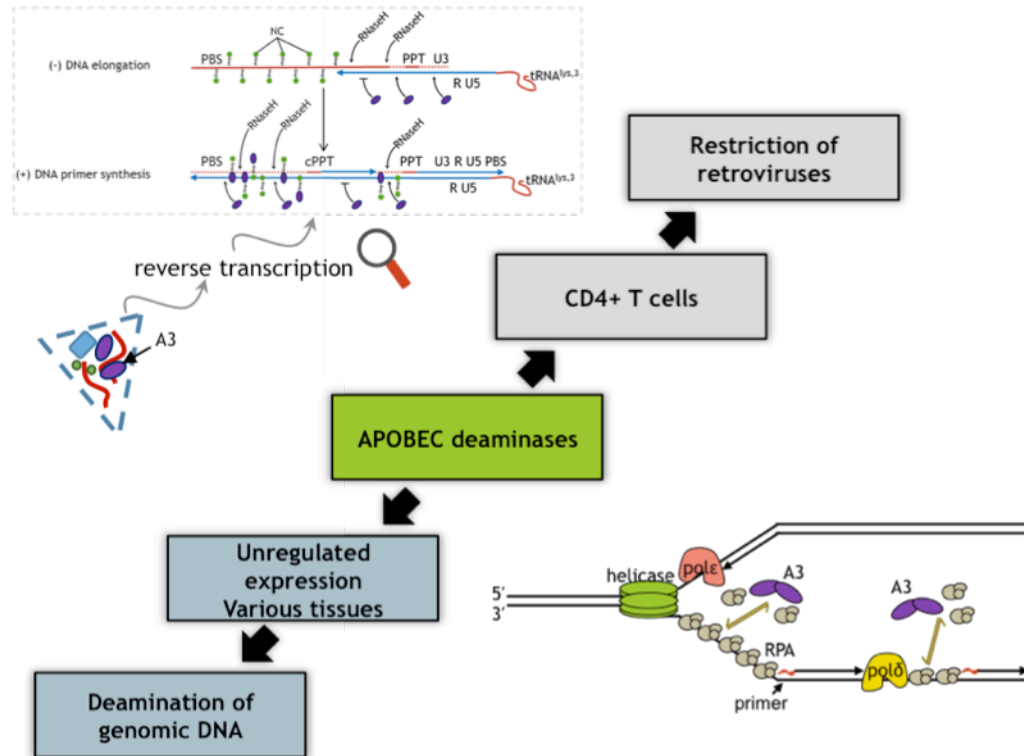


Figure 11.1 Model for A3 activity on diverse ssDNA substrates. As an innate immune factor in CD4+ T cells, A3 enzymes that are encapsidated into HIV virions can access and deaminate viral ssDNA during reverse transcription. Here, the A3 (purple) is able to displace the weakly associated nucleocapsid (NC, green) and is free to act processively to find target cytosines. Processive deamination ability of an A3 enzyme correlates with its mutagenic ability on HIV (-) DNA. When unregulated expression of A3s occurs in various tissues, certain A3s can deaminate genomic DNA. In this context, the A3 (purple) must displace the tightly associated RPA (grey) by efficiently cycling on and off ssDNA on the lagging strand of the replication fork, which can promote facilitated dissociation of RPA. In genomic DNA, processivity of A3 enzymes is not required.

lab, we discovered that a SNP existed in a subset of the human population that improved the restriction activity of A3C, namely hA3C S188I [27]. This SNP existed on helix 6, so it was initially expected to improve the processive mechanisms of the enzyme and therefore increase the mutagenic potential. However, we discovered that the improvement in restriction capacity observed was due to hA3C S188I being able to dimerize. Interestingly, chimpanzee and gorilla, which also possessed the less active residue at position 188, had higher processivity factors and increased restriction capacity despite this inactivating residue. This suggested that a separate residue was also important for the activity of A3C. The chimpanzee and gorilla A3Cs were also able to form dimer populations. We identified this residue involved in activity and dimerization of chimpanzee and gorilla A3C as amino acid 115 on β -strand 4. Mutation of this residue in hA3C to the chimpanzee/gorilla form stabilized the dimer in hA3C and further improved the activity. The converse was true for chimpanzee and gorilla, that processive movements could be reduced when the dimers were no longer present. We therefore concluded that dimerization alone was sufficient to improve the processivity of A3C, unlike the other A3s in which specific residues mediated the ssDNA scanning mechanisms.

We had also identified in studies with A3H, that residues along helix 6 allowed the enzyme to undergo intersegmental transfer, a type of three-dimensional translocation [110]. This type of movement requires a doubly-bound state, and therefore single domain enzymes such as A3H and A3C would require oligomerization in order to facilitate this type of movement. It is interesting to note that initial assumptions that enzymes that had two DNA binding domains could undergo intersegmental transfer are not entirely correct, as double-domain enzymes A3G and A3F are not observed to use this mechanism [110]. Therefore, the exact mechanics of intersegmental transfer remain to be elucidated. We examined intersegmental transfer through an assay in which the ratio of enzyme to ssDNA remains constant, but the concentration of these components increases. By crowding the reaction, intersegmental transfer can be promoted, as the presence of increased amounts of substrate will make an enzyme more likely to transfer to a distal site on the same substrate, or a second substrate than remain on the same site. We conclude that an enzyme is able to undergo intersegmental transfer if the processivity factor decreases, the reaction rate increases, or both. For A3C, intersegmental transfer required a monomer-dimer equilibrium, and A3C enzymes that were stable dimers were unable to undergo these types of movements, likely due to conformational restrictions. This is unique to A3C, as A3H is a stable

dimer and can undergo intersegmental transfer [110]. Further, intersegmental transfer in A3H can be disrupted through a single residue on helix 6, without disruption of dimerization [110].

Despite processive movements on ssDNA, A3s like A3F and A3C I188 still restrict HIV less well than A3G or A3H. Therefore, it is important for A3s to have multiple mechanisms in order to ensure that the virus is completely inactivated. We had previously demonstrated that A3G and A3F can delay RT initiation and decrease in the production of full-length products by [71, 74]. Therefore, it appears that the A3s are creating a roadblock to RT polymerization. This roadblock requires RT to circumvent the block in order to fully extend the template, and is observed as an increase in template switching. Similarly, this competitive binding was shown to hinder the proper sampling of dNTPs by RT, leading to an increase in mutation errors by RT. While alone these observations may not lead to drastic changes in the restriction of the virus, in combination with the mutational load imposed by the A3 or other A3s additively, it is expected the virus will be inactivated. The results observed in Chapter 7 also emphasize the importance of complete inactivation of the virus, as changes in template switching and mutational frequency if not accompanied by virus inactivation may lead to viral recombination and evolution. The effect of the A3s on these processes of RT is still not entirely clear, and more work on these deamination-independent mechanisms is warranted.

While processivity and oligomerization of the A3s on the ssDNA during HIV replication leads to higher levels of restriction, we found that these determinants of activity could not be extended to the more dynamic processes of replication and transcription of genomic DNA. In these systems, the environment of the replisome and the transcription bubble is much more competitive, and the ssDNA is available for very limited amounts of time. Therefore, maintaining processivity on these substrates would limit the ability to induce multiple deaminations. Interestingly, A3 enzymes that formed oligomers on ssDNA (A3G, A3B) were unable to deaminate within a moving transcription bubble. This suggests that while oligomerization may impart more processive movements for A3 enzymes in HIV restriction, this same mechanism excludes the A3s from the transcription complex. We found that A3s that are able to cycle readily through substrates have higher levels of observed deaminations in our model system of replication. This cycling mechanism also facilitated the A3 enzymes ability to displace bound RPA from the DNA, and A3G, which remains bound to ssDNA for longer periods of time and cannot cycle, was not able to compete with RPA as readily as the A3 enzymes that could cycle.

These observations also seem to suggest that A3s are more likely to induce mutations during replication breakdown, where replication stress creates more single-stranded due to stalled and misfired replication forks. While the data presented in Chapter 9 provides initial insight into how A3s are able to access and deaminate ssDNA during transcription and replication, further studies are required that utilize more of the replisome components in addition to RPA, as the replication fork is a complex and dynamic system. More complex assays including additional replication proteins such as Rad51, and assays in cells examining the dynamic replisome are required.

Despite different functions and substrates, A3 enzymes share many similar features with other DNA binding and modification enzymes. For example, many DNA repair enzymes scan DNA rapidly by facilitated diffusion and slow down to hover over their target motif or lesion, much like what has been observed for A3G [253, 353, 376, 377]. Where a DNA repair enzyme is charged with being an efficient enzyme that searches dsDNA for lesions that need to be repaired, A3 enzymes must search for deamination motifs before the ssDNA is covered up by the complementary strand or other DNA binding proteins. Considering the A3A, A3B, and A3H hap I data indicating that some A3 enzymes can actively compete with RPA and are not just opportunistic in how they access ssDNA in the genome, perhaps more parallels can be drawn between A3 enzymes and proteins involved in genomic DNA binding and modification. Merging the insights from A3 biochemical studies and the existing understanding of DNA replication proteins may offer key insights into mechanisms of A3-induced somatic mutagenesis.

12.0 References

1. Adolph, M.B., R.P. Love, and L. Chelico, *Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates*. ACS Infect Dis, 2018. **4**(3): p. 224-238.
2. Feng, Y., et al., *Suppression of APOBEC3-mediated restriction of HIV-1 by Vif*. Front Microbiol, 2014. **5**: p. 450.
3. Conticello, S.G., *The AID/APOBEC family of nucleic acid mutators*. Genome Biol, 2008. **9**(6): p. 229.
4. Muramatsu, M., et al., *Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells*. J Biol Chem, 1999. **274**(26): p. 18470-6.
5. Salter, J.D., R.P. Bennett, and H.C. Smith, *The APOBEC Protein Family: United by Structure, Divergent in Function*. Trends Biochem Sci, 2016. **41**(7): p. 578-94.
6. Sharma, S., et al., *APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages*. Nat Commun, 2015. **6**: p. 6881.
7. Sharma, S., et al., *The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme*. Sci Rep, 2016. **6**: p. 39100.
8. Ohtsubo, H., et al., *APOBEC2 negatively regulates myoblast differentiation in muscle regeneration*. Int J Biochem Cell Biol, 2017. **85**: p. 91-101.
9. Sato, Y., et al., *Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy*. J Biol Chem, 2010. **285**(10): p. 7111-8.
10. Severi, F., A. Chicca, and S.G. Conticello, *Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function*. Mol Biol Evol, 2011. **28**(3): p. 1125-9.
11. Peled, J.U., et al., *The biochemistry of somatic hypermutation*. Annu Rev Immunol, 2008. **26**: p. 481-511.
12. Suspène, R., et al., *Self-cytoplasmic DNA upregulates the mutator enzyme APOBEC3A leading to chromosomal DNA damage*. Nucleic Acids Res, 2017. **45**(6): p. 3231-3241.
13. Friedberg, E.C., *DNA repair and mutagenesis*. 2nd ed. 2006, Washington, DC: ASM Press. xxix, 1118 p.
14. Hoopes, J.I., et al., *Avoidance of APOBEC3B-induced mutation by error-free lesion bypass*. Nucleic Acids Res, 2017. **45**(9): p. 5243-5254.
15. Koito, A. and T. Ikeda, *Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases*. Front Microbiol, 2013. **4**: p. 28.
16. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer*. Nature, 2013. **500**(7463): p. 415-21.
17. Burns, M.B., N.A. Temiz, and R.S. Harris, *Evidence for APOBEC3B mutagenesis in multiple human cancers*. Nat Genet, 2013. **45**(9): p. 977-83.
18. Harris, R.S., et al., *AID is essential for immunoglobulin V gene conversion in a cultured B cell line*. Curr Biol, 2002. **12**(5): p. 435-8.
19. Muramatsu, M., et al., *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme*. Cell, 2000. **102**(5): p. 553-63.
20. Lada, A.G., et al., *Disruption of Transcriptional Coactivator Sub1 Leads to Genome-Wide Re-distribution of Clustered Mutations Induced by APOBEC in Active Yeast Genes*. PLoS Genet, 2015. **11**(5): p. e1005217.

21. Taylor, B.J., Y.L. Wu, and C. Rada, *Active RNAP pre-initiation sites are highly mutated by cytidine deaminases in yeast, with AID targeting small RNA genes*. Elife, 2014. **3**: p. e03553.
22. Revy, P., et al., *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)*. Cell, 2000. **102**(5): p. 565-75.
23. LaRue, R.S., et al., *The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals*. BMC Mol Biol, 2008. **9**: p. 104.
24. Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger, *AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification*. Nature, 2002. **418**(6893): p. 99-103.
25. Jern, P. and J.M. Coffin, *Host-retrovirus arms race: trimming the budget*. Cell Host Microbe, 2008. **4**(3): p. 196-7.
26. Kidd, J.M., et al., *Population stratification of a common APOBEC gene deletion polymorphism*. PLoS Genet, 2007. **3**(4): p. e63.
27. Duggal, N.K., et al., *Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations*. Virology, 2013. **443**(2): p. 329-37.
28. Wittkopp, C.J., et al., *A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses*. PLoS Pathog, 2016. **12**(10): p. e1005865.
29. Kharsany, A.B. and Q.A. Karim, *HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities*. Open AIDS J, 2016. **10**: p. 34-48.
30. Coffin, J.M., S.H. Hughes, and H.E. Varmus, *Retroviruses*. Retroviruses, ed. J.M. Coffin, S.H. Hughes, and H.E. Varmus. 1997, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. 843.
31. Lu, K., X. Heng, and M.F. Summers, *Structural determinants and mechanism of HIV-1 genome packaging*. J Mol Biol, 2011. **410**(4): p. 609-33.
32. Moore, M.D. and W.S. Hu, *HIV-1 RNA dimerization: It takes two to tango*. AIDS Rev, 2009. **11**(2): p. 91-102.
33. Peterlin, B.M. and D. Trono, *Hide, shield and strike back: how HIV-infected cells avoid immune eradication*. Nat Rev Immunol, 2003. **3**(2): p. 97-107.
34. Karn, J. and C.M. Stoltzfus, *Transcriptional and posttranscriptional regulation of HIV-1 gene expression*. Cold Spring Harb Perspect Med, 2012. **2**(2): p. a006916.
35. Collins, D.R. and K.L. Collins, *HIV-1 accessory proteins adapt cellular adaptors to facilitate immune evasion*. PLoS Pathog, 2014. **10**(1): p. e1003851.
36. Gramberg, T., N. Sunseri, and N.R. Landau, *Accessories to the crime: recent advances in HIV accessory protein biology*. Curr HIV/AIDS Rep, 2009. **6**(1): p. 36-42.
37. Malim, M.H. and M. Emerman, *HIV-1 accessory proteins--ensuring viral survival in a hostile environment*. Cell Host Microbe, 2008. **3**(6): p. 388-98.
38. Wen, X., et al., *The HIV1 protein Vpr acts to promote G2 cell cycle arrest by engaging a DDB1 and Cullin4A-containing ubiquitin ligase complex using VprBP/DCAF1 as an adaptor*. J Biol Chem, 2007. **282**(37): p. 27046-57.
39. Van Damme, N., et al., *The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein*. Cell Host Microbe, 2008. **3**(4): p. 245-52.

40. Schubert, U., et al., *CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway*. J Virol, 1998. **72**(3): p. 2280-8.
41. Desimmie, B.A., et al., *Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all*. J Mol Biol, 2014. **426**(6): p. 1220-45.
42. Arhel, N., *Revisiting HIV-1 uncoating*. Retrovirology, 2010. **7**: p. 96.
43. Aiken, C., *Viral and cellular factors that regulate HIV-1 uncoating*. Curr Opin HIV AIDS, 2006. **1**(3): p. 194-9.
44. Jacques, D.A., et al., *HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis*. Nature, 2016. **536**(7616): p. 349-53.
45. Sundquist, W.I. and H.G. Kräusslich, *HIV-1 assembly, budding, and maturation*. Cold Spring Harb Perspect Med, 2012. **2**(7): p. a006924.
46. Hu, W.S. and S.H. Hughes, *HIV-1 reverse transcription*. Cold Spring Harb Perspect Med, 2012. **2**(10).
47. Sarafianos, S.G., et al., *Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition*. J Mol Biol, 2009. **385**(3): p. 693-713.
48. Sheehy, A.M., et al., *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein*. Nature, 2002. **418**(6898): p. 646-50.
49. Apolonia, L., et al., *Promiscuous RNA binding ensures effective encapsidation of APOBEC3 proteins by HIV-1*. PLoS Pathog, 2015. **11**(1): p. e1004609.
50. York, A., et al., *The RNA Binding Specificity of Human APOBEC3 Proteins Resembles That of HIV-1 Nucleocapsid*. PLoS Pathog, 2016. **12**(8): p. e1005833.
51. Li, J., et al., *APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells*. J Mol Biol, 2014. **426**(6): p. 1296-307.
52. Yu, X., et al., *Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex*. Science, 2003. **302**(5647): p. 1056-60.
53. Jäger, S., et al., *Vif hijacks CBF- β to degrade APOBEC3G and promote HIV-1 infection*. Nature, 2011. **481**(7381): p. 371-5.
54. Zhang, W., et al., *T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction*. Nature, 2012. **481**(7381): p. 376-9.
55. Zhou, X., et al., *Characterization of the interaction of full-length HIV-1 Vif protein with its key regulator CBFbeta and CRL5 E3 ubiquitin ligase components*. PLoS One, 2012. **7**(3): p. e33495.
56. Albin, J.S. and R.S. Harris, *Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics*. Expert Rev Mol Med, 2010. **12**: p. e4.
57. Bruner, K.M., et al., *Defective proviruses rapidly accumulate during acute HIV-1 infection*. Nat Med, 2016. **22**(9): p. 1043-9.
58. Hultquist, J.F., et al., *Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1*. J Virol, 2011. **85**(21): p. 11220-34.
59. OhAinle, M., et al., *Antiretroelement activity of APOBEC3H was lost twice in recent human evolution*. Cell Host Microbe, 2008. **4**(3): p. 249-59.
60. Wang, X., et al., *Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity*. J Virol, 2011. **85**(7): p. 3142-52.

61. Rein, A., L.E. Henderson, and J.G. Levin, *Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication*. Trends Biochem Sci, 1998. **23**(8): p. 297-301.
62. Johnson, P.E., et al., *A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription*. Biochemistry, 2000. **39**(31): p. 9084-91.
63. Levin, J.G., et al., *Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription*. RNA Biol, 2010. **7**(6): p. 754-74.
64. Yu, Q., et al., *Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome*. Nat Struct Mol Biol, 2004. **11**(5): p. 435-42.
65. Liddament, M.T., et al., *APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo*. Curr Biol, 2004. **14**(15): p. 1385-91.
66. Ooms, M., et al., *HIV-1 Vif adaptation to human APOBEC3H haplotypes*. Cell Host Microbe, 2013. **14**(4): p. 411-21.
67. Mangeat, B., et al., *Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts*. Nature, 2003. **424**(6944): p. 99-103.
68. Harris, R.S., et al., *DNA deamination mediates innate immunity to retroviral infection*. Cell, 2003. **113**(6): p. 803-9.
69. Zhang, H., et al., *The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA*. Nature, 2003. **424**(6944): p. 94-8.
70. Lecossier, D., et al., *Hypermutation of HIV-1 DNA in the absence of the Vif protein*. Science, 2003. **300**(5622): p. 1112.
71. Adolph, M.B., J. Webb, and L. Chelico, *Retroviral restriction factor APOBEC3G delays the initiation of DNA synthesis by HIV-1 reverse transcriptase*. PLoS One, 2013. **8**(5): p. e64196.
72. Chaurasiya, K.R., et al., *Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid-binding protein*. Nat Chem, 2014. **6**(1): p. 28-33.
73. Iwatani, Y., et al., *Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G*. Nucleic Acids Res, 2007. **35**(21): p. 7096-108.
74. Ara, A., et al., *Mechanism of Enhanced HIV Restriction by Virion Coencapsidated Cytidine Deaminases APOBEC3F and APOBEC3G*. J Virol, 2017. **91**(3).
75. Pollpeter, D., et al., *Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G*. Nat Microbiol, 2017.
76. Janini, M., et al., *Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells*. J Virol, 2001. **75**(17): p. 7973-86.
77. Land, A.M., et al., *Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya*. J Virol, 2008. **82**(16): p. 8172-82.
78. Delviks-Frankenberry, K.A., et al., *Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation*. PLoS Pathog, 2016. **12**(5): p. e1005646.
79. Mulder, L.C., A. Harari, and V. Simon, *Cytidine deamination induced HIV-1 drug resistance*. Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5501-6.

80. Pollack, R.A., et al., *Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape*. Cell Host Microbe, 2017. **21**(4): p. 494-506 e4.
81. Grant, M. and M. Larijani, *Evasion of adaptive immunity by HIV through the action of host APOBEC3G/F enzymes*. AIDS Res Ther, 2017. **14**(1): p. 44.
82. Kim, E.Y., et al., *Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection*. PLoS Pathog, 2014. **10**(7): p. e1004281.
83. Chiu, Y.L., et al., *High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition*. Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15588-93.
84. Kozak, S.L., et al., *The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules*. J Biol Chem, 2006. **281**(39): p. 29105-19.
85. Wichroski, M.J., G.B. Robb, and T.M. Rana, *Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies*. PLoS Pathog, 2006. **2**(5): p. e41.
86. Bach, D., et al., *Characterization of APOBEC3G binding to 7SL RNA*. Retrovirology, 2008. **5**: p. 54.
87. Bogerd, H.P. and B.R. Cullen, *Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation*. RNA, 2008. **14**(6): p. 1228-36.
88. Burnett, A. and P. Spearman, *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, 2007. **81**(10): p. 5000-13.
89. Svarovskaia, E.S., et al., *Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs*. J Biol Chem, 2004. **279**(34): p. 35822-8.
90. Strebel, K. and M.A. Khan, *APOBEC3G encapsidation into HIV-1 virions: which RNA is it?* Retrovirology, 2008. **5**: p. 55.
91. Alce, T.M. and W. Popik, *APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein*. J Biol Chem, 2004. **279**(33): p. 34083-6.
92. Cen, S., et al., *The interaction between HIV-1 Gag and APOBEC3G*. J Biol Chem, 2004. **279**(32): p. 33177-84.
93. Douaisi, M., et al., *HIV-1 and MLV Gag proteins are sufficient to recruit APOBEC3G into virus-like particles*. Biochem Biophys Res Commun, 2004. **321**(3): p. 566-73.
94. Khan, M.A., et al., *Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions*. Retrovirology, 2007. **4**: p. 48.
95. Zhen, A., et al., *Reduced APOBEC3H variant anti-viral activities are associated with altered RNA binding activities*. PLoS One, 2012. **7**(7): p. e38771.
96. Ooms, M., et al., *The localization of APOBEC3H variants in HIV-1 virions determines their antiviral activity*. J Virol, 2010. **84**(16): p. 7961-9.
97. Wang, T., et al., *Interaction with 7SL RNA but not with HIV-1 genomic RNA or P bodies is required for APOBEC3F virion packaging*. J Mol Biol, 2008. **375**(4): p. 1098-112.
98. Huthoff, H., et al., *RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1*. PLoS Pathog, 2009. **5**(3): p. e1000330.

99. Chelico, L., et al., *Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G*. J Biol Chem, 2010. **285**(21): p. 16195-205.
100. Haché, G., M.T. Liddament, and R.S. Harris, *The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain*. J Biol Chem, 2005. **280**(12): p. 10920-4.
101. Navarro, F., et al., *Complementary function of the two catalytic domains of APOBEC3G*. Virology, 2005. **333**(2): p. 374-86.
102. Chelico, L., et al., *A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV*. J Biol Chem, 2008. **283**(20): p. 13780-91.
103. Shlyakhtenko, L.S., et al., *Atomic force microscopy studies of APOBEC3G oligomerization and dynamics*. J Struct Biol, 2013. **184**(2): p. 217-25.
104. Berg, O.G., R.B. Winter, and P.H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory*. Biochemistry, 1981. **20**(24): p. 6929-48.
105. von Hippel, P.H. and O.G. Berg, *Facilitated target location in biological systems*. J Biol Chem, 1989. **264**(2): p. 675-8.
106. Stanford, N.P., et al., *One- and three-dimensional pathways for proteins to reach specific DNA sites*. EMBO J, 2000. **19**(23): p. 6546-57.
107. Feng, Y. and L. Chelico, *Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain*. J Biol Chem, 2011. **286**(13): p. 11415-26.
108. Ara, A., R.P. Love, and L. Chelico, *Different mutagenic potential of HIV-1 restriction factors APOBEC3G and APOBEC3F is determined by distinct single-stranded DNA scanning mechanisms*. PLoS Pathog, 2014. **10**(3): p. e1004024.
109. Adolph, M.B., et al., *Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization*. Nucleic Acids Res, 2017.
110. Feng, Y., et al., *Natural Polymorphisms and Oligomerization of Human APOBEC3H Contribute to Single-stranded DNA Scanning Ability*. J Biol Chem, 2015. **290**(45): p. 27188-203.
111. Chelico, L., et al., *APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA*. Nat Struct Mol Biol, 2006. **13**(5): p. 392-9.
112. Chaipan, C., et al., *APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages*. J Virol, 2013. **87**(1): p. 444-53.
113. Bishop, K.N., et al., *Cytidine deamination of retroviral DNA by diverse APOBEC proteins*. Curr Biol, 2004. **14**(15): p. 1392-6.
114. Wiegand, H.L., et al., *A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins*. EMBO J, 2004. **23**(12): p. 2451-8.
115. Zheng, Y.H., et al., *Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication*. J Virol, 2004. **78**(11): p. 6073-6.
116. Song, C., et al., *Signals in APOBEC3F N-terminal and C-terminal deaminase domains each contribute to encapsidation in HIV-1 virions and are both required for HIV-1 restriction*. J Biol Chem, 2012. **287**(20): p. 16965-74.
117. Zennou, V. and P.D. Bieniasz, *Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates*. Virology, 2006. **349**(1): p. 31-40.

118. Sato, K., et al., *APOBEC3D and APOBEC3F potently promote HIV-1 diversification and evolution in humanized mouse model*. PLoS Pathog, 2014. **10**(10): p. e1004453.
119. Love, R.P., H. Xu, and L. Chelico, *Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A*. J Biol Chem, 2012. **287**(36): p. 30812-22.
120. Jern, P., et al., *Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance*. PLoS Pathog, 2009. **5**(4): p. e1000367.
121. Neogi, U., et al., *Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals*. J Int AIDS Soc, 2013. **16**: p. 18472.
122. Monajemi, M., et al., *Positioning of APOBEC3G/F mutational hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells*. PLoS One, 2014. **9**(4): p. e93428.
123. Squires, K.D., et al., *Impact of APOBEC Mutations on CD8+ T Cell Recognition of HIV Epitopes Varies Depending on the Restricting HLA*. J Acquir Immune Defic Syndr, 2015. **70**(2): p. 172-8.
124. Wood, N., et al., *HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC*. PLoS Pathog, 2009. **5**(5): p. e1000414.
125. Desimmie, B.A., et al., *APOBEC3 proteins can copackage and comutate HIV-1 genomes*. Nucleic Acids Res, 2016. **44**(16): p. 7848-65.
126. Dang, Y., et al., *Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family*. J Virol, 2006. **80**(21): p. 10522-33.
127. Duggal, N.K., H.S. Malik, and M. Emerman, *The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection*. J Virol, 2011. **85**(21): p. 11361-71.
128. Harari, A., et al., *Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H*. J Virol, 2009. **83**(1): p. 295-303.
129. Li, M.M. and M. Emerman, *Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity*. J Virol, 2011. **85**(16): p. 8197-207.
130. Ito, F., et al., *Understanding the Structure, Multimerization, Subcellular Localization and mC Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H*. Sci Rep, 2018. **8**(1): p. 3763.
131. Shaban, N.M., et al., *The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism*. Mol Cell, 2018. **69**(1): p. 75-86.e9.
132. Smith, H.C., *RNA binding to APOBEC deaminases; Not simply a substrate for C to U editing*. RNA Biol, 2017. **14**(9): p. 1153-1165.
133. Bohn, J.A., et al., *APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA*. Nat Commun, 2017. **8**(1): p. 1021.
134. Madani, N. and D. Kabat, *An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein*. J Virol, 1998. **72**(12): p. 10251-5.
135. Simon, J.H., et al., *Evidence for a newly discovered cellular anti-HIV-1 phenotype*. Nat Med, 1998. **4**(12): p. 1397-400.

136. Zhang, W., et al., *T-cell differentiation factor CBF- β regulates HIV-1 Vif-mediated evasion of host restriction*. Nature, 2011. **481**(7381): p. 376-9.
137. Liu, B., et al., *Influence of primate lentiviral Vif and proteasome inhibitors on human immunodeficiency virus type 1 virion packaging of APOBEC3G*. J Virol, 2004. **78**(4): p. 2072-81.
138. Mehle, A., et al., *Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation*. Genes Dev, 2004. **18**(23): p. 2861-6.
139. Luo, K., et al., *Primate lentiviral virion infectivity factors are substrate receptors that assemble with cullin 5-E3 ligase through a HCCH motif to suppress APOBEC3G*. Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11444-9.
140. Fribourgh, J.L., et al., *Core binding factor beta plays a critical role by facilitating the assembly of the Vif-cullin 5 E3 ubiquitin ligase*. J Virol, 2014. **88**(6): p. 3309-19.
141. Guo, Y., et al., *Structural basis for hijacking CBF- β and CUL5 E3 ligase complex by HIV-1 Vif*. Nature, 2014. **505**(7482): p. 229-33.
142. Aydin, H., M.W. Taylor, and J.E. Lee, *Structure-guided analysis of the human APOBEC3-HIV restrictome*. Structure, 2014. **22**(5): p. 668-84.
143. Letko, M., et al., *Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface*. Cell Rep, 2015. **13**(9): p. 1789-99.
144. Dang, Y., et al., *Identification of a novel WxSLVK motif in the N terminus of human immunodeficiency virus and simian immunodeficiency virus Vif that is critical for APOBEC3G and APOBEC3F neutralization*. J Virol, 2009. **83**(17): p. 8544-52.
145. Mehle, A., et al., *Identification of an APOBEC3G binding site in human immunodeficiency virus type 1 Vif and inhibitors of Vif-APOBEC3G binding*. J Virol, 2007. **81**(23): p. 13235-41.
146. Richards, C., et al., *The Binding Interface between Human APOBEC3F and HIV-1 Vif Elucidated by Genetic and Computational Approaches*. Cell Rep, 2015. **13**(9): p. 1781-8.
147. Ooms, M., M. Letko, and V. Simon, *The Structural Interface between HIV-1 Vif and Human APOBEC3H*. J Virol, 2017. **91**(5).
148. Huthoff, H. and M.H. Malim, *Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation*. J Virol, 2007. **81**(8): p. 3807-15.
149. Kitamura, S., et al., *The APOBEC3C crystal structure and the interface for HIV-1 Vif binding*. Nat Struct Mol Biol, 2012. **19**(10): p. 1005-10.
150. Feng, Y., R.P. Love, and L. Chelico, *HIV-1 viral infectivity factor (Vif) alters processive single-stranded DNA scanning of the retroviral restriction factor APOBEC3G*. J Biol Chem, 2013. **288**(9): p. 6083-94.
151. Albin, J.S., et al., *A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif*. J Biol Chem, 2010. **285**(52): p. 40785-92.
152. Zhen, A., et al., *A single amino acid difference in human APOBEC3H variants determines HIV-1 Vif sensitivity*. J Virol, 2010. **84**(4): p. 1902-11.
153. Li, M.M., L.I. Wu, and M. Emerman, *The range of human APOBEC3H sensitivity to lentiviral Vif proteins*. J Virol, 2010. **84**(1): p. 88-95.
154. Bélanger, K., et al., *Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses*. Nucleic Acids Res, 2013. **41**(15): p. 7438-52.

155. Bishop, K.N., et al., *APOBEC3G inhibits elongation of HIV-1 reverse transcripts*. PLoS Pathog, 2008. **4**(12): p. e1000231.
156. Gillick, K., et al., *Suppression of HIV-1 infection by APOBEC3 proteins in primary human CD4(+) T cells is associated with inhibition of processive reverse transcription as well as excessive cytidine deamination*. J Virol, 2013. **87**(3): p. 1508-17.
157. Guo, F., et al., *Inhibition of formula-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication*. J Virol, 2006. **80**(23): p. 11710-22.
158. Guo, F., et al., *The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA³Lys annealing to viral RNA*. J Virol, 2007. **81**(20): p. 11322-31.
159. Wang, X., et al., *The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication*. J Virol, 2012. **86**(7): p. 3777-86.
160. Mbisa, J.L., et al., *Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration*. J Virol, 2007. **81**(13): p. 7099-110.
161. Miyagi, E., et al., *Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1*. J Virol, 2007. **81**(24): p. 13346-53.
162. Kobayashi, T., et al., *Quantification of deaminase activity-dependent and -independent restriction of HIV-1 replication mediated by APOBEC3F and APOBEC3G through experimental-mathematical investigation*. J Virol, 2014. **88**(10): p. 5881-7.
163. Lee, Y.N., M.H. Malim, and P.D. Bieniasz, *Hypermutation of an ancient human retrovirus by APOBEC3G*. J Virol, 2008. **82**(17): p. 8762-70.
164. Quinn, J.P. and V.J. Bubb, *SVA retrotransposons as modulators of gene expression*. Mob Genet Elements, 2014. **4**: p. e32102.
165. Turelli, P., S. Vianin, and D. Trono, *The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay*. J Biol Chem, 2004. **279**(42): p. 43371-3.
166. Richardson, S.R., et al., *APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition*. Elife, 2014. **3**: p. e02008.
167. Horn, A.V., et al., *Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity*. Nucleic Acids Res, 2014. **42**(1): p. 396-416.
168. Liang, W., et al., *APOBEC3DE Inhibits LINE-1 Retrotransposition by Interacting with ORF1p and Influencing LINE Reverse Transcriptase Activity*. PLoS One, 2016. **11**(7): p. e0157220.
169. Feng, Y., et al., *Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H*. Sci Rep, 2017. **7**(1): p. 10881.
170. Swanton, C., et al., *APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity*. Cancer Discov, 2015. **5**(7): p. 704-12.
171. Roberts, S.A., et al., *An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers*. Nat Genet, 2013. **45**(9): p. 970-6.
172. Stephens, P.J., et al., *The landscape of cancer genes and mutational processes in breast cancer*. Nature, 2012. **486**(7403): p. 400-4.

173. Roberts, S.A. and D.A. Gordenin, *Clustered and genome-wide transient mutagenesis in human cancers: Hypermutation without permanent mutators or loss of fitness*. Bioessays, 2014.
174. Starrett, G.J., et al., *The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis*. Nat Commun, 2016. **7**: p. 12918.
175. Burns, M.B., et al., *APOBEC3B is an enzymatic source of mutation in breast cancer*. Nature, 2013. **494**(7437): p. 366-70.
176. Harris, R.S., *Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer*. Breast Cancer Res, 2015. **17**: p. 8.
177. Nikkila, J., et al., *Elevated APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells*. Br J Cancer, 2017.
178. Sieuwerts, A.M., et al., *Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers*. Horm Cancer, 2014. **5**(6): p. 405-13.
179. Law, E.K., et al., *The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer*. Sci Adv, 2016. **2**(10): p. e1601737.
180. Leonard, B., et al., *APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma*. Cancer Res, 2013. **73**(24): p. 7222-31.
181. Taylor, B.J., et al., *DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis*. Elife, 2013. **2**: p. e00534.
182. Chan, K., et al., *An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers*. Nat Genet, 2015. **47**(9): p. 1067-72.
183. Mussil, B., et al., *Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death*. PLoS One, 2013. **8**(8): p. e73641.
184. Landry, S., et al., *APOBEC3A can activate the DNA damage response and cause cell-cycle arrest*. EMBO Rep, 2011. **12**(5): p. 444-50.
185. Schutsky, E.K., et al., *APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA*. Nucleic Acids Res, 2017. **45**(13): p. 7655-7665.
186. Basu, U., et al., *The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates*. Cell, 2011. **144**(3): p. 353-63.
187. King, J.J. and M. Larijani, *A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer: Schrödinger's CATalytic Pocket*. Front Immunol, 2017. **8**: p. 351.
188. Wang, M., et al., *AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity*. Nat Struct Mol Biol, 2009. **16**(7): p. 769-76.
189. Adolph, M.B., et al., *Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B*. Nucleic Acids Res, 2017.
190. Hoopes, J.I., et al., *APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication*. Cell Rep, 2016. **14**(6): p. 1273-82.
191. Kazanov, M.D., et al., *APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions*. Cell Rep, 2015. **13**(6): p. 1103-9.

192. Kanu, N., et al., *DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer*. *Genome Biol*, 2016. **17**(1): p. 185.
193. Haradhvala, N.J., et al., *Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair*. *Cell*, 2016. **164**(3): p. 538-49.
194. Seplyarskiy, V.B., et al., *APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication*. *Genome Res*, 2016. **26**(2): p. 174-82.
195. Stenglein, M.D., et al., *APOBEC3 proteins mediate the clearance of foreign DNA from human cells*. *Nat Struct Mol Biol*, 2010. **17**(2): p. 222-9.
196. Suspene, R., et al., *Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism*. *Proc Natl Acad Sci U S A*, 2011. **108**(12): p. 4858-63.
197. Wu, Y., J. Lu, and T. Kang, *Human single-stranded DNA binding proteins: guardians of genome stability*. *Acta Biochim Biophys Sin (Shanghai)*, 2016. **48**(7): p. 671-7.
198. Symington, L.S., *End resection at double-strand breaks: mechanism and regulation*. *Cold Spring Harb Perspect Biol*, 2014. **6**(8).
199. Vesela, E., et al., *Common Chemical Inductors of Replication Stress: Focus on Cell-Based Studies*. *Biomolecules*, 2017. **7**(1).
200. Techer, H., et al., *The impact of replication stress on replication dynamics and DNA damage in vertebrate cells*. *Nat Rev Genet*, 2017. **18**(9): p. 535-550.
201. Nedelcheva-Velva, M.N., D.B. Krastev, and S.S. Stoyanov, *Coordination of DNA synthesis and replicative unwinding by the S-phase checkpoint pathways*. *Nucleic Acids Res*, 2006. **34**(15): p. 4138-46.
202. Gaillard, H., T. Garcia-Muse, and A. Aguilera, *Replication stress and cancer*. *Nat Rev Cancer*, 2015. **15**(5): p. 276-89.
203. Waters, C.E., et al., *FHIT loss-induced DNA damage creates optimal APOBEC substrates: Insights into APOBEC-mediated mutagenesis*. *Oncotarget*, 2015. **6**(5): p. 3409-19.
204. He, Y., et al., *Near-atomic resolution visualization of human transcription promoter opening*. *Nature*, 2016. **533**(7603): p. 359-65.
205. Harris, R.S. and J.P. Dudley, *APOBECs and virus restriction*. *Virology*, 2015. **479-480**: p. 131-45.
206. Jarmuz, A., et al., *An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22*. *Genomics*, 2002. **79**(3): p. 285-96.
207. Duggal, N.K. and M. Emerman, *Evolutionary conflicts between viruses and restriction factors shape immunity*. *Nat Rev Immunol*, 2012. **12**(10): p. 687-95.
208. Compton, A.A. and M. Emerman, *Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses*. *PLoS Pathog*, 2013. **9**(1): p. e1003135.
209. Yu, Q., et al., *APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication*. *J Biol Chem*, 2004. **279**(51): p. 53379-86.
210. Ahasan, M.M., et al., *APOBEC3A and 3C decrease human papillomavirus 16 pseudovirion infectivity*. *Biochem Biophys Res Commun*, 2015. **457**(3): p. 295-9.
211. Warren, C.J., et al., *APOBEC3A functions as a restriction factor of human papillomavirus*. *J Virol*, 2015. **89**(1): p. 688-702.

212. Zhang, W., et al., *Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression*. Cell Microbiol, 2008. **10**(1): p. 112-21.
213. Chen, H., et al., *APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons*. Curr Biol, 2006. **16**(5): p. 480-5.
214. Muckenfuss, H., et al., *APOBEC3 proteins inhibit human LINE-1 retrotransposition*. J Biol Chem, 2006. **281**(31): p. 22161-72.
215. Wissing, S., et al., *Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells*. J Biol Chem, 2011. **286**(42): p. 36427-37.
216. Niewiadomska, A.M., et al., *Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association*. J Virol, 2007. **81**(17): p. 9577-83.
217. Rose, K.M., et al., *Regulated production and anti-HIV type 1 activities of cytidine deaminases APOBEC3B, 3F, and 3G*. AIDS Res Hum Retroviruses, 2005. **21**(7): p. 611-9.
218. Langlois, M.A., et al., *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, 2005. **33**(6): p. 1913-23.
219. Kinomoto, M., et al., *All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition*. Nucleic Acids Res, 2007. **35**(9): p. 2955-64.
220. Sakurai, D., et al., *APOBEC3H polymorphisms associated with the susceptibility to HIV-1 infection and AIDS progression in Japanese*. Immunogenetics, 2015. **67**(4): p. 253-7.
221. Naruse, T.K., et al., *APOBEC3H polymorphisms and susceptibility to HIV-1 infection in an Indian population*. J Hum Genet, 2016. **61**(3): p. 263-5.
222. Smith, J.L. and V.K. Pathak, *Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif*. J Virol, 2010. **84**(24): p. 12599-608.
223. Refsland, E.W., et al., *Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction*. Nucleic Acids Res, 2010. **38**(13): p. 4274-84.
224. Compton, A.A., H.S. Malik, and M. Emerman, *Host gene evolution traces the evolutionary history of ancient primate lentiviruses*. Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1626): p. 20120496.
225. Smith, J.L., et al., *HIV-1 and HIV-2 Vif interact with human APOBEC3 proteins using completely different determinants*. J Virol, 2014. **88**(17): p. 9893-908.
226. Prado-Martinez, J., et al., *Great ape genetic diversity and population history*. Nature, 2013. **499**(7459): p. 471-5.
227. Berniell-Lee, G., et al., *Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages*. Mol Biol Evol, 2009. **26**(7): p. 1581-9.
228. Li, S., C. Schlebusch, and M. Jakobsson, *Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples*. Proc Biol Sci, 2014. **281**(1793).
229. Lachance, J., et al., *Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers*. Cell, 2012. **150**(3): p. 457-69.

230. Schuster, S.C., et al., *Complete Khoisan and Bantu genomes from southern Africa*. Nature, 2010. **463**(7283): p. 943-7.
231. Scally, A., et al., *Insights into hominid evolution from the gorilla genome sequence*. Nature, 2012. **483**(7388): p. 169-75.
232. Langergraber, K.E., et al., *Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution*. Proc Natl Acad Sci U S A, 2012. **109**(39): p. 15716-21.
233. Mangeat, B., et al., *A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action*. J Biol Chem, 2004. **279**(15): p. 14481-3.
234. Satta, Y. and N. Takahata, *Out of Africa with regional interbreeding? Modern human origins*. Bioessays, 2002. **24**(10): p. 871-5.
235. Cuevas, J.M., et al., *Extremely High Mutation Rate of HIV-1 In Vivo*. PLoS Biol, 2015. **13**(9): p. e1002251.
236. Ho, Y.C., et al., *Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure*. Cell, 2013. **155**(3): p. 540-51.
237. Kourteva, Y., et al., *APOBEC3G expression and hypermutation are inversely associated with human immunodeficiency virus type 1 (HIV-1) burden in vivo*. Virology, 2012. **430**(1): p. 1-9.
238. Yang, Z., *PAML 4: phylogenetic analysis by maximum likelihood*. Mol Biol Evol, 2007. **24**(8): p. 1586-91.
239. Kosakovsky Pond, S.L., et al., *A random effects branch-site model for detecting episodic diversifying selection*. Mol Biol Evol, 2011. **28**(11): p. 3033-43.
240. Yamashita, M. and M. Emerman, *Capsid is a dominant determinant of retrovirus infectivity in nondividing cells*. J Virol, 2004. **78**(11): p. 5670-8.
241. Xie, Y., et al., *Characterization of L1 retrotransposition with high-throughput dual-luciferase assays*. Nucleic Acids Res, 2011. **39**(3): p. e16.
242. Creighton, S., L.B. Bloom, and M.F. Goodman, *Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies*. Methods Enzymol, 1995. **262**: p. 232-56.
243. Yu, Q., et al., *Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome*. Nat Struct Mol Biol, 2004. **11**(5): p. 435-442.
244. Sheehy, A.M., et al., *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein*. Nature, 2002. **418**: p. 646-650.
245. Refsland, E.W., et al., *Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to affect viral G-to-A mutation levels and infectivity*. PLoS Genet, 2014. **10**(11): p. e1004761.
246. Etienne, L., et al., *Gene loss and adaptation to hominids underlie the ancient origin of HIV-1*. Cell Host Microbe, 2013. **14**(1): p. 85-92.
247. Mariani, R., et al., *Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif*. Cell, 2003. **114**(1): p. 21-31.
248. Stopak, K., et al., *HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability*. Mol Cell, 2003. **12**(3): p. 591-601.
249. Sheehy, A.M., N.C. Gaddis, and M.H. Malim, *The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif*. Nat Med, 2003. **9**(11): p. 1404-7.
250. Marin, M., et al., *HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation*. Nat Med, 2003. **9**(11): p. 1398-403.

251. Kao, S., et al., *The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity*. J Virol, 2003. **77**(21): p. 11398-407.
252. Jager, S., et al., *Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection*. Nature, 2012. **481**(7381): p. 371-5.
253. Senavirathne, G., et al., *Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution*. J Biol Chem, 2012. **287**(19): p. 15826-35.
254. Nowarski, R., et al., *Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase*. Nat Struct Mol Biol, 2008. **15**(10): p. 1059-1066.
255. Stauch, B., et al., *Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation*. Proc Natl Acad Sci U S A, 2009. **106**(29): p. 12079-84.
256. Refsland, E.W., et al., *Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction*. Nucleic Acids Res, 2010.
257. Koning, F.A., et al., *Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets*. J Virol, 2009. **83**(18): p. 9474-85.
258. Wang, T., et al., *Distinct viral determinants for the packaging of human cytidine deaminases APOBEC3G and APOBEC3C*. Virology, 2008. **377**(1): p. 71-9.
259. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Mol Syst Biol, 2011. **7**: p. 539.
260. Gouet, P., et al., *ESPrpt: analysis of multiple sequence alignments in PostScript*. Bioinformatics, 1999. **15**(4): p. 305-8.
261. Lieberman, B.A. and S.K. Nordeen, *DNA intersegment transfer, how steroid receptors search for a target site*. J Biol Chem, 1997. **272**(2): p. 1061-8.
262. Bohn, M.F., et al., *Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain*. Structure, 2013. **21**(6): p. 1042-50.
263. Siu, K.K., et al., *Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F*. Nat Commun, 2013. **4**: p. 2593.
264. Marx, A., M. Galilee, and A. Alian, *Zinc enhancement of cytidine deaminase activity highlights a potential allosteric role of loop-3 in regulating APOBEC3 enzymes*. Sci Rep, 2015. **5**: p. 18191.
265. Bohn, M.F., et al., *The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization*. Structure, 2015. **23**(5): p. 903-11.
266. King, J.J., et al., *Catalytic Pocket Inaccessibility of Activation-Induced Cytidine Deaminase Is a Safeguard against Excessive Mutagenic Activity*. Structure, 2015. **23**(4): p. 615-27.
267. Mak, C.H., et al., *A mathematical model for scanning and catalysis on single-stranded DNA, illustrated with activation-induced deoxycytidine deaminase*. J Biol Chem, 2013. **288**(41): p. 29786-95.
268. Senavirathne, G., et al., *Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution*. Nat Commun, 2015. **6**: p. 10209.
269. Lackey, L., et al., *Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination*. Cell Cycle, 2013. **12**(5): p. 762-72.

270. Byeon, I.J., et al., *Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity*. Biochemistry, 2016. **55**(21): p. 2944-59.
271. Mitra, M., et al., *Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties*. Nucleic Acids Res, 2014. **42**(2): p. 1095-110.
272. Byeon, I.J., et al., *NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity*. Nat Commun, 2013. **4**: p. 1890.
273. Pham, P., et al., *Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase*. J Biol Chem, 2008.
274. Opi, S., et al., *Monomeric APOBEC3G is catalytically active and has antiviral activity*. J Virol, 2006. **80**(10): p. 4673-82.
275. Lu, X., et al., *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, 2015. **290**(7): p. 4010-21.
276. Polevoda, B., et al., *Structural and functional assessment of APOBEC3G macromolecular complexes*. Methods, 2016.
277. Ara, A., et al., *Mechanism of enhanced HIV restriction by virion coencapsidated cytidine deaminases APOBEC3F and APOBEC3G*. J Virol, 2016.
278. Letko, M., et al., *Vif proteins from diverse primate lentiviral lineages use the same binding site in APOBEC3G*. J Virol, 2013. **87**(21): p. 11861-71.
279. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. J Mol Biol, 2007. **372**(3): p. 774-97.
280. Xiao, X., et al., *Crystal structures of APOBEC3G N-domain alone and its complex with DNA*. Nat Commun, 2016. **7**: p. 12193.
281. Nakashima, M., et al., *Structural Insights into HIV-1 Vif-APOBEC3F Interaction*. J Virol, 2015. **90**(2): p. 1034-47.
282. Bransteitter, R., et al., *Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase*. J Biol Chem, 2004. **279**(49): p. 51612-21.
283. Baig, T.T., Y. Feng, and L. Chelico, *Determinants of efficient degradation of APOBEC3 restriction factors by HIV-1 Vif*. J Virol, 2014. **88**(24): p. 14380-95.
284. Kimpton, J. and M. Emerman, *Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene*. J Virol, 1992. **66**(4): p. 2232-9.
285. Harris, R.S., J.F. Hultquist, and D.T. Evans, *The restriction factors of human immunodeficiency virus*. J Biol Chem, 2012. **287**(49): p. 40875-83.
286. Morse, M., et al., *Dimerization regulates both deaminase-dependent and deaminase-independent HIV-1 restriction by APOBEC3G*. Nat Commun, 2017. **8**(1): p. 597.
287. Schumacher, A.J., et al., *The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction*. J Virol, 2008. **82**(6): p. 2652-60.
288. Suspene, R., et al., *APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase*. Nucleic Acids Res, 2004. **32**(8): p. 2421-9.

289. Yu, Q., et al., *Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome*. Nat Struct Mol Biol, 2004. **11**(5): p. 435-42.
290. Coffin, J.M., *Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses*. J Gen Virol, 1979. **42**(1): p. 1-26.
291. Nowarski, R., et al., *Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase*. Nat Struct Mol Biol, 2008. **15**(10): p. 1059-66.
292. Li, X.Y., et al., *APOBEC3G inhibits DNA strand transfer during HIV-1 reverse transcription*. J Biol Chem, 2007. **282**(44): p. 32065-74.
293. Stenglein, M.D. and R.S. Harris, *APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism*. J Biol Chem, 2006. **281**(25): p. 16837-41.
294. Pollpeter, D., et al., *Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G*. Nat Microbiol, 2018. **3**(2): p. 220-233.
295. Belanger, K., et al., *Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses*. Nucleic Acids Res, 2013. **41**(15): p. 7438-52.
296. Basu, V.P., et al., *Strand transfer events during HIV-1 reverse transcription*. Virus Res, 2008. **134**(1-2): p. 19-38.
297. Temin, H.M., *Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation*. Proc Natl Acad Sci U S A, 1993. **90**(15): p. 6900-3.
298. Katz, R.A. and A.M. Skalka, *Generation of diversity in retroviruses*. Annu Rev Genet, 1990. **24**: p. 409-45.
299. Hu, W.S. and H.M. Temin, *Retroviral recombination and reverse transcription*. Science, 1990. **250**(4985): p. 1227-33.
300. Hu, W.S., et al., *Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference*. J Virol, 1997. **71**(8): p. 6028-36.
301. Zhang, J. and H.M. Temin, *Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication*. Science, 1993. **259**(5092): p. 234-8.
302. Parthasarathi, S., et al., *Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications*. J Virol, 1995. **69**(12): p. 7991-8000.
303. Pathak, V.K. and H.M. Temin, *Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations*. Proc Natl Acad Sci U S A, 1990. **87**(16): p. 6019-23.
304. Hwang, C.K., E.S. Svarovskaia, and V.K. Pathak, *Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching*. Proc Natl Acad Sci U S A, 2001. **98**(21): p. 12209-14.
305. Pfeiffer, J.K., M.M. Georgiadis, and A. Telesnitsky, *Structure-based moloney murine leukemia virus reverse transcriptase mutants with altered intracellular direct-repeat deletion frequencies*. J Virol, 2000. **74**(20): p. 9629-36.
306. Svarovskaia, E.S., et al., *Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching*. J Virol, 2000. **74**(15): p. 7171-8.
307. Kunkel, T.A. and K. Bebenek, *Recent studies of the fidelity of DNA synthesis*. Biochim Biophys Acta, 1988. **951**(1): p. 1-15.

308. Kunkel, T.A., et al., *Analysis of fidelity mechanisms with eukaryotic DNA replication and repair proteins*. Genome, 1989. **31**(1): p. 100-3.
309. Coffin, J.M., *HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy*. Science, 1995. **267**(5197): p. 483-9.
310. Abram, M.E., et al., *Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication*. J Virol, 2010. **84**(19): p. 9864-78.
311. Abram, M.E., et al., *Mutations in HIV-1 reverse transcriptase affect the errors made in a single cycle of viral replication*. J Virol, 2014. **88**(13): p. 7589-601.
312. Das, K., et al., *Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance*. J Biol Chem, 2009. **284**(50): p. 35092-100.
313. Halvas, E.K., E.S. Svarovskaia, and V.K. Pathak, *Role of murine leukemia virus reverse transcriptase deoxyribonucleoside triphosphate-binding site in retroviral replication and in vivo fidelity*. J Virol, 2000. **74**(22): p. 10349-58.
314. Luo, G.X. and J. Taylor, *Template switching by reverse transcriptase during DNA synthesis*. J Virol, 1990. **64**(9): p. 4321-8.
315. Arts, E.J., et al., *Comparison of deoxyoligonucleotide and tRNA(Lys-3) as primers in an endogenous human immunodeficiency virus-1 in vitro reverse transcription/template-switching reaction*. J Biol Chem, 1994. **269**(20): p. 14672-80.
316. Arts, E.J. and M.A. Wainberg, *Preferential incorporation of nucleoside analogs after template switching during human immunodeficiency virus reverse transcription*. Antimicrob Agents Chemother, 1994. **38**(5): p. 1008-16.
317. Liu, S., et al., *Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates*. Science, 2008. **322**(5904): p. 1092-7.
318. Liu, S., et al., *Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription*. Nat Struct Mol Biol, 2010. **17**(12): p. 1453-60.
319. Anderson, J.A., et al., *Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination*. J Virol, 1998. **72**(2): p. 1186-94.
320. Arts, E.J., Z. Li, and M.A. Wainberg, *Analysis of Primer Extension and the First Template Switch during Human Immunodeficiency Virus Reverse Transcription*. J Biomed Sci, 1995. **2**(4): p. 314-321.
321. Nikolenko, G.N., et al., *Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency*. J Virol, 2004. **78**(16): p. 8761-70.
322. Bertram, J.G., et al., *DNA polymerase fidelity: comparing direct competition of right and wrong dNTP substrates with steady state and pre-steady state kinetics*. Biochemistry, 2010. **49**(1): p. 20-8.
323. Boosalis, M.S., J. Petruska, and M.F. Goodman, *DNA polymerase insertion fidelity. Gel assay for site-specific kinetics*. J Biol Chem, 1987. **262**(30): p. 14689-96.
324. Goodman, M.F., *DNA replication fidelity: kinetics and thermodynamics*. Mutat Res, 1988. **200**(1-2): p. 11-20.
325. Bebenek, K., et al., *Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots*. J Biol Chem, 1993. **268**(14): p. 10324-34.

326. Roberts, J.D., K. Bebenek, and T.A. Kunkel, *The accuracy of reverse transcriptase from HIV-1*. Science, 1988. **242**(4882): p. 1171-3.
327. Malim, M.H., *APOBEC proteins and intrinsic resistance to HIV-1 infection*. Philos Trans R Soc Lond B Biol Sci, 2009. **364**(1517): p. 675-87.
328. Gao, F., et al., *The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin*. J Virol, 1996. **70**(10): p. 7013-29.
329. Le Grice, S.F. and F. Grüniger-Leitch, *Rapid purification of homodimer and heterodimer HIV-1 reverse transcriptase by metal chelate affinity chromatography*. Eur J Biochem, 1990. **187**(2): p. 307-14.
330. Saraconi, G., et al., *The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas*. Genome Biol, 2014. **15**(7): p. 417.
331. Casellas, R., et al., *Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity*. Nat Rev Immunol, 2016. **16**(3): p. 164-76.
332. Cescon, D.W. and B. Haibe-Kains, *DNA replication stress: a source of APOBEC3B expression in breast cancer*. Genome Biol, 2016. **17**(1): p. 202.
333. Periyasamy, M., et al., *APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer*. Cell Rep, 2015. **13**(1): p. 108-21.
334. Kosumi, K., et al., *APOBEC3B is an enzymatic source of molecular alterations in esophageal squamous cell carcinoma*. Med Oncol, 2016. **33**(3): p. 26.
335. Onguru, O., et al., *APOBEC3B expression in drug resistant MCF-7 breast cancer cell lines*. Biomed Pharmacother, 2016. **79**: p. 87-92.
336. Chou, W.C., et al., *B-Myb Induces APOBEC3B Expression Leading to Somatic Mutation in Multiple Cancers*. Sci Rep, 2017. **7**: p. 44089.
337. Rogozin, I.B., et al., *APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis*. Cell Cycle, 2005. **4**(9): p. 1281-5.
338. Lackey, L., et al., *APOBEC3B and AID have similar nuclear import mechanisms*. J Mol Biol, 2012. **419**(5): p. 301-14.
339. Sieuwerts, A.M., et al., *Progressive APOBEC3B mRNA expression in distant breast cancer metastases*. PLoS One, 2017. **12**(1): p. e0171343.
340. Fanourakis, G., et al., *Evidence for APOBEC3B mRNA and protein expression in oral squamous cell carcinomas*. Exp Mol Pathol, 2016. **101**(3): p. 314-319.
341. Chan, K. and D.A. Gordenin, *Clusters of Multiple Mutations: Incidence and Molecular Mechanisms*. Annu Rev Genet, 2015. **49**: p. 243-67.
342. Shi, K., et al., *Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain*. J Biol Chem, 2015. **290**(47): p. 28120-30.
343. Siriwardena, S.U., T.A. Guruge, and A.S. Bhagwat, *Characterization of the Catalytic Domain of Human APOBEC3B and the Critical Structural Role for a Conserved Methionine*. J Mol Biol, 2015. **427**(19): p. 3042-55.
344. Shi, K., et al., *Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B*. Nat Struct Mol Biol, 2017. **24**(2): p. 131-139.
345. Xiao, X., et al., *Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation*. Nucleic Acids Res, 2017.

346. Fu, Y., et al., *DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B*. *Biochem J*, 2015. **471**(1): p. 25-35.
347. McDougale, R.M., et al., *D316 is critical for the enzymatic activity and HIV-1 restriction potential of human and rhesus APOBEC3B*. *Virology*, 2013. **441**(1): p. 31-9.
348. Pham, P., et al., *A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer*. *J Biol Chem*, 2013. **288**(41): p. 29294-304.
349. Pham, P., et al., *Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase*. *J Biol Chem*, 2008. **283**(25): p. 17428-39.
350. Miller, L.K., *Baculoviruses as gene expression vectors*. *Annu Rev Microbiol*, 1988. **42**: p. 177-99.
351. Halford, S.E. and J.F. Marko, *How do site-specific DNA-binding proteins find their targets?* *Nucleic Acids Res*, 2004. **32**(10): p. 3040-52.
352. Schonhoft, J.D. and J.T. Stivers, *DNA translocation by human uracil DNA glycosylase: the case of single-stranded DNA and clustered uracils*. *Biochemistry*, 2013. **52**(15): p. 2536-44.
353. Friedman, J.I. and J.T. Stivers, *Detection of damaged DNA bases by DNA glycosylase enzymes*. *Biochemistry*, 2010. **49**(24): p. 4957-67.
354. Pham, P., et al., *Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation*. *Nature*, 2003. **424**(6944): p. 103-7.
355. Shlyakhtenko, L.S., et al., *Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G*. *J Biol Chem*, 2011. **286**(5): p. 3387-95.
356. Shlyakhtenko, L.S., et al., *APOBEC3G Interacts with ssDNA by Two Modes: AFM Studies*. *Sci Rep*, 2015. **5**: p. 15648.
357. Mertz, T.M., V. Harcy, and S.A. Roberts, *Risks at the DNA Replication Fork: Effects upon Carcinogenesis and Tumor Heterogeneity*. *Genes (Basel)*, 2017. **8**(1).
358. Reagin, M.J., et al., *TempliPhi: A sequencing template preparation procedure that eliminates overnight cultures and DNA purification*. *J Biomol Tech*, 2003. **14**(2): p. 143-8.
359. Shlyakhtenko, L.S., et al., *Interaction of APOBEC3A with DNA assessed by atomic force microscopy*. *PLoS One*, 2014. **9**(6): p. e99354.
360. Esadze, A., et al., *Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength*. *Nucleic Acids Res*, 2014. **42**(11): p. 7039-46.
361. Chen, R. and M.S. Wold, *Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair*. *Bioessays*, 2014. **36**(12): p. 1156-61.
362. Gibb, B., et al., *Concentration-dependent exchange of replication protein A on single-stranded DNA revealed by single-molecule imaging*. *PLoS One*, 2014. **9**(2): p. e87922.
363. Nguyen, B., et al., *Diffusion of human replication protein A along single-stranded DNA*. *J Mol Biol*, 2014. **426**(19): p. 3246-61.
364. Gotte, M., et al., *Reverse transcriptase in motion: conformational dynamics of enzyme-substrate interactions*. *Biochim Biophys Acta*, 2010. **1804**(5): p. 1202-12.

365. Pellegrini, L. and A. Costa, *New Insights into the Mechanism of DNA Duplication by the Eukaryotic Replisome*. Trends Biochem Sci, 2016. **41**(10): p. 859-71.
366. Chaudhuri, J., C. Khuong, and F.W. Alt, *Replication protein A interacts with AID to promote deamination of somatic hypermutation targets*. Nature, 2004. **430**(7003): p. 992-8.
367. Vuong, B.Q., et al., *Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination*. Nat Immunol, 2009. **10**(4): p. 420-6.
368. Lada, A.G., et al., *Replication protein A (RPA) hampers the processive action of APOBEC3G cytosine deaminase on single-stranded DNA*. PLoS One, 2011. **6**(9): p. e24848.
369. Bachl, J., et al., *Increased transcription levels induce higher mutation rates in a hypermutating cell line*. J Immunol, 2001. **166**(8): p. 5051-7.
370. Saini, N., et al., *APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast*. DNA Repair (Amst), 2017.
371. El Hage, A., et al., *Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria*. PLoS Genet, 2014. **10**(10): p. e1004716.
372. King, J.J. and M. Larijani, *A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer: Schrodinger's CATalytic Pocket*. Front Immunol, 2017. **8**: p. 351.
373. Kouno, T., et al., *Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity*. Nat Commun, 2017. **8**: p. 15024.
374. Harjes, S., et al., *NMR-based method of small changes reveals how DNA mutator APOBEC3A interacts with its single-stranded DNA substrate*. Nucleic Acids Res, 2017. **45**(9): p. 5602-5613.
375. Bransteitter, R., et al., *Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase*. Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4102-7.
376. Min, J.H. and N.P. Pavletich, *Recognition of DNA damage by the Rad4 nucleotide excision repair protein*. Nature, 2007. **449**(7162): p. 570-5.
377. Sugasawa, K., et al., *A molecular mechanism for DNA damage recognition by the xeroderma pigmentosum group C protein complex*. DNA Repair (Amst), 2002. **1**(1): p. 95-107.