

**STUDIES ON THE DETERMINANTS OF HIV MUTAGENESIS AND
STRATEGIES FOR ITS ENHANCEMENT**

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Jonathan M. Rawson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Louis Mansky

December 2015

© Jonathan M Rawson, 2015

Acknowledgements

I am tremendously grateful to my family and friends for their continual support and encouragement. Without you all, the journey toward my degree would have been far less bearable. Thank you for helping me see the light at the end of the tunnel and for pushing me forward in its pursuit.

I thank my advisor, Dr. Louis Mansky, for his mentorship, which has helped me develop into a better scientist. I can now confidently say that I am prepared for the next stage of what will hopefully be a lifelong career in virology research.

I would like to express gratitude to all the scientists that helped intellectually guide my dissertation research along the way. In this regard, I thank my thesis committee—Dr. Reuben Harris (chair), Dr. Daniel Voytas, Dr. Paul Jardine, and Dr. Nik Somia. I also thank Dr. Steve Patterson, who in many respects served as an unofficial committee member.

I would also like to acknowledge all of the members (both past and present) of the Mansky Lab. I particularly thank Dr. Christine Clouser and Dr. Iwen Grigsby for mentorship on research projects and for suggesting revisions to manuscripts.

In addition, I am enormously grateful to all of the collaborators that assisted me on various research projects. I am indebted to Sean Landman for development of bioinformatics tools that were instrumental to all of the mutational analyses presented in Chapters II-V. I thank Dr. Cavan Reilly for statistical analysis of next-generation sequencing data in Chapters II-V. I am also grateful to all of the staff of the University of Minnesota Genomics Center, particularly Dr. Kenneth Beckman and Dr. Daryl Gohl, for assistance planning, performing, and analyzing data from next-generation sequencing experiments. Further, I thank Dr. Laurent

Bonnac for synthesis of anti-viral compounds and for determination of compound purity and stability.

Lastly, I am truly appreciative of all the financial support I have received while pursuing my degree, including graduate school and doctoral dissertation fellowships from the University of Minnesota. I am also grateful for support from the National Institutes of Health, provided through the Institute of Molecular Virology Training Program and an F31 individual predoctoral fellowship.

Dedication

To my beautiful wife, Karina, and my ever-energetic son, Tristan, who constantly inspire me to become a better person, scientist, husband, father, and friend.

Abstract

Human immunodeficiency virus type-1 (HIV-1) is one of the fastest evolving entities on earth, due in part to a mutation rate that is at least 10,000-fold higher than that of eukaryotic genomic DNA. The adaptability of HIV-1 prevents clearance of the virus by the immune system, promotes drug resistance, and has impeded development of effective vaccines. The viral mutation rate is a composite result of mutations that are contributed by multiple host and viral enzymes involved in viral replication. However, the precise determinants of viral mutagenesis in HIV-1 and the degree of their contribution to HIV-1 genetic diversity remain incompletely understood. Furthermore, differences in viral mutagenesis among different HIV types (e.g., HIV-1 and human immunodeficiency virus type 2, HIV-2) and subtypes have not been explored to date. Lastly, the elimination of HIV-1 infectivity by increasing the viral mutation rate has not been extensively investigated – in terms of mechanism of action and in preclinical and clinical evaluations. Based upon these general observations, this dissertation research was conducted to test the following hypotheses: 1) HIV-1 and HIV-2 have different mutation rates; 2) decitabine, a HIV-1 mutagen, creates G-to-C transversion mutations in the absence of mutational hotspots; 3) 5-azacytidine causes HIV-1 mutagenesis after reduction to 5-aza-2'-deoxycytidine (i.e. decitabine); 4) the anti-HIV-1 activity of 5-azacytidine is antagonized by inhibitors of ribonucleotide reductase; 5) the anti-HIV-1 activity of ribonucleotide reductase inhibitors are potentiated by 5,6-dihydro-5-aza-2'-deoxycytidine. The studies conducted in this dissertation advance current understanding of the determinants for retroviral mutagenesis, the mechanisms by which small molecules promote mutations in HIV-1, and approaches for using combinations of small molecules to reduce viral infectivity by enhancing the mutagenesis of HIV-1. Supplementary data (figures, tables, and spreadsheets) associated with this thesis are available online.

TABLE OF CONTENTS

ABSTRACT	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF PUBLICATIONS	xii
CHAPTER I: GENERAL INTRODUCTION	
HIV and the AIDS Pandemic	2
Genetic Diversity of HIV	2
HIV-1 Recombination: Mechanisms and Determinants	6
HIV-1 Mutagenesis: Mechanisms and Determinants	9
The HIV-1 Mutation Rate as a Therapeutic Target	13
Dissertation Objectives	17
Figures	21
CHAPTER II: HIV-1 AND HIV-2 EXHIBIT SIMILAR MUTATION FREQUENCIES AND SPECTRA IN THE ABSENCE OF G-TO-A HYPERMUTATION	
Introduction	27
Results	29
Discussion	38
Materials and Methods	44
Figures	53
CHAPTER III: LACK OF MUTATIONAL HOTSPOTS DURING DECITABINE-MEDIATED HIV-1 MUTAGENESIS	
Introduction	59
Results	61
Discussion	67
Materials and Methods	72

Figures	76
CHAPTER IV: 5-AZACYTIDINE ENHANCES THE MUTAGENESIS OF HIV-1 BY REDUCTION TO 5-AZA-2'-DEOXYCYTIDINE	
Introduction	86
Results	88
Discussion	92
Materials and Methods	95
Figures	100
CHAPTER V: SYNERGISTIC REDUCTION OF HIV-1 INFECTIVITY BY 5-AZACYTIDINE AND INHIBITORS OF RIBONUCLEOTIDE REDUCTASE	
Introduction	107
Results	109
Discussion	118
Materials and Methods	122
Figures	130
CHAPTER VI: 5,6-DIHYDRO-5-AZA-2'-DEOXYCYTIDINE POTENTIATES THE ANTI-HIV-1 ACTIVITY OF RIBONUCLEOTIDE REDUCTASE INHIBITORS	
Introduction	142
Results and Discussion	143
Materials and Methods	149
Figures	155
CHAPTER VII: DISSERTATION SUMMARY AND FUTURE DIRECTIONS	161
BIBLIOGRAPHY	173

APPENDIX I: DETERMINATION OF HIV-1 AND HIV-2 MUTATION FREQUENCIES AND SPECTRA BY SINGLE-STRAND CONSENSUS SEQUENCING	202
APPENDIX II: RETROVIRAL VECTORS FOR ANALYSIS OF VIRAL MUTAGENESIS AND RECOMBINATION	229
APPENDIX III: RAPID DETERMINATION OF HIV-1 MUTANT FREQUENCIES AND MUTATION SPECTRA USING AN MCHERRY/EGFP DUAL REPORTER VIRAL VECTOR	274
APPENDIX IV: DECLARATION OF CONTRIBUTIONS TO CO-AUTHORED PUBLICATION: NOVEL INHIBITORS OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 INFECTIVITY	299
APPENDIX V: DECLARATION OF CONTRIBUTIONS TO CO-AUTHORED PUBLICATION: CHARACTERIZATION OF CYTOPLASMIC GAG-GAG INTERACTIONS BY DUAL-COLOR Z-SCAN FLUORESCENCE FLUCTUATION SPECTROSCOPY	313
APPENDIX VI: DECLARATION OF CONTRIBUTIONS TO CO-AUTHORED PUBLICATION: BIOPHYSICAL ANALYSIS OF HTLV-1 PARTICLES REVEALS NOVEL INSIGHTS INTO PARTICLE MORPHOLOGY AND GAG STOICHIOMETRY	339
APPENDIX VII: COPYRIGHT PERMISSIONS	368

LIST OF TABLES

Chapter III		Page
Table 3-1	Locations and effects of common decitabine-induced G-to-C transversions in HIV-1	83
Table 3-2	Locations and effects of common decitabine-induced C-to-G transversions in HIV-1	84
Chapter IV		
Table 4-1	Primer sequences for Illumina amplicon sequencing of HIV-1	105
Chapter V		
Table 5-1	Effects of ribonucleotide reductase inhibitors on the antiviral activity and cytotoxicity of 5-azacytidine	140
Chapter VI		
Table 6-1	KP-1212 improves the antiretroviral potency of RNRI without altering cytotoxicity	160
Chapter VII		
Table 7-1	HIV-1 and HIV-2 vectors for comparison of viral mutagenesis by single-strand consensus sequencing	172

LIST OF FIGURES

Chapter I		Page
Figure 1-1	Origins of HIV-1 and HIV-2	21
Figure 1-2	Mechanisms of HIV-1 recombination and mutation	22
Figure 1-3	Compounds under investigation for the lethal mutagenesis of HIV-1	24
Figure 1-4	Model for the anti-HIV-1 activity of decitabine and 5-azacytidine	25
Chapter II		
Figure 2-1	Experimental strategy for investigating HIV-1 and HIV-2 mutagenesis by Illumina DNA sequencing	53
Figure 2-2	HIV-2 has a lower mutation frequency and distinct mutation spectrum relative to HIV-1	54
Figure 2-3	HIV-1 demonstrates higher G-to-A hypermutant frequencies than HIV-2	55
Figure 2-4	HIV-1 and HIV-2 mutation frequencies and spectra are similar in the absence of G-to-A hypermutation	57
Chapter III		
Figure 3-1	Decitabine primarily induces G-to-C transversion mutations during HIV-1 replication	76
Figure 3-2	Most decitabine-induced G-to-C and C-to-G mutant read pairs have relatively low mutational loads	78
Figure 3-3	All amplicons are susceptible to decitabine-mediated HIV-1 mutagenesis	79
Figure 3-4	Decitabine-induced mutations in HIV-1 are not caused by DNA methyltransferase-mediated hydrolysis	80
Figure 3-5	Lack of decitabine-induced G-to-C transversion mutation hotspots	81

Figure 3-6	Virtually all guanine and cytosine positions are susceptible to decitabine-mediated mutagenesis	82
Chapter IV		
Figure 4-1	5-azacytidine and 5-aza-2'-deoxycytidine induce similar levels of G-to-C and C-to-G transversion mutations during HIV-1 replication	100
Figure 4-2	5-azacytidine and 5-aza-2'-deoxycytidine induce similar patterns of mutation during HIV-1 replication	101
Figure 4-3	5-aza-2'-deoxycytidine-triphosphate levels are comparable in cells treated with 5-azacytidine or 5-aza-2'-deoxycytidine	102
Figure 4-4	HIV-1 RT incorporates 5-azacytidine-triphosphate <i>in vitro</i> much less efficiently than 5-aza-2'-deoxycytidine-triphosphate	103
Figure 4-5	Model of 5-azacytidine and 5-aza-2'-deoxycytidine-mediated HIV-1 mutagenesis	104
Chapter V		
Figure 5-1	Effects of 5-azacytidine and ribonucleotide reductase inhibitor combinations on HIV-1 infectivity and cell viability	130
Figure 5-2	5-azacytidine and ribonucleotide reductase inhibitors exhibit synergy against HIV-1	131
Figure 5-3	5-azacytidine and resveratrol exhibit antiviral synergy in a pattern similar to that of 5-aza-2'-deoxycytidine and resveratrol	132
Figure 5-4	Differential depletion of endogenous dNTP pools by ribonucleotide reductase inhibitors	133
Figure 5-5	Low concentrations of ribonucleotide reductase inhibitors do not decrease the level of 5-aza-2'-deoxycytidine-triphosphate in cells treated with 5-azacytidine	134
Figure 5-6	Resveratrol does not potentiate 5-azacytidine-mediated increases in the HIV-1 mutation frequency	136
Figure 5-7	Combinations of 5-azacytidine and resveratrol reduce levels of HIV-1 reverse transcription products	138

Chapter VI

Figure 6-1	Vector and assay design for rapid determination of HIV-1 infectivity and mutant frequency	155
Figure 6-2	Effect of KP-1212 and RNRIs on HIV-1 infectivity and mutant frequency	156
Figure 6-3	KP-1212 potentiates the antiviral and mutagenic activities of RNRIs against HIV-1	157
Figure 6-4	The ability of KP-1212 to potentiate RNRIs is not due to enhanced cytotoxicity	158
Figure 6-5	Characterization of viral mutation spectra in the presence of KP-1212 and RNRIs	159

LIST OF PUBLICATIONS

1. **Rawson JM**, Clouser CL, Mansky LM. Rapid determination of HIV-1 mutant frequencies and mutation spectra using an mCherry/EGFP dual reporter viral vector. *Method Mol Biol* [In Press].
2. **Rawson JM**, Landman SR, Reilly CS, Bonnac L, Patterson SE, Mansky LM. Lack of mutational hotspots during decitabine-mediated HIV-1 mutagenesis. *Antimicrob Agents Ch* (2015) 59/11, 6834-43.
3. **Rawson JM**, Landman SR, Reilly CS, Mansky LM. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* (2015) 12/60.
4. **Rawson JM**, Mansky LM. Retroviral vectors for analysis of viral mutagenesis and recombination. *Viruses* (2014) 6/9, 3612-42.
5. Beach LB, **Rawson JM**, Kim B, Patterson SE, Mansky LM. Novel inhibitors of human immunodeficiency virus type 2 infectivity. *J Gen Virol* (2014) Epub 2014 Aug 7.
6. **Rawson JM**, Heineman RH, Beach LB, Martin JL, Schnettler EK, Dapp MJ, Patterson SE, Mansky LM. 5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors. *Bioorg Med Chem* (2013) 21/22, 7222-8.
7. Fogarty KH, Chen Y, Grigsby IF, Macdonald PJ, Smith EM, Johnson JL, **Rawson JM**, Mansky LM, Mueller JD. Characterization of cytoplasmic Gag-Gag interactions by dual-color z-scan fluorescence fluctuation spectroscopy. *Biophys J* (2011) 100/6, 1587-95.
8. Grigsby IF, Zhang W, Johnson JL, Fogarty KH, Chen Y, **Rawson JM**, Crosby AJ, Mueller JD, Mansky LM. Biophysical analysis of HTLV-1 particles reveals novel insights into particle morphology and Gag stoichiometry. *Retrovirology* (2010) 7/75.

CHAPTER I

GENERAL INTRODUCTION

Adapted in part from: Rawson JM, Mansky LM. Retroviral Vectors for Analysis of Viral Mutagenesis and Recombination. *Viruses* (2014) 6/9, 3612-42. This article is openly distributed by MDPI under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

HIV and the AIDS Pandemic

Human immunodeficiency virus type 1 (HIV-1) infects ~37 million individuals worldwide and has resulted in ~39 million deaths since the onset of the pandemic (www.unaids.org). While the advent of combination anti-retroviral therapy (ART) has saved millions of lives, only 37% (~13 million) of infected individuals are receiving ART globally. In recent years, the pandemic has stabilized and begun to improve in most regions, as new infections and deaths have gradually decreased to ~2.1 and ~1.5 million/year, respectively. Nonetheless, the AIDS pandemic has intensified in specific problematic sub-regions, such as Eastern Europe, the Middle East, and North Africa, where the rates of new infections and deaths have risen over the past few years. Efforts to end the AIDS pandemic are focused on expanding education and awareness, improving access to testing (for diagnosis, viral load monitoring, and genotyping), preventing transmission through pre-exposure and post-exposure prophylaxis methods, increasing anti-retroviral drug availability, and raising the overall standard of care for infected individuals. Additionally, scientists continue to perform research directed at key unsolved problems, such as the development of an effective vaccine, the identification of improved chemoprevention strategies (such as microbicides), and the establishment of methods to eliminate latent viral reservoirs, which could potentially enable viral eradication and provide a “functional cure” for infected individuals. Further, while 26 anti-retroviral drugs (representing 6 classes) have been approved for the treatment of HIV-1 infection, the search for improved anti-retroviral drugs and drug combinations continues, due to concerns over the transmission of drug-resistant virus, the development of multi-drug resistance, and long-term drug-associated toxicities.

Genetic Diversity of HIV

Human immunodeficiency virus type-1 (HIV-1) is the major cause of the worldwide pandemic, whereas human immunodeficiency virus type-2 (HIV-2)

infects only ~1-2 million individuals worldwide and has remained largely confined to West Africa (1). Even in West Africa, HIV-2 prevalence has declined while HIV-1 prevalence has steadily increased, demonstrating the lower epidemiological fitness of HIV-2 (2-4). HIV-2 is likely less prevalent due to its reduced transmissibility, which has been demonstrated for both sexual transmission and mother to child transmission (5-7). Further, HIV-2 is generally less pathogenic than HIV-1, leading to AIDS in only ~25% of untreated individuals (compared to >95% for HIV-1) (8, 9). The reduced transmissibility and attenuated pathogenicity of HIV-2 are associated with much lower viral loads for HIV-2 than HIV-1, both in plasma and in genital secretions (7, 10-12). However, when HIV-2 infection does lead to AIDS, the clinical course of disease progression is virtually indistinguishable from that induced by HIV-1 (13).

HIV-1 & 2 both originated from cross-species transmission of distinct lineages of simian immunodeficiency virus (SIV) from primates to humans (Figure 1-1). Specifically, HIV-1 resulted from transmission of SIV from chimpanzees or gorillas (SIVcpz and SIVgor), while HIV-2 originated from SIV in sooty mangabey monkeys (SIVsm). Both types of transmission events have occurred multiple times, resulting in four groups of HIV-1 (M—P) and eight groups of HIV-2 (A—H). HIV-1 Group M (for Major) is responsible for the worldwide pandemic and has further diversified into nine primary subtypes (A-D, F-H, J-K) and more than 70 circulating recombinant forms. At the amino acid level, HIV-1 & 2 are ~60% homologous in Gag and Pol, but only ~30-40% homologous in Env. Additionally, HIV-1 & 2 differ in the accessory proteins they encode. HIV-2 encodes Vpx, which counteracts the host restriction factor SAMHD1, allowing for efficient infection of monocyte-derived macrophages (14). HIV-1, in contrast, does not encode an accessory protein to antagonize SAMHD1 activity. Further, HIV-2 lacks Vpu, which in HIV-1 counteracts the host restriction factor tetherin, promoting efficient virus release (15). HIV-2 instead counteracts tetherin through Env (16, 17), thus imposing additional functional constraints on Env.

Many groups have attempted to explore mechanisms of the limited transmissibility and pathogenicity of HIV-2, as such efforts could inform vaccine design for HIV-1. In cell culture, HIV-2 exhibits reduced replicative fitness and transmission efficiency relative to HIV-1 Group M (18). In patients, HIV-2 Env diversifies more slowly than HIV-1 Env when compared to HIV-1-infected individuals with high viral loads (19). However, HIV-1 & 2 diversification rates are similar when compared to HIV-1-infected individuals with low viral loads. The authors also found that the HIV-2 Env evolves more slowly than HIV-1 Env and is subjected to strong purifying selection. In another study, HIV-2 Env was found to exhibit a lower rate of synonymous substitutions than HIV-1, implying reduced viral mutation and/or replication rates (20), though another group has reported contrasting results (21). Multiple components of the host immune response appear more effective in controlling HIV-2 infection, including superior CD4⁺ and CD8⁺ T-cell responses (22, 23). Further, HIV-2 infection frequently elicits potent and broadly neutralizing antibodies against Env, which rarely arise during HIV-1 infection (24). The authors proposed that these antibodies were induced due to a combination of: 1) reduced glycan shielding and, 2) reduced conformational masking, leading to increased exposure of multiple cross-reactive epitopes. Additionally, others have proposed that HIV-2 is better controlled because it infects monocyte-derived dendritic cells more efficiently than HIV-1 (by virtue of Vpx), leading to improved immunosurveillance (25-27). However, Vpx is clearly important for *in vivo* infection: Vpx is actively maintained in HIV-2 infected individuals (28), and SIV Vpx-deficient variants replicate with delayed kinetics in primate models (29, 30). Overall, the limited pathogenicity of HIV-2 is likely due to a complex combination of innate virological and host immunological factors.

HIV-1 has been successful in establishing a worldwide pandemic largely due to its ability to rapidly diversify within infected hosts (31-33). The transmission of HIV-1 from donor to recipient has long been known to be associated with an extreme reduction in genetic diversity, such that the initial virus population in a newly infected individual is mostly homogenous (34-36). In

fact, ~75-80% of primary infections resulting from sexual transmission are caused by a single virion, indicative of an extreme genetic bottleneck during transmission (37-39). In the other ~20-25% of cases, primary infection by sexual transmission can be explained by a minimum of 2—5 virions. Some evidence suggests that HIV-1 transmission by injection drug use may be associated with a higher rate of multiple-variant transmission, as would be expected since this transmission route bypasses the mucosal barriers (40). During acute infection, viremia initially increases dramatically, but activation of the adaptive immune response ~4 to 6 weeks after infection reduces the population size by several orders of magnitude. After the selection of viral variants that escape the immune system, a prolonged period of chronic infection begins, in which the virus continually replicates at a high and stable level (called the viral set point) but is clinically asymptomatic. Throughout chronic infection, the viral population diversifies into an increasingly heterogeneous population of related but unique and interacting variants, termed the viral quasispecies (41). HIV-1 intra-host genetic variability can eventually reach up to 5-10% at the nucleotide level within Env (42, 43). Because the virus is constantly subjected to strong positive selective pressures from the immune system, the high rates of diversification observed with HIV-1 correspond with high rates of evolution (i.e. rates of divergence) as well. Indeed, both HIV-1 diversity and divergence increase approximately linearly during chronic infection, eventually stabilizing near the time of disease progression (42, 44, 45). The intra-host genetic diversity of HIV-1 enables escape from the immune system, accelerates the emergence of drug resistance, and promotes changes in cell and/or tissue tropism. Ultimately, HIV-1 genetic variation has posed a tremendous barrier to the development of an effective vaccine and has necessitated the use of combination drug therapy to control viral infection.

The ability of HIV-1 to quickly diversify can be attributed to several key factors: large population sizes ($>10^{10}$ virions/mL) (46, 47), short turnover times (1-2 days) (46, 48-50), and high rates of recombination and mutation (33, 51, 52).

HIV-1 mutates at an average rate of $\sim 3.6 \times 10^{-5}$ mutations/base pair/cycle (m/bp/c), corresponding to roughly one mutation per three genomes synthesized (53-60). This mutation rate is ~ 10 – 1000 fold higher than DNA viruses (10^{-6} – 10^{-8} m/bp/c) (61) and at least 10,000-fold higher than for eukaryotic genomic DNA ($\leq 10^{-9}$ m/bp/c) (62). In addition to mutating rapidly, HIV-1 frequently undergoes recombination, an ability that stems from the co-packaging of precisely two RNA genomes into the vast majority of virions (63). However, only one DNA provirus is ultimately formed, such that retroviruses are considered pseudodiploid in nature (64). HIV-1 recombines at a rate of ~ 3 – 14 crossovers/genome/cycle (60, 65-70), which is at least 10-fold higher than the rate of mutation. However, recombination only results in novel viral variants when the co-packaged genomes are genetically distinct, as recombination serves to reshuffle pre-existing mutations. Figure 1-2 demonstrates how mutation (from a variety of different sources) and recombination can lead to emergence of new viral variants.

HIV-1 Recombination: Mechanisms and Determinants

As discussed above, HIV-1, like all retroviruses, co-packages two genomes into budding virions, which can be identical or non-identical (as in Figure 1-2) in sequence. Upon infection, reverse transcriptase (RT) converts the single-stranded genomic RNA into double-stranded DNA. During this process, RT can switch templates, either within the same RNA genome (intra-molecular) or between RNA genomes (inter-molecular). In fact, RT must switch between RNA templates at least once in order to complete reverse transcription (i.e. the first strand transfer). While this obligate strand transfer can be intra- or inter-molecular in nature (64, 71, 72), template switching is clearly not limited to this step of the reverse transcription cycle, as recombination can occur throughout the viral genome. The rate at which RT switches templates is thought to be controlled by the relative balance of polymerase and RNase H enzymatic activities, referred to as the dynamic copy choice model of recombination (73, 74). Factors that reduce the speed of RT-mediated DNA synthesis promote

template switching, such as low dNTP pool levels (73, 75-77), RNA secondary structures (78-80), and mutations in RT that impair processivity (73, 76, 77). In contrast, RT mutations that reduce RNase H activity decrease the level of template switching (73, 76, 81).

As a pre-requisite for recombination to occur between genetically distinct viruses, the producer cells must become dually infected, either as a result of co-infection (near simultaneous infection) or super-infection (sequential infection). Thus, factors that alter the incidence of dual infection through either mechanism will impact the frequency of recombination. HIV-1 readily co-infects primary CD4⁺ T-cells *ex vivo* with little evidence of interference (82-85). In fact, co-infection has been found to occur more frequently than expected from random interactions between viruses and cells (82, 83), but this may be due to reactivation of silent proviruses upon co-infection (86). In contrast to co-infection, cells infected by HIV-1 (and many other retroviruses) are resistant to re-infection, a phenomenon called super-infection resistance (87). Super-infection of humans at the organismal level has often been documented (88-93), though initial infection may be somewhat protective of re-infection (92). However, such individuals do not necessarily contain super-infected cells, as only a small fraction of CD4⁺ T-cells becomes infected by HIV-1. Of note, dually infected splenocytes can readily be observed in samples from HIV-1 infected individuals (94, 95) or SIV_{mac}-infected rhesus macaques (96). However, another study found that most peripheral blood CD4⁺ T-cells harbor only a single provirus (97). In a more recent extension of this work, ~90% of CD4⁺ T-cells were demonstrated to contain a single provirus in both peripheral blood and paired lymph node samples from five patients (98). Unfortunately, it is not yet clear whether the discrepancies between these studies are due to the different compartments sampled, unique features of the infected individuals, or other factors. Nonetheless, the wide prevalence of intra-subtype, inter-subtype, and inter-group recombinants of HIV-1 clearly demonstrates that recombination is a relevant mechanism of genetic variation that has shaped the worldwide AIDS

pandemic.

In addition to the incidence of dual infection, rates of retroviral recombination are influenced by factors that alter the ability of genetically distinct genomes to co-package into the same virus particle. For example, the HIV-1 genome contains the dimerization initiation signal (DIS), a six nucleotide palindromic sequence within the first stem loop of the 5' untranslated region. The DIS sequence is the dominant factor that drives HIV-1 genomic RNA co-dimerization and co-packaging (99-102). Thus, HIV-1 variants with matched DIS sequences co-package and ultimately recombine much more frequently than those with mismatched DIS sequences. Also, independent of the DIS sequence, the frequency of recombination correlates with the degree of sequence homology between templates, such that closely related sequences (e.g., from the same subtype of HIV-1) recombine more frequently than divergent sequences (e.g., from different subtypes of HIV-1) (102, 103).

HIV-1 recombination can both promote the completion of reverse transcription in the presence of RNA damage and also contribute to genetic diversity. While recombination events between identical co-packaged genomes are effectively silent, recombination events between distinct genomes (i.e. from heterozygous virions) can lead to novel viral variants through the reshuffling of mutations (Figure 1-2). For example, recombination can link beneficial mutations together, such as low-level drug resistance mutations into a highly drug-resistant complex (104) or single-drug resistance mutations into multi-drug resistance (105). Conversely, recombination can allow escape from deleterious or lethal mutations (106-108). While generally viewed as beneficial, recombination can also dissociate co-adapted mutations, leading to unfit viral variants (109). Overall, the importance of recombination to the establishment of the global AIDS pandemic is clear: 72 circulating recombinant forms (CRFs) have been identified to date (www.hiv.lanl.gov), accounting for ~20% of HIV-1 infections worldwide (110).

HIV-1 Mutagenesis: Mechanisms and Determinants

HIV-1 has been found to mutate at a rate of $\sim 3.6 \times 10^{-5}$ m/bp/c (average across all studies) (53-60), at least 10,000-fold faster than for eukaryotic genomic DNA (62). RT is thought to be a key driver of viral mutagenesis (Figure 1-2), primarily due to its high error rates *in vitro* (typically $\sim 10^{-4}$ m/bp) (111). RT is error-prone due to efficient extension from mismatched termini and lack of 5'-to-3' exonucleolytic proofreading activity. Further, the intermediates of reverse transcription (RNA-DNA hybrids in the cytosol) are thought to avoid host repair processes. RNA polymerase II can also generate mutations when transcribing the viral genomic RNA from the integrated proviral DNA (Figure 1-2), but limited evidence suggests this is a lesser source of error than RT (59). Cellular DNA polymerases could theoretically introduce mutations when replicating the integrated provirus during cell division, but the high fidelity of cellular DNA replication ($\leq 10^{-9}$ m/bp/c) argues that this is a relatively minor source of virus variation.

The fidelity of HIV-1 reverse transcription can be modulated by a variety of host and viral factors. Drug resistance-associated mutations in RT can increase or decrease the mutation rate of HIV-1 (54, 112-115). Approved anti-viral drugs, including nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTIs and NNRTIs) can elevate the HIV-1 mutation rate (112, 113, 115, 116). Experimental nucleoside analogs with altered base structures can lead to dramatic elevations of the viral mutation rate and shifts in mutational spectra (117-119). Anti-metabolites that alter endogenous dNTP pools can also increase HIV-1 mutant frequencies (120). In addition, primary sequence features and secondary structures can trigger mutations at a higher rate. For example, most insertions and deletions take place at homopolymeric runs during viral replication (53, 55, 121).

In addition to RNA and DNA polymerases, the APOBEC3 family of DNA deaminases can induce mutations in HIV-1 by performing C-to-U editing of minus strand viral DNA during reverse transcription (Figure 1-2) (122, 123). Ultimately,

APOBEC3 editing leads to the fixation of G-to-A mutations on the plus strand viral DNA. APOBEC3 proteins typically introduce numerous G-to-A mutations into the same viral DNA, termed G-to-A hypermutation. In fact, APOBEC3 proteins can mutate as many as 10% of the total guanines within certain regions of the HIV-1 genome (124). APOBEC3-mediated hypermutation is often lethal due to the many missense and nonsense mutations generated, restricting further viral propagation. The viral accessory protein Vif counteracts most APOBEC3 activity by targeting APOBEC3 proteins for degradation in producer cells (125-128). Specifically, Vif forms an E3 ubiquitin ligase complex by interacting with CBF- β (an obligate co-factor), elongins B and C (adapters), CUL5 (scaffold), and RBX2 (regulator of polyubiquitination) (129-132). This complex directs polyubiquitination and proteasomal degradation of APOBEC3 proteins, which precludes APOBEC3 packaging into budding virions and ultimately prevents APOBEC3 activity in infected target cells. Humans encode seven APOBEC3 family members: A3A, A3B, A3C, A3D, A3F, A3G, and A3H. Of these, A3D, A3F, A3G, and A3H (haplotypes II, V, and VII), are the most relevant for restriction of Vif-deficient HIV-1 in CD4⁺ T-cells (133-136). A3G preferentially mutates guanines within a GG dinucleotide context (137-139), whereas the other APOBEC3 proteins prefer a GA dinucleotide context (140-143). Thus, in some instances, the specific APOBEC3 family member responsible for hypermutation can be inferred based on the sequence context of edited sites.

Although Vif counteracts most APOBEC3 activity, multiple lines of evidence demonstrate that Vif-mediated protection is not absolute. In cell culture systems, the expression of Vif often does not fully restore the infectivity of virus produced in the presence APOBEC3 proteins (141, 144, 145). Vif alleles from different virus isolates have been shown to vary widely in their abilities to counteract various APOBEC3 proteins, with possible correlations depending on the HIV-1 subtype (144, 146). Vif variants that are partially or fully defective against specific APOBEC3 proteins can be readily identified in patient samples (145). In addition, G-to-A hypermutants have often been observed in proviral DNA from

patient samples (145, 147-153), which provides strong evidence that APOBEC3-mediated editing occurs *in vivo*. These hypermutants have been observed in both GG and GA dinucleotide contexts, implicating the activity of at least two APOBEC3 family members relevant to *in vivo* hypermutation.

APOBEC3-mediated editing of HIV-1 has traditionally been viewed as a lethal event, resulting in a hypermutated provirus that represents an evolutionary dead-end. Indeed, while G-to-A hypermutation is often detected in proviral DNA from patient samples, it is rarely observed in virion RNA (149, 154). Cell culture studies have found that hypermutants are subject to strong purifying selection, as the level of hypermutation in virion RNA is much lower than in proviral DNA or intracellular viral RNA (155). However, APOBEC3 proteins could potentially contribute to HIV-1 diversity, and thus adaptation and evolution, through two different mechanisms. First, APOBEC3-mediated mutation or hypermutation could be sub-lethal in rare instances, allowing further viral replication. Second, APOBEC3-driven hypermutation could be lethal but contribute to viral variation through recombination with a non-hypermutated genome. In the latter case, at least two viruses would be required to co-infect or super-infect the same cell, with only one of the viruses exposed to hypermutation by APOBEC3 activity.

Although still controversial, accumulating evidence suggests that both mechanisms are possible. Sadler *et al.* demonstrated that A3G sub-lethally mutates HIV-1 in the presence and absence of Vif (156). Mulder *et al.* showed that naturally occurring Vif variants defective against A3G promote the development of the M184I mutation in RT, which confers high-level resistance to the antiviral lamivudine (3TC) (157). Recombination between hypermutated and non-hypermutated virus populations enabled rapid 3TC escape. Another group demonstrated that A3G could promote evolution of 3TC resistance even in the context of wild-type Vif (158). In addition, G-to-A hypermutation has been observed in virion RNA, albeit at frequencies ~15-fold lower than in proviral DNA (155). Intriguingly, many of the hypermutated RNA genomes in virions contained mutations that would eliminate expression of functional Gag, which is required for

particle assembly and release. This finding strongly argues that hypermutated viruses were rescued through co-infection and complementation by non-hypermutated viruses. More recently, A3D and/or F were shown to promote diversification of HIV-1 (encoding a partially defective Vif variant) in a humanized mouse model of infection (159). Diversification was associated with emergence of mutations conferring CCR5/CXCR4 dual-tropism, suggesting that APOBEC3-mediated mutagenesis is capable of altering co-receptor usage. Lastly, studies of patient samples have found that G-to-A mutations occur more often in sequence contexts favored by specific APOBEC3 proteins than in disfavored contexts (154, 160). These results are consistent with the hypothesis that one or more APOBEC3 family members facilitate diversification of HIV-1 *in vivo*. Nonetheless, the relative importance of APOBEC3 proteins compared to other sources of error (such as RT) in driving the evolution of novel variants that confer drug resistance, immune escape, or altered co-receptor tropism, is still unclear.

In developed countries, the advent of highly active antiretroviral therapy (HAART) directed against HIV-1 has enabled the indefinite suppression of viral replication in the vast majority (>95%) of infected individuals, provided proper drug adherence is maintained. This raises the issue of the settings in which the high mutation and recombination rates of HIV-1 remain relevant. Prior to the initiation of treatment, rapid viral replication and diversification permit escape from CD8⁺ cytotoxic T-cell and neutralizing antibody responses, ultimately preventing effective immune system control of the virus. Defining the roles of viral mutation and recombination in these processes may inform efforts to develop an effective vaccine. Further, during HAART, a low level of ongoing viral replication may persist in areas of limited drug penetration, such as the gut-associated lymphoid tissue and central nervous system (161, 162), permitting continual viral evolution. However, the presence of ongoing viral replication during HAART has remained controversial, and the predominant mechanism of viral persistence is thought to be latent infection of resting CD4⁺ T-cells (163). HAART is also not completely suppressive in all individuals, as ~3% of infected

individuals develop triple-class virological failure (164), though this may result from suboptimal drug adherence. Additionally, high rates of mutation and recombination are features often shared by other RNA viruses (61, 165-167), many of which cannot currently be effectively countered by drug treatment. The investigation of these processes in HIV-1 will serve as a useful model system for studying the genetic diversification of other RNA viruses.

The HIV-1 Mutation Rate as a Therapeutic Target

As discussed earlier, HIV-1 rapidly diversifies within infected individuals into a viral quasispecies, a population of related but unique variants that allow for the exploration of sequence space (41, 168). Quasispecies theory was originally developed to describe the behavior of self-replicating short RNA molecules (called replicators), which are thought to have preceded the evolution of life on earth. At approximately the same time quasispecies theory was first being formulated, RNA viruses such as the bacteriophages MS2 and Q β were found to spontaneously mutate at high rates during maintenance of viral stocks (169-171). Over time, quasispecies theory was shown to apply to most RNA viruses, due to shared properties such as high mutation rates, small genomes, and fast turnover times. However, there are rare notable exceptions, such as the retrovirus HTLV-1, which develops little intra-patient diversity due to propagation primarily by oligoclonal expansion rather than *de novo* viral replication (172-175). It should also be noted that development of viral quasispecies does not necessarily always correspond with rapid evolution of the consensus viral sequence, as the consensus sequence can remain unchanged in the absence of positive selective pressures. However, HIV-1 is continually subject to such pressures *in vivo*, primarily from the immune response, such that diversity and divergence increase throughout chronic infection until the time of disease progression.

According to quasispecies theory, genetic information can only be stably preserved under a critical mutation rate, termed the error threshold (176, 177). The error threshold is thought to depend on genome size, relative viral fitness,

and population size. Above the error threshold, the viral quasispecies transitions into error catastrophe, a process in which the quasispecies loses informative genetic content and collapses. RNA viruses were postulated to naturally mutate close to the error threshold due to their high mutation rates and the finding that many virus particles are non-infectious. Thus, mutagenic agents resulting in even a small-fold increase in viral mutation rate may be able to drive the viral quasispecies toward extinction, referred to as lethal mutagenesis (119). John Holland and colleagues were the first to experimentally demonstrate lethal mutagenesis of a virus (178). Specifically, they showed that mutagenic nucleoside analogs (5-fluorouracil, 5-azacytidine) and mutagenic chemicals (ethyl methanesulfonate, nitrous acid) result in extreme reductions of poliovirus and vesicular stomatitis virus infectivities. Lethal mutagenesis has now been explored using a wide variety of RNA viruses and mutagenic agents (41, 179), primarily focusing on mutagenic nucleoside or nucleobase analogs, due to the prohibitive toxicity of mutagenic chemicals. Notably, HIV-1 lethal mutagenesis based on mutagenic nucleosides is similar to APOBEC3-mediated hypermutation that occurs naturally to restrict viral replication in the absence of Vif. Thus, targeting the Vif-APOBEC3 interaction (or other necessary interactions for APOBEC3 degradation, such as between Vif and CBF- β) represents an additional strategy to achieve lethal mutagenesis, and this avenue is actively being explored (180-183).

Although results in cell culture have been promising, the clinical translation of lethal mutagenesis has been a slow process, due largely to a lack of strong lead compounds. Notably, ribavirin may inhibit hepatitis C virus *in vivo* at least in part through a lethal mutagenesis mechanism, but this area remains controversial (184-187). The mutagenic nucleobase analog favipiravir (T-705) exerts broad-spectrum anti-viral activity against a number of RNA viruses in cell culture (188). T-705 is currently being investigated in phase III clinical trials for influenza in the United States and, more recently, in phase II clinical trials for Ebola virus in West Africa. T-705 has been shown to be mutagenic for influenza

in cell culture and for norovirus in a mouse model of infection (189, 190). However, T-705 has also been shown to directly inhibit the influenza viral RNA polymerase (191-193). Therefore, it remains unclear whether T-705 acts against influenza *in vivo* primarily as a lethal mutagen or as a direct inhibitor of viral replication.

The lethal mutagenesis of HIV-1 in cell culture was first demonstrated by Lawrence Loeb and colleagues using the nucleoside analog 5-hydroxy-2'-deoxycytidine (5-OH-dC), which was found to induce primarily G-to-A transitions in the virus (119). HIV-1 replication in the presence of 5-OH-dC resulted in virus extinction after 9-24 passages, with accompanying increases in the frequency of G-to-A mutations. Later, the combination of 5-OH-dC and a direct RT inhibitor (AZT) was shown to accelerate the extinction of HIV-1 relative to 5-OH-dC alone (194). More recently, the ribonucleoside analog 5-azacytidine (5-AZC, Figure 1-3A) and the related deoxyribonucleoside analog 5-aza-2'-deoxycytidine (5-dAZC, or decitabine, Figure 1-3A) have been demonstrated to induce lethal mutagenesis of HIV-1 (117, 118). 5-AZC can be incorporated into viral RNA during the late phase of replication, resulting in C-to-G transversions, or into viral DNA by RT during the early phase of replication, producing G-to-C transversions (Figure 1-4) (118). In order to be incorporated by RT, 5-AZC likely must first be reduced to decitabine by the cellular enzyme ribonucleotide reductase (RNR), but this hypothesis has not yet been formally tested. Decitabine also causes striking increases in G-to-C transversion frequencies as a result of being incorporated during the early phase of viral replication by RT (Figure 1-4) (117). For both 5-AZC and decitabine, transversions likely result from rapid hydrolysis and deformylation, which result in ring-opened derivatives predicted to base pair with cytidine or deoxycytidine, respectively (117, 118, 195-197). These structural rearrangements have been proposed to result from either spontaneous hydrolysis (195, 197) or DNA methyltransferase-mediated hydrolysis (196). In the case of HIV-1, 5-AZC and decitabine-mediated G-to-C transversions are thought to result from incorporation and mispairing during reverse transcription in

the cytosol (117, 118), suggesting that spontaneous hydrolysis is the more relevant mechanism. Additionally, the nucleoside analog 5,6-dihydro-5-aza-2'-deoxycytidine (KP-1212, Figure 1-3A) has been shown to inhibit HIV-1 replication by lethal mutagenesis (198). However, in phase II clinical trials, KP-1461 (a prodrug of KP-1212) did not reduce viral loads and demonstrated only slight evidence of mutagenic effects, despite sufficient bioavailability (199, 200). In agreement with this, others have found that KP-1212 has little antiviral or mutagenic effects against HIV-1 in cell culture (120), though the reasons for these discrepancies remain unclear.

Nucleoside analogs with anti-HIV-1 activity, including conventional chain-terminating analogs as well as lethal mutagens, can often be potentiated by anti-metabolites targeting host enzymes involved in biosynthetic pathways (117, 201-204). Most of the anti-metabolites investigated for this purpose thus far have been ribonucleotide reductase inhibitors (RNRIs, see Figure 1-3B for examples), though other enzymes could potentially be targeted as well. In the absence of nucleoside analogs, RNRIs can reduce HIV-1 infectivity by depleting endogenous dNTP pool levels, which can increase viral mutant frequencies and/or inhibit viral DNA synthesis (120, 205). In combinations, RNRIs can potentiate nucleoside analogs, resulting in synergistic reductions in viral infectivity. Although the mechanisms of these combinations have not been elucidated in many cases, RNRIs likely enhance nucleoside analog transport into the cell, phosphorylation by cellular kinases, and/or incorporation by RT through reduced competition with endogenous nucleotides. For example, hydroxyurea (Figure 1-3B) exhibits synergy with the chain-terminating nucleoside analog didanosine (ddI, which is converted intracellularly into ddATP) by depleting endogenous dATP pools (204, 206). In fact, this combination has been extensively investigated in clinical trials, although hydroxyurea was ultimately not adopted as a common component of HAART due to dose-limiting toxicities (207-212). The RNRIs gemcitabine and resveratrol (Figure 1-3B) have been shown to synergize with decitabine in cell culture, without corresponding increases in cellular toxicity (117, 201). The

combination of gemcitabine and decitabine has further been shown to be effective in an MLV-based mouse model of AIDS (213). Likewise, combinations of KP-1212 with gemcitabine or resveratrol in cell culture exert greater antiviral activity than either compound alone (120). In some cases, RNRIs have even been shown to restore the activity of nucleoside analogs against analog-resistant viruses, effectively increasing the genetic barrier to resistance (214).

Dissertation Objectives

The research described in this dissertation was performed to address several over-arching hypotheses: First, that HIV-1 and HIV-2 would exhibit distinct mutation frequencies and spectra; Second, that decitabine would promote G-to-C transversions in HIV-1 in the absence of mutational hotspots; Third, that 5-AZC would act primarily as decitabine to enhance HIV-1 mutagenesis; Fourth, that the anti-HIV-1 activity of 5-azacytidine would be antagonized by RNRIs; Fifth, that KP-121 would potentiate the anti-HIV-1 activity of RNRIs.

More specifically, in Chapter II, viral mutagenesis was compared between HIV-1 and HIV-2. We hypothesized that HIV-2 would exhibit a lower mutation frequency than HIV-1, as HIV-2 has previously been found to exhibit a lower evolutionary rate, reduced viral fitness, and better control by the host immune response compared to HIV-1 (9, 215). To address this hypothesis, Illumina sequencing was performed of multiple homologous amplicons in HIV-1 and HIV-2 that were prepared from large quantities of proviral DNA. HIV-1 was found to display significantly higher mutation frequencies than HIV-2, with a mutational spectrum more heavily biased toward G-to-A mutations. These differences were largely driven by G-to-A hypermutants, which occurred primarily in the GA dinucleotide context, suggesting they were caused by APOBEC3-mediated editing. While HIV-2 G-to-A hypermutants were also observed, they occurred at much lower frequencies than for HIV-1 in all amplicons examined. In the absence of G-to-A hypermutants, HIV-1 and HIV-2 total mutation frequencies

were not significantly different, and the resulting mutation spectra of the viruses were much more similar. While these results suggest that HIV-2 might be less susceptible than HIV-1 to G-to-A hypermutation, these findings need to be confirmed in cell types expressing physiologically relevant levels of APOBEC3 proteins. Overall, these data support the conclusion that differences in general replication fidelity (i.e. in the absence of APOBEC3 proteins) are likely not a primary driver of the unique clinical features observed during HIV-2 infection.

Next, in Chapter III, the mutagenic effects of decitabine (Figure 1-3A), a nucleoside analog with potent anti-HIV-1 activity, were investigated in tremendous depth using Illumina next-generation sequencing to test the hypothesis that decitabine-mediated mutagenesis of HIV-1 would not result in clear mutational hotspots. Illumina sequencing permitted detailed analyses of mutation frequencies, spectra, sequence context preferences, mutational hotspots, and analysis of drug susceptibility at the level of individual sequence positions. This analysis showed that decitabine resulted in an extreme increase (~155-fold) in the frequency of G-to-C transversions in HIV-1, with a corresponding shift in mutational spectra. In addition, decitabine was found to induce a significant increase (~29-fold) in the frequency of C-to-G transversions. Surprisingly, G-to-C mutational hotspots (defined as upper outliers within the mutation distribution) were not observed. Notably, every single guanine position (134/134) and nearly every cytosine position (163/169) were found to be significantly susceptible to decitabine-mediated mutagenesis, demonstrating that incorporation of this compound into HIV-1 is highly promiscuous. Overall, these data support the use of decitabine-mediated mutagenesis as an antiviral strategy for extinguishing the infectivity of HIV-1.

Additionally, in Chapter IV, experiments were performed to test the hypothesis that the ribonucleoside analog 5-AZC (Figure 1-3A) would act primarily after reduction to 5-aza-2'-deoxycytidine (i.e. decitabine) to reduce HIV-1 infectivity by lethal mutagenesis. 5-AZC has previously been shown to be active during both the early and late phases of viral replication (118), thus

implying that it can be incorporated into viral DNA (by RT) and viral RNA (by RNA polymerase II), respectively. 5-AZC could potentially be incorporated by RT directly as a ribonucleotide (i.e. as 5-AZC-TP) or as a deoxyribonucleotide (i.e. as decitabine-TP), which would first require intracellular conversion to decitabine by RNR. While HIV-1 RT has been shown to selectively exclude ribonucleotides (216-218), HIV-1 RT has also been found to incorporate significant levels of endogenous ribonucleotides in cell types with low levels of deoxyribonucleotides, such as in macrophages (219, 220). Notably, high concentrations of 5-AZC are required to elicit antiviral activity (118), which could potentially alter the ratio of NTPs to dNTPs enough to allow for significant incorporation of 5-AZC-TP. We used a combination of *in vitro* RT incorporation assays, liquid chromatography-tandem mass spectrometry (LC-MS/MS), and Illumina sequencing to convincingly demonstrate that 5-AZC acts primarily as decitabine-TP against HIV-1, such that 5-AZC and decitabine result in extremely similar patterns of enhanced mutagenesis in HIV-1.

Further, in chapter V, we examined the effects of RNRIs on the potency of 5-AZC, hypothesizing that RNRIs would inhibit the intracellular conversion of 5-AZC to decitabine and thus antagonize the activity of 5-AZC. Surprisingly, we found that low concentrations of four different RNRIs all significantly potentiated the antiviral activity of 5-AZC, resulting in antiviral synergy and improvement of the selectivity index of 5-AZC. The mechanisms of these novel antiviral combinations were addressed using LC-MS/MS to measure levels of decitabine-TP and endogenous dNTPs, Illumina sequencing to determine mutation frequencies and spectra, and qPCR to measure levels of reverse transcription products in the presence of the drug combinations. We found that RNRIs did not effectively inhibit the reduction of 5-AZC to decitabine, suggesting that 5-AZC acts primarily as decitabine against HIV-1 even in the presence of RNRIs. The mechanism of antiviral synergy was further investigated for the combination of 5-AZC and resveratrol and was found to be primarily due to reduction of RT products rather than enhanced viral mutagenesis. Overall, these findings

demonstrate that low concentrations of RNRI can be applied in drug combinations to potentiate the activity of a ribonucleoside analog against HIV-1, which may hold true for other ribonucleoside analogs as well and thus open new avenues for the design of synergistic drug combinations directed against HIV-1 and other viruses.

Lastly, in Chapter VI, the experimental mutagenic nucleoside KP-1212 (Figure 1-3A) was studied alone or in combination with various RNRI. The primary goal of this project was to identify a strategy to significantly improve the antiviral activity of KP-1212, as a prodrug form of KP-1212 (KP-1461) failed to reduce viral loads in phase II clinical trials (199, 221). KP-1212 was found to substantially potentiate the anti-HIV-1 activities of multiple RNRI, despite having little antiviral or mutagenic activity on its own. The combinations were associated with significant increases in HIV-1 mutant frequencies, arguing that enhanced viral mutagenesis was the primary antiviral mechanism of the combinations. The combination of resveratrol and KP-1212 was found to strongly shift the mutational spectrum of HIV-1 toward G-to-C transversions. These findings will ultimately inform future attempts to clinically translate KP-1212 (and related nucleoside analogs) for the treatment of HIV-1 infection by lethal mutagenesis.

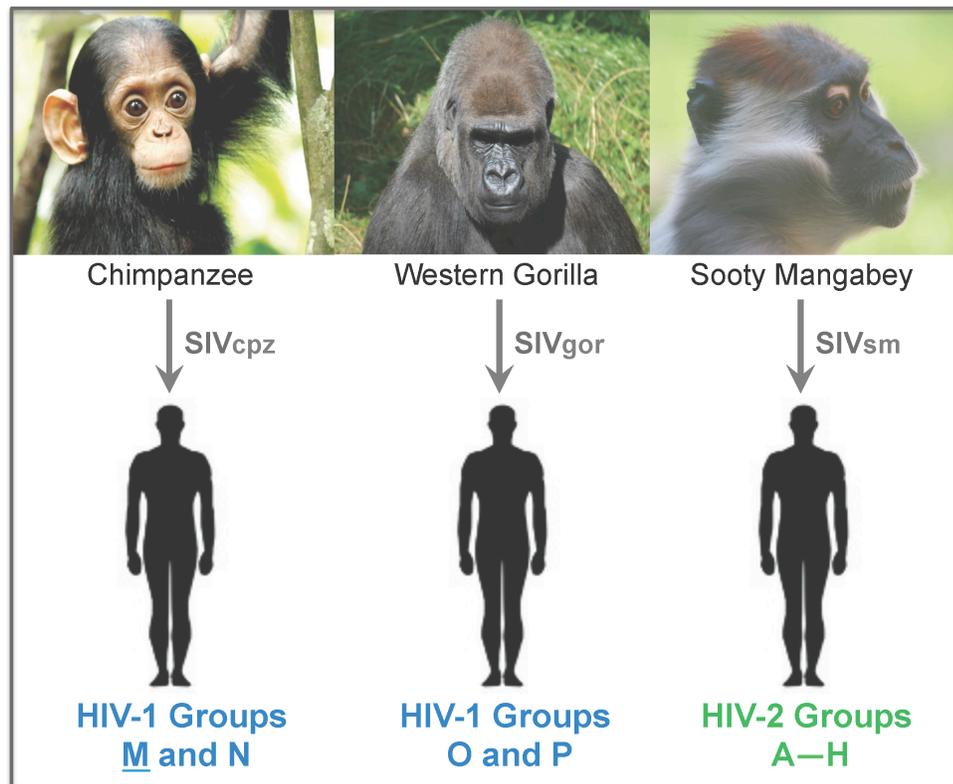


Figure 1-1. Origins of HIV-1 and HIV-2. HIV type-1 (HIV-1) resulted from the transmission of simian immunodeficiency virus (SIV) from chimpanzees or western gorillas into humans. Cross-species transmission has occurred at least four times, resulting in groups M-P, although group M (for Major) is primarily responsible for the worldwide pandemic. Group M has further diversified into nine primary subtypes and >70 circulating recombinant forms. HIV type-2 (HIV-2) resulted from the cross-species transmission of SIV from sooty mangabeys in West Africa. Transmission has occurred at least eight times, resulting in groups A-H, although only groups A and B appear widespread.

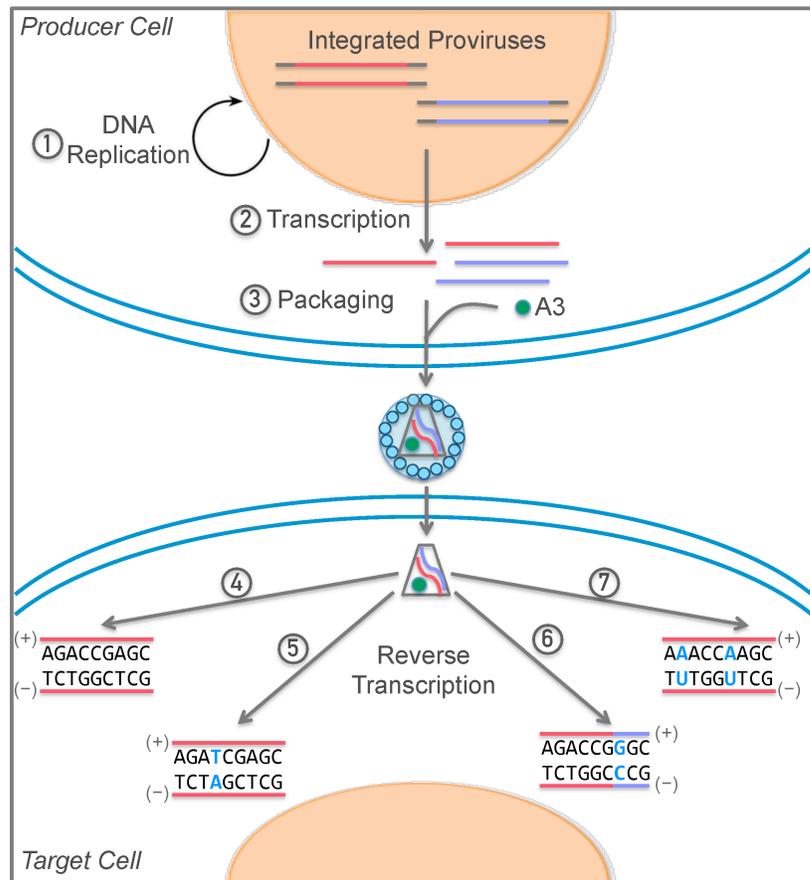


Figure 1-2. Mechanisms of HIV-1 recombination and mutation. Producer cells that become dually infected (either as a result of co-infection or super-infection) can result in the transcription of genetically distinct viral genomes (2) that may be able to co-package into virions (3). Upon infection of target cells, multiple outcomes are possible: Reverse transcription may occur in the absence of mutation or recombination (4). Alternatively, the error-prone reverse transcriptase may directly introduce novel mutations into the genome (5) or transfer pre-existing mutations from one viral genome to another by recombination (6). Further, APOBEC3 proteins can perform C-to-U editing of minus strand DNA in a processive fashion during reverse transcription (7), which ultimately results in the fixation of multiple G-to-A mutations into the genome,

termed G-to-A hypermutation. For this to occur, APOBEC3 proteins must be packaged into virions from the producer cells (3), which typically only takes place in the absence of functional Vif. Cellular DNA replication (1) and genome transcription (2) can also introduce mutations in HIV-1, but the available evidence suggests that these are less important sources of error.

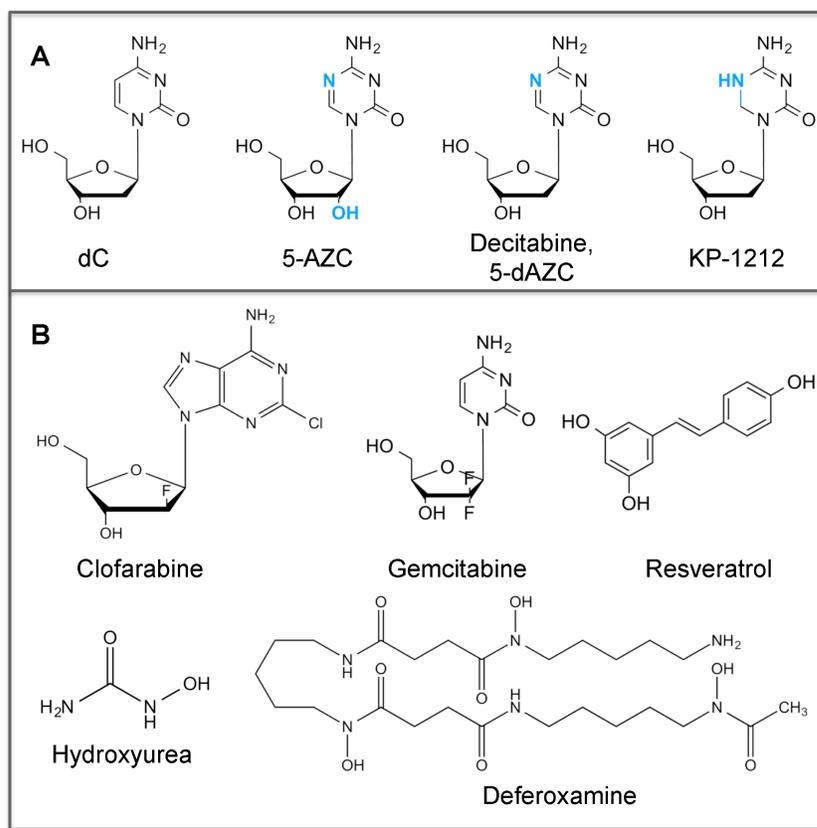


Figure 1-3. Compounds under investigation for the lethal mutagenesis of HIV-1. (A) Mutagenic nucleoside analogs. 5-azacytidine (5-AZC) and 5-aza-2'-deoxycytidine (5-dAZC, or decitabine) are incorporated in place of cytidine or deoxycytidine (dC), respectively, and undergo hydrolysis, which results in ring opening and base pairing with cytosine on the opposite strand. 5-aza-5,6-dihydro-2'-deoxycytidine (KP-1212) is also an analog of dC but is thought to cause mutations as a result of tautomerization. (B) Ribonucleotide reductase inhibitors (RNRIs). RNRIs deplete endogenous dNTP pools and often exhibit anti-HIV-1 activity by inhibiting viral DNA synthesis and elevating the viral mutation rate. However, they are primarily used in lethal mutagenesis strategies for their ability to potentiate the activity of mutagenic nucleoside analogs.

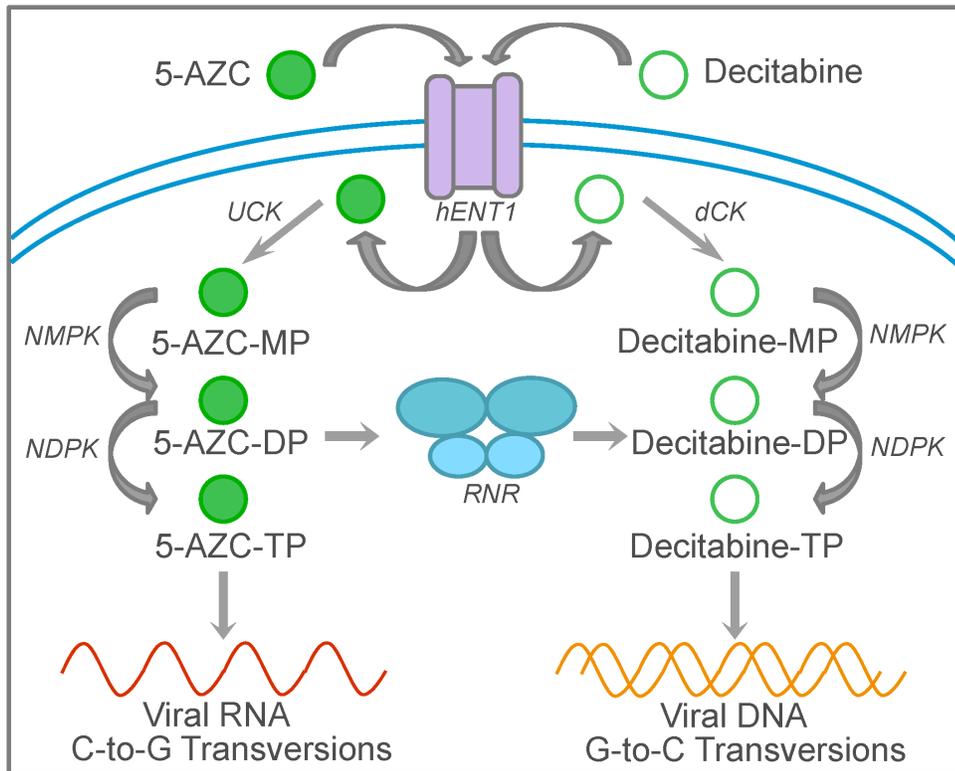


Figure 1-4. Model for the anti-HIV-1 activity of 5-azacytidine and decitabine.

Both 5-azacytidine (5-AZC) and decitabine are transported into the cell by facilitated diffusion through the human equilibrative nucleoside transporter 1 (hENT1). Next, they are phosphorylated by uridine-cytidine kinase (UCK) or deoxycytidine kinase (dCK), respectively, into the monophosphate (MP) forms. Nucleoside mono- and di-phosphate kinases (NMPK, NDPK) convert the monophosphate forms of 5-AZC and decitabine into the diphosphate (DP) and triphosphate (TP) forms, respectively. 5-AZC-TP can then be incorporated into viral genomic RNA, resulting in fixation of C-to-G transversions, while decitabine-TP can be incorporated into viral DNA, resulting in fixation of G-to-C transversions. 5-AZC-TP and decitabine-TP cause mutations by hydrolysis into ring-opened remnants, which then mispair with cytosine during synthesis of the complementary strand. 5-AZC may potentially also be reduced to decitabine by the cellular enzyme ribonucleotide reductase (RNR), promoting incorporation into viral DNA during reverse transcription.

CHAPTER II
HIV-1 AND HIV-2 EXHIBIT SIMILAR MUTATION FREQUENCIES AND
SPECTRA IN THE ABSENCE OF G-TO-A HYPERMUTATION

Reprinted with permission from: Rawson JM, Landman SR, Reilly CS, Mansky LM. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* (2015) 12/60. This article is openly distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)

J.M.R. and S.R.L. contributed equally to this work.

Introduction

Human immunodeficiency virus type-1 (HIV-1) infects approximately 35 million individuals worldwide and has resulted in about 39 million deaths since the onset of the AIDS pandemic (<http://www.unaids.org>). Within infected individuals, HIV-1 undergoes rapid genetic diversification, promoting the acquisition of drug resistance, evasion of the host immune response, and alterations in cell tropism. Genetic diversification is, in turn, driven by large population sizes and high rates of replication, recombination, and mutation. HIV-1 has been found to mutate on the order of 10^{-4} - 10^{-5} mutations/base pair/cycle (m/bp/c), corresponding to ~0.1-1 mutations per genome synthesized (53, 55, 57, 59, 222). HIV-1 thus mutates about 10,000-100,000 times faster than eukaryotic genomic DNA (62). Multiple viral and cellular factors influence both the rates and types of mutations produced during viral replication. Reverse transcriptase (RT) is likely a major generator of mutations *in vivo*, as it is tremendously error-prone *in vitro* (mutation rates $> 10^{-4}$ m/bp/c), primarily due to a lack of proofreading activity (111, 223). RNA polymerase II can also introduce mutations when transcribing the viral RNA genome, but less than that of RT (59). Cellular DNA polymerases can introduce mutations when replicating the proviral DNA integrated into the cellular genome; however, these enzymes have been assumed to minimally contribute to HIV-1 variation due to the high fidelity of genomic DNA replication (62). In addition, apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (APOBEC3) proteins can induce G-to-A hypermutation, particularly in the absence of Vif (122, 123). APOBEC3-mediated hypermutation is often lethal to virus replication, but accumulating evidence suggests that APOBEC3 proteins can in some cases promote genetic diversification and the evolution of new variants conferring drug resistance or altered cell tropism (156-160). Other factors, such as dNTP pool levels (224), other viral proteins (Vpr) (58, 113), and cell type (225) have been shown to influence HIV-1 mutagenesis as well.

Relative to HIV-1, human immunodeficiency virus type 2 (HIV-2) infection is often marked clinically by lower viral loads, reduced transmissibility, and longer asymptomatic periods in infected individuals (9, 215). Differences in the mutation frequencies of HIV-1 and HIV-2 have been hypothesized to contribute to the attenuated progression of HIV-2 observed clinically. Specifically, a lower mutation rate for HIV-2 would be expected to limit genetic diversification, which could in turn reduce viral fitness and/or attenuate viral pathogenicity. Consistent with this hypothesis, viral variants with increased replication fidelity have been shown to result in impaired viral fitness and virulence in many other RNA viruses (226-230). In further support of this hypothesis, HIV-2 has been found to evolve less quickly than HIV-1 (19, 20), though one report has found the opposite (21). However, it should be noted that evolutionary rates depend on a wide variety of factors besides the mutation rate, such as replication rate, population size, and selective pressures. In addition, the HIV-2 RT contains a highly conserved V75I polymorphism, a drug resistance-associated mutation in HIV-1 that improves RT fidelity (231, 232). Lastly, HIV-2 has been shown to be less sensitive than HIV-1 to APOBEC3G activity, potentially diminishing the contribution of APOBEC3G to viral mutagenesis (233).

In order to compare mutation frequencies and spectra between HIV-1 & 2, we performed Illumina high-throughput sequencing of proviral DNA from cells infected with HIV-1 or 2. We found that HIV-2 displayed lower total, substitution, and transition mutation frequencies than HIV-1, particularly due to reduced levels of G-to-A transition mutations. We also observed low-level G-to-A hypermutation in both HIV-1 and HIV-2 that was consistent with the activity of APOBEC3 proteins. Intriguingly, HIV-2 demonstrated significantly lower levels of G-to-A hypermutation than HIV-1. After exclusion of G-to-A hypermutants, total mutation frequencies were not significantly different between HIV-1 and HIV-2. Together, these data support the conclusion that differences in general replication fidelity are likely not a primary driver of the unique clinical features of HIV-2 infection.

Results

Characterization of background errors induced by PCR and Illumina sequencing. In order to compare mutation frequencies and spectra between HIV-1 and HIV-2, single cycle infections with HIV-1 or HIV-2 were performed at a high MOI (1 million U373-MAGI-X4 cells per replicate infected at an MOI of 1.0, see Materials and Methods and Figure 2-1). In this assay, producer cells cannot be re-infected due to a lack of the appropriate receptor and co-receptor, and target cells likewise cannot be re-infected due to disruption of the *env* genes in the HIV-1 and HIV-2 vectors. Genomic DNA was purified from infected cells and first subjected to quantitative PCR (qPCR) in order to determine the level of plasmid carryover from transfections. Plasmid carryover was quantified either by: 1) determining the plasmid backbone copy number (by measuring the ampicillin resistance gene) and dividing by the proviral copy number, or 2) determining the proviral copy number from heat-inactivated viral infections and dividing by the proviral copy number from un-treated infections (see Materials and Methods). We found that the level of plasmid carryover for HIV-1 was 0.2% when measured by either method, while the level of carryover was 2.8% or 1.4% for HIV-2, depending on the approach used (Table S2-1). The significantly higher level of plasmid carryover for HIV-2 likely reflects the reduced infectivity of HIV-2 viral stocks, which resulted in larger volumes of viral stocks being used during infection. These results are comparable to those obtained in another study (222) and are too low to significantly impact measured mutation frequencies. Next, amplicons were prepared from proviral DNA for Illumina sequencing. In total, 12 samples were analyzed--three experimental replicates each of HIV-1, HIV-2, and HIV-1 and HIV-2 plasmid amplifications as controls to determine levels of background errors. Further, for each sample, five amplicons were prepared (Gag, Vif, HSA, EGFP-1, and EGFP-2), representing a mixture of viral (Gag, Vif) and marker (HSA, EGFP-1, EGFP-2) gene targets. Libraries were prepared individually from samples in order to prevent inter-sample

recombination during library construction. Following this, all libraries were pooled and subjected to 2×150 paired-end sequencing on the Illumina MiSeq, resulting in ~4.7 million total read pairs after processing, or an average of ~79,000 read pairs/amplicon/sample (Table S2-2). After stringent filtering of Illumina data, the mutation frequencies (expressed as mutations per base pair, or m/bp) were determined for all samples, both in terms of total mutations and every possible subdivision (i.e. substitutions, transitions, transversions, etc.). Mutation counts, frequencies, and relative percentages are listed in Dataset S2-1, both combined across all five amplicons and separated by amplicon.

After sequencing, the first objective was to utilize the plasmid controls to characterize the frequencies and spectra of background errors (i.e. errors from PCR or sequencing) in order to determine the extent to which biological mutations could be detected above the level of the background. The average mutation frequencies of the plasmid controls were 2.8 (HIV-1) and 2.6 (HIV-2) × 10⁻⁴ m/bp (Figure 2-2A and Dataset S2-1), consistent with a recent investigation into background error during amplicon sequencing on the Illumina MiSeq (234). Most of the background errors observed were substitutions, with insertions and deletions comprising only 4.3-4.9% of total mutations. Of the substitutions, transversions and transitions were observed at similar frequencies: 1.2-1.4 × 10⁻⁴ m/bp for transversions and ~1.2 × 10⁻⁴ m/bp for transitions, such that each category composed ~ half of all mutations (Figure 2-2C). However, as described in Materials and Methods, plasmid error hotspots (i.e. positions highly prone to background error) were masked prior to this analysis, and in the absence of such masking transversions occurred ~2.5-fold more often than transitions. Thus, our findings are consistent with previous reports that have observed a bias of Illumina sequencing (using a variety of different platforms and library preparation methods) toward transversion types (234-237). Interestingly, among the eight possible transversion types, background transversion errors were non-randomly distributed (Figure S2-1). Specifically, background transversions were strongly biased toward C-to-A and G-to-T transversions, which composed 74% (HIV-1) or

73% (HIV-2) of all transversions. C-to-A and G-to-T are reciprocal mutational types, and these mutations may be due to oxidative conversion of guanine to 8-oxoguanine or due to Illumina imaging artifacts (see Discussion). Background errors were distributed quite evenly across the five amplicons (Figure 2-2B), demonstrating that no individual amplicon was particularly prone to background errors.

In addition to analyzing background error frequencies and spectra, the frequency of intra-sample recombination due to PCR was also examined. While PCR-mediated recombination is not known to be mutagenic, recombination could affect associations between mutations. We attempted to minimize recombination from PCR by stopping reactions after 30 cycles of amplification, corresponding to the ~end of the log-linear phase of amplification, after which recombination occurs at a much higher rate due to saturation of dNTPs and primers (238-240). Intra-sample recombination frequencies were determined using genetic markers that were incorporated into the plasmid control amplifications (see Materials and Methods). Under these amplification conditions, intra-sample recombination was observed at frequencies of ~2-3% of the maximum observable in the assay, demonstrating that recombination from PCR was relatively rare (Table S2-3).

We next compared the mutation frequencies and spectra of the HIV-1 & 2 biological samples to that of the plasmid controls in order to determine which types of mutations could be detected above background levels. HIV-1 and HIV-2 exhibited average mutation frequencies of 6.9 (HIV-1) and 3.1 (HIV-2) $\times 10^{-4}$ m/bp (Figure 2-2A and Dataset S2-1). The total mutation frequency of HIV-1 was significantly higher than the corresponding plasmid control ($p < 0.001$). In contrast, the HIV-2 total mutation frequency was not significantly higher than the plasmid control ($p = 0.40$). Upon separating out the major classes of mutations, we found that this was primarily due to high levels of transversions in the plasmid controls. Transversions occurred at frequencies of 1.4 (HIV-1) or 1.2 (HIV-2) $\times 10^{-4}$ m/bp in the plasmid controls (Figure 2-2A), values that were not significantly different from that of the biological samples. Similar to transversions, insertion

frequencies did not significantly vary between biological samples and plasmid controls, and deletion frequencies were only significantly higher than plasmid for HIV-2 ($p = 0.02$, Dataset S2-1). However, insertions and deletions had little impact on overall mutation frequencies because they were much less frequent (3-6% of total mutations) than substitutions (Figure 2-2C). Notably, although the HIV-2 total mutation frequency was not significantly higher than its corresponding plasmid control, transition frequencies (5.8 [HIV-1] and 2.0 [HIV-2] $\times 10^{-4}$ m/bp) were significantly higher than the plasmid controls for both viruses (HIV-1: $p < 0.001$; HIV-2: $p = 0.038$). These differences were also reflected in the mutation spectra, as transitions comprised 84% (HIV-1) or 66% (HIV-2) of total mutations in the biological samples but only 44% (HIV-1) or 48% (HIV-2) in the plasmid controls (Figure 2-2C). Importantly, many previous reports have demonstrated that transitions predominate during the replication of HIV-1, generally comprising 70-90% of all substitutions (53, 55, 57-59, 225, 241). Thus, considering that transitions are more relevant to HIV-1 replication and that they were detected at levels significantly higher than the background, most downstream analyses focused on transition mutational types.

HIV-2 exhibits a lower mutation frequency and an altered mutation spectrum relative to that of HIV-1. We next compared HIV-1 and HIV-2 mutation frequencies in order to test our initial hypothesis that HIV-2 would display a lower mutation frequency than HIV-1. We found that HIV-1 had a significantly higher total mutation frequency, as well as higher frequencies of substitutions and transitions, than HIV-2 (relative differences of 2.3, 2.3, and 2.9-fold, respectively; all p -values < 0.001) (Dataset S2-1 and Figure 2-2A). In contrast, the levels of transversion mutations were not significantly different between HIV-1 and HIV-2 ($p = 0.32$). The observed differences in transition frequencies were primarily due to an approximately 6.9-fold higher frequency ($p < 0.001$) of G-to-A transition mutations with HIV-1 than with HIV-2 (Dataset S2-1). Less striking, but statistically significant, differences were observed for several other types of transitions as well. HIV-1 displayed a 1.1-fold higher frequency of A-to-G

transitions ($p = 0.049$) than HIV-2, while HIV-2 displayed a 1.5-fold higher frequency of C-to-T transitions than HIV-1 ($p = 0.0019$). The frequency of T-to-C transitions was not significantly different between HIV-1 and HIV-2. Consistent with these results, the mutation spectrum of HIV-1 was much more heavily biased toward G-to-A transition mutations than for HIV-2 (66% vs. 22% of total mutations, Figure 2-2C). Overall, these data demonstrate that the HIV-1 mutation frequency is significantly higher than that of HIV-2, predominantly due to substantially higher levels of G-to-A transition mutations.

Mutation frequencies vary across amplicons for HIV-1 and HIV-2. We hypothesized that specific amplicons might be more or less error-prone than others due to features of the primary sequence (such as homopolymeric runs), secondary structures, or the relative positioning of the amplicon within the viral genome. To address this, transition frequencies were compared across the five amplicons analyzed in this study. As previously indicated, transitions in the plasmid controls were distributed relatively evenly across the five amplicons, demonstrating that no individual amplicon was particularly prone to background transition mutations (Figure 2-2B). In contrast, for HIV-1, it was found that most transitions were concentrated in the EGFP-1 amplicon and, to a lesser extent, the Vif amplicon (Figure 2-2B). The EGFP-1 and Vif amplicons together accounted for about 74% of all transitions in HIV-1. The relative order of transition frequencies among the amplicons was found to be EGFP-1 > Vif > EGFP-2 \approx HSA \approx Gag, with all indicated differences between amplicon pairs being statistically significant ($p < 0.001$). Since G-to-A transitions predominated for HIV-1, the observed differences could potentially be explained by varying nucleoside content between amplicons. However, the amplicons contained relatively similar frequencies of deoxyguanosine, ranging from 19.6-31.2%. Further, the Vif amplicon actually contained the lowest deoxyguanosine content of the five amplicons, despite having the second highest transition frequency. For HIV-2, transition frequencies were much more evenly distributed between amplicons than for HIV-1, with a maximal difference of \sim 1.3-fold (between Gag

and EGFP-2). We also compared transition frequencies between HIV-1 and HIV-2 at the level of individual amplicons. HIV-1 was found to have a higher frequency of transitions than HIV-2 in all amplicons except Gag ($p = 0.12$ for Gag; $p < 0.001$ for all other amplicons); however, the greatest differences were observed in the EGFP-1 and Vif amplicons (7.2 and 2.9-fold differences, respectively).

Detection of G-to-A hypermutation in HIV-1 and HIV-2. Considering the high levels of G-to-A transition mutations (particularly for HIV-1) that we observed (Figure 2-2C), we next investigated whether HIV-1 or 2 displayed evidence of G-to-A hypermutation. G-to-A hypermutation could potentially arise from the activity of APOBEC3 proteins, though relatively little APOBEC3 activity was expected in this particular experimental system (see Discussion). In these analyses, hypermutants were defined as read pairs (~120 bp in length) that contained two or more mutations of the same type within the read pair. In addition to analyzing G-to-A hypermutants, we determined frequencies of other possible types of transition hypermutants (A-to-G, C-to-T, and T-to-C) as well, as HIV-1 A-to-G hypermutants have occasionally been reported in the literature (54, 55). All hypermutant counts and frequencies (defined as hypermutant read pairs/all read pairs) are tabulated in Dataset S2-2, either combined across all amplicons or separated by amplicon.

In HIV-1, the frequency of G-to-A hypermutants was approximately 7.9×10^{-3} (or ~1 of 127 read pairs), while in HIV-2 the frequency was much lower, approximately 2.8×10^{-4} (or 1 of 3623 read pairs) (Figure 2-3A). The average G-to-A hypermutation frequencies were about 1060-fold higher for HIV-1 or about 17-fold higher for HIV-2 as compared to their corresponding plasmid controls, respectively. HIV-1 and HIV-2 G-to-A hypermutant frequencies were significantly higher than the plasmid controls in all five amplicons examined (p -values ranging from <0.001 to 0.01). The G-to-A hypermutation frequency was much higher (i.e., about 28-fold) in HIV-1 than observed in HIV-2 ($p < 0.001$), demonstrating that HIV-2 was less susceptible to G-to-A hypermutation in this experimental system.

This trend was significant in all five amplicons examined ($p < 0.001$). Other types of transition hypermutants were also observed (Figure 2-3A), but they were far less frequent than G-to-A hypermutants and were not observed at levels significantly higher than that of the plasmid controls. In HIV-1, G-to-A hypermutants comprised about 98% of all transition hypermutants, indicating a clear dominance over other possible types of transition hypermutants. When subdivided by amplicon, the distribution of HIV-1 G-to-A hypermutation was found to resemble that of all transitions (Figure 2-2B), with most G-to-A hypermutation concentrated in the EGFP-1 and Vif amplicons (Figure 2-3B). For example, after removal of G-to-A hypermutants, transition frequencies were very consistent across all five amplicons, demonstrating that G-to-A hypermutants were responsible for the elevated transition frequencies in EGFP-1 and Vif (Figure S2-2). The overall observed trend of G-to-A hypermutation among amplicons in HIV-1 was EGFP-1 > Vif > EGFP-2 > HSA > Gag (all p -values < 0.001). Despite being much less frequent, HIV-2 G-to-A hypermutants were also concentrated mainly in the EGFP-1 and Vif amplicons, although differences between amplicons did not reach statistical significance. These observations indicate that G-to-A hypermutation can significantly vary between different genes and even between different regions of the same gene (i.e. EGFP-1 vs. EGFP-2).

G-to-A hypermutation was further analyzed by determining the number of G-to-A mutations per hypermutant read pair. Most G-to-A hypermutants contained low numbers of G-to-A mutations (medians of 4 for both viruses), as expected for these short (~120 bp) read pairs (Figure 2-3C). However, a minority of G-to-A hypermutants contained high numbers of G-to-A mutations (maxima of 22 and 21 for HIV-1 and HIV-2, respectively). In contrast to the biological samples, the ultra-rare G-to-A hypermutants observed in the plasmid controls nearly all contained only two G-to-A mutations (one triple G-to-A mutant was observed for the HIV-2 plasmid). Next, the dinucleotide contexts of G-to-A mutations in hypermutants were analyzed, as APOBEC3 proteins are strongly associated with GG or GA dinucleotide contexts (137, 141-143), depending on

the specific APOBEC3 family member. Most G-to-A mutations were found within the GA dinucleotide context for both viruses, and, strikingly, only 1% (HIV-1) or 4% (HIV-2) were found to occur in either the GT or GC sequence contexts (Figure 2-3D). These dinucleotide contexts were markedly different from the contexts of the rare G-to-A hypermutants identified in the plasmid controls (Figure 2-3D) as well as from the contexts of single G-to-A mutants (Figure S2-3). Although the dinucleotide context appears to differ between the HIV-1 and HIV-2 plasmid controls, it should be noted that these contexts are based on extremely low numbers of hypermutants for the plasmids (10 for HIV-1 and 17 for HIV-2). The observed bias in the virological samples was not caused by a higher prevalence of the GA dinucleotide (relative to GG, GC, and GT) within the amplicon sequences, as GA dinucleotides only accounted for 24% of all guanine-containing dinucleotides. Taken together, these findings support the conclusion that the low level of G-to-A hypermutation observed was primarily caused by one or more APOBEC3 family members, despite the presence of Vif in our experimental system.

The patterns of G-to-A hypermutation were further characterized by investigating whether G-to-A mutations in hypermutants occurred at hotspots and, if so, whether sequence preferences (beyond the dinucleotide context) could be identified. In this analysis, we defined G-to-A hypermutation hotspots as statistical upper outliers using the $1.5 \times$ interquartile range (IQR) rule. Using this criterion, 18 (HIV-1) or 16 (HIV-2) hotspots for G-to-A hypermutation were identified, all of which were located in the Vif or EGFP-1 amplicons (Table S2-4). Of these, 14 (13 in EGFP-1 and 1 in Vif) were shared between HIV-1 and HIV-2, suggesting a common mechanism of mutation. Because specific APOBEC3 proteins have been shown to exhibit additional sequence preferences at the -1 and +2 positions (relative to the mutated guanine), sequence logos for total G-to-A hypermutation hotspots, GA context hotspots, and GG context hotspots were generated. However, statistically significant preferences at other positions were not observed for HIV-1 or HIV-2 in any of these categories. Furthermore, in

order to gauge the potential effects of hypermutation on protein expression and/or function, the coding changes introduced at G-to-A hypermutation hotspots were examined (Table S2-4). G-to-A hypermutation did not result in the introduction of premature stop codons at any of the examined hotspots. Missense mutations were typically R-to-K (semi-conservative), D-to-N (semi-conservative), or E-to-K (non-conservative); however, other types were occasionally observed. Interestingly, several of the Vif mutations occurring in G-to-A hypermutants at these hotspots have been previously characterized (see Discussion) (159).

HIV-1 and HIV-2 mutation frequencies and spectra are not significantly different in the absence of G-to-A hypermutants. The HIV-1 and HIV-2 mutational data was analyzed with and without G-to-A hypermutants in order to: 1) estimate the extent to which G-to-A hypermutation contributed to the total mutational data, and 2) determine whether G-to-A hypermutation was primarily responsible for observed differences between HIV-1 and HIV-2 mutation frequencies and spectra. While removal of G-to-A hypermutants had little effect on HIV-2 mutation frequencies (reducing the total mutation frequency by only 5%), their removal reduced the overall HIV-1 mutation frequency by 53% and the G-to-A transition mutation frequency by 81% (Figure 2-4A). Thus, G-to-A hypermutation was responsible for approximately half of all mutations observed in HIV-1, despite being a relatively rare event (~1 of 127 read pairs were G-to-A hypermutants). In the absence of G-to-A hypermutants, HIV-1 and HIV-2 total mutation frequencies were found to be 3.2 or 2.9×10^{-4} m/bp respectively, a 1.1-fold difference that was not statistically significant ($p = 0.14$; Figure 2-4A). However, as noted earlier, HIV-2 displayed an ~1.1-fold lower frequency of A-to-G transitions ($p = 0.049$) and an ~1.5-fold higher frequency of C-to-T transitions ($p = 0.0019$) than HIV-1, and these findings were not altered by the removal of G-to-A hypermutants. Further, HIV-1 still demonstrated an ~1.8-fold higher frequency of G-to-A transitions ($p < 0.001$) than HIV-2 after removal of G-to-A hypermutants. As expected, the mutation spectra for HIV-1 and HIV-2 were also

much more comparable in the absence of the G-to-A hypermutation data (compare Figures 2-4B and 2-2C), although HIV-1 displayed a somewhat higher percentage of G-to-A (27% vs 17%) and lower percentage of C-to-T (14% vs 23%) transitions than HIV-2. Taken together, these observations suggest that most of the observed differences in viral mutation patterns between HIV-1 & 2 were due to the highly reduced levels of G-to-A hypermutants for HIV-2. These findings imply that the fidelities of other mutational sources, such as RT, are relatively similar for HIV-1 and HIV-2.

Discussion

Detection of rare mutations is often difficult with next-generation sequencing technologies due to high error rates, which for Illumina platforms typically range from $\sim 10^{-2}$ m/bp for unfiltered data and $\sim 10^{-3}$ - 10^{-4} m/bp for stringently filtered data (234-236, 242-245). We adopted numerous measures to minimize background errors due to PCR and sequencing, as well as to minimize PCR-mediated recombination (see Materials and Methods). We also performed control amplifications from plasmids in which we matched conditions as closely as possible with the biological samples. After stringent filtering, we obtained error frequencies of 2.8 (HIV-1) or 2.6 (HIV-2) $\times 10^{-4}$ m/bp for the plasmid controls (Figure 2-2A). The frequencies and distribution of background errors were nearly identical between the HIV-1 & 2 plasmid controls (Figure 2-2A-C), as expected considering that the amplicons were either identical (HSA, EGFP-1, EGFP-2) or $\sim 60\%$ homologous (Gag, Vif) in sequence. Consistent with previous reports on Illumina background errors (234-237, 245), the spectra of background errors were more heavily biased toward transversions than biological samples (Figure 2-2C), specifically C-to-A and G-to-T transversions (Figure S2-1). This trend was observed despite prior masking of plasmid error hotspots, which were also mostly ($\sim 83\%$) C-to-A and G-to-T transversions (see Materials and Methods and Table S2-5). Two mechanisms for these transversion types have been suggested in the literature (234-237, 246): 1. Oxidative conversion of guanine to

8-oxoguanine during sample processing, which would lead to fixation of C-to-A and G-to-T mutations during PCR, and 2. Imaging errors that result from crosstalk between the C and A channels or between the G and T channels, as the fluorophores attached to these bases are excited by the same lasers and then distinguished using different filters. Due to these issues, we unfortunately were not able to make a detailed comparison of transversion frequencies between HIV-1 & 2. However, transition frequencies (5.8 [HIV-1] and 2.0 [HIV-2] $\times 10^{-4}$ m/bp) were significantly higher than the plasmid controls for both viruses (HIV-1: $p < 0.001$; HIV-2: $p = 0.038$), and many previous reports have established transitions as the dominant type of mutations that occur during HIV-1 replication (53, 55, 57-59, 225, 241). Thus, further analyses focused primarily on these mutational types.

Relative to previous estimates of the HIV-1 mutation rate (which range from 1.4 to 8.5×10^{-5} mutations/bp/cycle) (53-59, 222), we obtained a somewhat higher mutation frequency for HIV-1 of 6.9×10^{-4} m/bp (or 5.8×10^{-4} m/bp if considering only transitions). However, these discrepancies may be due to several key differences between the assays used and the ways in which the data were analyzed. First, the level of background error in previous studies was likely much lower. Most previous estimates of the HIV-1 mutation rate were determined using the LacZ α assay (53-55, 57, 58), in which a proviral *lacZ α* sequence is purified (using the Lac repressor or Hirt extraction), directly transformed into *E. coli* (without using PCR), and subjected to blue/white color screening, followed by Sanger sequencing of mutants. While this assay results in little background, it is low-throughput and can only detect and measure mutations within the *lacZ α* marker gene. Second, previously used marker gene assays (such as the LacZ α assay) necessarily underestimate true mutation rates because they only detect mutations leading to a phenotypic change (silent mutations are not detected unless co-occurring with other mutations). For the LacZ α assay, measured mutation rates are thought to underestimate actual

mutation rates by ~2 to 3-fold (55). In contrast, the assay we used based on direct PCR and Illumina sequencing of multiple amplicons should detect all mutations, provided that they do not prevent amplification. Third, most previous studies used a second marker gene (typically an antibiotic resistance gene) to select for infected cells or to select for transformed *E. coli* (53-59). This approach will not detect heavily mutated sequences (such as hypermutants) in which both marker genes are mutated and functionally disrupted. Fourth, actual mutation frequencies may vary between the genes analyzed here (*gag*, *vif*, *hsa*, and *egfp*) and mutational targets from previous studies. Fifth, many previous studies scored mutants harboring multiple mutations as single mutants for the purposes of calculating mutation rates (53-58). Also, in some cases background error frequencies were subtracted from the mutation frequencies of the biological samples in order to calculate mutation rates (222).

Unexpectedly, we identified G-to-A hypermutants for both HIV-1 & 2, though the frequency was much higher (~28-fold) for HIV-1 than HIV-2 (Figure 2-3A; $p < 0.001$). The presence of G-to-A hypermutants could not be attributed to background errors, as their frequencies were much higher in the biological samples than in the plasmid controls. Also, the patterns of G-to-A hypermutation (in terms of dinucleotide context preferences and hypermutation hotspots) were markedly distinct between the biological samples and plasmid controls (Figure 2-3D and data not shown). Unlike single G-to-A mutants from the biological samples, G-to-A hypermutants exhibited a strong bias toward GA dinucleotide contexts (Figures 2-3D and Figure S2-3). In sum, these findings support the idea that the G-to-A hypermutants were caused by low-level activity of one or more APOBEC3 proteins. However, APOBEC3G cannot be primarily responsible, due to its strong preference for GG dinucleotides (137, 141, 142, 247). Further experiments in which specific APOBEC3 proteins are down-regulated (through knockout or knockdown) are required to prove that the G-to-A hypermutants we observed are APOBEC3-mediated, as well as to identify the specific APOBEC3 protein(s) responsible.

The observation of G-to-A hypermutation was somewhat surprising because: 1) the HIV-1 and HIV-2 vectors used in these studies encoded for a functional Vif protein, 2) the expression levels of most APOBEC3 proteins are relatively low in 293T (i.e., the cells that produced the HIV-1 and HIV-2 vector viruses) (225), and 3) Vif-proficient and Vif-deficient viruses exhibit similar infectivities when produced in 293T cells (156, 159, 233, 248). Upon observing G-to-A hypermutation, we confirmed via Sanger sequencing that the *vif* genes in the HIV-1 and HIV-2 vectors were intact (data not shown). Notably, previous studies have demonstrated that Vif does not always fully neutralize the activity of APOBEC3 proteins (140, 141, 144, 145, 156). In addition, Vif from different subtypes or containing naturally occurring polymorphisms have been shown to vary widely in their neutralization capacities (144-146). Further, multiple APOBEC3 proteins (including B, C, D, and F) have been detected at the mRNA level in 293T cells and could be involved in the observed G-to-A hypermutation (225, 249). Previous studies by our group have identified rare G-to-A hypermutants occurring at GA dinucleotides by Sanger sequencing of clones (in which case the virus stocks were also generated in 293T cells) (120, 225). Taken together, the G-to-A hypermutation observed here was likely due to the failure of HIV-1 or HIV-2 Vif to fully neutralize low levels of APOBEC3 proteins present in 293T producer cells. Consistent with this hypothesis, G-to-A hypermutation was observed very infrequently even for HIV-1 (~1 of 127 read pairs were G-to-A hypermutants), despite contributing to about half of all mutations (Figure 2-4A). Although HIV-2 was less susceptible than HIV-1 to G-to-A hypermutation in this experimental system, further investigation will be required to determine whether these trends hold true under conditions of higher APOBEC3 expression. Intriguingly, HIV-2 Δvif has been reported to be less sensitive than HIV-1 Δvif to APOBEC3G (233), but susceptibilities to other APOBEC3 proteins (in the presence or absence of Vif) have not yet been compared. Further, HIV-1 and HIV-2 Vif were recently shown to interact with APOBEC3F and APOBEC3G through completely separate sequence

determinants, and differences in HIV-1 and HIV-2 Vif-induced degradation of specific APOBEC3 proteins were also noted (250).

Most of the G-to-A hypermutants occurred within the EGFP-1 and Vif amplicons (Figure 2-3B). This may be due in part to higher frequencies of GA dinucleotides in these particular amplicons. For HIV-1, the frequencies of GA dinucleotides (relative to GG, GT, and GC) were 13% (Gag), 32% (Vif), 9% (HSA), 32% (EGFP-1), and 30% (EGFP-2). Thus, the GA dinucleotide frequency varied maximally by 3.6-fold, whereas G-to-A hypermutation frequencies varied up to 26.3-fold (EGFP-1 vs Gag). However, APOBEC3-mediated hypermutation can also be influenced by broader sequence contexts and secondary structures (141, 142, 159, 160, 251-253). Further, APOBEC3 activity follows a twin gradient along the HIV-1 genome corresponding to the amount of time the minus strand viral DNA remains single-stranded, such that the positioning of the amplicon could affect hypermutant frequencies (124, 254). Thus, one or more of these other features may have favored hypermutation in the EGFP-1 and Vif amplicons. Full genome sequencing will be required to address the distribution of G-to-A hypermutation in more detail. G-to-A hypermutation hotspots were also identified (Table S2-4), which all occurred in the EGFP-1 and Vif amplicons. G-to-A hotspots did not result in any nonsense mutations, consistent with the notion that GA dinucleotide-biased hypermutation generates fewer stop codons than GG-biased hypermutation. However, the G-to-A hotspots did introduce a number of missense mutations, particularly D-to-N, E-to-K, and R-to-K. Surprisingly, all of the mutations resulting from the four hotspots in HIV-1 Vif were recently identified in another study in which humanized mice were infected with an HIV-1 variant that cannot neutralize APOBEC3D or F (159). One of these Vif mutants (E134K) lost the ability to neutralize APOBEC3G, raising the possibility that hypermutation itself can influence susceptibility to hypermutation in further rounds of replication.

We initially hypothesized that HIV-2 would exhibit a lower mutation frequency than HIV-1 due to its attenuated pathogenicity, reduced evolutionary

rate (19, 20), a conserved fidelity-improving V75I polymorphism in RT (231, 232), and reduced susceptibility to APOBEC3G (233). More specifically, HIV-2 Env has been found to diversify more slowly than HIV-1 Env when compared to HIV-1-infected individuals with high viral loads (19). However, HIV-1 and HIV-2 diversification rates were similar when compared to HIV-1-infected individuals with low viral loads. In the same report, HIV-2 Env was found to evolve more slowly than HIV-1 Env and to be subject to strong purifying selection. Indeed, HIV-2 infection appears to elicit broad and potent neutralizing antibody responses against Env more frequently than for HIV-1 (24). In another report, HIV-2 Env was found to exhibit a lower rate of synonymous substitutions than HIV-1, implying reduced viral mutation and/or replication rates (20), though another group has reported opposing findings (21). Unfortunately, it is difficult to compare the results of our analyses to these published reports due to these contradicting results, and larger studies of HIV-2 intra-patient diversification and evolution are clearly warranted. Nonetheless, we found that HIV-1 and HIV-2 exhibited similar total mutation frequencies in the absence of G-to-A hypermutants, suggesting that differences in replication fidelity do not have a major impact on differences in evolutionary or synonymous substitution rates between HIV-1 and HIV-2.

Conclusions

In sum, we have found that HIV-2 exhibited significantly lower total and transition mutation frequencies than HIV-1, as well as a mutation spectrum less biased toward G-to-A transitions. However, these differences were mostly due to a significantly higher G-to-A hypermutation frequency for HIV-1 than HIV-2. These findings raise the intriguing possibility that HIV-2 might be less sensitive than HIV-1 to APOBEC3-mediated hypermutation, consistent with a previous report (233), but additional experiments in other cell types will be required to fully address this question. After removal of all G-to-A hypermutants, HIV-1 and HIV-2 total mutation frequencies were not significantly different, although small but

significant differences were still observed in the frequencies of G-to-A and C-to-T transitions. Overall, these results suggest that the fidelities of other mutagenic processes (such as reverse transcription) are relatively similar between the two viruses. Nevertheless, we cannot rule out the possibility that HIV-1 and HIV-2 exhibit more minor differences in mutation frequencies or spectra that we were not able to detect or that differences would be observed in other cell types. Overall, these data imply that differences in replication fidelity are likely not a major contributing factor to the unique clinical features of HIV-2 infection.

Materials and Methods

Plasmids, cell lines, and reagents. The HIV-1 vector, pNL4-3 HIG, has been previously described (118). This vector contains a cassette encoding heat stable antigen (HSA), an internal ribosomal entry site (IRES), and enhanced green fluorescent protein (EGFP). The HIV-2 vector, pROD HIG, was created from pHIV-2 H0G (255), a kind gift from Dr. Wei-Shau Hu (HIV Drug Resistance Program, Frederick National Laboratory for Cancer Research, Frederick, MD). The pHIV-2 H0G vector does not express Vpr or EGFP due to multiple point mutations. In order to construct pROD HIG, the *vpr* and *egfp* genes were restored by site-directed mutagenesis using the QuikChange II XL kit (Agilent Technologies, Inc.; Santa Clara, CA), and the resulting vector was verified by DNA sequencing. Both HIV-1 and HIV-2 vectors contain intact open reading frames for all genes except for the *env* and *nef* genes. Vector viral stocks from pNL4-3 HIG were produced by co-transfecting with pNL4-3 Env, a kind gift from Dr. Eric Freed (HIV Drug Resistance Program, Frederick National Laboratory for Cancer Research, Frederick, MD). Vector viral stocks of pROD HIG were produced by co-transfecting pROD10-Env (256), a kind gift from Dr. Paula Cannon (University of Southern California, Los Angeles, CA). The human embryonic kidney (HEK 293T) cells were purchased from American Type Culture Collection (Manassas, VA) and maintained in Dulbecco's Modified Eagle's Medium (DMEM) from Cellgro (Manassas, VA) with 10% HyClone FetalClone III

(FC3) from Thermo Scientific (Waltham, MA) and 1% penicillin/streptomycin from Life Technologies (Grand Island, NY). U373-MAGI-CXCR4_{CEM} cells were obtained from Dr. Michael Emerman through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (257). U373-MAGI cells were maintained similarly to 293T cells but with addition of 1.0 µg/mL puromycin, 0.1 mg/mL hygromycin B, and 0.2 mg/mL G418 to the medium. For transfections, poly-L-lysine was from Newcomer Supply (Middleton, WI) and polyethylenimine (PEI) was from Polysciences, Inc. (Warrington, PA).

Virus production and titering. Virus was produced by co-transfecting pNL4-3 Hig or pROD Hig with pNL4-3 Env or pROD10 Env, respectively, into 293T cells via the PEI method (258). PEI stocks were prepared at 1 mg/mL by dissolving PEI in water, adjusting the pH to 7.0, and filtering through a 0.2 µm filter. 24 h before transfection, 4 million 293T cells/plate were seeded onto 10 cm plates pre-coated for 5 min with poly-L-lysine. For each plate, 10 µg of pNL4-3 Hig or pROD Hig + 5 µg Env expression plasmid + 45 µL 1 mg/mL PEI were combined with serum-free DMEM to a final volume of 1 mL. After 20 min of incubation, the medium on the 293T cells was replaced and the DNA-PEI mixture was added. The medium was replaced 16 h post-transfection, and viral stocks were collected 48 h post-transfection by filtering the supernatants through a 0.2 µm filter. For each viral stock, five plates were transfected, and the resulting supernatants were pooled and concentrated (~10-fold) using 100,000 MWCO filtration columns (Vivaproducts; Littleton, MA). Viral stocks were then treated with 10 U/mL of DNase I (New England Biolabs; Ipswich, MA) for 2 hr at 37 °C to degrade residual plasmid DNA from transfections. Viral stocks were then divided into 1.0 mL aliquots and frozen at -80 °C.

Prior to large-scale infections, viral stocks were first titered in U373-MAGI cells based on EGFP expression. The day before infection, 31,250 cells/well were plated in 24-well plates. After 24 h, the media was replaced and varying volumes of virus ranging from 15.625 to 500 µL (2-fold dilution series) were added. To improve infectivity, the cells were infected by spinoculation (1200×g

for 2 h at 24 °C). The media was replaced again 24 h post-infection and cells were collected at 72 h post-infection for analysis of EGFP expression by flow cytometry. The cells were treated with trypsin, transferred to 96-well plates, pelleted at 500×g for 5 min, and resuspended in 200 µL Dulbecco's phosphate-buffered saline (DPBS) + 2% FC3/well. EGFP expression from at least 10,000 gated cells was analyzed using a BD LSR II flow cytometer (BD Biosciences; San Jose, CA). EGFP was excited with a blue 488-nm laser and emission detected using 505LP and 525/50 filters. Virus titers (expressed as infectious units/mL) were calculated based upon EGFP expression at low infectivities (<40%) as previously described (259).

Infections for Illumina sequencing. In order to prepare samples for Illumina sequencing, 1×10^6 U373-MAGI cells were infected at a multiplicity of infection (MOI) of 1.0. These infections were performed in 24-well plates (31,250 cells/well) in order to avoid any potential effect of the plate format on infectious titer. Uninfected cells and cells infected with heat-inactivated viruses (i.e. virus stocks that were incubated at 65 °C for 1 h) were included as negative controls. The cells were infected by spinoculation, and the medium was replaced 24 h post-infection. Cells were collected for genomic DNA extraction at 72 h post-infection by treating with trypsin, pelleting, and washing three times with DPBS to further reduce plasmid carryover. Extra wells of infected cells were analyzed by flow cytometry to verify infectivity. In this assay, all proviruses result from a single cycle of infection, as neither producer cells nor target cells can be re-infected, due to a lack of the appropriate receptor and co-receptor or to a lack of envelope expression, respectively (Figure 2-1).

Genomic DNA extraction and quantification of plasmid carryover. Genomic DNA was extracted from all collected cells using the High Pure PCR Template Preparation Kit (Roche; Basel, Switzerland) following the manufacturer's instructions and eluted in 150 µL buffer. Genomic DNA was treated with DpnI for 1 h at 37 °C to further reduce plasmid carryover from the

transfections, after which DpnI was heat-inactivated at 80 °C for 20 min. In order to quantify any residual plasmid carryover, two approaches based on quantitative PCR (qPCR) were adopted: 1. The ampicillin resistance gene copy number was determined and divided by the proviral copy number (as measured using the HSA Illumina amplicon), or 2. The proviral copy number from heat-inactivated virus infections was determined and divided by the proviral copy number from the corresponding un-treated infections. For both approaches, qPCR was performed using 4 µL water, 6.25 µL 2X Power SYBR Green Master Mix (Life Technologies), 0.625 µL each primer (500 nM final concentration), and 1 µL template. The cycling conditions used were an initial denaturation of 95 °C 10 m, 40 cycles of 95 °C 15 s/55 °C 15 s/72 °C 30 s, and a final extension of 72 °C 7 min. The sequences of the primers to the ampicillin resistance gene were 5'-ACTTTATCCGCCTCCATCCAGTC-3' and 5'-GAGCGTGACACCACGATGC-3'. Absolute standard curve series (from 10¹-10⁶ copies/µL) were constructed by quantifying the pNL4-3 HIG and pROD HIG plasmids with the Qubit dsDNA HS Assay Kit (Life Technologies). We found that the level of plasmid carryover for HIV-1 was 0.2% when measured by either method, while the level of carryover was 2.8% or 1.4% for HIV-2, depending on the approach used (Table S2-1). These results are comparable to those obtained in a similar study of HIV-1 mutagenesis (222) and are too low to significantly affect measured mutation frequencies.

Amplifications for Illumina sequencing. PCR was performed on five small (~150-170 bp) amplicons (Gag, Vif, HSA, EGFP-1, EGFP-2) for each sample, with HIV-1 & 2 amplicons positioned in homologous locations. All forward primers contained 5-base barcodes (differing by at least two bases) to allow demultiplexing of pooled PCR products. All primer and barcode sequences are listed in Table S2-6. Barcodes were randomly generated using a program written by Luca Comai and Tyson Howell (http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator) (260).

PCR was performed using the Phusion Hot Start II High-Fidelity DNA Polymerase (Fisher Scientific; Pittsburgh, PA). PCR reactions were performed with 8 μ L of genomic DNA (~50,000 target copies), 500 nM of each primer, and a final volume of 50 μ L/reaction. The cycling conditions used were an initial denaturation of 98 °C 30 s, 30 cycles of 98 °C 10 s/56 °C 30 s/72 °C 15 s, and a final extension of 72 °C 10 min. All PCR reactions were performed in triplicate to reduce the risk of clonal amplification, and the products were pooled after amplification. Negative control reactions were performed lacking template or with genomic DNA from uninfected cells.

In order to assess the degree of background error due to PCR and Illumina sequencing, control amplifications were performed from pNL4-3 HIG or pROD HIG plasmids using the same cycling conditions. For these reactions, 50,000 copies of plasmid per 50 μ L reaction was used, a target level found via qPCR to be similar to that of the other samples. Genomic DNA (8 μ L/reaction) from uninfected cells was added to the plasmid PCR reactions to account for any potential impact of genomic DNA on amplification. Further, the degree of intra-sample PCR-mediated recombination was investigated in the plasmid controls, as recombination could affect the association between mutations and thus alter hypermutant frequencies. To determine recombination frequencies, EGFP-1 was amplified from a 1:1 mixture (i.e. 25,000 copies each) of wild-type and mutant EGFP-1 sequences. The mutant EGFP-1 sequence contained two genetic markers positioned ~50 bp apart, and each marker consisted of two mutations to facilitate distinction from PCR and sequencing-induced errors. Intra-sample recombinants were detected by identifying read pairs from the plasmid controls with a single marker set, rather than the expected zero or two marker sets present in the wild-type and mutant EGFP-1 sequences, respectively. Importantly, all mutations utilized as recombination markers were masked before mutational analysis.

Library preparation and Illumina sequencing. Amplicons were gel-purified using the Promega SV Gel Extraction Kit (Promega Corp.; Madison, WI). After

gel purification, all amplicons were quantified using the Qubit dsDNA HS Assay Kit (Life Technologies) and a Qubit 2.0 fluorometer. For each sample, all amplicons were pooled together in an equimolar fashion to normalize coverage between amplicons. Next, 100 ng of each sample (12 samples in total) were submitted to the University of Minnesota Genomics Center for library preparation. The libraries were constructed with the TruSeq Nano DNA Sample Preparation Kit following the manufacturer's instructions but using AMPure XP beads (Beckman Coulter, Inc.; Indianapolis, IN) at a bead to sample ratio of 1.8. The 15 libraries were again quantified using the Qubit dsDNA BR Assay Kit, size-confirmed with Agilent DNA 1000 chips (Agilent Technologies; Santa Clara, CA), and pooled in an equimolar fashion to normalize coverage between libraries. Samples were pooled after library preparation, rather than before, because it had been previously determined that significant inter-sample recombination can occur during the 10 cycles of PCR typically utilized in library construction (data not shown). In order to improve sequence diversity and quality, a PhiX library was added in at ~25% of total mass. Next, 2 μ L of the 10 nM library pool was denatured and diluted to a final concentration of 4 pM for DNA sequencing. Sequencing of proviral DNA was conducted by using the Illumina MiSeq with 2 \times 150 paired-end sequencing.

Processing of proviral DNA sequencing data. First, paired-end reads from the Illumina MiSeq reaction were demultiplexed based on perfect barcode matches, and barcode sequences were trimmed off during the process. Second, poor quality reads were filtered out using quality criteria found to reduce Illumina background error rates (236). Specifically, read pairs were discarded in which either read contained a B-tail (i.e. one or more low quality bases at the end of the read), contained at least one uncalled base, had less than two-thirds of bases with Q-score \geq 30 in the first half of the read, or had an average Q-score $<$ 30 in the first 30% of the read. Third, reads from each of the 12 samples were mapped to the appropriate reference sequence (pNL4-3 HIG or pROD HIG) with GSNAP (261) using default parameters. Finally, a small number of read pairs

(~35,000) that were aligned either partially or fully outside of the appropriate amplicon regions were excluded. For each sample, the numbers of initial read pairs, read pairs lost during mapping or filtering, and final read pairs are listed in Table S2-2. We obtained ~4.7 million total read pairs after all processing steps, which removed primarily PhiX read pairs or HIV read pairs with imperfect barcodes or poor quality. About 319,000-461,000 read pairs were obtained per sample (average of 395,000 read pairs/sample), or 46,000-111,000 read pairs per amplicon per sample (average of 79,000 read pairs/amplicon/sample).

In order to identify mutations present in read pairs passing the above processing steps, a custom algorithm was developed to compile mutation frequency data for each sample, built using the Genome Analysis Toolkit (GATK) walker framework (262). This algorithm determines both the frequency of total mutations as well as of specific mutational types (substitutions, transitions, transversions, every type of transition or transversion, insertions, deletions, etc.). In order to minimize background error rates, mutations were required to be identified on both sequences in a read pair, which was possible because forward and reverse reads were mostly overlapping due to small amplicon sizes (~150-170 bp). Furthermore, substitutions and insertions were only counted if they had a Q-score ≥ 30 for the relevant base(s) on both reads. Wild-type bases had to meet the same criteria as mutations (i.e. identified as wild-type and Q-score ≥ 30 on both sequences of a read pair). Non-overlapping segments of read pairs as well as single reads were excluded from mutational analyses. All primer sequences were also excluded from mutational analysis, as errors within primers would not represent biologically meaningful mutations. We also examined the distribution of all background errors (from PCR and sequencing) in the plasmid controls and identified numerous plasmid error hotspots (defined as upper outliers within the distribution of frequencies of individual mutations based on the $1.5 \times \text{IQR}$ rule). Most plasmid error hotspots (~83%) were G-to-T or C-to-A transversions. Within identical amplicons (HSA/EGFP-1/EGFP-2), many common plasmid error hotspots were identified in the HIV-1 and HIV-2 plasmid

controls (~88% overlap), whereas the degree of overlap was much lower for the ~60% homologous viral amplicons (~49% overlap in Gag and ~31% overlap in Vif). Plasmid error hotspots that were shared between the HIV-1 and HIV-2 plasmid controls (Table S2-5) were masked before further downstream analysis, as the presence of these mutations in the biological samples would most likely represent background errors. Rather than masking all mutational types at error hotspots, only the problematic type(s) (e.g. G-to-T) were masked at the indicated positions. Insertions and deletions were scored as single events regardless of the number of bases inserted or deleted. Mutation frequencies (defined as mutations/bp) were calculated by dividing the number of mutations passing filters by all reference bases (mutations + wild-type bases) passing filters.

In addition to examining mutation frequencies and spectra, hypermutants were identified and characterized within the Illumina sequencing data. Using a custom GATK walker, hypermutant counts for each type of transition were collected (G-to-A, A-to-G, T-to-C, and C-to-T). Hypermutants were defined as individual read pairs containing two or more of the same type of transition. All transitions had to be identified on both sequences with Q-scores ≥ 30 , and a single read pair could theoretically count as two different types of hypermutants if it contained multiple instances of two different transition types. Hypermutant frequencies were calculated by dividing the number of hypermutant read pairs by all (hypermutant + non-hypermutant) read pairs that passed the processing steps. In order to examine G-to-A hypermutation hotspots, ranked lists of G-to-A mutation frequencies were generated at individual bases within G-to-A hypermutants. Hypermutation hotspots were then defined as statistical upper outliers within the distribution of frequencies using the $1.5 \times \text{IQR}$ rule.

Biostatistical analysis of Illumina DNA sequencing data. To test for factors that may influence mutation frequencies, generalized linear mixed effects models were applied to the data that came from our Illumina data processing pipeline. The raw counts for each type of mutation (e.g. transitions) were modeled as Poisson random variables with an offset given by the total number of reference

bases. The type of sample (i.e. HIV-1, HIV-2, and plasmid controls for HIV-1 and HIV-2), the type of amplicon, and their interactions were treated as fixed effects while the replicate was treated as a random effect. The logarithmic link was used, as is standard for Poisson outcomes, and penalized quasilielihood was used to estimate the model parameters (263). These computations were conducted using R v 3.1.0 and the MASS package. All figures and tables were created in GraphPad Prism v 6.0 (GraphPad Software, Inc.; La Jolla, CA) or Microsoft Office for Mac 2011 v 14.3.8 (Microsoft Corp.; Redmond, WA).

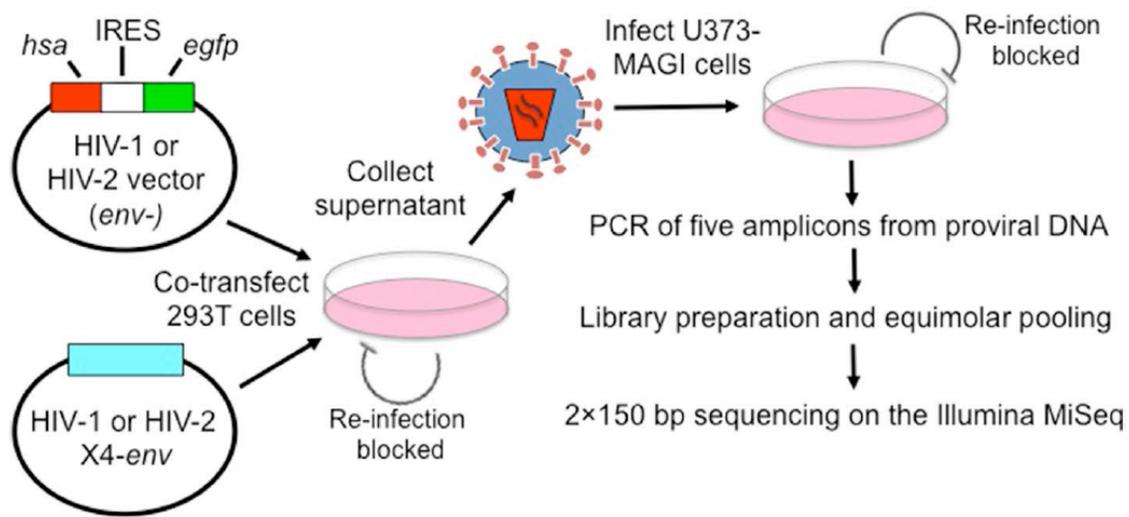


Figure 2-1. Experimental strategy for investigating HIV-1 and HIV-2 mutagenesis by Illumina DNA sequencing. Vector virus stocks were produced by co-transfecting 293T cells with HIV-1 or HIV-2 Env-deficient vectors and HIV-1 or HIV-2 CXCR4-tropic Env expression constructs. Virus stocks were concentrated, DNase I-treated to reduce plasmid carryover, and titered in U373-MAGI cells. To prepare samples for Illumina sequencing, 1×10^6 U373-MAGI cells were infected at an MOI of 1.0, generating approximately 1×10^6 proviruses per experimental replicate. This assay represents a single round of viral replication, as producer cells and target cells cannot be re-infected, due to a lack of receptor or Env expression, respectively. Polymerase chain reaction (PCR) of five amplicons (Gag, Vif, HSA, EGFP-1, and EGFP-2) was performed from the proviral DNA. Amplicons from the HIV-1 and HIV-2 proviral DNAs were either identical (HSA, EGFP-1 & 2) or homologous (Gag and Vif) in sequence. The EGFP-1 and EGFP-2 amplicons represent non-overlapping segments of the *egfp* gene. Sequencing libraries were prepared from the amplicons, pooled in an equimolar fashion to normalize coverage, and subjected to 2x150 bp sequencing on the Illumina MiSeq, generating approximately 4.7 million read pairs after data processing.

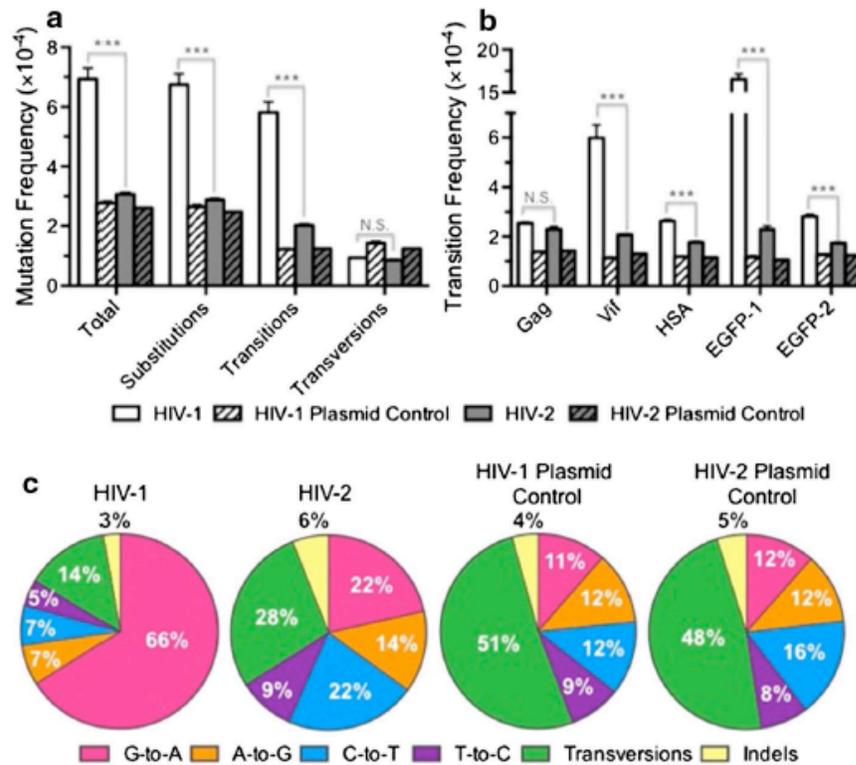


Figure 2-2. HIV-2 has a lower mutation frequency and distinct mutation spectrum relative to HIV-1. A. Mutation frequency analysis. Mutation frequencies were calculated by dividing the number of mutations by the number of reference bases (mutations + wild-type bases) and are expressed as mutations/bp, or m/bp. Mutation frequencies were determined for HIV-1 and HIV-2, as well as for plasmid controls to assess background error levels. B. Transition frequency analysis. Transition frequencies were compared across the five different amplicons subjected to Illumina DNA sequencing. C. Mutation spectra analysis. Mutation spectra were determined by dividing the frequency of each type of mutation by the total mutation frequency, with the results expressed as a percentage of total mutations. Data in all panels represent the mean of three experimental replicates, with error bars indicating the standard deviation. Asterisks denote statistically significant differences between HIV-1 & 2 (* $p < 0.05$, *** $p < 0.001$, N.S. = not significant). The actual numbers of read pairs, mutations, and reference bases are listed in Dataset S2-1.

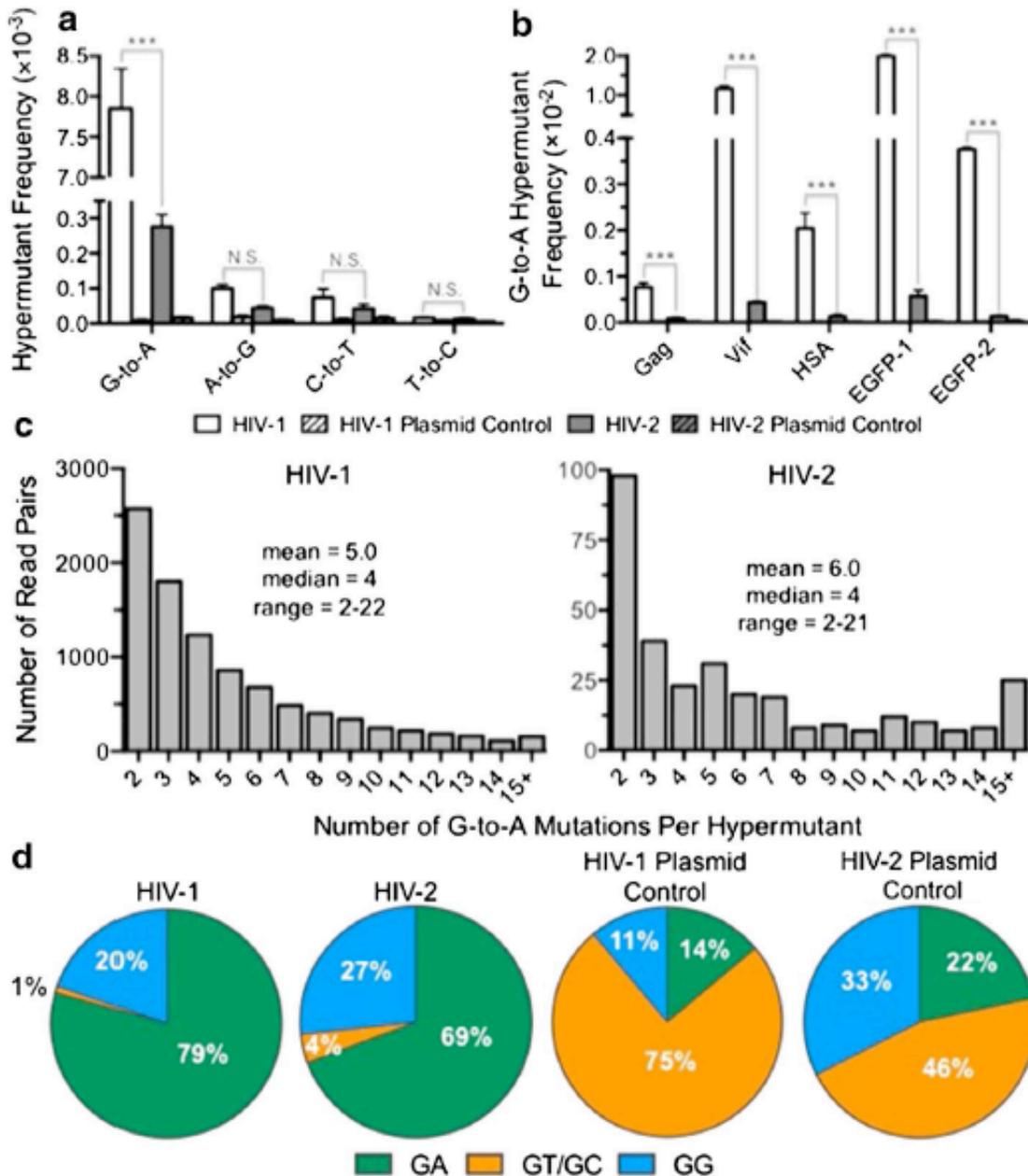


Figure 2-3. HIV-1 demonstrates higher G-to-A hypermutant frequencies than HIV-2. A. The frequencies of each type of transition hypermutant were compared between HIV-1, HIV-2, and the plasmid controls. For this analysis, hypermutants were defined as read pairs containing two or more mutations of the indicated type within an individual read pair (approximately 120 bp in length).

The frequency of hypermutation was then calculated by dividing the number of hypermutant read pairs by all read pairs. B) The frequency of G-to-A hypermutation was compared across all five amplicons examined by Illumina DNA sequencing. C) The degree of G-to-A hypermutation was analyzed by determining the numbers of G-to-A mutations within hypermutant read pairs. D) The dinucleotide context of G-to-A mutations from G-to-A hypermutants was determined and expressed as a percentage of the total. Data in panels A, B, and D represent the mean of three experimental replicates, with error bars representing standard deviation, while data in panel C represent the total (i.e. compiled) data. Asterisks denote statistically significant differences between HIV-1 & 2 (* $p < 0.05$, *** $p < 0.001$, N.S. = not significant). The actual numbers of G-to-A hypermutant read pairs and reference read pairs are listed in Dataset S2-2.

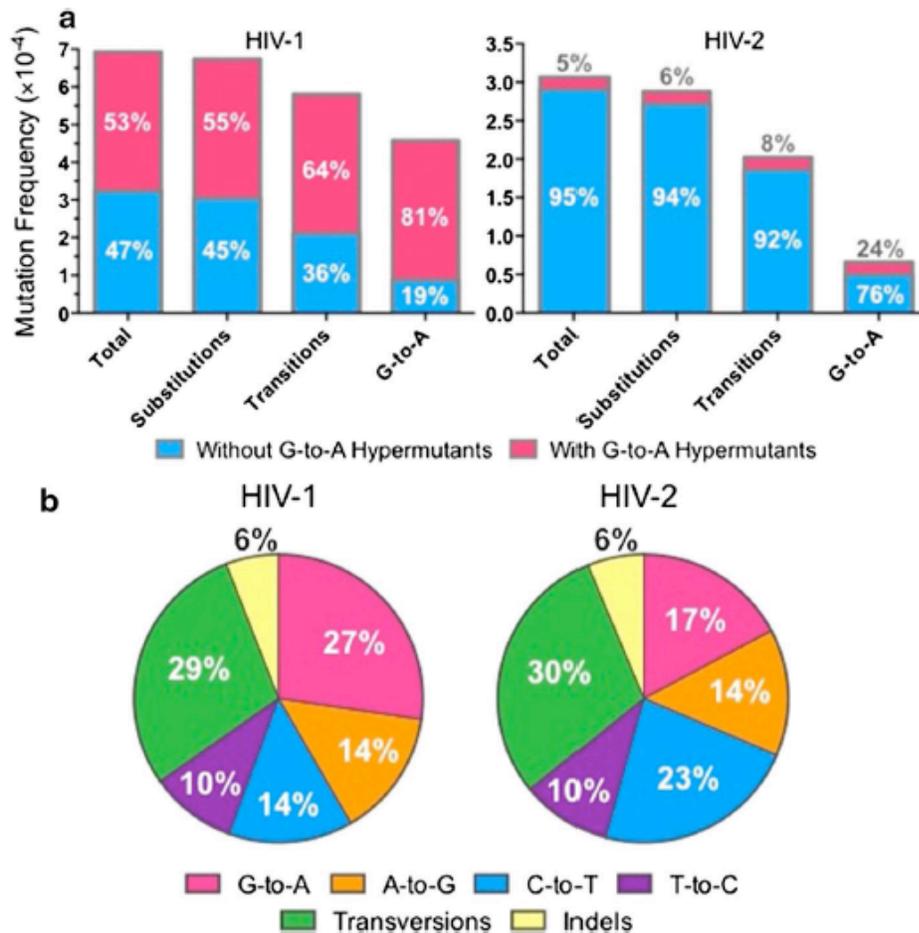


Figure 2-4. HIV-1 and HIV-2 mutation frequencies and spectra are similar in the absence of G-to-A hypermutation. A. Analysis of mutation frequency in the absence of G-to-A hypermutation. HIV-1 and HIV-2 mutation frequencies were determined either including or excluding G-to-A hypermutants, with the results superimposed. The relative percentage of the total data that can be attributed (or not attributed) to G-to-A hypermutation is indicated within the bars. B. Analysis of HIV-1 and HIV-2 mutation spectra in the absence of G-to-A hypermutation. HIV-1 and HIV-2 mutation spectra were examined after excluding all G-to-A hypermutants. Mutation spectra were determined by dividing the frequency of each type of mutation by the total mutation frequency, with the results expressed as a percentage. Data in both panels represent the mean of three experimental replicates.

CHAPTER III
LACK OF MUTATIONAL HOTSPOTS DURING DECITABINE-MEDIATED
HIV-1 MUTAGENESIS

Reprinted with permission from: Rawson JM, Landman SR, Reilly CS, Bonnac L, Patterson SE, Mansky LM. 5 Lack of mutational hotspots during decitabine-mediated HIV-1 mutagenesis. *Antimicrob Agents Ch* (2015) 59/11, 6834-43. Copyright © 2015, American Society for Microbiology.

J.M.R. and S.R.L. contributed equally to this work.

Introduction

RNA viruses often mutate at high rates, ranging from 10^{-4} to 10^{-6} mutations/bp/cycle (61), enabling rapid adaptation to constantly changing environments. However, rapid adaptability comes at a price—relatively small increases in the mutation rate can drive the viral population into error catastrophe, during which genetic information cannot be stably maintained and the viral population collapses (41, 176, 177). As a result, RNA viruses have been hypothesized to be susceptible to compounds that would elevate the viral mutation rate, a therapeutic strategy referred to as lethal mutagenesis (119). Lethal mutagenesis has now been demonstrated in cell culture with a wide variety of RNA viruses and mutagenic agents (179, 264). While experimental results in cell culture have been promising, the clinical translation of lethal mutagenesis has been a slow process, due largely to a lack of strong lead compounds. Notably, ribavirin may inhibit hepatitis C virus *in vivo* at least in part through a lethal mutagenesis mechanism, but this topic remains controversial (184-187). The nucleoside analog KP-1212 has been shown to inhibit human immunodeficiency virus type 1 (HIV-1) replication by lethal mutagenesis in cell culture (198). However, in phase II clinical trials, KP-1461 (a prodrug of KP-1212) did not reduce viral loads and demonstrated only slight evidence of mutagenic effects, despite sufficient bioavailability (199, 200). The nucleobase analog favipiravir (T-705) exerts broad-spectrum anti-viral activity in cell culture and is currently being investigated in phase III clinical trials for influenza in the United States and in phase II clinical trials for Ebola virus in West Africa (188, 265). T-705 has been shown to elicit antiviral mutagenesis against influenza in cell culture and against norovirus in a mouse model of infection (189, 190). However, T-705 has also been shown to directly inhibit viral RNA polymerases (191-193), and it therefore remains unclear whether T-705 acts *in vivo* primarily as a lethal mutagen or as a direct inhibitor of viral replication.

Most lethal mutagens investigated thus far have been nucleobase or nucleoside analogs that base-pair promiscuously, due to properties such as

conformational flexibility, tautomerization, structural rearrangement, or ionization (264). For HIV-1 lethal mutagenesis, several active compounds have been identified, including decitabine (5-aza-2'-deoxycytidine), 5-azacytidine (5-AZC, the ribonucleoside equivalent of decitabine), 5,6-dihydro-5-aza-2'-deoxycytidine (KP-1212), and 5-hydroxy-2'-deoxycytidine (117-119, 198). Notably, HIV-1 lethal mutagenesis based on mutagenic nucleosides is similar to APOBEC3-mediated hypermutation that occurs naturally to restrict viral replication in the absence of Vif (122, 123). Thus, targeting the Vif-APOBEC3 interaction (or other necessary interactions for APOBEC3 degradation, such as between Vif and CBF- β) represents an additional strategy to achieve lethal mutagenesis, and this avenue is also actively being explored (180-183).

Decitabine is a particularly promising candidate for the lethal mutagenesis of HIV-1 because it exhibits potent antiviral activity (EC_{50} of ≈ 200 nM) in the absence of significant cytotoxicity (117). Furthermore, decitabine synergizes with low, non-cytotoxic concentrations of several anti-metabolites, including gemcitabine and resveratrol (117, 201). The combination of decitabine and gemcitabine was also shown to exhibit potent antiretroviral activity in an MLV-based mouse model for AIDS (213). Although decitabine exhibits poor bioavailability when administered orally, a prodrug form called decitabine divalerate has recently been synthesized that demonstrates improved stability and permeability while retaining anti-HIV-1 activity (266). Previous studies found that decitabine increased the mutant frequency of HIV-1 by inducing predominantly G-to-C transversions, which rarely occurred in the absence of drug (117). However, the factors that determine the susceptibilities of individual sequence positions in HIV-1 to decitabine have not yet been analyzed in great depth. To examine this, we performed Illumina high-throughput sequencing of five amplicons prepared from proviral DNA that was recovered from untreated or decitabine-treated cells infected with HIV-1. We found that decitabine induced an ≈ 4.1 -fold increase in the total mutation frequency of HIV-1, largely due to an ≈ 155 -fold increase in the frequency of G-to-C transversions. Intriguingly,

decitabine also resulted in an ≈ 29 -fold increase in the frequency of C-to-G transversions. G-to-C frequencies varied significantly (up to ≈ 80 -fold) between different sequence positions, but, unexpectedly, mutational hotspots (defined as upper outliers within the distribution of mutation frequencies) were not observed. We also found that every single guanine position examined (134/134) and nearly every cytosine position examined (163/169) were significantly susceptible to the mutagenic effects of decitabine. Taken together, these observations show for the first time that decitabine-mediated lethal mutagenesis of HIV-1 is promiscuous and occurs without a clear bias for mutational hotspots. These data imply that decitabine-mediated mutagenesis is a highly effective approach for extinguishing the infectivity of HIV-1 and support further evaluation of decitabine in animal models of HIV-1 infection.

Results

Illumina sequencing of HIV-1 proviral DNA produced in the presence or absence of decitabine. In order to investigate the impact of decitabine on HIV-1 mutagenesis, we prepared samples for Illumina amplicon sequencing as described previously (see Materials and Methods and (267)). Briefly, we produced viral stocks by co-transfecting an envelope-deficient HIV-1 vector with an HIV-1 X4-tropic Env expression plasmid and then infected 1 million U373-MAGI-X4 cells at an MOI of 1.0. In this assay, only a single cycle of viral replication can occur, as neither producer cells nor target cells can be re-infected. U373-MAGI-X4 cells were pre-treated (or not pre-treated) with 2 μ M (an $\approx EC_{65}$ concentration) of decitabine divalate, a prodrug form of decitabine, for 2 h before infection. The media was replaced 24 h post-infection, and cells were collected 72 h post-infection. Genomic DNA was purified from infected cells and used to prepare five amplicons by PCR, including viral (Gag, Vif) and marker (HSA, EGFP-1, EGFP-2) gene targets. The EGFP-1 and EGFP-2 amplicons represent non-overlapping portions of the *egfp* gene. Plasmid amplifications were included to measure the level of background error due to

PCR and Illumina sequencing. Amplicons from the same sample were pooled and used to prepare sequencing libraries, which were then subjected to 2×150 paired-end sequencing on the Illumina MiSeq. In total, nine samples were analyzed, including three experimental replicates each of HIV-1, HIV-1 exposed to decitabine, and HIV-1 plasmid controls. For the decitabine samples, Illumina sequencing resulted in ≈425,000 read pairs/sample, representing ≈125,000 mutations and ≈45 million reference bases (wild-type bases + mutations) per sample after all bioinformatics processing steps (Table S3-2). The Illumina sequencing data for the controls (HIV-1 without drug, HIV-1 plasmid control) has previously been described elsewhere but was re-analyzed for the purposes of this study (267). Using this data, we determined mutation frequencies (expressed as mutations per base pair, or m/bp) for all samples, both for total mutations and every possible mutational subclass (e.g. substitutions, transitions, transversions, etc.). The data were analyzed both combined across all five amplicons and separated by amplicon.

Decitabine induces primarily G-to-C transversion mutations during HIV-1 replication. We found that decitabine increased the total mutation frequency of HIV-1 from $\approx 6.9 \times 10^{-4}$ m/bp to $\approx 2.9 \times 10^{-3}$ m/bp (Figure 3-1A), a difference of ≈4.1-fold ($p < 0.001$). Decitabine increased the transversion frequency of HIV-1 by ≈23-fold ($p < 0.001$), while transition frequencies were not significantly affected by the presence of the compound ($p = 0.23$). Upon dividing out the eight possible types of transversions, we found that decitabine increased the frequency of G-to-C transversions from 1.1×10^{-5} m/bp to 1.7×10^{-3} m/bp, a striking difference of ≈155-fold ($p < 0.001$). As a result, the HIV-1 mutational spectrum shifted significantly in the presence of the drug, with an increase in the relative percentage of G-to-C transversions from 2% to 57% (Figure 3-1B). Intriguingly, we also observed an ≈29-fold increase ($p = 0.003$) in the frequency of C-to-G transversions in the presence of decitabine (Figure 3-1A), which was not observed in a previous report analysing the activity of decitabine against HIV-1 (117) (see also Discussion). The relative percentage of C-to-G transversions

increased from 2% in the absence of decitabine to 10% in the presence of decitabine (Figure 3-1B). Decitabine-mediated G-to-C transversions have previously been hypothesized to result from incorporation into the minus strand viral DNA, followed by structural rearrangement and mispairing during plus strand synthesis (117), while C-to-G transversions could be caused by plus strand incorporation of decitabine or minus strand incorporation of decitabine decomposition products (see Discussion). The preponderance of G-to-C transversions over C-to-G transversions is consistent with the hypothesis that most of the decitabine incorporated into the plus strand viral DNA is removed by host repair processes (see Discussion).

Most decitabine-induced G-to-C and C-to-G mutant read pairs have relatively low mutational loads. Decitabine could potentially lead to high levels of G-to-C and C-to-G transversions by inducing high numbers of mutations in relatively few read pairs or by inducing low numbers of mutations in many read pairs. To address this, we determined the numbers of G-to-C or C-to-G mutations per mutant read pair (Figure 3-2). We found that 12% of all read pairs ($\approx 149,000/1,272,000$) contained at least one G-to-C transversion in the presence of decitabine (Figure 3-2A), whereas 3% ($\approx 33,000/1,272,000$) contained at least one C-to-G transversion (Figure 3-2B). As expected, the total numbers of G-to-C and C-to-G mutants were much higher in the presence of decitabine than in its absence. Most of the decitabine-induced mutant read pairs contained a single G-to-C or C-to-G transversion, with single mutants representing $\approx 71\%$ of all G-to-C mutants and $\approx 83\%$ of all C-to-G mutants. However, decitabine treatment also led to a significant number of read pairs with multiple G-to-C or C-to-G transversion mutations. While most of these multiply mutated sequences contained two or three mutations, we identified rare sequences that were excessively mutated, with as many as 13 G-to-C or 9 C-to-G transversions in a single read pair. Decitabine-induced mutants with multiple G-to-C or C-to-G transversions accounted for $\approx 53\%$ of all G-to-C transversions and $\approx 33\%$ of all C-to-G transversions, indicating their important contribution to decitabine-mediated

HIV-1 mutagenesis. Notably, given the small size of Illumina read pairs (≈ 120 bp after processing), the full-length proviral genomes (≈ 11.2 kb including marker genes) likely contained much higher levels of G-to-C and C-to-G transversions in the presence of decitabine.

All amplicons are susceptible to decitabine-mediated HIV-1 mutagenesis.

In order to further examine the distribution of decitabine-induced mutations, we analysed the mutational data separately for each of the five amplicons (Gag, Vif, HSA, EGFP-1, and EGFP-2). We hypothesized that differences in amplicon primary sequences, nucleic acid folding, or positioning within the viral genome could potentially lead to varying susceptibilities to the drug. We found that all five amplicons were highly sensitive to decitabine-induced G-to-C transversions ($p < 0.001$), with increases ranging from 77-fold (Gag) to 248-fold (EGFP-2) (Figure 3-3A). The EGFP-1 and EGFP-2 amplicons were most sensitive to G-to-C transversions, both in terms of absolute G-to-C frequencies and in terms of fold-induction. The absolute G-to-C frequencies were 2-3-fold higher for these amplicons than for the Gag, Vif, and HSA amplicons ($p < 0.001$). Likewise, all five amplicons were significantly susceptible to decitabine-mediated C-to-G transversions ($p < 0.001$), with increases ranging from 15-fold (EGFP-2) to 45-fold (Gag) (Figure 3-3A). The absolute frequencies of C-to-G transversions varied between amplicons up to ≈ 2.1 -fold (Gag vs. EGFP-2). Small differences in the susceptibilities of the five amplicons to G-to-C or C-to-G transversions could potentially be due to varying nucleotide content of the amplicon sequences. We found that guanine content varied from 20% to 31% while cytosine content varied from 15% to 43% between amplicons. To account for these differences, we divided G-to-C and C-to-G transversion frequencies by the guanine or cytosine content of each amplicon, respectively, and normalized the resulting data to the amplicon with the highest transversion frequency (EGFP-1 for G-to-C, Vif for C-to-G) (Figure 3-3B). After normalization, the general trends remained unchanged: EGFP-1 and EGFP-2 demonstrated ≈ 2.0 - 2.5 -fold higher G-to-C

frequencies than Gag, Vif, and HSA, while conversely EGFP-1 and EGFP-2 demonstrated the lowest C-to-G frequencies.

Decitabine-induced mutations in HIV-1 are not caused by DNA methyltransferase-mediated hydrolysis. Decitabine-induced mutations in genomic DNA have been previously found by one group to be heavily biased toward CpG dinucleotides, suggesting that DNA methyltransferases (DNMTs) were involved in their mechanism of formation (196). DNMTs were proposed to attack the C6 position of decitabine, leading to ring opening and formation of a pre-mutagenic lesion capable of base-pairing with cytosine. However, decitabine is also known to be highly unstable in aqueous solution, undergoing rapid spontaneous hydrolysis and deformylation into a variety of ring-opened derivatives (197, 268, 269). To determine the mechanism of hydrolysis most relevant to the antiviral activity of decitabine, we analysed the dinucleotide context of all G-to-C and C-to-G transversions that occurred in the presence of the drug. We found that only 19% of G-to-C transversions occurred within CpG dinucleotides (as opposed to TpG, ApG, and GpG dinucleotides, sites of mutations underlined) (Figure 3-4). In particular, after taking into account the varying frequencies of the dinucleotides within the amplicon sequences, there was no significant bias of G-to-C transversions toward any particular dinucleotide (Figure S3-1). Likewise, only 11% of C-to-G transversions occurred within the CpG dinucleotide context (as opposed to CpA, CpC, and CpT dinucleotides, sites of mutation underlined) (Figure 3-4). Even after taking into account the lower frequency of CpG dinucleotides within the amplicon sequences, only 18% of C-to-G transversions occurred within the CpG context, although a possible bias toward CpT dinucleotides was observed (Figure S3-1). Taken together, these data support the hypothesis that most decitabine-mediated transversions in HIV-1 are due to spontaneous hydrolysis, deformylation, and mispairing rather than DNMT-mediated hydrolysis.

Lack of decitabine-induced G-to-C transversion hotspots. In order to investigate whether certain sequence positions were more susceptible to

decitabine-mediated mutagenesis than others, the G-to-C frequency was determined at every individual guanine position (134 in total) and the C-to-G frequency at every individual cytosine position (169 in total). G-to-C frequencies were highly variable, ranging from $\approx 2.0 \times 10^{-4}$ m/bp to $\approx 1.7 \times 10^{-2}$ m/bp (median of 7.1×10^{-3} m/bp), a maximal difference of ≈ 83 -fold (Figure 3-5A). However, we failed to identify G-to-C mutational hotspots, which we defined as upper outliers within the mutation frequency distribution using the $1.5 \times$ interquartile range (IQR) rule. For C-to-G transversions, we also identified a wide range of frequencies in the presence of decitabine, varying from 0 to 3.5×10^{-3} m/bp (median of 9.0×10^{-4} m/bp) (Figure 3-5B). Using the $1.5 \times$ IQR rule, we identified four decitabine hotspots for C-to-G transversions, all located within the Gag and Vif amplicons. We next examined the locations and coding effects of the most frequently mutated guanines and cytosines in order to determine whether they occurred in specific amplicons and to gauge the effects of the mutations on gene function. We found that the twenty most frequent decitabine-induced G-to-C transversions all occurred within the EGFP-1 and EGFP-2 amplicons (Table 3-1), consistent with the higher G-to-C frequencies observed earlier in these amplicons (Figure 3-3). Most (18/20) of these G-to-C transversions would lead to non-synonymous missense mutations, and notably G-to-C mutations cannot introduce premature stop codons. However, decitabine induced a wide array of semi-conservative and non-conservative amino acid changes that could potentially disrupt protein function. In contrast to G-to-C transversions, the most common C-to-G transversions were located in the Gag and Vif amplicons (Table 3-2), in agreement with the higher C-to-G frequencies observed in these amplicons (Figure 3-3). Although C-to-G mutations could potentially introduce stop codons, the most frequent C-to-G transversions did not result in stop codons but instead introduced missense mutations. Overall, these findings demonstrate that 1) decitabine-induced G-to-C transversion mutation frequencies vary widely

among sequence positions but without clear mutational hotspots and that 2) decitabine triggers HIV-1 lethal mutagenesis mainly through missense mutations.

Virtually all guanine and cytosine positions are susceptible to decitabine-mediated mutagenesis. Intriguingly, we observed that the minimal G-to-C and C-to-G frequencies in the presence of decitabine were similar to the highest G-to-C and C-to-G frequencies in the absence of drug (Figure 3-5), suggesting that most guanines and cytosines are susceptible to decitabine-mediated mutagenesis. In order to more thoroughly examine this possibility, we determined the fold difference (decitabine/no drug) in G-to-C or C-to-G transversion frequency at every individual guanine or cytosine position, respectively. Fold differences could not be calculated for a handful of positions (1 of 134 for G-to-C and 6 of 169 for C-to-G) because they were not mutated at all in the absence of decitabine. We found that G-to-C transversion frequencies increased by ≈ 15 to ≈ 1010 -fold (median of ≈ 178 -fold) in the presence of decitabine (Figure 3-6). Likewise, C-to-G transversion frequencies increased by ≈ 3 to ≈ 152 -fold (median of ≈ 27 -fold) in the presence of decitabine. In order to determine whether these increases were statistically significant, we compared G-to-C and C-to-G frequencies in the presence or absence of decitabine at every possible sequence position using the Mantel-Haenszel test (see Materials and Methods). We found that 100% of the guanine positions (134/134) were significantly susceptible to decitabine-mediated G-to-C transversions ($p < 0.001$). Similarly, 96% of the cytosine positions (163/169) were significantly susceptible to decitabine-induced C-to-G transversions ($p < 0.001$). Overall, these data suggest that virtually all guanines and cytosines within the HIV-1 genome are susceptible to the mutagenic effects of decitabine, although full genome sequencing will be required to address this in more detail (see Discussion).

Discussion

In this study, we have shown that decitabine is strongly mutagenic of HIV-1, inducing ≈ 4.1 , 23, and 155-fold increases in total, transversion, and G-to-C

mutation frequencies, respectively (Figure 3-1). We also observed an ≈ 29 -fold increase in C-to-G transversions in the presence of decitabine, which was not found in a previous report on the mutagenic and antiviral effects of decitabine (117). These discrepancies are likely due to differences in the sequencing methodologies applied. Previously, cells harbouring mutant proviruses (HSA+ EGFP-) were isolated, and the mutated marker gene (*egfp*) was amplified and analysed by Sanger sequencing. In total, 82 mutations were identified, 31 (i.e. 38%) of which were G-to-C transversions. Here, we found that the frequency of C-to-G transversions in the presence of decitabine was $\approx 3.0 \times 10^{-4}$ m/bp, while the G-to-C transversion frequency was $\approx 1.7 \times 10^{-3}$ m/bp, a 5.5-fold difference (Figure 3-1). However, we also found that G-to-C mutations were over-represented and C-to-G mutations were under-represented in the EGFP-1 and EGFP-2 amplicons compared to the other amplicons we examined (Figure 3-3). In these amplicons, the G-to-C transversion frequency was ≈ 12 -fold higher than the C-to-G transversion frequency. Thus, on the basis of our results, we would expect only two to three C-to-G transversions to have been identified in the earlier study involving Sanger sequencing, a level that would likely be too low to result in a statistically significant increase in C-to-G transversions. Additionally, it is possible that decitabine and decitabine divalerate (the prodrug form used here) may exhibit small differences in mutagenicity due to the improved permeability and stability of decitabine divalerate (266).

The observation that decitabine induces primarily G-to-C transversions in HIV-1 is consistent with a previously proposed model (117) for decitabine: 1. Decitabine is incorporated in place of deoxycytidine during minus strand synthesis by the HIV-1 reverse transcriptase (RT), 2. Decitabine undergoes structural rearrangement (hydrolysis and deformylation), resulting in a pre-mutagenic lesion, and 3. The ring-opened derivative of decitabine mispairs with deoxycytidine during plus strand synthesis, leading to fixation of G-to-C transversions. In contrast, while decitabine is presumably also incorporated during plus strand synthesis by the HIV-1 RT, the ring-opened derivatives of

decitabine can be correctly removed and replaced by host repair process after import of viral DNA into the nucleus, preventing fixation of C-to-G transversions. However, we have demonstrated here that decitabine increased the C-to-G transversion frequency of HIV-1 by ≈ 29 -fold, clearly demonstrating that this process is more complex. While the mechanism responsible for decitabine-induced C-to-G transversions is presently unknown, one possible explanation is that they result from the incorporation of decitabine during plus strand synthesis, followed by structural rearrangement and inefficient repair by host processes. In the absence of repair by host processes, ring-opened derivatives of decitabine could lead to the fixation of C-to-G transversions upon replication of the eukaryotic genomic DNA during cell division or during PCR. In some cases, decitabine may be incorporated but not immediately undergo structural rearrangement, potentially evading host repair processes. However, the *in vitro* half-life of decitabine is quite short (≈ 11 h) under physiological conditions (197), and cells were not collected for genomic DNA extraction until 72 h post-infection. Nonetheless, the stability of decitabine may significantly vary between double-stranded DNA, single-stranded DNA, and unincorporated forms. In the future, it would be of great interest to better define the host repair processes and enzymes that act on decitabine and its derivatives, as such processes could greatly influence the mutagenicity of decitabine against HIV-1 proviral DNA and eukaryotic genomic DNA. Notably, 2'-deoxyriboguanylurea (dRGU), the primary decomposition product of decitabine, was recently shown to be mutagenic in human cells and found to base pair with both deoxycytidine and deoxyguanosine in Taq-based PCR reactions (270). Thus, decitabine-mediated transversions (G-to-C and C-to-G) in HIV-1 could also result from direct incorporation of decomposition products by RT during minus strand synthesis. For example, C-to-G transversions could result from the incorporation of dRGU in place of deoxyguanosine during minus strand synthesis, followed by base pairing with deoxyguanosine during plus strand synthesis. However, the ability of dRGU to

directly induce the mutagenesis of HIV-1 has not yet been examined in cell culture or in assays with purified RT.

Decitabine is primarily used clinically for the management of myelodysplastic syndromes as an inhibitor of DNA methyltransferases (DNMTs), by which decitabine leads to the reactivation of aberrantly silenced genes (271). DNMTs can attack decitabine at the C6 position, resulting in the formation of trapped covalent complexes and depletion of functional DNMTs in the cell. In a previous study, decitabine was found to induce C-to-G and G-to-C transversions in eukaryotic genomic DNA mostly within CpG dinucleotides, suggesting that the mutations resulted from DNMT attack, followed by hydrolysis and deformylation (196). However, other studies found either that decitabine did not increase the mutation frequency of eukaryotic genomic DNA (272), or that decitabine was mutagenic through spontaneous rather than DNMT-mediated hydrolysis (273). We predicted that most G-to-C transversions would not be caused by DNMT attack, as the preponderance of G-to-C over C-to-G transversions suggests that they mostly result from decitabine incorporation and rearrangement during reverse transcription, which takes place in the cytoplasm. Nonetheless, DNMTs could still be involved in the formation of C-to-G transversions and of some G-to-C transversions. In this study, we found that there was no bias of G-to-C or C-to-G transversions toward CpG dinucleotides, even after correcting for the varying frequencies of dinucleotides in our amplicon sequences (Figures 3-4 & S3-1). This data supports the hypothesis that most decitabine-induced mutations result from spontaneous hydrolysis and deformylation.

Upon examining the distribution of decitabine-mediated mutations, we observed significant variability between amplicons and between individual sequence positions (Figures 3-3 & 3-5). The positions most susceptible to G-to-C transversions were located in the EGFP-1 and EGFP-2 amplicons, whereas the positions most susceptible to C-to-G transversions were located in the Gag and Vif amplicons (Tables 3-1 & 3-2). Notably, many previous studies on HIV-1 mutagenesis have reported wide variability in the distribution of mutations across

the sequences of mutational targets, even in the absence of a mutagen (53, 55, 57). Unfortunately, the factors that create favorable positions for RT-mediated substitutions are still very poorly understood, although homopolymeric runs are known to trigger most insertions and deletions. We hypothesize that certain sequence positions were more susceptible to decitabine than others because: 1. They occurred within primary sequence contexts that favoured decitabine incorporation, 2. They occurred within nucleic acid secondary structures that favoured decitabine incorporation, or 3. They occurred within regions of the HIV-1 genome more susceptible than others to decitabine-mediated mutagenesis. More specifically, the positioning of the amplicon within the viral genome impacts the amount of time the minus strand viral DNA remains single-stranded during reverse transcription, which has previously been proposed to drive differences in APOBEC3-mediated hypermutation across the genome (124, 254). Decitabine that is incorporated into regions of the genome that remain single-stranded longer than others may more efficiently undergo structural rearrangement prior to plus strand synthesis. The differences in G-to-C frequencies we observed between amplicons are in general agreement with this hypothesis (Figure 3-3). However, full genome sequencing will be required to determine the extent to which these factors can explain the variability of decitabine-mediated mutagenesis across amplicons and sequence positions.

In order to more thoroughly investigate the susceptibility of individual sequence positions to decitabine, we determined fold changes in mutation frequencies at individual positions and also evaluated the results statistically. Strikingly, we found that every guanine position examined (134/134) and nearly every cytosine position examined (163/169) were significantly susceptible to the mutagenic effects of decitabine, with increases in G-to-C and C-to-G transversion frequencies ranging from ≈ 15 to 1000-fold and ≈ 3 to 150-fold, respectively (Figure 3-6). These results were obtained by sequencing five different amplicons representing multiple genes, nucleotide contents, and positions within the genome. Thus, these findings support the hypothesis that the vast majority of

guanines and cytosines within the HIV-1 genome are susceptible to decitabine-mediated mutagenesis, although full genome sequencing will be required to examine this possibility in more detail. Overall, these findings demonstrate for the first time that decitabine-mediated mutagenesis of HIV-1 is widely promiscuous and occurs in the clear absence of mutational hotspots. These data imply that decitabine-mediated G-to-C mutagenesis is a highly effective mechanism for extinguishing HIV-1 infectivity and encourage further evaluation of decitabine in animal models for the induction of HIV-1 lethal mutagenesis.

Materials and Methods

Production of viral stocks. The production and titering of HIV-1 viral stocks for single-cycle infections and Illumina sequencing were previously described (267). Briefly, we used the envelope-deficient HIV-1 vector, pNL4-3 HIG (118), which contains a cassette inserted into *nef* that encodes heat stable antigen (HSA), an internal ribosomal entry site (IRES), and enhanced green fluorescent protein (EGFP). Viral stocks were produced by co-transfecting 10 µg of pNL4-3 HIG and 5 µg of pNL4-3 Env per 10 cm plate of 293T cells. For each viral stock, five 10 cm plates were transfected, and the resulting supernatants were pooled and concentrated (≈10-fold) using 100,000 MWCO filtration columns (Vivaproducts; Littleton, MA). Viral stocks were then treated with 10 U/mL of DNase I (New England Biolabs; Ipswich, MA) for 2 h at 37 °C to degrade residual plasmid DNA from transfections. Lastly, viral stocks were titered in U373-MAGI-X4 cells based on EGFP expression at low infectivities as described (267).

Preparation of genomic DNA from infected cells. In order to prepare samples for Illumina sequencing, 1×10^6 U373-MAGI-X4 cells were infected at a multiplicity of infection (MOI) of 1.0 as described (267). Cells were pre-treated (or not pre-treated) with 2 µM decitabine divalate (an ≈EC₆₅ concentration) for 2 h before infection. Uninfected cells and cells infected with heat-inactivated viruses (i.e. virus stocks that were incubated at 65 °C for 1 h) were included as negative controls. The medium was replaced 24 h post-infection, and the cells

were harvested for genomic DNA extraction at 72 h post-infection. Notably, only a single round of replication can occur within this assay, due to the lack of envelope expression in target cells. All infections were performed in triplicate, using independently produced viral stocks. Infections in the presence or absence of decitabine were performed using the same viral stock for each replicate. Genomic DNA was isolated from all collected cells using the High Pure PCR Template Preparation Kit (Roche; Basel, Switzerland) following the manufacturer's instructions and eluted in 150 μ L buffer. Genomic DNA was analyzed by qPCR in order to quantify the amount of plasmid carryover from transfections and found to be \approx 0.2% of the proviral copy number, a level that would not significantly impact measured mutation frequencies.

Proviral DNA amplification and Illumina sequencing. PCR was performed as described previously (267) to generate five small (150-170 bp) amplicons (Gag, Vif, HSA, EGFP-1, EGFP-2) for each sample, representing a mixture of viral and marker gene targets. All forward primers contained barcodes for later demultiplexing of samples. Primer and barcode sequences for the decitabine samples are listed in Table S3-1, while others have been previously reported (267). PCR was performed using the Phusion Hot Start II High-Fidelity DNA Polymerase (Fisher Scientific; Pittsburgh, PA). PCR reactions were performed with 8 μ L of genomic DNA per 50 μ L reaction and cycling conditions of 98 $^{\circ}$ C 30 s, 30 cycles of 98 $^{\circ}$ C 10 s/56 $^{\circ}$ C 30 s/72 $^{\circ}$ C 15 s, and a final extension of 72 $^{\circ}$ C 10 min. All PCR reactions were performed in triplicate, pooled after amplification, and gel-purified. Plasmid control amplifications from pNL4-3 HIG were performed in parallel under closely matched conditions to measure background error due to PCR and Illumina sequencing. For each sample, all amplicons were pooled together in an equimolar fashion to normalize coverage between amplicons. Next, sequencing libraries were constructed from each sample (9 samples in total; 3 each for HIV-1, HIV-1 + decitabine divaleryl, and HIV-1 plasmid control) using the TruSeq Nano DNA Sample Preparation Kit and following the manufacturer's instructions. Libraries were quantified using the

Qubit dsDNA BR Assay Kit, size-confirmed with Agilent DNA 1000 chips (Agilent Technologies; Santa Clara, CA), and pooled in an equimolar fashion to normalize library coverage. Before sequencing, a PhiX library was added in at 25% of total mass to improve sequence diversity. The pooled libraries were then subjected to 2×150 paired-end sequencing on the Illumina MiSeq. All Illumina sequencing data supporting the results of this manuscript have been deposited into the NCBI Sequence Read Archive (SRA) under accession code SRP059680.

Illumina sequencing data processing. Illumina sequencing data were subjected to multiple read-level and base-level quality filtering steps, as described previously (267). Samples were first demultiplexed based on perfectly matching barcodes. Read pairs with poor quality or improper mapping or alignment to target amplicon sequences were excluded from analysis. After all filtering steps, we obtained ≈413,000-494,000 read pairs per sample (see Table S3-2 and (267)). We then used a custom algorithm based on the Genome Analysis Toolkit (GATK) walker framework (262) to identify mutations and calculate mutation frequencies for total mutations as well as every possible mutational subclass. Mutations and wild-type bases were required to be identified on both forward and reverse reads, as read pairs were mostly overlapping. Mutations and wild-type bases were also required to have a Q-score ≥ 30 on both reads. Furthermore, background error hotspots were identified using the plasmid controls and masked prior to mutational analysis, as described previously (267). Most (≈83%) of the mutations excluded by this process were G-to-T or C-to-A transversions, and only one of the excluded mutations was a G-to-C or C-to-G transversion. Mutation frequencies (defined as mutations/bp) were calculated by dividing the number of mutations passing filters by all reference bases (mutations + wild-type bases) passing filters, both combined across amplicons and separated by amplicon. We also determined the relative percentage of each mutational type (mutation spectra), numbers of mutants, dinucleotide contexts of mutants, and the mutation frequencies at individual sequence positions within the target amplicons.

Biostatistical analysis of Illumina DNA sequencing data. To test for variables that may influence mutation frequencies, generalized linear mixed effects models were applied to the data that came from our Illumina data processing pipeline. The raw counts for each type of mutation (e.g. transitions) were modeled as Poisson random variables with an offset given by the total number of reference bases. The type of sample (i.e. HIV-1 without drug, HIV-1 + dicitabine, and HIV-1 plasmid control), the type of amplicon, and their interactions were treated as fixed effects while the replicate was treated as a random effect. The logarithmic link was used, as is standard for Poisson outcomes, and penalized quasilielihood was used to estimate the model parameters (263). These computations were conducted using R v 3.1.0 and the MASS package. Mutation frequencies at individual sequence positions were compared using the Mantel-Haenszel test. All figures and tables were created in GraphPad Prism v 6.0 (GraphPad Software, Inc.; La Jolla, CA) or Microsoft Office for Mac 2011 v 14.3.8 (Microsoft Corp.; Redmond, WA).

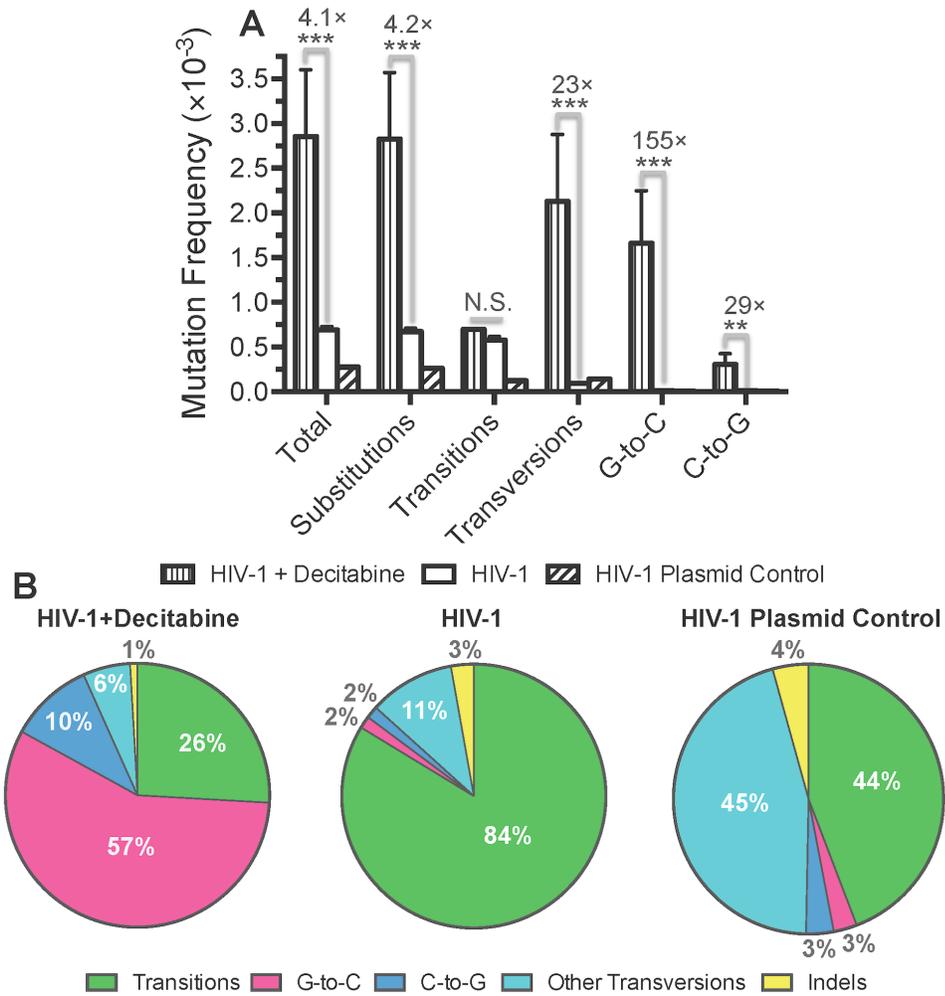


Figure 3-1. Decitabine primarily induces G-to-C transversion mutations during HIV-1 replication. A. Mutation frequency analysis. Using Illumina sequencing data, mutation frequencies were calculated by dividing the number of mutations by the number of reference bases (mutations + wild-type bases) and are expressed as mutations/bp, or m/bp. Mutation frequencies were determined for HIV-1 in the presence or absence of decitabine (2 μ M, an \approx EC₆₅ value), as well as for a plasmid control to determine the level of background errors. Fold increases in mutation frequencies in the presence of decitabine are indicated. B. Mutation spectra analysis. Mutation spectra were determined by dividing the frequency of each type of mutation by the total mutation frequency, with the results expressed as a percentage of total mutations. Data in both panels

represent the mean of three experimental replicates, with error bars indicating the standard deviation. Asterisks denote statistically significant differences between HIV-1 and HIV-1 + decitabine (** $p < 0.01$, *** $p < 0.001$, N.S. = not significant). The actual numbers of read pairs, mutations, and reference bases are listed in Table S3-2 and in a previous report (267).

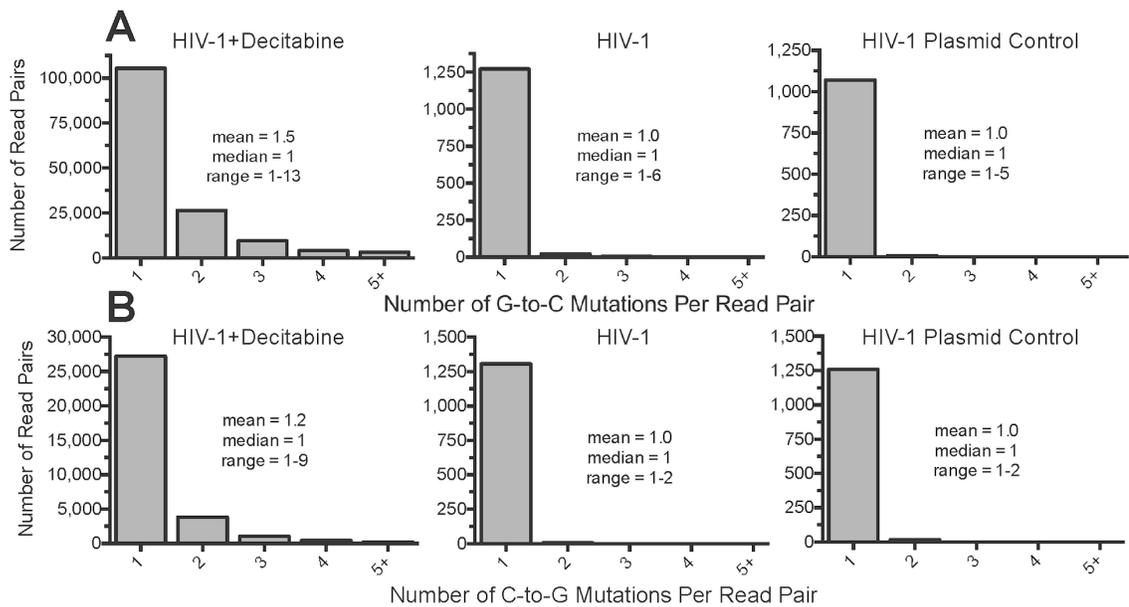


Figure 3-2. Most decitabine-induced G-to-C and C-to-G mutant read pairs have relatively low mutational loads. For all G-to-C (A) or C-to-G (B) mutants, the numbers of G-to-C or C-to-G mutations per read pair were determined in order to evaluate mutational loads. The means, medians, and ranges are also indicated. Note that due to the short length of the Illumina read pairs (≈ 120 bp after processing), the total numbers of G-to-C and C-to-G mutations in the full-length viral genome may be much higher. The figure represents compiled data across three experimental replicates.

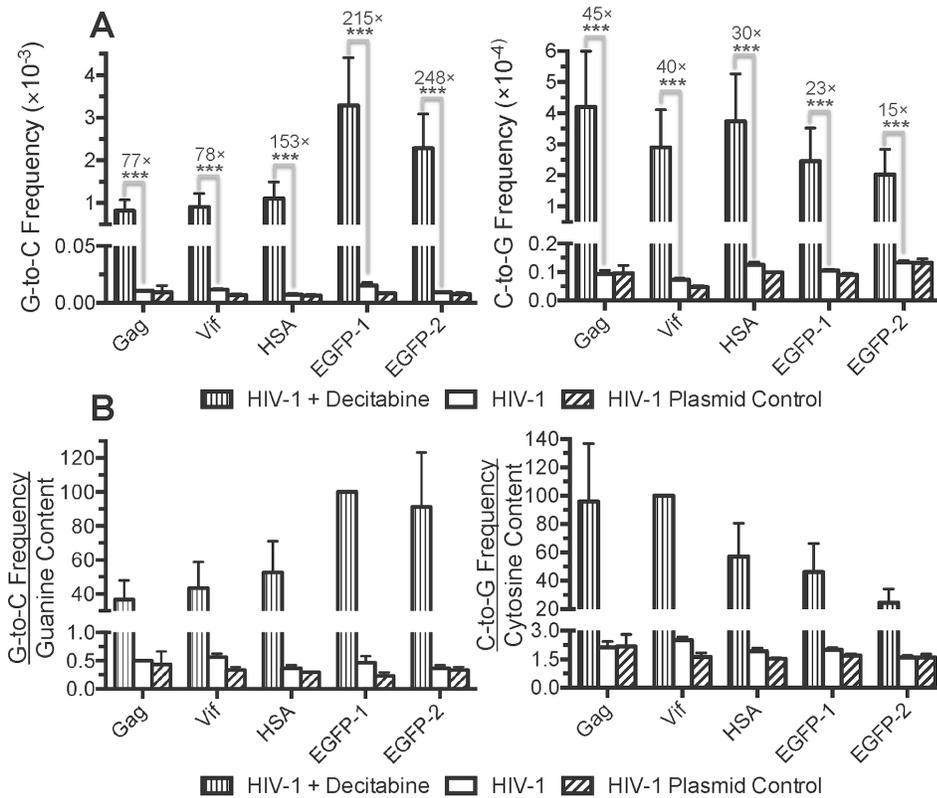


Figure 3-3. All amplicons are susceptible to decitabine-mediated HIV-1 mutagenesis. A. Analysis of mutation frequencies separated by amplicon. G-to-C and C-to-G transversion frequencies were determined in all five amplicons subjected to Illumina sequencing. The EGFP-1 and EGFP-2 amplicons represent non-overlapping portions of the *egfp* gene. Fold increases in mutation frequencies in the presence of decitabine are indicated. B. Analysis of amplicon mutation frequencies normalized for variations in amplicon nucleoside content. To account for varying nucleoside content across amplicons, G-to-C and C-to-G transversion frequencies were divided by the proportion of guanines or cytosines, respectively, within each amplicon. These values were then normalized to the amplicon with the highest G-to-C or C-to-G frequency (EGFP-1 and Vif, respectively). Data in both panels represent the mean of three experimental replicates, with error bars indicating the standard deviation. Asterisks denote statistically significant differences between HIV-1 and HIV-1 + decitabine (** $p < 0.001$).

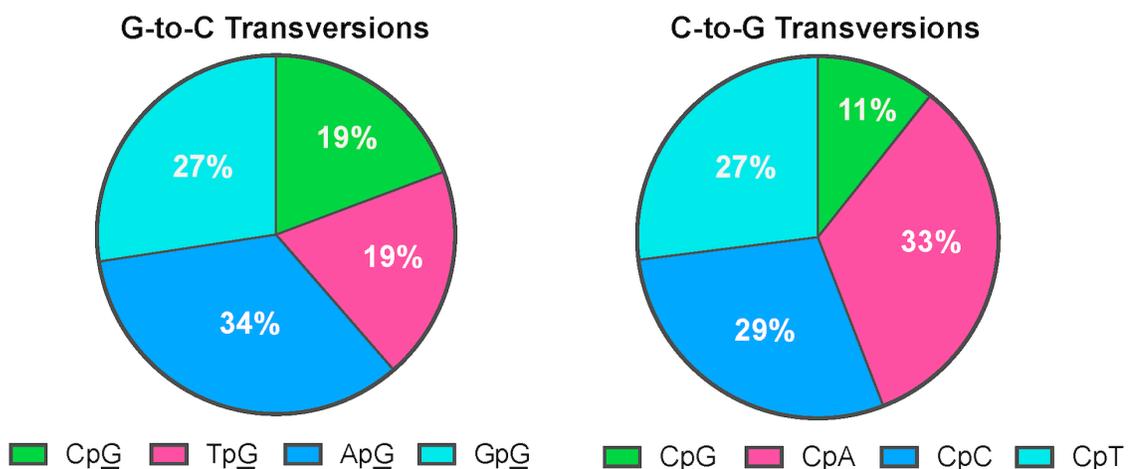


Figure 3-4. Decitabine-induced mutations in HIV-1 are not caused by DNA methyltransferase-mediated hydrolysis. Decitabine-induced mutations in eukaryotic genomic DNA can either be biased toward CpG dinucleotides, implicating formation via DNA methyltransferases (196), or not be biased toward CpG dinucleotides, implicating spontaneous structural rearrangement (273). In order to investigate the mechanism most important for decitabine-mediated HIV-1 mutagenesis, the dinucleotide contexts of all G-to-C and C-to-G transversions observed in the presence of decitabine were determined. Note that G-to-C transversions likely arise from decitabine incorporation in place of deoxycytidine in the minus strand viral DNA, such that a 5'-CpG-3' context on the minus strand corresponds to a 5'-CpG-3' context on the plus strand (sites of mutation underlined). The data represent the mean of three experimental replicates.

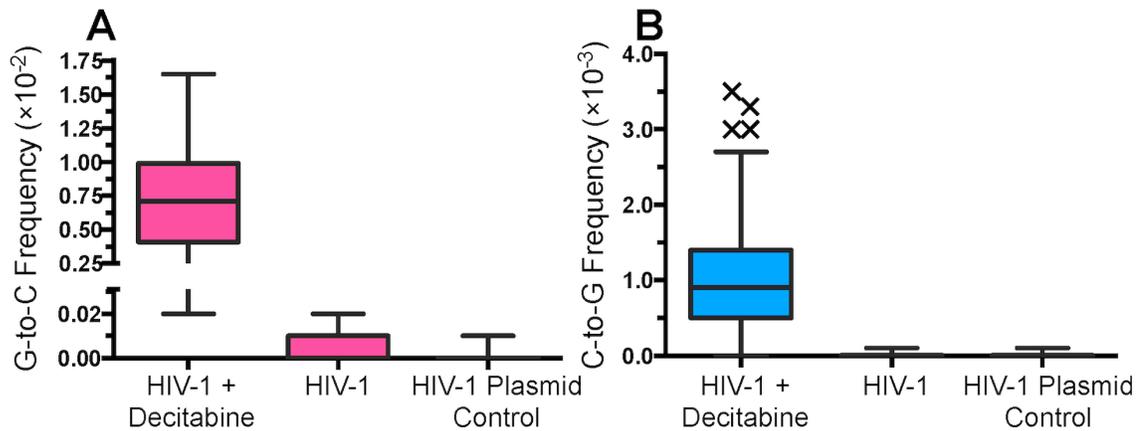


Figure 3-5. Lack of decitabine-induced G-to-C transversion mutation hotspots. In order to determine whether certain sequence positions in HIV-1 were more susceptible to decitabine than others, the G-to-C frequency at each individual guanine position (134 in total, panel A) and the C-to-G frequency at each individual cytosine position (169 in total, panel B) were determined. The boxplots indicate the observed minima, maxima, and medians. We also investigated whether decitabine induced mutational hotspots, which we defined as upper outliers within the mutation frequency distributions using the $1.5 \times$ interquartile range rule. Decitabine was not found to induce any G-to-C mutational hotspots, although several C-to-G mutational hotspots (denoted by \times) were observed. The data represent the mean of three experimental replicates.

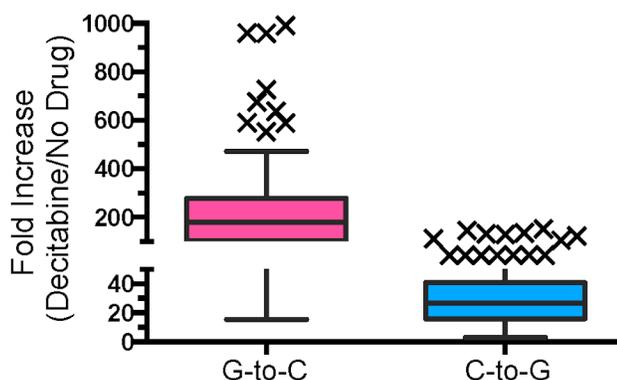


Figure 3-6. Virtually all guanine and cytosine positions are susceptible to decitabine-mediated mutagenesis. In order to further investigate the susceptibilities of individual nucleotides in HIV-1 to decitabine, fold differences in mutation frequencies (decitabine/no drug) were calculated at every possible sequence position. Fold differences could not be calculated for a small number of positions (1 of 134 for G-to-C and 6 of 169 for C-to-G), as they were not mutated in the absence of decitabine. Mutation frequencies at individual sequence positions were also compared statistically using the Mantel-Haenszel test (see Results).

Table 3-1. Locations and effects of common decitabine-induced G-to-C transversions in HIV-1.^a

<i>Rank</i>	<i>Position^b</i>	<i>Amplicon</i>	<i>Frequency^c</i>	<i>Mutation Type^d</i>	<i>Amino Acid Change</i>
1	324	EGFP-1	1.65E-02	N	K108N
2	417	EGFP-1	1.56E-02	N	E139D
3	383	EGFP-1	1.49E-02	N	G128A
4	608	EGFP-2	1.45E-02	N	S203T
5	427	EGFP-1	1.43E-02	N	E143Q
6	416	EGFP-1	1.41E-02	N	G139A
7	429	EGFP-1	1.40E-02	N	E143D
8	388	EGFP-1	1.34E-02	N	D130H
9	415	EGFP-1	1.33E-02	N	G139R
10	368	EGFP-1	1.32E-02	N	R123P
11	606	EGFP-2	1.29E-02	S	N/A
12	555	EGFP-2	1.29E-02	N	Q185H
13	552	EGFP-2	1.25E-02	N	Q184H
14	352	EGFP-1	1.22E-02	N	D118H
15	350	EGFP-1	1.21E-02	N	G117A
16	626	EGFP-2	1.19E-02	N	S209T
17	534	EGFP-2	1.18E-02	N	Q178H
18	624	EGFP-2	1.17E-02	S	N/A
19	592	EGFP-2	1.16E-02	N	D198H
20	615	EGFP-2	1.15E-02	N	Q205H

a. The data indicate the positions, frequencies, and coding effects of the 20 most common G-to-C transversions observed in HIV-1 exposed to decitabine, although these were not technically defined as mutational hotspots by our statistical criterion.

b. All numbers refer to the position relative to the start of the *egfp* gene.

c. The G-to-C transversion frequency (mutations/bp) was determined at every individual guanine position, though only the top 20 are shown here. The data represent the average of three experimental replicates.

d. N = non-synonymous mutation, S = synonymous mutation

Table 3-2. Locations and effects of common decitabine-induced C-to-G transversions in HIV-1.^a

<i>Rank</i>	<i>Position^b</i>	<i>Amplicon</i>	<i>Frequency^c</i>	<i>Mutation Type^d</i>	<i>Amino Acid Change</i>
1*	438	Gag	3.50E-03	S	N/A
2*	446	Gag	3.30E-03	N	P149R
3*	392	Vif	3.00E-03	N	P131R
4*	431	Vif	3.00E-03	N	S144C
5	452	Gag	2.70E-03	N	T151S
6	379	Vif	2.60E-03	N	R127G
7	412	Gag	2.60E-03	N	L138V
8	388	Gag	2.40E-03	N	Q130E
9	356	Vif	2.30E-03	N	A119G
10	369	Vif	2.30E-03	S	N/A
11	437	Gag	2.20E-03	N	A146G
12	398	Gag	2.20E-03	N	P133R
13	414	Gag	2.20E-03	S	N/A
14	415	Vif	2.10E-03	N	H139D
15	118	HSA	2.00E-03	N	P40A
16	368	Vif	2.00E-03	N	T123S
17	325	Vif	2.00E-03	N	L109V
18	445	Gag	2.00E-03	N	P149A
19	160	HSA	1.90E-03	N	L54V
20	461	Gag	1.90E-03	N	A154G

a. The data indicate the positions, frequencies, and coding effects of the 20 most common C-to-G transversions observed in HIV-1 exposed to decitabine, although only the top four of these (those marked by asterisks) were mutational hotspots, defined as upper outliers within the mutation frequency distribution using the $1.5 \times$ interquartile range rule.

b. All numbers refer to the position relative to the start of the indicated gene (*gag*, *vif*, or *hsa*).

c. The C-to-G transversion frequency (mutations/bp) was determined at every individual cytosine position, though only the top 20 are shown here. The data represent the average of three experimental replicates.

d. N = non-synonymous mutation, S = synonymous mutation

CHAPTER IV
5-AZACYTIDINE ENHANCES THE MUTAGENESIS OF HIV-1
BY REDUCTION TO 5-AZA-2'-DEOXYCYTIDINE

Jonathan M.O. Rawson, Michele B. Daly, Jiashu Xie, Christine L. Clouser, Sean R. Landman, Cavan S. Reilly, Laurent Bonnac, Baek Kim, Steven E. Patterson, and Louis M. Mansky. *Submitted.*

Introduction

RNA viruses exhibit high mutation rates and have been postulated to replicate near the error threshold—the maximal mutation rate compatible with the maintenance of genetic information (41, 177). Thus, they may be particularly sensitive to small molecules that promote viral mutations, an antiviral strategy called lethal mutagenesis (119). Lethal mutagenesis has been pursued as a potential antiviral approach for many different RNA viruses (274). Most small molecule candidates for lethal mutagenesis identified thus far have been nucleoside analogs with altered base-pairing properties. These nucleoside analogs base-pair promiscuously due to ionization, structural rearrangement, tautomerization, or conformational flexibility (264). Alternatively, small molecules can be used to promote viral mutagenesis by leveraging host nucleic acid editing enzymes that are part of the innate immune response. For example, molecules have been identified that block the degradation of APOBEC3 enzymes by the human immunodeficiency virus type-1 (HIV-1) Vif accessory protein (180-183). These molecules ultimately promote incorporation of APOBEC3 proteins into virions, resulting in lethal G-to-A hypermutation during the next cycle of replication.

The ribonucleoside analog 5-azacytidine (5-aza-C) reduces the infectivity of HIV-1 by inducing lethal mutagenesis (118). 5-aza-C is active during both the early and late phases of viral replication, reflecting incorporation during reverse transcription and during transcription of viral genomic RNA, respectively. When added during the late phase of viral replication, 5-aza-C induces primarily C-to-G transversions in HIV-1. In contrast, during the early phase of replication, 5-aza-C induces primarily G-to-C transversions in the virus. These G-to-C transversions are thought to be caused by the incorporation of 5-aza-C into minus strand viral DNA, followed by its hydrolysis and deformylation into ring-opened remnants (118). These ring-opened remnants can then mispair with deoxycytidine during plus strand synthesis, leading to the fixation of G-to-C transversions in proviral DNA. However, it remains possible that 5-aza-C hydrolysis products are directly

incorporated by HIV-1 reverse transcriptase (RT) as well. Notably, 5-aza-C is closely related to 5-aza-2'-deoxycytidine (5-aza-dC), another nucleoside analog that has been explored for the lethal mutagenesis of HIV-1 (117, 275), with the primary difference being that 5-aza-dC is much more potent than 5-aza-C and likely only incorporated into viral DNA.

Two different mechanisms could account for the antiviral activity of 5-aza-C during the early phase of replication: First, 5-aza-C could be incorporated during reverse transcription primarily as a deoxyribonucleotide (i.e. as 5-aza-2'-deoxycytidine-5'-triphosphate, or 5-aza-dCTP). For this to occur, the cellular enzyme ribonucleotide reductase (RNR) would have to first convert 5-aza-C (in its diphosphate form) to 5-aza-dC, which could then be phosphorylated to form 5-aza-dCTP. Notably, one previous study has demonstrated that ~10-20% of 5-aza-C is reduced to 5-aza-dC by RNR (276), suggesting that 5-aza-dCTP would likely be available for incorporation during reverse transcription. However, the reduction of 5-aza-C to 5-aza-dC has not yet been demonstrated in cell types for which antiviral activity has been reported. In further support of this possibility, HIV-1 RT has been shown to selectively exclude ribonucleotides using a residue (Y115) that acts as a steric gate (216-218). Alternatively, 5-aza-C could be incorporated during reverse transcription directly as a ribonucleotide (i.e. as 5-azacytidine-5'-triphosphate, or 5-aza-CTP). Notably, HIV-1 RT has been found to incorporate significant levels of endogenous ribonucleotides when the levels of deoxyribonucleotides are very low (resulting in high NTP:dNTP ratios), such as in macrophages (219, 220). Previous studies have found that high concentrations of 5-aza-C are required to elicit antiviral activity in cell culture (118), potentially skewing the NTP:dNTP ratio enough to allow for significant 5-aza-CTP incorporation. However, these findings could also reflect the inefficient reduction of 5-aza-C to 5-aza-dC.

To determine the primary form of 5-aza-C active during HIV-1 reverse transcription, Illumina high-throughput sequencing was performed to compare viral mutagenesis in the presence of 5-aza-C or 5-aza-dC. 5-aza-C and 5-aza-

dC caused similar levels of G-to-C and C-to-G transversions in HIV-1. Further, G-to-C and C-to-G transversions were distributed at individual sequence positions in a highly similar pattern for 5-aza-C and 5-aza-dC. In addition, 5-aza-dCTP was detected in cells treated with 5-aza-C, indicating that 5-aza-C was reduced to 5-aza-dC by RNR. Importantly, 5-aza-dCTP levels were similar in cells treated with equivalent antiviral effect concentrations of 5-aza-C or 5-aza-dC. Lastly, 5-aza-CTP was incorporated *in vitro* by HIV-1 RT, but much less efficiently than 5-aza-dCTP. Overall, the data support the conclusion that 5-aza-C enhances HIV-1 mutagenesis and diminishes HIV-1 infectivity primarily by acting as 5-aza-dCTP, which may have important implications for the design of antiretroviral ribonucleoside analogs.

Results

5-azacytidine and 5-aza-2'-deoxycytidine induce similar levels of G-to-C and C-to-G transversions in HIV-1. In order to determine the primary form of 5-aza-C active against HIV-1, the effects of 5-aza-C and 5-aza-dC on HIV-1 mutagenesis were compared using Illumina amplicon sequencing. The direct incorporation of 5-aza-C as a ribonucleotide (5-aza-CTP) was predicted to lead to a distinct mutational pattern relative to 5-aza-dCTP, because incorporated ribonucleotides can: 1. be replaced (correctly or incorrectly) by RNase H2-mediated repair, 2. cause mutations if left unrepaired, particularly short (2-5 bp) deletions, and 3. elevate the HIV-1 RT mismatch extension frequency on the opposite strand (277-279). To prepare samples for sequencing, U373-MAGI cells were treated with DMSO (i.e. no drug), 5-aza-C (EC_{75} , ~260 μ M) or 5-aza-dC (EC_{75} , ~3.8 μ M). Cells were then infected 2 h after drug addition at an MOI of 1.0 with NL4-3 MIG-VSVG. Genomic DNA was purified from cells collected 72 h post-infection, and PCR was performed of multiple amplicons (Gag, Pol, Vif, Env, Nef) from proviral DNA. Plasmid control amplifications were included to measure the level of background error from PCR and sequencing. The amplicons were pooled, used to prepare libraries, and subjected to 2×250 paired-end sequencing

on the Illumina MiSeq. Illumina sequencing generated ~5.5 million read pairs after all bioinformatics processing steps, containing ~570,000 mutations and ~630 million reference bases. 5-aza-C and 5-aza-dC were found to significantly increase the frequencies of G-to-C and C-to-G transversions relative to the no drug control in all amplicons examined ($p < 0.0001$ for all comparisons), although C-to-G transversions were ~5 to 6-fold less prevalent than G-to-C transversions (Figure 4-1). In the presence of 5-aza-C, ~67% of all mutations were G-to-C transversions and ~12% were C-to-G transversions (compared to 64% and 12% for 5-aza-dC). C-to-G transversions were not previously observed when 5-aza-C was added during the early phase of viral replication, likely due to the lower sequencing depth of earlier studies (see Discussion). Notably, 5-aza-C and 5-aza-dC increased the levels of G-to-C transversions (Figure 4-1A) and C-to-G transversions (Figure 4-1B) to similar extents in all five amplicons, and the resulting mutation frequencies did not significantly differ ($p > 0.05$ for all comparisons). Depending on the amplicon, 5-aza-C and 5-aza-dC raised G-to-C frequencies by ~45 to 75-fold and C-to-G frequencies by ~25 to 60-fold. Further, we did not observe short (2-5 bp) deletions that are characteristic of ribonucleoside incorporation in the 5-aza-C-treated sample (data not shown). These data demonstrate that 5-aza-C and 5-aza-dC cause very similar changes in viral mutation frequencies and spectra, arguing that 5-aza-C acts primarily as 5-aza-dCTP during reverse transcription.

5-azacytidine and 5-aza-2'-deoxycytidine cause highly similar HIV-1 mutational patterns. 5-aza-dC has previously been demonstrated to cause highly variable levels of G-to-C and C-to-G transversions depending upon the specific sequence position (275). If 5-aza-C acts against HIV-1 after reduction to 5-aza-dC, the susceptibilities of individual bases to 5-aza-C and 5-aza-dC should closely resemble one another. To address this, mutation frequencies were determined at individual sequence positions in the five amplicons. G-to-C frequencies were determined at every guanine position (124 in total), while C-to-G frequencies were determined at every cytosine position (116 in total). The

extent of correlation between 5-aza-C and 5-aza-dC-induced mutation frequencies was then examined. For both 5-aza-C and 5-aza-dC, significant variability (up to ~20-fold) was observed between G-to-C and C-to-G frequencies at individual sequence positions (Figure 4-2). 5-aza-C and 5-aza-dC-mediated mutation frequencies were found to be strongly positively correlated, both for G-to-C transversions (Pearson r of 0.93 [95% CI: 0.90-0.95], $p < 0.0001$) and C-to-G transversions (Pearson r of 0.96 [95% CI: 0.94-0.97], $p < 0.0001$). This correlation was able to explain most of the variability in the observed data (R^2 of 0.87 [G-to-C] or 0.91 [C-to-G]). The slopes of the best-fit linear regression lines were also determined, as they would equal 1.0 if 5-aza-C and 5-aza-dC mutation frequencies perfectly matched. The regression lines exhibited slopes of 0.71 (G-to-C, 95% CI: 0.66-0.76) and 0.88 (C-to-G, 95% CI: 0.83-0.93), indicating that 5-aza-C-mediated mutation frequencies were slightly higher than for 5-aza-dC. These findings indicate that patterns of viral mutagenesis in the presence of 5-aza-C- and 5-aza-dC closely resembled one another, further indicating that 5-aza-C likely reduces HIV-1 infectivity after conversion to 5-aza-dC.

Levels of 5-aza-2'-deoxycytidine-triphosphate are similar in cells treated with 5-azacytidine or 5-aza-2'-deoxycytidine. While a previous report indicated that ~10-20% of 5-aza-C is reduced to 5-aza-dC in one particular cell line (276), the conversion of 5-aza-C to 5-aza-dC has not yet been assessed in cell lines for which antiviral activity has been demonstrated. To address this, cells were treated with multiple concentrations of either 5-aza-C or 5-aza-dC for 4 h (corresponding to a time point during which reverse transcription should be actively occurring), followed by extraction of nucleotides and LC-MS/MS to determine the relative quantities of 5-aza-CTP and 5-aza-dCTP (see Materials and Methods). The relative levels of RGU-TP and dRGU-TP were also determined, which are the final hydrolysis products of 5-aza-CTP and 5-aza-dCTP, respectively, and may be relevant to antiviral activity. Cells were treated with matched antiviral concentrations (EC_{25} , EC_{50} , or EC_{75}) of 5-aza-C or 5-aza-dC. If 5-aza-C acts primarily as 5-aza-dCTP against HIV-1, cells treated with 5-

aza-C or 5-aza-dC should contain approximately equal levels of 5-aza-dCTP at equivalent antiviral effect levels. 5-aza-CTP and RGU-TP were only detected in cells treated with 5-aza-C (data not shown), as expected considering the lack of cellular pathways for conversion of deoxyribonucleotides to ribonucleotides. In contrast, 5-aza-dCTP was detected in cells treated with either 5-aza-C or 5-aza-dC (Figure 4-3A), indicating that 5-aza-C (in its diphosphate form) was reduced to 5-aza-dC by RNR. Further, the level of 5-aza-dCTP was not significantly different between cells treated with 5-aza-C or 5-aza-dC (EC₂₅ and EC₇₅) or was higher in cells treated with 5-aza-C (EC₅₀, $p < 0.05$). Similar trends were observed for dRGU-TP, the final hydrolysis product of 5-aza-dCTP (Figure 4-3B). These findings further indicate that 5-aza-C primarily reduces the infectivity of HIV-1 after reduction to 5-aza-dC, as HIV-1 RT should incorporate 5-aza-dCTP much more readily than 5-aza-CTP.

HIV-1 reverse transcriptase incorporates 5-azacytidine-triphosphate in vitro much less efficiently than 5-aza-2'-deoxycytidine-triphosphate. Although similar levels of 5-aza-dCTP were observed in cells treated with 5-aza-C or 5-aza-dC, the extent to which 5-aza-CTP and 5-aza-dCTP are incorporated by HIV-1 RT has not yet been determined. 5-aza-CTP was hypothesized to be incorporated only weakly, while 5-aza-dCTP was hypothesized to be incorporated relatively efficiently, as 5-aza-dCTP can be efficiently incorporated by cellular DNA polymerases (280, 281). To address this, the incorporation of 5-aza-CTP and 5-aza-dCTP was compared using an HIV-1 RT *in vitro* single nucleotide extension assay, with dCTP and CTP included for comparative purposes. Additional control reactions were performed by incubating with all four standard dNTPs (50 μ M, +) or by omitting RT (-). It was found that HIV-1 RT incorporated 5-aza-dCTP relatively efficiently (Figure 4-4), resulting in total primer extension at 5 μ M (lane 10) and some detectable extension as low as 50 nM (lane 8). However, 5-aza-dCTP was incorporated somewhat less effectively than dCTP (compare lanes 7-9 with 19-21). In contrast, HIV-1 RT incorporated 5-aza-CTP and CTP only at the highest concentration tested (500 μ M) (Figure 4-

4A, lanes 6 and 18). These data demonstrate that 5-aza-CTP can be directly incorporated by the HIV-1 RT, but only at very high concentrations and in the absence of other competing deoxyribonucleotides. 5-aza-dCTP was incorporated at least 10,000-fold more efficiently than 5-aza-CTP (compare lanes 6 and 8). These observations are consistent with previous findings that HIV-1 RT exhibits high selectivity for deoxyribonucleotides, largely due to a steric gate (Y115) that blocks incoming ribonucleotides (216-218). Considering that similar levels of 5-aza-dCTP were observed in cells treated with equivalent antiviral concentrations of 5-aza-C or 5-aza-dC (Figure 4-3), these findings strongly indicate that 5-aza-C primarily exerts activity against HIV-1 through its reduction to 5-aza-dC.

Discussion

The ribonucleoside analog 5-aza-C could potentially act as a ribonucleotide or as a deoxyribonucleotide to reduce HIV-1 infectivity during the early phase of viral replication. To address this, Illumina sequencing was performed in order to compare the enhancement of HIV-1 mutagenesis by 5-aza-C and 5-aza-dC. 5-aza-C and 5-aza-dC were found to cause similar changes in viral mutation frequencies regardless of the amplicon examined (Figure 4-1). Both drugs primarily elicited G-to-C transversions in HIV-1, but also significant levels of C-to-G transversions, which were not observed in a previous study of 5-aza-C (118). 5-aza-C was not found to induce other types of mutations that are considered indicative of ribonucleotide incorporation, such as short (2-5 bp) deletions (278). While the mechanism of C-to-G transversions remains unclear, they may result from the incorporation of 5-aza-dCTP into plus strand viral DNA or, alternatively, from the direct incorporation of 5-aza-dCTP hydrolysis products (e.g. dRGU-TP) into minus strand viral DNA. 5-aza-C-mediated C-to-G transversions might not have been observed in a previous study (118) due to the reduced depth of Sanger sequencing, the lower concentration of 5-aza-C used, or other differences in the mutational assay. Additionally, 5-aza-C and 5-aza-dC-

mediated mutational patterns strongly correlated with one another even at the level of individual sequence positions in HIV-1 (Figure 4-2), further indicating that 5-aza-C exerts antiviral activity after reduction to 5-aza-dC.

5-aza-dCTP was detected in cells treated with 5-aza-C, indicating that 5-aza-C was reduced to 5-aza-dC (Figure 4-3), and, importantly, similar levels of 5-aza-dCTP were observed in cells treated with equivalent antiviral effect concentrations of 5-aza-C and 5-aza-dC. Notably, the absolute concentrations of 5-aza-C were ~70 to 300-fold higher than for 5-aza-dC in these experiments, indicating that intracellular conversion of 5-aza-C to 5-aza-dCTP was much less efficient than conversion of 5-aza-dC to 5-aza-dCTP. Using an *in vitro* incorporation assay, 5-aza-CTP was found to be successfully incorporated at very high concentrations by HIV-1 RT (Figure 4-4). However, 5-aza-dCTP was incorporated at least 10,000-fold more efficiently than 5-aza-CTP, further supporting 5-aza-dCTP as the principal form of 5-aza-C active against HIV-1.

Taken together, these findings clearly indicate that 5-aza-dCTP is the primary form of 5-aza-C active during the early phase of HIV-1 replication. Further, 5-aza-C has been shown to exert much more potent antiviral activity during the early phase of replication than during the late phase of replication (118), indicating that 5-aza-C likely acts primarily as 5-aza-dCTP during spreading infections as well. These observations lead to the following model for the antiviral activities of 5-aza-C and 5-aza-dC (Figure 4-5): 5-aza-C and 5-aza-dC enter the cell by facilitated diffusion through a transporter protein and are then phosphorylated by the appropriate cellular kinases. 5-aza-CTP is incorporated into viral RNA during the late phase of replication, leading to the fixation of C-to-G transversions during the following round of reverse transcription. 5-aza-C (in its diphosphate form) is also converted to 5-aza-dC by RNR, which can then be phosphorylated to form 5-aza-dCTP. Lastly, 5-aza-dCTP is incorporated into viral DNA during reverse transcription, resulting in the fixation of G-to-C and (to a lesser extent) C-to-G transversions in HIV-1.

In the future, it would be of interest to investigate the antiviral activity of 5-aza-C in other cell types, such as primary CD4+ T-cells and macrophages. 5-aza-C would likely act primarily as 5-aza-dCTP in activated CD4+ T-cells, as they rapidly divide and thus have high levels of RNR and dNTPs (219, 282). In contrast, 5-aza-CTP could potentially be directly incorporated at significant levels in macrophages, which have low levels of RNR, low dNTP pools, and much higher ratios of NTPs to dNTPs (219, 282). Indeed, previous work has shown that HIV-1 RT can incorporate significant levels of endogenous ribonucleotides in macrophages and *in vitro* under macrophage-like conditions (219, 220). Further, the ribonucleoside analog 3'-deoxyadenosine, which likely cannot be reduced, exhibits antiviral activity in macrophages but not in CD4+ T-cells, indicating that ribonucleoside analog triphosphates can be directly incorporated in macrophages (219).

Notably, the findings in this study have important implications for the design of ribonucleoside analogs directed against retroviruses. Ribonucleoside analogs are of interest as antiretroviral agents because their synthesis is often less complicated, less time-intensive, and more affordable than for their deoxyribonucleoside counterparts. Ribonucleoside analogs may also offer other advantages in certain instances, such as better activation by cellular enzymes or improved antiviral activity by the targeting of both viral RNA and DNA. However, the findings in this study suggest that it is important to consider the efficiency at which ribonucleoside analogs are reduced – at least for the targeting of viral replication in cell types with high dNTP pools, such as CD4+ T-cells. In contrast, efficient reduction may not be necessary or even desired for the targeting of viral replication in cell types with low dNTP pools, such as macrophages. Ribonucleoside analogs that cannot be reduced (such as those lacking a 3' hydroxyl group, e.g. 3'-deoxyadenosine) could be used to specifically target viral replication in macrophages, an important viral reservoir (219). In sum, these findings inform efforts to identify additional nucleoside analogs that target HIV-1

RT, which continue due to significant concerns regarding the development of drug resistance and off-target effects.

Materials and Methods

Plasmids, cell lines, and reagents. For single-cycle infections, viral stocks were produced by co-transfecting the HIV-1 envelope-deficient vector pNL4-3 MIG (120), which expresses mCherry and EGFP, and pHCMV-G, a kind gift from J. Burns (University of California, San Diego). Viral stocks were produced in human embryonic kidney (HEK 293T) cells from the American Type Culture Collection (Manassas, VA). Viral infections were performed in U373-MAGI-CXCR4_{CEM} cells obtained from Michael Emerman through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (257). 5-aza-C was purchased from Sigma-Aldrich (St. Louis, MO), while 5-aza-dC divalate (hereafter referred to simply as 5-aza-dC), a more stable prodrug form of 5-aza-dC, was synthesized by the Center for Drug Design at the University of Minnesota as previously described (266). 5-aza-CTP and 5-aza-dCTP were purchased from American Advanced Scientific (College Station, TX) and Jena Bioscience (Jena, Germany), respectively. All drugs were dissolved in the appropriate solvent and stored in aliquots at -20 °C.

Production and titration of viral stocks for Illumina sequencing. HIV-1 viral stocks for Illumina sequencing were produced by co-transfecting 10 µg of pNL4-3 MIG and 1 µg of pHCMV-G per 10 cm plate of 293T cells. Transfections were carried out using the PEI method, as previously described (120). Viral supernatants were collected 48 h post-transfection, treated with 10 U/mL of DNase I for 2 h at 37 °C to reduce plasmid carryover from transfections, and frozen in aliquots at -80 °C. Viral stocks were titered by infecting U373-MAGI cells (31,250 cells/well in 24-well plates) with volumes of virus ranging from 1.25 to 40 µL (2-fold series). Viral infectivity was determined by performing flow cytometry of cells collected 72 h post-infection. Viral titers were determined by

plotting the volume of virus against the percentage of infected cells (i.e. expressing mCherry and/or EGFP).

Determination of viral mutation frequencies by Illumina sequencing. Samples were prepared for Illumina amplicon sequencing as described previously (283), but with several minor modifications. Briefly, DMSO (for the no drug control), 5-aza-C (EC₇₅: 260 μ M) or 5-aza-dC (EC₇₅: 3.75 μ M) was added to U373-MAGI cells (1 million cells/sample) 2 h before infection. Cells were then infected at an MOI of 1.0 by adding the appropriate volume of NL4-3 MIG-VSVG. The medium was replaced 24 h post-infection, and cells were collected 72 h post-infection. Genomic DNA was purified from cell pellets, and plasmid carryover (from transfections) was then assessed by qPCR. Levels of plasmid carryover were determined by dividing the starting quantity of ampicillin resistance gene by the starting quantity of HIV-1 vif, using previously published primers (283). The average level of plasmid carryover was found to be ~0.4% (range 0.1-0.7%). Five small (~160-170 bp) amplicons (Gag, Pol, Vif, Env, Nef) were then prepared from each sample by PCR, using the primers listed in Table 4-1. PCR reactions were also performed from purified plasmid to determine the level of background error due to PCR and Illumina sequencing. For each sample, all amplicons were gel-purified, quantified, and pooled together in an equimolar fashion to normalize coverage between amplicons. Sequencing libraries were prepared from each sample using the TruSeq Nano DNA Kit, pooled in an equimolar fashion, and subjected to 2 \times 250 paired-end sequencing on the Illumina MiSeq. Sequencing reads were analyzed as described before (283) but were demultiplexed based on library indices instead of internal barcodes. As before, background error hotspots (mostly G-to-T and C-to-A transversions) were identified using the plasmid controls and masked prior to mutational analysis of the biological samples. The final sequencing data were also used in separate analyses that were focused on drug interactions between 5-aza-C and RNR inhibitors (Rawson J, et al. Submitted).

Determination of 5-azacytidine-triphosphate and 5-aza-2'-deoxycytidine-triphosphate levels by LC-MS/MS. Samples were prepared for liquid chromatography-tandem mass spectrometry (LC-MS/MS) by splitting U373-MAGI cells onto 10 cm plates (1.3 million cells/plate), using two plates per treatment group. The medium was replaced 24 h later, and 5-aza-C, 5-aza-dC, or DMSO (for the no drug control) was added. 5-aza-C and 5-aza-dC were added at equivalent antiviral effect levels (i.e. EC₂₅, EC₅₀, or EC₇₅). Cells were collected 4 h after 5-aza-C or 5-aza-dC addition, corresponding to the time of early reverse transcription (data not shown). Cell pellets were then resuspended in 750 µL of 60% methanol (stored at -20 °C) and incubated at -20 °C for 18 h. The samples were vortexed, heated at 95 °C for 3 min, and centrifuged at 16,000×g for 5 min. The supernatants were transferred to new microcentrifuge tubes, dried using a Savant SPD1010 SpeedVac Concentrator (Thermo Fisher Scientific, Inc.) at maximum pressure, and frozen at -80 °C until the time of analysis.

To perform LC-MS/MS, dried extracts were dissolved in 200 µL of water containing 10 µM of an internal standard, 5-iodo-dCTP. The samples were centrifuged at 14,000 rpm for 5 min at 4 °C, and the supernatants were analyzed by LC-MS/MS. 5-aza-C and 5-aza-dC samples were reconstituted immediately prior to LC-MS/MS injection to minimize their degradation during the procedure. LC-MS/MS was used to determine levels of 5-aza-CTP, 5-aza-dCTP, riboguanylurea-5'-triphosphate (RGU-TP, the final hydrolysis product of 5-aza-CTP), and 2'-deoxyriboguanylurea-5'-triphosphate (dRGU-TP, the final hydrolysis product of 5-aza-dCTP), using a previously published method (284) with minor modifications. Purified 5-aza-CTP and 5-aza-dCTP were included as standards to verify the procedure. The LC-MS/MS system consisted of an AB Sciex QTrap 5500 mass spectrometer and an Agilent 1260 Infinity HPLC. Chromatographic separation of analytes was achieved using a Thermo Scientific Hypercarb column (100 × 3 mm, 5 µm). The two eluents were: (A) 0.5 % diethylamine in water, pH adjusted to 10 with acetic acid; and (B) 50% acetonitrile in water. The

mobile phase was delivered at a flow rate of 0.5 mL/min using stepwise gradients of A and B: 0–20 min, 0–25% B (v/v); 20–28 min, 25–50% B (v/v); 28–28.5 min, 50–95% B (v/v); 28.5–30.5 min, 95–95% B (v/v); 30.5–31 min, 95–0% B, (v/v); 31–39 min, 0–0% B (v/v). Only eluate from 10–30 min was diverted into the mass spectrometer for analysis. MS/MS detection of the analytes was conducted using an ESI ion source with MRM detection in negative mode, with the curtain gas set to 20 psi. The ionspray voltage was set at -4500 V, and the temperature at 650 °C. The nebulizer gas (GS1) and turbo gas (GS2) were both set at 45 psi.

HIV-1 RT in vitro incorporation assay. The HIV-1 RT single nucleotide extension assay was performed similarly to a previous study (114) but with minor modifications. Briefly, a ³²P 5'-end-labeled DNA primer of 18 nt (5'-GTCCCTCTTCGGGCGCCA-3') was annealed to a DNA template of 19 nt (3'-CAGGGAGAAGCCCGCGGTG-5') at a 1:2 ratio. Single nucleotide extension indicates that the compound of interest (5-aza-CTP, 5-aza-dCTP, CTP, or dCTP) was successfully incorporated. The reactions contained 200 fmol template/primer, 2 µL of the test compound at the indicated concentrations (or 2 µL of dNTPs for a final concentration of 50 µM for the positive control), 4 µL of purified RT (HIV-1 NL4-3), 25 mM Tris–HCl, pH 8.0, 2 mM dithiothreitol, 100 mM KCl, 5 mM MgCl₂, and 10 µM oligo(dT), for a final volume of 20 µL/reaction. Reactions were incubated at 37 °C for 5 min and then quenched with 10 µL of 40 mM EDTA and 99% (vol/vol) formamide at 95 °C for 2 min. The reactions were resolved on a 14% urea-PAGE gel (AmericanBio, Inc.; Natick, MA) and analyzed using a PharosFX Molecular Imager (Biorad) and Image Lab software (Biorad).

Statistical analyses. All figures were created in Microsoft Office for Mac 2011 v 14.5.2 (Redmond, WA) or GraphPad Prism v 5.0 (GraphPad Software, Inc.; La Jolla, CA). To determine the EC₂₅, EC₅₀, and EC₇₅ of 5-aza-C and 5-aza-dC, infectivity data were normalized to the no drug control, plotted against log-transformed drug concentrations, and subjected to non-linear regression in GraphPad Prism. In order to determine whether differences in 5-aza-dCTP levels between cells treated with 5-aza-C and 5-aza-dC were statistically

significant, normalized data were analyzed by one-way repeated measures ANOVA (without assuming equal variability of differences) in GraphPad Prism. Sidak's post-test was used to compare the relative quantity of 5-aza-dCTP at each effect level (i.e. EC₂₅, EC₅₀, or EC₇₅). To test for variables that may influence mutation frequencies, generalized linear mixed effects models were applied to processed Illumina sequencing data. The raw counts for each type of mutation (e.g. transitions) were modeled as overdispersed Poisson random variables with an offset given by the total number of reference bases. The type of sample, the type of amplicon, and their interactions were treated as fixed effects while the replicate was treated as a random effect. The logarithmic link was used, as is standard for Poisson outcomes, and penalized quasilielihood was used to estimate the model parameters (263). These computations were conducted using R v 3.1.0 and the MASS package.

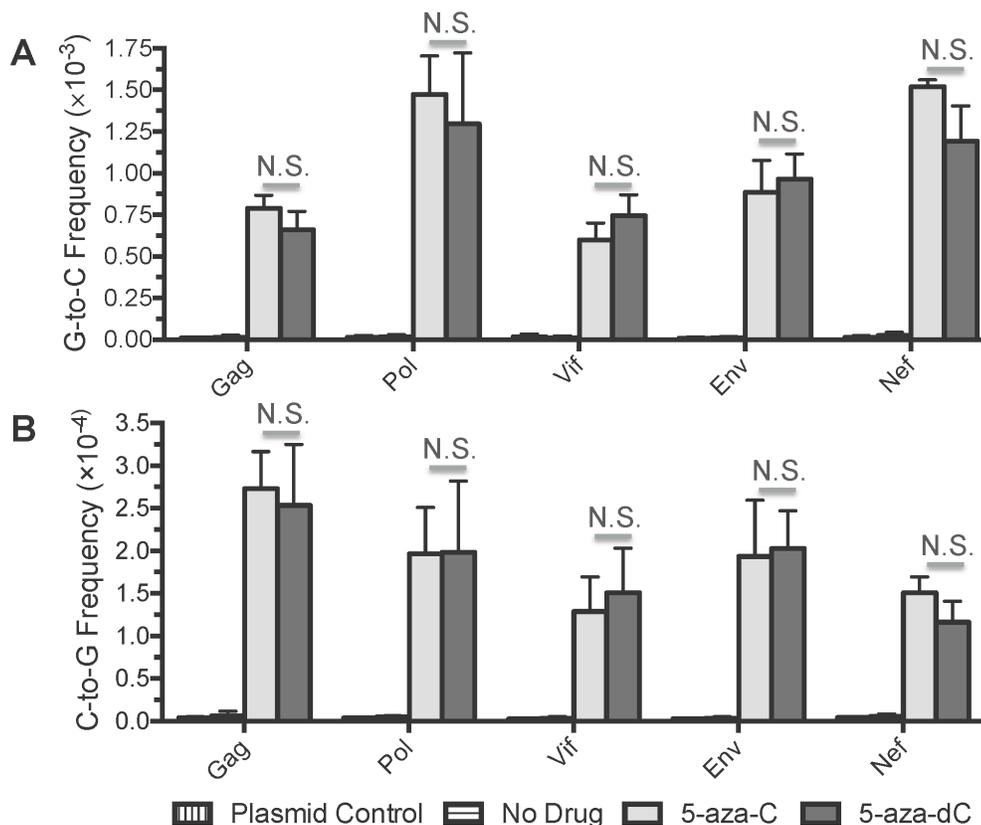


Figure 4-1. 5-azacytidine and 5-aza-2'-deoxycytidine induce similar levels of G-to-C and C-to-G transversion mutations during HIV-1 replication. In order to determine whether 5-azacytidine (5-aza-C) and 5-aza-2'-deoxycytidine (5-aza-dC) induce similar changes in HIV-1 mutation frequencies and spectra, U373-MAGI cells were treated with DMSO (for the no drug control), 5-aza-C, or 5-aza-dC. 5-aza-C and 5-aza-dC were added 2 h before infection at an EC₇₅ concentration (~260 or 3.8 μ M, respectively). Cells were infected at an MOI of 1.0 with NL4-3 MIG-VSVG and collected 72 h post-infection for purification of genomic DNA. PCR was performed to prepare multiple amplicons (Gag, Pol, Vif, Env, Nef) from proviral DNA, which were then pooled, used to prepare libraries, and analysed by 2 \times 250 paired-end sequencing on the Illumina MiSeq. Plasmid control amplifications were performed to determine the levels of background errors resulting from PCR and sequencing. Mutation frequencies for each

amplicon were calculated by dividing the number of mutations by the number of reference bases (mutations + wild-type bases) and are represented as mutations/bp, or m/bp. The data represent the mean \pm s.d. of three independent biological replicates; N.S.: not significant ($p > 0.05$).

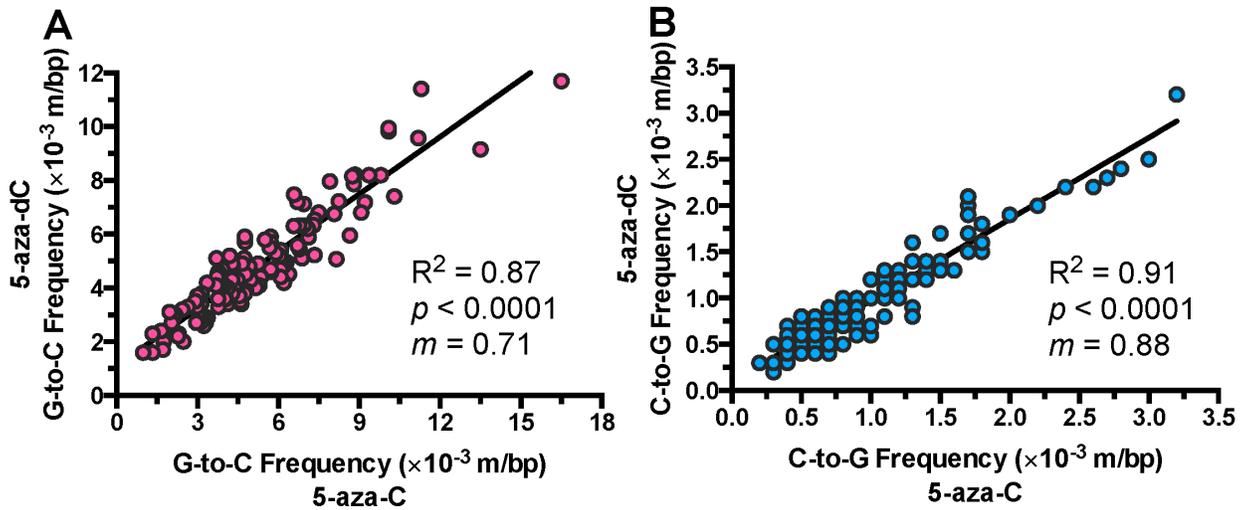


Figure 4-2. 5-azacytidine and 5-aza-2'-deoxycytidine induce similar patterns of mutation during HIV-1 replication. Using the Illumina sequencing data, G-to-C and C-to-G transversion frequencies were determined at every individual guanine position (124 in total) or cytosine position (116 in total) within the sequences of the five amplicons. 5-aza-C and 5-aza-dC mutation frequencies were then plotted against each other for every individual sequence position, and the resulting data were subjected to linear regression and correlation analyses. The data represent the average of three independent biological replicates, with R^2 denoting the extent to which the best-fit regression line explains observed variability in the data, p indicating the significance of the correlation, and m indicating the slope of the best-fit regression line.

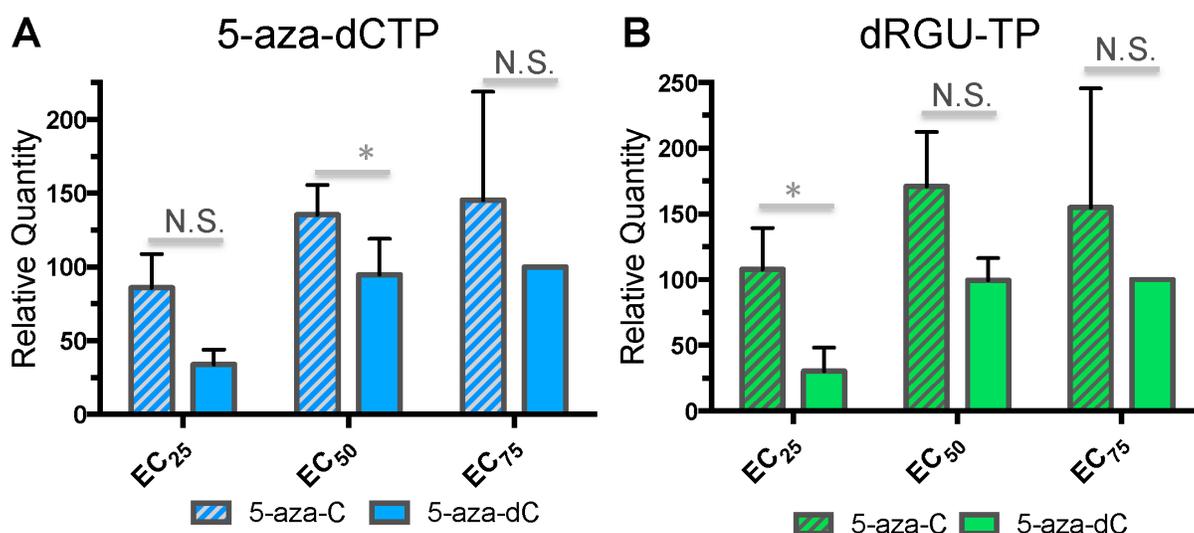


Figure 4-3. 5-aza-2'-deoxycytidine-triphosphate levels are comparable in cells treated with 5-azacytidine or 5-aza-2'-deoxycytidine. In order to determine the extent to which 5-aza-C is reduced intracellularly to 5-aza-dC, U373-MAGI cells were incubated with varying concentrations (EC₂₅, EC₅₀, or EC₇₅) of 5-aza-C or 5-aza-dC. Cells were collected for analysis four hours after drug addition, corresponding to the expected time of reverse transcription. LC-MS/MS was then used to determine the relative levels of 5-aza-2'-deoxycytidine-triphosphate (5-aza-dCTP) and 2'-deoxyriboguanylurea-5'-triphosphate (dRGU-TP). dRGU-TP is the final hydrolysis product of 5-aza-dCTP, and it could be incorporated directly by HIV-1 RT and thus relevant to antiviral activity. The data represent the mean \pm s.d. of three independent experiments, normalized to the EC₇₅ concentration of 5-aza-dCTP; N.S.: not significant, * $p < 0.05$.

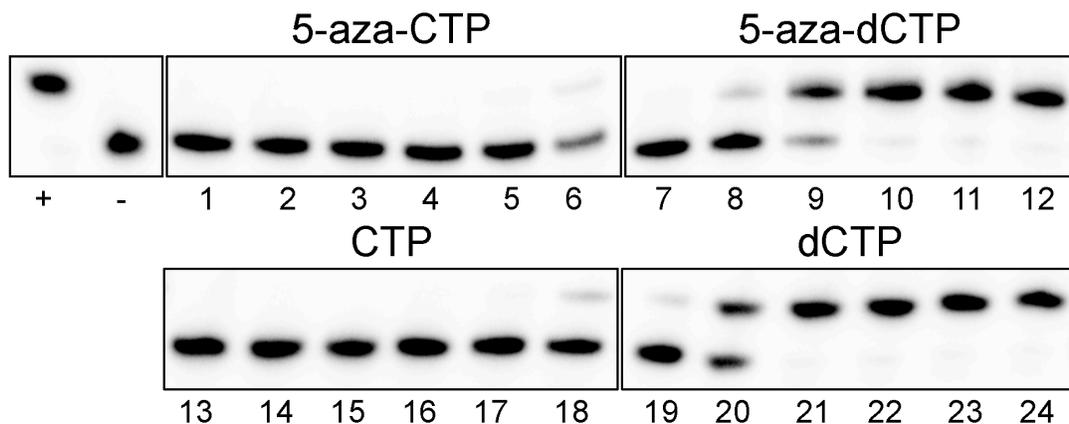


Figure 4-4. HIV-1 RT incorporates 5-azacytidine-triphosphate *in vitro* much less efficiently than 5-aza-2'-deoxycytidine-triphosphate. The relative ability of HIV-1 RT to incorporate 5-azacytidine-triphosphate (5-aza-CTP), 5-aza-2'-deoxycytidine-triphosphate (5-aza-dCTP), CTP, and dCTP was determined using an *in vitro* single nucleotide extension assay. The HIV-1 RT was incubated with a radiolabeled primer (18 nt) annealed to a DNA template (19 nt) in the presence of 5 nM to 500 μ M (10-fold series from left to right) of each compound. Additional control reactions were performed by incubating with all four standard dNTPs (50 μ M, +) or by omitting RT (-). All of the reactions were analyzed on the same gel.

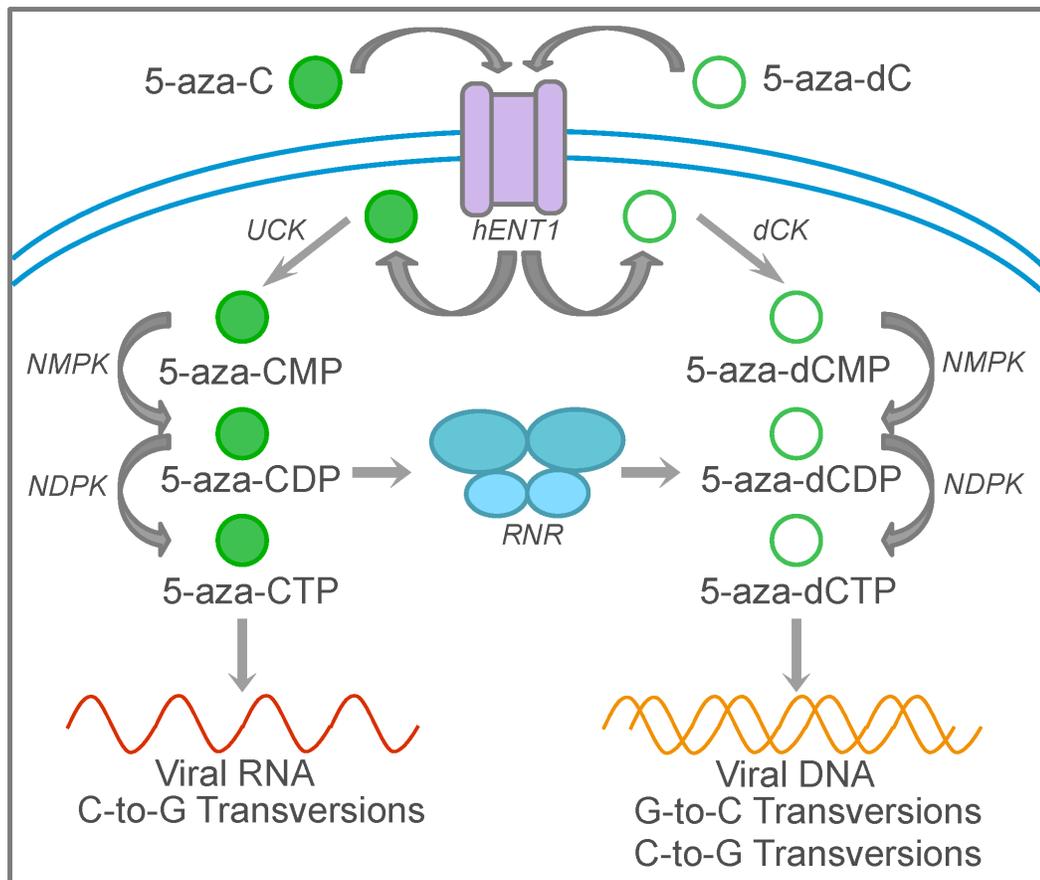


Figure 4-5. Model of 5-azacytidine and 5-aza-2'-deoxycytidine-mediated HIV-1 mutagenesis. 5-aza-C and 5-aza-dC are first transported into the cell by facilitated diffusion through the human equilibrative nucleoside transporter 1 (hENT1). 5-aza-C and 5-aza-dC are then phosphorylated by uridine-cytidine kinase (UCK) or deoxycytidine kinase (dCK), respectively, to form monophosphates. Nucleoside monophosphate and diphosphates kinases (NMPK and NDPK) phosphorylate the monophosphate and diphosphate forms, respectively, resulting in the formation of 5-aza-CTP and 5-aza-dCTP. 5-aza-CTP can be incorporated during transcription of viral genomic RNA, resulting in C-to-G transversions. 5-azacytidine-diphosphate (5-aza-CDP) is also reduced to 5-aza-2'-deoxycytidine-diphosphate (5-aza-dCDP), ultimately forming 5-aza-dCTP. 5-aza-dCTP is incorporated during reverse transcription, resulting in primarily G-to-C transversions (reflecting minus strand incorporation), but also

low levels of C-to-G transversions. The mechanism by which C-to-G transversions are formed is still unclear but may be due to plus strand incorporation of 5-aza-dCTP or minus strand incorporation of 5-aza-dCTP hydrolysis products.

Table 4-1. Primer sequences for Illumina amplicon sequencing of HIV-1.^a

<i>Amplicon</i>	<i>Sense</i>	<i>Forward Primer</i>
Gag	+	NNAAGCAGCAGCTGACACAGGAA
	-	NNTTCTGGGCTGAAAGCCTTCTCT
Pol	+	NNAGCAGTTCATGTAGCCAGTGGA
	-	NNACAGGCGGCCTTAACTGTAGTA
Vif	+	NNACCCTGACCTAGCAGACCAACTAA
	-	NNGCTGCTAGTGCCAAGTACTGTAGA
Env	+	NNACAGCTGAACACATCTGTAG
	-	NNAGTGGCATTCCATTTTGCTC
Nef	+	NNGCAGCTGTAGATCTTAGCCA
	-	NNTCAGTGGATATCTGACCC

^a Illumina primers were designed to prepare five different amplicons (~160-170 bp in length) from different viral genes. Primers were designed with two 5' terminal degenerate bases (NN) to promote similar efficiency of ligation to Illumina adapters. The Gag and Vif primers are the same as those utilized in a previous study (283) but lacking barcodes, as samples were demultiplexed based on library index rather than internal barcode for this study.

CHAPTER V
SYNERGISTIC REDUCTION OF HIV-1 INFECTIVITY BY 5-AZACYTIDINE AND
INHIBITORS OF RIBONUCLEOTIDE REDUCTASE

Jonathan M.O. Rawson, Megan E. Roth, Jiashu Xie, Christine L. Clouser, Sean R. Landman, Cavan S. Reilly, Laurent Bonnac, Steven E. Patterson, and Louis M. Mansky. *Submitted.*

Introduction

Human immunodeficiency virus type-1 (HIV-1) infects ~37 million individuals worldwide and has resulted in ~39 million deaths since the onset of the pandemic (www.unaids.org). Despite the approval of almost 30 drugs for the treatment of HIV-1 infection, current therapies are still limited by issues such as toxicity, cross-resistance, and the transmission of drug resistance variants (285-287). Like most RNA viruses (61), HIV-1 exhibits a high mutation rate (53), which enables rapid adaptation but may also confer susceptibility to therapeutics that further elevate the viral mutation rate, an antiviral strategy termed lethal mutagenesis (119). Lethal mutagenesis was first demonstrated by Holland and colleagues using a variety of mutagenic ribonucleosides and alkylating agents in poliovirus and vesicular stomatitis virus (178), but has since been demonstrated in cell culture for many other RNA viruses (274). Lethal mutagenesis has also been explored *in vivo* using a small number of mutagenic drugs against a limited number of viruses, but clinical translation has proceeded slowly. Ribavirin may block hepatitis C virus replication *in vivo* by inducing lethal mutagenesis, but these findings remain controversial because ribavirin has several other reported antiviral mechanisms as well (184-187). Notably, viral variants with modest decreases in fidelity are often severely attenuated *in vivo* in terms of viral fitness and pathogenicity, further supporting lethal mutagenesis as an antiviral strategy (288-293). Additionally, lethal mutagenesis using small molecule mutagens closely mimics innate antiviral systems based on nucleic acid editing, such as APOBEC3-mediated hypermutation that restricts the replication of HIV-1 (in the absence of Vif) and other viruses (294, 295).

Most of the drugs investigated thus far for HIV-1 lethal mutagenesis have been mutagenic ribonucleoside or deoxyribonucleoside analogs with altered base-pairing properties, such as 5-azacytidine (5-aza-C), 5-aza-2'-deoxycytidine (5-aza-dC), 5,6-dihydro-5-aza-2'-deoxycytidine (KP-1212), and 5-hydroxy-2'-deoxycytidine (5-OH-dC) (117-119, 198). Mutagenic deoxyribonucleoside analogs can be potentiated in some cases by small molecule inhibitors of

ribonucleotide reductase (RNR). RNR inhibitors prevent the reduction of ribonucleotide diphosphates to deoxyribonucleoside diphosphates and thus ultimately deplete endogenous deoxyribonucleotide triphosphate (dNTP) pools and introduce dNTP pool imbalances, as certain dNTPs are reduced more dramatically than others (296). RNR inhibitors typically reduce HIV-1 infectivity even in the absence of nucleoside analogs, likely by inhibiting viral DNA synthesis (due to a lack of dNTPs) and increasing the viral mutant frequency (due to dNTP pool imbalances) (117, 120, 201, 297-299). In combination experiments, low concentrations of RNR inhibitors (i.e. concentrations with little or no antiviral activity) often potentiate mutagenic deoxyribonucleoside analogs, resulting in antiviral synergy (117, 201). This strategy has also been used to potentiate the activity of conventional chain-terminating nucleoside analogs against HIV-1 (203, 204, 214, 297). While the mechanisms of many of these combinations have not been fully resolved, RNR inhibitors may upregulate the import, phosphorylation, and/or incorporation of deoxyribonucleoside analogs by depleting the competing endogenous dNTPs.

The mutagenic ribonucleoside analog 5-aza-C is active during both the early phase (i.e. reverse transcription) and the late phase (i.e. transcription) of HIV-1 replication (118). Although HIV-1 reverse transcriptase (RT) can incorporate significant levels of ribonucleotides in certain circumstances (219, 220), 5-aza-C was recently demonstrated to act primarily after reduction to 5-aza-dC during HIV-1 reverse transcription (Rawson J, *et al. Submitted*). Thus, RNR inhibitors were predicted to effectively antagonize the antiviral activity of 5-aza-C by inhibiting its reduction to 5-aza-dC. Surprisingly, 5-aza-C and RNR inhibitors exhibited antiviral synergy rather than antagonism, and RNR inhibitors were able to improve the selectivity index of 5-aza-C. RNR inhibitors caused subtle changes in endogenous dNTP pools but failed to efficiently inhibit reduction of 5-aza-C to 5-aza-dC, suggesting that 5-aza-C acted primarily as 5-aza-dC against HIV-1 even in the combinations. The combination of resveratrol and 5-aza-C was investigated further, and antiviral synergy was found to be

primarily due to the reduced accumulation of RT products rather than enhanced viral mutagenesis. Taken together, these observations demonstrate the synergistic reduction of HIV-1 infectivity by 5-aza-C and RNR inhibitors. These findings suggest new potential avenues for enhancing anti-HIV-1 therapy through the use of synergistic drug interactions.

Results

5-azacytidine synergizes with multiple ribonucleotide reductase inhibitors to reduce HIV-1 infectivity. Considering that 5-aza-C acts primarily after reduction to 5-aza-dC during HIV-1 reverse transcription (Rawson J, et al. Submitted), 5-aza-C activity was predicted to be antagonized by the down-regulation or inhibition of RNR. To address this, RNA interference was performed to decrease expression of the large subunit of RNR (RRM1), but RRM1 knockdown was extremely cytotoxic and, unlike most RNR inhibitors, did not exhibit antiviral activity in the absence of cytotoxicity (data not shown). Further, low, non-cytotoxic concentrations of siRNAs targeting RRM1 did not significantly alter the potency of 5-aza-C against HIV-1 (data not shown). As an alternative strategy, the potency of 5-aza-C was determined in combination with four different RNR inhibitors—clofarabine, gemcitabine, hydroxyurea, and resveratrol—that inhibit RNR through a variety of different mechanisms (296, 300-304). Multiple RNR inhibitors were utilized in these experiments to exclude the possibility of off-target effects that might alter the potency of 5-aza-C. In these combination experiments, cells were pre-incubated for 2 h with a fixed concentration of each RNRI that exhibited little antiviral activity or cytotoxicity on its own (Figure S5-1), followed by the addition of varying concentrations of 5-aza-C. Contrary to initial predictions that RNR inhibitors would antagonize the activity of 5-aza-C, all four RNR inhibitors significantly potentiated the activity of 5-aza-C against HIV-1 (Figure 5-1A). RNR inhibitors reduced the EC₅₀ of 5-aza-C from 136 μ M to 10.0-53.5 μ M (Table 5-1), corresponding to relative increases in potency of 2.5 to 13.6-fold. We also examined the effect of RNR inhibitors on the

cytotoxicity of 5-aza-C, although RNR inhibitors alone did not exhibit significant cytotoxicity at the concentrations used in combination experiments (Figure S5-1). Most of the RNR inhibitors (all except gemcitabine) resulted in small but statistically significant increases in 5-aza-C cytotoxicity (Figure 5-1B), with relative decreases in the CC_{50} ranging from 1.5 to 2.7-fold (Table 5-1). However, all of the RNR inhibitors increased the distance between the antiviral activity and cytotoxicity of 5-aza-C, ultimately improving the selectivity index (CC_{50}/EC_{50}) of 5-aza-C by 1.9 to 5.2-fold (Table 5-1) and demonstrating that the antiviral activity of the combinations cannot be attributed to cytotoxicity. Notably, the combination of 5-aza-C with the natural product resveratrol resulted in the greatest improvement in antiviral selectivity index.

Combination index analysis was performed in order to determine whether the combinations of 5-aza-C and RNR inhibitors exhibited additive or synergistic antiviral activity. The combination index (CI) is a simple quantitative measure of whether two drugs interact in an additive ($CI = 1$), antagonistic ($CI > 1$) or synergistic ($CI < 1$) fashion. The CI is calculated by first normalizing the drug concentration that produces a particular effect alone (e.g. EC_{50}) to the drug concentration that produces the same effect in a combination, followed by taking the sum of the values for the two drugs (see Materials and Methods). For the combinations of 5-aza-C and RNR inhibitors, CIs were calculated at the EC_{50} , EC_{75} , and EC_{90} (levels relevant to antiviral activity), and the resulting values were then averaged. Combinations of 5-aza-C and RNR inhibitors resulted in CI values ranging from ~0.33 to 0.55, indicative of moderate synergy (Figure 5-2). Thus, 5-aza-C and RNR inhibitors interacted in a synergistic fashion to reduce HIV-1 infectivity, in stark contrast to the initial prediction of antagonism.

5-azacytidine and resveratrol exhibit antiviral synergy in a pattern similar to that of 5-aza-2'-deoxycytidine and resveratrol. In the initial combination experiments, RNR inhibitors may not have antagonized the antiviral activity of 5-aza-C because their concentrations or pre-incubation times were insufficient. To address this, additional combination experiments were performed in which the

concentrations of both drugs were varied, and the resulting data were analysed by isobologram analysis, another method commonly used for the analysis of drug interactions (305, 306). These experiments were performed for the combination of 5-aza-C and resveratrol, as this combination demonstrated the best selectivity index (Table 5-1). Further, resveratrol is not a nucleoside analog (unlike gemcitabine and clofarabine), reducing the likelihood of off-target metabolic interactions with 5-aza-C. Isobologram analysis was performed by pre-incubating cells with varying concentrations of resveratrol (0 to 300 μ M) for 2 h and then adding 5-aza-C (0 to 300 μ M), resulting in a matrix of 80 different treatments. Isobologram analysis was also performed for the combination of resveratrol with 5-aza-dC divalrate, a prodrug of 5-aza-dC with improved stability and permeability (hereafter referred to simply as 5-aza-dC) (266). The combination of resveratrol and 5-aza-dC has previously been shown to synergistically reduce the infectivity of HIV-1 (201). 5-aza-C and resveratrol exhibited significant antiviral synergy across multiple resveratrol concentrations, most notably at 25, 50, or 100 μ M resveratrol (Figure 5-3). Peak antiviral synergy was observed at 50 μ M of each drug. Notably, even at higher concentrations (150 to 300 μ M), resveratrol did not significantly antagonize the activity of 5-aza-C. However, resveratrol alone exhibits substantial antiviral activity at these concentrations (120, 201), such that it would be statistically difficult to observe antagonism or synergy in these combinations. The pattern of synergy between 5-aza-C and resveratrol closely mirrored that of 5-aza-dC and resveratrol (Figure 5-3), even though 5-aza-dC does not need to be reduced by RNR in order to decrease HIV-1 infectivity. For this combination, peak antiviral synergy was observed at 25 μ M resveratrol and 1 μ M 5-aza-dC. While multiple concentrations of resveratrol potentiated 5-aza-C, it remains possible that 5-aza-C was simply reduced to 5-aza-dC by RNR faster than resveratrol was able to inhibit RNR, despite the addition of resveratrol 2 h before 5-aza-C in these experiments. To address this, the potency of 5-aza-C was determined in

combination with 50 μ M resveratrol using varying resveratrol pre-incubation times. The duration of pre-incubation time (2, 4, 6, or 8 h) did not significantly alter the potency of resveratrol alone or the potency of 5-aza-C in combination with resveratrol (Figure S5-2). Thus, 5-aza-C and resveratrol exhibited synergy against HIV-1 under a wide variety of experimental conditions, and, surprisingly, the overall synergy pattern of 5-aza-C and resveratrol closely matched that of 5-aza-dC and resveratrol.

Ribonucleotide reductase inhibitors deplete the endogenous dNTP pools differentially. Other studies have demonstrated that 5-aza-C acts primarily as 5-aza-dC against HIV-1 during reverse transcription (Rawson J, et al. Submitted). Thus, in 5-aza-C and RNR inhibitor combinations, RNR inhibitors were hypothesized to either: 1. not effectively inhibit reduction of 5-aza-C to 5-aza-dC, or 2. deplete endogenous dNTPs enough to allow for significant incorporation of 5-aza-CTP by HIV-1 RT. In order to distinguish between these possibilities, the effects of all individual drugs (RNR inhibitors as well as 5-aza-C and 5-aza-dC) on endogenous dNTP pools were examined using LC-MS/MS. In these experiments, 5-aza-C and 5-aza-dC were analysed at an EC₇₅ concentration, while RNR inhibitors were analysed at two concentrations: the low (L) concentrations correspond to levels with little antiviral activity and were used in earlier combination experiments (Figures 5-1 & 5-2), while the high (H) concentrations exhibit significant antiviral activity (> EC₇₅) but were not utilized in combination experiments (due to the difficulty in observing statistically significant drug interactions at high effect levels). Unexpectedly, 5-aza-C and 5-aza-dC significantly depleted the endogenous dCTP pool, resulting in decreases of ~50-55% relative to the no drug control (Figure 5-4). This depletion would presumably result in self-potential, as sequencing data suggests that 5-aza-dCTP is primarily incorporated in place of dCTP during reverse transcription (117, 275). RNR inhibitors exhibited differential effects on dNTP pools depending upon the specific inhibitor, concentration, and dNTP (Figure 5-4). RNR inhibitors reduced the dATP pool much more effectively than other dNTP

pools, consistent with several previous reports and possibly due to inefficient formation of dATP by nucleotide salvage pathways (see Discussion) (205, 297, 298, 307). RNR inhibitors (except hydroxyurea) also significantly depleted the dGTP and dCTP pools at high concentrations. However, only resveratrol significantly depleted the dTTP pool. In previous reports, RNR inhibitors have been found in some cases to result in either unaltered or even increased dTTP pools due to the upregulation of multiple salvage pathways (see Discussion) (302, 308, 309). Overall, RNR inhibitors depleted endogenous dNTP pools in the order of dATP > dCTP = dGTP > dTTP. Low concentrations of RNR inhibitors did not significantly deplete dTTP, dGTP, or dCTP, but did significantly deplete levels of dATP (except for hydroxyurea), with decreases ranging from ~30 to 50%. Interestingly, these results suggest that low concentrations of RNR inhibitors might not effectively inhibit reduction of 5-aza-C to 5-aza-dC, as the dCTP pool was not depleted.

Low concentrations of ribonucleotide reductase inhibitors do not reduce the level of 5-aza-2'-deoxycytidine-triphosphate in cells treated with 5-azacytidine.

In order to determine whether RNR inhibitors prevented the reduction of 5-aza-C to 5-aza-dC, cells were treated with combinations of RNR inhibitors (at low or high concentrations) and 5-aza-C or 5-aza-dC (at an EC₇₅ concentration). In these experiments, RNR inhibitors were pre-incubated for 2 h prior to addition of 5-aza-C or 5-aza-dC (as before), and cells were collected 4 h later for analysis by LC-MS/MS. High concentrations of three RNR inhibitors (all except hydroxyurea) significantly depleted 5-aza-dCTP levels in cells treated with 5-aza-C (Figure 5-5A). However, RNR inhibitors at low concentrations did not deplete 5-aza-dCTP levels in cells treated with 5-aza-C, similar to observations for dCTP in cells treated with low concentrations of RNR inhibitors alone (Figure 5-4). In contrast, in cells treated with 5-aza-dC, RNR inhibitors either significantly increased or had no effect on the 5-aza-dCTP levels (Figure 5-5A). In addition, the levels of dRGU-TP in the combinations were measured, as dRGU-TP may contribute to

the antiviral activity of 5-aza-dC. However, the overall trends in dRGU-TP levels were similar to those observed for 5-aza-dCTP, both for the combinations of RNR inhibitors with 5-aza-C (Figure S5-3A) and with 5-aza-dC (Figure S5-3B). Additionally, the levels of dCTP and the ratio of 5-aza-dCTP to dCTP were examined in these drug combinations. Based on sequencing results from previous studies, 5-aza-dCTP likely competes primarily with dCTP for incorporation during reverse transcription (117, 275). Thus, the ratio of 5-aza-dCTP to dCTP may be a more important determinant of antiviral antagonism or synergy than simply the level of 5-aza-dCTP alone. As demonstrated earlier, 5-aza-C and 5-aza-dC both depleted the endogenous dCTP pool alone relative to the no drug control (Figure 5-5B). The addition of RNR inhibitors at high concentrations to cells treated with 5-aza-C resulted in significant further depletions of the dCTP pool relative to 5-aza-C alone. However, low concentrations of RNR inhibitors did not significantly alter the level of dCTP in cells treated with 5-aza-C. Also, most of the RNR inhibitors did not significantly alter the ratio of 5-aza-dCTP to dCTP in cells treated with 5-aza-C (Figure 5-5C). In contrast, in cells treated with 5-aza-dC, RNR inhibitors often resulted in more pronounced depletions of dCTP relative to 5-aza-dC alone (Figure 5-5B) and corresponding increases in the ratio of 5-aza-dCTP to dCTP (Figure 5-5C). Overall, these results help explain the lack of antagonism between RNR inhibitors and 5-aza-C in terms of antiviral activity, as RNR inhibitors (at low concentrations) did not significantly reduce levels of 5-aza-dCTP or the ratio of 5-aza-dCTP to dCTP in cells treated with 5-aza-C. However, they also demonstrate that the mechanism of antiviral synergy between 5-aza-C and RNR inhibitors is complex and cannot simply be attributed to higher levels of 5-aza-dCTP (see Discussion).

Resveratrol does not potentiate 5-azacytidine-mediated increases in the HIV-1 mutation frequency. In order to better determine the mechanism of antiviral synergy in combinations of 5-aza-C and RNR inhibitors, the effects of single drugs and drug combinations on HIV-1 mutation frequencies were

examined using Illumina sequencing. Illumina sequencing was performed only for the combination of 5-aza-C and resveratrol, as resveratrol increased the selectivity index of 5-aza-C to the greatest degree. Further, resveratrol is not a nucleoside analog, reducing the likelihood of off-target interactions. To prepare samples for sequencing, U373-MAGI cells were first treated with DMSO (i.e. no drug), resveratrol (50 μ M) alone, 5-aza-C (EC_{75} , \sim 260 μ M) alone, or 5-aza-C in combination with resveratrol. For comparative purposes, 5-aza-dC (EC_{75} , \sim 3.8 μ M) was also examined, alone or in combination with resveratrol. Cells were infected at an MOI of 1.0 with NL4-3 MIG-VSVG and collected 72 h post-infection for genomic DNA extraction, followed by PCR of multiple amplicons (Gag, Pol, Vif, Env, Nef) from proviral DNA. Plasmid control amplifications were included to measure the level of background error due to PCR and sequencing. Amplicons were pooled, subjected to library preparation, and analysed by paired-end sequencing (2 \times 250) on the Illumina MiSeq. The numbers of read pairs, mutations, and reference bases resulting from Illumina sequencing for each sample and amplicon are listed in Table S5-2. 5-aza-C alone was found to increase the total mutation frequency of HIV-1 from 2.5×10^{-4} mutations/base pair (m/bp) to 1.7×10^{-3} m/bp, a significant difference of 6.8-fold ($p < 0.001$; Figure 5-6A). Consistent with a previous report (118), 5-aza-C induced primarily G-to-C transversions in HIV-1, increasing the G-to-C frequency by \sim 67-fold ($p < 0.001$; Figure 5-6A) and the relative percentage of G-to-C transversions by 60% (Figure 5-6B). 5-aza-C also increased the frequency of C-to-G transversions by \sim 43-fold ($p < 0.001$; Figure 5-6A) and the relative percentage of C-to-G transversions by 10% (Figure 5-6B). Resveratrol alone was found to increase the total mutation frequency of HIV-1 by \sim 1.5-fold (Figure 5-6A), a difference that was statistically significant ($p < 0.001$). Resveratrol significantly increased the frequencies of both transitions ($p = 0.002$; 1.3-fold difference) and transversions ($p < 0.001$; 1.8-fold difference). Resveratrol significantly increased the frequencies of six different types of transversions (C-to-A: 1.3-fold, $p = 0.03$; A-to-

C: 1.8-fold, $p = 0.001$; A-to-T: 1.8-fold, $p < 0.001$; T-to-A: 1.8-fold, $p < 0.001$; G-to-C: 2.0-fold, $p = 0.002$; T-to-G: 3.2-fold, $p < 0.001$). Further, resveratrol increased the frequency of T-to-C transitions by 2.0-fold ($p < 0.001$). Ultimately, resveratrol shifted the mutational spectrum of HIV-1 toward transversions (Figure 5-6B), as transversions were induced to a greater extent than transitions. Overall, these data indicate that resveratrol (at 50 μM) was weakly mutagenic and induced complex changes in the viral mutation spectrum, although higher concentrations of resveratrol could potentially have more pronounced effects. Surprisingly, the combination of resveratrol with 5-aza-C or 5-aza-dC led to decreased total mutation frequencies relative to 5-aza-C or 5-aza-dC alone ($p < 0.001$), although mutation frequencies were still significantly higher than for the no drug control ($p < 0.001$; Figure 5-6A). This effect was observed for both G-to-C and C-to-G transversions ($p < 0.001$; Figure 5-6A), and was further consistently observed across all five amplicons (data not shown). As a result, the combinations exhibited decreased relative percentages of G-to-C and C-to-G transversions relative to 5-aza-C or 5-aza-dC alone (Figure 5-6B). Thus, the ability of resveratrol to potentiate the antiviral activity of 5-aza-C cannot be attributed to induction of synergistic increases in the HIV-1 mutation frequency (see Discussion).

Combinations of 5-azacytidine and resveratrol reduce levels of HIV-1 reverse transcription products. The RNR inhibitor hydroxyurea has previously been demonstrated to inhibit viral DNA synthesis in HIV-1 by depleting endogenous dNTP pools (297, 298). Thus, resveratrol was hypothesized to potentiate the antiviral activity of 5-aza-C by inhibiting viral DNA synthesis rather than by enhancing viral mutagenesis. To address this, U373-MAGI cells were pre-treated with individual drugs or drug combinations, infected, and collected at various time-points after infection (0, 2, 6, 12, or 72 h post-infection). Genomic DNA was then extracted, and qPCR was performed to measure levels of RT products or cellular genes (for normalization). All drugs were examined at multiple concentrations ranging from little antiviral activity (EC_5 - EC_{15}) to high

antiviral activity ($> EC_{90}$), as determined by flow cytometry. The assay was first validated using two drugs as controls: emtricitabine (FTC), a deoxycytidine analog that inhibits RT by chain termination, and raltegravir (RAL), an integrase inhibitor. Emtricitabine was found to significantly reduce the level of late RT product (LRT; U5-gag) at 6, 12, and 72 h in a concentration-dependent manner (Figure 5-7A). In contrast, raltegravir did not significantly reduce the level of LRT product at any concentration at 6 or 12 h post-infection. At 72 h post-infection, raltegravir significantly reduced the level of LRT product (by ~40%) at 50 or 100 nM, consistent with previous observations by others (310) and likely due to the progressive loss of unintegrated viral DNA during cell division (311). Next, we examined the effects of resveratrol, 5-aza-C, and 5-aza-dC individually on LRT product levels (Figure 5-7B). Resveratrol reduced the level of LRT product at 6, 12, and 72 h post-infection, similar to previous observations with another RNRI (hydroxyurea) and likely due to depletion of endogenous dNTP pools (205, 298). Notably, the concentration of resveratrol used in earlier combination experiments (50 μ M) significantly reduced the level of LRT product at 6 h post-infection but not 12 or 72 h post-infection. 5-aza-C and 5-aza-dC also significantly decreased levels of LRT product, although these molecules are thought to act primarily as mutagens against HIV-1 (117, 118). Previous groups have found that the incorporation of 5-aza-dCTP does not directly inhibit DNA synthesis by cellular DNA polymerases (280, 281). However, 5-aza-C and 5-aza-dC could reduce levels of RT products through a variety of indirect mechanisms (see Discussion). Lastly, LRT product levels were analysed in cells treated with 5-aza-C or 5-aza-dC in combination with 50 μ M resveratrol (Figure 5-7C). Resveratrol potentiated the ability of 5-aza-C or 5-aza-dC to reduce levels of LRT product. Additionally, qPCR was performed on the same samples to determine levels of a second RT product (gag, an intermediate product), and similar trends were observed (Figure S5-4), further supporting the conclusion that resveratrol potentiates 5-aza-C and 5-aza-dC-mediated decreases in levels of RT products. Overall, these findings indicate that the primary mechanism of antiviral synergy between resveratrol and

5-aza-C is likely the reduced formation of RT products rather than enhanced mutagenesis.

Discussion

This study has shown that RNR inhibitors do not antagonize the antiviral activity of 5-aza-C, as originally predicted, but instead synergize with 5-aza-C to reduce the infectivity of HIV-1 (Figures 5-1, 5-2, & 5-3). 5-aza-C exhibited antiviral synergy with four different RNR inhibitors, which act by different mechanisms and include both nucleoside analogs and non-nucleoside analogs (296), indicating that synergy was not due to potential off-target effects of RNR inhibitors. While initial combination experiments used a single, fixed concentration and pre-incubation time for each RNR inhibitor, additional experiments with 5-aza-C and resveratrol demonstrated synergy under a wider array of conditions (Figure 5-3), including multiple resveratrol concentrations (25, 50, or 100 μ M) and pre-incubation times (Figure S5-2). Further, antiviral synergy was demonstrated using multiple methods for the analysis of drug-drug interactions (i.e. combination index analysis and isobologram analysis). RNR inhibitors ultimately improved the selectivity index of 5-aza-C, demonstrating that antiviral synergy could not be attributed to cytotoxicity, with the addition of resveratrol improving the selectivity index of 5-aza-C to the greatest degree (Table 5-1). These findings closely match the antiviral synergy observed when combining 5-aza-dC and RNR inhibitors, observed both here (Figure 5-3) and in previous reports (117, 201). However, these findings are quite surprising considering that 5-aza-C acts primarily as 5-aza-dCTP in the absence of RNR inhibitors (Rawson J, *et al. Submitted*), implying that 5-aza-C must be reduced in order to effectively inhibit viral replication.

Low concentrations of RNR inhibitors did not significantly reduce levels of 5-aza-dCTP or the ratio of 5-aza-dCTP to dCTP in cells treated with 5-aza-C (Figure 5-5), indicating that 5-aza-C likely acted predominantly as 5-aza-dCTP in synergistic drug combinations. In fact, even high concentrations of RNR

inhibitors (defined as $> EC_{75}$ in terms of antiviral activity alone) only depleted 5-aza-dCTP levels by ~60% in cells treated with 5-aza-C, suggesting that RNR inhibitors can exert high levels of antiviral activity without completely inhibiting the activity of RNR. In terms of antiviral activity, the effects of high concentrations of one particular RNRI (resveratrol) on the potency of 5-aza-C were examined, but only additive effects were observed (Figure 5-3). However, resveratrol (as well as the other RNR inhibitors examined here) independently exhibit strong antiviral activity (due to depletion of endogenous dNTP pools) at high concentrations, such that drug interactions would be statistically difficult to detect. High concentrations of RNR inhibitors could potentially antagonize the activity of 5-aza-C by inhibiting the reduction of 5-aza-C to 5-aza-dC or, alternatively, could exhibit additive or synergistic effects by promoting significant direct incorporation of 5-aza-CTP during reverse transcription. Indeed, previous work has shown that HIV-1 RT can incorporate significant levels of ribonucleotides (both endogenous ribonucleotides as well as chain-terminating analogs) in macrophages (219, 220), where low levels of intracellular dNTPs lead to much higher rNTP:dNTP ratios than in activated CD4+ T-cells (219, 282). The treatment of rapidly dividing immortalized cell lines (which contain high levels of dNTPs) with high levels of RNR inhibitors could potentially mimic conditions within macrophages, and it would be of great interest to investigate 5-aza-C activity and mechanism of action in macrophages in the future.

Intriguingly, both 5-aza-C and 5-aza-dC resulted in reproducible decreases (of ~50-55%) in the intracellular dCTP pool (Figure 5-4), which to our knowledge has not been previously reported and would be expected to result in self-potential. While the mechanism for this decrease is not yet clear, it could be due to competition for cellular enzymes involved in the production of dCTP (e.g. RNR or deoxycytidine kinase) and/or feedback inhibition mechanisms. Consistent with the findings reported in this study, one previous group found that the anti-HIV-1 drug 2',3'-dideoxycytidine (ddC) also depleted the endogenous dCTP pool (297). RNR inhibitors clearly depleted endogenous dNTP pools to

varying extents as well, with decreases following a trend of dATP > dGTP = dCTP > dTTP. In most previous reports, RNR inhibitors have been found to deplete dATP the most effectively (205, 297, 298, 302, 307), possibly reflecting the inefficiency of the dATP salvage pathway (298), although others have found the dGTP or dCTP pool to be the most affected (301, 303). These observations have been used to explain the finding that RNR inhibitors exhibit stronger anti-HIV-1 synergy when combined with chain-terminating deoxyadenosine analogs (such as didanosine and tenofovir) than other types of nucleoside analogs (203, 214, 297). However, there are not yet any reported mutagenic deoxyadenosine analogs with anti-HIV-1 activity. RNR inhibitors were found to have little effect on the dTTP pool, which has also been observed by others and has been demonstrated in one case to result from the salvage of dTTP from dUMP and dCMP (302, 308, 309).

The mechanism of antiviral synergy in combinations of 5-aza-C and RNR inhibitors cannot be attributed simply to an increase in the amount of 5-aza-dCTP relative to dCTP, as most RNR inhibitors did not significantly alter this ratio (Figure 5-5). However, low concentrations of most RNR inhibitors did significantly deplete the dATP pool (by ~30-50%) (Figure 5-5), which could contribute to antiviral synergy in the combinations. In contrast, RNR inhibitors often significantly increased the ratio of 5-aza-dCTP to dCTP in cells treated with 5-aza-dC, which could potentially enhance the incorporation of 5-aza-dCTP and lead to antiviral synergy (Figure 5-5). The mechanism of antiviral synergy for the combination of 5-aza-C and resveratrol was further examined by Illumina sequencing and qPCR, with 5-aza-dC and resveratrol included for comparative purposes. 5-aza-C and 5-aza-dC were found to induce primarily G-to-C and, to a lesser extent, C-to-G transversions in HIV-1 (Figure 5-6). Resveratrol caused a mild increase in the total mutation frequency of HIV-1, resulting from small increases in several different types of substitutions. While resveratrol was hypothesized to increase the frequency of 5-aza-C-mediated transversions, the combinations were found to exhibit decreased mutation frequencies relative to 5-

aza-C alone, both for G-to-C and C-to-G transversions. This data indicates that the antiviral synergy between resveratrol and 5-aza-C cannot be attributed to enhanced viral mutagenesis, suggesting that another mechanism was responsible for antiviral synergy.

The mechanism of antiviral synergy between resveratrol and 5-aza-C was further addressed by qPCR to measure levels of intermediate and late RT products (Figures 5-7 and S5-4). Resveratrol depleted levels of RT products, similar to previous observations with hydroxyurea (205, 298). 5-aza-C and 5-aza-dC also resulted in decreased levels of RT products, although previous groups have found that 5-aza-dCTP does not directly inhibit DNA synthesis by cellular DNA polymerases (280, 281). However, 5-aza-C and 5-aza-dC could reduce levels of RT products through a variety of indirect mechanisms. 5-aza-dC has been found by multiple groups to induce alkali-labile sites (ALS) in DNA, which have been proposed to result from either the spontaneous hydrolysis of 5-aza-dC or from the removal of 5-aza-dC by glycosylases, resulting in abasic sites (281, 312, 313). ALS can ultimately lead to both single-stranded and double-stranded breaks, which could in turn decrease detected levels of RT products. Alternatively, sequences mutagenized by 5-aza-dC may not amplify well during PCR due to mutations in the template sequence that interfere with annealing of primers. However, decreased levels of RT products were observed even when using primers with mixed bases designed to better amplify mutagenized sequences (Figure S5-4). It is also possible that internal, unrepaired 5-aza-C (or hydrolysis products) interferes with DNA polymerization (on the opposite strand) during PCR. Notably, another group previously found indirect evidence suggesting that the final hydrolysis product of 5-aza-dC (dRGU) acts as a replication block to Taq polymerase (270). Further studies will be required to fully elucidate the mechanism by which 5-aza-C and 5-aza-dC result in decreased levels of RT products. Notably, resveratrol potentiated 5-aza-C-mediated decreases in levels of RT products (Figure 5-7), indicating that the

primary mechanism of antiviral synergy in the combinations was the reduction of RT products.

In total, these observations demonstrate that low concentrations of RNR inhibitors potentiate the antiviral activity of the ribonucleoside analog 5-aza-C, even though 5-aza-C acts primarily as a deoxyribonucleotide against HIV-1 in the absence of RNR inhibitors (Rawson J, *et al. Submitted*). Further studies will be required to determine whether RNR inhibitors are able to potentiate the antiretroviral activity of other mutagenic and/or chain-terminating ribonucleoside analogs as well. It would also be of great interest to investigate the activities of RNR inhibitors and 5-aza-C (individually or in combination) in cell types with naturally low dNTP pool levels, such as macrophages. 5-aza-C may be directly incorporated as a ribonucleotide at significant levels in macrophages. Even in this case, RNR inhibitors would likely still potentiate the activity of 5-aza-C, as further depletion of endogenous dNTP pools might promote increased incorporation of 5-aza-CTP. In addition, it would be of interest to determine the effects of RNR inhibitors on the potency of antiviral ribonucleoside analogs that cannot undergo reduction (i.e. those lacking a 3'-hydroxyl group, such as 3'-deoxyadenosine). These types of ribonucleoside analogs have been proposed as a means to specifically target the replication of HIV-1 in macrophages (an important viral reservoir), as they are not active in proliferating CD4+ T-cells (219). Together, the observations here further support the use of anti-metabolites such as RNR inhibitors to potentiate both ribonucleoside and deoxyribonucleoside analogs, ultimately opening additional avenues for the design of synergistic drug combinations directed against HIV-1 and other viruses.

Materials and Methods

Plasmids, cell lines, and reagents. For single-cycle infections, viral stocks were produced using the HIV-1 envelope-deficient vector pNL4-3 MIG (120), which expresses mCherry and EGFP reporter proteins upon infection. Viral stocks were pseudotyped with VSV-G expressed from pHCMV-G, a kind gift from

J. Burns (University of California, San Diego). The human embryonic kidney (HEK 293T) cells were purchased from American Type Culture Collection (Manassas, VA) and maintained in Dulbecco's Modified Eagle's Medium (DMEM) from Mediatech, Inc. (Manassas, VA) with 10% HyClone FetalClone III (FC3) and 1% penicillin/streptomycin from Thermo Fisher Scientific, Inc. (Waltham, MA). U373-MAGI-CXCR4_{CEM} cells were obtained from Michael Emerman through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (257). U373-MAGI cells were maintained in 293T medium supplemented with 1.0 µg/mL puromycin, 0.1 mg/mL hygromycin B, and 0.2 mg/mL G-418. Antiviral drugs were obtained from Sigma-Aldrich (5-aza-C, hydroxyurea, and resveratrol; St. Louis, MO), Carbosynth (clofarabine and gemcitabine; Compton, UK), Selleck Chemicals (raltegravir; Houston, TX), or the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (emtricitabine). 5-aza-dC divalate, a more stable prodrug form of 5-aza-dC, was synthesized by the Center for Drug Design at the University of Minnesota, as previously described (266). 5-azacytidine-5'-triphosphate (5-aza-CTP) and 5-aza-2'-deoxycytidine-5'-triphosphate (5-aza-dCTP) were used as standards for LC-MS/MS and were provided by American Advanced Scientific (College Station, TX) and Jena Bioscience (Jena, Germany), respectively. All drugs were dissolved in the appropriate solvent and stored in aliquots at -20 °C. For transfections, poly-L-lysine was from Newcomer Supply (Middleton, WI), and polyethylenimine (PEI) was from Polysciences, Inc. (Warrington, PA).

Production and titration of viral stocks. HIV-1 viral stocks for single-cycle infections were produced by co-transfecting 10 µg of pNL4-3 MIG and 1 µg of pHCMV-G per 10 cm plate of 293T cells. Transfections were performed using the PEI method, as previously described (120). Viral supernatants were collected ~48 hours (h) post-transfection and frozen in aliquots at -80 °C. For qPCR or Illumina sequencing experiments, viral stocks were treated with 10 U/mL of DNase I (New England Biolabs; Ipswich, MA) for 2 h at 37 °C to degrade residual plasmid DNA from transfections prior to freezing at -80 °C. Viral stocks

were titrated by infecting U373-MAGI cells (31,250 cells/well in 24-well plates) with varying amounts of virus ranging from 1.25 to 40 μ L. The medium was replaced 24 h post-infection, and cells were collected 72 h post-infection for analysis by flow cytometry. The infectious titer was calculated from the flow cytometry data by adding all positive quadrants (mCherry+ only, EGFP+ only, and mCherry+/EGFP+) to determine infectivity and then plotting the volume of virus against infectivity.

Antiviral infectivity assays. All drug treatments were performed using U373-MAGI cells in 24-well plates (31,250 cells/well). For single drug treatments, 5-aza-C and 5-aza-dC were added 2 h before infection, while RNR inhibitors (clofarabine, gemcitabine, hydroxyurea, and resveratrol) were added 4 h before infection. For combination drug treatments, these times remained unchanged, such that RNR inhibitors were added 2 h before 5-aza-C or 5-aza-dC. Most combination drug treatments were performed using a fixed concentration of RNRI (clofarabine-50 nM, gemcitabine-2.5 nM, hydroxyurea-500 μ M, resveratrol-50 μ M) and varying concentrations of 5-aza-C or 5-aza-dC. Drugs (1-2 μ L) were added to a final volume of 0.5 mL medium/well, with uninfected cells and infected cells without drug (i.e. DMSO-treated cells) included as controls. At the time of infection, virus and additional media were added to a final volume of 1 mL/well. All indicated drug concentrations are based on the final volume of 1 mL. The amount of virus added was targeted to achieve ~20% infection (for the no drug control) based on the previously determined viral titer. Drugs were removed by replacing the medium 24 h post-infection, and the cells were collected 72 h post-infection for determination of viral infectivity by flow cytometry.

Cell viability analysis. The effects of single drugs or drug combinations on cell viability were assessed using the CellTiter-Glo Luminescent Cell Viability Assay from Promega (Madison, WI). Five thousand U373-MAGI cells/well were plated in 96-well plates the day before drug treatments. The medium was replaced 24 h later and drugs (alone or in combination) were added to a final

volume of 100 μ L/well. Drugs were added with the same timing as for infectivity assays (e.g. RNR inhibitors added 2 h before 5-aza-C). No drug (DMSO only) and no cell wells were included as controls. The medium was replaced 24 h after the addition of 5-aza-C, and the cell viability assay was performed 72 h after the addition of 5-aza-C according to the manufacturer's instructions. The luminescence was recorded using an Orion microplate luminometer from Berthold Detection Systems (Huntsville, AL). The data were analyzed by first subtracting the background signal (the average no cell control value) and then normalizing to the no drug control.

Combination index analysis of drug combinations. The combination indices (CIs) of drug combinations were calculated similarly to previous reports (305, 306, 314). First, the EC₅₀, EC₇₅, and EC₉₀ values for each single drug and drug combination were determined by non-linear regression in GraphPad Prism v 5.0 (GraphPad Software, Inc.; La Jolla, CA). Next, the CI was determined by the following formula: $CI = C_{a,x}/IC_{x,a} + C_{b,x}/IC_{x,b}$, where a and b represent each drug, x represents a particular effect level (e.g. EC₅₀), and C or IC represent the concentration of drug required to produce effect x individually or in a combination, respectively. CIs were determined at the EC₅₀, EC₇₅, and EC₉₀ of each combination (i.e. levels relevant to antiviral activity), with the results averaged for simplicity.

Isobologram analysis of drug combinations. Cells were treated with 5-aza-C or 5-aza-dC alone or in combination with resveratrol as described before. However, in these experiments, the concentrations of both 5-aza-C (or 5-aza-dC) and resveratrol were varied simultaneously. Resveratrol was tested at 10, 25, 50, 100, 150, 200, or 300 μ M. The resulting data were analyzed using the MacSynergy II program (315) to determine the percent inhibition above additivity in drug combinations (i.e. the extent of synergy). The lower limits of the 99% confidence intervals were used to generate isobolograms, such that points above 0% represent areas of statistically significant synergy.

Determination of HIV-1 mutation frequencies by Illumina sequencing. Illumina sequencing was performed as described previously (283), but with several minor modifications. Briefly, 1 million U373-MAGI cells were treated with DMSO or 50 μ M resveratrol at 4 h pre-infection, followed by the addition of DMSO, 5-aza-C (EC₇₅: 258.5 μ M) or 5-aza-dC (EC₇₅: 3.751 μ M) at 2 h pre-infection. Cells were then infected at an MOI of 1.0 with NL4-3 MIG-VSVG, and genomic DNA was purified from cells collected 72 h post-infection. The amount of plasmid carryover from transfections was determined by performing qPCR of the ampicillin resistance gene and of HIV-1 vif (using the Illumina Vif primers) in the samples without drug and was found to be ~0.4% (range 0.1-0.7%). Next, PCR was performed to generate five small (160-170 bp) amplicons (Gag, Pol, Vif, Env, Nef) for each sample, using the primers listed in Table S5-1. Plasmid control amplifications from pNL4-3 MIG were performed in parallel under closely matched conditions to measure background error due to PCR and Illumina sequencing. For each sample, all amplicons were gel-purified and pooled together in an equimolar fashion to normalize coverage between amplicons. Illumina sequencing libraries were then constructed from each sample (21 libraries in total), pooled in an equimolar fashion, and analyzed by 2 \times 250 paired-end sequencing on the Illumina MiSeq. The sequencing reads were analyzed as described before (283), but were demultiplexed based on library indices rather than internal barcodes. As before, background error hotspots (mostly G-to-T and C-to-A transversions) were identified using the plasmid controls and masked prior to mutational analysis. The final sequencing data for this project was also used in separate analyses for another project focused on comparing the mutational patterns of 5-aza-C and 5-aza-dC (Rawson J, et al. Submitted).

Determination of intracellular RT product levels by qPCR. U373-MAGI cells were pre-treated with individual drugs or drug combinations and infected as described before, but were collected at various time points after infection (0, 2, 6, 12, or 72 h post-infection). Uninfected cells and cells infected with heat-treated viruses (i.e. virus stocks that were incubated at 65 °C for 1 h) were included as

negative controls. Viral DNA was extracted by incubating cells at 50 °C for 1 h in 50 µL lysis buffer (50 mM KCl, 10 mM Tris pH 8.3, 1.8 mM MgCl₂, 0.45% IGEPAL CA-630, 0.45% Tween 20, and 0.12 mg/mL proteinase K), followed by heat inactivation of proteinase K at 95 °C for 15 min. Next, qPCR was performed using 4 µL water, 6.25 µL 2X Power SYBR Green Master Mix (Thermo Fisher Scientific, Inc.), 0.625 µL each primer (500 nM final concentration), and 1 µL template. The cycling conditions used were an initial denaturation of 95 °C 10 m, 40 cycles of 95 °C 15 s/55 °C 15 s/72 °C 30 s, and a final extension of 72 °C 7 min. The sequences of the late RT (LRT; U5-gag) and cellular 18S rRNA primers have been previously reported (316, 317), while gag (an intermediate RT product) was amplified using primers containing mixed bases (S = G/C) to promote amplification from templates mutagenized by 5-aza-C or 5-aza-dC: 5'-ACATCAASCASCATSCAAAT-3' (forward) and 5'-GTASTAGTAGTTSSTGSTATG-3' (reverse). The quantity of each target was determined using a plasmid standard curve series (10¹-10⁶ copies/µL, pNL4-3 MIG for LRT or gag, pCR-18S for 18S rRNA), followed by normalization of RT products to the cellular 18S rRNA gene.

Determination of 5-aza-CTP, 5-aza-dCTP, and dNTP levels using LC-MS/MS. To prepare samples for liquid chromatography-tandem mass spectrometry (LC-MS/MS), 1.3 million U373-MAGI cells/ were plated on 10 cm plates, using two plates per treatment group. The following day the medium was replaced and RNR inhibitors were added. 5-aza-C or 5-aza-dC were added 2 h later, and cells were harvested 4 h after 5-aza-C or 5-aza-dC addition (corresponding to the 2 h post-infection time point in earlier antiviral infectivity assays). Cell pellets were then resuspended in 750 µL of 60% methanol (stored at -20 °C) and incubated at -20 °C for ~18 h. The samples were then vortexed, heated at 95 °C for 3 min, and centrifuged at 16,000×g for 5 min. The supernatants were transferred to new microcentrifuge tubes and dried using a Savant SPD1010 SpeedVac Concentrator (Thermo Fisher Scientific, Inc.) at

maximum pressure. Dried samples were frozen at -80 °C until the time of analysis.

Dried extracts were reconstituted in 200 µL of water containing 10 µM of an internal standard, 5-iodo-dCTP. The samples were then centrifuged at 14,000 rpm for 5 min at 4 °C, and the supernatants were subjected to LC-MS/MS. Samples containing 5-aza-CTP or 5-aza-dCTP were reconstituted immediately before each injection into LC-MS/MS to minimize their degradation by hydrolysis. The levels of 5-aza-CTP, 5-aza-dCTP, riboguanylurea-5'-triphosphate (RGU-TP, a hydrolysis product of 5-aza-CTP), 2'-deoxyriboguanylurea-5'-triphosphate (dRGU-TP, a hydrolysis product of 5-aza-dCTP), and the four endogenous dNTPs (dTTP, dGTP, dCTP and dATP) were determined on an LC-MS/MS system following a previous published method (284) with minor modifications. The LC-MS/MS system consisted of an AB Sciex QTrap 5500 mass spectrometer and an Agilent 1260 Infinity HPLC. The chromatographic separation of analytes was achieved using a Thermo Scientific Hypercarb column (100 × 3 mm, 5 µm). The two eluents were: (A) 0.5 % diethylamine in water, pH adjusted to 10 with acetic acid; and (B) 50% acetonitrile in water. The mobile phase was delivered at a flow rate of 0.5 mL/min using stepwise gradients of A and B: 0–20 min, 0-25% B (v/v); 20-28 min, 25-50% B (v/v); 28-28.5 min, 50-95% B (v/v); 28.5-30.5 min, 95-95% B (v/v); 30.5-31 min, 95-0% B, (v/v); 31-39 min, 0-0% B (v/v). Only eluate from 10-30 min was diverted into the mass spectrometer for analysis.

MS/MS detection of the analytes was conducted using an ESI ion source with MRM detection in negative mode. The curtain gas was set at 20 psi. The ionspray voltage was set at -4500 V, and the temperature at 650 °C. The nebulizer gas (GS1) and turbo gas (GS2) were both set at 45 psi. Under these conditions, the two analytes ATP and dGTP, which share the same MRM transition (m/z : 506 → 159), were baseline separated in the chromatogram (data not shown). dTTP, CTP and UTP, which exhibit cross-channel interference with each other, were also well separated (data not shown).

Statistical analyses. All figures were created in Microsoft Office for Mac 2011 v 14.5.2 (Redmond, WA) or GraphPad Prism v 5.0. To determine the potency of single drugs and drug combinations, the infectivity data were normalized to the no drug control, plotted against log-transformed drug concentrations, and subjected to non-linear regression to determine the EC_{50} of each drug or drug combination (with 95% confidence intervals) in GraphPad Prism v 5.0. Likewise, to determine the cytotoxicity of individual drugs and drug combinations, the cell viability data were normalized to the no drug control, plotted against log-transformed drug combinations, and subjected to non-linear regression to determine CC_{50} values. In order to determine whether differences in dNTP pools or RT product levels were statistically significant, the normalized data were analyzed by one-way repeated measures ANOVA (without assuming equal variability of differences) in GraphPad Prism v 5.0. Dunnett's post-test was used to compare drug treatments to either the no drug control (for single drug treatments) or the single drug control (for combination treatments). To test for factors that may impact mutation frequencies, generalized linear mixed effects models were applied to processed Illumina data. The raw counts for each type of mutation were modelled as overdispersed Poisson random variables with an offset given by the total number of reference bases. The type of sample, the type of amplicon, and their interactions were treated as fixed effects, and the replicate was treated as a random effect. The logarithmic link was used, as is standard for Poisson outcomes, and penalized quasiliikelihood was used to estimate the model parameters (263). These calculations were performed using R v 3.1.0 and the MASS package.

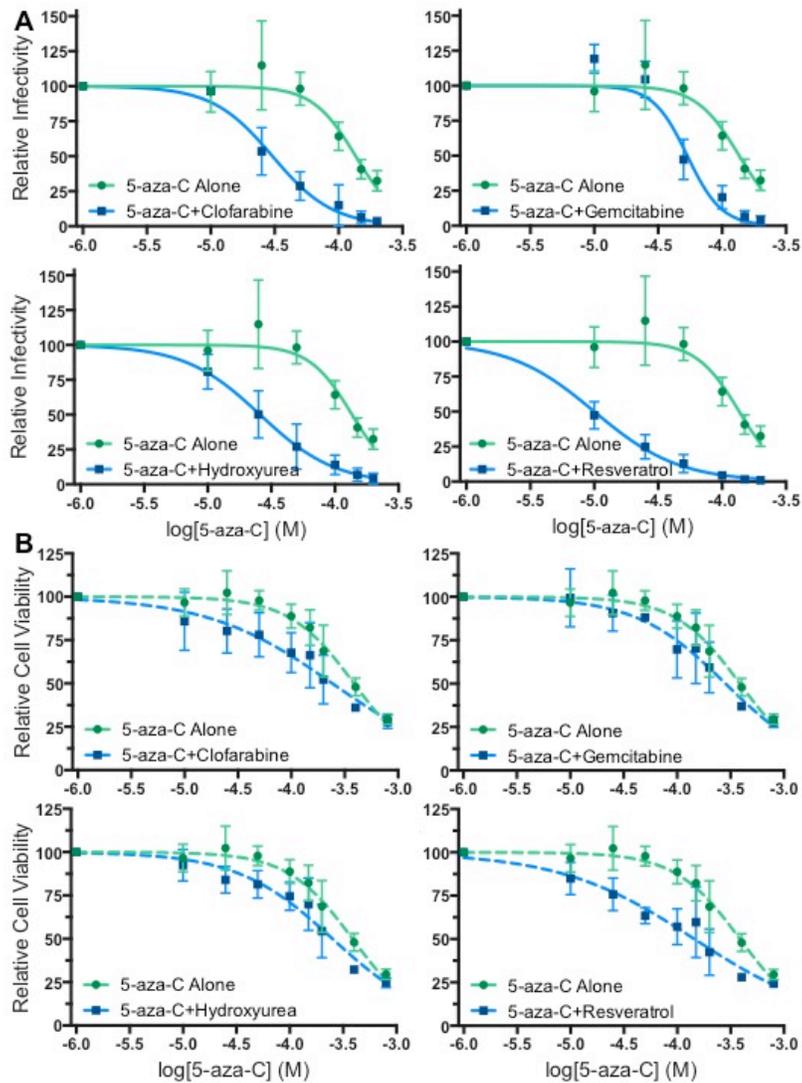


Figure 5-1. Effects of 5-azacytidine and ribonucleotide reductase inhibitor combinations on HIV-1 infectivity and cell viability. A. Cells were pre-treated for 2 h with a fixed concentration of four different ribonucleotide reductase (RNR) inhibitors (clofarabine-50 nM, gemcitabine-2.5 nM, hydroxyurea-500 μ M, resveratrol-50 μ M) that exhibited little antiviral activity or cytotoxicity alone (see Figure S5-1). 5-azacytidine (5-aza-C) was then added at varying concentrations, and 2 h later the cells were infected using a single-cycle HIV-1 vector. Cells were then collected 72 h post-infection for analysis by flow cytometry, allowing for the determination of viral infectivity. B. In parallel, cells were collected for

determination of cell viability using an ATP-based luciferase assay. The data represent the mean \pm s.d. of three independent experiments, normalized to the no drug control.

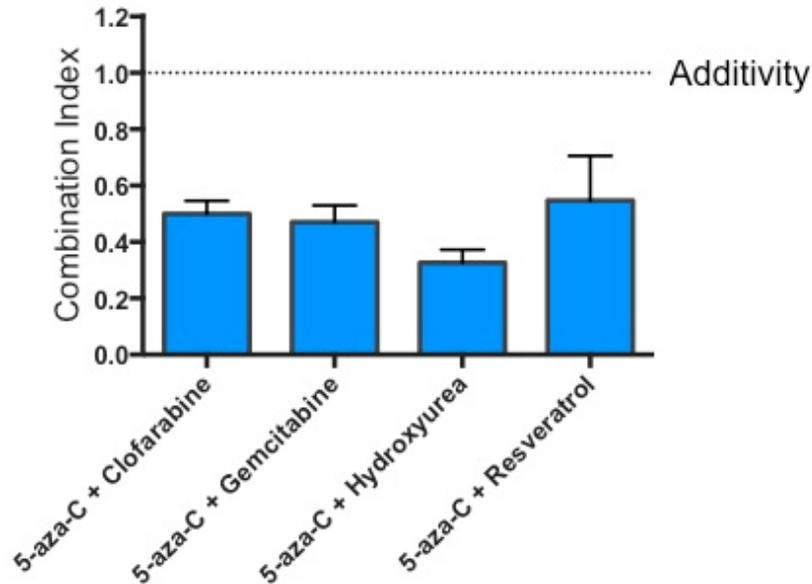


Figure 5-2. 5-azacytidine and ribonucleotide reductase inhibitors exhibit synergy against HIV-1. In order to determine whether combinations of 5-aza-C and RNR inhibitors display additive or synergistic antiviral activity, combination index analysis was performed for each combination. The combination index (CI) was calculated by dividing the drug concentration that produced a desired effect level alone by the drug concentration that produced the same effect level in a combination, with the resulting values for each drug added together (see Materials and Methods). The CI indicates whether the drug combination is additive (CI = 1), antagonistic (CI > 1), or synergistic (CI < 1). The data represent the mean \pm s.d. of the CI values calculated at the EC₅₀, EC₇₅, and EC₉₀ (i.e. levels relevant to antiviral activity), using the infectivity data from Figure 5-1A.

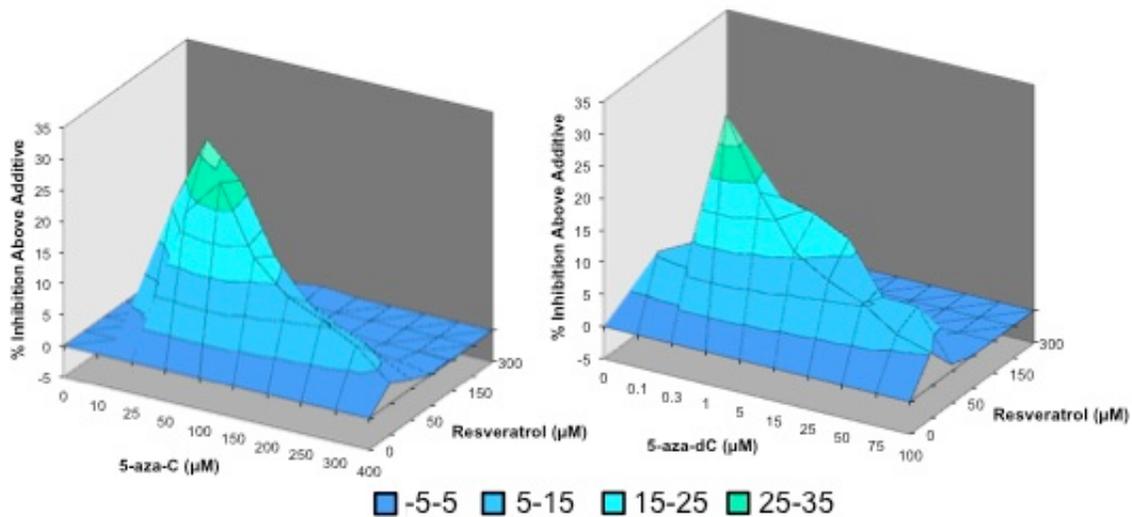


Figure 5-3. 5-azacytidine and resveratrol exhibit antiviral synergy in a pattern similar to that of 5-aza-2'-deoxycytidine and resveratrol. To investigate the extent of synergy between 5-aza-C and resveratrol, combination experiments were performed in which the concentrations of both drugs were varied, and the resulting data were used to generate isobolograms in the MacSynergy II program (315). The combination of 5-aza-2'-deoxycytidine (5-aza-dC) and resveratrol was included for comparative purposes, as this combination has previously been shown to exhibit synergy against HIV-1 (201). Both 5-aza-C and 5-aza-dC synergized most effectively with lower concentrations of resveratrol (25, 50, or 100 μM). The data graphed on the plots represent the lower limits of the 99% confidence intervals of the mean from three independent experiments.

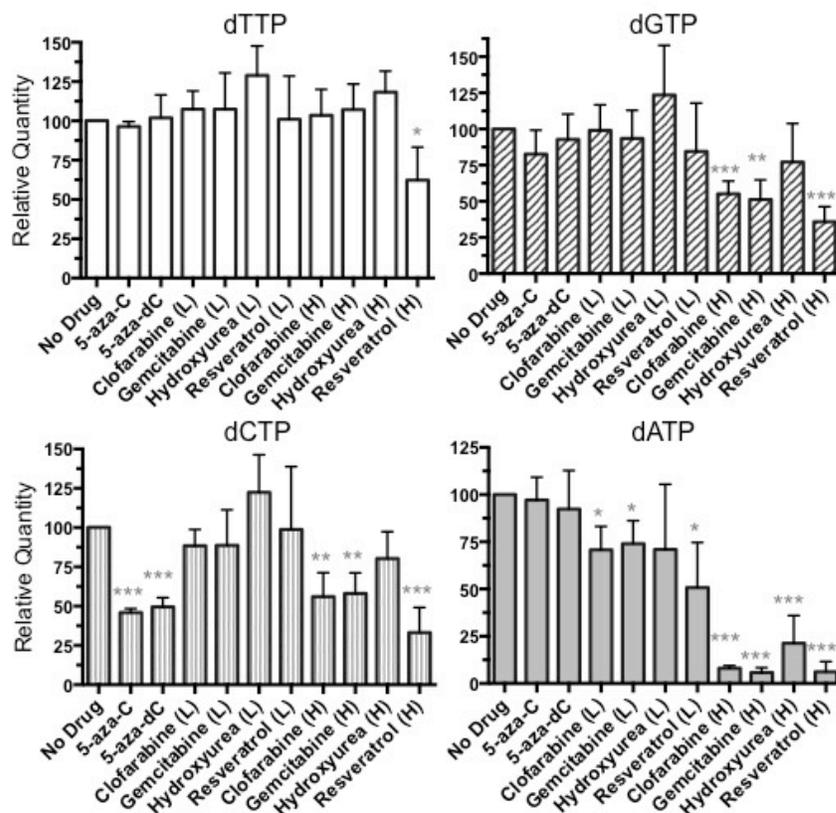


Figure 5-4. Differential depletion of endogenous dNTP pools by ribonucleotide reductase inhibitors. In order to determine the extent to which RNR inhibitors deplete endogenous dNTP pools, cells were incubated with two different concentrations (L = low, H = high) of RNR inhibitors for 6 h (clofarabine: 50 or 200 nM, gemcitabine: 2.5 or 25 nM, hydroxyurea: 0.5 or 2.0 mM, resveratrol: 50 or 200 μ M). Cells were also incubated with 5-aza-C or 5-aza-dC (at an EC₇₅ concentration) for 4 h to determine whether these drugs could alter endogenous dNTP pools. Next, dNTPs were isolated by methanol extraction and subjected to LC-MS/MS analysis to determine the relative levels of dTTP, dGTP, dCTP, and dATP. The data represent the mean \pm s.d. of six independent experiments, normalized to the no drug control; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

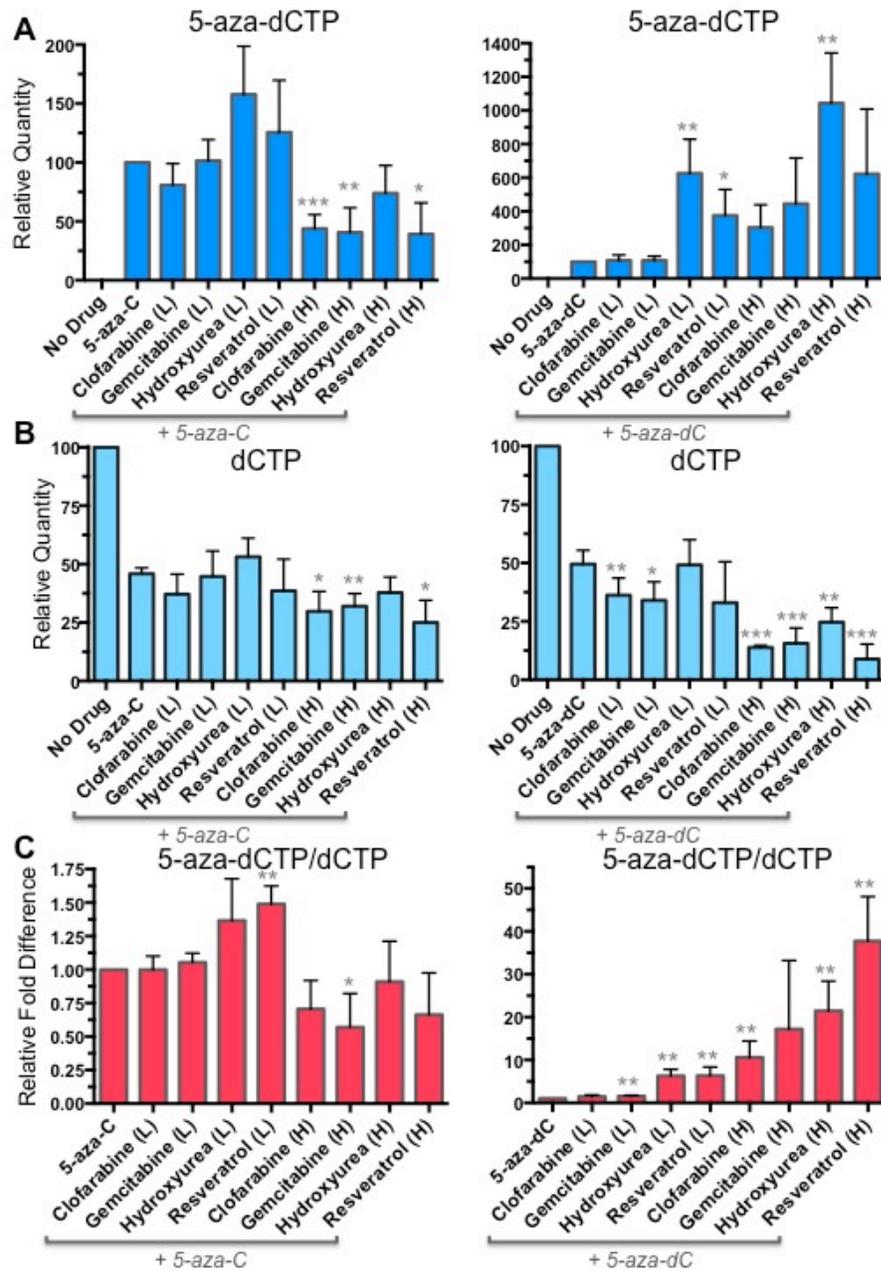


Figure 5-5. Low concentrations of ribonucleotide reductase inhibitors do not decrease the level of 5-aza-2'-deoxycytidine-triphosphate in cells treated with 5-azacytidine. In order to determine the effect of RNR inhibitors on 5-aza-2'-deoxycytidine-5'-triphosphate (5-aza-dCTP) levels in cells treated with

5-aza-C or 5-aza-dC, cells were incubated with two different concentrations (L = low, H = high) of RNR inhibitors for 2 h. Next, 5-aza-C or 5-aza-dC (at an EC₇₅ concentration) was added, and cells were collected 4 h later for dNTP extraction. LC-MS/MS analysis was used to quantify the levels of 5-aza-dCTP (A) and dCTP (B), which in turn were used to determine the ratio of 5-aza-dCTP to dCTP (C). The data represent the mean \pm s.d. of six independent experiments, normalized to either 5-aza-C or 5-aza-dC alone (A & C) or to the no drug control (B). All symbols of statistical significance indicate differences relative to the single drug (5-aza-C or 5-aza-dC) control; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

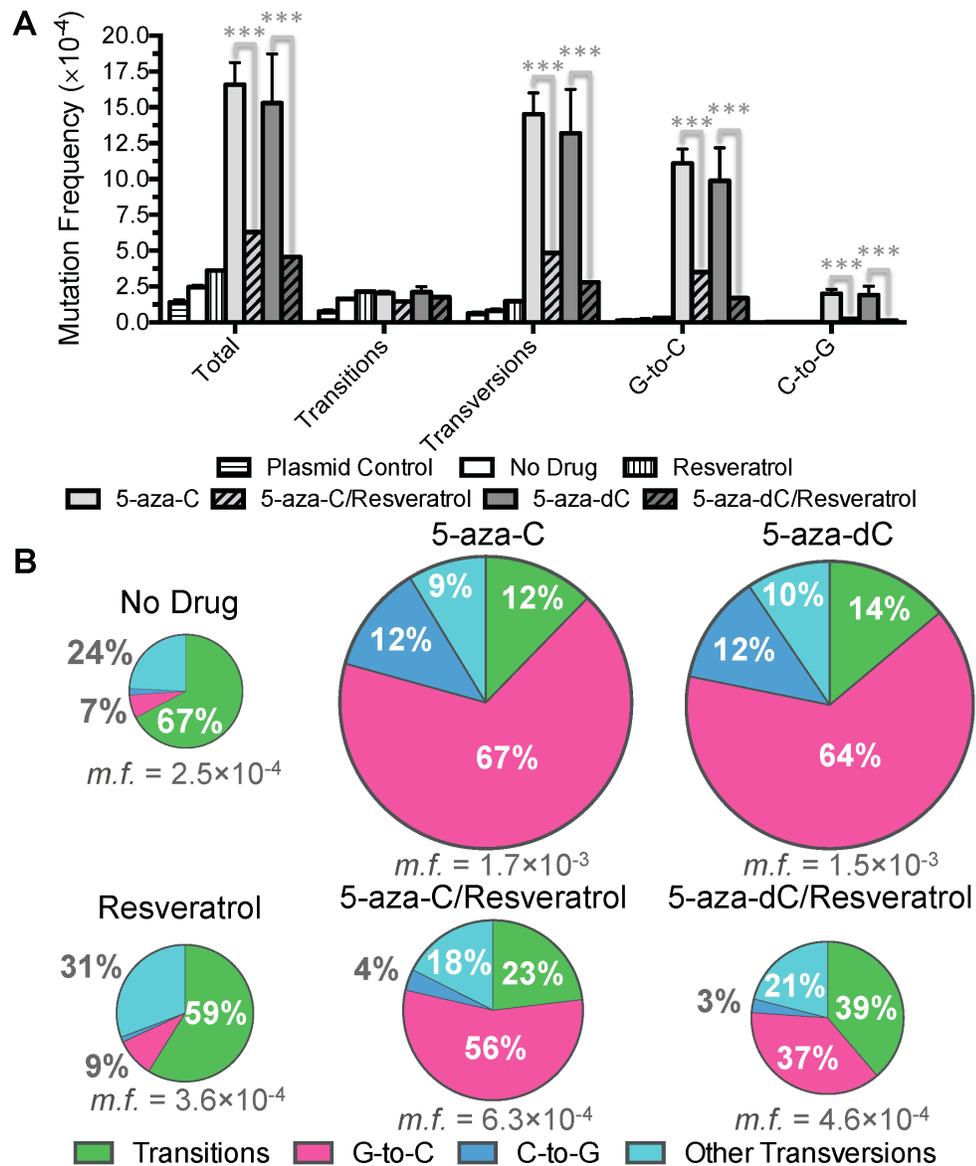


Figure 5-6. Resveratrol does not potentiate 5-azacytidine-mediated increases in the HIV-1 mutation frequency. In order to further investigate the mechanism of synergistic antiviral combinations, U373-MAGI cells were treated with 50 μ M resveratrol (or DMSO), followed by 5-aza-C or 5-aza-dC at an EC_{75} concentration (~ 260 or 3.8μ M, respectively). Cells were then infected at an MOI of 1.0 with NL4-3 MIG-VSVG and collected 72 h post-infection for genomic DNA extraction. Multiple amplicons (Gag, Pol, Vif, Env, Nef) were then prepared from proviral DNA by PCR and subjected to library preparation and paired-end

sequencing (2×250) on the Illumina MiSeq. Plasmid control amplifications were included to determine the levels of background errors resulting from PCR and sequencing. A. Mutation frequency analysis. Mutation frequencies were calculated by dividing the number of mutations by the number of reference bases (mutations + wild-type bases) and are represented as mutations/bp, or m/bp. B. Mutation spectra analysis. Mutation spectra were determined by dividing the frequency of each type of mutation by the total mutation frequency, with the results expressed as a percentage of total mutations. Data in both panels represent the mean of three independent biological replicates, with error bars in panel A indicating standard deviation. All symbols of statistical significance indicate differences of the drug combinations to the single drug (5-aza-C or 5-aza-dC) control; *** $p < 0.001$. The actual numbers of read pairs, mutations, and reference bases are listed in Table S5-2.

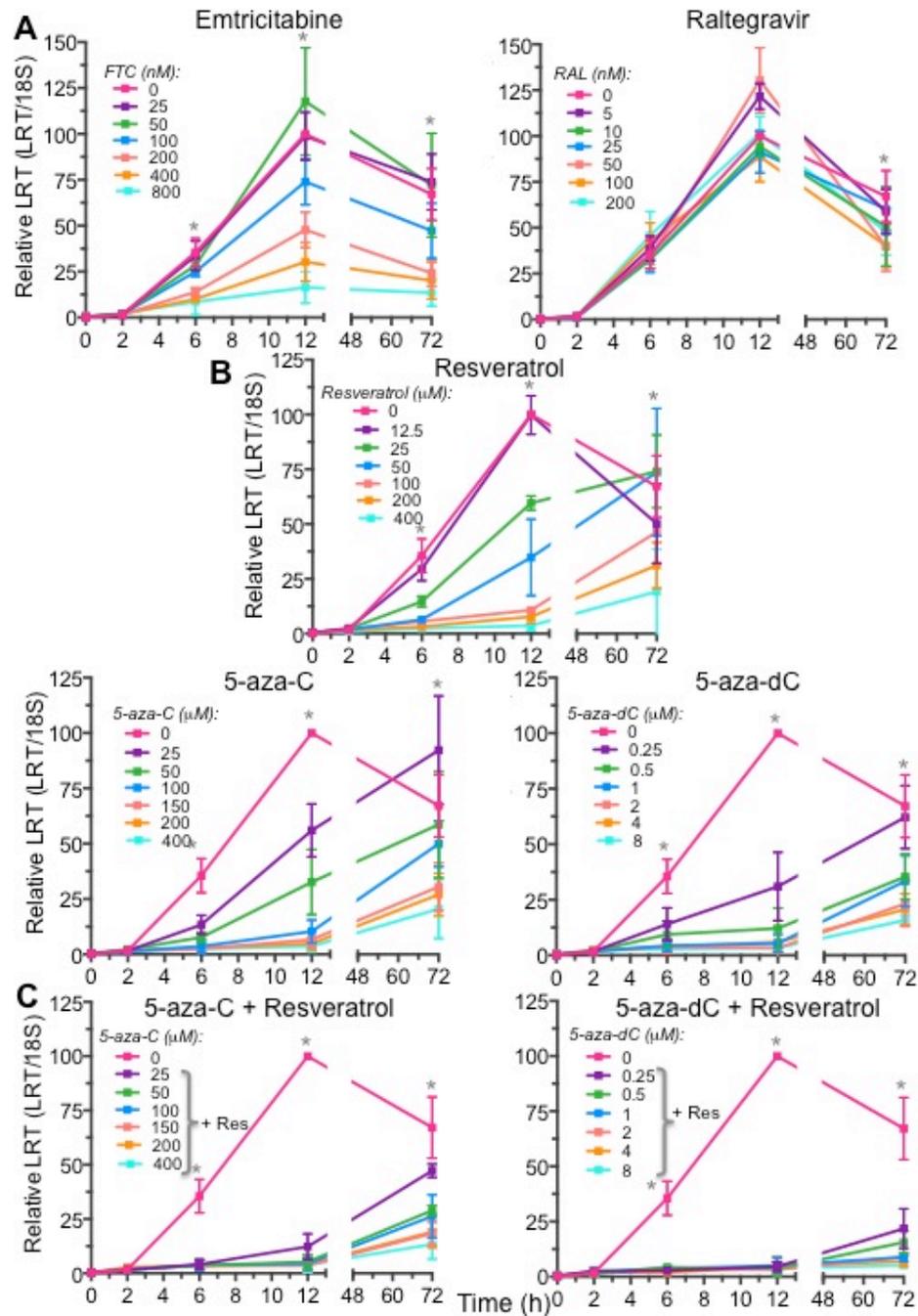


Figure 5-7. Combinations of 5-azacytidine and resveratrol reduce levels of HIV-1 reverse transcription products. In order to determine the effects of single drugs and drug combinations on reverse transcription, cells were pre-incubated with drugs as before, infected with NL4-3 MIG-VSVG, and collected at various time-point after infection (0, 2, 6, 12, or 72 h), for analysis of late reverse

transcription product (LRT, U5-*gag*) by qPCR. All drugs were examined at multiple concentrations ranging from little antiviral activity (EC_5 - EC_{15}) to high antiviral activity ($> EC_{90}$), as determined by flow cytometry. A. The effects of emtricitabine (FTC, a chain-terminating nucleoside analog) and raltegravir (RAL, an integrase inhibitor) on reverse transcription were determined as controls to validate the assay. B. The effects of 5-aza-C, 5-aza-dC, or resveratrol alone on levels of LRT product were determined. C. The effects of 5-aza-C or 5-aza-dC in combination with 50 μ M resveratrol on LRT product were determined. The data represent the mean \pm s.d. of three independent experiments, normalized to the quantity of cellular 18S rRNA gene (to account for variations in cell number or extraction efficiency) and set relative to the no drug control at 12 h post-infection. Asterisks indicate that the drug significantly reduced ($p < 0.05$) the level of late RT product at one or more concentrations relative to the no drug control.

Table 5-1. Effects of ribonucleotide reductase inhibitors on the antiviral activity and cytotoxicity of 5-azacytidine.^a

	<i>5-aza-C</i>	<i>5-aza-C</i> <i>+Clofarabine</i>	<i>5-aza-C</i> <i>+Gemcitabine</i>	<i>5-aza-C</i> <i>+Hydroxyurea</i>	<i>5-aza-C</i> <i>+Resveratrol</i>
EC ₅₀ (μM)	136 (115-162)	30.4 (26.0-35.7)	53.5 (45.9-62.3)	25.8 (21.4-31.0)	10.0 (8.6-11.7)
CC ₅₀ (μM)	387 (327-457)	243 (181-325)	284 (225-358)	252 (206-307)	145 (114-185)
SI (CC ₅₀ /EC ₅₀)	2.8	8.0	5.3	9.8	14.5

a. The antiviral activity and cytotoxicity of 5-azacytidine (5-aza-C) was determined in the absence or presence of ribonucleotide reductase inhibitors. The concentrations of 5-aza-C required to reduce viral infectivity or cell viability by 50% (i.e. the EC₅₀ or CC₅₀, respectively) are indicated, as determined from non-linear regression of the data presented in Figure 5-1. The data represent the mean of three independent experiments, with 95% confidence intervals indicated in parentheses. The selectivity index (SI) for each treatment was determined by dividing the CC₅₀ by the EC₅₀.

CHAPTER VI
5,6-DIHYDRO-5-AZA-2'-DEOXYCYTIDINE POTENTIATES THE ANTI-HIV-1
ACTIVITY OF RIBONUCLEOTIDE REDUCTASE INHIBITORS

Reprinted with permission from: Rawson JM, Heineman RH, Beach LB, Martin JL, Schnettler EK, Dapp MJ, Patterson SE, Mansky LM. 5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors. *Bioorg Med Chem* (2013) 21/22, 7222-8. Copyright © 2013, Elsevier Ltd.

Introduction

Human immunodeficiency virus type-1 (HIV-1) infects about 34 million individuals globally and resulted in the deaths of 1.7 million in 2011 alone (www.unaids.org). Despite great strides in antiretroviral therapy, current drugs are still limited by toxicity, cross-resistance, and the transmission of drug-resistant viral variants (318). HIV-1 has a high mutation rate (3 to 9×10^{-5} mutations/base pair/cycle) (53, 59, 319), and it has been postulated to replicate near the error threshold—a theoretical upper limit to the mutation rate beyond which genetic information cannot be maintained (120, 320, 321). Thus, it is conceivable to exceed the error threshold with small molecule viral mutagens. This antiviral strategy, called lethal mutagenesis, was first investigated using mutagenic ribonucleoside analogs and RNA-damaging agents in poliovirus and vesicular stomatitis virus by Holland and colleagues in the 1990s (178, 322). It was later demonstrated in HIV-1 using 5-hydroxy-2'-deoxycytidine (119), and subsequently explored in several other RNA viruses (323-326).

HIV-1 lethal mutagenesis has not been clinically exploited, but a prodrug of 5,6-dihydro-5-aza-2'-deoxycytidine (KP-1212) has advanced to clinical trials. KP-1212 potently inhibited HIV-1 replication in culture (EC_{50} of ~ 10 nM) with low cellular, mitochondrial, and genome toxicities, as well as activity against both wildtype (wt) and drug resistant strains (198). KP-1212 was shown to induce primarily G-to-A and A-to-G (and to a lesser extent T-to-C and C-to-T) transition mutations within HIV-1. The prodrug of KP-1212 (KP-1461) was well tolerated in clinical trials, but was not effective in reducing viral loads despite adequate drug bioavailability (199, 221). Nonetheless, there was a detectable increase in the number of private mutations (i.e., those most likely to have occurred in the latest round of replication) and evidence of an altered mutation spectrum in clinical samples (200).

We have previously demonstrated that combinations of nucleoside analogs that act as viral mutagens and ribonucleotide reductase inhibitors (RNRIs) can result in synergistic decreases in viral infectivity (117, 201).

Similarly, combining nucleoside reverse transcriptase inhibitors (NRTIs) with RNRI results in potentiation of anti-HIV-1 activity (203, 297, 327). RNRI likely have a threefold antiretroviral mechanism in such combinations: 1) inhibition of retroviral DNA synthesis *via* depletion of deoxynucleoside triphosphate (dNTP) pools (205, 301, 309, 327); 2) many RNRI do not inhibit synthesis of all four natural dNTPs equally, thus generating dNTP pool imbalances that increase the viral mutation rate (297, 301, 309, 328); and 3) decreases in cellular dNTP concentrations result in the upregulation of nucleotide kinases, leading to more efficient activation of nucleoside analogues such as KP-1212 (117, 201, 203, 297, 327). Given this, the combination of an RNRI with KP-1212 or a KP-1212 prodrug should result in potent inhibition of HIV-1 infectivity.

We report here that while KP-1212 alone had minimal antiviral activity in our single round replication assay, it potentiated the activity of the RNRI gemcitabine and resveratrol. The combination of KP-1212 with resveratrol, in particular, reduced the EC₅₀ by 1.9-fold relative to resveratrol alone (i.e., from 99.6 μ M to 52.8 μ M). The KP-1212-RNRI combinations led to corresponding increases in the viral mutant frequency in the absence of cell cytotoxicity relative to that of RNRI alone. Taken together, these observations represent the first demonstration of a mild anti-HIV-1 mutagen having the ability to potentiate the antiretroviral activity of RNRI. These findings may improve the potential for clinical translation of KP-1212 in the treatment of HIV-1 infection.

Results and Discussion

Development of an mCherry/GFP dual reporter HIV-1 vector. In order to efficiently assess both HIV-1 infectivity and mutant frequency, we constructed a HIV-1 vector expressing two fluorescent proteins—mCherry and the green fluorescence protein (GFP) (Figure 6-1A). The resulting vector expresses all viral proteins except Env and Nef. Infectious vector virus was produced by trans-complementing with the vesicular stomatitis virus G protein (VSV-G) via co-transfection with the vector virus plasmid. This strategy limits virus replication to

a single cycle of replication. The utility of the construct for detecting drug-induced increases in virus mutant frequency was confirmed using 5-azacytidine (5-AZC), a ribonucleoside analog previously shown to have antiretroviral activity against HIV-1 (118). We found that treatment of target cells with 5-AZC reduced HIV-1 infectivity in a concentration-dependent manner, with an EC₅₀ of 53.0 μM, comparable to the previously reported value of 57.0 μM (Figure 6-1B) (118). We also observed a concentration-dependent elevation of the viral mutant frequency to a maximum of 19.8-fold relative to the no drug control, confirming previous findings with 5-AZC (118).

Individual effects of RNRI and KP-1212 on HIV-1 infectivity & mutant frequency. We initially combined KP-1212 at 100 μM with a panel of several anti-metabolites and found that KP-1212 potentiated the antiretroviral activity of gemcitabine (2',2'-difluoro-2'-deoxycytidine, dFdC), resveratrol, and deferoxamine (data not shown). These three compounds have all been previously reported to have RNRI activity as well as antiretroviral activity, though inhibition of ribonucleotide reductase has not been clearly demonstrated to be their primary mechanism of action against HIV-1 (117, 201-203, 300, 301, 329, 330). Resveratrol, in particular, has been reported to interact with a wide variety of cellular targets in exerting its anti-inflammatory, anti-tumorigenic, and antiviral effects (331). We next characterized the individual effects of each compound on HIV-1 infectivity and mutant frequency. We found that gemcitabine, resveratrol, and deferoxamine all reduced HIV-1 infectivity (Figure 6-2A & Table 6-1) and elevated the viral mutant frequency in a concentration-dependent fashion (Figure 6-2B).

We observed that KP-1212 exerted only a mild antiviral effect alone, without concentration dependence in the concentration range of 5 to 200 μM (Figure 6-2A). However, we observed a slight dose-dependent increase in mutant frequency (up to ~1.6-fold) (Figure 6-2B). Our observations using this assay are somewhat different than previous findings of KP-1212 being a potent inhibitor of HIV-1 replication (198). We tested multiple batches of KP-1212 to

rule out the possibility of degradation of our liquid drug stocks (data not shown). Additionally, ^1H NMR and combustion analysis confirmed the purity and stability of KP-1212 in our hands (data not shown). The differences in observed activity of KP-1212 may be due to the previous study using a very low multiplicity of infection (MOI), ranging from 1:1000 to 1:5000 (198). These authors found that higher MOIs led to higher EC_{50} values, and were not able to obtain a reproducible EC_{50} at MOIs of 1:300 or greater. In line with this, we used a much higher MOI (~1:3 to 1:5) and were unable to obtain an EC_{50} . Additionally, the previous study used a different cell line (MT-2) and centrifuged the drug onto cells, potentially altering the uptake and/or phosphorylation of KP-1212. Finally, the virus was likely not limited to a single round of replication, which potentially enhanced the observed effects of KP-1212. Despite the apparent discrepancies in antiretroviral potency, our findings are in relatively good agreement with the Phase II clinical trials on KP-1212, which found no clinically significant decline in viral loads but some evidence of mutagenicity (increases in private mutations and altered mutation spectra) (199, 200).

KP-1212 potentiates the effects of RNRI on HIV-1 infectivity & mutant frequency. To further investigate the interaction between RNRI and KP-1212, we combined a fixed low concentration of each RNRI with varying concentrations of KP-1212 (i.e., ranging from 5 to 200 μM). For the gemcitabine and KP-1212 drug combination, we observed an unexpected trend: low concentrations of KP-1212 (5 to 50 μM) enhanced the antiviral activity of gemcitabine, whereas high concentrations partially (100 μM) or fully (200 μM) antagonized the antiviral activity (Figure 6-3A). Similarly, the corresponding mutant frequencies increased to a maximum of 4.8-fold at 25 μM KP-1212 and then decreased to levels lower than for gemcitabine alone (Figure 6-3B). KP-1212 likely antagonized gemcitabine at high concentrations because they are both deoxycytidine analogs that require transport and metabolic conversion to their active forms. Specifically, they may utilize the same transporter proteins to breach the plasma membrane (either through facilitated diffusion or active transport), and they are

both phosphorylated by deoxycytidine kinase to become active within cells (332-335). Furthermore, the drugs were simultaneously added to cells, and KP-1212 was used at a much higher concentration than gemcitabine (10,000-fold higher for 200 μ M KP-1212). Our observation of potential drug interference could likely be circumvented by 1) using a more potent mutagenic deoxycytidine analog, 2) using a mutagenic nucleoside analog that is not phosphorylated by deoxycytidine kinase, 3) applying the drugs sequentially rather than simultaneously, or 4) using an RNRI that is not a nucleoside analog.

We therefore examined the interactions of KP-1212 with two non-nucleoside RNRIs—resveratrol and deferoxamine. The addition of KP-1212 (5 to 200 μ M) to resveratrol (50 μ M) resulted in a concentration-dependent decline in viral infectivity, with the greatest potentiation observed at 200 μ M KP-1212 (Figure 6-3A). We observed a similar, though less concentration-dependent, trend for the combination of KP-1212 with deferoxamine. For both of these combinations, we also observed increases in the viral mutant frequencies (to a maximum of 6.2-fold for resveratrol and 4.7-fold for deferoxamine relative to that of no drug) (Figure 6-3B). These results indicate, in contrast to gemcitabine, high concentrations of KP-1212 were superior compared of that of low concentrations for potentiating the antiviral activity of resveratrol and deferoxamine. The elevation of the viral mutant frequency by the drug combinations is consistent with an increase in virus mutational load as a major mechanism of antiretroviral activity.

KP-1212 potentiates the anti-HIV-1 activity of RNRIs without enhancing cytotoxicity. To further characterize these drug combinations, we investigated the extent to which KP-1212 can augment the potency of RNRIs. We found that the addition of KP-1212 at a fixed amount (5 μ M for gemcitabine or 200 μ M for resveratrol) enhanced the potency of both gemcitabine and resveratrol, as evidenced by the significant decrease in EC_{50} values upon addition of KP-1212 (Table 6-1). The greatest enhancement was observed for resveratrol, with the EC_{50} approximately halved. The effect was much less for gemcitabine, as only

low concentrations of KP-1212 could be used due to antagonism. The effect of KP-1212 in combination with deferoxamine was not statistically significant compared to deferoxamine alone. We then examined the effect of the drugs (alone and in combination) on cell viability in order to determine whether drug cytotoxicity could explain the antiviral effects observed. For gemcitabine, some degree of cytotoxicity was observed at all concentrations tested, but the CC_{50} was higher than the EC_{50} (62.0 μ M vs. 27.5 nM), indicating that the antiviral activity of gemcitabine cannot be explained by cytotoxicity (Figure 6-4 & Table 6-1). We also performed combustion analysis to confirm high purity of drug stocks, and conducted a separate cytotoxicity assay (trypan blue staining) to confirm cytotoxicity results (data not shown). Taken together, we conclude that gemcitabine has low cytotoxicity in this cell line, possibly due to inactivation by rapid deamination of gemcitabine to its major metabolite, 2',2'-difluoro-2'-deoxyuridine (dFdU) (336, 337). Unlike gemcitabine, KP-1212 was non-toxic up to 200 μ M (data not shown), and the addition of KP-1212 (5 μ M) did not alter the cytotoxicity of gemcitabine (Figure 6-4 & Table 6-1). These results indicate that enhanced cytotoxicity cannot explain the potentiation of the antiviral activity of gemcitabine by KP-1212. Similarly, the antiviral activity of resveratrol also could not be explained by cytotoxicity, as the CC_{50} was markedly higher than the EC_{50} (>400 μ M; Figure 6-4 & Table 6-1). Furthermore, the addition of KP-1212 (200 μ M) did not significantly alter the cytotoxicity of resveratrol. In summary, KP-1212 potentiates the antiviral activity of gemcitabine and resveratrol without augmenting cytotoxicity, which improves their selectivity indices against HIV-1 (Table 6-1).

Characterization of viral mutation spectra in the presence of KP-1212 and RNRIs. To determine the types of mutations resulting from these drugs (i.e., individually and in combination), we sorted out mCherry+/GFP- cells infected in the presence or absence of drug. We then amplified and performed Sanger sequencing of the *gfp* gene. We chose not to include deferoxamine in this analysis due to its poor potency against HIV-1 (Table 6-1). We obtained a total

of 376 unique mutations across all treatments, but several G-to-A hypermutants (likely a result of APOBEC3 activity) were excluded from further analysis (see Section 3.7 for details), lowering the total number of mutations to 303. In the absence of drug, G-to-A mutations were the most common, followed by A-to-G, T-to-C, and C-to-T transitions, with the remainder representing transversions (Figure 6-5). In the presence of only gemcitabine or resveratrol, no statistically significant ($p < 0.05$; 2-way Fisher's exact test) change in the relative appearance of each mutation type within the HIV-1 mutation spectrum was observed compared to the no drug control, suggesting that these drugs did not cause an increase in any particular mutation type. KP-1212, despite having only a modest ability to elevate the mutant frequency of HIV-1, caused a significant increase (p -value = 0.004) in the frequency of G-to-C mutations (from 1% to 22% of all mutations identified, relative to that of no drug), a transversion mutation rarely observed during HIV-1 replication. Strikingly, the combination of KP-1212 with resveratrol led to an even greater increase in the frequency of G-to-C mutations (up to 41% of all substitutions identified), which is significantly different from either resveratrol alone (p -value < 0.0001) or KP-1212 alone (p -value = 0.048). The combination of KP-1212 and gemcitabine did not result in a significant increase in the frequency of G-to-C mutations (p -value = 0.130), likely due to a lower concentration of KP-1212 used in this combination (i.e., 5 μ M versus 200 μ M) due to drug antagonism observed at higher concentrations of KP-1212.

KP-1212 has not been previously found to induce G-to-C transversions, but 5-aza-2'-deoxycytidine (decitabine) and its ribonucleoside counterpart (5-AZC), have been reported to cause G-to-C transversions in HIV-1 and cellular genomic DNA (117, 118, 196). These molecules decompose in water at physiological temperature and pH by hydrolytic opening and deformylation of the triazine ring (195, 197, 268, 269, 338, 339). The resulting ring-opened guanylurea derivatives are thought to pair with cytosine leading to G-to-C transversions (196). It is unclear how KP-1212 might induce G-to-C transversions, as KP-1212 should be resistant to hydrolysis.

Conclusions. While KP-1212 exerted minimal antiviral and mutagenic activity against HIV-1 in our assay, it was able to enhance the antiviral activities of the RNRIs gemcitabine and resveratrol, leading to robust reductions in HIV-1 infectivity and improvement of their selectivity indices. We found that the combinations simultaneously elevated the mutant frequency of HIV-1, suggesting enhancement of virus mutational loads as a major mechanism of the drug combinations. While the RNRIs gemcitabine and resveratrol do not appear to cause a specific type of mutation, the addition of KP-1212 to resveratrol led to a pronounced increase in G-to-C transversion mutations. Further studies of the antiretroviral mechanisms of these combinations are warranted, as are efforts to identify prodrug forms and derivatives exhibiting reduced toxicity, improved potency, and prolonged stability. Our findings suggest that the enhancement of viral mutant frequency is a major contributor to these antiviral mechanisms but do not preclude other activities, such as inhibition of reverse transcription. Resveratrol, while well-tolerated, has low bioavailability due to rapid and extensive metabolism in a number of clinical trials (340) but may serve as a lead for further development. It would also be of great interest to investigate the interplay of RNRIs \pm KP-1212 with conventional anti-HIV-1 inhibitors. Previous studies have shown that these inhibitors can accelerate mutagen-driven viral extinction. This has been demonstrated for HIV-1 using AZT and 5-hydroxy-2'-deoxycytidine in combination and in foot-and-mouth disease virus, in which case sequential treatments were superior to combination therapy (194, 341). Overall, our findings are the first to demonstrate that a mild anti-HIV-1 mutagen can significantly potentiate the antiretroviral activity of RNRIs. These observations may enhance the potential for clinical translation of KP-1212 for the treatment of HIV-1 infection.

Materials and Methods

Plasmids, cell lines, and reagents. The molecular clone of HIV-1 we used to produce virus, pNL4-3 MIG, was created from pNL4-3 HIG, which has been

previously described (117). The virus encoded by this plasmid contains mCherry, an internal ribosome entry site (IRES), and enhanced green fluorescent protein (GFP). We pseudotyped virions with VSV-G expressed from pHCMV-G, a kind gift from J. Burns (University of California, San Diego). The human embryonic kidney (HEK 293T) cells were purchased from American Type Culture Collection (Manassas, VA) and maintained in Dulbecco's Modified Eagle's Medium (DMEM) from Cellgro (Manassas, VA) with 10% HyClone FetalClone III (FC3) from Thermo Scientific (Waltham, MA) and 1% penicillin/streptomycin from Invitrogen (Carlsbad, CA). U373-MAGI-CXCR4_{CEM} cells were obtained from Michael Emerman through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (257). U373-MAGI cells were maintained similarly to 293T cells but with addition of 1.0 µg/mL puromycin, 0.1 mg/mL hygromycin B, and 0.2 mg/mL neomycin to the medium. The chemicals investigated were obtained from Sigma-Aldrich (deferroxamine and resveratrol; St. Louis, MO), Carbosynth (5-AZC and gemcitabine; Compton, UK), and Berry & Associates (KP-1212; Dexter, MI), dissolved in the appropriate solvent, and stored at -20 °C. For cloning, PCR was performed using the Platinum PCR Supermix from Invitrogen, while restriction enzymes, Klenow fragment, and T4 DNA ligase were from New England Biolabs (Ipswich, MA). For transfections, poly-L-lysine was from Newcomer Supply (Middleton, WI), and polyethylenimine (PEI) was from Polysciences, Inc. (Warrington, PA).

Construction of an mCherry/GFP HIV-1 vector for rapid mutation detection. The mCherry gene was swapped with the mouse heat stable antigen (hsa) in pNL4-3 HIG to generate pNL4-3 MIG as follows. From NL4-3 HIG, an IRES-GFP fragment was amplified via PCR and cloned into a pEGFP-N1-based mCherry vector using NotI and XbaI. The NotI site was then destroyed by fill-in with large Klenow fragment. The entire MIG cassette was then amplified using primers that added a NotI site to the 5' end and a XhoI site to the 3' end of the cassette. The PCR product was directly digested with NotI and XhoI and ligated into NotI and XhoI-digested pNL4-3 HIG using T4 DNA ligase. The resulting

plasmid, pNL4-3 MIG, expresses mCherry and GFP as well as all viral proteins except Env and Nef.

Virus production and titering. Virus was produced by co-transfecting pNL4-3 MIG and pHCMV-G into 293T cells via the PEI method (258). PEI stocks were prepared at 1 mg/mL by dissolving PEI in water, adjusting the pH to 7.0, and filtering through a 0.2 μ M filter. Stocks were divided into 1 mL aliquots and frozen at -80 °C. On the day before transfection, 4 million 293T cells were plated on each 10 cm plate (pre-coated with poly-L-lysine) to be transfected. For each plate, 10 μ g pNL4-3 MIG + 1 μ g pHCMV-G + 33 μ L 1 mg/mL PEI were mixed in a 1.6 mL microcentrifuge tube to a final volume of 1 mL using serum-free DMEM. After 20 minutes of incubation, the medium on the 293T cells was replaced and the DNA-PEI mixture was added. The medium was replaced the morning after transfection, and virus was collected ~48 hours post-transfection by filtering the supernatant through a 0.2 μ m filter. Viral stocks were divided into 1 mL aliquots and frozen at -80 °C.

Prior to any drug treatments, viral stocks were first titered in U373-MAGI cells. The day before infection, 62,500 cells/well were plated in 12-well plates. The next day the media was replaced and varying amounts of virus ranging from 1.25 to 40 μ L were added. The media was replaced again 24 hours post-infection and cells were collected at 72 hours post-infection for analysis by flow cytometry. The infectious titer was calculated from the flow data by adding all positive quadrants (mCherry+ only, GFP+ only, and mCherry+/GFP+) to determine infectivity and then plotting the volume of virus against infectivity.

Drug treatments and infections. Drug treatments were performed using U373-MAGI cells in 12-well plates similar to the way in which viral stocks were titered (Section 3.3). Two hours prior to infection, drugs (1-2 μ L) were added to a final volume of 0.5 mL media in each well. Uninfected cells and no drug (DMSO only-treated cells) were included as controls. After two hours, virus and additional media were added, resulting in a final volume of 1 mL in each well. Drug concentrations were based on the final volume of 1 mL. The amount of

virus added was targeted to achieve 15-30% infection (for the no drug control) based on the previously determined viral titer. Drugs were removed by replacing the media 24 hours post-infection. The cells were collected 72 hours post-infection for flow cytometry by treating with trypsin, pelleting at 500xg for 5 minutes, resuspending in 200 μ L Dulbecco's Phosphate-Buffered Saline (DPBS) + 2% FC3, and transferring to 96-well plates.

Infectivity and mutant frequency determination by flow cytometry. Cells were analyzed for mCherry and GFP expression on a BD LSR II flow cytometer (BD Biosciences; San Jose, CA). Cells were first gated based on forward scatter and side scatter, and a minimum of 10,000 gated cells were analyzed per sample. GFP was excited with a blue 488 nm laser and emission detected using 505LP and 525/50 filters. mCherry was excited by a custom green 561 nm laser and emission detected using 595LP and 610/20 filters. It should be noted that early on in our work we determined that very little fluorescent compensation was necessary with this particular combination of fluorophores, lasers, and filters (data not shown). These control experiments were performed by transfecting mCherry or GFP into 293T cells and then analyzing the transfected cells (either kept separately or mixed together) by flow cytometry. Flow cytometry data were examined in FlowJo v. 7.6.1 (Ashland, OR). Infectivity was determined by adding all three positive populations: mCherry+ only, GFP+ only, and mCherry+/GFP+. Mutant frequency was calculated by dividing the single positive populations (mCherry+ only and GFP+ only) by all infected cells. Double negative (mCherry-/GFP-) cells were excluded due to the difficulty in detecting these rare events; mutant frequency calculations are therefore underestimates. Infectivity and mutant frequency data were normalized to the no drug controls.

Cellular toxicity analysis. The cytotoxicity of drugs was assessed using the CellTiter-Glo Luminescent Cell Viability Assay from Promega (Madison, WI) following the manufacturer's instructions. 5000 U373-MAGI cells/well were plated in 96-well plates the day before drug treatment. The following day the medium was replaced and drugs (alone or in combination) were added to a final

volume of 200 μ L/well. No drug (DMSO only) and no cell wells were included as controls. The luminescence was recorded using an Orion microplate luminometer from Berthold Detection Systems (Huntsville, AL). The data were analyzed by first subtracting the no cell control value and then normalizing to the no drug control.

Characterization of mutation spectra in HIV-1. Infected mCherry+/GFP-U373-MAGI cells were sorted at the Masonic Cancer Center at the University of Minnesota using a BD FACSAria II. A minimum of 3000 cells were collected per sample. Genomic DNA was isolated by the High Pure PCR Template kit from Roche (Indianapolis, IN). GFP was amplified in a single PCR reaction using the Platinum PCR Supermix from Invitrogen and the primers 5'-CAAGCGTATTCAACAAGG-3' and 5'-GCAATACAGCAGCTAACAATG-3'. The PCR product was ligated into pGEM-T (Promega) and the *gfp* gene was sequenced at Functional Biosciences, Inc. (Madison, WI) using the T7 promoter primer. Sequences were aligned to the reference and mutations identified using Lasergene (DNASTAR, Inc., Madison, WI). Duplicate mutations, presumably arising from PCR amplification, were excluded from analyses. Several G-to-A hypermutants (multiple G-to-A mutations within the same sequence) were observed, and these likely resulted from APOBEC3 activity, as they occurred in a GA or GG dinucleotide context and were present in both the no drug as well as drug treatment samples. These rare events were excluded from the analyses due to their potential for creating false differences in the rate of G-to-A mutations between samples.

Statistical analyses. All graphs were created in Microsoft Excel v 12.3.5 (Redmond, WA) or GraphPad Prism v 5.0 (GraphPad Software, Inc.; La Jolla, CA). EC₅₀ values of drugs (drug concentration at which there is a 50% reduction in viral infectivity) were determined by plotting normalized infectivity against the log-transformed drug concentrations and fitting a non-linear regression curve to the data. Likewise, CC₅₀ values of drugs (drug concentration at which there is 50% cytotoxicity) were determined by plotting normalized cell viability against the

log-transformed drug concentrations and fitting a non-linear regression curve to the data. Two-way ANOVA was performed using the statistical package R to compare the impact of drug combinations against single drugs on infectivity and mutant frequency. All analyses were performed on the log-transformed data without normalization. Differences in mutation spectra were assessed using Fisher's exact test in GraphPad Prism.

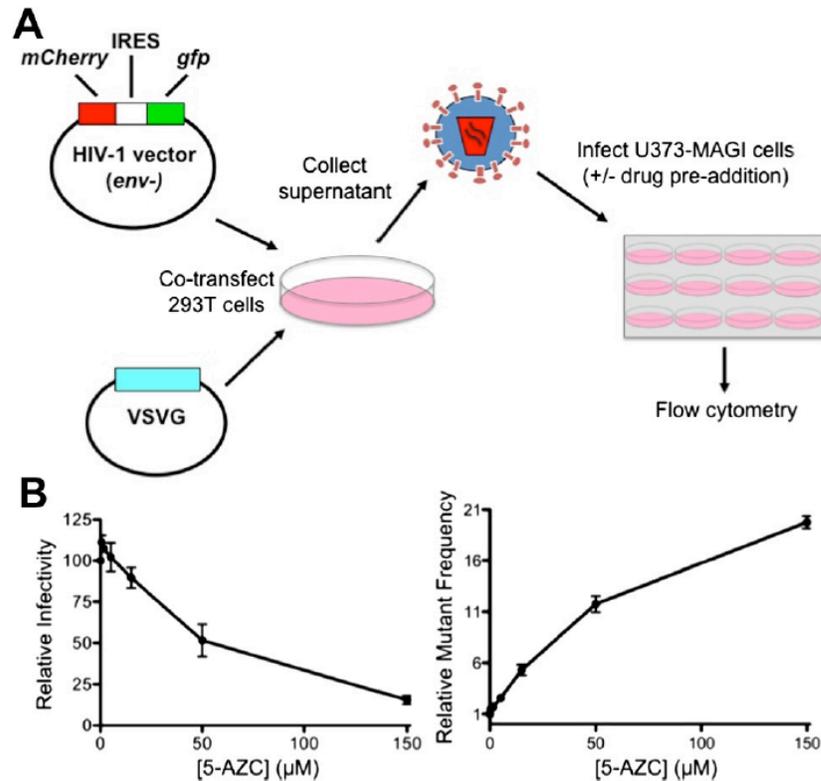


Figure 6-1. Vector and assay design for rapid determination of HIV-1 infectivity and mutant frequency. (A) A dual-reporter (mCherry/GFP) HIV-1 vector was created and co-transfected with VSVG to produce infectious virus. Target cells were pre-incubated with drugs for 2 h before infection. Drug was removed 24 h post-infection and cells were collected 72 h post-infection for determination of infectivity and mutant frequency *via* flow cytometry, as described in Section 3.5. (B) The utility of the dual reporter HIV-1 vector for detecting alterations in infectivity and mutant frequency was validated using 5-azacytidine, a previously characterized potent mutagen of HIV-1 (118). The data were normalized to the no drug (DMSO only) control and represent the mean \pm sd of three independent experiments.

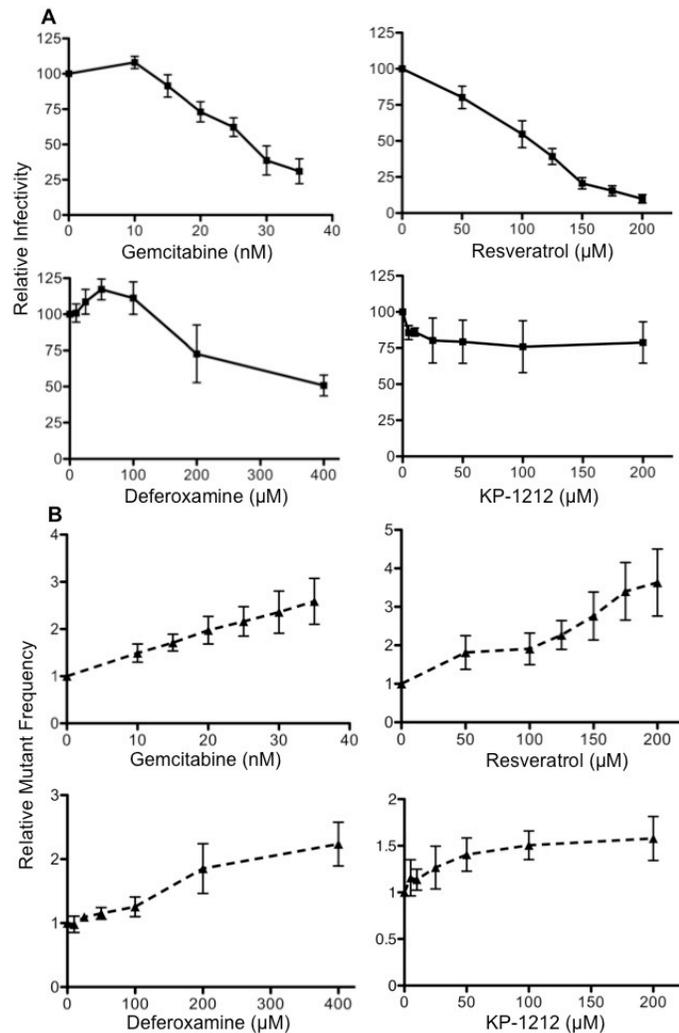


Figure 6-2. Effect of KP-1212 and RNRIs on HIV-1 infectivity and mutant frequency. Cells were infected with HIV-1 after being treated with KP-1212 or selected ribonucleotide reductase inhibitors (RNRIs) for 2 h. Treatments were concluded at 24 h post-infection by replacing the media, and cells were collected for analysis by flow cytometry at 72 h post-infection. The effects of the drugs on infectivity (A) and mutant frequency (B) were determined and normalized to the no drug control. The data shown are the mean \pm sd of three independent experiments.

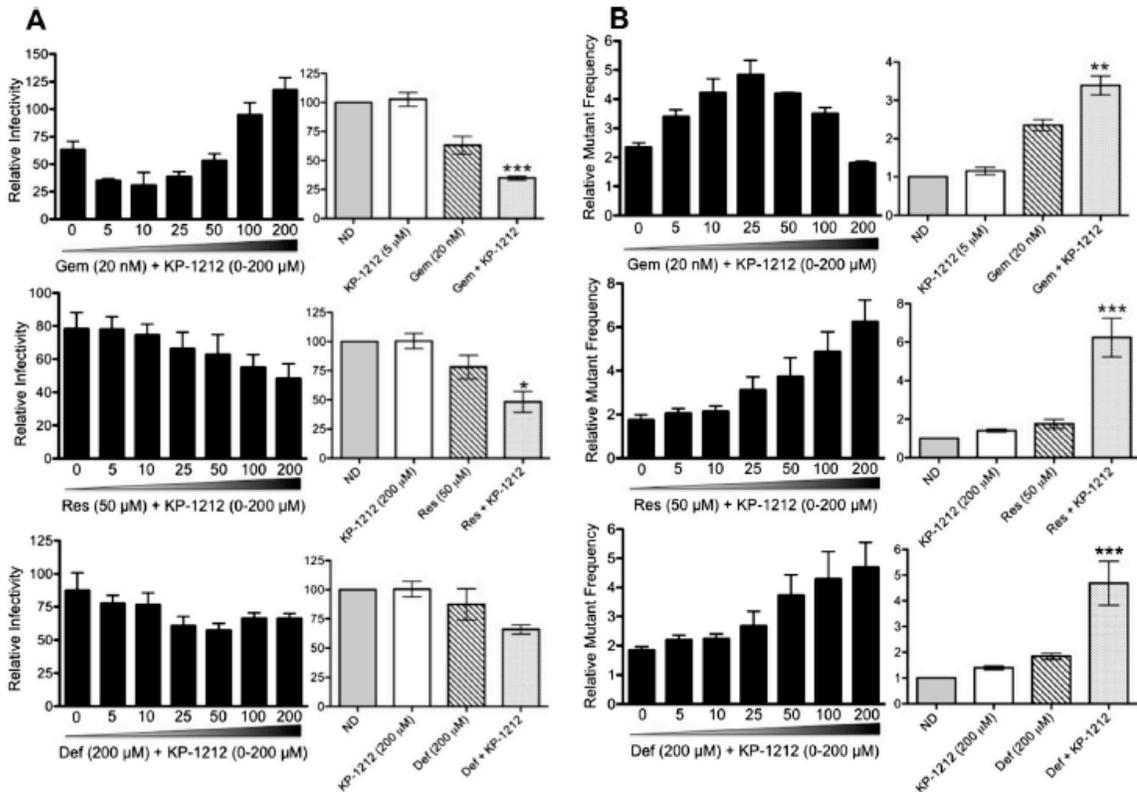


Figure 6-3. KP-1212 potentiates the antiviral and mutagenic activities of RNRIs against HIV-1. Cells were pre-treated with a set amount of each ribonucleotide reductase inhibitor (RNRI) and varying amounts of KP-1212 (from 0 to 200 μM) as described in Section 3.4. Flow cytometry was used to determine HIV-1 infectivity (A) and mutant frequency (B), which were plotted relative to no drug. Statistical significance of drug combinations relative to both single drugs was assessed using a two-way ANOVA. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001. The data shown represent the mean ± sd of three independent experiments.

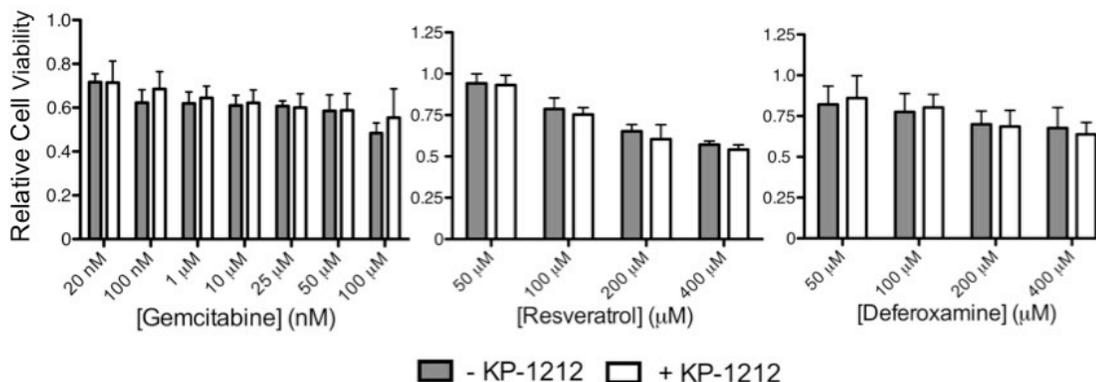


Figure 6-4. The ability of KP-1212 to potentiate RNRIs is not due to enhanced cytotoxicity. U373-MAGI cells were treated with ribonucleotide reductase inhibitors (RNRIs) \pm KP-1212 (5 μ M for gemcitabine, 200 μ M for resveratrol and deferoxamine) for 24 h, then washed and replenished with fresh media. Cell viability was assessed 24 h after molecule removal using a luminescent ATP-based kit, and the data were normalized to the no drug control. The data shown are the mean \pm sd of four independent experiments.

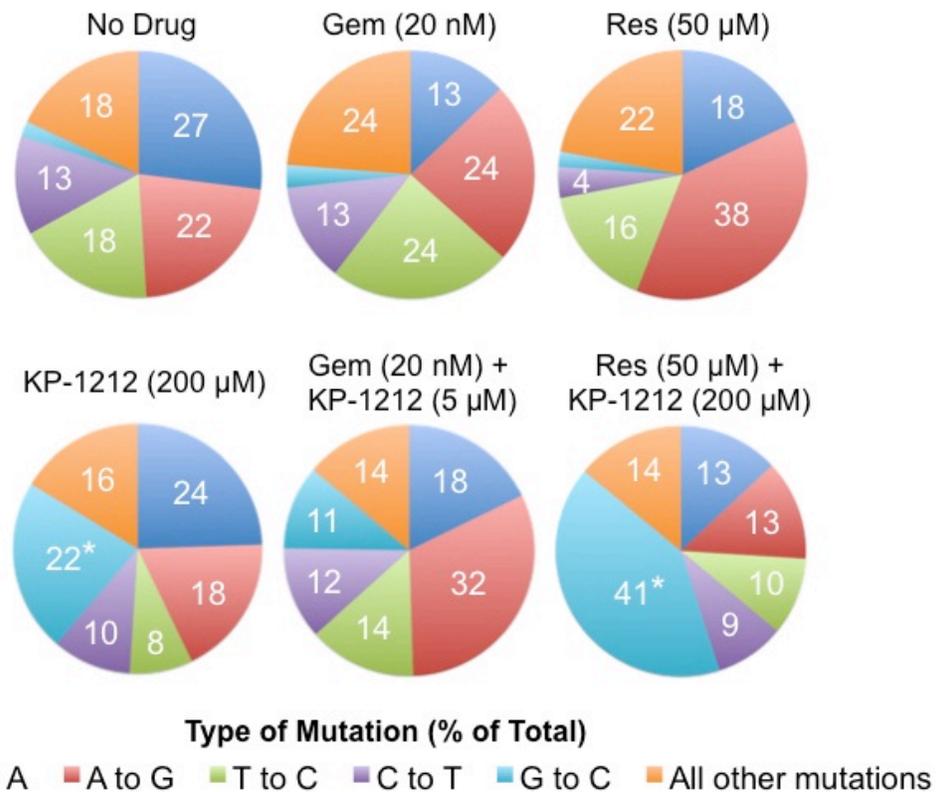


Figure 6-5. Characterization of viral mutation spectra in the presence of KP-1212 and RNRIs. U373-MAGI cells were pre-treated with KP-1212 and/or ribonucleotide reductase inhibitors (RNRIs) and infected with HIV-1. mCherry+/GFP- cells were sorted to enrich for mutants, and the *gfp* gene was amplified and sequenced. Each number in a wedge of a pie chart represents the specific corresponding mutational type, expressed as the percentage of the total mutations observed for that treatment. The total number of mutations sequenced were: no drug, 45, gemcitabine (Gem), 38; resveratrol (Res), 45; KP-1212, 49; gemcitabine + KP-1212, 57; resveratrol + KP-1212, 69 (303 mutations in total). Statistically significant (p -value < 0.05) changes relative to the no drug control are indicated with an asterisk.

Table 6-1. KP-1212 improves the antiretroviral potency of RNRIs without altering cytotoxicity.

Drug	-KP-1212			+KP-1212		
	EC ₅₀	CC ₅₀	SI	EC ₅₀	CC ₅₀	SI
Gemcitabine	27.5 nM (26.0–29.1)	62.0 μM (59.0–65.2)	2250	21.6 nM (20.9–22.3)	64.2 μM (60.8–67.8)	2970
Resveratrol	99.6 μM (92.5–107.4)	>400 μM	>4.0	52.8 μM (39.1–71.2)	>400 μM	>7.6
Deferoxamine	384 μM (288–514)	>400 μM	>1.0	248 μM (194–316)	>400 μM	>1.6

The EC₅₀ and CC₅₀ values for ribonucleotide reductase inhibitors (RNRIs) in the presence or absence of KP-1212 (5 μM for gemcitabine; 200 μM for resveratrol/deferaxamine) in U373-MAGI cells were calculated as described in Materials and Methods. Selectivity indices were determined by dividing the CC₅₀ by the respective EC₅₀. The 95% confidence intervals are indicated in parentheses. The data shown represent the mean ± sd of at least three independent experiments.

CHAPTER VII
DISSERTATION SUMMARY AND FUTURE DIRECTIONS

This dissertation has demonstrated that the major difference in viral mutagenesis between HIV-1 and HIV-2 is a lower frequency of G-to-A hypermutants for HIV-2 (Chapter II). In the absence of G-to-A hypermutants, HIV-2 did not exhibit a significantly different total mutation frequency than HIV-1, although small but statistically significant differences in G-to-A, A-to-G, and C-to-T transition frequencies were still noted. The pattern of G-to-A hypermutation was consistent with the editing activity of APOBEC3 proteins, suggesting that in this particular experimental system HIV-2 was less susceptible to APOBEC3 editing than HIV-1. These findings suggest that differences in general replication fidelity (i.e. in the absence of APOBEC3 proteins) may not contribute to the unique clinical features of HIV-2 infection, namely attenuated pathogenesis, reduced viral loads, and better control by the host immune response (9, 215).

These findings raise a number of unresolved issues that will be addressed in future studies. First, it is not yet clear whether HIV-2 exhibits reduced susceptibility to APOBEC3 proteins under physiologically relevant expression levels. In fact, surprisingly little is known about the restriction and hypermutation of HIV-2 by APOBEC3 proteins. Previously, one group demonstrated that HIV-1 Vif and HIV-2 Vif interact with APOBEC3 proteins using distinct binding surfaces and that they vary in their ability to degrade specific APOBEC3 proteins (250). Further, a second group found that Vif-deficient HIV-2 is restricted by APOBEC3G, although not quite as potently as for Vif-deficient HIV-1 (233). The ability of other APOBEC3 proteins to restrict HIV-2 has not yet been examined, such that the restrictive repertoire of APOBEC3 proteins for HIV-2 remains almost entirely undefined. In the study described in this dissertation, 293T cells were used for the production of HIV-1 and HIV-2, as large quantities of highly infectious virus stocks were needed for Illumina sequencing. However, 293T cells express low levels of most APOBEC3 proteins, particularly those capable of restricting Vif-deficient HIV-1, such that viruses produced in the presence or absence of Vif in 293T cells exhibit similar infectivities (156, 159, 225, 233, 248). This was not initially viewed as a limitation for the comparison of HIV-1 and HIV-2

mutagenesis since the primary interest was in comparing the fidelity of reverse transcription between the two viruses. Given the finding of differences in hypermutation frequencies between HIV-1 and HIV-2, it would be of interest to re-examine the mutagenesis of HIV-1 and HIV-2 using virus stocks produced in the presence of physiologically relevant levels of APOBEC3 expression, such as in immortalized or primary CD4⁺ T-cells. Second, the specific APOBEC3 protein(s) responsible for the differences in hypermutation has not yet been determined. Most G-to-A mutations in hypermutants were positioned at GA dinucleotides, indicating that A3G (which prefers almost exclusively GG dinucleotides) cannot be the primary driver of the observed hypermutation (137, 141, 142, 247). Several other APOBEC3 proteins, including A3C, A3D, and A3F have been detected at the mRNA level in 293T cells (225), although protein expression could not be examined due to a lack of specific antibodies. Unlike A3D and A3F, A3C does not restrict the replication of Vif-deficient HIV-1 in T-cells (134), but it remains possible that A3C could induce the low levels of hypermutation (<1%) observed in our study. The specific APOBEC3 protein responsible for the observed hypermutation could be identified by knock-out or knock-down of select APOBEC3 proteins in 293T cells. However, this would be difficult, time-consuming, and of limited significance considering the low levels of APOBEC3 expression in 293T cells. Instead, HIV-1 and HIV-2 hypermutation should first be compared using immortalized or primary CD4⁺ T-cells, followed by knock-out or knock-down of select APOBEC3 proteins to examine the basis for any observed differences. For example, HIV-1 and HIV-2 hypermutation could be compared in CEM2n cells, for which a variety of stable APOBEC3 knock-out and knock-down derivatives have already been constructed (135).

If HIV-2 is found to exhibit reduced susceptibility to APOBEC3 proteins in CD4⁺ T-cells, further experiments could be performed to determine the mechanistic basis for differences in susceptibility. The reduced susceptibility of HIV-2 to G-to-A hypermutation could potentially be due to decreased levels of the responsible APOBEC3 protein(s) in virions, which could result from either: 1.

more efficient degradation of APOBEC3 protein(s) by HIV-2 Vif than by HIV-1 Vif, or 2. less efficient packaging of APOBEC3 proteins into the viral core for HIV-2 than for HIV-1. Additionally, experiments could be performed to examine the potential significance of reduced hypermutation to HIV-2 replication, diversification, and evolution. The reduced susceptibility of HIV-2 to APOBEC3 proteins could promote viral replication by leading to decreased instances of lethal hypermutation but, on the other hand, could prevent HIV-2 from capitalizing on APOBEC3 proteins as a source of genetic variation.

While several differences in viral mutagenesis between HIV-1 and HIV-2 were identified in this initial study, a detailed comparison of all mutational types could not be performed due to high levels of background error from PCR and/or sequencing. To address this, a method based on single-strand consensus sequencing (SSCS) (342, 343) was developed and shown to efficiently reduce the frequencies of most types of background errors (Appendix I). More specifically, SSCS reduced background error frequencies for insertions, deletions, and all types of substitutions except C-to-A transversions. SSCS also completely eliminated background G-to-A hypermutants (defined as reads with at least two G-to-A mutations), although the background level of G-to-A hypermutants was already quite low during standard Illumina sequencing. SSCS was used to collect preliminary data comparing the mutagenesis of HIV-1 and HIV-2, and HIV-2 was again found to exhibit a lower total mutation frequency than HIV-1, due in part to lower levels of G-to-A hypermutants (Appendix I). In continuing studies, SSCS is being used to compare viral mutagenesis using viral vectors that contain reverse transcriptase (RT) from 16 different viral isolates (10 from HIV-1, 6 from HIV-2; see Table 7-1). Notably, these vectors will allow comparison of viral mutagenesis between HIV types as well as between the most prevalent subtypes of HIV-1 Group M. Several previous studies have reported differences in transmissibility and pathogenicity depending on HIV-1 subtype (344-348), and SSCS will test the hypothesis that these differences could relate to variable mutation frequencies and spectra. These viral vectors were created

in an isogenic manner (in which only the RT was altered) so that any observed differences in mutagenesis could be attributed to RT. In contrast, swapping the entire viral sequence could lead to differences in mutagenesis due to variability in Vif or in the template sequence. These studies using SSCS and multiple viral vectors will expand our initial report by: 1. enabling the comparison of HIV-1 and HIV-2 mutagenesis under conditions of much lower background error, which may resolve additional differences in mutagenesis between HIV-1 and HIV-2, 2. determining whether initial findings were truly representative of HIV-1 and HIV-2 as a whole (as multiple RTs will be examined from each type), and 3. examining whether there are differences in viral mutagenesis between the most prevalent subtypes of HIV-1 Group M (A, B, and C).

Although SSCS significantly enhanced the examination of viral mutagenesis by high-throughput sequencing, there are still areas where further improvements to the method could be made. SSCS did not reduce levels of C-to-A transversions, the most common type of background error during our experiments, thus leading to only mild reductions (2- to 3-fold) in the total background error frequency (Appendix I). SSCS also lowered background G-to-A transitions somewhat less effectively than C-to-T, A-to-G, and T-to-C transitions. It was hypothesized that background C-to-A and G-to-A mutations in consensus sequences were created by low levels of DNA damage in the starting template. If so, SSCS could likely be greatly improved by the use of double-stranded tags (i.e. duplex consensus sequencing) (246) or the incorporation of DNA repair enzymes into sample preparation protocols. Further, SSCS currently only examines one region of interest (~250 bp) in the provirus at a time, and the number of targets cannot easily be expanded due to the laborious nature of the protocols involved. In the future, it may be possible to amplify and sequence multiple targets of interest at the same time, either by multiplexing SSCS reactions or by isolating proviral DNA and performing SSCS by ligating adapters with unique tags to sheared proviral DNA.

Illumina high-throughput sequencing was also used to characterize the lethal mutagenesis of HIV-1 by decitabine in great detail (Chapter III). Decitabine was previously shown to potently reduce the infectivity of HIV-1 by inducing primarily G-to-C transversions in the virus (117). G-to-C transversions were hypothesized to be caused by incorporation of decitabine in place of deoxycytidine during minus-strand synthesis, followed by hydrolysis and base-pairing with deoxycytidine during plus-strand synthesis. In contrast, decitabine incorporated during plus-strand synthesis was hypothesized to be repaired appropriately by host enzymes in the nucleus, explaining the lack of C-to-G transversions. Using Illumina sequencing, previous observations were confirmed by the finding that decitabine induced primarily G-to-C transversions, but decitabine also induced (to a lesser extent) C-to-G transversions. C-to-G transversions were likely observed in the present study due to the much greater sequencing depth of Illumina sequencing compared to Sanger sequencing. While the mechanism of C-to-G transversions was not determined, it is hypothesized that C-to-G transversions were caused by the incorporation of decitabine during plus-synthesis, followed by improper repair or simply a lack of repair. Alternatively, C-to-G transversions could have been caused by the direct incorporation of decitabine hydrolysis products during minus-strand synthesis. Illumina sequencing was also used to examine the susceptibility of individual sequence positions to decitabine and to determine whether certain sequence positions acted as decitabine hotspots or coldspots. It was found that every guanine position examined (134/134) and nearly every cytosine position examined (163/169) was significantly susceptible to the mutagenic effects of decitabine, indicating extreme promiscuity of decitabine incorporation during reverse transcription. However, the degree of susceptibility varied greatly among sequence positions. Due to this extensive variability, decitabine-mediated G-to-C mutational hotspots were not identified, which were defined as upper outliers within the mutation frequency distribution.

Additional studies will be required to explain the mechanistic basis for decitabine-induced C-to-G transversions in HIV-1. As part of this process, it will be important to determine which host repair enzymes remove and replace incorporated decitabine, the kinetics by which repair proceeds, and the fidelity of the repair process. It will also be important to try to separate the effects of decitabine and decitabine-derived hydrolysis products (both in cell-based assays and in assays performed with HIV-1 RT *in vitro*), as C-to-G transversions may be caused by direct incorporation of hydrolysis products. However, this will likely prove challenging due to the instability of decitabine and hydrolysis intermediates, lack of commercial availability of some of the required compounds, and difficulties in synthesizing some of the required compounds (and oligonucleotides containing those compounds). Further, while decitabine-mediated mutation frequencies varied between amplicons and individual sequence positions, susceptibility could not be clearly associated with sequence context or amplicon position within the genome. Illumina sequencing of the entire genome could reveal whether decitabine exhibits subtle sequence context preferences and whether certain regions of the genome (such as those that remain single-stranded for a longer period of time) are more susceptible to decitabine than others. Also, while decitabine was previously found to be non-cytotoxic and non-mutagenic of the host cellular DNA at antiviral concentrations (117), the potential mutagenicity of decitabine of host cellular DNA (mitochondrial or genomic) has not yet been examined by direct high-throughput sequencing. In future studies, Illumina sequencing could be performed of total, sheared genomic DNA in order to examine the potential mutagenicity of decitabine throughout the entire genome at concentrations relevant to antiviral activity. Lastly, the total mutational load in proviruses exposed to decitabine could not be assessed by Illumina sequencing, and therefore the average number of decitabine-induced mutations required to eliminate viral infectivity could not be determined. The total mutational load could not be determined by Illumina sequencing because the read pairs were short (~120 bp in length) and un-linked between amplicons.

Additional experiments using single genome amplification and Sanger sequencing will be required to determine the total mutational load in proviruses exposed to decitabine.

The ribonucleoside analog 5-azacytidine (5-AZC) has previously been reported to reduce HIV-1 infectivity in the early phase of replication by inducing G-to-C transversions (118). 5-AZC could potentially be incorporated during reverse transcription as a ribonucleotide (i.e. as 5-AZC-triphosphate, or 5-AZC-TP) or as a deoxyribonucleotide (i.e. as 5-aza-2'-deoxycytidine-triphosphate, or decitabine-TP). For the latter to occur, 5-AZC-diphosphate would first have to be reduced to decitabine-diphosphate by the cellular enzyme ribonucleotide reductase. While HIV-1 RT selectively excludes ribonucleotides using a steric gate mechanism (216-218), HIV-1 RT can incorporate substantial levels of endogenous ribonucleotides under certain circumstances (219, 220). Nonetheless, it was hypothesized that 5-AZC would act primarily as decitabine-TP during HIV-1 reverse transcription. In Chapter IV, using a combination of *in vitro* incorporation assays, analytical methods, and high-throughput sequencing, 5-AZC was clearly shown to act primarily as decitabine-TP against HIV-1. These data have important implications for the design of anti-retroviral ribonucleoside analogs, which are of general interest because they are often easier to synthesize, purify, and modify than deoxyribonucleoside analogs, and are also capable of targeting both viral RNA and DNA. These findings suggest that the efficient reduction of ribonucleoside analogs to deoxyribonucleoside analogs is likely a crucial determinant of their antiviral potency, at least for targeting viral replication in cell types with high dNTP pools, such as CD4+ T-cells (282). Further studies will be required to determine whether 5-AZC can be directly incorporated as a ribonucleotide in cell types with low dNTP pools, such as macrophages. It would also be of interest to examine the antiviral activity of chain-terminating or mutagenic ribonucleoside analogs that cannot be reduced, as these compounds may provide a means to specifically target viral replication in macrophages, an important viral reservoir.

In Chapter V, experiments were performed to test the hypothesis that the anti-HIV-1 activity of 5-AZC would be antagonized by ribonucleotide reductase inhibitors (RNRIs). RNRIs were expected to inhibit reduction of 5-AZC to decitabine, ultimately decreasing the amount of decitabine-TP available for incorporation during reverse transcription. Surprisingly, 5-AZC instead exhibited antiviral synergy with low concentrations of four different RNRIs (clofarabine, gemcitabine, hydroxyurea, and resveratrol), similar to previous observations with combinations of RNRIs and decitabine (117, 201). RNRIs did not significantly inhibit reduction of 5-AZC to decitabine, explaining the lack of antagonism in the combinations. For one particular combination, 5-AZC and resveratrol, the mechanism of antiviral synergy was assessed in more detail by high-throughput sequencing and qPCR. Resveratrol potentiated 5-AZC-mediated decreases in RT products but not 5-AZC-mediated increases in viral mutation frequencies. These data indicate that the primary mechanism of antiviral synergy for this combination was the depletion of RT products rather than lethal mutagenesis.

Collectively, these findings demonstrate that RNRIs can potentiate the activity of a ribonucleoside analog against HIV-1, which has not previously been reported and may generally apply to other antiviral ribonucleoside analogs as well. In the future, it would be of interest to investigate whether RNRIs can potentiate the activity of 5-AZC in cell types with naturally low levels of endogenous dNTPs, such as macrophages (282). Even if 5-AZC is directly incorporated as a ribonucleotide in macrophages, RNRIs could likely further enhance its incorporation by further depleting endogenous dNTP pools. It would also be of interest to determine whether RNRIs potentiate other antiviral ribonucleoside analogs, including ribonucleosides that cannot be reduced within the cell. These compounds have been proposed as a means to specifically target HIV-1 replication in macrophages (219). Lastly, further studies are required to fully elucidate the mechanism by which resveratrol and 5-AZC reduced levels of RT products. Resveratrol and 5-AZC could decrease RT products by directly inhibiting viral DNA synthesis or, more likely, by a variety of

indirect mechanisms, such as depletion of dNTP pools or mutation-induced degradation of viral DNA.

Lastly, in Chapter VI, the anti-HIV-1 activity of the deoxyribonucleoside analog KP-1212 was investigated. KP-1212 was initially reported to potently inhibit HIV-1 replication in cell culture by lethal mutagenesis (198). However, in clinical trials, KP-1461 (a prodrug of KP-1212) unfortunately did not exhibit significant antiviral activity and demonstrated only mild mutagenicity against HIV-1 (199, 200). In Chapter VI, it was hypothesized that the antiviral activity and mutagenicity of KP-1212 would be improved by combining KP-1212 with RNRIs. It was found that KP-1212 alone exhibited very little antiviral activity, in agreement with results from clinical trials (199) but in apparent contrast to previous findings in cell culture (198). The stocks of KP-1212 used in this study were analysed by NMR and elemental combustion analysis, which excluded compound impurity or degradation as the factor responsible for the lack of antiviral activity. However, it was found that KP-1212 significantly potentiated the antiviral activity of multiple RNRIs. The combination of KP-1212 and resveratrol was shown to increase the HIV-1 mutant frequency by inducing G-to-C transversions, consistent with enhanced mutagenesis as the primary mechanism of the combination. However, the impact of the combinations on levels of RT products was not assessed, such that an effect on viral DNA synthesis cannot be excluded. Overall, these findings may enhance the clinical translation of KP-1212 and related nucleoside analogs for the treatment of HIV-1 or other viruses.

It is not yet clear why KP-1212 did not exhibit significant antiviral activity in the experiments performed in Chapter VI or in phase II clinical trials (199), as KP-1212 was originally reported to exhibit potent antiviral activity (198). Although there were a number of minor differences between the assays used (e.g. cell type, viral vector, etc.), none of these differences would be expected to result in a major difference in the potency of KP-1212. Nonetheless, KP-1212 activity has not yet been tested in exactly the same manner in which its activity was originally identified. Importantly, the stocks of KP-1212 used were confirmed to be pure

and not degraded. It remains possible that the KP-1212 stocks used in the original study contained impurities or degradation products that were responsible for the observed activity of the compound. While it did not exhibit antiviral activity, KP-1212 was found to potentiate the antiviral activity of multiple RNRIs, most notably the natural product resveratrol, resulting in induction of G-to-C transversions in HIV-1. Further studies will be required to identify the mechanism by which resveratrol and KP-1212 caused G-to-C transversions in HIV-1. It is hypothesized that KP-1212 was converted intracellularly to decitabine, a closely related nucleoside analog that induces primarily G-to-C transversions. Even if the conversion of KP-1212 to decitabine was extremely inefficient, it would likely be enough to explain the observed results, as decitabine exhibits fairly potent anti-HIV-1 activity (117). Further, additional studies should be performed to identify pro-drug forms of resveratrol with prolonged stability, as resveratrol was found to interact the most favourably with 5-AZC and KP-1212 but is rapidly metabolized within the body (340).

In sum, the research performed in this dissertation has demonstrated that:

1. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra, at least in the absence of G-to-A hypermutation,
2. Decitabine is incorporated promiscuously by HIV-1 RT and does not induce G-to-C mutational hotspots,
3. 5-AZC primarily enhances the mutagenesis of HIV-1 after reduction to decitabine,
4. Combinations of 5-AZC and RNRIs (at low concentrations) unexpectedly lead to anti-HIV-1 synergy, and
5. KP-1212 potentiates the anti-HIV-1 activity of RNRIs.

Taken together, these findings advance our understanding of determinants of HIV mutagenesis, a key process that promotes viral evolution and acts as a barrier to vaccines and antiviral drugs. Further, this work has offered support for the use of lethal mutagenesis as an effective strategy for eliminating HIV-1 infectivity and has opened additional avenues for the design of synergistic drug combinations directed against HIV-1 and other retroviruses.

Table 7-1. HIV-1 and HIV-2 vectors for comparison of viral mutagenesis by single-strand consensus sequencing.^a

<i>HIV Type</i>	<i>Number</i>	<i>Subtype (HIV-1) or Group (HIV-2)</i>	<i>Isolate Name</i>
HIV-1	1	A	94CY017.41
HIV-1	2	B	NL4-3
HIV-1	3	B	YU2
HIV-1	4	B	BH10
HIV-1	5	B	93BR029.4
HIV-1	6	C	MJ4
HIV-1	7	C	96ZM651.8
HIV-1	8	C	98IS002.5
HIV-1	9	D	94UG114.1.6
HIV-1	10	G	92NG003.1
HIV-2	11	A	ROD
HIV-2	12	A	ST
HIV-2	13	A	ISY
HIV-2	14	A	D194
HIV-2	15	A	CBL-20
HIV-2	16	A	GH

a. A panel of 10 HIV-1 and 6 HIV-2 vectors were created by sub-cloning the reverse transcriptase regions from other viral isolates into pNL4-3 MIG (HIV-1) or pROD MIG (HIV-2) in an isogenic fashion. HIV-1 vectors were engineered with RTs from multiple subtypes of Group M, especially subtypes B (the predominant subtype in the U.S.) and C (the predominant subtype globally). HIV-2 vectors were engineered with RTs only from Group A. HIV-2 Group B is also widespread, although significantly less prevalent than Group A, but only one Group B molecular clone could be obtained, and it was found to be highly unstable in *E. coli* and not suitable for sub-cloning.

BIBLIOGRAPHY

1. **Gottlieb GS, Eholie SP, Nkengasong JN, Jallow S, Rowland-Jones S, Whittle HC, Sow PS.** 2008. A call for randomized controlled trials of antiretroviral therapy for HIV-2 infection in West Africa. *AIDS* **22**:2069-2072; discussion 2073-2064.
2. **da Silva ZJ, Oliveira I, Andersen A, Dias F, Rodrigues A, Holmgren B, Andersson S, Aaby P.** 2008. Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? *AIDS* **22**:1195-1202.
3. **Tienen C, van der Loeff MS, Zaman SM, Vincent T, Sarge-Njie R, Peterson I, Leligdowicz A, Jaye A, Rowland-Jones S, Aaby P, Whittle H.** 2010. Two distinct epidemics: the rise of HIV-1 and decline of HIV-2 infection between 1990 and 2007 in rural Guinea-Bissau. *J Acquir Immune Defic Syndr* **53**:640-647.
4. **van der Loeff MF, Awasana AA, Sarge-Njie R, van der Sande M, Jaye A, Sabally S, Corrah T, McConkey SJ, Whittle HC.** 2006. Sixteen years of HIV surveillance in a West African research clinic reveals divergent epidemic trends of HIV-1 and HIV-2. *Int J Epidemiol* **35**:1322-1328.
5. **Adjorlolo-Johnson G, De Cock KM, Ekpini E, Vetter KM, Sibailly T, Brattegaard K, Yavo D, Doorly R, Whitaker JP, Kestens L, et al.** 1994. Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast. *JAMA* **272**:462-466.
6. **Kanki PJ, Travers KU, S MB, Hsieh CC, Marlink RG, Gueye NA, Siby T, Thior I, Hernandez-Avila M, Sankale JL, et al.** 1994. Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**:943-946.
7. **O'Donovan D, Ariyoshi K, Milligan P, Ota M, Yamuah L, Sarge-Njie R, Whittle H.** 2000. Maternal plasma viral RNA levels determine marked differences in mother-to-child transmission rates of HIV-1 and HIV-2 in The Gambia. MRC/Gambia Government/University College London Medical School working group on mother-child transmission of HIV. *AIDS* **14**:441-448.
8. **Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, Traore I, Hsieh CC, Dia MC, Gueye EH, et al.** 1994. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**:1587-1590.
9. **de Silva TI, Cotten M, Rowland-Jones SL.** 2008. HIV-2: the forgotten AIDS virus. *Trends Microbiol* **16**:588-595.
10. **Andersson S, Norrgren H, da Silva Z, Biague A, Bamba S, Kwok S, Christopherson C, Biberfeld G, Albert J.** 2000. Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Arch Intern Med* **160**:3286-3293.

11. **Popper SJ, Sarr AD, Travers KU, Gueye-Ndiaye A, Mboup S, Essex ME, Kanki PJ.** 1999. Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. *J Infect Dis* **180**:1116-1121.
12. **Gottlieb GS, Hawes SE, Agne HD, Stern JE, Critchlow CW, Kiviat NB, Sow PS.** 2006. Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: implications for differences in transmission. *AIDS* **20**:895-900.
13. **Schim van der Loeff MF, Jaffar S, Aveika AA, Sabally S, Corrah T, Harding E, Alabi A, Bayang A, Ariyoshi K, Whittle HC.** 2002. Mortality of HIV-1, HIV-2 and HIV-1/HIV-2 dually infected patients in a clinic-based cohort in The Gambia. *AIDS* **16**:1775-1783.
14. **Schaller T, Bauby H, Hue S, Malim MH, Goujon C.** 2014. New insights into an X-traordinary viral protein. *Front Microbiol* **5**:126.
15. **Sauter D.** 2014. Counteraction of the multifunctional restriction factor tetherin. *Front Microbiol* **5**:163.
16. **Hauser H, Lopez LA, Yang SJ, Oldenburg JE, Exline CM, Guatelli JC, Cannon PM.** 2010. HIV-1 Vpu and HIV-2 Env counteract BST-2/tetherin by sequestration in a perinuclear compartment. *Retrovirology* **7**:51.
17. **Le Tortorec A, Neil SJ.** 2009. Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. *J Virol* **83**:11966-11978.
18. **Arien KK, Abraha A, Quinones-Mateu ME, Kestens L, Vanham G, Arts EJ.** 2005. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J Virol* **79**:8979-8990.
19. **MacNeil A, Sankale JL, Meloni ST, Sarr AD, Mboup S, Kanki P.** 2007. Long-term inpatient viral evolution during HIV-2 infection. *J Infect Dis* **195**:726-733.
20. **Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A.** 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput Biol* **3**:e29.
21. **Skar H, Borrego P, Wallstrom TC, Mild M, Marcelino JM, Barroso H, Taveira N, Leitner T, Albert J.** 2010. HIV-2 genetic evolution in patients with advanced disease is faster than that in matched HIV-1 patients. *J Virol* **84**:7412-7415.
22. **Duvall MG, Precopio ML, Ambrozak DA, Jaye A, McMichael AJ, Whittle HC, Roederer M, Rowland-Jones SL, Koup RA.** 2008. Polyfunctional T cell responses are a hallmark of HIV-2 infection. *Eur J Immunol* **38**:350-363.
23. **Duvall MG, Jaye A, Dong T, Brenchley JM, Alabi AS, Jeffries DJ, van der Sande M, Togun TO, McConkey SJ, Douek DC, McMichael AJ, Whittle HC, Koup RA, Rowland-Jones SL.** 2006. Maintenance of HIV-

- specific CD4+ T cell help distinguishes HIV-2 from HIV-1 infection. *J Immunol* **176**:6973-6981.
24. **Kong R, Li H, Bibollet-Ruche F, Decker JM, Zheng NN, Gottlieb GS, Kiviat NB, Sow PS, Georgiev I, Hahn BH, Kwong PD, Robinson JE, Shaw GM.** 2012. Broad and potent neutralizing antibody responses elicited in natural HIV-2 infection. *J Virol* **86**:947-960.
 25. **Goujon C, Riviere L, Jarrosson-Wuilleme L, Bernaud J, Rigal D, Darlix JL, Cimarelli A.** 2007. SIVSM/HIV-2 Vpx proteins promote retroviral escape from a proteasome-dependent restriction pathway present in human dendritic cells. *Retrovirology* **4**:2.
 26. **Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, Yatim A, Emiliani S, Schwartz O, Benkirane M.** 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**:654-657.
 27. **Manel N, Hogstad B, Wang Y, Levy DE, Unutmaz D, Littman DR.** 2010. A cryptic sensor for HIV-1 activates antiviral innate immunity in dendritic cells. *Nature* **467**:214-217.
 28. **Yu H, Usmani SM, Borch A, Kramer J, Sturzel CM, Khalid M, Li X, Krnavek D, van der Ende ME, Osterhaus AD, Gruters RA, Kirchhoff F.** 2013. The efficiency of Vpx-mediated SAMHD1 antagonism does not correlate with the potency of viral control in HIV-2-infected individuals. *Retrovirology* **10**:27.
 29. **Gibbs JS, Lackner AA, Lang SM, Simon MA, Sehgal PK, Daniel MD, Desrosiers RC.** 1995. Progression to AIDS in the absence of a gene for vpr or vpx. *J Virol* **69**:2378-2383.
 30. **Hirsch VM, Sharkey ME, Brown CR, Brichacek B, Goldstein S, Wakefield J, Byrum R, Elkins WR, Hahn BH, Lifson JD, Stevenson M.** 1998. Vpx is required for dissemination and pathogenesis of SIV(SM) PBj: evidence of macrophage-dependent viral amplification. *Nat Med* **4**:1401-1408.
 31. **Ndung'u T, Weiss RA.** 2012. On HIV diversity. *AIDS* **26**:1255-1260.
 32. **Salemi M.** 2013. The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. *Infect Dis Rep* **5**:e3.
 33. **Smyth RP, Davenport MP, Mak J.** 2012. The origin of genetic diversity in HIV-1. *Virus Res* **169**:415-429.
 34. **Wolfs TF, Zwart G, Bakker M, Goudsmit J.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* **189**:103-110.
 35. **Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, Rosenblum LL, Kunstman KJ, Furtado MR, Munoz JL.** 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* **255**:1134-1137.

36. **Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, Ho DD.** 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* **261**:1179-1181.
37. **Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N, Salazar-Gonzalez JF, Bhattacharya T, Chu H, Hoffman I, Galvin S, Mapanje C, Kazembe P, Thebus R, Fiscus S, Hide W, Cohen MS, Karim SA, Haynes BF, Shaw GM, Hahn BH, Korber BT, Swanstrom R, Williamson C, Team CAIS, Center for HIVAVIC.** 2009. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J Virol* **83**:3556-3567.
38. **Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM.** 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* **105**:7552-7557.
39. **Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, Manigart O, Mulenga J, Keele BF, Shaw GM, Hahn BH, Allen SA, Derdeyn CA, Hunter E.** 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog* **5**:e1000274.
40. **Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, Sun C, Wang S, Learn GH, Morgan CJ, Schumacher JE, Haynes BF, Keele BF, Hahn BH, Shaw GM.** 2010. Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J Virol* **84**:6241-6247.
41. **Domingo E, Sheldon J, Perales C.** 2012. Viral quasispecies evolution. *Microbiol Mol Biol Rev* **76**:159-216.
42. **Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI.** 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* **73**:10489-10502.
43. **Rambaut A, Posada D, Crandall KA, Holmes EC.** 2004. The causes and consequences of HIV evolution. *Nat Rev Genet* **5**:52-61.
44. **Mani I, Gilbert P, Sankale JL, Eisen G, Mboup S, Kanki PJ.** 2002. Inpatient diversity and its correlation with viral setpoint in human immunodeficiency virus type 1 CRF02_AG-IbNG infection. *J Virol* **76**:10745-10755.
45. **Shankarappa R, Gupta P, Learn GH, Jr., Rodrigo AG, Rinaldo CR, Jr., Gorry MC, Mullins JI, Nara PL, Ehrlich GD.** 1998. Evolution of human

- immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. *Virology* **241**:251-259.
46. **Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**:1582-1586.
 47. **Piatak M, Jr., Saag MS, Yang LC, Clark SJ, Kappes JC, Luk KC, Hahn BH, Shaw GM, Lifson JD.** 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**:1749-1754.
 48. **Fu YX.** 2001. Estimating mutation rate and generation time from longitudinal samples of DNA sequences. *Mol Biol Evol* **18**:620-626.
 49. **Rodrigo AG, Shpaer EG, Delwart EL, Iversen AK, Gallo MV, Brojatsch J, Hirsch MS, Walker BD, Mullins JI.** 1999. Coalescent estimates of HIV-1 generation time in vivo. *Proc Natl Acad Sci U S A* **96**:2187-2191.
 50. **Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, Lifson JD, Bonhoeffer S, Nowak MA, Hahn BH, et al.** 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**:117-122.
 51. **Delviks-Frankenberry K, Galli A, Nikolaitchik O, Mens H, Pathak VK, Hu WS.** 2011. Mechanisms and factors that influence high frequency retroviral recombination. *Viruses* **3**:1650-1680.
 52. **Hu WS, Hughes SH.** 2012. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**.
 53. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**:5087-5094.
 54. **Abram ME, Ferris AL, Das K, Quinones O, Shao W, Tuske S, Alvord WG, Arnold E, Hughes SH.** 2014. Mutations in HIV-1 Reverse Transcriptase Affect the Errors Made in a Single Cycle of Viral Replication. *J Virol* **88**:7589-7601.
 55. **Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH.** 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* **84**:9864-9878.
 56. **Huang KJ, Wooley DP.** 2005. A new cell-based assay for measuring the forward mutation rate of HIV-1. *J Virol Methods* **124**:95-104.
 57. **Mansky LM.** 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res Hum Retroviruses* **12**:307-314.
 58. **Mansky LM.** 1996. The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**:391-400.
 59. **O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD.** 2002. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J Biol Chem* **277**:38053-38061.

60. **Schlub TE, Grimm AJ, Smyth RP, Cromer D, Chopra A, Mallal S, Venturi V, Waugh C, Mak J, Davenport MP.** 2014. Fifteen to twenty percent of HIV substitution mutations are associated with recombination. *J Virol* **88**:3837-3849.
61. **Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R.** 2010. Viral mutation rates. *J Virol* **84**:9733-9748.
62. **McCulloch SD, Kunkel TA.** 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**:148-161.
63. **Chen J, Nikolaitchik O, Singh J, Wright A, Bencsics CE, Coffin JM, Ni N, Lockett S, Pathak VK, Hu WS.** 2009. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc Natl Acad Sci U S A* **106**:13535-13540.
64. **Hu WS, Temin HM.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**:1556-1560.
65. **Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP.** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol* **74**:1234-1240.
66. **Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, Preston BD, Dougherty JP.** 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* **76**:11273-11282.
67. **Rhodes TD, Nikolaitchik O, Chen J, Powell D, Hu WS.** 2005. Genetic recombination of human immunodeficiency virus type 1 in one round of viral replication: effects of genetic distance, target cells, accessory genes, and lack of high negative interference in crossover events. *J Virol* **79**:1666-1677.
68. **Schlub TE, Smyth RP, Grimm AJ, Mak J, Davenport MP.** 2010. Accurately measuring recombination between closely related HIV-1 genomes. *PLoS Comput Biol* **6**:e1000766.
69. **Rhodes T, Wargo H, Hu WS.** 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. *J Virol* **77**:11193-11200.
70. **Onafuwa A, An W, Robson ND, Telesnitsky A.** 2003. Human immunodeficiency virus type 1 genetic recombination is more frequent than that of Moloney murine leukemia virus despite similar template switching rates. *J Virol* **77**:4577-4587.
71. **van Wamel JL, Berkhout B.** 1998. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. *Virology* **244**:245-251.

72. **Yu H, Jetzt AE, Ron Y, Preston BD, Dougherty JP.** 1998. The nature of human immunodeficiency virus type 1 strand transfers. *J Biol Chem* **273**:28384-28391.
73. **Svarovskaia ES, Delviks KA, Hwang CK, Pathak VK.** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. *J Virol* **74**:7171-7178.
74. **Hwang CK, Svarovskaia ES, Pathak VK.** 2001. 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* **98**:12209-12214.
75. **Pfeiffer JK, Topping RS, Shin NH, Telesnitsky A.** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription. *J Virol* **73**:8441-8447.
76. **Nikolenko GN, Svarovskaia ES, Delviks KA, Pathak VK.** 2004. Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. *J Virol* **78**:8761-8770.
77. **Operario DJ, Balakrishnan M, Bambara RA, Kim B.** 2006. Reduced dNTP interaction of human immunodeficiency virus type 1 reverse transcriptase promotes strand transfer. *J Biol Chem* **281**:32113-32121.
78. **Galetto R, Moumen A, Giacomoni V, Veron M, Charneau P, Negroni M.** 2004. The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *J Biol Chem* **279**:36625-36632.
79. **DeStefano JJ, Mallaber LM, Rodriguez-Rodriguez L, Fay PJ, Bambara RA.** 1992. Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J Virol* **66**:6370-6378.
80. **Wu W, Blumberg BM, Fay PJ, Bambara RA.** 1995. Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation. *J Biol Chem* **270**:325-332.
81. **Brincat JL, Pfeiffer JK, Telesnitsky A.** 2002. RNase H activity is required for high-frequency repeat deletion during Moloney murine leukemia virus replication. *J Virol* **76**:88-95.
82. **Chen J, Dang Q, Unutmaz D, Pathak VK, Maldarelli F, Powell D, Hu WS.** 2005. Mechanisms of nonrandom human immunodeficiency virus type 1 infection and double infection: preference in virus entry is important but is not the sole factor. *J Virol* **79**:4140-4149.
83. **Dang Q, Chen J, Unutmaz D, Coffin JM, Pathak VK, Powell D, KewalRamani VN, Maldarelli F, Hu WS.** 2004. Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. *Proc Natl Acad Sci U S A* **101**:632-637.

84. **Del Portillo A, Tripodi J, Najfeld V, Wodarz D, Levy DN, Chen BK.** 2011. Multiploid inheritance of HIV-1 during cell-to-cell infection. *J Virol* **85**:7169-7176.
85. **Levy DN, Aldrovandi GM, Kutsch O, Shaw GM.** 2004. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci U S A* **101**:4204-4209.
86. **Bregnard C, Pacini G, Danos O, Basmaciogullari S.** 2012. Suboptimal provirus expression explains apparent nonrandom cell coinfection with HIV-1. *J Virol* **86**:8810-8820.
87. **Nethe M, Berkhout B, van der Kuyl AC.** 2005. Retroviral superinfection resistance. *Retrovirology* **2**:52.
88. **Chohan B, Lavreys L, Rainwater SM, Overbaugh J.** 2005. Evidence for frequent reinfection with human immunodeficiency virus type 1 of a different subtype. *J Virol* **79**:10701-10708.
89. **Piantadosi A, Chohan B, Chohan V, McClelland RS, Overbaugh J.** 2007. Chronic HIV-1 infection frequently fails to protect against superinfection. *PLoS Pathog* **3**:e177.
90. **Piantadosi A, Ngayo MO, Chohan B, Overbaugh J.** 2008. Examination of a second region of the HIV type 1 genome reveals additional cases of superinfection. *AIDS Res Hum Retroviruses* **24**:1221.
91. **Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, Ricklefs SM, Tobian AA, Xiao C, Grabowski MK, Nalugoda F, Kigozi G, Laeyendecker O, Kagaayi J, Sewankambo N, Gray RH, Porcella SF, Wawer MJ, Quinn TC.** 2012. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J Infect Dis* **206**:267-274.
92. **Ronen K, McCoy CO, Matsen FA, Boyd DF, Emery S, Odem-Davis K, Jaoko W, Mandaliya K, McClelland RS, Richardson BA, Overbaugh J.** 2013. HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLoS Pathog* **9**:e1003593.
93. **Smith DM, Wong JK, Hightower GK, Ignacio CC, Koelsch KK, Daar ES, Richman DD, Little SJ.** 2004. Incidence of HIV superinfection following primary infection. *JAMA* **292**:1177-1178.
94. **Gratton S, Cheynier R, Dumaurier MJ, Oksenhendler E, Wain-Hobson S.** 2000. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers. *Proc Natl Acad Sci U S A* **97**:14566-14571.
95. **Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, Fischer U, Meese E, Wain-Hobson S, Meyerhans A.** 2002. Recombination: Multiply infected spleen cells in HIV patients. *Nature* **418**:144.
96. **Schultz A, Sopper S, Sauermann U, Meyerhans A, Suspene R.** 2012. Stable multi-infection of splenocytes during SIV infection--the basis for continuous recombination. *Retrovirology* **9**:31.
97. **Josefsson L, King MS, Makitalo B, Brannstrom J, Shao W, Maldarelli F, Kearney MF, Hu WS, Chen J, Gaines H, Mellors JW, Albert J, Coffin**

- JM, Palmer SE.** 2011. Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc Natl Acad Sci U S A* **108**:11199-11204.
98. **Josefsson L, Palmer S, Faria NR, Lemey P, Casazza J, Ambrozak D, Kearney M, Shao W, Kottlil S, Sneller M, Mellors J, Coffin JM, Maldarelli F.** 2013. Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS Pathog* **9**:e1003432.
99. **Chin MP, Rhodes TD, Chen J, Fu W, Hu WS.** 2005. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci U S A* **102**:9002-9007.
100. **Chin MP, Chen J, Nikolaitchik OA, Hu WS.** 2007. Molecular determinants of HIV-1 intersubtype recombination potential. *Virology* **363**:437-446.
101. **Chin MP, Lee SK, Chen J, Nikolaitchik OA, Powell DA, Fivash MJ, Jr., Hu WS.** 2008. Long-range recombination gradient between HIV-1 subtypes B and C variants caused by sequence differences in the dimerization initiation signal region. *J Mol Biol* **377**:1324-1333.
102. **Nikolaitchik OA, Galli A, Moore MD, Pathak VK, Hu WS.** 2011. Multiple barriers to recombination between divergent HIV-1 variants revealed by a dual-marker recombination assay. *J Mol Biol* **407**:521-531.
103. **An W, Telesnitsky A.** 2002. Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J Virol* **76**:7897-7902.
104. **Kellam P, Larder BA.** 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *J Virol* **69**:669-674.
105. **Moutouh L, Corbeil J, Richman DD.** 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc Natl Acad Sci U S A* **93**:6106-6111.
106. **Kim EY, Busch M, Abel K, Fritts L, Bustamante P, Stanton J, Lu D, Wu S, Glowczwskie J, Rourke T, Bogdan D, Piatak M, Jr., Lifson JD, Desrosiers RC, Wolinsky S, Miller CJ.** 2005. Retroviral recombination in vivo: viral replication patterns and genetic structure of simian immunodeficiency virus (SIV) populations in rhesus macaques after simultaneous or sequential intravaginal inoculation with SIVmac239Deltavpx/Deltavpr and SIVmac239Deltanef. *J Virol* **79**:4886-4895.
107. **Quan Y, Liang C, Brenner BG, Wainberg MA.** 2009. Multidrug-resistant variants of HIV type 1 (HIV-1) can exist in cells as defective quasispecies and be rescued by superinfection with other defective HIV-1 variants. *J Infect Dis* **200**:1479-1483.
108. **Quan Y, Xu H, Wainberg MA.** 2014. Defective HIV-1 quasispecies in the form of multiply drug-resistant proviral DNA within cells can be rescued by

- superinfection with different subtype variants of HIV-1 and by HIV-2 and SIV. *J Antimicrob Chemother* **69**:21-27.
109. **Galli A, Kearney M, Nikolaitchik OA, Yu S, Chin MP, Maldarelli F, Coffin JM, Pathak VK, Hu WS.** 2010. Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J Virol* **84**:7651-7661.
 110. **Hemelaar J, Gouws E, Ghys PD, Osmanov S.** 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**:679-689.
 111. **Menendez-Arias L.** 2009. Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses* **1**:1137-1165.
 112. **Mansky LM, Bernard LC.** 2000. 3'-Azido-3'-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J Virol* **74**:9532-9539.
 113. **Mansky LM, Le Rouzic E, Benichou S, Gajary LC.** 2003. Influence of reverse transcriptase variants, drugs, and Vpr on human immunodeficiency virus type 1 mutant frequencies. *J Virol* **77**:2071-2080.
 114. **Weiss KK, Chen R, Skasko M, Reynolds HM, Lee K, Bambara RA, Mansky LM, Kim B.** 2004. A role for dNTP binding of human immunodeficiency virus type 1 reverse transcriptase in viral mutagenesis. *Biochemistry* **43**:4490-4500.
 115. **Chen R, Yokoyama M, Sato H, Reilly C, Mansky LM.** 2005. Human immunodeficiency virus mutagenesis during antiviral therapy: impact of drug-resistant reverse transcriptase and nucleoside and nonnucleoside reverse transcriptase inhibitors on human immunodeficiency virus type 1 mutation frequencies. *J Virol* **79**:12045-12057.
 116. **Mansky LM.** 2003. Mutagenic outcome of combined antiviral drug treatment during human immunodeficiency virus type 1 replication. *Virology* **307**:116-121.
 117. **Clouser CL, Patterson SE, Mansky LM.** 2010. Exploiting drug repositioning for discovery of a novel HIV combination therapy. *J Virol* **84**:9301-9309.
 118. **Dapp MJ, Clouser CL, Patterson S, Mansky LM.** 2009. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J Virol* **83**:11950-11958.
 119. **Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI.** 1999. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc Natl Acad Sci U S A* **96**:1492-1497.
 120. **Rawson JM, Heineman RH, Beach LB, Martin JL, Schnettler EK, Dapp MJ, Patterson SE, Mansky LM.** 2013. 5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors. *Bioorg Med Chem* **21**:7222-7228.
 121. **Bebenek K, Abbotts J, Roberts JD, Wilson SH, Kunkel TA.** 1989. Specificity and mechanism of error-prone replication by human

- immunodeficiency virus-1 reverse transcriptase. *J Biol Chem* **264**:16948-16956.
122. **Refsland EW, Harris RS.** 2013. The APOBEC3 family of retroelement restriction factors. *Curr Top Microbiol Immunol* **371**:1-27.
 123. **Desimmie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK.** 2014. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol* **426**:1220-1245.
 124. **Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, Richman D, Coffin JM, Landau NR.** 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* **11**:435-442.
 125. **Conticello SG, Harris RS, Neuberger MS.** 2003. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* **13**:2009-2013.
 126. **Marin M, Rose KM, Kozak SL, Kabat D.** 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* **9**:1398-1403.
 127. **Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D.** 2004. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J Biol Chem* **279**:7792-7798.
 128. **Sheehy AM, Gaddis NC, Malim MH.** 2003. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **9**:1404-1407.
 129. **Jager S, Kim DY, Hultquist JF, Shindo K, LaRue RS, Kwon E, Li M, Anderson BD, Yen L, Stanley D, Mahon C, Kane J, Franks-Skiba K, Cimermancic P, Burlingame A, Sali A, Craik CS, Harris RS, Gross JD, Krogan NJ.** 2012. Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481**:371-375.
 130. **Zhang W, Du J, Evans SL, Yu Y, Yu XF.** 2012. T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction. *Nature* **481**:376-379.
 131. **Kobayashi M, Takaori-Kondo A, Miyauchi Y, Iwai K, Uchiyama T.** 2005. Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. *J Biol Chem* **280**:18573-18578.
 132. **Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF.** 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**:1056-1060.
 133. **Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, Brown WL, Reilly C, Emerman M, Harris RS.** 2014. Natural Polymorphisms in Human APOBEC3H and HIV-1 Vif Combine in Primary T Lymphocytes to Affect Viral G-to-A Mutation Levels and Infectivity. *PLoS Genet* **10**:e1004761.

134. **Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS.** 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**:11220-11234.
135. **Refsland EW, Hultquist JF, Harris RS.** 2012. Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n. *PLoS Pathog* **8**:e1002800.
136. **Chaipan C, Smith JL, Hu WS, Pathak VK.** 2013. APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol* **87**:444-453.
137. **Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH.** 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**:803-809.
138. **Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D.** 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**:99-103.
139. **Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L.** 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**:94-98.
140. **Doehle BP, Schafer A, Cullen BR.** 2005. Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. *Virology* **339**:281-288.
141. **Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH.** 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**:1392-1396.
142. **Liddament MT, Brown WL, Schumacher AJ, Harris RS.** 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr Biol* **14**:1385-1391.
143. **Harari A, Ooms M, Mulder LC, Simon V.** 2009. Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. *J Virol* **83**:295-303.
144. **Binka M, Ooms M, Steward M, Simon V.** 2012. The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. *J Virol* **86**:49-59.
145. **Simon V, Zennou V, Murray D, Huang Y, Ho DD, Bieniasz PD.** 2005. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* **1**:e6.
146. **Iwabu Y, Kinomoto M, Tatsumi M, Fujita H, Shimura M, Tanaka Y, Ishizaka Y, Nolan D, Mallal S, Sata T, Tokunaga K.** 2010. Differential anti-APOBEC3G activity of HIV-1 Vif proteins derived from different subtypes. *J Biol Chem* **285**:35350-35358.
147. **Fourati S, Lambert-Niclot S, Soulie C, Wirlden M, Malet I, Valantin MA, Tubiana R, Simon A, Katlama C, Carcelain G, Calvez V, Marcelin AG.**

2014. Differential impact of APOBEC3-driven mutagenesis on HIV evolution in diverse anatomical compartments. *AIDS* **28**:487-491.
148. **Janini M, Rogers M, Birx DR, McCutchan FE.** 2001. Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. *J Virol* **75**:7973-7986.
149. **Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, Siliciano RF.** 2005. G->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. *J Virol* **79**:1975-1980.
150. **Kijak GH, Janini LM, Tovanabutra S, Sanders-Buell E, Arroyo MA, Robb ML, Michael NL, Birx DL, McCutchan FE.** 2008. Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells. *Virology* **376**:101-111.
151. **Pace C, Keller J, Nolan D, James I, Gaudieri S, Moore C, Mallal S.** 2006. Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. *J Virol* **80**:9259-9269.
152. **Wei M, Xing H, Hong K, Huang H, Tang H, Qin G, Shao Y.** 2004. Biased G-to-A hypermutation in HIV-1 proviral DNA from a long-term non-progressor. *AIDS* **18**:1863-1865.
153. **Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, Daniels M, Ferrari G, Haynes BF, McMichael A, Shaw GM, Hahn BH, Korber B, Seoighe C.** 2009. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog* **5**:e1000414.
154. **Armitage AE, Deforche K, Welch JJ, Van Laethem K, Camacho R, Rambaut A, Iversen AK.** 2014. Possible footprints of APOBEC3F and/or other APOBEC3 deaminases, but not APOBEC3G, on HIV-1 from patients with acute/early and chronic infections. *J Virol* **88**:12882-12894.
155. **Russell RA, Moore MD, Hu WS, Pathak VK.** 2009. APOBEC3G induces a hypermutation gradient: purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA. *Retrovirology* **6**:16.
156. **Sadler HA, Stenglein MD, Harris RS, Mansky LM.** 2010. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J Virol* **84**:7396-7404.
157. **Mulder LC, Harari A, Simon V.** 2008. Cytidine deamination induced HIV-1 drug resistance. *Proc Natl Acad Sci U S A* **105**:5501-5506.
158. **Kim EY, Bhattacharya T, Kunstman K, Swantek P, Koning FA, Malim MH, Wolinsky SM.** 2010. Human APOBEC3G-mediated editing can

- promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J Virol* **84**:10402-10405.
159. **Sato K, Takeuchi JS, Misawa N, Izumi T, Kobayashi T, Kimura Y, Iwami S, Takaori-Kondo A, Hu WS, Aihara K, Ito M, An DS, Pathak VK, Koyanagi Y.** 2014. APOBEC3D and APOBEC3F potentially promote HIV-1 diversification and evolution in humanized mouse model. *PLoS Pathog* **10**:e1004453.
 160. **Kim EY, Lorenzo-Redondo R, Little SJ, Chung YS, Phalora PK, Maljkovic Berry I, Archer J, Penugonda S, Fischer W, Richman DD, Bhattacharya T, Malim MH, Wolinsky SM.** 2014. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. *PLoS Pathog* **10**:e1004281.
 161. **Chun TW, Nickle DC, Justement JS, Meyers JH, Roby G, Hallahan CW, Kottlilil S, Moir S, Mican JM, Mullins JI, Ward DJ, Kovacs JA, Mannon PJ, Fauci AS.** 2008. Persistence of HIV in gut-associated lymphoid tissue despite long-term antiretroviral therapy. *J Infect Dis* **197**:714-720.
 162. **Gras G, Kaul M.** 2010. Molecular mechanisms of neuroinvasion by monocytes-macrophages in HIV-1 infection. *Retrovirology* **7**:30.
 163. **Ruelas DS, Greene WC.** 2013. An integrated overview of HIV-1 latency. *Cell* **155**:519-529.
 164. **Pursuing Later Treatment Options Ilpt, Collaboration of Observational HIVREG, Nakagawa F, Lodwick R, Costagliola D, van Sighem A, Torti C, Podzamczar D, Mocroft A, Ledergerber B, Dorrucchi M, Cozzi-Lepri A, Jansen K, Masquelier B, Garcia F, De Wit S, Stephan C, Obel N, Fatkenhaeuer G, Castagna A, Sambatakou H, Mussini C, Ghosn J, Zangerle R, Duval X, Meyer L, Perez-Hoyos S, Fabre Colin C, Kjaer J, Chene G, Grarup J, Phillips A.** 2012. Calendar time trends in the incidence and prevalence of triple-class virologic failure in antiretroviral drug-experienced people with HIV in Europe. *J Acquir Immune Defic Syndr* **59**:294-299.
 165. **Galli A, Bukh J.** 2014. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends Microbiol* **22**:354-364.
 166. **Lin CC, Yang ZW, lang SB, Chao M.** 2014. Reduced genetic distance and high replication levels increase the RNA recombination rate of hepatitis delta virus. *Virus Res* doi:10.1016/j.virusres.2014.08.011.
 167. **Lowry K, Woodman A, Cook J, Evans DJ.** 2014. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates. *PLoS Pathog* **10**:e1004191.
 168. **Biebricher CK, Eigen M.** 2006. What is a quasispecies? *Curr Top Microbiol Immunol* **299**:1-31.

169. **Batschelet E, Domingo E, Weissmann C.** 1976. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* **1**:27-32.
170. **Domingo E, Sabo D, Taniguchi T, Weissmann C.** 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**:735-744.
171. **Valentine RC, Ward R, Strand M.** 1969. The replication cycle of RNA bacteriophages. *Adv Virus Res* **15**:1-59.
172. **Cavrois M, Gessain A, Wain-Hobson S, Wattel E.** 1996. Proliferation of HTLV-1 infected circulating cells in vivo in all asymptomatic carriers and patients with TSP/HAM. *Oncogene* **12**:2419-2423.
173. **Cavrois M, Wain-Hobson S, Gessain A, Plumelle Y, Wattel E.** 1996. Adult T-cell leukemia/lymphoma on a background of clonally expanding human T-cell leukemia virus type-1-positive cells. *Blood* **88**:4646-4650.
174. **Gessain A, Gallo RC, Franchini G.** 1992. Low degree of human T-cell leukemia/lymphoma virus type I genetic drift in vivo as a means of monitoring viral transmission and movement of ancient human populations. *J Virol* **66**:2288-2295.
175. **Pedroza Martins L, Chenciner N, Wain-Hobson S.** 1992. Complex inpatient sequence variation in the V1 and V2 hypervariable regions of the HIV-1 gp 120 envelope sequence. *Virology* **191**:837-845.
176. **Biebricher CK, Eigen M.** 2005. The error threshold. *Virus Res* **107**:117-127.
177. **Eigen M.** 2002. Error catastrophe and antiviral strategy. *Proc Natl Acad Sci U S A* **99**:13374-13376.
178. **Holland JJ, Domingo E, de la Torre JC, Steinhauer DA.** 1990. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J Virol* **64**:3960-3962.
179. **Perales C, Martin V, Domingo E.** 2011. Lethal mutagenesis of viruses. *Curr Opin Virol* **1**:419-422.
180. **Matsui M, Shindo K, Izumi T, Ito K, Shinohara M, Komano J, Kobayashi M, Kadowaki N, Harris RS, Takaori-Kondo A.** 2014. Small molecules that inhibit Vif-induced degradation of APOBEC3G. *Virol J* **11**:122.
181. **Nathans R, Cao H, Sharova N, Ali A, Sharkey M, Stranska R, Stevenson M, Rana TM.** 2008. Small-molecule inhibition of HIV-1 Vif. *Nat Biotechnol* **26**:1187-1192.
182. **Zuo T, Liu D, Lv W, Wang X, Wang J, Lv M, Huang W, Wu J, Zhang H, Jin H, Zhang L, Kong W, Yu X.** 2012. Small-molecule inhibition of human immunodeficiency virus type 1 replication by targeting the interaction between Vif and ElonginC. *J Virol* **86**:5497-5507.
183. **Pery E, Sheehy A, Nebane NM, Brazier AJ, Misra V, Rajendran KS, Buhrlage SJ, Mankowski MK, Rasmussen L, White EL, Ptak RG, Gabuzda D.** 2015. Identification of a novel HIV-1 inhibitor targeting Vif-

- dependent degradation of human APOBEC3G protein. *J Biol Chem* **290**:10504-10517.
184. **Asahina Y, Izumi N, Enomoto N, Uchihara M, Kurosaki M, Onuki Y, Nishimura Y, Ueda K, Tsuchiya K, Nakanishi H, Kitamura T, Miyake S.** 2005. Mutagenic effects of ribavirin and response to interferon/ribavirin combination therapy in chronic hepatitis C. *J Hepatol* **43**:623-629.
 185. **Chevaliez S, Brillet R, Lazaro E, Hezode C, Pawlotsky JM.** 2007. Analysis of ribavirin mutagenicity in human hepatitis C virus infection. *J Virol* **81**:7732-7741.
 186. **Cuevas JM, Gonzalez-Candelas F, Moya A, Sanjuan R.** 2009. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. *J Virol* **83**:5760-5764.
 187. **Dietz J, Schelhorn SE, Fitting D, Mihm U, Susser S, Welker MW, Fuller C, Daumer M, Teuber G, Wedemeyer H, Berg T, Lengauer T, Zeuzem S, Herrmann E, Sarrazin C.** 2013. Deep sequencing reveals mutagenic effects of ribavirin during monotherapy of hepatitis C virus genotype 1-infected patients. *J Virol* **87**:6172-6181.
 188. **Furuta Y, Gowen BB, Takahashi K, Shiraki K, Smee DF, Barnard DL.** 2013. Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Res* **100**:446-454.
 189. **Arias A, Thorne L, Goodfellow I.** 2014. Favipiravir elicits antiviral mutagenesis during virus replication in vivo. *Elife* **3**:e03679.
 190. **Baranovich T, Wong SS, Armstrong J, Marjuki H, Webby RJ, Webster RG, Govorkova EA.** 2013. T-705 (favipiravir) induces lethal mutagenesis in influenza A H1N1 viruses in vitro. *J Virol* **87**:3741-3751.
 191. **Furuta Y, Takahashi K, Kuno-Maekawa M, Sangawa H, Uehara S, Kozaki K, Nomura N, Egawa H, Shiraki K.** 2005. Mechanism of action of T-705 against influenza virus. *Antimicrob Agents Chemother* **49**:981-986.
 192. **Jin Z, Smith LK, Rajwanshi VK, Kim B, Deval J.** 2013. The ambiguous base-pairing and high substrate efficiency of T-705 (Favipiravir) Ribofuranosyl 5'-triphosphate towards influenza A virus polymerase. *PLoS One* **8**:e68347.
 193. **Sangawa H, Komeno T, Nishikawa H, Yoshida A, Takahashi K, Nomura N, Furuta Y.** 2013. Mechanism of action of T-705 ribosyl triphosphate against influenza virus RNA polymerase. *Antimicrob Agents Chemother* **57**:5202-5208.
 194. **Tapia N, Fernandez G, Parera M, Gomez-Mariano G, Clotet B, Quinones-Mateu M, Domingo E, Martinez MA.** 2005. Combination of a mutagenic agent with a reverse transcriptase inhibitor results in systematic inhibition of HIV-1 infection. *Virology* **338**:1-8.
 195. **Beisler JA.** 1978. Isolation, characterization, and properties of a labile hydrolysis product of the antitumor nucleoside, 5-azacytidine. *J Med Chem* **21**:204-208.

196. **Jackson-Grusby L, Laird PW, Magge SN, Moeller BJ, Jaenisch R.** 1997. Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. *Proc Natl Acad Sci U S A* **94**:4681-4685.
197. **Rogstad DK, Herring JL, Theruvathu JA, Burdzy A, Perry CC, Neidigh JW, Sowers LC.** 2009. Chemical decomposition of 5-aza-2'-deoxycytidine (Decitabine): kinetic analyses and identification of products by NMR, HPLC, and mass spectrometry. *Chem Res Toxicol* **22**:1194-1204.
198. **Harris KS, Brabant W, Styrchak S, Gall A, Daifuku R.** 2005. KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. *Antiviral Res* **67**:1-9.
199. **Hicks C, Clay P, Redfield R, Lalezari J, Liporace R, Schneider S, Sension M, McRae M, Laurent JP.** 2013. Safety, tolerability, and efficacy of KP-1461 as monotherapy for 124 days in antiretroviral-experienced, HIV type 1-infected subjects. *AIDS Res Hum Retroviruses* **29**:250-255.
200. **Mullins JI, Heath L, Hughes JP, Kicha J, Styrchak S, Wong KG, Rao U, Hansen A, Harris KS, Laurent JP, Li D, Simpson JH, Essigmann JM, Loeb LA, Parkins J.** 2011. Mutation of HIV-1 genomes in a clinical population treated with the mutagenic nucleoside KP1461. *PLoS One* **6**:e15135.
201. **Clouser CL, Chauhan J, Bess MA, van Oploo JL, Zhou D, Dimick-Gray S, Mansky LM, Patterson SE.** 2012. Anti-HIV-1 activity of resveratrol derivatives and synergistic inhibition of HIV-1 by the combination of resveratrol and decitabine. *Bioorg Med Chem Lett* **22**:6642-6646.
202. **Georgiou NA, van der Bruggen T, Oudshoorn M, Nottet HS, Marx JJ, van Asbeck BS.** 2000. Inhibition of human immunodeficiency virus type 1 replication in human mononuclear blood cells by the iron chelators deferoxamine, deferiprone, and bleomycin. *J Infect Dis* **181**:484-490.
203. **Heredia A, Davis C, Redfield R.** 2000. Synergistic inhibition of HIV-1 in activated and resting peripheral blood mononuclear cells, monocyte-derived macrophages, and selected drug-resistant isolates with nucleoside analogues combined with a natural product, resveratrol. *J Acquir Immune Defic Syndr* **25**:246-255.
204. **Malley SD, Grange JM, Hamedi-Sangsari F, Vila JR.** 1994. Synergistic anti-human immunodeficiency virus type 1 effect of hydroxamate compounds with 2',3'-dideoxyinosine in infected resting human lymphocytes. *Proc Natl Acad Sci U S A* **91**:11017-11021.
205. **Gao WY, Cara A, Gallo RC, Lori F.** 1993. Low levels of deoxynucleotides in peripheral blood lymphocytes: a strategy to inhibit human immunodeficiency virus type 1 replication. *Proc Natl Acad Sci U S A* **90**:8925-8928.
206. **Bakshi RP, Hamzeh F, Frank I, Eron JJ, Jr., Bosch RJ, Rosenkranz SL, Cramer YS, Ussery M, Flexner C.** 2007. Effect of hydroxyurea and

- dideoxyinosine on intracellular 3'-deoxyadenosine-5'-triphosphate concentrations in HIV-infected patients. *AIDS Res Hum Retroviruses* **23**:1360-1365.
207. **Frank I, Bosch RJ, Fiscus S, Valentine F, Flexner C, Segal Y, Ruan P, Gulick R, Wood K, Estep S, Fox L, Nevin T, Stevens M, Eron JJ, Jr., Team AP.** 2004. Activity, safety, and immunological effects of hydroxyurea added to didanosine in antiretroviral-naive and experienced HIV type 1-infected subjects: a randomized, placebo-controlled trial, ACTG 307. *AIDS Res Hum Retroviruses* **20**:916-926.
208. **Havir DV, Gilbert PB, Bennett K, Collier AC, Hirsch MS, Tebas P, Adams EM, Wheat LJ, Goodwin D, Schnittman S, Holohan MK, Richman DD, Group AS.** 2001. Effects of treatment intensification with hydroxyurea in HIV-infected patients with virologic suppression. *AIDS* **15**:1379-1388.
209. **Lori F, Pollard RB, Whitman L, Bakare N, Blick G, Shalit P, Foli A, Peterson D, Tennenberg A, Schrader S, Rashbaum B, Farthing C, Herman D, Norris D, Greiger P, Frank I, Groff A, Lova L, Asmuth D, Lisziewicz J.** 2005. Lowering the dose of hydroxyurea minimizes toxicity and maximizes anti-HIV potency. *AIDS Res Hum Retroviruses* **21**:263-272.
210. **Rutschmann OT, Vernazza PL, Bucher HC, Opravil M, Ledergerber B, Telenti A, Malinverni R, Bernasconi E, Fagard C, Leduc D, Perrin L, Hirschel B.** 2000. Long-term hydroxyurea in combination with didanosine and stavudine for the treatment of HIV-1 infection. Swiss HIV Cohort Study. *AIDS* **14**:2145-2151.
211. **Stebbing J, Nelson M, Orkin C, Mandalia S, Bower M, Pozniak A, Gazzard B.** 2004. A randomized trial to investigate the recycling of stavudine and didanosine with and without hydroxyurea in salvage therapy (RESTART). *J Antimicrob Chemother* **53**:501-505.
212. **Zala C, Salomon H, Ochoa C, Kijak G, Federico A, Perez H, Montaner JS, Cahn P.** 2002. Higher rate of toxicity with no increased efficacy when hydroxyurea is added to a regimen of stavudine plus didanosine and nevirapine in primary HIV infection. *J Acquir Immune Defic Syndr* **29**:368-373.
213. **Clouser CL, Holtz CM, Mullett M, Crankshaw DL, Briggs JE, O'Sullivan MG, Patterson SE, Mansky LM.** 2012. Activity of a novel combined antiretroviral therapy of gemcitabine and decitabine in a mouse model for HIV-1. *Antimicrob Agents Chemother* **56**:1942-1948.
214. **Heredia A, Davis CE, Reitz MS, Le NM, Wainberg MA, Foulke JS, Wang LX, Redfield RR.** 2013. Targeting of the purine biosynthesis host cell pathway enhances the activity of tenofovir against sensitive and drug-resistant HIV-1. *J Infect Dis* **208**:2085-2094.
215. **Campbell-Yesufu OT, Gandhi RT.** 2011. Update on human immunodeficiency virus (HIV)-2 infection. *Clin Infect Dis* **52**:780-787.

216. **Boyer PL, Sarafianos SG, Arnold E, Hughes SH.** 2000. Analysis of mutations at positions 115 and 116 in the dNTP binding site of HIV-1 reverse transcriptase. *Proc Natl Acad Sci U S A* **97**:3056-3061.
217. **Cases-Gonzalez CE, Gutierrez-Rivas M, Menendez-Arias L.** 2000. Coupling ribose selection to fidelity of DNA synthesis. The role of Tyr-115 of human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem* **275**:19759-19767.
218. **Nguyen LA, Domaal RA, Kennedy EM, Kim DH, Schinazi RF, Kim B.** 2015. Pre-steady state kinetic analysis of HIV-1 reverse transcriptase for non-canonical ribonucleoside triphosphate incorporation and DNA synthesis from ribonucleoside-containing DNA template. *Antiviral Res* **115**:75-82.
219. **Kennedy EM, Gavegnano C, Nguyen L, Slater R, Lucas A, Fromentin E, Schinazi RF, Kim B.** 2010. Ribonucleoside triphosphates as substrate of human immunodeficiency virus type 1 reverse transcriptase in human macrophages. *J Biol Chem* **285**:39380-39391.
220. **Kennedy EM, Amie SM, Bambara RA, Kim B.** 2012. Frequent incorporation of ribonucleotides during HIV-1 reverse transcription and their attenuated repair in macrophages. *J Biol Chem* **287**:14280-14288.
221. **Clay PG, McRae M, Laurent JP.** 2011. Safety, Tolerability, and Pharmacokinetics of KP-1461 in Phase I Clinical Studies: A Single Oral Dose Study in Non-HIV-Infected Adults, and a 14-Day Dose-Escalating Study in Highly Antiretroviral-Experienced HIV-Infected Adults. *J Int Assoc Physicians AIDS Care (Chic)* **10**:232-238.
222. **Schlub TE, Grimm AJ, Smyth RP, Cromer D, Chopra A, Mallal S, Venturi V, Waugh C, Mak J, Davenport MP.** 2014. 15-20% of HIV substitution mutations are associated with recombination. *J Virol* doi:10.1128/JVI.03136-13.
223. **Svarovskaia ES, Cheslock SR, Zhang WH, Hu WS, Pathak VK.** 2003. Retroviral mutation rates and reverse transcriptase fidelity. *Front Biosci* **8**:d117-134.
224. **Martinez MA, Vartanian JP, Wain-Hobson S.** 1994. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc Natl Acad Sci U S A* **91**:11787-11791.
225. **Holtz CM, Mansky LM.** 2013. Variation of HIV-1 mutation spectra among cell types. *J Virol* **87**:5296-5299.
226. **Coffey LL, Beeharry Y, Borderia AV, Blanc H, Vignuzzi M.** 2011. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc Natl Acad Sci U S A* **108**:16038-16043.
227. **Meng T, Kwang J.** 2014. Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. *J Virol* **88**:5803-5815.

228. **Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R.** 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**:344-348.
229. **Vignuzzi M, Wendt E, Andino R.** 2008. Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* **14**:154-161.
230. **Zeng J, Wang H, Xie X, Li C, Zhou G, Yang D, Yu L.** 2014. Ribavirin-resistant variants of foot-and-mouth disease virus: the effect of restricted quasispecies diversity on viral virulence. *J Virol* **88**:4008-4020.
231. **Matamoros T, Kim B, Menendez-Arias L.** 2008. Mechanistic insights into the role of Val75 of HIV-1 reverse transcriptase in misinsertion and mispair extension fidelity of DNA synthesis. *J Mol Biol* **375**:1234-1248.
232. **Alvarez M, Matamoros T, Menendez-Arias L.** 2009. Increased thermostability and fidelity of DNA synthesis of wild-type and mutant HIV-1 group O reverse transcriptases. *J Mol Biol* **392**:872-884.
233. **Ribeiro AC, Maia e Silva A, Santa-Marta M, Pombo A, Moniz-Pereira J, Goncalves J, Barahona I.** 2005. Functional analysis of Vif protein shows less restriction of human immunodeficiency virus type 2 by APOBEC3G. *J Virol* **79**:823-833.
234. **Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C.** 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* **43**:e37.
235. **Dohm JC, Lottaz C, Borodina T, Himmelbauer H.** 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**:e105.
236. **Minoche AE, Dohm JC, Himmelbauer H.** 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**:R112.
237. **Abnizova I, Leonard S, Skelly T, Brown A, Jackson D, Gourtovaia M, Qi G, Te Boekhorst R, Faruque N, Lewis K, Cox T.** 2012. Analysis of context-dependent errors for illumina sequencing. *J Bioinform Comput Biol* **10**:1241005.
238. **Judo MS, Wedel AB, Wilson C.** 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res* **26**:1819-1825.
239. **Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S, Davenport MP, Mak J.** 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**:45-51.
240. **Thompson JR, Marcelino LA, Polz MF.** 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res* **30**:2083-2088.
241. **Mansky LM, Preveral S, Selig L, Benarous R, Benichou S.** 2000. The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate. *J Virol* **74**:7039-7047.
242. **Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-UI-Amin M,**

- Ogasawara N, Kanaya S.** 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* **39**:e90.
243. **Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y.** 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**:341.
244. **Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB.** 2013. Characterizing and measuring bias in sequence data. *Genome Biol* **14**:R51.
245. **Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X.** 2015. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC Genomics* **16**:79.
246. **Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA.** 2012. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* **109**:14508-14513.
247. **Langlois MA, Beale RC, Conticello SG, Neuberger MS.** 2005. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* **33**:1913-1923.
248. **Piroozmand A, Yamamoto Y, Khamsri B, Fujita M, Uchiyama T, Adachi A.** 2007. Generation and characterization of APOBEC3G-positive 293T cells for HIV-1 Vif study. *J Med Invest* **54**:154-158.
249. **Bourara K, Liegler TJ, Grant RM.** 2007. Target cell APOBEC3C can induce limited G-to-A mutation in HIV-1. *PLoS Pathog* **3**:1477-1485.
250. **Smith JL, Izumi T, Borbet TC, Hagedorn AN, Pathak VK.** 2014. HIV-1 and HIV-2 Vif interact with human APOBEC3 proteins using completely different determinants. *J Virol* **88**:9893-9908.
251. **Holtz CM, Sadler HA, Mansky LM.** 2013. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. *Nucleic Acids Res* **41**:6139-6148.
252. **Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS.** 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol* **337**:585-596.
253. **Wiegand HL, Doehle BP, Bogerd HP, Cullen BR.** 2004. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J* **23**:2451-2458.
254. **Suspene R, Rusniok C, Vartanian JP, Wain-Hobson S.** 2006. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucleic Acids Res* **34**:4677-4684.
255. **Chen J, Powell D, Hu WS.** 2006. High frequency of genetic recombination is a common feature of primate lentivirus replication. *J Virol* **80**:9651-9658.

256. **Abada P, Noble B, Cannon PM.** 2005. Functional domains within the human immunodeficiency virus type 2 envelope protein required to enhance virus production. *J Virol* **79**:3627-3638.
257. **Vodicka MA, Goh WC, Wu LI, Rogel ME, Bartz SR, Schweickart VL, Raport CJ, Emerman M.** 1997. Indicator cell lines for detection of primary strains of human and simian immunodeficiency viruses. *Virology* **233**:193-198.
258. **Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, Demeneix B, Behr JP.** 1995. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc Natl Acad Sci U S A* **92**:7297-7301.
259. **Kutner RH, Zhang XY, Reiser J.** 2009. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat Protoc* **4**:495-505.
260. **Tsai H, Missirian V, Ngo KJ, Tran RK, Chan SR, Sundaresan V, Comai L.** 2013. Production of a high-efficiency TILLING population through polyploidization. *Plant Physiol* **161**:1604-1614.
261. **Wu TD, Nacu S.** 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**:873-881.
262. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**:1297-1303.
263. **Wolfinger R, O'Connell M.** 1993. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**:233-243.
264. **Bonnac LF, Mansky LM, Patterson SE.** 2013. Structure-activity relationships and design of viral mutagens and application to lethal mutagenesis. *J Med Chem* **56**:9403-9414.
265. **Furuta Y, Takahashi K, Fukuda Y, Kuno M, Kamiyama T, Kozaki K, Nomura N, Egawa H, Minami S, Watanabe Y, Narita H, Shiraki K.** 2002. In vitro and in vivo activities of anti-influenza virus compound T-705. *Antimicrob Agents Chemother* **46**:977-981.
266. **Clouser CL, Bonnac L, Mansky LM, Patterson SE.** 2013. Characterization of permeability, stability and anti-HIV-1 activity of decitabine and gemcitabine divaleryl prodrugs. *Antivir Chem Chemother* doi:10.3851/IMP2682.
267. **Rawson JML, S. R.; Reilly, C. S.; Mansky, L. M. .** 2015. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* **12**.
268. **Lin KT, Momparler RL, Rivard GE.** 1981. High-performance liquid chromatographic analysis of chemical stability of 5-aza-2'-deoxycytidine. *J Pharm Sci* **70**:1228-1232.

269. **Liu Z, Marcucci G, Byrd JC, Grever M, Xiao J, Chan KK.** 2006. Characterization of decomposition products and preclinical and low dose clinical pharmacokinetics of decitabine (5-aza-2'-deoxycytidine) by a new liquid chromatography/tandem mass spectrometry quantification method. *Rapid Commun Mass Spectrom* **20**:1117-1126.
270. **Lamparska K, Clark J, Babilonia G, Bedell V, Yip W, Smith SS.** 2012. 2'-Deoxyriboguanylurea, the primary breakdown product of 5-aza-2'-deoxyribocytidine, is a mutagen, an epimutagen, an inhibitor of DNA methyltransferases and an inducer of 5-azacytidine-type fragile sites. *Nucleic Acids Res* **40**:9788-9801.
271. **Quintas-Cardama A, Santos FP, Garcia-Manero G.** 2010. Therapy with azanucleosides for myelodysplastic syndromes. *Nat Rev Clin Oncol* **7**:433-444.
272. **Oz S, Raddatz G, Rius M, Blagitko-Dorfs N, Lubbert M, Maercker C, Lyko F.** 2014. Quantitative determination of decitabine incorporation into DNA and its effect on mutation rates in human cancer cells. *Nucleic Acids Res* **42**:e152.
273. **Maslov AY, Lee M, Gundry M, Gravina S, Strogonova N, Tazearslan C, Bendebury A, Suh Y, Vijg J.** 2012. 5-aza-2'-deoxycytidine-induced genome rearrangements are mediated by DNMT1. *Oncogene* **31**:5172-5179.
274. **Perales C, Domingo E.** 2015. Antiviral Strategies Based on Lethal Mutagenesis and Error Threshold. *Curr Top Microbiol Immunol* doi:10.1007/82_2015_459.
275. **Rawson JM, Landman SR, Reilly CS, Bonnac L, Patterson SE, Mansky LM.** 2015. Lack of mutational hotspots during decitabine-mediated HIV-1 mutagenesis. *Antimicrob Agents Chemother* doi:10.1128/AAC.01644-15.
276. **Li LH, Olin EJ, Buskirk HH, Reineke LM.** 1970. Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. *Cancer Res* **30**:2760-2769.
277. **Daddacha W, Noble E, Nguyen LA, Kennedy EM, Kim B.** 2013. Effect of ribonucleotides embedded in a DNA template on HIV-1 reverse transcription kinetics and fidelity. *J Biol Chem* **288**:12522-12532.
278. **Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundstrom EB, Johansson E, Chabes A, Kunkel TA.** 2010. Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* **6**:774-781.
279. **Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A, Kunkel TA.** 2010. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A* **107**:4949-4954.
280. **Bouchard J, Momparler RL.** 1983. Incorporation of 5-Aza-2'-deoxycytidine-5'-triphosphate into DNA. Interactions with mammalian DNA polymerase alpha and DNA methylase. *Mol Pharmacol* **24**:109-114.

281. **Covey JM, D'Incalci M, Tilchen EJ, Zaharko DS, Kohn KW.** 1986. Differences in DNA damage produced by incorporation of 5-aza-2'-deoxycytidine or 5,6-dihydro-5-azacytidine into DNA of mammalian cells. *Cancer Res* **46**:5511-5517.
282. **Diamond TL, Roshal M, Jamburuthugoda VK, Reynolds HM, Merriam AR, Lee KY, Balakrishnan M, Bambara RA, Planelles V, Dewhurst S, Kim B.** 2004. Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* **279**:51545-51553.
283. **Rawson JM, Landman SR, Reilly CS, Mansky LM.** 2015. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* **12**:60.
284. **Cohen S, Megherbi M, Jordheim LP, Lefebvre I, Perigaud C, Dumontet C, Guitton J.** 2009. Simultaneous analysis of eight nucleoside triphosphates in cell lines by liquid chromatography coupled with tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **877**:3831-3840.
285. **Pennings PS.** 2013. HIV Drug Resistance: Problems and Perspectives. *Infect Dis Rep* **5**:e5.
286. **Iyidogan P, Anderson KS.** 2014. Current perspectives on HIV-1 antiretroviral drug resistance. *Viruses* **6**:4095-4139.
287. **Looney D, Ma A, Johns S.** 2015. HIV therapy-the state of art. *Curr Top Microbiol Immunol* **389**:1-29.
288. **Gnadig NF, Beaucourt S, Campagnola G, Borderia AV, Sanz-Ramos M, Gong P, Blanc H, Peersen OB, Vignuzzi M.** 2012. Coxsackievirus B3 mutator strains are attenuated in vivo. *Proc Natl Acad Sci U S A* **109**:E2294-2303.
289. **Graham RL, Becker MM, Eckerle LD, Bolles M, Denison MR, Baric RS.** 2012. A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. *Nat Med* **18**:1820-1826.
290. **Korboukh VK, Lee CA, Acevedo A, Vignuzzi M, Xiao Y, Arnold JJ, Hemperly S, Graci JD, August A, Andino R, Cameron CE.** 2014. RNA virus population diversity, an optimum for maximal fitness and virulence. *J Biol Chem* **289**:29531-29544.
291. **Rozen-Gagnon K, Stapleford KA, Mongelli V, Blanc H, Failloux AB, Saleh MC, Vignuzzi M.** 2014. Alphavirus mutator variants present host-specific defects and attenuation in mammalian and insect models. *PLoS Pathog* **10**:e1003877.
292. **Xie X, Wang H, Zeng J, Li C, Zhou G, Yang D, Yu L.** 2014. Foot-and-mouth disease virus low-fidelity polymerase mutants are attenuated. *Arch Virol* **159**:2641-2650.
293. **Van Slyke GA, Arnold JJ, Lugo AJ, Griesemer SB, Moustafa IM, Kramer LD, Cameron CE, Ciota AT.** 2015. Sequence-Specific Fidelity

- Alterations Associated with West Nile Virus Attenuation in Mosquitoes. *PLoS Pathog* **11**:e1005009.
294. **Harris RS, Dudley JP.** 2015. APOBECs and virus restriction. *Virology* **479-480**:131-145.
 295. **Willems L, Gillet NA.** 2015. APOBEC3 Interference during Replication of Viral Genomes. *Viruses* **7**:2999-3018.
 296. **Shao J, Zhou B, Chu B, Yen Y.** 2006. Ribonucleotide reductase inhibitors and future drug design. *Curr Cancer Drug Targets* **6**:409-431.
 297. **Gao WY, Johns DG, Mitsuya H.** 1994. Anti-human immunodeficiency virus type 1 activity of hydroxyurea in combination with 2',3'-dideoxynucleosides. *Mol Pharmacol* **46**:767-772.
 298. **Bouchonnet F, Dam E, Mammano F, de Soultrait V, Hennere G, Benech H, Clavel F, Hance AJ.** 2005. Quantification of the effects on viral DNA synthesis of reverse transcriptase mutations conferring human immunodeficiency virus type 1 resistance to nucleoside analogues. *J Virol* **79**:812-822.
 299. **Beach LB, Rawson JM, Kim B, Patterson SE, Mansky LM.** 2014. Novel inhibitors of human immunodeficiency virus type 2 infectivity. *J Gen Virol* **95**:2778-2783.
 300. **Fontecave M, Lepoivre M, Elleingand E, Gerez C, Guittet O.** 1998. Resveratrol, a remarkable inhibitor of ribonucleotide reductase. *FEBS Lett* **421**:277-279.
 301. **Heinemann V, Xu YZ, Chubb S, Sen A, Hertel LW, Grindey GB, Plunkett W.** 1990. Inhibition of ribonucleotide reduction in CCRF-CEM cells by 2',2'-difluorodeoxycytidine. *Mol Pharmacol* **38**:567-572.
 302. **Parker WB, Shaddix SC, Chang CH, White EL, Rose LM, Brockman RW, Shortnacy AT, Montgomery JA, Secrist JA, 3rd, Bennett LL, Jr.** 1991. Effects of 2-chloro-9-(2-deoxy-2-fluoro-beta-D-arabinofuranosyl)adenine on K562 cellular metabolism and the inhibition of human ribonucleotide reductase and DNA polymerases by its 5'-triphosphate. *Cancer Res* **51**:2386-2394.
 303. **Xie KC, Plunkett W.** 1996. Deoxynucleotide pool depletion and sustained inhibition of ribonucleotide reductase and DNA synthesis after treatment of human lymphoblastoid cells with 2-chloro-9-(2-deoxy-2-fluoro-beta-D-arabinofuranosyl) adenine. *Cancer Res* **56**:3030-3037.
 304. **Moore EC, Hurlbert RB.** 1985. The inhibition of ribonucleoside diphosphate reductase by hydroxyurea, guanazole and pyrazoloimidazole (IMPY). *Pharmacol Ther* **27**:167-196.
 305. **Zhao L, Au JL, Wientjes MG.** 2010. Comparison of methods for evaluating drug-drug interaction. *Front Biosci (Elite Ed)* **2**:241-249.
 306. **Bijnsdorp IV, Giovannetti E, Peters GJ.** 2011. Analysis of drug interactions. *Methods Mol Biol* **731**:421-434.

307. **Tyrsted G.** 1982. Effect of hydroxyurea and 5-fluorodeoxyuridine on deoxyribonucleoside triphosphate pools early in phytohemagglutinin-stimulated human lymphocytes. *Biochem Pharmacol* **31**:3107-3113.
308. **Akerblom L, Reichard P.** 1985. Azidocytidine is a specific inhibitor of deoxyribonucleotide synthesis in 3T6 cells. *J Biol Chem* **260**:9197-9202.
309. **Bianchi V, Pontis E, Reichard P.** 1986. Changes of deoxyribonucleoside triphosphate pools induced by hydroxyurea and their relation to DNA synthesis. *J Biol Chem* **261**:16037-16042.
310. **Munir S, Thierry S, Subra F, Deprez E, Delelis O.** 2013. Quantitative analysis of the time-course of viral DNA forms during the HIV-1 life cycle. *Retrovirology* **10**:87.
311. **Sloan RD, Wainberg MA.** 2011. The role of unintegrated DNA in HIV infection. *Retrovirology* **8**:52.
312. **D'Incalci M, Covey JM, Zaharko DS, Kohn KW.** 1985. DNA alkali-labile sites induced by incorporation of 5-aza-2'-deoxycytidine into DNA of mouse leukemia L1210 cells. *Cancer Res* **45**:3197-3202.
313. **Limonta M, Colombo T, Damia G, Catapano CV, Conter V, Gervasoni M, Masera G, Liso V, Specchia G, Giudici G, et al.** 1993. Cytotoxic activity and mechanism of action of 5-Aza-2'-deoxycytidine in human CML cells. *Leuk Res* **17**:977-982.
314. **Zhao J, Kelnar K, Bader AG.** 2014. In-depth analysis shows synergy between erlotinib and miR-34a. *PLoS One* **9**:e89105.
315. **Prichard MN, Prichard LE, Shipman C, Jr.** 1993. Strategic design and three-dimensional analysis of antiviral drug combinations. *Antimicrob Agents Chemother* **37**:540-545.
316. **Schmittgen TD, Zakrajsek BA.** 2000. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J Biochem Biophys Methods* **46**:69-81.
317. **Butler SL, Hansen MS, Bushman FD.** 2001. A quantitative assay for HIV DNA integration in vivo. *Nat Med* **7**:631-634.
318. **Kim DZ, R.; Saduvala, N.; Kline, R.; Banez Ocfemia, C.; Prejean, J.; Zhang, X.; Pearson, M.; Hall, I.; Variant, Atypical, and Resistant HIV Surveillance Group.** Trend in Transmitted HIV-1 ARV Drug Resistance-associated Mutations: 10 HIV Surveillance Areas, US, 2007–2010, p. *In* (ed),
319. **Gao F, Chen Y, Levy DN, Conway JA, Kepler TB, Hui H.** 2004. Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious. *J Virol* **78**:2426-2433.
320. **Anderson JP, Daifuku R, Loeb LA.** 2004. Viral error catastrophe by mutagenic nucleosides. *Annu Rev Microbiol* **58**:183-205.
321. **Smith RA, Loeb LA, Preston BD.** 2005. Lethal mutagenesis of HIV. *Virus Res* **107**:215-228.

322. **Lee CH, Gilbertson DL, Novella IS, Huerta R, Domingo E, Holland JJ.** 1997. Negative effects of chemical mutagenesis on the adaptive behavior of vesicular stomatitis virus. *J Virol* **71**:3636-3640.
323. **Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, Andino R, Cameron CE.** 2000. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat Med* **6**:1375-1379.
324. **Crotty S, Cameron CE, Andino R.** 2001. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci U S A* **98**:6895-6900.
325. **Sierra S, Davila M, Lowenstein PR, Domingo E.** 2000. Response of foot-and-mouth disease virus to increased mutagenesis: influence of viral load and fitness in loss of infectivity. *J Virol* **74**:8316-8323.
326. **Ruiz-Jarabo CM, Ly C, Domingo E, de la Torre JC.** 2003. Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). *Virology* **308**:37-47.
327. **Lori F, Malykh A, Cara A, Sun D, Weinstein JN, Lisziewicz J, Gallo RC.** 1994. Hydroxyurea as an inhibitor of human immunodeficiency virus-type 1 replication. *Science* **266**:801-805.
328. **Julias JG, Pathak VK.** 1998. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *J Virol* **72**:7941-7949.
329. **Baker CH, Banzon J, Bollinger JM, Stubbe J, Samano V, Robins MJ, Lippert B, Jarvi E, Resvick R.** 1991. 2'-Deoxy-2'-methylencytidine and 2'-deoxy-2',2'-difluorocytidine 5'-diphosphates: potent mechanism-based inhibitors of ribonucleotide reductase. *J Med Chem* **34**:1879-1884.
330. **Nyholm S, Mann GJ, Johansson AG, Bergeron RJ, Graslund A, Thelander L.** 1993. Role of ribonucleotide reductase in inhibition of mammalian cell growth by potent iron chelators. *J Biol Chem* **268**:26200-26205.
331. **Smoliga JM, Baur JA, Hausenblas HA.** 2011. Resveratrol and health--a comprehensive review of human clinical trials. *Mol Nutr Food Res* **55**:1129-1141.
332. **Heinemann V, Hertel LW, Grindey GB, Plunkett W.** 1988. Comparison of the cellular pharmacokinetics and toxicity of 2',2'-difluorodeoxycytidine and 1-beta-D-arabinofuranosylcytosine. *Cancer Res* **48**:4024-4031.
333. **Huang P, Chubb S, Hertel LW, Grindey GB, Plunkett W.** 1991. Action of 2',2'-difluorodeoxycytidine on DNA synthesis. *Cancer Res* **51**:6110-6117.
334. **Mackey JR, Mani RS, Selner M, Mowles D, Young JD, Belt JA, Crawford CR, Cass CE.** 1998. Functional nucleoside transporters are required for gemcitabine influx and manifestation of toxicity in cancer cell lines. *Cancer Res* **58**:4349-4357.
335. **Murakami E, Basavapathruni A, Bradley WD, Anderson KS.** 2005. Mechanism of action of a novel viral mutagenic covert nucleotide:

- molecular interactions with HIV-1 reverse transcriptase and host cell DNA polymerases. *Antiviral Res* **67**:10-17.
336. **Heinemann V, Xu YZ, Chubb S, Sen A, Hertel LW, Grindey GB, Plunkett W.** 1992. Cellular elimination of 2',2'-difluorodeoxycytidine 5'-triphosphate: a mechanism of self-potential. *Cancer Res* **52**:533-539.
337. **Abbruzzese JL, Grunewald R, Weeks EA, Gravel D, Adams T, Nowak B, Mineishi S, Tarasoff P, Satterlee W, Raber MN, et al.** 1991. A phase I clinical, plasma, and cellular pharmacology study of gemcitabine. *J Clin Oncol* **9**:491-498.
338. **Chan KK, Giannini DD, Staroscik JA, Sadee W.** 1979. 5-Azacytidine hydrolysis kinetics measured by high-pressure liquid chromatography and ¹³C-NMR spectroscopy. *J Pharm Sci* **68**:807-812.
339. **Yoo CB, Jeong S, Egger G, Liang G, Phiasivongsa P, Tang C, Redkar S, Jones PA.** 2007. Delivery of 5-aza-2'-deoxycytidine to cells using oligodeoxynucleotides. *Cancer Res* **67**:6400-6408.
340. **Patel KR, Scott E, Brown VA, Gescher AJ, Steward WP, Brown K.** 2011. Clinical trials of resveratrol. *Ann N Y Acad Sci* **1215**:161-169.
341. **Perales C, Agudo R, Tejero H, Manrubia SC, Domingo E.** 2009. Potential benefits of sequential inhibitor-mutagen treatments of RNA virus infections. *PLoS Pathog* **5**:e1000658.
342. **Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R.** 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A* **108**:20166-20171.
343. **Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B.** 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**:9530-9535.
344. **Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L, Mandaliya K, Jaoko W, Overbaugh J.** 2007. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect Dis* **195**:1177-1180.
345. **Kiwanuka N, Laeyendecker O, Quinn TC, Wawer MJ, Shepherd J, Robb M, Kigozi G, Kagaayi J, Serwadda D, Makumbi FE, Reynolds SJ, Gray RH.** 2009. HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS* **23**:2479-2484.
346. **Kiwanuka N, Laeyendecker O, Robb M, Kigozi G, Arroyo M, McCutchan F, Eller LA, Eller M, Makumbi F, Birx D, Wabwire-Mangen F, Serwadda D, Sewankambo NK, Quinn TC, Wawer M, Gray R.** 2008. Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J Infect Dis* **197**:707-713.
347. **Renjifo B, Gilbert P, Chaplin B, Msamanga G, Mwakagile D, Fawzi W, Essex M, Tanzanian V, Group HIVS.** 2004. Preferential in-utero

transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS* **18**:1629-1636.

348. **Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M, Fawzi W, Hunter D.** 2006. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin Infect Dis* **42**:843-852.

APPENDIX I
DETERMINATION OF HIV-1 AND HIV-2 MUTATION FREQUENCIES AND
SPECTRA BY SINGLE-STRAND CONSENSUS SEQUENCING

Jonathan M.O. Rawson,^{a,d} Daryl M. Gohl,^e Sean R. Landman,^f Cavan S. Reilly,^{a,g}
Kenneth B. Beckman,^e and Louis M. Mansky^{a-d}

Institute for Molecular Virology,^a Department of Diagnostic and Biological Sciences, School of Dentistry,^b Department of Microbiology,^c Molecular, Cellular, Developmental Biology & Genetics Graduate Program,^d University of Minnesota Genomics Center,^e Department of Computer Science and Engineering,^f Division of Biostatistics, School of Public Health,^g University of Minnesota, Minneapolis, Minnesota, USA

Abstract

The high mutability of retroviruses is critical to their ability to evade the immune response, develop drug resistance, and modulate host cell tropism. While direct amplification and high-throughput sequencing of viral genes can be used to study viral mutagenesis, these experiments often lead to high levels of background errors that hinder identification of true viral mutations. To help overcome this limitation, a method was devised based on single-strand consensus sequencing (SSCS) for the identification of mutations in proviral DNA and was applied to compare viral mutagenesis between HIV type-1 (HIV-1) and HIV type-2 (HIV-2). We found that SSCS efficiently reduced all types of background errors with the exception of C-to-A transversion mutations, although the fold reductions varied widely depending upon the specific error type. Using SSCS, HIV-2 displayed a lower overall mutation frequency as well as lower substitution and transition mutation frequencies than that of HIV-1, due in part to a reduced recovery of HIV-2 G-to-A hypermutants. HIV-1 G-to-A hypermutants were heavily biased toward GA dinucleotides, consistent with the editing activity of APOBEC3 proteins. After exclusion of G-to-A hypermutants, HIV-2 still exhibited a somewhat lower total mutation and G-to-A transition mutation frequency than HIV-1. Taken together, these findings demonstrate that SSCS improves the detection of viral mutations compared to traditional high-throughput sequencing approaches. Furthermore, this study demonstrates that there are multiple differences between HIV-1 and HIV-2 mutagenesis, which may impact the relative abilities of these closely related viruses to evolve within hosts.

Introduction

Retroviruses have typically been found to mutate on the order of 10^{-4} - 10^{-5} mutations/base pair/cycle (m/bp/c) (1-7), a rate approximately 10,000-100,000 times faster than eukaryotic genomic DNA (8). In the past, retroviral mutation rates have primarily been determined using reporter gene assays, as direct amplification and Sanger sequencing of viral genes is typically far too low-

throughput to survey sufficient numbers of mutations. For example, assuming a mutation rate of 10^{-4} m/bp/c (corresponding to the approximate upper end of previous estimates of retroviral mutation rates) and a sequence length of 800 bp, ~12.5 sequences would be required to identify a single mutation (or ~1250 sequences to identify 100 mutations). Thus, it is difficult to use Sanger sequencing of viral genes to examine appreciable numbers of mutations, especially for experiments involving multiple samples and biological replicates. High-throughput sequencing of viral amplicons has also recently been used to determine retroviral mutation frequencies (7, 9), but PCR and high-throughput sequencing often lead to high levels of background errors, typically $\sim 10^{-2}$ m/bp for raw sequencing data and 10^{-3} to 10^{-4} m/bp for stringently filtered data (10-16). Thus, although large numbers of mutations can readily be identified, it becomes difficult to distinguish biologically meaningful mutations from PCR and/or sequencing artifacts.

In the last few years, several different approaches have been developed that allow for the identification and exclusion of most PCR and sequencing errors that arise during high-throughput sequencing experiments, thus greatly reducing the background error rates of such experiments. These error-correcting techniques can broadly be divided into three categories: single-strand consensus sequencing (SSCS) (17-20), duplex consensus sequencing (DCS) (21, 22), and circle sequencing (CirSeq) (23, 24). For SSCS and DCS, starting templates are uniquely tagged (usually by a string of random bases), exponentially amplified, redundantly sequenced, and consensus sequences are constructed from reads containing identical tags. For CirSeq, starting templates are subjected to rolling circle amplification, resulting in multiple, physically-linked copies that can then be used to construct consensus sequences. Thus far, these approaches have been used for the analysis of viral populations (18, 23, 25), the identification of mutations in eukaryotic mitochondrial or chromosomal DNA (19, 21, 22, 24, 26), determination of background error rates during PCR or oligonucleotide synthesis (19), characterization of microbiota (i.e. 16S rRNA

sequencing) (17, 20), and analysis of rare circulating tumor DNA (27). In regard to analysis of viral populations, an SCS-related method (referred to as Primer ID) has been used to identify rare polymorphisms (including drug resistance-associated mutations) in HIV-1 protease and to study their evolution within patients over time (18, 25). Additionally, CirSeq has been used to determine the mutation rate and spectrum of poliovirus in cell culture, as well as to determine the effects of individual mutations on viral fitness by tracking their frequency during viral passaging (23). Notably, these error-correcting approaches also mitigate the effects of PCR-mediated recombination and PCR-mediated skewing of mutation frequencies.

Previously, Illumina high-throughput sequencing was used to compare viral mutation frequencies and spectra between human immunodeficiency virus type-1 (HIV-1) and type-2 (HIV-2) (9). HIV-2 was hypothesized to exhibit a lower mutation frequency, which might aid in explaining its reduced viral fitness, transmissibility, and pathogenicity (28, 29). We found that HIV-2 displayed lower total, substitution, and transition mutation frequencies than HIV-1, particularly due to reduced levels of G-to-A hypermutants that were consistent with the activity of APOBEC3 proteins. However, high levels of background errors were observed during this approach, which made it difficult to perform a detailed comparison between HIV-1 and HIV-2 for every possible error type. To address this limitation, a high-throughput sequencing method based on SCS was developed for the determination of mutation frequencies and spectra in HIV-1 and HIV-2 proviral DNA. SCS effectively reduced most types of background errors, although fold reductions varied widely among error types. Using SCS, HIV-2 was found to exhibit a lower total mutation frequency than HIV-1, due in part to reduced levels of G-to-A hypermutation for HIV-2. Overall, SCS greatly improved the determination of retroviral mutation frequencies and spectra by high-throughput sequencing. This technical advancement enables future studies of factors that may influence retroviral mutagenesis, including virus type or subtype, cell type, and antiviral drugs.

Results

Development of SSCS for identification of mutations in HIV-1 and HIV-2 proviral DNA. While a previous study described an SSCS method for the detection of mutations in HIV-1 protease (18), their approach was developed for sequencing viral RNA, whereas retroviral mutation frequencies and spectra are typically calculated by analysis of proviral DNA. Thus, a similar method was developed for the amplification and sequencing of HIV-1 and HIV-2 proviral DNA, which incorporated select features of SSCS protocols from other published studies as well (17, 19, 20). The SSCS protocol is described in detail in Figure A1-1 and in the Materials and Methods. Briefly, U373-MAGI-X4 cells were first infected by HIV-1 or HIV-2 in a single cycle of replication, and genomic DNA (containing proviral DNA) was purified from cells collected 72 h post-infection. Proviral DNA was then subjected to linear extension using a forward primer (specific to HIV-1 or HIV-2 integrase) that contains a unique identifier (UID) composed of fourteen degenerate bases. The number of possible UIDs (4^{14} , or ~268 million) far exceeded the number of proviruses per sample (~375,000), such that it is highly unlikely two different templates were assigned the same UID. The complementary strand was then synthesized from the uniquely tagged copies of starting templates using an HIV-1 or HIV-2 integrase-specific reverse primer. Next, the doubly-extended templates were amplified by exponential qPCR using primers specific to the tails (corresponding to Illumina adapter sequence) of earlier tagging primers. The results of initial qPCR reactions were used to estimate the approximate starting copy number of doubly-tagged templates, which then allowed for targeting of specific copy numbers in additional downstream qPCR reactions. Further PCR was then performed to add sample indices and the remaining Illumina adapter sequences. Last, redundant sequencing was performed on the Illumina MiSeq, leading to the generation of consensus families of variable sizes (mean family size varied from ~11 to 27 members/family). Consensus sequences were constructed from all consensus

families containing at least three members, and bases (whether wild-type or mutations) had to be present in at least 75% of members in order to be represented in the consensus sequence. Ambiguous bases were converted to uncalled bases and were not included in subsequent analysis of mutational events. Consensus sequences were then mapped to the appropriate reference sequence, and mutation frequencies and spectra were assessed as described before (9).

SSCS effectively reduces background error frequencies for most error types. After performing SSCS, the extent to which SSCS reduced the levels of various types of background errors was determined using the HIV-1 and HIV-2 plasmid controls. Plasmid control reactions were performed by tagging and amplifying purified plasmid DNA that was diluted into genomic DNA from uninfected cells. Mutation frequencies of the plasmid controls were determined by both standard analysis (involving read-level and base-level quality filters) of reads prior to building of consensus sequences as well as by SSCS (see Materials and Methods). SSCS decreased the total mutation frequency by ~1.8-fold for the HIV-1 plasmid control (3.1 to 1.8×10^{-4} m/bp) and by ~3.3-fold for the HIV-2 plasmid control (2.8×10^{-4} to 8.5×10^{-5} m/bp) (Figure A1-2A). However, SSCS reduced background transition frequencies by 6.4-fold [HIV-1] or 10.7-fold [HIV-2], much more effectively than for transversion frequencies. As transversions comprised the majority of background errors (~70% using standard analysis, Figure A1-2D), they had a greater impact on total mutation frequencies than transitions. These results indicate that analysis of biological-occurring transversions might be problematic even with SSCS. However, SSCS effectively reduced background transition frequencies, and many previous reports have demonstrated that transitions are the primary mutation type during HIV-1 replication (3, 4, 30-34).

Interestingly, SSCS reduced frequencies of specific types of transitions and transversions more effectively than others. G-to-A transitions were reduced

to a smaller degree than other types of transitions (Figure A1-2B), possibly due to low level DNA damage in the starting template, as has been observed by others using these error-reducing techniques (21, 24). Spontaneous deamination of deoxycytidine to deoxyuridine could lead to G-to-A or C-to-T errors during PCR, depending upon the strand(s) amplified. In the protocol described here, only the antisense strand of the starting template is amplified, such that spontaneous deamination would result in G-to-A transitions. Likewise, C-to-A and, to a lesser extent, G-to-T background transversion errors were reduced much less effectively than other types of transversion errors (Figure A1-2C), a phenomenon which has also been observed by others (21, 24). Notably, C-to-A transversions comprised the majority of background errors (~60%) observed during standard analysis, and they represented ~90% of all background errors using SSCS (Figure A1-2D). Thus, the limited reduction of total mutation frequencies observed during SSCS (Figure A1-2A) was primarily due to C-to-A transversions. These C-to-A transversions may have resulted from low-level oxidative damage of the starting template that converted guanine to 8-oxoguanine (8-oxoG). During linear extension, 8-oxoG in the antisense template could result in a first-cycle C-to-A error, which would then become fixed within all members of a consensus family (see Discussion). Although SSCS did not provide much benefit for analysis of G-to-T or C-to-A errors, the other six types of transversion errors were effectively reduced by SSCS (Figure A1-2C), thus facilitating their detection in biological samples.

HIV-1 and HIV-2 exhibit different mutation frequencies and spectra. Next, SSCS was utilized to compare mutation frequencies and spectra between HIV-1 and HIV-2. HIV-1 and HIV-2 viral stocks were used to infect 375,000 U373-MAGI cells at an MOI of 1.0 in a single cycle of replication. Genomic DNA was purified from cells collected 72 h post-infection and subjected to SSCS. In addition to applying SSCS, background error hotspots (defined as upper outliers within the mutation frequency distribution of the plasmid controls) were identified

and masked from the consensus sequences as well (see Materials and Methods). Notably, all masked errors were C-to-A transversions, again reflecting the dominance of C-to-A transversions over other mutational types within the plasmid controls (Figure A1-2). HIV-1 exhibited a 2.2-fold higher total mutation frequency than HIV-2 (8.0 vs. 3.6×10^{-5} m/bp) (Figure A1-3A). This difference was primarily due a 2.5-fold higher transition frequency for HIV-1, although HIV-1 also displayed a 1.6-fold higher frequency of transversions than HIV-2. For transitions, the primary difference observed was an ~5.0-fold higher frequency of G-to-A transitions for HIV-1 (Figure A1-3B), consistent with our previous report (9). We also observed smaller differences (~1.5 to 1.8-fold) between HIV-1 and HIV-2 in C-to-T, A-to-G, and T-to-C frequencies. HIV-1 exhibited higher frequencies than HIV-2 for all transition types except A-to-G transitions. For transversions, two types (G-to-T and C-to-A) were not compared between HIV-1 and HIV-2 because they were not detected at significantly higher levels in the biological samples than the plasmid controls. For the remaining types, the most striking differences observed were ~3.8- and 4.9-fold higher T-to-A and G-to-C frequencies, respectively, for HIV-1 (Figure A1-3C). The relative distribution of mutation types, or mutation spectrum, also varied between HIV-1 and HIV-2, with HIV-2 exhibiting a lower relative percentage of G-to-A transitions (25% vs. 56%) and a higher relative percentage of A-to-G transitions (21% vs. 6%) compared to HIV-1 (Figure A1-3D). The frequencies of insertions and deletions for HIV-1 and HIV-2 were also determined, but insertions and deletions were much less frequent than substitutions (composing ~2 to 4% of all mutations, Figure A1-3D) and were detected at similar levels in HIV-1 and HIV-2 (data not shown).

Differences in mutation frequencies between HIV-1 and HIV-2 are due in part to HIV-1 G-to-A hypermutants. Previously, we found that differences between HIV-1 and HIV-2 mutation frequencies and spectra were primarily due to higher levels of G-to-A hypermutants for HIV-1, although subtle but statistically significant differences were observed in G-to-A, A-to-G, and C-to-T frequencies

even after removal of G-to-A hypermutants (9). G-to-A hypermutants were biased toward GA dinucleotides, consistent with APOBEC3 activity. To determine whether the higher total, transition, and G-to-A frequencies observed here for HIV-1 (Figure A1-3) were again due to hypermutants, hypermutant frequencies were determined for all four possible types of transition hypermutants. Hypermutants were defined as individual consensus reads (~200 bp in length after processing) containing at least two mutations of the appropriate type. In total, 88 G-to-A hypermutants were identified for HIV-1 and 6 G-to-A hypermutants were identified for HIV-2. HIV-1 exhibited a G-to-A hypermutant frequency of $\sim 1.1 \times 10^{-3}$ (hypermutant reads/all reads), which was ~22-fold higher than the G-to-A hypermutant frequency for HIV-2 (5.1×10^{-5}) (Figure A1-4A). HIV-1 G-to-A hypermutants were heavily biased toward GA dinucleotides (Figure A1-4A; inset), consistent with the preferences of multiple APOBEC3 family members.

Next, in order to determine whether differences in HIV-1 and HIV-2 were primarily due to G-to-A hypermutants, mutation frequencies and spectra were calculated in the absence of G-to-A hypermutants. After excluding G-to-A hypermutants, HIV-1 and HIV-2 exhibited total mutation frequencies of 6.0×10^{-5} [HIV-1] or 3.6×10^{-5} m/bp [HIV-2] (Figure A1-4B). Thus, HIV-1 had an ~1.7-fold higher total mutation frequency than HIV-2, less than the 2.2-fold difference originally observed (Figure A1-3A). Even after removing G-to-A hypermutants, HIV-1 had an ~2.9-fold higher G-to-A frequency than HIV-2, a somewhat more pronounced difference than observed in a previous study (see Discussion) (9). The mutation spectra of HIV-1 and HIV-2 were also more similar after the removal of G-to-A hypermutants (compare Figures A1-3D and A1-4C). However, HIV-2 still demonstrated a lower relative percentage of G-to-A transitions (23% vs. 40%) and a higher relative percentage of A-to-G transitions (22% vs. 8%) than HIV-1 (Figure A1-4C).

Discussion

The advent of next-generation sequencing has enabled the detection of large numbers of mutations in viral genes, which can then be used to examine mutation frequencies and spectra in great depth. However, rates of background errors from PCR and sequencing are often similar to or even higher than published estimates of retroviral mutation rates, making it difficult to distinguish true viral mutations from experimental artifacts. To address these issues, a high-throughput sequencing method was developed that utilizes single-strand consensus sequencing (SSCS) for the identification of mutations in a targeted region of HIV-1 or HIV-2 proviral DNA. SSCS effectively reduced background error frequencies for most types of substitutions (Figure A1-2). However, C-to-A transversions comprised the bulk of background errors and were not significantly reduced by SSCS, such that total mutation frequencies did not decrease as much as originally anticipated. Indeed, while SSCS reduced the total mutation frequency by only 1.7-fold [HIV-1 plasmid] or 3.3-fold [HIV-2 plasmid], SSCS reduced the total mutation frequency by 6.9-fold [HIV-1 plasmid] or 13.0-fold [HIV-2 plasmid] in the absence of C-to-A transversions. Based on these results and the work of others (21, 24), C-to-A transversions may have been primarily caused by low-level oxidative damage of the starting template, which could convert guanine to 8-oxoG and ultimately lead to the propagation of C-to-A transversions to all members of consensus families.

Although SSCS successfully reduced most types of background errors, there are still several limitations of the assay that will hopefully be addressed by future improvements. First, C-to-A (and to a lesser extent G-to-A and G-to-T) errors were not as effectively reduced as other error types, likely due at least in part to damage of the starting template before (and possibly during) linear extension and exponential amplification. These errors could be prevented by using duplex consensus sequencing (DCS) instead of SSCS, as DCS excludes mutations unless they are present in both original template strands (21). DCS of HIV-1 proviral DNA might prove technically challenging, because DCS requires

the ligation of double-stranded tags to sheared genomic DNA, and proviral sequences would then have to be specifically enriched prior to exponential PCR. Alternatively, the starting template for SSCS could be treated with repair enzymes (e.g. 8-oxoG glycosylase, uracil DNA glycosylase, DNA polymerase, DNA ligase) prior to linear extension and PCR. Additionally, our current method can only analyze a small region of proviral DNA, and mutation frequencies and spectra may vary across viral genes. In the future, it may be possible to sequence multiple regions or even the entire proviral genome either by multiplexing linear extension reactions or by shearing genomic DNA, enriching proviral DNA using a probe-based capture assay, and adding UIDs by ligation. As these technologies continue to rapidly develop, our ability to detect ultra-rare mutations in viral populations will likely continue to improve.

Using SSCS, the primary difference between HIV-1 and HIV-2 was found to be an ~5.0-fold higher G-to-A frequency for HIV-1 (Figure A1-3). Smaller differences (~1.5 to 1.8-fold) in the frequencies of other transitions were also observed. In this study, transversion frequencies between HIV-1 and HIV-2 for all error types except G-to-T and C-to-A were compared, whereas in a previous study transversions could not be assessed at all due to high background error (9). HIV-1 was found to have an ~3.8-fold and an ~4.9-fold higher frequency of T-to-A and G-to-C errors than HIV-2, respectively. HIV-1 had a higher total mutation frequency in part due to a higher frequency of G-to-A hypermutants (Figure A1-4A), as observed previously (9). G-to-A hypermutants were heavily biased toward GA dinucleotides (Figure A1-4A), unlike single G-to-A mutants (data not shown), indicating they were likely the result of APOBEC3 activity. However, even in the absence of these hypermutants, HIV-1 still had a higher total and G-to-A frequency than HIV-2 (Figure A1-4B). Additionally, the removal of G-to-A hypermutants did not impact differences in mutation frequencies that were observed for many other types of substitutions (Figure A1-3). Although the total mutation frequencies of HIV-1 and HIV-2 did not differ after removal of G-to-A hypermutants in a previous study (9), the differences observed here were likely

due to the much lower background error rate of SSCS. Further, there were several other differences between the methods used that hinder direct comparison of results, such as the amplicon (integrase in this study), PCR protocols, sequencing format, and data processing steps. Taken together, the findings described here demonstrate that there are multiple differences in mutation frequencies and spectra between HIV-1 and HIV-2, which may impact the relative abilities of the viruses to adapt within hosts. These results suggest that both APOBEC3 proteins and reverse transcriptase may contribute to the variable fidelity of HIV-1 and HIV-2 replication. Additional studies are in progress to determine: 1) whether these differences are representative of HIV-1 and HIV-2 as a whole, 2) whether there are differences in mutation frequencies and spectra between subtypes of HIV-1, 3) whether HIV-1 and HIV-2 mutation frequencies vary in relevant primary cell types (i.e. macrophages and CD4+ T-cells), and 4) whether HIV-1 and HIV-2 vary in their susceptibility to APOBEC3 hypermutation using higher, physiologically-relevant levels of APOBEC3 expression. Overall, the development of SSCS for detection of mutations in proviral DNA will likely greatly improve the feasibility of such studies, leading to new insights into retroviral mutagenesis.

Materials and Methods

Preparation of HIV-1 and HIV-2 proviral DNA. Viral stocks were prepared using pNL4-3 MIG (HIV-1) and pROD MIG (HIV-2) vectors (35, 36). These vectors express two reporter proteins, mCherry and EGFP, which enable determination of infectivity by flow cytometry. The vectors contain intact open reading frames for all viral genes except for *env* and *nef*. Viral stocks were produced, concentrated, DNaseI-treated, and titered in U373-MAGI cells as described previously (9). Next, large-scale infections were performed in which 375,000 U373-MAGI cells were infected at a multiplicity of infection (MOI) of 1.0. Uninfected cells were included as negative controls. All infections were performed three times, using independently produced viral stocks. Cells were

collected at 72 h post-infection, and genomic DNA was purified using the High Pure PCR Template Preparation Kit (Roche; Basel, Switzerland), eluting in 100 μ L buffer.

Assignment of unique identifiers (UIDs) by linear extension. To prepare samples for SSCS, 50 μ L linear extensions were performed using 16 μ L water, 20 μ L genomic DNA, 10 μ L 5X HF buffer, 1 μ L dNTPs (10 mM), 2.5 μ L forward primer (10 μ M), and 0.5 μ L Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific; Waltham, MA). The cycling conditions used were an initial denaturation of 98 $^{\circ}$ C 30 s, and 10 cycles of 98 $^{\circ}$ C 10 s/61 $^{\circ}$ C 30 s/72 $^{\circ}$ C 30 s. Although it would be preferable to use a single cycle of linear extension (to minimize re-extension from the same starting template), extension efficiencies in a single cycle were usually very low (~1-5%), thus necessitating the use of multiple extension cycles. The forward primers in these reactions contained a region specific to integrase, a UID composed of 14 degenerate bases (Ns), and Illumina adapter sequence (Table A1-1). HIV-1 and HIV-2 primers were homologous in sequence.

The newly synthesized and tagged strand was purified to remove free unextended primer. Free primers could lead to further extension from proviral DNA and also lead to the formation of primer heterodimers that would be difficult to remove, a common problem when using long primers containing Illumina adapter sequences (due to partial complementarity between the adapter sequences). The tagged strand was purified using AMPure XP beads (Beckman Coulter; Brea, CA) at a 0.8:1 (beads:DNA) ratio and eluted in 42 μ L water, followed by transfer of 36 μ L/sample to a new 96-well plate. Next, the complementary strand was synthesized using a reverse primer specific to integrase by mixing 36 μ L purified DNA, 10 μ L 5X HF buffer, 1 μ L dNTPs, 2.5 μ L primer, and 0.5 μ L Phusion. The reaction was incubated at 98 $^{\circ}$ C 40 s, 58 $^{\circ}$ C 30 s, and 72 $^{\circ}$ C 30 s. During this step, the reverse primer added part of the Illumina

adapter sequence. Lastly, the doubly-tagged DNA was again purified using AMPure XP beads at a 0.8:1 ratio, eluting in 41 μL water. The purified DNA (~ 35 μL) was then transferred to a new 96-well plate. Three extensions (forward + reverse) were performed per sample and pooled after the final bead purification.

Exponential qPCR of tagged DNA. Tagged DNA was next subjected to exponential qPCR using primers specific to the tails of earlier primers, which represent Illumina adapter sequences. Thus, templates lacking one or both tags could not be exponentially amplified. Initially, small-scale qPCR reactions were performed in order to estimate the absolute copy number of starting molecules, as this information would allow for the targeting of specific family sizes during consensus sequencing. To accomplish this, 12.5 μL qPCR reactions were performed using 4.25 μL water, 2.5 μL 5X HF buffer, 0.25 μL dNTPs, 0.625 μL each primer, 0.625 μL SYBR Green I (Life Technologies; Carlsbad, CA) diluted 1:2000 in DMSO, 3.5 μL purified DNA, and 0.125 μL Phusion polymerase. The cycling conditions were 98 $^{\circ}\text{C}$ 30 s, 40 cycles of 98 $^{\circ}\text{C}$ 10 s/70 $^{\circ}\text{C}$ 30 s/72 $^{\circ}\text{C}$ 30 s, and a final extension of 72 $^{\circ}\text{C}$ 2 min. To create appropriate standard curves for qPCR, we produced fully tailed PCR products from plasmids using nested PCR and Platinum PCR SuperMix (Thermo Fisher Scientific). These A-tailed PCR products were then ligated to the pGEM-T vector (Promega; Madison, WI), purified, and fully verified by Sanger sequencing. Absolute standard curve series (from 10^2 - 10^6 copies/ μL) were constructed by quantifying the resulting plasmids with the Qubit dsDNA HS Assay Kit (Life Technologies) and diluting appropriately.

Based on results of the initial qPCR, additional 50 μL qPCR reactions were performed using 25,000 copies of starting template per reaction and 4 reactions per sample. Thus, $\sim 100,000$ starting copies were amplified per sample, which was expected to result in consensus families of ~ 10 -12 members/family based on typical MiSeq 2 \times 250 sequencing outputs. These

reactions were performed using 31 μL of water + template (25,000 copies), 10 μL 5X HF buffer, 1 μL dNTPs, 2.5 μL each primer, 2.5 μL SYBR Green, and 0.5 μL Phusion polymerase. The cycling conditions were 98 °C 30 s, 30 cycles of 98 °C 10 s/70 °C 30 s/72 °C 30 s, and a final extension of 72 °C 2 min. While SYBR Green could potentially be omitted from these reactions, it was included in order to ensure that the observed starting quantity was close to that expected. Quadruplicate reactions were then pooled, purified by gel extraction, quantified by the Qubit dsDNA HS assay, and normalized to a concentration of 2.5 ng/ μL .

Library preparation and Illumina sequencing. 100 ng (40 μL) of each gel-purified amplicon was submitted to the University of Minnesota Genomics Center for library preparation and sequencing. Additional PCR was performed on the amplicons in order to add library indices and remaining Illumina adapter sequences. This amplification was performed using 5 μL template DNA, 1 μL water, 2 μL 5x KAPA HiFi buffer (Kapa Biosystems, Woburn, MA), 0.3 μL dNTPs (10 mM), 0.5 μL DMSO, 0.5 μL each primer (10 μM), and 0.2 μL KAPA HiFi Polymerase (Kapa Biosystems). The cycling conditions were 95 °C for 5 min, followed by 10 cycles of 98 °C for 20 s, 55 °C for 15 s, 72°C for 1 min, and a final extension at 72 °C for 10 min. The primers used were standard Illumina Nextera dual-indexing primers (Table A1-1). PCR products were quantified using the Qubit dsDNA BR Assay Kit (Life Technologies), and the libraries were normalized, pooled, purified with 1.8X AMPure XP beads, and eluted in 20 μL of EB buffer (10 mM Tris-HCl, pH 8.5). The final library pool was again quantified using the Qubit dsDNA BR assay and further assessed by a Bioanalyzer High Sensitivity chip (Agilent Technologies, Inc. Santa Clara, CA). The library pool was denatured with 0.2N NaOH, diluted to 8 pM, spiked with 15% PhiX to improve sequence diversity and quality, and subjected to 2 \times 250 paired-end sequencing on the Illumina MiSeq.

Bioinformatics processing of sequencing data. Samples were demultiplexed using the standard indices added during library preparation. Illumina adapters were trimmed using cutadapt (37), and paired-end reads were merged using PANDAseq (38). The amplicons were small enough (~254 bp, not including Illumina adapter sequences) such that forward and reverse reads were almost totally overlapping. If a base did not match between forward and reverse reads, the base with higher quality was chosen by PANDAseq. Next, consensus families were generated using BioPython (5) and a custom Python script. The script was used to create a list of the unique identifier (UID) sequences corresponding to the first 14 bp of each merged read. For each UID that was present at least three times in the dataset, a list of sequences containing the UID was generated and the consensus sequence at each position was determined by a simple pileup. The threshold used for making a consensus call was 75%. In cases where there was not a consensus base, the position was converted to an uncalled base (X) and was not considered during mutational analysis.

For the identification of mutations by standard (i.e. non-SSCS) methods, merged reads were analyzed prior to the construction of consensus sequences. Reads were mapped to the appropriate reference sequence (pNL4-3 MIG or pROD MIG) using GSNAP (39), and a small number of misaligned reads were discarded. Mutation frequencies and hypermutant frequencies were determined using a custom algorithm based on the Genome Analysis Toolkit walker framework (40). Both mutations and wild-type bases were required to have a quality score of at least 30 in the merged read in order to be considered. All primer sequences were excluded from mutational analysis, as errors within primers would not represent biologically meaningful mutations. For initial analyses (Figure A1-2), plasmid error hotspots (defined as upper outliers within the distribution of frequencies of individual mutations based on the $1.5 \times \text{IQR}$ rule) were not masked, a strategy previously used to reduce the level of background error (9). For later analyses (Figures A1-3 & A1-4), plasmid error hotspots were masked in order to further reduce the level of background error

during SSCS. All plasmid error hotspots were C-to-A transversions, and thus their removal only influenced the C-to-A frequency. Rather than masking all mutational types at hotspot positions, only the problematic type (i.e. C-to-A) was masked at error hotspots.

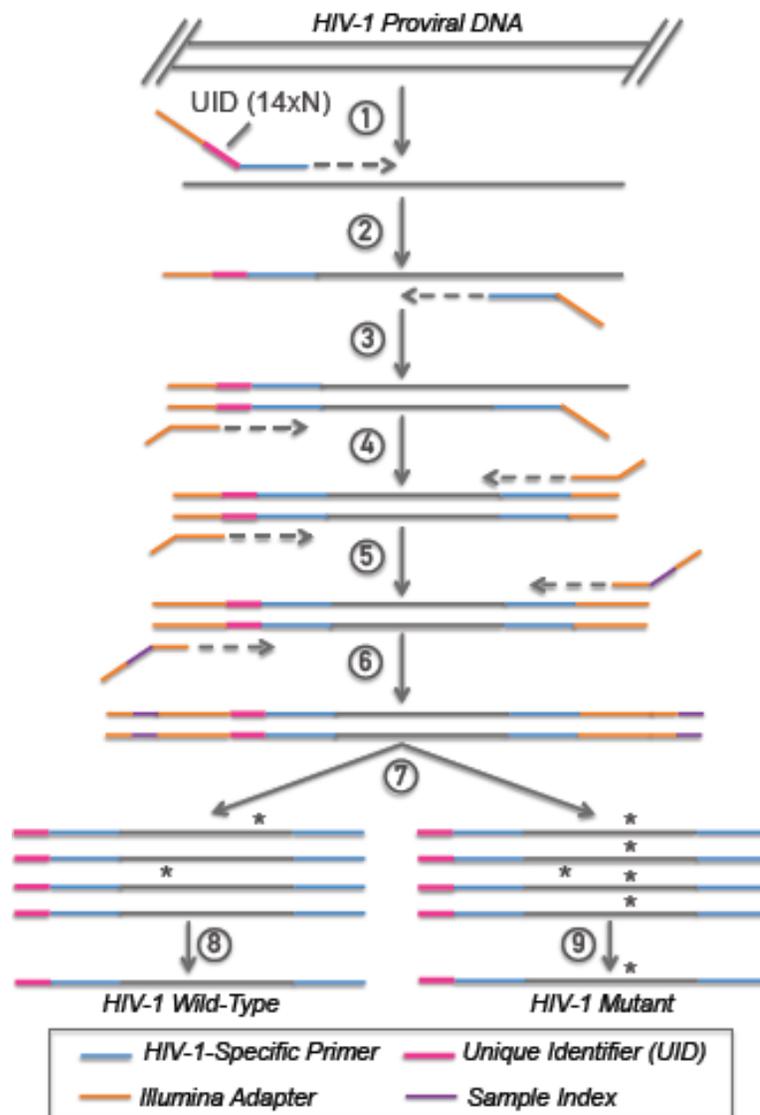


Figure A1-1. Experimental strategy for performing single-strand consensus sequencing of proviral DNA. Single-strand consensus sequencing (SSCS) was performed as follows: 1) HIV-1 proviral DNA was mixed in a linear extension reaction containing a forward primer specific to HIV-1 integrase. The forward primer contained a unique identifier (UID) composed of fourteen degenerate bases. 2) Linear extension resulted in unique tagging of sequence complementary to the starting template. 3) The tagged strand was purified to remove free primer, and the complementary strand was synthesized using a reverse primer specific to HIV-1 integrase. 4-5) The doubly-extended template

was again purified and subjected to exponential PCR using primers specific to the tails of earlier primers. 6) The amplicons were purified, and sample indices were added by further PCR. 7) Redundant sequencing was performed on the Illumina MiSeq (2×250) and consensus families were generated. 8-9) Consensus sequences were constructed from all families with at least three members, with mutations required to be identified in at least 75% of members. Examples of wild-type (8) and mutant (9) consensus sequences are illustrated, with several PCR or sequencing errors being excluded. SSCS of HIV-2 proviral DNA was carried out in a similar manner using primers to the homologous region of integrase.

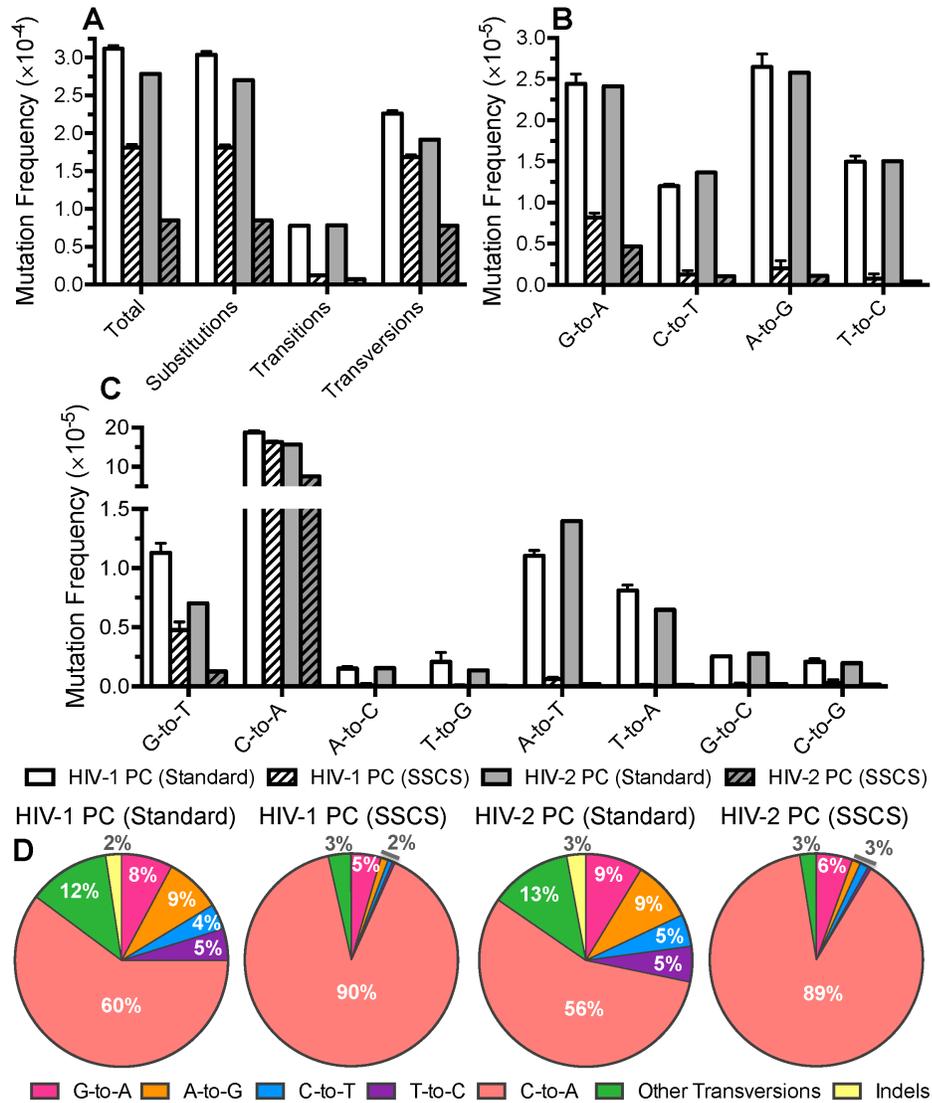


Figure A1-2. SCS effectively reduces background error frequencies for most error types. A-C. Mutation frequency analysis. Mutation frequencies were calculated by dividing the number of mutations by the number of reference bases (mutations + wild-type bases) and are expressed as mutations/bp, or m/bp. Mutation frequencies were determined for the HIV-1 and HIV-2 plasmid controls (PC) in order to assess background error levels, using either standard analysis of merged reads prior to building of consensus families or using SCS. Mutation

frequencies were calculated for total mutations and other broad categories of mutations (A), individual types of transitions (B), and individual types of transversions (C). D. Mutation spectra analysis. Mutation spectra were determined by dividing the frequency of each type of mutation by the total mutation frequency, with the results expressed as a percentage of total mutations. Data in all panels represent the mean of three experimental replicates, with error bars indicating the standard deviation.

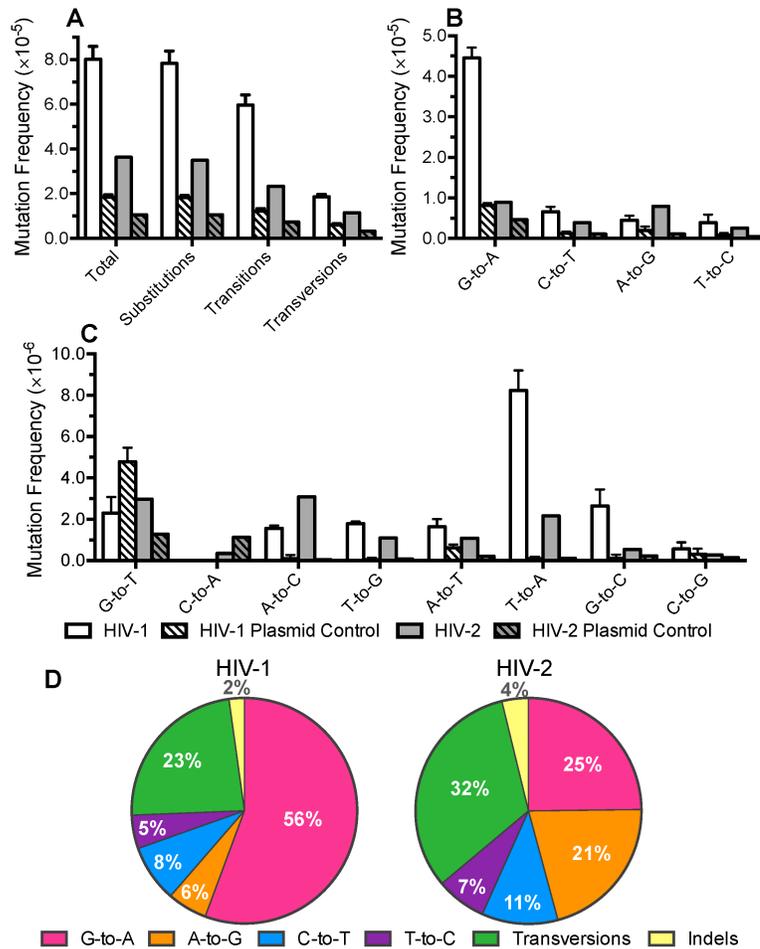


Figure A1-3. HIV-1 and HIV-2 exhibit different mutation frequencies and spectra. A-C. Mutation frequency analysis. Mutation frequencies were determined for HIV-1, HIV-2, and corresponding plasmid controls, using SSCS to exclude background errors. Mutation frequencies were calculated for total mutations and other broad categories of mutations (A), individual types of transitions (B), and individual types of transversions (C). D. Mutation spectra analysis. Mutation spectra were determined for HIV-1 and HIV-2 by dividing the frequency of each type of mutation by the total mutation frequency, with the results expressed as a percentage of total mutations. Data in all panels represent the mean of three experimental replicates, with error bars indicating the standard deviation.

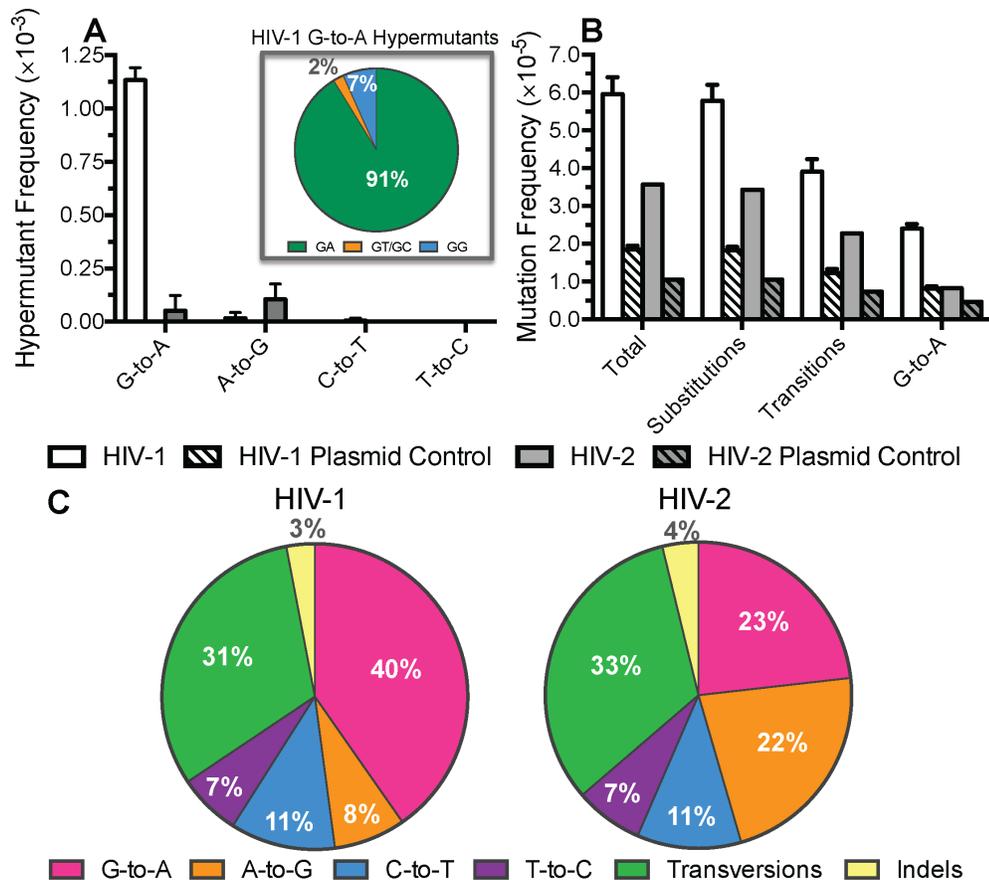


Figure A1-4. Differences in mutation frequencies between HIV-1 and HIV-2 are due in part to G-to-A hypermutants. A. The frequencies of each type of transition hypermutant were compared between HIV-1, HIV-2, and the plasmid controls. For this analysis, hypermutants were defined as individual consensus sequences (~200 bp in length after processing) containing two or more mutations of the indicated type. Hypermutant frequencies were calculated by dividing the number of hypermutant reads by all reads. The inset indicates the dinucleotide context distribution for HIV-1 G-to-A hypermutants. B. Mutation frequencies were calculated in the absence of G-to-A hypermutants. C. HIV-1 and HIV-2 mutation spectra were determined after excluding all G-to-A hypermutants. The data in all panels represent the mean of three experimental replicates, with error bars denoting standard deviation.

Table A1-1. Primer sequences for SSCS of HIV-1 and HIV-2 proviral DNA.^a

#	Primer Name	Sense	Sequence
1	HIV-1 Int-For	+	CAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNNNNNTGGAATAGATAAGGCCCAAG
2	HIV-1 Int-Rev	-	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATCCACTGGCTACATGAAC
3	HIV-2 Int-For	+	CAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNNNNNAAAAATAGAGCCCGCTCAGG
4	HIV-2 Int-Rev	-	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATCCACTTGCAACATGTAC
5	SSCS Exp-For	+	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
6	SSCS Exp-Rev	-	GTCTCGTGGGCTCGGAG
7	Nextera Indexing-For ^b	+	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTCGTCGGCAGCGTC
8	Nextera Indexing-Rev ^b	-	CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTCTCGTGGGCTCGG

a. Primers #1-4 were used in linear extension reactions for the unique tagging of templates and synthesis of complementary strands from HIV-1 or HIV-2 proviral DNA. Primers #5-6 were used to exponentially amplify doubly-tagged DNA via qPCR. Primers #7-8 were standard Illumina Nextera dual-indexing primers containing various 8-base indices for library demultiplexing. Blue = HIV-1 or HIV-2-specific region; Pink = Unique identifiers (14 random bases); Orange = Illumina adapter sequences; Purple = Indices.

b. Oligonucleotide sequences © 2007-2013 Illumina, Inc. All rights reserved.

References

1. **Leider JM, Palese P, Smith FI.** 1988. Determination of the mutation rate of a retrovirus. *J Virol* **62**:3084-3091.
2. **Monk RJ, Malik FG, Stokesberry D, Evans LH.** 1992. Direct determination of the point mutation rate of a murine retrovirus. *J Virol* **66**:3683-3689.
3. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**:5087-5094.
4. **O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD.** 2002. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J Biol Chem* **277**:38053-38061.
5. **Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ.** 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**:1422-1423.
6. **Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R.** 2010. Viral mutation rates. *J Virol* **84**:9733-9748.
7. **Schlub TE, Grimm AJ, Smyth RP, Cromer D, Chopra A, Mallal S, Venturi V, Waugh C, Mak J, Davenport MP.** 2014. 15-20% of HIV substitution mutations are associated with recombination. *J Virol* doi:10.1128/JVI.03136-13.
8. **McCulloch SD, Kunkel TA.** 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**:148-161.
9. **Rawson JM, Landman SR, Reilly CS, Mansky LM.** 2015. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* **12**:60.
10. **Dohm JC, Lottaz C, Borodina T, Himmelbauer H.** 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**:e105.
11. **Minoche AE, Dohm JC, Himmelbauer H.** 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**:R112.
12. **Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S.** 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* **39**:e90.
13. **Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y.** 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**:341.

14. **Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB.** 2013. Characterizing and measuring bias in sequence data. *Genome Biol* **14**:R51.
15. **Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C.** 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* **43**:e37.
16. **Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X.** 2015. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC Genomics* **16**:79.
17. **Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI.** 2013. The long-term stability of the human gut microbiota. *Science* **341**:1237439.
18. **Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R.** 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A* **108**:20166-20171.
19. **Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B.** 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**:9530-9535.
20. **Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangi JL.** 2013. Practical innovations for high-throughput amplicon sequencing. *Nat Methods* **10**:999-1002.
21. **Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA.** 2012. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* **109**:14508-14513.
22. **Gregory MT, Bertout JA, Ericson NG, Taylor SD, Mukherjee R, Robins HS, Drescher CW, Bielas JH.** 2015. Targeted single molecule mutation detection with massively parallel sequencing. *Nucleic Acids Res* doi:10.1093/nar/gkv915.
23. **Acevedo A, Brodsky L, Andino R.** 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**:686-690.
24. **Lou DI, Hussmann JA, McBee RM, Acevedo A, Andino R, Press WH, Sawyer SL.** 2013. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc Natl Acad Sci U S A* **110**:19872-19877.
25. **Zhou S, Jones C, Mieczkowski P, Swanstrom R.** 2015. Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. *J Virol* **89**:8540-8555.
26. **Ahn EH, Hirohata K, Kohn BF, Fox EJ, Chang CC, Loeb LA.** 2015. Detection of Ultra-Rare Mitochondrial Mutations in Breast Stem Cells by Duplex Sequencing. *PLoS One* **10**:e0136216.
27. **Kukita Y, Matoba R, Uchida J, Hamakawa T, Doki Y, Imamura F, Kato K.** 2015. High-fidelity target sequencing of individual molecules identified

- using barcode sequences: de novo detection and absolute quantitation of mutations in plasma cell-free DNA from cancer patients. *DNA Res* **22**:269-277.
28. **Campbell-Yesufu OT, Gandhi RT.** 2011. Update on human immunodeficiency virus (HIV)-2 infection. *Clin Infect Dis* **52**:780-787.
 29. **de Silva TI, Cotten M, Rowland-Jones SL.** 2008. HIV-2: the forgotten AIDS virus. *Trends Microbiol* **16**:588-595.
 30. **Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH.** 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* **84**:9864-9878.
 31. **Holtz CM, Mansky LM.** 2013. Variation of HIV-1 mutation spectra among cell types. *J Virol* **87**:5296-5299.
 32. **Mansky LM.** 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res Hum Retroviruses* **12**:307-314.
 33. **Mansky LM.** 1996. The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**:391-400.
 34. **Mansky LM, Preveral S, Selig L, Benarous R, Benichou S.** 2000. The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate. *J Virol* **74**:7039-7047.
 35. **Beach LB, Rawson JM, Kim B, Patterson SE, Mansky LM.** 2014. Novel inhibitors of human immunodeficiency virus type 2 infectivity. *J Gen Virol* **95**:2778-2783.
 36. **Dapp MJ, Clouser CL, Patterson S, Mansky LM.** 2009. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J Virol* **83**:11950-11958.
 37. **Martin M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**:10-12.
 38. **Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD.** 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**:31.
 39. **Wu TD, Nacu S.** 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**:873-881.
 40. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**:1297-1303.

APPENDIX II
RETROVIRAL VECTORS FOR ANALYSIS OF VIRAL
MUTAGENESIS AND RECOMBINATION

Reprinted with permission from: Rawson JM, Mansky LM. Retroviral Vectors for Analysis of Viral Mutagenesis and Recombination. *Viruses* (2014) 6/9, 3612-42. This article is openly distributed by MDPI under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Abstract

Retrovirus population diversity within infected hosts is commonly high due in part to elevated rates of replication, mutation, and recombination. This high genetic diversity often complicates the development of effective diagnostics, vaccines, and antiviral drugs. This review highlights the diverse vectors and approaches that have been used to examine mutation and recombination in retroviruses. Retroviral vectors for these purposes can broadly be divided into two categories: those that utilize reporter genes as mutation or recombination targets and those that utilize viral genes as targets of mutation or recombination. Reporter gene vectors greatly facilitate the detection, quantification, and characterization of mutants and/or recombinants, but may not fully recapitulate the patterns of mutagenesis or recombination observed in native viral gene sequences. In contrast, the detection of mutations or recombination events directly in viral genes is more biologically relevant but also typically more challenging and inefficient. We will highlight the advantages and disadvantages of the various vectors and approaches used as well as propose ways in which they could be improved.

Introduction

Retroviruses share a remarkable capacity to rapidly evolve, which can be attributed to several key variables: high rates of mutation, recombination, and replication, large numbers of infected cells, and strong positive selective pressures (1-3). These factors act together to drive the expansion of a small number of viruses that are initially transmitted to a particular host into a population of diverse but interacting variants, termed the viral quasispecies. Genetic diversification enables escape from the host immune response, accelerates the emergence of drug resistance, and promotes cross-species transmission. The ability of human retroviruses like human immunodeficiency virus type-1 (HIV-1) to rapidly evolve poses a tremendous challenge to the development of effective prophylactics and therapeutics.

Retroviruses mutate at an average rate of $\sim 3 \times 10^{-5}$ mutations/base pair (bp)/cycle, corresponding to roughly one mutation per three genomes synthesized (4-8). This mutation rate is ~ 10 - 1000 fold higher than DNA viruses (10^{-6} - 10^{-8} mutations/bp) (9) and at least $10,000$ -fold higher than for eukaryotic DNA ($\leq 10^{-9}$ mutations/bp) (10). Mutations could arise during several different steps of the retroviral life cycle. Reverse transcriptase (RT) converts the single-stranded viral RNA into double-stranded DNA and is thought to be one of the key drivers of viral mutagenesis, primarily due to its high error rates (typically 10^{-4} - 10^{-5} mutations/bp) (11) *in vitro*. However, RNA polymerase II (Pol II) can also generate mutations when transcribing the viral genomic RNA from the integrated proviral DNA (12), though the relative contribution of Pol II compared to RT has not been well defined. Cellular DNA polymerases can also create mutations when replicating the integrated provirus during cell division, but the high fidelity of cellular DNA replication argues that this is a relatively minor source of virus variation. Lastly, nucleic acid-editing enzymes such as the APOBEC3 family of cytosine deaminases can edit minus strand viral DNA during reverse transcription (13, 14), ultimately leading to G-to-A mutations on the plus strand viral DNA. APOBEC3-mediated editing is often lethal, as multiple G-to-A mutations are usually induced in the same provirus (i.e. G-to-A hypermutation), but editing has been reported to accelerate viral evolution in certain contexts (15, 16). Furthermore, while HIV-1 expresses an accessory protein called Vif to counteract APOBEC3 proteins, Vif alleles have been shown to vary widely in their abilities to counteract various APOBEC3 proteins (17-19). In addition, G-to-A hypermutants have often been observed in patient samples (17, 20-26), which provides strong evidence that APOBEC3-mediated editing occurs *in vivo*. All of these processes together contribute to the high mutation rates of retroviruses, leading to the evolution of variants that may confer drug resistance, improve transmissibility, or allow for cytotoxic T-cell or neutralizing antibody escape. Many additional variables have been shown to influence

retroviral mutagenesis, including RT variants (27-31), antiviral drugs (27, 28, 30, 32), and accessory proteins (28, 33-35).

In addition to mutating at high rates, retroviruses are also able to recombine at high rates (e.g. ~3-14 crossovers/genome/cycle for HIV-1) (36-42). Retroviruses are able to recombine rapidly because (unlike other RNA viruses) they co-package two copies of the RNA genome into every viral particle. During reverse transcription, recombination between RNA genomes often occurs but only one DNA provirus is ultimately formed, such that retroviruses are considered pseudodiploid in nature (43). Recombination events between identical co-packaged genomes are effectively silent since new viral sequences are not generated. In contrast, recombination events between distinct genomes (i.e., from heterozygous virions) can lead to novel viral variants. Recombination permits the completion of DNA synthesis even in the presence of RNA damage and also promotes the generation of novel variants through the shuffling of viral mutations. For example, recombination can link beneficial mutations together, such as low-level drug resistance mutations into a highly drug-resistant complex (44) or single-drug resistance mutations into multi-drug resistance (45). Conversely, recombination can allow escape from deleterious or lethal mutations (46-48). While often beneficial, recombination can also dissociate co-adapted mutations, leading to unfit viral variants (49). The importance of recombination to the establishment of the global AIDS pandemic is clear: 65 circulating recombinant forms (CRFs) have been identified to date (www.hiv.lanl.gov), accounting for ~ 20% of HIV-1 infections worldwide (50).

As a pre-requisite for recombination to occur between distinct viruses, the producer cells must become dually infected, either as a result of co-infection (near simultaneous infection) or super-infection (sequential infection). Thus, factors that alter the incidence of dual infection through either mechanism will impact the frequency of recombination. HIV-1 readily co-infects primary CD4⁺ T-cells *ex vivo* with little evidence of interference (51-54). In fact, co-infection has been found to occur more frequently than expected

from random interactions between viruses and cells (51, 52), but this may be due to reactivation of silent proviruses upon co-infection (55). In contrast to co-infection, cells infected by HIV-1 (and many other retroviruses) are resistant to re-infection, a phenomenon called super-infection resistance (56). Super-infection of humans at the organismal level has often been documented (57-62), though initial infection may be somewhat protective of re-infection (61). However, such individuals do not necessarily contain super-infected cells, as only a small fraction of CD4⁺ T-cells become infected. Nonetheless, the wide prevalence of intra-subtype, inter-subtype, and inter-group recombinants of HIV-1 demonstrate that cellular super-infection resistance is not absolute. Of note, dually infected splenocytes can readily be observed in samples from HIV-1 infected individuals (63, 64) or SIV_{mac}-infected rhesus macaques (65). However, another study found that most peripheral blood CD4⁺ T-cells harbor only a single provirus (66). Unfortunately, it is not yet clear whether these discrepancies are due to the different compartments sampled, unique features of the infected individuals, or other factors.

In addition to the incidence of dual infection, rates of retroviral recombination are influenced by factors that alter the ability of genetically distinct genomes to co-package into the same virus particle. For example, the HIV-1 genome contains the dimerization initiation signal (DIS), a six nucleotide palindromic sequence within the first stem loop of the 5' untranslated region. The DIS sequence is the dominant factor that drives HIV-1 genomic RNA co-dimerization and co-packaging (67-70). Thus, HIV-1 variants with matched DIS sequences ultimately recombine much more frequently than those with mismatched DIS sequences. Additionally, the rate of recombination depends on the frequency of template switching during reverse transcription. Template switching is thought to be controlled by the relative balance of polymerase and RNase H activities of RT, referred to as the dynamic copy choice model of recombination (71, 72). Factors that reduce the speed of RT-mediated DNA synthesis promote template switching, including low dNTP pool levels (71, 73-75), RNA secondary structures (76-

78), and mutations in RT that impair processivity (71, 74, 75). In contrast, RT mutations that reduce RNase H activity decrease the level of template switching (71, 74, 79). The level of template switching also correlates fairly well with the degree of sequence homology between templates, such that closely related sequences (e.g. from the same subtype of HIV-1) recombine more frequently than divergent sequences (e.g. from different subtypes of HIV-1) (70, 80).

In developed countries, the advent of highly active antiretroviral therapy (HAART) directed against HIV-1 has enabled the indefinite suppression of viral replication in the vast majority (>95%) of infected individuals, provided proper drug adherence is maintained. This raises the issue of the settings in which the high mutation and recombination rates of HIV-1 remain relevant. Prior to the initiation of treatment, rapid viral replication and diversification permit escape from CD8⁺ cytotoxic T-cell and neutralizing antibody responses, ultimately preventing effective immune system control of the virus. Defining the roles of viral mutation and recombination in these processes may inform efforts to develop an effective vaccine. Further, during HAART, a low level of ongoing viral replication may persist in areas of limited drug penetration, such as the gut-associated lymphoid tissue and central nervous system (81, 82), permitting continual viral evolution. However, the presence of ongoing viral replication during HAART has remained controversial, and the predominant mechanism of viral persistence is thought to be latent infection of resting CD4⁺ T-cells (83). HAART is also not completely suppressive in all individuals, as ~3% of infected individuals develop triple-class virological failure (84), though this may result from suboptimal drug adherence. Additionally, high rates of mutation and recombination are features often shared by other RNA viruses (9, 85-87), many of which cannot be effectively countered by drug treatment. The investigation of these processes in HIV-1 will serve as a useful model system for studying the genetic diversification of other RNA viruses.

Although many advances have been made, there are still a number of fundamental questions surrounding the nature of retroviral mutagenesis and recombination that remain unanswered. For example, very few studies have examined these processes in primary cells rather than immortalized cell lines. Differences in dNTP pool levels and expression of relevant cellular factors between immortalized cell lines and primary cells could have a large impact on viral mutagenesis and recombination. Even primary cell types can differ substantially in this regard, as the most relevant targets of HIV-1 infection, activated CD4⁺ T-cells and macrophages, differ by ~130-250-fold in dNTP pool levels (88). In addition, very few studies have examined mutagenesis or recombination in native viral genes rather than foreign reporter genes, primarily due to the increased difficulty of detecting such events. Given this, there remains a clear need for continued dissection of these processes in retroviruses, and improved vectors and methodologies to aid in these investigations.

Many different vectors and approaches have been used to examine retroviral mutagenesis and recombination. Retroviral vectors can broadly be divided into two categories: those that utilize reporter genes (e.g. antibiotic resistance genes, β -galactosidase, or fluorescent proteins) and those that utilize viral genes as targets for mutation or recombination. Reporter gene vectors greatly facilitate the detection, quantification, and characterization of mutants or recombinants, but the results may not be representative of native viral genes in regards to the frequency or types of mutation and/or recombination hotspots. In contrast, the detection of mutations or recombination events directly in viral genes is more biologically relevant, but the methods to do so are also much less efficient. Most of the retroviral vectors utilized for studies of viral mutagenesis and recombination are not replication-competent and must be trans-complemented in order to produce infectious vector virus. The benefit of this approach is that it allows for the more accurate quantification of mutation and recombination rates, as all events must occur in a single cycle of replication. In other instances,

replication-competent vectors have been used to examine retroviral mutagenesis and recombination, but they have been limited to a single round of replication by neutralizing antibodies or antiviral drugs. In contrast, multi-cycle replication assays are valuable for understanding how natural selection acts upon emerging mutants and recombinants, and some of these studies will briefly be addressed. Here, the design and application of a number of different retroviral vectors that have been employed in assays to investigate retroviral mutagenesis and recombination is reviewed. The benefits and drawbacks of each approach are highlighted and potential ways to improve current vectors and methodologies are explored.

Vectors for Examination of Retroviral Mutagenesis

Reporter Gene Vectors for Reversion Assays

Some of the earliest vectors for quantification of retroviral mutation rates relied upon the detection of mutations that restored defective drug resistance genes (see Table A2-1). Many of these initial studies utilized spleen necrosis virus (SNV)-based vectors. SNV is an avian gammaretrovirus capable of infecting several types of mammalian cells and has often been employed as a model system for retrovirus biology. The first mutation rate of a retrovirus was determined using a SNV-based vector containing a dysfunctional neomycin resistance gene (*neo*) and a functional hygromycin B resistance gene (*hph*), as shown in Figure A2-1A (89). The neomycin resistance gene had a single point mutation introducing a premature stop codon, and *neo* revertants could be selected for using the neomycin analog G-418. The reversion mutation frequency of SNV was determined by calculating the ratio of G-418-resistant (*neo*^r) to hygromycin-resistant (*hygro*^r) colonies and was found to be 2.2×10^{-5} . In this particular assay, the reversion mutation frequency was equivalent to the reversion mutation rate (defined as mutations/bp/cycle), as the size of the mutational target was a single base and the vector was limited to one round of replication.

Viral mutagenesis has been investigated in several other retroviruses using reversion vectors as well. The mutation rate of murine leukemia virus (MLV), another gammaretrovirus, has been determined using a *neo/hph* cassette similar to the one used for SNV (90). More recently, the mutagenesis of HIV-1 has been investigated using a defective luciferase gene containing an insertion of eight T residues near the 5' end of the gene (30). This vector was designed specifically to detect frameshift events, which are often triggered by homopolymeric runs (91-94). Relative mutation frequencies were determined by comparing ratios of relative light units to the viral titers, which were determined using an integrated LTR-driven marker gene in the same cell line. In addition to luciferase, vectors containing an inactivated blasticidin resistance gene (*bsr*) and enhanced yellow fluorescent protein (EYFP) have been used to measure mutation rates of several retroviruses, including avian leukosis virus (ALV), MLV, and HIV-1 (95). In these assays, reversion frequencies were defined as the ratio of blasticidin-resistant/EYFP⁺ cells to all EYFP⁺ cells.

Although used infrequently, reversion reporter vectors do greatly facilitate the detection and quantification of viral mutants, and sequencing can determine the nature of reverting mutations. However, these vectors can only detect the few specific mutations that are able to restore a functional gene phenotype. Thus, many types of mutations cannot be detected at all, and those mutations that can be observed are limited to very narrow sequence contexts. Therefore, calculated mutation rates may not be representative of the overall mutation rate within the reporter gene or virus. In addition, the influence of factors on the mutation spectra cannot readily be addressed. For these reasons, vectors that instead rely on the elimination of functional reporter genes (i.e. forward mutation assays) have been utilized much more frequently than vectors that measure reversion mutation frequencies.

2.2. Reporter Gene Vectors for Forward Mutational Assays

2.2.1 LacZ α /LacZ (β -galactosidase)

Rather than monitoring the restoration of defective reporter genes, forward mutational assays detect the elimination of functional reporter gene products. Thus, these assays can detect every mutational class, including the twelve types of substitutions, insertions, deletions, and more complex events. Furthermore, mutations can be detected in a variety of sequence contexts. Though many different reporter genes have been used in these assays (see Table A2-1), the *lacZ α* peptide gene (or, in some cases, the full-length *lacZ*) is one of the most widely used reporters for analyses of retroviral mutagenesis. The *lacZ α* gene encodes a small fragment (~150-170 bp) of β -galactosidase that can interact with another fragment (*lacZ ω*) to reconstitute full-length β -galactosidase. The formation of functional β -galactosidase can be detected by blue-white color screening in *E. coli*. This system has been used to determine forward mutation rates and spectra for a number of different retroviruses (see Table A2-1). Furthermore, this system has been used to characterize the impact of factors such as drugs, drug resistance-associated mutations, and viral accessory proteins on viral mutagenesis.

While specific details vary between laboratories, the general steps of the *lacZ α* assay are as follows: First, producer cells are established that contain an integrated packaging construct with a shuttle cassette. The shuttle cassette often consists of *lacZ α* , an antibiotic resistance gene (for selection of producer and target cells), and a bacterial origin of replication (see Figure A2-1B). Virus production is initiated by transient transfection with helper plasmids, and the collected virus is used to infect target cells. The virus can only replicate a single time in the target cells due to the lack of essential viral genes. Next, the *lacZ α* DNA is recovered, usually through a process that avoids the use of PCR, as PCR can generate background errors. For example, the *lacZ α* DNA can be recovered by purification with the Lac repressor protein (4) or by Hirt extraction (96), which recovers

extrachromosomal DNA. For Hirt extraction, vectors have been designed with inactivating mutations in integrase to facilitate the recovery of *lacZ α* DNAs. The DNA can then be transformed into bacteria, selected with antibiotics, and subjected to blue-white color screening using the chromogenic substrate X-gal. Mutation frequencies are calculated by dividing the number of mutant colonies (white or light blue) by the total number of colonies. Forward mutation rates can be calculated by dividing the mutation frequencies by the size of *lacZ α* or by the known number of detectable sites (97, 98) for mutations in *lacZ α* (this is possible because *lacZ α* has been extensively mutated). Mutation spectra can be determined by targeted sequencing of white colonies, leading to identification of a wide array of mutations.

2.2.2. Thymidine Kinase (TK)

In addition to *lacZ α* , the herpes simplex virus thymidine kinase (*tk*) gene has been used in forward reporter vectors in analyses of retroviral mutagenesis (Table A2-1). Thymidine kinase is a nucleotide salvage enzyme that converts thymidine into thymidine monophosphate (TMP). Cells expressing thymidine kinase are resistant to inhibitors of *de novo* TMP synthesis (e.g. aminopterin) but sensitive to the thymidine analog bromodeoxyuridine (BrdU). These vectors contain both *tk* as well as an antibiotic resistance gene (*hygro* or *neo*) for determination of viral titers (see Figure A2-1C). To use these vectors, producer cells are established, virus production is initiated by transfection of helper plasmids, and TK⁻ target cells are infected, such that the viral vector is the only source of TK. Forward mutation frequencies can be calculated by dividing the number of antibiotic-resistant (*hygro*^r or *neo*^r) and BrdU-resistant (BrdU^r) colonies by all antibiotic-resistant colonies. Forward mutation rates can be calculated by dividing mutation frequencies by the size of *tk* (~1.2 kb), and spectra can be examined by PCR and sequencing. Relative to *lacZ α* , *tk* vectors offer several advantages, such as the ability to propagate and further characterize mutant proviruses that survive drug selection. As *tk* is substantially longer than *lacZ α* (~1.2 kb and ~0.2 kb, respectively), the general likelihood of observing

mutations is greater. However, the increased length of *tk* also necessitates that more sequencing be performed to identify all mutations within the reporter gene. In addition, the use of PCR to recover *tk* sequences results in some level of background errors. Further, many types of cells cannot be used as target cells of infection in this assay because they express endogenous TK.

2.2.3. Fluorescent Proteins & Cell Surface Markers

HIV-1 mutagenesis has been investigated using several single-cycle reporter vectors that encode fluorescent proteins and/or cell surface markers, which can be stained by antibodies conjugated to fluorescent dyes. To construct the first vector (99), the mouse heat stable antigen (*hsa*, a cell surface marker) and enhanced green fluorescent protein (*egfp*) were inserted in place of *nef* within an HIV-1 envelope-deficient molecular clone. The reporter genes were separated by an internal ribosome entry site (IRES) element necessary for translation of EGFP. More recently, the *hsa* gene has been replaced with *mCherry* (Figure A2-1D), eliminating the need for antibody staining (100). In addition, a similar *mCherry/egfp* vector has been constructed for mutagenesis studies in human immunodeficiency virus type-2 (HIV-2) (101). In order to determine mutant frequencies, these vectors are co-transfected with an envelope plasmid into 293T cells. Viral stocks are collected and used to infect target cells, and flow cytometry is performed to quantify the number of infected cells (cells expressing at least one reporter) as well as the number of cells with mutant proviruses (cells expressing only one reporter). Mutant frequencies are calculated by dividing the number of mutants by the total number of infected cells. To assess mutational spectra, fluorescence-activated cell sorting (FACS) can be performed to isolate cells expressing only one reporter gene, from which genomic DNA can be purified. Lastly, the defective gene (e.g. *egfp* from *mCherry*⁺/*EGFP*⁻ cells) can be amplified by PCR, ligated into a cloning vector, and sequenced. To date, these vectors have been employed to examine the impact of small molecule mutagens (99-105), drug resistance-associated mutations (106), cell type

(107), and APOBEC3G (105, 108) on HIV-1 mutagenesis. The primary advantage of these vectors is that they allow for rapid determination of infectivities and mutant frequencies, and thus are ideally suited for the screening of small molecules for antiviral and/or mutagenic activities. However, cell sorting is required to enrich for mutant sequences prior to further processing, as the vast majority of sequences are wild-type in the absence of sorting. Further, PCR is required to amplify, clone, and sequence defective genes. Some mutations that eliminate expression of EGFP may occur within the IRES element necessary for EGFP translation, such that the entire IRES-EGFP region (~1.3 kb) must be sequenced to identify all responsible mutations.

2.3. Detection of Mutations in Viral Genes

2.3.1. Biochemical & Sanger Sequencing-Based Approaches

Rather than detecting mutations in reporter genes, some studies have assessed viral mutagenesis by directly detecting mutations in viral genes (Table A2-1). As reporter genes are unnecessary for this approach, a wide variety of retroviral vectors can be used, provided that they can be limited to a single round of replication. In addition to sequencing, mutations can be identified in viral genes by biochemical assays, including heteroduplex tracking (HTA), RNase T1 fingerprinting, and detection of single-strand conformation polymorphisms (SSCP). In HTA (109), viral RNA is annealed to a radiolabeled probe and subjected to denaturing-gradient gel electrophoresis. Mismatches caused by viral mutations, particularly in low-melting regions, can alter the thermodynamic stability of the heteroduplex and lead to altered migration during electrophoresis. For RNase T1 fingerprinting (110), radiolabeled viral RNA is digested with RNase T1, which cleaves at G residues in single-stranded RNA. The digested viral RNA is subjected to two-dimensional electrophoresis-homochromatography, generating a map, or fingerprint, of RNase T1-resistant oligonucleotides. Mutations within the larger RNase T1-resistant oligonucleotides often alter migration, allowing for their detection. Lastly, in SSCP (12), PCR of viral genes is performed from

infected cell clones. The PCR products are radiolabeled, denatured, and analyzed by non-denaturing gel electrophoresis. Again, mutations are detected by altered mobility during electrophoresis. The primary drawback of these biochemical approaches is that they are labor-intensive, and are thus not conducive to identifying and characterizing large numbers of viral mutants. Additionally, they cannot detect all mutations (RNase T1 fingerprinting and SSCP have been estimated to detect ~80% of changes) (12, 110), and thus calculated mutation rates underestimate true mutation rates.

In place of biochemical assays, other groups have simply amplified viral genes by PCR and identified mutations by Sanger sequencing (104, 111). In this instance, nearly all mutations can be identified, provided that they do not prevent amplification. Unfortunately, like biochemical assays, this approach is extremely low-throughput. For example, assuming a mutation rate of 3×10^{-5} mutations/bp/cycle (the retroviral average) and a target size of ~800 bp (the length of a typical Sanger read), ~0.024 mutations will be identified per read. Thus, in order to obtain 50 mutations, ~2000 Sanger reads would be required. Accordingly, studies following this approach have reported very low numbers of mutations, which hinders a robust determination of mutation rates and spectra. Additionally, without enriching for mutated sequences, the presence of background errors due to PCR and sequencing becomes a much greater concern, and the necessary controls to account for this must be included. To circumvent the inefficiency of Sanger sequencing, some groups have sequenced virus passaged multiple (~5-10) times (112, 113), but mutation rates cannot easily be calculated in these experiments due to purifying selection and uncertainty regarding the number of rounds of replication.

2.3.2. Next-Generation Sequencing-Based Approaches

The advent of next-generation sequencing (NGS) technologies has revolutionized biology, and retrovirology has been no exception. NGS has been used to infer co-receptor usage (114, 115), examine immune escape variants (116), and identify minority variants (117-120) that may contribute to drug resistance in samples from HIV-1-infected individuals. Illumina platforms

currently dominate the market due to their high outputs and low costs. Illumina can presently generate ~45-50 million 2×300 bp paired-end reads (MiSeq) or ~600 million 2×150 bp paired-end reads (HiSeq 2500 in rapid mode). While NGS can identify thousands of variants, NGS technologies are currently hindered by high error rates, due to both PCR and sequencing. Although sequencing libraries can be created in the absence of amplification, PCR is often employed to generate more material for sequencing and to enrich for adapter-ligated templates. The background error rates from NGS have been estimated to range from 10^{-2} mutations/bp for unprocessed data to 10^{-3} or 10^{-4} mutations/bp for processed data (121-125). Given this, the average retroviral mutation rate is minimally ~10 fold lower than the background error rate, which prevents distinction of biological mutations from background errors. This problem is illustrated by a recent study (40) in which HIV-1 was allowed to replicate for a single cycle in primary CD4⁺ T-cells, amplified by PCR, and subjected to 454 sequencing. The background error rate was determined by amplifying and sequencing plasmid DNA and found to be 7.4×10^{-5} mutations/bp, while the error rate in the biological sample was found to be 1.2×10^{-4} mutations/bp. Thus, assuming equivalent background levels, ~60% of the mutations were artifactual, and these mutations could not be distinguished from those induced during the viral life cycle. Nonetheless, the mutation rate of HIV-1 was estimated by subtracting the background error rate from the error rate of the biological sample, leading to an estimate of $\sim 4.6 \times 10^{-5}$ mutations/bp/cycle. Recently, two key strategies have been developed to help circumvent the high error rates of NGS. In the first strategy, which has been referred to as low-error amplicon sequencing (LEA-Seq) (126), unique identifier tags (UIDs), such as twelve random bases, are assigned to each template molecule, either through ligation, reverse transcription, or linear PCR (119, 126-129). Next, exponential PCR of the tagged molecules is performed, followed by redundant sequencing. Consensus families can be built from sequences with identical tags, allowing the identification and exclusion of most PCR and sequencing errors. LEA-Seq has been used for 16S rRNA

sequencing in microbial metagenomics (126, 128), identification of mutations in human mitochondrial and nuclear DNA (127, 129), determination of DNA polymerase fidelity during PCR (129), and tracking minority variants in the HIV-1 protease during drug exposure (119). In the second strategy, called circular resequencing (CirSeq), rolling circle amplification of template molecules is used to generate three copies from the same starting template molecule that are linked in tandem (130-132). After the linked copies are sequenced, consensus families are again built, permitting the removal of most background errors. CirSeq was recently used to determine the mutation rate and spectrum of poliovirus, as well as to track the fitness of poliovirus variants across serially passaged populations (132). While neither method has yet been used to determine mutation rates or spectra of retroviruses, both methods are likely to advance our understanding of how mutations arise in viral genes during retroviral replication.

3. Vectors for Examination of Retroviral Recombination

3.1. Reporter Gene Vectors

3.1.1. Antibiotic Resistance Genes

Similar to the initial vectors used to investigate retroviral mutation rates, the first vectors used to quantify a retroviral recombination rate relied upon a cassette containing two antibiotic resistance genes (*neo* and *hph*) in an SNV-derived vector (see Table A2-2 and Figure A2-2A) (43). Two similar vectors were engineered: one with a dysfunctional *neo* gene, and the other with a dysfunctional *hph* gene. In both cases, the respective antibiotic resistance genes were disrupted by 4-bp insertions (~1 kb apart), which were found to spontaneously revert at low rates. To examine recombination, dually-infected producer cell clones were created by sequential infection with the parental viruses at low multiplicities of infection (MOI). PCR and enzyme digestion were performed to ensure that producer cell clones contained a single copy of each parental virus and that recombination did not occur during the generation of producer cell clones. While more laborious than simply co-transfecting viral vectors, this procedure eliminates transfection-induced recombination and

more closely mimics natural viral replication. Viruses were collected from producer cell clones and used to infect target cells at low MOIs. The recombination rate was determined by dividing the titer of double drug-resistant (*neo^r/hygro^r*) colonies by the lower of the two single drug-resistant titers. Using this method, the recombination rate of SNV was found to be ~2%/kilobase/cycle (43). Similar MLV-based vectors were later devised and used to show that recombination rates increase with the distance between genetic markers (133). Additionally, MLV-based vectors have been used to characterize the effects of moving the dimerization linkage structure on recombination rates (134). More recently, HIV-1-based *neo/hph* vectors were used to demonstrate that the HIV-1 recombination rate is ~10-fold higher than in gammaretroviruses (41). Notably, experiments with these vectors must be performed at low MOIs, as co-infection can produce the same phenotype as a recombinant. Often, Southern blotting of recombinants is required to definitively demonstrate that double-resistant cell clones are not due to co-infection. In addition, this system is not easily amenable to other cell lines and primary cell types, as antibiotic sensitivities can vary widely between cell types.

3.1.2. Fluorescent Proteins

Retroviral recombination rates have often been measured with fluorescent proteins (Table A2-2), eliminating the need for antibiotic selection of target cells. Importantly, these vectors facilitate the analysis of recombination in a variety of cell types (including primary cells), as the antibiotic sensitivities of different cell lines do not need to be considered. In one vector system that has often been used for HIV-1, recombination rates are measured using combinations of vectors with distinct inactivating mutations in *gfp* (38). Specifically, these vectors contain cassettes encoding a cell surface marker (HSA or Thy-1) used to track infectivity, an IRES element, and an inactivated GFP as the recombination target. The expression of GFP can be eliminated by introducing mutations at various positions allowing for measurement of

recombination at distances ranging from ~100 to 600 bp. To determine recombination rates, producer cell clones were first created by sequential infection at low MOIs (0.05-0.1). Cell sorting of infected cells was used to purify dually infected producer cells, which express both cell surface markers (HSA and Thy-1). Virus production from producer cells was initiated by the transfection of helper plasmids, and collected viruses were used to infect target cells. Recombination rates in target cells were calculated by comparing the frequency of GFP⁺ cells to all infected cells (HSA⁺ and/or Thy-1⁺). As the experiments were performed at relatively high MOIs (0.4-0.5), some cells were very likely co-infected. In contrast to the dual-antibiotic resistance vectors described in Section 3.1.1., co-infection with the inactivated GFP vectors does not result in a recombinant phenotype. However, co-infection does influence the calculation of the total number of infected cells, which in turn could alter calculated recombination rates. Therefore, the numbers of infected cells were converted to MOIs based upon a Poisson distribution of co-infection in order to more accurately estimate recombination rates (38). In some studies, further analyses were done to confirm recombination events in *gfp* and map recombination breakpoints in viral genes (69, 135). In order to accomplish this, single GFP⁺ cell clones were isolated by cell sorting and propagated independently in order to prevent PCR-mediated recombination between proviruses. PCR and restriction mapping were used to verify reconstitution of functional *gfp* genes, as the inactivating mutations in *gfp* were designed to introduce specific restriction sites. Sanger sequencing of PCR products was used to further confirm recombination in *gfp*. Additional PCR and Sanger sequencing of GFP⁺ cell clones has often been performed to identify crossovers in viral sequences as well. Of note, rather than isolating single cell clones, a modified type of PCR called single-genome sequencing (SGS) has also been used to prevent PCR-mediated recombination (49). In SGS, templates (e.g. viral RNA or proviral DNA) are diluted such that <30% of reactions result in products. Under these conditions, the vast majority of

reactions contain only a single template, thus preventing most PCR-mediated recombination between templates.

Initially, vectors based on inactivated GFP were used to examine the impact of distance, cell type (primary vs immortalized T-cells), and accessory genes on the recombination rate of HIV-1 (38). Similar vectors were later used to measure recombination rates in non-B subtypes of HIV-1 as well as rates of intersubtype or intergroup recombination (49, 67-70). Further, these vectors have been used to study the impact of cell type (T-cells, monocytes, and macrophages) (136) on recombination. In addition to HIV-1, analogous vectors have been engineered for HIV-2 and simian immunodeficiency virus from African green monkeys (SIV_{agm}), both of which were found to recombine at similar rates to HIV-1 (137). These vectors have also been used to demonstrate that HIV-1 and HIV-2 are able to co-package and recombine (135), albeit at very low rates (recombination frequency of 0.2%, ~35-fold lower than for intrasubtype recombination in HIV-1).

In addition to GFP, several other fluorescent proteins have been used in recombination vectors such as enhanced yellow and cyan fluorescent proteins (EYFP and ECFP). These vectors take advantage of the high level of homology between different fluorescent proteins. Specifically, infectious molecular clones of HIV-1 have been developed that encode EYFP or ECFP in place of the accessory protein Nef, with Nef expressed from a downstream IRES element (see Figure A2-2B) (54). As the critical residues in *yfp* and *cfp* are ~400 bp apart, recombination between these residues gives rise to a modified *gfp* gene (called *gfp**) that can be spectrally separated from *yfp* with the appropriate filters. Thus, recombination frequencies were calculated by determining the ratio of GFP⁺ to EYFP or ECFP⁺ cells during flow cytometry. To map recombination breakpoints in *gfp**, single GFP⁺ cell clones were isolated by FACS, amplified by PCR, and sequenced. Sequencing was also performed to map crossovers in viral genes, which could easily be identified as the vectors were engineered in two distinct isolates of HIV-1. Although these vectors express all viral genes and are replication-competent,

replication can still be limited to a single round of replication through use of an antiviral drug.

3.1.3. LacZ (β -galactosidase)

Recombination rates have occasionally been determined using other reporter genes such as the full-length *lacZ* gene encoding β -galactosidase (42). For example, HIV-1 and MLV vector pairs have been designed in which one encodes a functional *lacZ* gene but no antibiotic resistance gene, while the other encodes a defective *lacZ* gene and a puromycin resistance gene (*puro*). Recombination rates were calculated by determining the ratio of *lacZ*⁺/puromycin-resistant (*puro*^r) colonies (i.e. *puro*^r colonies that stain blue in the presence of X-gal) to all *puro*^r colonies. These vectors have been used to compare frequencies of recombination between MLV and HIV-1, as well as to investigate factors responsible for observed differences between these viruses.

3.1.4. Direct Repeat Reporter Vectors

Directly repeated sequences in retroviral genomes and vectors have long been known to delete at high frequencies during reverse transcription (138-141). Deletions occur at high rates regardless of whether the repeats are in tandem or separated by additional sequences. In one of the earliest observations of this phenomenon, the *src* oncogene was found to be frequently lost during RSV replication due to flanking direct repeats (139, 140). Direct repeat deletions also contribute to the variable copy number of repeats that have been observed in the enhancer sequences of retroviral LTRs (142, 143). These deletions have been shown to occur primarily through intramolecular template switching events, rather than through homologous recombination between separate viral genomes (144). Retroviral vectors have been developed that take advantage of the high efficiency of direct repeat deletion to assay the impact of various factors on template switching. These vectors have large (>100 bp) inactivating duplications in reporter genes, such as *tk*, *lacZ*, or *gfp*. Direct repeat deletion eliminates the

duplicated sequence, thus restoring the reporter gene phenotype. In one early study, an MLV vector was engineered with a *tk/neo* cassette in which *tk* contained an internal 700 bp duplication (145). Deletion frequencies were first assessed by comparing the ratio of *tk*⁺ (i.e. aminopterin-resistant)/*neo*^r colonies to all *neo*^r colonies. Southern blotting and densitometry were performed to determine deletion frequencies in a more quantitative fashion. In several other studies, MLV and HIV-1-based vectors were engineered in which *lacZ* reporter genes were inactivated by internal duplications (42, 73, 79, 146, 147). Deletion frequencies were calculated by comparing titers of *lacZ*⁺/*puro*^r cell colonies to all *puro*^r colonies. Lastly, MLV and HIV-1 vectors with internal duplications (of 200-250 bp) in *gfp* have been developed (71, 72, 74). With these vectors, direct repeat deletion leads to reconstitution of functional GFP, and deletion frequencies can be determined as the ratio of GFP⁺/antibiotic-resistant cell colonies to all antibiotic-resistant cell colonies. More recently, a similar HIV-1 vector was developed in which the antibiotic resistance gene was replaced with the cell surface marker *hsc*, facilitating use in primary cells (148). As direct repeat deletion is very efficient during reverse transcription, the ability of various factors to impact deletion frequencies can readily be determined, even when such effects are relatively small. However, as discussed, direct repeat deletions primarily occur through intramolecular template switching events rather than intermolecular template switching events (144). Thus, the outcomes observed in these studies may not necessarily be representative of retroviral recombination, which requires intermolecular template switching events between distinct viruses.

3.2. Detection of Recombinants in Viral Genes

3.2.1. Gag Reconstitution Assay

Recently, a new HIV-1 vector system has been engineered that is based on the reconstitution of functional Gag protein (Figure A2-2C) (70). This vector system combines many of the best features of reporter assays with the increased biological relevance of viral genes. The concept behind the vector system was straightforward: Gag is expressed at high levels on the plasma

membranes of infected target cells, such that Gag itself can be treated as a reporter gene with a reliable anti-Gag antibody. More specifically, multiple pairs of vectors were engineered in different subtypes and groups of HIV-1 with distinct inactivating mutations separated by ~300 bp in the capsid portion of the Gag polyprotein. In addition, the *gfp* genes were inactivated in these vectors, allowing for simultaneous detection of recombination in identical (*gfp*) and non-identical (*gag*) target sequences. Cell-surface markers such as HSA and B7 (i.e. human CD80) permitted tracking of viral infectivity. These vectors were sequentially introduced into producer cells, complemented with Gag-Pol in trans to produce infectious virus, and used to infect target cells. Target cells were stained with a PE-conjugated anti-p24 (capsid) antibody to detect recombination events that led to expression of full-length Gag. The authors showed that the particular antibody they used could reliably detect a variety of parental Gag proteins from different subtypes and groups of HIV-1, as well as recombinants between subtypes and groups. Recombination rates were determined by comparing the frequency of Gag⁺ cells to all infected cells (as determined by the cell surface markers). Furthermore, since these vectors also contain defective forms of *gfp*, recombination rates could simultaneously be determined in identical (albeit foreign) gene sequences across viruses. While representing a tremendous improvement over standard reporter gene vectors, these vectors are limited by a small dynamic range of the assay due to the fixed distance (~300 bp of capsid) of the mutations. However, suitable antibodies that recognize other regions of the unprocessed Gag polyprotein may be identified in the future, allowing for analysis of a larger region of Gag. Generally, this approach would likely not be amenable for identifying recombinants in other viral genes and non-coding regions of the genome. Additionally, some Gag recombinants may not be identified if, for example, they prevent proper folding or recognition by the antibody.

3.2.2. Use of LacZ (β -galactosidase) to Detect Recombinants in Viral Genes

Additional retroviral vectors have been developed that take advantage of reporter gene phenotypes but actually detect recombination events within viral

targets. Specifically, vector pairs for different groups and subtypes of HIV-1 have been designed that contain either a functional *lacZ* or a completely unrelated sequence in its place (such as a defective *malT* gene in reverse orientation), as demonstrated in Figure A2-2D (76, 149, 150). After the *lacZ* gene, *env* genes from various viral isolates were inserted, followed by a BamHI restriction site (in the *lacZ* vector) or no restriction site (in the *lacZ*⁺ vector). After PCR amplification, proviral DNA was digested with the BamHI restriction enzyme (in conjunction with PstI), ligated into a cloning vector, and subjected to blue-white color screening. Recombination in *env* results in *lacZ*⁺ proviruses with the BamHI restriction site, which can then efficiently ligate into the cloning vector. The recombination frequency is then computed by determining the ratio of *lacZ*⁺ colonies to all colonies. In this approach, all observed recombination events must have occurred within *env* rather than *lacZ*, as one vector completely lacks the *lacZ* gene. When *env* genes from different viral isolates are used, sequencing can be performed to identify the exact breakpoints. While this approach is able to identify recombinants in viral genes, it is somewhat laborious and involves the placement of viral genes outside of their native sequence contexts.

3.2.3. Heteroduplex Tracking, Quantitative PCR, and Restriction Mapping

Numerous biochemistry and molecular biology-based techniques have been used to directly identify and quantify retroviral recombinants occurring within viral genes, without any assistance from reporter genes. Most often, these assays rely upon the detection of sequence differences between distinct viral isolates, either from the same subtype (intrasubtype) or different subtypes (intersubtype), groups, or types of HIV. Thus, recombination frequencies from these assays may somewhat underestimate recombination rates between highly homologous viruses. Even when these assays are able to detect recombination between nearly identical viruses (by, for example, restriction mapping), the recombination breakpoints cannot be better defined unless silent point mutations have been introduced to create intervening intervals. These types of approaches have been used to identify

recombinants in viral genes during both single cycle and multi-cycle infections. In spreading infections, viral recombinants are subject to strong purifying selection, which can restrict observable recombination events. In some cases, parental retroviral vectors have been used that contain different defects in essential viral genes (such as deletions in *pol* and *env*), such that only recombinants are able to spread after the initial round of replication (49, 151, 152).

In several reports, heteroduplex tracking (HTA) has been performed to identify recombinants in viral genes (36, 37, 153), which can detect recombinants (as well as mutants) when there are sufficient sequence differences between the parental viruses. Recombination can alter the affinity of a viral sequence for a radioactive probe (relative to the two parental DNAs), which can be detected as a shift in band pattern during gel electrophoresis. HTA has been estimated to reliably detect sequence divergence when the mismatches exceed ~1% (154, 155), and thus this technique can only identify crossovers introducing >1% mismatch relative to the parental viruses. Therefore, some recombinants cannot be detected by this method, particularly when the parental isolates are quite similar (e.g from the same HIV-1 subtype). This method also requires sequencing of fragments to confirm that gel shifts were not caused by mutations, insertions, or deletions. Restriction mapping, Southern blotting (with short oligonucleotides specific to one of the parental isolates), and/or Sanger sequencing can be used to further refine recombination breakpoints between the viruses.

As an alternative to HTA, several studies have used PCR or restriction mapping to detect recombination (37, 45, 149, 153, 156, 157). PCR has been performed using primers that are specific to one of the two parental viruses. By combining one primer from each virus, recombinants can be detected. Additionally, different forms of quantitative PCR have been used to determine starting amounts of recombinant and parental DNAs, enabling calculation of recombination frequencies. As for HTA, Southern blotting or Sanger sequencing can be used to more precisely identify crossover points between

distinct viral isolates. These approaches have been used to track HIV-1 intersubtype viral recombinants, and have led to identification of recombination hotspots as well as characterization of the potential mutagenicity of recombination.

3.2.4. Direct Sequencing of Viral Genes-Sanger

As discussed above, Sanger sequencing of viral genes has often been employed to further characterize recombinants identified by techniques such as HTA, PCR, or restriction mapping. Sanger sequencing has often also been used to find crossovers in viral genes from proviruses that are recombinants in reporter genes. In addition, recombinants have been identified simply by direct amplification and sequencing of viral genes from infected cells (37, 39, 157). The low-throughput nature of direct Sanger sequencing is somewhat less problematic for studies of retroviral recombination than for mutagenesis. This is because some retroviruses, such as lentiviruses, recombine ~10 times more frequently than they mutate. While sequencing is often used to map recombination breakpoints between heterogeneous viruses, HIV-1 vectors have recently been engineered that allow detection of recombination between highly homologous viruses, thus more accurately representing the viral population found within infected individuals (39). Specifically, silent mutations that could be detected by PCR and Sanger sequencing were introduced throughout the *gag* gene of an HIV-1 infectious molecular clone. These silent mutations served as markers for recombination and did not alter the protein sequences or any known cis elements. Further, they were shown to have no effect on viral infectivity. Most markers actually consisted of two mutations, typically spaced three bases apart, to facilitate differentiation of markers from mutations or background errors. Notably, the vectors they designed encode functional open reading frames for all viral proteins. In order to investigate HIV-1 recombination, viral stocks were first produced by co-transfecting the wild-type and marker plasmids into 293T cells. This is in contrast to many earlier studies in which producer cell clones were generated by sequential infection or, in some cases, co-infection with parental viruses (38, 41, 43, 67-

69, 133-137, 144, 158, 159). While sequential infection is more time-consuming, it eliminates the possibility of transfection-induced recombination (TIR) and more closely resembles native viral replication. However, the investigators of this study measured the frequency of TIR by direct sequencing of plasmid DNA from co-transfected cells and found that the frequency of TIR was much lower than the recombination rates measured for HIV-1. Next, the authors collected virus from co-transfected cells, infected primary T-cells, performed PCR, and identified recombinants by Sanger sequencing. Virus replication was limited to a single round using the T-20 fusion inhibitor. All PCR reactions were terminated in the log-linear phase of amplification, which reduces the frequency of PCR-mediated recombination (160). Additional controls were performed to show that PCR-induced and interviral recombination did not substantially contribute to the observed data. Further, the authors estimated the fraction of heterozygous virions directly from the sequencing data and developed a mathematical model that accounted for the possibility of multiple template switches, based on distances between genetic markers and observed recombination rates. Using this vector system, the HIV-1 recombination rate was observed to be 0.81×10^{-3} recombination events per nucleotide (REPN), corresponding to a true recombination rate of 1.35×10^{-3} REPN (or 12.5 crossovers/genome) after considering the possibility of multiple template switches.

3.2.5. Direct Sequencing of Viral Genes-NGS

In several subsequent studies, similar HIV-1 vectors with marker mutations throughout the *gag* and *pol* genes have been applied to 454 sequencing (40, 161). As described above, wild-type and marker vectors were co-transfected into 293T cells, and viruses were collected and used to infect primary blood mononuclear cells. Although replication-competent, the virus was limited to a single round of replication with fusion inhibitor. Proviral DNA was amplified by PCR, sequencing libraries were prepared, and 454 sequencing was performed, generating ~1 million reads per run. Experimental sources of recombination, such as transfection-induced recombination, interviral

recombination, and PCR, were carefully measured in parallel. These background sources were found to result in recombination rates less than 10% of the recombination rate measured in the biological samples. Using these vectors and 454 sequencing, the overall rates of recombination in HIV-1 *gag* and *pol* were found to be $1.5\text{-}2 \times 10^{-3}$ REPN, or ~14-19 crossovers per genome. In addition, using the large amounts of data generated by 454, recombination hot and cold spots could be identified in viral genes (161). Further, ~15-20% of mutations were found to be associated with recombination, though the direction of causality could not be determined (40). As the association of recombination and mutation is itself a topic of great research interest, the high background error rates of NGS technologies are also somewhat problematic for studies of viral recombination. Thus, recent improvements aimed at lowering NGS error rates (see Section 2.3.2.) will surely benefit recombination studies as well. Additionally, these improvements often permit the identification and exclusion of PCR-induced crossovers as well as errors, thus limiting the effects of PCR-mediated recombination on the observed data.

4. Conclusions

Given the wide variety of retroviral vectors available, choosing a particular vector to assess retroviral mutagenesis and/or recombination can be daunting. Retroviral vectors based on reporter genes offer a number of distinct benefits that have led to their widespread use in mutation and recombination studies. First, mutants or recombinants can be efficiently detected based on phenotypes that are easily observed and quantified, while mutants or recombinants in viral targets must be detected through much more laborious means. Targeted sequencing of mutant viruses can determine the nature of the underlying mutations, while targeted sequencing of recombinants can confirm that recombination took place (though the breakpoint locations cannot be precisely defined). In addition, reporter gene vectors are generally somewhat less dependent on PCR for identification of mutants or recombinants, and thus relatively less impeded by PCR-induced

errors or crossovers. Nonetheless, reporter gene-based approaches are limited in that they may not necessarily be reflective of viral gene sequences. Large amounts of evidence have shown that patterns of mutation and recombination are affected by factors like homopolymeric runs and secondary structures (76-78, 91-94, 153), which may vary between template sequences. Given this, the degree to which reporter genes accurately represent viral genes cannot be easily ascertained at present, as so few studies have examined large numbers of mutations or crossovers in viral genes. Additionally, reporter gene vectors typically do not express all viral genes, though some exceptions do exist. In contrast, native viral genes are viewed as more desirable targets of mutation and recombination, but detection of mutation or recombination is often very laborious and time-consuming. Next-generation sequencing (NGS) technologies offer an obvious solution to these barriers, but standard NGS methods result in high background error rates, hindering identification of true mutations. Additionally, NGS is heavily reliant on PCR for sample preparation, which, under standard conditions, generates high levels of artifactual mutations and recombination events. However, several strategies for NGS sample preparation have recently been established that allow for identification and exclusion of the vast majority of background errors and crossovers due to PCR or sequencing (119, 127, 129, 130, 132). While these methods have not yet been used to quantify retroviral mutation or recombination rates, they will be key to advancing our understanding of retroviral mutagenesis and recombination in viral genes. In the near future, standardized retroviral protocols, vectors, and assays will hopefully be developed that allow for application of these improvements to a wide variety of questions in retrovirology.

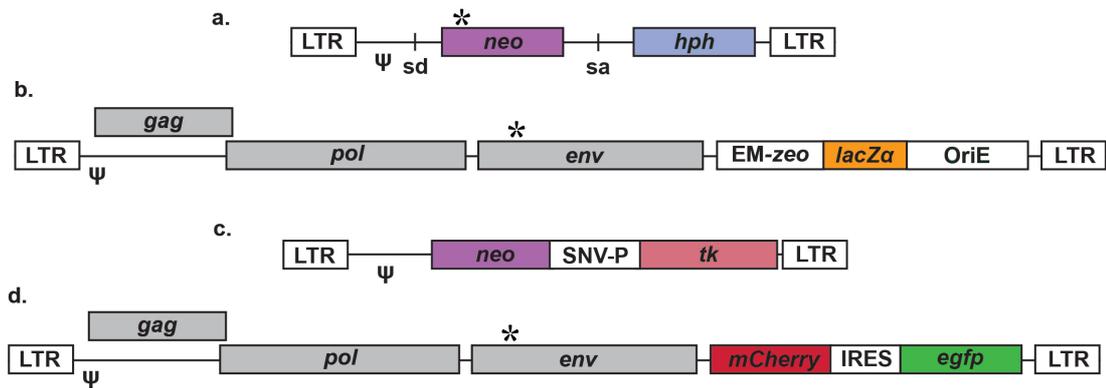


Figure A2-1. Vectors for the examination of retroviral mutagenesis. Most accessory genes and cis elements are not shown for simplicity; * represents inactivating mutations, while ψ represents packaging signals. **(a)** SNV vector to detect mutations that restore *neo*, conferring G-418 resistance (89); sd = splice donor, sa = splice acceptor. **(b)** HIV-1 vector that detects mutations in *lacZ α* that prevent complementation and expression of functional β -galactosidase in *E. coli*, ultimately resulting in white bacterial colonies in the presence of X-gal (96); EM = bacterial promoter, *zeo* = zeomycin resistance gene, OriE = *E. coli* origin of replication. **(c)** MLV vector to monitor mutations that eliminate thymidine kinase (*tk*), granting resistance to the thymidine analog bromodeoxyuridine (BrdU) (8); SNV-P = SNV promoter (U3). **(d)** HIV-1 vector to detect mutations that eliminate expression of mCherry or enhanced green fluorescent protein (EGFP) by flow cytometry (100); IRES = internal ribosome entry site.

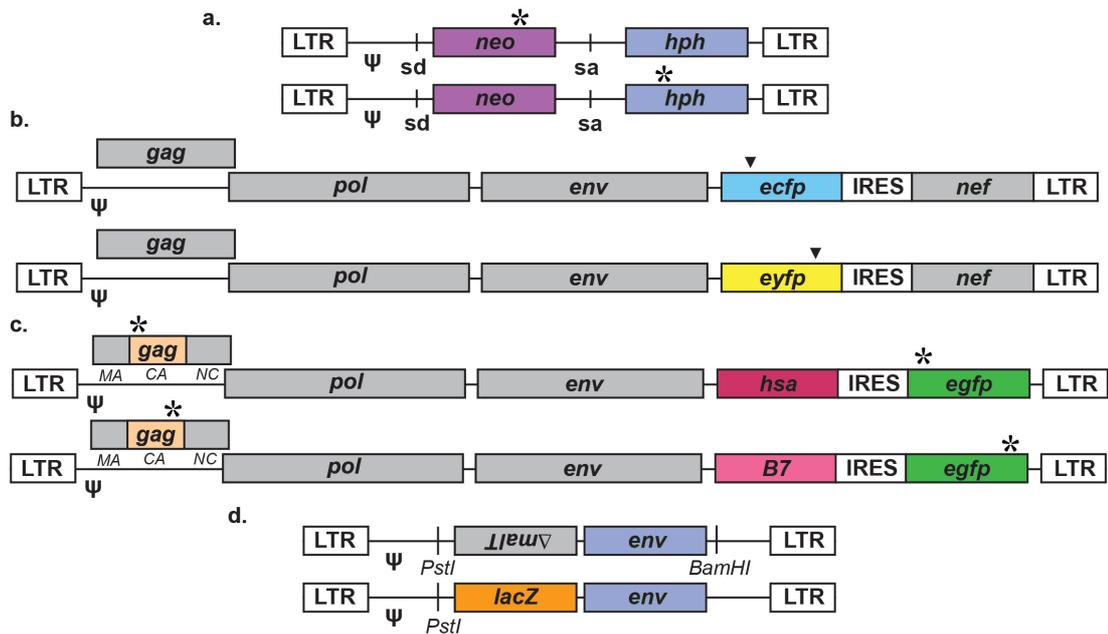


Figure A2-2. Vector pairs for the investigation of retroviral recombination. Most accessory genes and cis elements are not shown for simplicity; * represents inactivating mutations, while ψ represents packaging signals. **(a)** SNV vector pair that detects recombination events that restore *neo* and *hph*, resulting in viruses that confer G-418 and hygromycin B double drug-resistance to infected cells (43); sd = splice donor, sa = splice acceptor. **(b)** HIV-1 vector pair that detects recombination events between *ecfp* and *eyfp*, which leads to expression of a modified GFP (54); \blacktriangledown = essential mutations in *ecfp* and *eyfp*. **(c)** HIV-1 vector pair that can simultaneously detect recombination events that restore the capsid (CA) portion of the Gag polyprotein (by staining with an appropriate antibody) and enhanced green fluorescent protein (EGFP) (70). Viral infection is monitored with cell surface markers like the mouse heat stable antigen (HSA) and B7; IRES = internal ribosome entry site. **(d)** HIV-1 vector pair that detects recombination events that occur within a region of homology (such as the viral *env*) inserted between β -galactosidase (*lacZ*) and a BamHI restriction site (76).

Table A2-1. Targets for analysis of retroviral mutagenesis.

Target¹	Assay Type²	Virus	References
<i>neo</i>	Reporter (R)	SNV, MLV	(89, 90)
<i>tk</i>	Reporter (F)	MLV, HIV-1	(8, 162, 163)
<i>lacZ</i> / <i>lacZ</i>	Reporter (F)	SNV, BLV, HTLV-1, HIV-1	(4-7, 27-29, 31-35, 96, 164-166)
<i>luc</i>	Reporter (R)	HIV-1	(30)
<i>bsr</i>	Reporter (R)	ALV, MLV, HIV-1	(95)
<i>hsa</i>	Reporter (F)	HIV-1	(107, 108)
<i>egfp</i>	Reporter (F)	HIV-1	(99, 100, 102, 105)
Viral	HTA	RSV	(109)
	RNase T ₁	MLV	(110)
	SSCP	HIV-1	(12)
	Sequencing (Sanger)	HIV-1	(104, 111)
	Sequencing (NGS)	HIV-1	(40)

¹Reporter gene targets include antibiotic resistance genes (*neo*, *bsr*), thymidine kinase (*tk*), β -galactosidase (*lacZ* α or full-length *lacZ*), the cell surface marker heat stable antigen (*hsa*), and enhanced green fluorescent protein (*egfp*).

²Reporter gene targets are included for both reversion assays (R) as well as forward mutational assays (F). Assays that can detect mutations directly in viral genes include heteroduplex tracking (HTA), RNase T₁ fingerprinting, single-strand conformation polymorphisms (SSCP), Sanger sequencing, and next-generation sequencing (NGS) technologies.

Table A2-2. Targets for analysis of retroviral recombination.

Target ¹	Assay Type ²	Virus	References
<i>neo/hph</i>	Reporter Gene	SNV, MLV, HIV-1	(41, 43, 133, 134, 144, 158, 159)
<i>tk</i>	Reporter Gene	MLV	(145)
<i>lacZ</i>	Reporter Gene	MLV, HIV-1	(42)
<i>gfp</i>	Reporter Gene	MLV, HIV-1, HIV-2, SIV _{agm}	(38, 49, 67-70, 135-137)
<i>ecfp/eyfp</i>	Reporter Gene	HIV-1	(54)
viral	Gag	HIV-1	(70)
	Reconstitution		
	Modified LacZ	HIV-1	(76, 149, 150)
	PCR/qPCR	HIV-1	(149, 153, 156, 157)
	HTA	HIV-1	(36, 37, 153)
	Restriction	HIV-1	(37, 45)
	Mapping		
	Sequencing (Sanger)	HIV-1	(37, 39, 45, 157)
	Sequencing (NGS)	HIV-1	(40, 161)

¹Reporter gene targets include antibiotic resistance genes (*neo*, *hph*), thymidine kinase (*tk*), β -galactosidase (*lacZ*), and fluorescent proteins (*gfp*, *ecfp*, *eyfp*).

²Reporter gene assays typically detect recombination events that rescue defective gene products. Other assays have been used to detect recombination events in viral genes, such as a Gag reconstitution assay (in which Gag itself is treated as a reporter gene), a modified LacZ assay, PCR or quantitative PCR, heteroduplex tracking (HTA), restriction digest mapping, Sanger sequencing, and next-generation sequencing (NGS) technologies.

References

1. **Delviks-Frankenberry K, Galli A, Nikolaitchik O, Mens H, Pathak VK, Hu WS.** 2011. Mechanisms and factors that influence high frequency retroviral recombination. *Viruses* **3**:1650-1680.
2. **Smyth RP, Davenport MP, Mak J.** 2012. The origin of genetic diversity in HIV-1. *Virus Res* **169**:415-429.
3. **Hu WS, Hughes SH.** 2012. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**.
4. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**:5087-5094.
5. **Pathak VK, Temin HM.** 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc Natl Acad Sci U S A* **87**:6019-6023.
6. **Mansky LM.** 2000. In vivo analysis of human T-cell leukemia virus type 1 reverse transcription accuracy. *J Virol* **74**:9525-9531.
7. **Mansky LM, Temin HM.** 1994. Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus. *J Virol* **68**:494-499.
8. **Varela-Echavarria A, Prorock CM, Ron Y, Dougherty JP.** 1993. High rate of genetic rearrangement during replication of a Moloney murine leukemia virus-based vector. *J Virol* **67**:6357-6364.
9. **Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R.** 2010. Viral mutation rates. *J Virol* **84**:9733-9748.
10. **McCulloch SD, Kunkel TA.** 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**:148-161.
11. **Menendez-Arias L.** 2009. Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses* **1**:1137-1165.
12. **O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD.** 2002. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J Biol Chem* **277**:38053-38061.
13. **Refsland EW, Harris RS.** 2013. The APOBEC3 family of retroelement restriction factors. *Curr Top Microbiol Immunol* **371**:1-27.
14. **Desimie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK.** 2014. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol* **426**:1220-1245.
15. **Mulder LC, Harari A, Simon V.** 2008. Cytidine deamination induced HIV-1 drug resistance. *Proc Natl Acad Sci U S A* **105**:5501-5506.
16. **Kim EY, Bhattacharya T, Kunstman K, Swantek P, Koning FA, Malim MH, Wolinsky SM.** 2010. Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J Virol* **84**:10402-10405.

17. **Simon V, Zennou V, Murray D, Huang Y, Ho DD, Bieniasz PD.** 2005. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* **1**:e6.
18. **Iwabu Y, Kinomoto M, Tatsumi M, Fujita H, Shimura M, Tanaka Y, Ishizaka Y, Nolan D, Mallal S, Sata T, Tokunaga K.** 2010. Differential anti-APOBEC3G activity of HIV-1 Vif proteins derived from different subtypes. *J Biol Chem* **285**:35350-35358.
19. **Binka M, Ooms M, Steward M, Simon V.** 2012. The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. *J Virol* **86**:49-59.
20. **Fourati S, Lambert-Niclot S, Soulie C, Wirden M, Malet I, Valantin MA, Tubiana R, Simon A, Katlama C, Carcelain G, Calvez V, Marcelin AG.** 2014. Differential impact of APOBEC3-driven mutagenesis on HIV evolution in diverse anatomical compartments. *AIDS* **28**:487-491.
21. **Janini M, Rogers M, Birx DR, McCutchan FE.** 2001. Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. *J Virol* **75**:7973-7986.
22. **Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, Siliciano RF.** 2005. G-->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. *J Virol* **79**:1975-1980.
23. **Kijak GH, Janini LM, Tovanabutra S, Sanders-Buell E, Arroyo MA, Robb ML, Michael NL, Birx DL, McCutchan FE.** 2008. Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells. *Virology* **376**:101-111.
24. **Pace C, Keller J, Nolan D, James I, Gaudieri S, Moore C, Mallal S.** 2006. Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. *J Virol* **80**:9259-9269.
25. **Wei M, Xing H, Hong K, Huang H, Tang H, Qin G, Shao Y.** 2004. Biased G-to-A hypermutation in HIV-1 proviral DNA from a long-term non-progressor. *AIDS* **18**:1863-1865.
26. **Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, Daniels M, Ferrari G, Haynes BF, McMichael A, Shaw GM, Hahn BH, Korber B, Seoighe C.** 2009. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog* **5**:e1000414.
27. **Mansky LM, Bernard LC.** 2000. 3'-Azido-3'-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J Virol* **74**:9532-9539.
28. **Mansky LM, Le Rouzic E, Benichou S, Gajary LC.** 2003. Influence of reverse transcriptase variants, drugs, and Vpr on human

- immunodeficiency virus type 1 mutant frequencies. *J Virol* **77**:2071-2080.
29. **Weiss KK, Chen R, Skasko M, Reynolds HM, Lee K, Bambara RA, Mansky LM, Kim B.** 2004. A role for dNTP binding of human immunodeficiency virus type 1 reverse transcriptase in viral mutagenesis. *Biochemistry* **43**:4490-4500.
 30. **Chen R, Yokoyama M, Sato H, Reilly C, Mansky LM.** 2005. Human immunodeficiency virus mutagenesis during antiviral therapy: impact of drug-resistant reverse transcriptase and nucleoside and nonnucleoside reverse transcriptase inhibitors on human immunodeficiency virus type 1 mutation frequencies. *J Virol* **79**:12045-12057.
 31. **Abram ME, Ferris AL, Das K, Quinones O, Shao W, Tuske S, Alvord WG, Arnold E, Hughes SH.** 2014. Mutations in HIV-1 Reverse Transcriptase Affect the Errors Made in a Single Cycle of Viral Replication. *J Virol* **88**:7589-7601.
 32. **Mansky LM.** 2003. Mutagenic outcome of combined antiviral drug treatment during human immunodeficiency virus type 1 replication. *Virology* **307**:116-121.
 33. **Mansky LM.** 1996. The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**:391-400.
 34. **Mansky LM, Preveral S, Selig L, Benarous R, Benichou S.** 2000. The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate. *J Virol* **74**:7039-7047.
 35. **Chen R, Le Rouzic E, Kearney JA, Mansky LM, Benichou S.** 2004. Vpr-mediated incorporation of UNG2 into HIV-1 particles is required to modulate the virus mutation rate and for replication in macrophages. *J Biol Chem* **279**:28419-28425.
 36. **Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP.** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol* **74**:1234-1240.
 37. **Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, Preston BD, Dougherty JP.** 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* **76**:11273-11282.
 38. **Rhodes TD, Nikolaitchik O, Chen J, Powell D, Hu WS.** 2005. Genetic recombination of human immunodeficiency virus type 1 in one round of viral replication: effects of genetic distance, target cells, accessory genes, and lack of high negative interference in crossover events. *J Virol* **79**:1666-1677.
 39. **Schlub TE, Smyth RP, Grimm AJ, Mak J, Davenport MP.** 2010. Accurately measuring recombination between closely related HIV-1 genomes. *PLoS Comput Biol* **6**:e1000766.
 40. **Schlub TE, Grimm AJ, Smyth RP, Cromer D, Chopra A, Mallal S, Venturi V, Waugh C, Mak J, Davenport MP.** 2014. Fifteen to twenty

- percent of HIV substitution mutations are associated with recombination. *J Virol* **88**:3837-3849.
41. **Rhodes T, Wargo H, Hu WS.** 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. *J Virol* **77**:11193-11200.
 42. **Onafuwa A, An W, Robson ND, Telesnitsky A.** 2003. Human immunodeficiency virus type 1 genetic recombination is more frequent than that of Moloney murine leukemia virus despite similar template switching rates. *J Virol* **77**:4577-4587.
 43. **Hu WS, Temin HM.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**:1556-1560.
 44. **Kellam P, Larder BA.** 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *J Virol* **69**:669-674.
 45. **Moutouh L, Corbeil J, Richman DD.** 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc Natl Acad Sci U S A* **93**:6106-6111.
 46. **Kim EY, Busch M, Abel K, Fritts L, Bustamante P, Stanton J, Lu D, Wu S, Glowczwskie J, Rourke T, Bogdan D, Piatak M, Jr., Lifson JD, Desrosiers RC, Wolinsky S, Miller CJ.** 2005. Retroviral recombination in vivo: viral replication patterns and genetic structure of simian immunodeficiency virus (SIV) populations in rhesus macaques after simultaneous or sequential intravaginal inoculation with SIVmac239Deltavpx/Deltavpr and SIVmac239Deltanef. *J Virol* **79**:4886-4895.
 47. **Quan Y, Liang C, Brenner BG, Wainberg MA.** 2009. Multidrug-resistant variants of HIV type 1 (HIV-1) can exist in cells as defective quasispecies and be rescued by superinfection with other defective HIV-1 variants. *J Infect Dis* **200**:1479-1483.
 48. **Quan Y, Xu H, Wainberg MA.** 2014. Defective HIV-1 quasispecies in the form of multiply drug-resistant proviral DNA within cells can be rescued by superinfection with different subtype variants of HIV-1 and by HIV-2 and SIV. *J Antimicrob Chemother* **69**:21-27.
 49. **Galli A, Kearney M, Nikolaitchik OA, Yu S, Chin MP, Maldarelli F, Coffin JM, Pathak VK, Hu WS.** 2010. Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J Virol* **84**:7651-7661.
 50. **Hemelaar J, Gouws E, Ghys PD, Osmanov S.** 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**:679-689.
 51. **Chen J, Dang Q, Unutmaz D, Pathak VK, Maldarelli F, Powell D, Hu WS.** 2005. Mechanisms of nonrandom human immunodeficiency virus type 1 infection and double infection: preference in virus entry is important but is not the sole factor. *J Virol* **79**:4140-4149.

52. **Dang Q, Chen J, Unutmaz D, Coffin JM, Pathak VK, Powell D, KewalRamani VN, Maldarelli F, Hu WS.** 2004. Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. *Proc Natl Acad Sci U S A* **101**:632-637.
53. **Del Portillo A, Tripodi J, Najfeld V, Wodarz D, Levy DN, Chen BK.** 2011. Multiploid inheritance of HIV-1 during cell-to-cell infection. *J Virol* **85**:7169-7176.
54. **Levy DN, Aldrovandi GM, Kutsch O, Shaw GM.** 2004. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci U S A* **101**:4204-4209.
55. **Bregnard C, Pacini G, Danos O, Basmaciogullari S.** 2012. Suboptimal provirus expression explains apparent nonrandom cell coinfection with HIV-1. *J Virol* **86**:8810-8820.
56. **Nethe M, Berkhout B, van der Kuyl AC.** 2005. Retroviral superinfection resistance. *Retrovirology* **2**:52.
57. **Chohan B, Lavreys L, Rainwater SM, Overbaugh J.** 2005. Evidence for frequent reinfection with human immunodeficiency virus type 1 of a different subtype. *J Virol* **79**:10701-10708.
58. **Piantadosi A, Chohan B, Chohan V, McClelland RS, Overbaugh J.** 2007. Chronic HIV-1 infection frequently fails to protect against superinfection. *PLoS Pathog* **3**:e177.
59. **Piantadosi A, Ngayo MO, Chohan B, Overbaugh J.** 2008. Examination of a second region of the HIV type 1 genome reveals additional cases of superinfection. *AIDS Res Hum Retroviruses* **24**:1221.
60. **Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, Ricklefs SM, Tobian AA, Xiao C, Grabowski MK, Nalugoda F, Kigozi G, Laeyendecker O, Kagaayi J, Sewankambo N, Gray RH, Porcella SF, Wawer MJ, Quinn TC.** 2012. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J Infect Dis* **206**:267-274.
61. **Ronen K, McCoy CO, Matsen FA, Boyd DF, Emery S, Odem-Davis K, Jaoko W, Mandaliya K, McClelland RS, Richardson BA, Overbaugh J.** 2013. HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLoS Pathog* **9**:e1003593.
62. **Smith DM, Wong JK, Hightower GK, Ignacio CC, Koelsch KK, Daar ES, Richman DD, Little SJ.** 2004. Incidence of HIV superinfection following primary infection. *JAMA* **292**:1177-1178.
63. **Gratton S, Cheynier R, Dumaurier MJ, Oksenhendler E, Wain-Hobson S.** 2000. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers. *Proc Natl Acad Sci U S A* **97**:14566-14571.
64. **Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, Fischer U, Meese E, Wain-Hobson S, Meyerhans A.** 2002. Recombination: Multiply infected spleen cells in HIV patients. *Nature* **418**:144.

65. **Schultz A, Sopper S, Sauermann U, Meyerhans A, Suspene R.** 2012. Stable multi-infection of splenocytes during SIV infection--the basis for continuous recombination. *Retrovirology* **9**:31.
66. **Josefsson L, King MS, Makitalo B, Brannstrom J, Shao W, Maldarelli F, Kearney MF, Hu WS, Chen J, Gaines H, Mellors JW, Albert J, Coffin JM, Palmer SE.** 2011. Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc Natl Acad Sci U S A* **108**:11199-11204.
67. **Chin MP, Rhodes TD, Chen J, Fu W, Hu WS.** 2005. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci U S A* **102**:9002-9007.
68. **Chin MP, Chen J, Nikolaitchik OA, Hu WS.** 2007. Molecular determinants of HIV-1 intersubtype recombination potential. *Virology* **363**:437-446.
69. **Chin MP, Lee SK, Chen J, Nikolaitchik OA, Powell DA, Fivash MJ, Jr., Hu WS.** 2008. Long-range recombination gradient between HIV-1 subtypes B and C variants caused by sequence differences in the dimerization initiation signal region. *J Mol Biol* **377**:1324-1333.
70. **Nikolaitchik OA, Galli A, Moore MD, Pathak VK, Hu WS.** 2011. Multiple barriers to recombination between divergent HIV-1 variants revealed by a dual-marker recombination assay. *J Mol Biol* **407**:521-531.
71. **Svarovskaia ES, Delviks KA, Hwang CK, Pathak VK.** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. *J Virol* **74**:7171-7178.
72. **Hwang CK, Svarovskaia ES, Pathak VK.** 2001. 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* **98**:12209-12214.
73. **Pfeiffer JK, Topping RS, Shin NH, Telesnitsky A.** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription. *J Virol* **73**:8441-8447.
74. **Nikolenko GN, Svarovskaia ES, Delviks KA, Pathak VK.** 2004. Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. *J Virol* **78**:8761-8770.
75. **Operario DJ, Balakrishnan M, Bambara RA, Kim B.** 2006. Reduced dNTP interaction of human immunodeficiency virus type 1 reverse transcriptase promotes strand transfer. *J Biol Chem* **281**:32113-32121.
76. **Galetto R, Moumen A, Giacomoni V, Veron M, Charneau P, Negroni M.** 2004. The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *J Biol Chem* **279**:36625-36632.
77. **DeStefano JJ, Mallaber LM, Rodriguez-Rodriguez L, Fay PJ, Bambara RA.** 1992. Requirements for strand transfer between internal

- regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J Virol* **66**:6370-6378.
78. **Wu W, Blumberg BM, Fay PJ, Bambara RA.** 1995. Strand transfer mediated by human immunodeficiency virus reverse transcriptase in vitro is promoted by pausing and results in misincorporation. *J Biol Chem* **270**:325-332.
 79. **Brincat JL, Pfeiffer JK, Telesnitsky A.** 2002. RNase H activity is required for high-frequency repeat deletion during Moloney murine leukemia virus replication. *J Virol* **76**:88-95.
 80. **An W, Telesnitsky A.** 2002. Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J Virol* **76**:7897-7902.
 81. **Chun TW, Nickle DC, Justement JS, Meyers JH, Roby G, Hallahan CW, Kottlilil S, Moir S, Mican JM, Mullins JI, Ward DJ, Kovacs JA, Mannon PJ, Fauci AS.** 2008. Persistence of HIV in gut-associated lymphoid tissue despite long-term antiretroviral therapy. *J Infect Dis* **197**:714-720.
 82. **Gras G, Kaul M.** 2010. Molecular mechanisms of neuroinvasion by monocytes-macrophages in HIV-1 infection. *Retrovirology* **7**:30.
 83. **Ruelas DS, Greene WC.** 2013. An integrated overview of HIV-1 latency. *Cell* **155**:519-529.
 84. **Pursuing Later Treatment Options Iipt, Collaboration of Observational HIVREG, Nakagawa F, Lodwick R, Costagliola D, van Sighem A, Torti C, Podzamczar D, Mocroft A, Ledergerber B, Dorrucci M, Cozzi-Lepri A, Jansen K, Masquelier B, Garcia F, De Wit S, Stephan C, Obel N, Fatkenhaeuer G, Castagna A, Sambatakou H, Mussini C, Ghosn J, Zangerle R, Duval X, Meyer L, Perez-Hoyos S, Fabre Colin C, Kjaer J, Chene G, Grarup J, Phillips A.** 2012. Calendar time trends in the incidence and prevalence of triple-class virologic failure in antiretroviral drug-experienced people with HIV in Europe. *J Acquir Immune Defic Syndr* **59**:294-299.
 85. **Galli A, Bukh J.** 2014. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends Microbiol* **22**:354-364.
 86. **Lin CC, Yang ZW, Iang SB, Chao M.** 2014. Reduced genetic distance and high replication levels increase the RNA recombination rate of hepatitis delta virus. *Virus Res* doi:10.1016/j.virusres.2014.08.011.
 87. **Lowry K, Woodman A, Cook J, Evans DJ.** 2014. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates. *PLoS Pathog* **10**:e1004191.
 88. **Diamond TL, Roshal M, Jamburuthugoda VK, Reynolds HM, Merriam AR, Lee KY, Balakrishnan M, Bambara RA, Planelles V, Dewhurst S, Kim B.** 2004. Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* **279**:51545-51553.

89. **Dougherty JP, Temin HM.** 1988. Determination of the rate of base-pair substitution and insertion mutations in retrovirus replication. *J Virol* **62**:2817-2822.
90. **Varela-Echavarria A, Garvey N, Preston BD, Dougherty JP.** 1992. Comparison of Moloney murine leukemia virus mutation rate with the fidelity of its reverse transcriptase in vitro. *J Biol Chem* **267**:24681-24688.
91. **Bebenek K, Abbotts J, Roberts JD, Wilson SH, Kunkel TA.** 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J Biol Chem* **264**:16948-16956.
92. **Boyer JC, Bebenek K, Kunkel TA.** 1992. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc Natl Acad Sci U S A* **89**:6919-6923.
93. **Eckert KA, Kunkel TA.** 1993. Fidelity of DNA synthesis catalyzed by human DNA polymerase alpha and HIV-1 reverse transcriptase: effect of reaction pH. *Nucleic Acids Res* **21**:5212-5220.
94. **Ji J, Loeb LA.** 1994. Fidelity of HIV-1 reverse transcriptase copying a hypervariable region of the HIV-1 env gene. *Virology* **199**:323-330.
95. **Laakso MM, Sutton RE.** 2006. Replicative fidelity of lentiviral vectors produced by transient transfection. *Virology* **348**:406-417.
96. **Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH.** 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* **84**:9864-9878.
97. **Bebenek K, Kunkel TA.** 1995. Analyzing fidelity of DNA polymerases. *Methods Enzymol* **262**:217-232.
98. **Boyer JC, Bebenek K, Kunkel TA.** 1996. Analyzing the fidelity of reverse transcription and transcription. *Methods Enzymol* **275**:523-537.
99. **Dapp MJ, Clouser CL, Patterson S, Mansky LM.** 2009. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J Virol* **83**:11950-11958.
100. **Rawson JM, Heineman RH, Beach LB, Martin JL, Schnettler EK, Dapp MJ, Patterson SE, Mansky LM.** 2013. 5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors. *Bioorg Med Chem* **21**:7222-7228.
101. **Beach LB, Rawson JM, Kim B, Patterson SE, Mansky LM.** 2014. Novel inhibitors of human immunodeficiency virus type 2 infectivity. *J Gen Virol* doi:10.1099/vir.0.069864-0.
102. **Clouser CL, Patterson SE, Mansky LM.** 2010. Exploiting drug repositioning for discovery of a novel HIV combination therapy. *J Virol* **84**:9301-9309.
103. **Clouser CL, Chauhan J, Bess MA, van Oploo JL, Zhou D, Dimick-Gray S, Mansky LM, Patterson SE.** 2012. Anti-HIV-1 activity of resveratrol derivatives and synergistic inhibition of HIV-1 by the combination of resveratrol and decitabine. *Bioorg Med Chem Lett* **22**:6642-6646.

104. **Dapp MJ, Bonnac L, Patterson SE, Mansky LM.** 2014. Discovery of novel ribonucleoside analogs with activity against human immunodeficiency virus type 1. *J Virol* **88**:354-363.
105. **Dapp MJ, Holtz CM, Mansky LM.** 2012. Concomitant lethal mutagenesis of human immunodeficiency virus type 1. *J Mol Biol* **419**:158-170.
106. **Dapp MJ, Heineman RH, Mansky LM.** 2013. Interrelationship between HIV-1 fitness and mutation rate. *J Mol Biol* **425**:41-53.
107. **Holtz CM, Mansky LM.** 2013. Variation of HIV-1 mutation spectra among cell types. *J Virol* **87**:5296-5299.
108. **Sadler HA, Stenglein MD, Harris RS, Mansky LM.** 2010. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J Virol* **84**:7396-7404.
109. **Leider JM, Palese P, Smith FI.** 1988. Determination of the mutation rate of a retrovirus. *J Virol* **62**:3084-3091.
110. **Monk RJ, Malik FG, Stokesberry D, Evans LH.** 1992. Direct determination of the point mutation rate of a murine retrovirus. *J Virol* **66**:3683-3689.
111. **Gao F, Chen Y, Levy DN, Conway JA, Kepler TB, Hui H.** 2004. Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious. *J Virol* **78**:2426-2433.
112. **Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI.** 1999. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc Natl Acad Sci U S A* **96**:1492-1497.
113. **Harris KS, Brabant W, Styrchak S, Gall A, Daifuku R.** 2005. KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. *Antiviral Res* **67**:1-9.
114. **Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, Thielen A, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR.** 2011. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. *J Infect Dis* **203**:237-245.
115. **Vandenbroucke I, Van Marck H, Mostmans W, Van Eygen V, Rondelez E, Thys K, Van Baelen K, Franssen K, Vaira D, Kabeya K, De Wit S, Florence E, Moutschen M, Vandekerckhove L, Verhofstede C, Stuyver LJ.** 2010. HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. *AIDS Res Ther* **7**:4.
116. **Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT.** 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One* **5**:e12303.

117. **Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, Baxter JD, Huang C, Lubeski C, Turenchalk GS, Braverman MS, Desany B, Rothberg JM, Egholm M, Kozal MJ, Terry Beirn Community Programs for Clinical Research on A.** 2009. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis* **199**:693-701.
118. **Hedskog C, Mild M, Jernberg J, Sherwood E, Bratt G, Leitner T, Lundeberg J, Andersson B, Albert J.** 2010. Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS One* **5**:e11345.
119. **Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R.** 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A* **108**:20166-20171.
120. **Li JZ, Chapman B, Charlebois P, Hofmann O, Weiner B, Porter AJ, Samuel R, Vardhanabhuti S, Zheng L, Eron J, Taiwo B, Zody MC, Henn MR, Kuritzkes DR, Hide W, Team AAS, Wilson CC, Berzins BI, Acosta EP, Bastow B, Kim PS, Read SW, Janik J, Meres DS, Lederman MM, Mong-Kryspin L, Shaw KE, Zimmerman LG, Leavitt R, De La Rosa G, Jennings A.** 2014. Comparison of illumina and 454 deep sequencing in participants failing raltegravir-based antiretroviral therapy. *PLoS One* **9**:e90485.
121. **Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-UI-Amin M, Ogasawara N, Kanaya S.** 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* **39**:e90.
122. **Gilles A, Meglec E, Pech N, Ferreira S, Malausa T, Martin JF.** 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **12**:245.
123. **Vandenbroucke I, Van Marck H, Verhasselt P, Thys K, Mostmans W, Dumont S, Van Eygen V, Coen K, Tuefferd M, Aerssens J.** 2011. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. *Biotechniques* **51**:167-177.
124. **Minoche AE, Dohm JC, Himmelbauer H.** 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**:R112.
125. **Dohm JC, Lottaz C, Borodina T, Himmelbauer H.** 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**:e105.
126. **Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI.** 2013. The long-term stability of the human gut microbiota. *Science* **341**:1237439.

127. **Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA.** 2012. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* **109**:14508-14513.
128. **Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL.** 2013. Practical innovations for high-throughput amplicon sequencing. *Nat Methods* **10**:999-1002.
129. **Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B.** 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**:9530-9535.
130. **Lou DI, Hussmann JA, McBee RM, Acevedo A, Andino R, Press WH, Sawyer SL.** 2013. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc Natl Acad Sci U S A* **110**:19872-19877.
131. **Acevedo A, Andino R.** 2014. Library preparation for highly accurate population sequencing of RNA viruses. *Nat Protoc* **9**:1760-1769.
132. **Acevedo A, Brodsky L, Andino R.** 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**:686-690.
133. **Anderson JA, Bowman EH, Hu WS.** 1998. Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation. *J Virol* **72**:1195-1202.
134. **Anderson JA, Pathak VK, Hu WS.** 2000. Effect of the murine leukemia virus extended packaging signal on the rates and locations of retroviral recombination. *J Virol* **74**:6953-6963.
135. **Motomura K, Chen J, Hu WS.** 2008. Genetic recombination between human immunodeficiency virus type 1 (HIV-1) and HIV-2, two distinct human lentiviruses. *J Virol* **82**:1923-1933.
136. **Chen J, Rhodes TD, Hu WS.** 2005. Comparison of the genetic recombination rates of human immunodeficiency virus type 1 in macrophages and T cells. *J Virol* **79**:9337-9340.
137. **Chen J, Powell D, Hu WS.** 2006. High frequency of genetic recombination is a common feature of primate lentivirus replication. *J Virol* **80**:9651-9658.
138. **Czernilofsky AP, Levinson AD, Varmus HE, Bishop JM, Tischler E, Goodman HM.** 1980. Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product. *Nature* **287**:198-203.
139. **Hughes S, Kosik E.** 1984. Mutagenesis of the region between env and src of the SR-A strain of Rous sarcoma virus for the purpose of constructing helper-independent vectors. *Virology* **136**:89-99.
140. **Omer CA, Pogue-Geile K, Guntaka R, Staskus KA, Faras AJ.** 1983. Involvement of directly repeated sequences in the generation of deletions of the avian sarcoma virus src gene. *J Virol* **47**:380-382.
141. **Rhode BW, Emerman M, Temin HM.** 1987. Instability of large direct repeats in retrovirus vectors. *J Virol* **61**:925-927.

142. **Lovmand S, Kjeldgaard NO, Jorgensen P, Pedersen FS.** 1990. Enhancer functions in U3 of Akv virus: a role for cooperativity of a tandem repeat unit and its flanking DNA sequences. *J Virol* **64**:3185-3191.
143. **Trainor CD, Scott ML, Josephs SF, Fry KE, Reitz MS, Jr.** 1984. Nucleotide sequence of the large terminal repeat of two different strains of gibbon ape leukemia virus. *Virology* **137**:201-205.
144. **Hu WS, Bowman EH, Delviks KA, Pathak VK.** 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *J Virol* **71**:6028-6036.
145. **Delviks KA, Pathak VK.** 1999. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching. *J Virol* **73**:7923-7932.
146. **Pfeiffer JK, Georgiadis MM, Telesnitsky A.** 2000. Structure-based moloney murine leukemia virus reverse transcriptase mutants with altered intracellular direct-repeat deletion frequencies. *J Virol* **74**:9629-9636.
147. **An W, Telesnitsky A.** 2001. Frequency of direct repeat deletion in a human immunodeficiency virus type 1 vector during reverse transcription in human cells. *Virology* **286**:475-482.
148. **Nguyen LA, Kim DH, Daly MB, Allan KC, Kim B.** 2014. Host SAMHD1 protein promotes HIV-1 recombination in macrophages. *J Biol Chem* **289**:16642.
149. **Baird HA, Gao Y, Galetto R, Lalonde M, Anthony RM, Giacomoni V, Abreha M, Destefano JJ, Negroni M, Arts EJ.** 2006. Influence of sequence identity and unique breakpoints on the frequency of intersubtype HIV-1 recombination. *Retrovirology* **3**:91.
150. **Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefevre P, Martin DP, Robertson DL, Negroni M.** 2009. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. *PLoS Pathog* **5**:e1000418.
151. **Clavel F, Hoggan MD, Willey RL, Strebel K, Martin MA, Repaske R.** 1989. Genetic recombination of human immunodeficiency virus. *J Virol* **63**:1455-1459.
152. **St Louis DC, Gotte D, Sanders-Buell E, Ritchey DW, Salminen MO, Carr JK, McCutchan FE.** 1998. Infectious molecular clones with the nonhomologous dimer initiation sequences found in different subtypes of human immunodeficiency virus type 1 can recombine and initiate a spreading infection in vitro. *J Virol* **72**:3991-3998.
153. **Quinones-Mateu ME, Gao Y, Ball SC, Marozsan AJ, Abreha A, Arts EJ.** 2002. In vitro intersubtype recombinants of human immunodeficiency virus type 1: comparison to recent and circulating in vivo recombinant forms. *J Virol* **76**:9600-9613.
154. **Delwart EL, Shpaer EG, Louwagie J, McCutchan FE, Grez M, Rubsamen-Waigmann H, Mullins JI.** 1993. Genetic relationships

- determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. *Science* **262**:1257-1261.
155. **Zhu T, Wang N, Carr A, Nam DS, Moor-Jankowski R, Cooper DA, Ho DD.** 1996. Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. *J Virol* **70**:3098-3107.
 156. **Kuwata T, Miyazaki Y, Igarashi T, Takehisa J, Hayami M.** 1997. The rapid spread of recombinants during a natural in vitro infection with two human immunodeficiency virus type 1 strains. *J Virol* **71**:7088-7091.
 157. **Iglesias-Sanchez MJ, Lopez-Galindez C.** 2002. Analysis, quantification, and evolutionary consequences of HIV-1 in vitro recombination. *Virology* **304**:392-402.
 158. **Anderson JA, Teufel RJ, 2nd, Yin PD, Hu WS.** 1998. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol* **72**:1186-1194.
 159. **Bircher LA, Rigano JC, Ponferrada VG, Wooley DP.** 2002. High fidelity of homologous retroviral recombination in cell culture. *Arch Virol* **147**:1665-1683.
 160. **Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S, Davenport MP, Mak J.** 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**:45-51.
 161. **Smyth RP, Schlub TE, Grimm AJ, Waugh C, Ellenberg P, Chopra A, Mallal S, Cromer D, Mak J, Davenport MP.** 2014. Identifying recombination hot spots in the HIV-1 genome. *J Virol* **88**:2891-2902.
 162. **Parthasarathi S, Varela-Echavarria A, Ron Y, Preston BD, Dougherty JP.** 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. *J Virol* **69**:7991-8000.
 163. **Huang KJ, Wooley DP.** 2005. A new cell-based assay for measuring the forward mutation rate of HIV-1. *J Virol Methods* **124**:95-104.
 164. **Kim T, Mudry RA, Jr., Rexrode CA, 2nd, Pathak VK.** 1996. Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. *J Virol* **70**:7594-7602.
 165. **Mansky LM.** 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res Hum Retroviruses* **12**:307-314.
 166. **Pathak VK, Temin HM.** 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc Natl Acad Sci U S A* **87**:6024-6028.

APPENDIX III
RAPID DETERMINATION OF HIV-1 MUTANT FREQUENCIES AND
MUTATION SPECTRA USING AN MCHERRY/EGFP DUAL REPORTER
VIRAL VECTOR

Rawson JM, Clouser CL, Mansky LM. Rapid determination of HIV-1 mutant frequencies and mutation spectra using an mCherry/EGFP dual reporter viral vector. *Method Mol Biol* (2015) *In Press*.

Summary

The high mutation rate of human immunodeficiency virus type-1 (HIV-1) has been a pivotal factor in its evolutionary success as a human pathogen, driving the emergence of drug resistance, immune system escape, and invasion of distinct anatomical compartments. Extensive research has focused on understanding how various cellular and viral factors alter the rates and types of mutations produced during viral replication. Here, we describe a single-cycle dual-reporter vector assay that relies upon the detection of mutations that eliminate either expression of mCherry or enhanced green fluorescent protein (EGFP). The reporter-based method can be used to efficiently quantify changes in mutant frequencies and mutation spectra that arise due to a variety of factors, including viral mutagens, drug resistance mutations, cellular physiology, and APOBEC3 proteins.

Introduction

Human immunodeficiency virus type-1 (HIV-1) mutates at a rate of $\sim 3.6 \times 10^{-5}$ mutations/base pair (bp)/cycle (average across all studies; range of $1.4\text{--}8.5 \times 10^{-5}$), corresponding to about one mutation per three genomes synthesized (1-8). The rapid mutation rate, in conjunction with high rates of recombination and replication (9, 10), drives viral evolution within infected individuals and poses a significant barrier to the development of effective vaccines and drug regimens. Several studies have demonstrated that a variety of factors can affect the frequency and/or types of mutations that arise during viral replication, thus potentially altering the evolutionary path of the virus within infected individuals. These factors include experimental and approved antiviral therapeutics (e.g. nucleoside analogs that cause chain termination or altered base-pairing) (11-19), cell type (20), drug resistance-associated mutations in reverse transcriptase (8, 11, 13, 14, 21), viral accessory proteins (Vif and Vpr) (3, 13, 22), and host proteins of the APOBEC3 family (23, 24). The examination of factors influencing the mutagenesis of HIV-1 remains an area of active investigation, necessitating

the development of improved vectors and methodologies for quantifying and characterizing viral mutants.

Several methods have been used to determine mutant frequencies, mutation rates, and mutation spectra in HIV-1, most of which rely on reporter genes that confer a phenotype that can be easily scored, such as *lacZ* α (1, 6), *egfp* (15, 16), and *thymidine kinase* (5). Direct amplification and sequencing of viral genes has also been used to determine HIV-1 mutation rates and spectra (7, 25), but Sanger sequencing of viral genes is relatively inefficient (as most sequences lack mutations) and next-generation sequencing methods are currently too error-prone (error rates of 10^{-2} to 10^{-4} mutations/bp depending on how the data is processed) to be of much utility (26-28). Using reporter genes, both the restoration of defective reporter gene phenotypes (reversion assays) as well as the elimination of functional reporter gene phenotypes (forward mutation assays) can be detected. Forward mutation assays have generally been preferred because they allow for the detection of numerous types of mutations at many different positions within the reporter sequence. Using these assays, forward mutant frequencies can be calculated by determining the ratio of mutant proviruses (i.e. those lacking LacZ α or EGFP expression) to all proviruses. As these assays rely on the detection of mutations that eliminate reporter gene expression, silent mutations will typically not be detected, and thus the calculated mutant frequencies underestimate the intrinsic mutation rates. In the case of *lacZ* α , mutant frequencies can be converted to mutation rates, as saturation mutagenesis has determined the target bases at which substitutions or frameshifts can be detected within the *lacZ* α gene region (29, 30). Currently, other reporter genes have not yet been characterized to the same extent such that this type of conversion is not possible. In addition to mutant frequencies, these mutation assays enable targeted Sanger sequencing of defective reporter genes for the characterization of mutation spectra.

Here, we describe a replication-defective mCherry/EGFP dual-reporter vector that has been used in single-cycle assays for determination of HIV-1

mutant frequencies and mutation spectra (18, 19). Recently, a similar vector based on HIV type-2 (HIV-2) has been engineered, enabling comparative analyses between HIV-1 and HIV-2 (31). The vector does not express Env and must be complemented *in trans* with an envelope gene to produce infectious virus. Within *nef*, a cassette containing *mCherry* and *egfp* has been inserted, separated by an internal ribosome entry site necessary for translation of EGFP. Thus, the final molecular clone has functional open reading frames for all genes except *env* and *nef*. The resulting vector, called pNL4-3 MIG, can be used to determine viral mutant frequencies as follows: First, infectious viral stocks are produced by co-transfecting the vector and the vesicular stomatitis virus envelope glycoprotein (VSVG) into 293T cells (Figure A3-1A). The supernatants are collected, filtered, DNase I-treated, and titered to determine infectivity. Next, viral stocks are used to infect U373-MAGI-CXCR4 cells, which are then analyzed by flow cytometry to measure mCherry and EGFP expression. Mutant frequencies are calculated by dividing the number of cells expressing only mCherry or EGFP by all infected cells. Lastly, the mutation spectra can be analyzed by sorting out single positive cells (mCherry⁺/EGFP⁻ and/or mCherry⁻/EGFP⁺), isolating genomic DNA, performing PCR, and sequencing the mutated reporter gene(s). The entire assay can be performed in approximately two to three weeks, permitting rapid evaluation of the impact of mutations in reverse transcriptase, APOBEC3 proteins, viral mutagens, or alterations in cell physiology on HIV-1 mutant frequencies and mutation spectra.

Materials

Cell Culture, Production of Viral Stocks, Infections

1. Cells: 293T (ATCC; Manassas, VA) and U373-MAGI-CXCR4 (NIH AIDS Reagent Program; Germantown, MD) (32). In addition to U373-MAGI-CXCR4, many other cell lines can successfully be used as target cells in this assay (see Note 1).

2. 293T Medium: Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin.
3. U373-MAGI-CXCR4 Medium (MAGI Medium): DMEM supplemented with 10% FBS, 1% penicillin/streptomycin, 0.2 mg/mL G418, 0.1 mg/mL hygromycin B, and 1.0 µg/mL puromycin.
4. Dulbecco's Phosphate-Buffered Saline (DPBS) without Ca²⁺ or Mg²⁺.
5. FACS Buffer: DPBS, 2% FC3, 10 mM HEPES (pH 7.2), 10 U/mL DNase I, 5 mM MgCl₂.
6. Trypsin-EDTA.
7. Serum-free DMEM.
8. Poly-L-Lysine (10X). Dilute 1:10 in sterile water and store at 4 °C. Stocks can be re-used up to five times.
9. Polyethylenimine (PEI; Polysciences, Inc.; Warrington, PA). Other transfection reagents can also be used (see Note 2). Prepare liquid stocks by dissolving the PEI in sterile water to a final concentration of 1 mg/mL, adjusting the pH to 7.0 with 1 M HCl, and filtering through a 0.2 µm filter. Divide stocks into 1 mL aliquots and freeze at -80 °C. Maintain thawed stocks at 4 °C; stocks should retain full activity for up to three weeks.
10. Plasmids: pNL4-3 MIG (available upon request) and pHCMV-G (a kind gift from J. Burns, University of California, San Diego) or a comparable plasmid, such as pHEF-VSVG (available through the NIH AIDS Reagent Program). HIV-1 Env constructs may be used in place of VSVG (see Note 3).
11. DNase I.
12. Syringes: 10 mL, 20 mL, or 60 mL, depending on the volume of virus to be produced.
13. 0.2 µm filters.
14. Hemocytometer.
15. 96-well polystyrene plates.
16. Polystyrene round-bottom tubes with strainer caps .

DNA Isolation and PCR

1. Dpn I.
2. Genomic DNA purification kit, such as the High Pure PCR Template Preparation Kit (Roche; Basel, Switzerland).
3. High-fidelity amplification system, such as Phusion Hot Start II (Thermo Fisher Scientific, Inc.; Waltham, MA).
4. Taq DNA polymerase.
5. dNTPs.
6. Primers: EGFP-Forward (5'-TCAAGCGTATTCAACAAGG-3') and EGFP-Reverse (5'-CATTGTTAGCTGCTGTATTGC-3').
7. Gel extraction kit.
8. T/A cloning kit, such as the pGEM-T Vector Kit II (Promega; Madison, WI).
9. LB plates with ampicillin/IPTG/X-Gal: LB-agar, 100 µg/mL ampicillin, 0.5 mM IPTG, and 80 µg/mL X-Gal.
10. LB broth with ampicillin: LB-broth, 100 µg/mL ampicillin.
11. Standard DNA miniprep kit or 96-well miniprep kit, depending on desired throughput.
12. Sequencing Primers: T7 promoter (5'-TAATACGACTCACTATAGGG-3') and SP6 upstream (5'-ATTTAGGTGACACTATAG-3'). Sequencing facilities often provide these primers at no additional cost.

Equipment and Software

1. Flow cytometer capable of simultaneously detecting and separating mCherry and EGFP signals (see Note 4), preferably attached to a high-throughput sample (HTS) unit.
2. Access to a flow cytometer capable of fluorescence-activated cell sorting (FACS); this cytometer must also be able to simultaneously detect and separate mCherry and EGFP signals.
3. FlowJo (Tree Star Inc.; Ashland, OR) or similar program for analyzing flow cytometry data.
4. GraphPad Prism (GraphPad Software, Inc.; La Jolla, CA) or similar program for data analysis.

5. Program capable of assembling many Sanger sequences, such as SeqMan Pro (Lasergene Core Suite; DNASTAR, Inc.; Madison, WI) or Vector NTI (Life Technologies; Carlsbad, CA).

Methods

Production of Viral Stocks

1. Plate 4 million 293T cells/10 cm plate ~24 hours before transfection. Optional: Pre-coat plates with poly-L-lysine to minimize cell detachment during the transfection procedure. Pre-coat plates by adding sufficient poly-L-lysine to cover the plates (~5 mL/plate), waiting 5-10 minutes, aspirating or pipetting off the poly-L-lysine, and drying the plates for ~1 hour in a certified biosafety cabinet. Plates should be ~90-95% confluent at the time of transfection.
2. Begin the transfection procedure ~24 hours after plating the cells by preparing the transfection mixture. For each 10 cm plate, mix 10 µg pNL4-3 MIG, 1 µg pHCMV-G, and serum-free DMEM to a final volume of 0.5 mL in a microcentrifuge tube. Mix 33 µL 1 mg/mL PEI (3:1 ratio of µL PEI to µg of DNA) and 467 µL serum-free DMEM in a separate microcentrifuge tube. Add the DNA/DMEM all at once to the PEI/DMEM, mix by pipetting, and incubate 15-20 minutes.
3. Replace the medium on the 293T cells (10 mL/plate).
4. Add the DNA-PEI transfection mixture dropwise to the plates of 293T cells; mix with gentle swirling.
5. Replace the medium again ~12-16 hours post-transfection (7-10 mL/plate).
6. Collect viral stocks ~36-48 hours post-transfection by filtering the supernatants through 0.2 µm filters. Treat viral stocks with 10 U/mL of DNase I (in 1X DNase I buffer) at 37 °C for 3 hours to reduce plasmid contamination (see Note 5). Freeze stocks in 0.5-1.0 mL aliquots at -80 °C until the time of infection. If desired, the transfected cells can be collected for analysis of transfection efficiency by flow cytometry.

Titering Viral Stocks

1. Plate 31,250 U373-MAGI-CXCR4 cells/well in a 24-well plate ~24 hours before infection, using 1 mL MAGI medium/well.
2. At the time of infection, aspirate the medium and replace with variable volumes of the viral stock to be titered and 293T medium, keeping a total volume in each well of 1 mL (see Note 6). Note that 293T medium is used to avoid any effect that the selection compounds in the MAGI medium could have on infectivity. Perform all infections in triplicate and include uninfected controls.
3. Replace the medium 24 hours post-infection.
4. Collect the cells for flow cytometry 72 hours post-infection. Aspirate the medium and add 0.2 mL trypsin-EDTA/well. Transfer the trypsinized cells to a 96-well plate and centrifuge at $500 \times g$ for 5 minutes at 4 °C. Remove the trypsin-EDTA and wash the samples with 0.2 mL DPBS/2% FBS/well before centrifuging the cells for a second time. Remove the DPBS/2% FBS and again add 0.2 mL DPBS/2% FBS/well, wrap in paraffin film, and store at 4 °C until samples are assessed by flow cytometry.
5. Perform flow cytometry to quantify the number of cells expressing mCherry and/or EGFP. Optimize flow cytometer settings before recording data (see Note 7). Analyze a minimum of 10,000 gated cells/sample. Gates should be consistent across all samples. Determine the percent infection in each sample by taking the sum of the following quadrants: mCherry⁺/EGFP⁻, mCherry⁻/EGFP⁺, and mCherry⁺/EGFP⁺. Alternatively, subtract the double negative population (mCherry⁻/EGFP⁻) from 100.
6. In GraphPad Prism, plot the volume of virus (in μL) against the percent infection. The relationship between the volume of virus and the percent infection should be linear up to ~40%, and we recommend excluding data above this value (see note 8). If less than five data points are below 40% infectivity, repeat the titer using less virus. Fit a trend-line to the data using linear regression; the r^2 value should be > 0.99 . The equation of the trend line allows calculation of predicted volumes of viral stocks to achieve targeted infectivities in future experiments. Although not necessary for this protocol, one can also calculate

infectious titers from the data, though these titers likely underestimate true infectious titers (Note 9).

Small-Scale Infections to Determine Viral Mutant Frequencies

1. Plate 31,250 U373-MAGI-CXCR4 cells/well in a 24-well plate ~24 hours before infection, using 1 mL MAGI medium/well.
2. At the time of infection, replace the medium with 293T medium and the amount of virus predicted to result in 20% infection based upon the viral titer (see Note 10). Include uninfected controls. If testing the impact of small molecules on viral mutant frequency, add variable amounts of the compounds two hours before infection (see Note 11).
3. Replace the medium 24 hours post-infection.
4. Collect the cells for analysis by flow cytometry 72 hours post-infection as described before.
5. Perform flow cytometry to quantify the number of cells expressing mCherry and/or EGFP. Determine viral infectivity to ensure that it is close to the targeted 20% infection. Determine mutant frequency by adding the single positive cell populations and dividing by all infected cells: $(mCherry^+/EGFP^- + mCherry^-/EGFP^+) / (100 - mCherry^-/EGFP^-)$. Sample flow data with corresponding infectivities and mutant frequencies are illustrated in Figure A3-1B. HIV-1 wild-type mutant frequencies typically fall within the range of 0.04—0.06 in this assay. Using the length of the entire cassette (mCherry-IRES-EGFP; 2072 bp) as the size of the mutational target (a conservative means of estimating viral mutation rates), these mutant frequencies correspond to mutation rates of $1.9\text{--}2.9 \times 10^{-5}$ mutations/bp/cycle.
6. In GraphPad Prism, graph the infectivity and mutant frequency of the infected control (i.e. untreated wild-type virus) against the other samples (e.g. cells treated with compounds of interest or infected with mutant viruses), as shown in Figure A3-2A. Perform at least three independent experiments and graph the means \pm standard deviations. Although absolute infectivities and mutant frequencies can be graphed, we prefer to normalize the data relative to the

control virus (set to 100% for infectivity or 1 for mutant frequency) to simplify data interpretation.

7. Perform statistical analysis to assess significance of differences in mutant frequencies between samples. One-way ANOVA followed by Tukey's post-test using GraphPad Prism or similar software is usually sufficient to examine differences in mutant frequencies across multiple treatment groups. ANOVA should be performed using absolute (rather than normalized) mutant frequencies. As ANOVA tests rely upon certain assumptions, the data should ideally be transformed prior to one-way ANOVA (see Note 12). Other types of ANOVAs may be ideal for more complex comparisons of treatment groups (see Note 13).

Large-Scale Infections to Determine Viral Mutation Spectra

1. Plate 1.2 million U373-MAGI-CXCR4 cells/10 cm plate ~24 hours before infection, using 10 mL MAGI medium/plate. Prepare two plates/treatment and include uninfected controls.
2. At the time of infection, replace the medium with fresh 293T medium and the amount of virus predicted to result in 20% infection based upon the viral titer (see Note 14). If examining the impact of small molecules on viral mutant frequency, add the compounds two hours before infection.
3. Replace the medium 24 hours after infection.
4. Collect the cells for fluorescence-activated cell sorting (FACS) 72 hours post-infection. Before collecting samples, we advise discussing the procedure with the operator of the FACS instrument, as they may suggest modifications to this protocol. Begin by aspirating the medium and adding 3 mL trypsin-EDTA/plate. Neutralize the trypsin with 7 mL 293T medium/plate and centrifuge in 15 or 50 mL conical tubes for 5 min at $500 \times g$ and 4°C . Aspirate the trypsin-EDTA/medium.
5. Wash the cells twice with 10 mL DPBS/2% FBS/plate. Resuspend the final cell pellet in 1 mL FACS buffer/plate of cells and transfer to 5 mL polystyrene tubes.

6. Bring the following to the FACS instrument: samples for analysis, FACS buffer (~10 mL in a 15 mL conical tube), polystyrene tubes with strainer caps (one/sample), and collection tubes (microcentrifuge tubes with 0.1 mL DPBS/2% FBS/tube; bring two/sample).

7. Subject samples to FACS to sort out the mCherry⁺/EGFP⁻ and mCherry⁻/EGFP⁺ cell populations (see Note 15). Each sample should be filtered into a polystyrene tube through a strainer cap immediately before FACS to remove cell aggregates. The speed and efficiency of sorting can vary widely depending on numerous factors (infectivity, degree of cell death, etc.). We typically collect at least 2000 mCherry⁺/EGFP⁻ cells per sample, though the ideal number depends on the amount of sequencing data required for analysis (see Note 16).

Extraction of genomic DNA, PCR, and Sequencing

1. Purify genomic DNA from infected cells using Roche's High Pure PCR Template Preparation Kit (or a similar kit), following the manufacturer's instructions. Elute in 50 μ L sterile water/sample. Isolate genomic DNA from uninfected cells as a negative control for the PCR reaction.

2. Treat genomic DNA with 0.2 U/ μ L DpnI in 1X DpnI buffer at 37 °C for 2 hours to degrade any residual plasmid DNA.

3. Perform PCR amplification of *egfp* using a high-fidelity amplification system that is based on a proofreading polymerase, such as Phusion Hot Start II. Proofreading polymerases are recommended because they exhibit much lower error rates than Taq polymerase, which lacks proofreading activity (error rates of 10^{-6} - 10^{-7} mutations/bp/cycle or 10^{-4} - 10^{-5} mutations/bp/cycle, respectively). High-fidelity polymerases minimize background errors that could otherwise contribute to the mutation spectra data (see Note 17 for further discussion). Using Phusion Hot Start II, set up the PCR reactions with 28.5 μ L nuclease-free water, 10 μ L 5X HF buffer, 1 μ L dNTPs (10 mM), 2.5 μ L of each primer (EGFP-Forward and EGFP-Reverse primers; 10 μ M stocks), 5 μ L genomic DNA, and 0.5 μ L Phusion Hot Start II. We advise performing the PCR reactions from each sample in triplicate and pooling the resulting products in order to increase the number of

unique mutations that can be observed by sequencing. The *egfp* gene from the pNL4-3 MIG plasmid should be amplified as a positive control reaction. Use water and genomic DNA from uninfected cells as negative control reactions.

4. Perform PCR using the following cycling parameters: 98 °C for 30 seconds, 40 cycles of 98 °C 10 seconds/60 °C 30 seconds/72 °C 30 seconds, and a final extension of 72 °C for 10 min. In some cases, the number of PCR cycles can be reduced to 30 or 35 in order to further decrease PCR-mediated background errors, but in general this is not recommended because: 1) the benefit in terms of reducing background errors is relatively small (see Note 17), and 2) the PCR product may not be sufficiently visible on a gel to allow for DNA purification.

5. Gel-purify the PCR products (981 bp in length) to remove any non-specific products or primer dimers. Gel extraction also removes the proofreading polymerase (i.e. Phusion Hot Start II), which is crucial for efficient A-tailing of PCR products.

6. Add A-tails to the blunt-end PCR products by mixing 6 µL purified PCR product, 1 µL 10X Taq buffer, 2 µL dATP (1 mM), and 1 µL (5U) Taq polymerase. Incubate at 72 °C for 20 minutes.

7. Ligate the A-tailed PCR products into a T/A cloning vector such as pGEM-T following the manufacturer's instructions. Use 3 µL of PCR product in each ligation reaction and ligate overnight at 4 °C for maximum efficiency.

8. Transform 5 µL of each ligation reaction into JM109 cells. Other high-efficiency competent cells ($\geq 10^8$ colony forming units/µg) can be used as well, provided they are compatible with blue-white color screening. Plate several different volumes (e.g. 50, 200, 500 µL) onto LB-amp plates containing IPTG and X-Gal for blue-white color screening.

9. Start 1.5 mL liquid cultures in LB broth with ampicillin from white colonies (i.e. those that contain an insert disrupting the *lacZ* α reading frame in pGEM-T) the following day.

10. Purify plasmid DNA from the liquid cultures the next day. For smaller numbers of cultures (< 48), we recommend using standard miniprep kits. For larger numbers of cultures (≥ 48), use 96-well high-throughput miniprep kits.
11. Submit the samples for Sanger sequencing using primers that flank the ligated insert, such as the T7 promoter and/or SP6 upstream primers (see Note 18).
12. Align resulting sequences to the wild-type *egfp* sequence (this will exclude the primer sequences from analysis, as the primers anneal outside of *egfp*) using a program like SeqMan Pro (part of the Lasergene Core Suite) or Vector NTI. Perform trimming to remove low-quality sequence data.
13. Use the program to identify all putative mutations within the assembled sequences. Verify each mutation by viewing the chromatogram. Compile a list of information about each mutation—sequence identifier, position, and mutation type. Exclude identical mutations that are found in multiple clones from the same sample, as these may have arisen due to sequencing of amplified products from the same original provirus. It is important to note that even if the identical mutations occur in distinct sequences (e.g., with different co-occurring mutations), they may still have arisen from one initial provirus, as recombination can occur during PCR. Occasional G-to-A hypermutants (likely due to the activity of APOBEC3 proteins) may be identified, and the mutation spectra should be analyzed both in the presence and absence of these G-to-A hypermutants (see Note 19).
14. From the list of verified mutations, compile the mutation spectra by determining the absolute numbers and relative percentages that each mutational type contributes to the total. The mutation spectra data can be presented in table or pie chart format (as shown in Figure A3-2B). Mutation spectra can be compiled both for total mutations as well as for specific classes of mutations, such as substitutions. While insertions and deletions can be observed in this assay, we have found that they occur much less frequently than substitutions,

and therefore it is difficult to make meaningful comparisons of their frequencies between samples.

15. Analyze statistical significance of differences in mutation spectra using the Cochran-Mantel-Haenszel test or Fisher's Exact test (see Note 20). These analyses should be performed on the absolute numbers of mutants from each class, rather than relative percentages of the total.

Notes

1. Many other cell lines may successfully be used as targets for infection, particularly when viral stocks are pseudotyped with VSVG. Besides U373-MAGI-CXCR4, we have successfully used immortalized T-cell lines, such as CEM and SupT1. Infectivity of viral stocks can vary widely between cell lines, such that they must be titered in each target cell line.

2. We have also produced viral stocks using standard calcium phosphate protocols and lipid-based reagents (such as GenJet and Lipofectamine 2000), but the PEI method has provided the best balance between cost, efficiency, and speed. If viral titers are insufficient, lipid-based methods are recommended. Alternatively, larger volumes of virus may be produced by PEI transfection and then concentrated.

3. VSVG is usually preferable for pseudotyping the HIV-1 vector as it results in high titer viral stocks that can infect a variety of cell lines. However, when using VSVG, it is possible for the virus to re-infect the producer cells (293T). While this may elevate the absolute mutant frequency, it should not affect relative mutant frequency comparisons between samples. In place of VSVG, constructs expressing CXCR4-tropic HIV-1 Env may be utilized (the NIH AIDS Reagent Program has many available). While the use of these constructs prevents re-infection of producer cells, infectivity of viral stocks will likely be reduced.

4. For our experiments, we have used a BD LSR II flow cytometer (BD Biosciences) equipped with 488, 561, and 640 nm lasers. EGFP is excited by the standard 488 nm laser and detected with 505LP and 525/50 filters, while

mCherry is excited with the 561 nm laser and detected with 595LP and 610/20 filters. With this particular setup, no fluorescence compensation was necessary. We recommend optimizing the flow cytometer setup with single-color controls (such as 293T cells transfected with EGFP or mCherry) before using the pNL4-3 MIG construct. If the mCherry and EGFP signals cannot be satisfactorily separated (even with compensation), a similar molecular clone of HIV-1 (pNL4-3 HIG) that expresses the mouse heat stable antigen (HSA) and EGFP can be used instead (15, 16). In this case, an additional step must be performed in which HSA is stained with a phycoerythrin (PE)-conjugated anti-HSA antibody.

5. DNase I treatment of viral stocks is highly recommended to reduce plasmid contamination of genomic DNA used for PCR and sequencing. Perform this step before freezing and titering viral stocks, as the DNase I treatment itself may somewhat reduce viral titers. If viral stocks will only be used to determine mutant frequencies by flow cytometry, DNase I treatment is unnecessary.

6. For NL4-3 MIG-VSVG, infect with 1.25-80 μ L of virus (2-fold dilution series). For NL4-3 MIG complemented with NL4-3 Env, infect with 31.25-500 μ L of virus. These ranges should comfortably span the target of 20% infection for further experiments. Titers may vary considerably if other target cell lines or envelopes are used.

7. When initially performing flow cytometry of infected cells, include several extra samples (uninfected and infected) for optimization of flow settings (voltages and thresholds). We suggest optimizing the settings in tube mode and then proceeding with the analysis of all remaining samples in HTS mode. After gaining familiarity with the procedure, these extra samples can be omitted.

8. Within this assay, cells infected by a single virus or by multiple viruses cannot reliably be distinguished. However, at low infectivities, the probability of co-infection is small such that the relationship between the volume of virus added and percent infection is highly linear. We have found that linearity is maintained up to ~40% infection.

9. Calculate the titer (infectious units/mL) as follows: $(F \times N \times 1000)/V$, where F is the percentage of fluorescent cells (expressed in decimal form), N is the number of cells at the time of infection, and V is the volume (μL) of virus stock added. Determine the titer for each volume of virus within the linear range of infection and average the resulting values. These titers may underestimate true infectious titers because: 1) some cells may be co-infected, and 2) some cells may be infected but fail to express mCherry or EGFP (due to mutations in both *mCherry* and *egfp*, mutations in *tat*, or integration into a transcriptionally-repressive region of the genome).

10. We highly recommend performing experiments within the range of 15-30% infection (for the untreated control) whenever possible. At lower infectivities, mutant frequencies become noticeably more variable due to the low number of mutants detected. At higher infectivities, changes in mutant frequencies become harder to detect due to high levels of co-infection (in which case wild-type proviruses can mask mutant proviruses).

11. When testing the impact of small molecules on viral mutant frequencies, we generally pre-incubate the cells with the compounds of interest for two hours prior to infection. However, some compounds may require longer activation periods, necessitating optimization of the time of compound addition. At the desired time point, replace the medium and add the compounds of interest (typically 1-2 μL) to a final volume of 0.5 mL/well. Test a range that spans at least six concentrations of each compound. Include equivalent volumes of the compound solvent in the untreated (infected and uninfected) controls. At the time of infection, add virus and medium to bring the final volume to 1 mL/well. Compound concentrations are based on the final volume of 1 mL/well; thus, adding 1 μL of compound results in a 1000-fold dilution of the compound stock.

12. ANOVA tests assume that all treatment groups are normally distributed and that all treatment groups have equal variance. The small fractions obtained using this reporter assay can invalidate these assumptions and therefore invalidate the ANOVA. Variance stabilization can be performed to account for

unequal variance among treatment groups that may arise from small fractions. Arcsin transformation is one method of variance stabilization that does not require specialized software for statistical analysis. To analyze the data by this method, first transform mutant frequencies (arcsin(square root of mutant frequency)) prior to analysis by ANOVA with Tukey's post-test. We have used ANOVA to evaluate both transformed and untransformed data and found little difference in the statistical results, suggesting that the ANOVA is a fairly robust procedure.

13. To compare samples with two factors (e.g. a drug combination) against samples with a single factor (e.g. single drug treatments), first transform the data using the arcsin square root equation (as described in Note 12) and then perform either a one-way or two-way ANOVA. Although both the one-way and two-way ANOVA are valid methods for evaluating the data, the two-way ANOVA is a more powerful method for estimating interactions. To compare multiple treatment groups to an untreated control, perform one-way ANOVA with Dunnett's multiple comparison test.

14. Assay results are typically highly comparable across different plate formats as long as the numbers of cells and volumes of viruses are scaled based on surface area differences. The scaling factor between 10 cm plates and 24-well plates is 39.27, though this may vary depending on manufacturer. Therefore, one can calculate the amount of virus necessary per 10 cm plate by taking the amount needed per well of a 24-well plate and multiplying by 39.27.

15. Although previous analyses have only included sequences from the mCherry⁺/EGFP⁻ population, other populations (typically up to four total) can be sorted without adding any additional time or expense to the experiment. Thus, we suggest isolating the mCherry⁻/EGFP⁺ cell population as well. Genomic DNA from these cells can be used to identify additional unique mutations in *mCherry*, which may be useful in cases where mCherry⁺/EGFP⁻ genomic DNA is exhausted.

16. The number of cells isolated by FACS should be maximized such that the chance of re-sampling identical mutants is low. Re-sampling of sequences can be minimized by increasing the number of cells sorted and/or by decreasing the number of colonies selected. For example, very little, if any, re-sampling would be expected if 100 colonies were sequenced from 2,000 cells. Increasing the number of colonies screened to 200 would increase the probability of re-sampling, but still less than 10% of the sequences would be expected to be identical due to re-sampling.

17. The frequency of background errors in this assay due to PCR can be estimated based on the published fidelities of thermostable DNA polymerases. While error rates vary considerably depending on the specific polymerase, reaction conditions, and template sequence, error rates for Taq generally range from 10^{-4} to 10^{-5} mutations/bp/cycle (33-38) while error rates for proofreading polymerases range from 10^{-6} to 10^{-7} mutations/bp/cycle (36, 37, 39). Phusion Hot Start II exhibits an error rate of 4.4×10^{-7} mutations/bp/cycle according to the manufacturer, while published estimates of its error rate range from 4.2×10^{-7} to 2.6×10^{-6} mutations/bp/cycle (36, 37, 39). Based on these error rates, the size of the mutational target (*egfp*-720 bp), and 40 cycles of amplification, the use of Phusion should result in ~0.02-0.07 errors/sequence. As we typically observe an average of ~1 mutation/sequence, these calculations suggest that PCR-mediated errors would contribute <10% of the resulting data. In contrast, Taq would be expected to result in ~0.3-3.0 errors/sequence, and PCR errors would thus substantially contribute to resulting mutational spectra data. PCR errors can also be reduced by limiting the number of cycles to 30 or 35, but the benefit is marginal (e.g. 30 cycles with Phusion would lower expected errors to 0.01-0.06 errors/sequence), and the resulting products may not be readily visible by gel electrophoresis.

18. Most of the mutations in the PCR products can be identified by single sequencing reactions with primers that anneal to regions of the vector flanking the insert, such as the T7 and SP6 promoters. However, as the PCR product is

981 bp in length, both sequencing reactions are necessary to cover the entire 981 bp region. The overlapping region of these sequences can also be used to ensure that the same mutation is identified in both directions.

19. In this assay, we have observed occasional G-to-A hypermutants, characterized by 2-20 G-to-A mutations in a single sequencing read (18, 20). The high numbers of G-to-A mutations as well as their dinucleotide context (heavily biased toward GA dinucleotides) suggest that these hypermutants result from the activity of APOBEC3 proteins. Although these sequences are relatively rare, they can contain many G-to-A mutations and thus substantially skew the mutation spectra. Therefore, we suggest compiling and comparing mutation spectra with and without the G-to-A hypermutants from each sample.

20. If three independent experiments were performed, analyze the data using the Conchran-Mantel-Haenszel test, as this procedure can assess differences in mutation types and also account for differences among the independent experiments. If the experiment was not performed three independent times, evaluate the data using Fisher's Exact test. While these analyses are useful, they may not be powerful enough to detect small differences between treatment groups. To increase the likelihood of detecting differences in mutation types, it is always preferable to have a hypothesis that focuses on one specific mutation type. For example, studies focused on APOBEC3 proteins may only require detecting differences in the G-to-A mutations rather than identifying differences in all types of mutations.

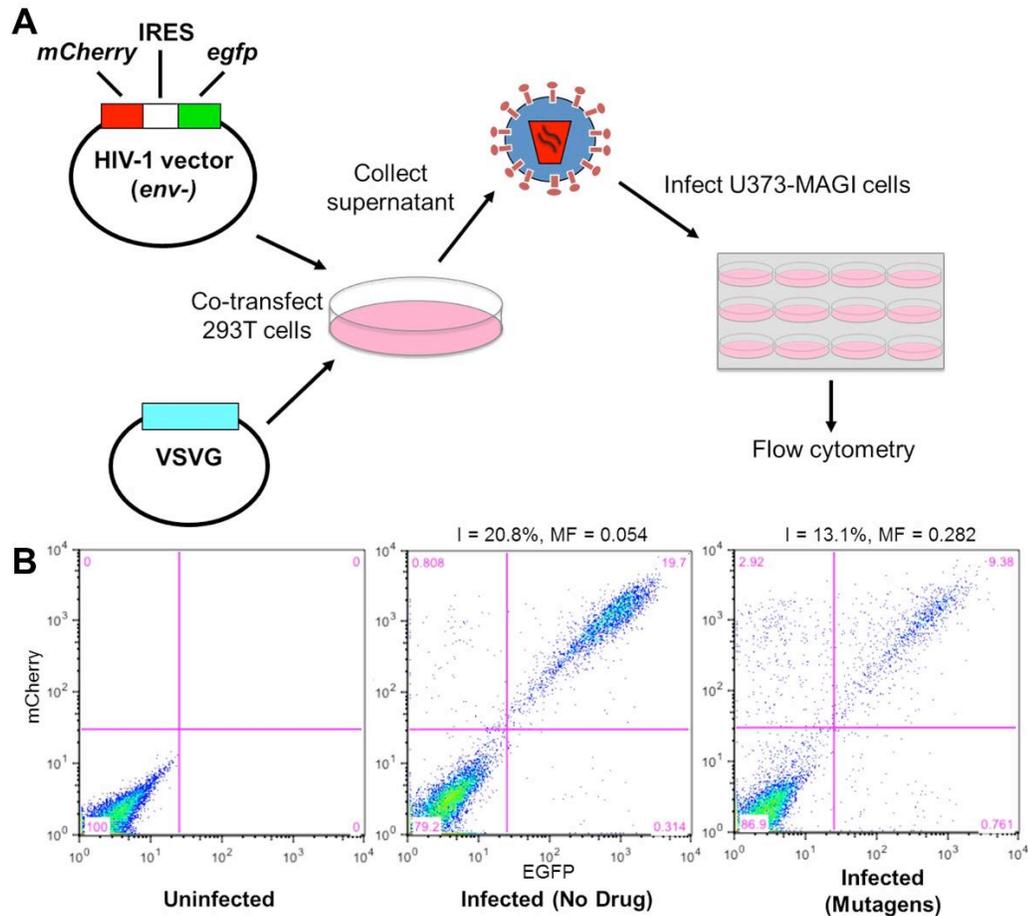


Figure A3-1. Vector and assay design for rapid determination of HIV-1 infectivities and mutant frequencies. (A) An HIV-1 vector expressing mCherry and EGFP (pNL4-3 MIG) was co-transfected with VSVG to produce infectious virus. The supernatant was used to infect U373-MAGI-CXCR4 target cells. Cells were collected 72 hours post-infection for analysis of infectivity and mutant frequency *via* flow cytometry. Reprinted in adapted form from Rawson, J.M. *et al.* 2013 (18) with permission from Elsevier. (B) Examples of representative flow cytometry data (EGFP vs. mCherry) from uninfected cells, infected untreated cells, and infected cells treated with a mutagenic drug combination (50 μ M resveratrol + 200 μ M KP-1212). The infectivities (I) were calculated by adding the percentages of cells in all three positive quadrants, whereas mutant frequencies (MF) were calculated by adding the single-positive quadrants (mCherry⁺/EGFP⁻ + mCherry⁻/EGFP⁺) and dividing by the infectivities.

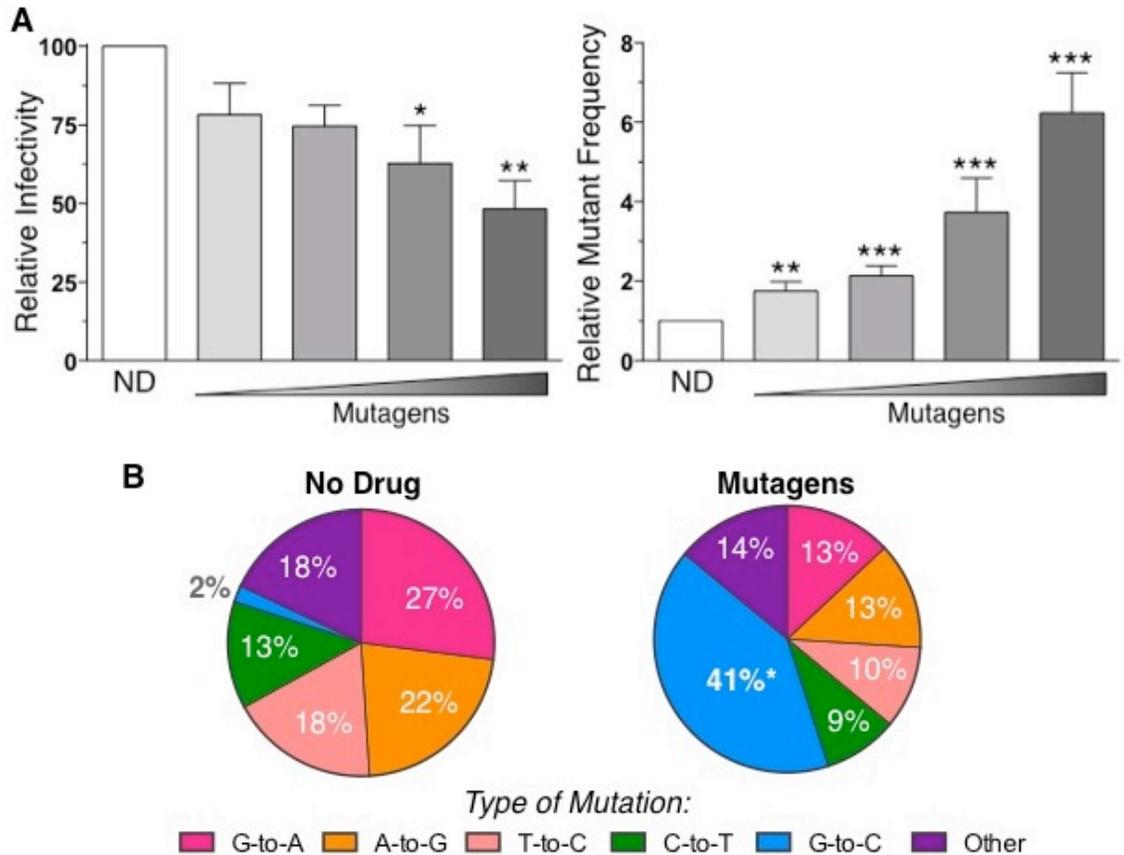


Figure A3-2. Use of the dual-reporter HIV-1 vector to analyze viral infectivities, mutant frequencies, and mutation spectra. (A) The impact of a mutagenic drug combination (50 μ M resveratrol + 0, 5, 50, or 200 μ M KP-1212) on viral infectivities and mutant frequencies was determined. The data were normalized to the no drug (ND) control, which was set to 100 (infectivity) or 1 (mutant frequency). The data represent the means of three independent experiments \pm standard deviations. One-way ANOVA with Dunnett's post-test was performed on the un-normalized data to assess statistical significance. *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001. (B) To examine mutation spectra, mCherry⁺/EGFP⁻ cells were isolated by fluorescence-activated cell sorting (FACS). Genomic DNA was extracted, and *egfp* from the proviral DNA was amplified, cloned, and analyzed by Sanger sequencing. After alignments and identification of mutations, the percentage of mutations belonging to each class was determined for the no drug (ND) control and cells treated with a

mutagenic drug combination (50 μ M resveratrol + 200 μ M KP-1212). The drug combination caused an increase in G-to-C mutations from 2% to 41% of total mutations, which was statistically significant (p -value < 0.0001). Total mutations analyzed: 45 (no drug) and 69 (drug combination). Reprinted in adapted form from Rawson, J.M. *et al.* 2013 (18) with permission from Elsevier.

References

1. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**:5087-5094.
2. **Mansky LM.** 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res Hum Retroviruses* **12**:307-314.
3. **Mansky LM.** 1996. The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**:391-400.
4. **O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD.** 2002. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J Biol Chem* **277**:38053-38061.
5. **Huang KJ, Wooley DP.** 2005. A new cell-based assay for measuring the forward mutation rate of HIV-1. *J Virol Methods* **124**:95-104.
6. **Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH.** 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* **84**:9864-9878.
7. **Schlub TE, Grimm AJ, Smyth RP, Cromer D, Chopra A, Mallal S, Venturi V, Waugh C, Mak J, Davenport MP.** 2014. 15-20% of HIV substitution mutations are associated with recombination. *J Virol* doi:10.1128/JVI.03136-13.
8. **Abram ME, Ferris AL, Das K, Quinones O, Shao W, Tuske S, Alvord WG, Arnold E, Hughes SH.** 2014. Mutations in HIV-1 Reverse Transcriptase Affect the Errors Made in a Single Cycle of Viral Replication. *J Virol* **88**:7589-7601.
9. **Smyth RP, Davenport MP, Mak J.** 2012. The origin of genetic diversity in HIV-1. *Virus Res* **169**:415-429.
10. **Delviks-Frankenberry K, Galli A, Nikolaitchik O, Mens H, Pathak VK, Hu WS.** 2011. Mechanisms and factors that influence high frequency retroviral recombination. *Viruses* **3**:1650-1680.
11. **Mansky LM, Bernard LC.** 2000. 3'-Azido-3'-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J Virol* **74**:9532-9539.
12. **Mansky LM.** 2003. Mutagenic outcome of combined antiviral drug treatment during human immunodeficiency virus type 1 replication. *Virology* **307**:116-121.
13. **Mansky LM, Le Rouzic E, Benichou S, Gajary LC.** 2003. Influence of reverse transcriptase variants, drugs, and Vpr on human immunodeficiency virus type 1 mutant frequencies. *J Virol* **77**:2071-2080.
14. **Chen R, Yokoyama M, Sato H, Reilly C, Mansky LM.** 2005. Human immunodeficiency virus mutagenesis during antiviral therapy: impact of drug-resistant reverse transcriptase and nucleoside and nonnucleoside

- reverse transcriptase inhibitors on human immunodeficiency virus type 1 mutation frequencies. *J Virol* **79**:12045-12057.
15. **Dapp MJ, Clouser CL, Patterson S, Mansky LM.** 2009. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J Virol* **83**:11950-11958.
 16. **Clouser CL, Patterson SE, Mansky LM.** 2010. Exploiting drug repositioning for discovery of a novel HIV combination therapy. *J Virol* **84**:9301-9309.
 17. **Clouser CL, Chauhan J, Bess MA, van Oploo JL, Zhou D, Dimick-Gray S, Mansky LM, Patterson SE.** 2012. Anti-HIV-1 activity of resveratrol derivatives and synergistic inhibition of HIV-1 by the combination of resveratrol and decitabine. *Bioorg Med Chem Lett* **22**:6642-6646.
 18. **Rawson JM, Heineman RH, Beach LB, Martin JL, Schnettler EK, Dapp MJ, Patterson SE, Mansky LM.** 2013. 5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors. *Bioorg Med Chem* **21**:7222-7228.
 19. **Dapp MJ, Bonnac L, Patterson SE, Mansky LM.** 2014. Discovery of novel ribonucleoside analogs with activity against human immunodeficiency virus type 1. *J Virol* **88**:354-363.
 20. **Holtz CM, Mansky LM.** 2013. Variation of HIV-1 mutation spectra among cell types. *J Virol* **87**:5296-5299.
 21. **Dapp MJ, Heineman RH, Mansky LM.** 2013. Interrelationship between HIV-1 fitness and mutation rate. *J Mol Biol* **425**:41-53.
 22. **Chen R, Le Rouzic E, Kearney JA, Mansky LM, Benichou S.** 2004. Vpr-mediated incorporation of UNG2 into HIV-1 particles is required to modulate the virus mutation rate and for replication in macrophages. *J Biol Chem* **279**:28419-28425.
 23. **Refsland EW, Harris RS.** 2013. The APOBEC3 family of retroelement restriction factors. *Curr Top Microbiol Immunol* **371**:1-27.
 24. **Munk C, Jensen BE, Zielonka J, Haussinger D, Kamp C.** 2012. Running loose or getting lost: how HIV-1 counters and capitalizes on APOBEC3-induced mutagenesis through its Vif protein. *Viruses* **4**:3132-3161.
 25. **Gao F, Chen Y, Levy DN, Conway JA, Kepler TB, Hui H.** 2004. Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious. *J Virol* **78**:2426-2433.
 26. **Dohm JC, Lottaz C, Borodina T, Himmelbauer H.** 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**:e105.
 27. **Minoche AE, Dohm JC, Himmelbauer H.** 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**:R112.

28. **Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-UI-Amin M, Ogasawara N, Kanaya S.** 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* **39**:e90.
29. **Bebenek K, Kunkel TA.** 1995. Analyzing fidelity of DNA polymerases. *Methods Enzymol* **262**:217-232.
30. **Boyer JC, Bebenek K, Kunkel TA.** 1996. Analyzing the fidelity of reverse transcription and transcription. *Methods Enzymol* **275**:523-537.
31. **Beach LB, Rawson JM, Kim B, Patterson SE, Mansky LM.** 2014. Novel inhibitors of human immunodeficiency virus type 2 infectivity. *J Gen Virol* doi:10.1099/vir.0.069864-0.
32. **Vodicka MA, Goh WC, Wu LI, Rogel ME, Bartz SR, Schweickart VL, Raport CJ, Emerman M.** 1997. Indicator cell lines for detection of primary strains of human and simian immunodeficiency viruses. *Virology* **233**:193-198.
33. **Cline J, Braman JC, Hogrefe HH.** 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* **24**:3546-3551.
34. **Flaman JM, Frebourg T, Moreau V, Charbonnier F, Martin C, Ishioka C, Friend SH, Iggo R.** 1994. A rapid PCR fidelity assay. *Nucleic Acids Res* **22**:3259-3260.
35. **Keohavong P, Thilly WG.** 1989. Fidelity of DNA polymerases in DNA amplification. *Proc Natl Acad Sci U S A* **86**:9253-9257.
36. **Li M, Diehl F, Dressman D, Vogelstein B, Kinzler KW.** 2006. BEAMing up for detection and quantification of rare sequence variants. *Nat Methods* **3**:95-97.
37. **McInerney P, Adams P, Hadi MZ.** 2014. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Mol Biol Int* **2014**:287430.
38. **Tindall KR, Kunkel TA.** 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27**:6008-6013.
39. **Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B.** 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**:9530-9535.

APPENDIX IV
DECLARATION OF CONTRIBUTIONS TO CO-AUTHORED PUBLICATION:
NOVEL INHIBITORS OF HUMAN IMMUNODEFICIENCY
VIRUS TYPE 2 INFECTIVITY

I am a co-author on the following publication:

Beach LB, **Rawson JM**, Kim B, Patterson SE, Mansky LM. Novel inhibitors of human immunodeficiency virus type 2 infectivity. *J Gen Virol* (2014) Epub 2014 Aug 7. Reprinted with permission; copyright © 2014, the Microbiology Society.

My contributions to this publication are as follows:

- Designed, constructed, and verified the HIV-2 dual-reporter vector (pROD-MIG)
- Assisted in writing the manuscript

Abstract

Human immunodeficiency virus type 2 (HIV-2) infects about 2 million people worldwide. HIV-2 has fewer treatment options than HIV-1, yet may evolve drug resistance more quickly. We have analyzed several novel drugs for anti-HIV-2 activity. It was observed that 5-azacytidine, clofarabine, gemcitabine and resveratrol have potent anti-HIV-2 activity. The EC₅₀ values for 5-azacytidine, clofarabine and resveratrol were found to be significantly lower with HIV-2 compared to that of HIV-1. A time-of-addition assay was used to analyze the ability of these drugs to interfere with HIV-2 replication. Reverse transcription was the likely target for antiretroviral activity. Taken together, several novel drugs have been discovered to have activity against HIV-2. Based upon their known activities, these drugs may elicit enhanced HIV-2 mutagenesis and therefore be useful for inducing HIV-2 lethal mutagenesis. In addition, the data are consistent with HIV-2 reverse transcriptase (RT) being more sensitive than HIV-1 RT to dNTP pool alterations.

Text

Individuals infected with human immunodeficiency virus type 2 (HIV-2) are primarily of west African descent and many of the HIV-2 cases worldwide are attributed to immigrant populations of west Africans living abroad (1-4). The low prevalence of HIV-2 compared to HIV-1 is attributed to its low infectivity. HIV-2 is considered a naturally attenuated infection and HIV-2-infected individuals are more likely to have lower viral RNA levels and are less likely to progress to AIDS compared to those infected with HIV-1 (2, 4-6). However, when HIV-2 infection induces AIDS, there are fewer treatment options than those infected with HIV-1, because not all anti-HIV-1 drugs inhibit replication of HIV-2. First generation NNRTIs, the fusion inhibitor enfuvirtide, and several protease inhibitors are known to be ineffective against HIV-2, while the clinical efficacy of the entry inhibitor maraviroc is unknown (7). Furthermore, HIV-2 has been reported to have lower genetic barriers to the evolution of multidrug resistance than HIV-1, further narrowing the already-limited HIV-2 drug treatment options (7-12).

Previous clinical trials to treat HIV-1 infection using drugs or drug candidates that enhance HIV-1 mutagenesis were associated with setbacks. For example, hydroxyurea (an inhibitor of cellular ribonucleotide reductase that has also been shown to enhance HIV-1 mutagenesis) has been tested clinically, but was found to have significant side effects (13-15). KP-1212 acts as a viral mutagen by being incorporated into viral DNA during reverse transcription and causing mispairing via tautomerization (16). However, a prodrug version (KP-1461) was not able to significantly reduce viral loads in clinical studies – though an altered spectrum of mutations was observed (17).

Several significant advancements have been made recently to raise renewed enthusiasm to therapeutic approaches that seek to reduce HIV-1 infectivity by enhancing the viral mutation rate via lethal mutagenesis in cell culture (18). First, 5-azacytidine was shown to induce HIV-1 lethal mutagenesis by the specific induction of G-to-C transversion mutations (19). The combination of 5-aza-2'-deoxycytidine (decitabine) and gemcitabine was found to synergistically reduce viral infectivity by enhanced viral mutagenesis (20). Decitabine and gemcitabine, both alone and in combination, were found to reduce viral loads in an AIDS mouse model in the absence of toxicity; stable prodrug derivatives have been identified to aid in clinical translation of these drugs for treatment of HIV-1 infection (21, 22). Resveratrol, a phytoalexin, has been shown to enhance viral mutation (likely via inhibition of ribonucleotide reductase), and has been shown to potentiate the activity of KP-1212 – which could enhance the likelihood of the clinical utility of KP-1212 in the treatment of HIV-1 infection (21, 23). In this study, we examined the ability of the ribonucleoside mutagen 5-azacytidine and the ribonucleotide reductase inhibitors gemcitabine, resveratrol, and clofarabine to reduce HIV-2 infectivity, as compared to HIV-1. A time-of-addition drug assay was utilized in order to identify the step(s) in the HIV-2 life cycle in which viral replication was perturbed.

An HIV-2 vector, pROD-MIG, was used in the drug susceptibility assays. This HIV-2 vector is an envelope-minus vector in which *env* gene sequences

were deleted and a gene cassette composed of the *mCherry* gene, an internal ribosomal entry site (IRES), and the *green fluorescence protein (gfp)* gene were cloned into the remaining *env* gene sequence. As a control, a HIV-1 vector, pNL43-MIG (*env*-minus HIV-1 vector with a mCherry-IRES-GFP expression cassette), was used in parallel to assess HIV-1 infectivity (23). The molecular clone pROD-MIG was created by exchanging the *mCherry* gene from pHIV-1 MIG with the mouse *heat stable antigen* from pROD10-HIG (Rawson *et al.*, submitted). Briefly, a fragment of pHIV-2 HIG spanning *env*, *hsa*, and IRES was amplified and ligated into pGEM-T (Promega Corp; Madison, WI). The resulting construct and pHIV-1 MIG were digested with BamHI and AvrII (New England Biolabs; Ipswich, MA). The appropriate fragments were purified and ligated together using T4 DNA ligase (New England Biolabs), thus exchanging *hsa* for *mCherry* within the pGEM-T subclone. The pGEM-T subclone and pHIV-2 HIG were then digested with PmlI, cleaving within *env* and IRES. The vector was treated with Antarctic phosphatase (New England Biolabs), and the appropriate fragments were ligated together using T4 DNA ligase. Like pNL43-MIG, pROD-MIG expresses mCherry and EGFP and all viral proteins except Env and Nef.

When treating cells with drugs (Figure A4-1), VSV-G pseudotyped HIV-1 and HIV-2 vector virus stocks were used to infect 10,000 Magi-U373-CXCR4_{CEM} cells/well in a 96-well plate format that had been pre-treated for 2 hours with increasing concentrations of drugs or with vehicle, DMSO. Flow cytometry was used to quantify infectivity and EC₅₀ values were calculated in GraphPad Prism6 (San Diego, California, USA). Table **A4-A4-1** shows the EC₅₀ values calculated for each drug under study along with several drugs that have known mechanisms of action (i.e., raltegravir, tenofovir, zidovudine, nevirapine). Dose response curves and EC₅₀ values were obtained for each independent experiment using nonlinear regression models that generated non-ambiguous EC₅₀ values and that gave acceptable fits in a combined replicates test. These replicate EC₅₀ values were subjected to an unpaired two-tailed t-test to generate a compiled EC₅₀ value for each drug and to compare whether the EC₅₀ values differed for HIV-1 and

HIV-2. All drugs tested, except nevirapine, possessed potent anti-HIV-1 and HIV-2 activity. Consistent with previous publications, nevirapine inhibited HIV-1, but not HIV-2 replication (24). Resveratrol, clofarabine, and 5-azacytidine showed statistically different EC_{50} values, as indicated by a p value of 0.05 or less. In contrast, zidovudine, tenofovir and raltegravir did not show any significant differences in their EC_{50} values between HIV-1 and HIV-2, in agreement with previously published reports (10, 25, 26). This is, to our knowledge, the first report of clofarabine, a known inhibitor of ribonucleotide reductase, having activity against both HIV-1 and HIV-2.

Previous studies have described the cytotoxicity of 5-azacytidine (19), resveratrol, gemcitabine (20, 23, 27), and clofarabine (Beach *et al.*, unpublished data) using the same cell line as reported here. Based on these previous reports, the observed antiviral activity cannot be attributed to cytotoxic effects of these drugs. In support of this, microscopic inspection of the cells as well as forward and side scatter by flow cytometric analysis did not reveal any evidence of increased cell death or abnormalities with any of the drug treatments used. The finding that nevirapine inhibits HIV-1, but not HIV-2, suggests that the antiviral activity is specific and not due to cell cytotoxicity.

The finding of similar antiviral efficacy for zidovudine, tenofovir, and raltegravir, but not for nevirapine, correlates well with previously reported findings that compared the anti-HIV-1 and anti-HIV-2 activities of these drugs (10, 28, 29). Though some of the confidence intervals were relatively large, Table A4-1 indicates that HIV-2 is more susceptible to gemcitabine, clofarabine, and resveratrol than HIV-1 by approximately 2-fold, 3.5-fold, or 4.2-fold, respectively. Since all 3 drugs are known to be inhibitors of ribonucleotide reductase, these observations provide one line of evidence for HIV-2 being more sensitive to ribonucleotide reductase inhibitors than HIV-1 in cell culture. Intriguingly, only resveratrol and clofarabine showed a statistically significant decrease in EC_{50} values, but all showed a general trend of inhibiting HIV-2 with greater potency than with HIV-1. It was also observed that HIV-2 was approximately 1.4-fold

more sensitive than HIV-1 to 5-azacytidine. Taken together, these data also suggest HIV-2 may be more sensitive than HIV-1 to viral mutagens that can induce lethal mutagenesis.

Time-of-addition drug assays were next done in order to identify the phase in the HIV-2 life cycle perturbed by each drug (Figure A4-2). Infections of permissive target cells were done as described above, except that rather than pre-treating with drugs for 2 h before infection, drugs were added to cell culture media immediately at the time of infection, or at varying time points post infection. Cell culture media was changed 24 h after each drug treatment for each time point. Cells were incubated at drug concentrations that could extinguish viral infectivity. The infectivity observed at the 0 h time point for each drug was subtracted and infectivity was subsequently normalized to 24 h no drug infectivity.

The last time point where a drug was observed to consistently extinguish viral replication was interpreted to correlate with the step of the viral life cycle that the drug likely interferes with (30). Drugs that interfered with replication at or before 1 h were defined as targeting entry or fusion; at 3 to 4 h were defined as interfering with reverse transcription; and after 6 h were defined as interfering with integration (30). Similar to previous publications, we qualitatively defined the time-of-drug inhibition in our assays as the time point after which infectivity was consistently detected above baseline (30). For an inhibitor of ribonucleotide reductase, it was hypothesized that depending on when dNTP pool imbalances occur, the inhibition of viral infectivity could occur after fusion but prior to completion of reverse transcription.

At the 6 h timepoint, gemcitabine, resveratrol and clofarabine were found to reduce HIV-2 replication, which was interpreted as likely influencing the reverse transcription phase of the virus life cycle (Figure A4-2). It was observed that gemcitabine and resveratrol typically lost their antiviral efficacy at earlier time points than that of clofarabine, but within a similar time frame as the loss of inhibition of tenofovir and zidovudine. This suggests that 2 to 4 h was the

window of time where an antiviral that affects reverse transcription and would extinguish viral replication. The size of this window for interfering with viral replication (i.e., 2h) could be a reflection of the different timing of cellular uptake and phosphorylation of nucleoside analogs in the Magi-U373-CXCR4_{CEM} cell line. This could also be true for the nucleoside analogs analyzed in this study (i.e., 5-azacytidine, gemcitabine, clofarabine). The timing of RT perturbation was distinct from the inhibition of integrase in our assay. The integrase inhibitor raltegravir inhibited HIV-2 through 8 h (Figure A4-2), the penultimate time point tested. Raltegravir was the only drug tested that inhibited HIV-2 replication beyond 6 h.

Several lines of evidence suggest that HIV-2 RT may be more sensitive than HIV-1 RT to alterations in dNTP pools. First, HIV-2 RT is less processive than HIV-1 RT (12, 31, 32), and this effect is enhanced under reduced dNTP pools (32). Thus, dNTP pool depletion by ribonucleotide reductase inhibitors could more greatly affect HIV-2 reverse transcription than that of HIV-1. Second, HIV-2 encodes the Vpx protein (33), which degrades the cellular triphosphorhydrolase SAMHD1 (34). SAMHD1 degradation results in an increase in dNTP pool concentrations, which allows HIV-2 to replicate in cells with low dNTP concentrations such as macrophages (35-38). Third, HIV-1 (which does not encode Vpx) can readily replicate in macrophages without counteracting SAMHD1 (39, 40). This argues for an important role of Vpx in maintaining dNTP pool levels for efficient HIV-2 DNA synthesis. Fourth, fidelity differences may exist between HIV-1 RT and HIV-2 RT that could influence the sensitivity to dNTP pool alterations. Taken together, these observations suggest that HIV-2 may be more sensitive to dNTP pool alterations. Therefore, drugs that perturb nucleotide pools could have greater potential for treating HIV-2 rather than HIV-1 infection.

Drug combinations of nucleoside reverse transcriptase inhibitors and ribonucleotide reductase inhibitors may be able to slow the emergence of HIV-2 drug resistance since, unlike HIV-1, HIV-2 RT cannot excise nucleoside reverse

transcriptase inhibitors and instead relies only on the exclusion of dNTP analogs to develop antiviral drug resistance (32, 41, 42). Despite a more limited drug repertoire for the treatment of HIV-2 infection due to natural occurring resistance polymorphisms, and the potential promise of ribonucleotide reductase inhibitors to selectively inhibit HIV-2 replication, no studies have examined the efficacy of ribonucleotide reductase inhibitors (or other classes of viral mutagens) against HIV-2 (7-12).

The observations made in this study indicate the antiviral activities of 5-azacytidine, gemcitabine, resveratrol and clofarabine are selective for the reverse transcription phase of HIV-2 replication. These findings also indicate 5-azacytidine, resveratrol, and clofarabine have greater antiviral activity against HIV-2 than HIV-1. Nevirapine, as expected, had antiviral activity against HIV-1, but not HIV-2. No statistical difference could be detected in the potency of gemcitabine, zidovudine, tenofovir or raltegravir. Recent reports have demonstrated that HIV-2, but not HIV-1, must degrade the dNTP pool regulator SAMHD1 for successful infection of macrophages and other myeloid cells, which are known to have extremely low dNTP pools (38, 39). The increased dNTP pools that arise through Vpx-induced SAMHD1 degradation has been hypothesized to be responsible for the rapid emergence of nucleoside reverse transcriptase inhibitor drug resistance in HIV-2 infected individuals (43). It is formally possible that treatment of HIV-2 infection with ribonucleotide reductase inhibitors could decrease dNTP pool levels below the K_m of HIV-2 RT in macrophages, resting CD4 T cells, and dendritic cells, which would disrupt the progression of HIV-related pathogenesis and dampen the emergence of drug resistant virus.

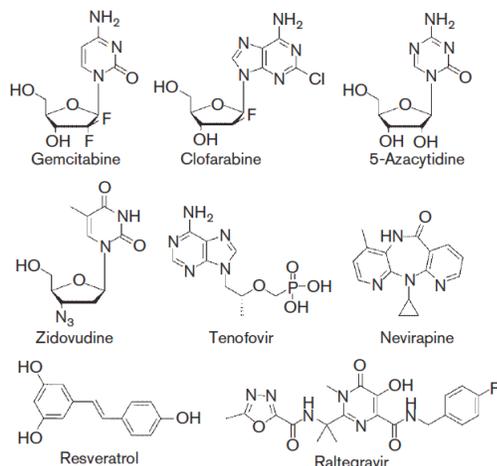


Figure A4-1. Drugs investigated for activity against HIV-2. Drug structures are indicated.

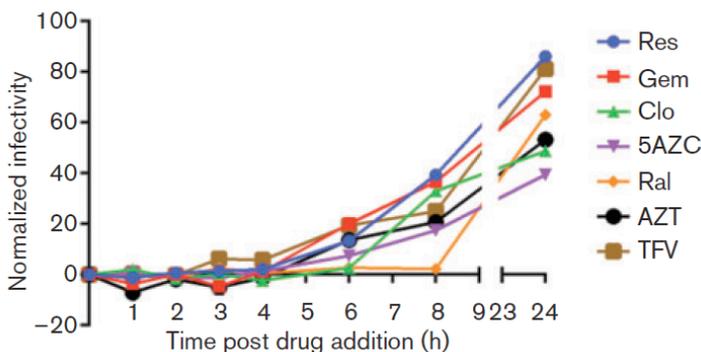


Figure A4-2. Time-of-addition drug assay. Magi cells were seeded into 96 well plates and were then infected with VSV-G pseudotyped HIV-2 vector virus. At 0, 1, 2, 3, 4, 6, 8 and 24 h post infection (penultimate time point tested was at 8 h), cells were treated with no drug (DMSO only) or with the IC_{100} values of clofarabine (Clo), zidovudine (AZT), tenofovir (TFV), resveratrol (Res), raltegravir (Ral), 5-azacytidine (5AZC), or gemcitabine (Gem). Infections were performed in triplicate, background infectivity at the 0 h time point was subtracted, and infectivity for drug treatments was normalized to no drug (DMSO only) infectivity at the 24 h time point (which was 55%). Values represent the averages of three independent experiments, with each treatment at each time point treated in duplicate. Nevirapine was not analyzed due to the lack of antiviral activity against HIV-2.

Table A4-1. EC₅₀ values of drugs under study for HIV-1 and HIV-2 infection.

Drug	HIV-1 EC ₅₀ (μM) [95% confidence interval]*	HIV-2 EC ₅₀ (μM) [95% confidence interval]*	Significant difference in EC ₅₀ values (P<0.05)†	EC ₅₀ ratio, HIV-1/ HIV-2
Gemcitabine‡	1.3 × 10 ⁻³ [2.28 × 10 ⁻⁴ –2.81 × 10 ⁻³]	6.36 × 10 ⁻⁴ [3.95 × 10 ⁻⁴ –1.67 × 10 ⁻³]	N	2.04
Clofarabine§	1.12 × 10 ⁻¹ [3.06 × 10 ⁻² –1.94 × 10 ⁻¹]	3.21 × 10 ⁻² [9.52 × 10 ⁻⁴ –6.32 × 10 ⁻²]	Y	3.78
Resveratrol‡	25.6 [20.1–31.0]	6.08 [4.49–7.68]	Y	4.21
5-Azacytidine§	20.2 [18.7–21.8]	14.5 [10.3–18.8]	Y	1.39
Raltegravir	1.72 × 10 ⁻³ [8.47 × 10 ⁻⁴ –4.30 × 10 ⁻³]	1.85 × 10 ⁻³ [4.43 × 10 ⁻³ –4.15 × 10 ⁻²]	N	0.93
Tenofovir	1.9 × 10 ⁻¹ [49.8 × 10 ⁻² –3.31 × 10 ⁻¹]	1.30 × 10 ⁻¹ [1.40 × 10 ⁻² –2.73 × 10 ⁻¹]	N	1.46
Zidovudine‡	3.67 × 10 ⁻² [2.72 × 10 ⁻² –4.62 × 10 ⁻²]	4.46 × 10 ⁻² [2.73 × 10 ⁻² –6.19 × 10 ⁻²]	N	0.823
Nevirapine‡	6.2 [5.31–7.12]	No activity	–	–

*Curve fitting was performed using log(inhibitor) vs normalized response. If the goodness of fit was not appropriate as determined by the replicates test and residuals, the curves were fitted using either log(inhibitor) vs response or log(inhibitor) vs response (variable slope) as indicated.

†EC₅₀ values for each replicate were used to perform an unpaired two-tailed *t*-test to determine differences between HIV-1 and HIV-2. Experiments were performed in parallel with HIV-1 and HIV-2 for each drug at least four times.

‡log(inhibitor) versus normalized response was used to determine EC₅₀.

§log response (inhibitor) versus response (variable slope) was used to determine EC₅₀.

||log(inhibitor) versus response (three parameters) was used to determine EC₅₀.

References

1. **van der Ende ME, Kroes AC, Buitenwerf J, van der Poel CL.** 1990. [Aids caused by HIV-2 in The Netherlands]. *Ned Tijdschr Geneeskd* **134**:495-497.
2. **Campbell-Yesufu OT, Gandhi RT.** 2011. Update on human immunodeficiency virus (HIV)-2 infection. *Clin Infect Dis* **52**:780-787.
3. **Rey MA, Krust B, Laurent AG, Guetard D, Montagnier L, Hovanessian AG.** 1989. Characterization of an HIV-2-related virus with a smaller sized extracellular envelope glycoprotein. *Virology* **173**:258-267.
4. **Costarelli S, Torti C, Rodella A, Baldanti F, Paolucci S, Lapadula G, Manca N, Quiros-Roldan E, Izzo I, Carosi G.** 2008. Screening and Management of HIV-2-Infected Individuals in Northern Italy. *AIDS Patient Care STDS* **22**:489-494.
5. **Soares RS, Tendeiro R, Foxall RB, Baptista AP, Cavaleiro R, Gomes P, Camacho R, Valadas E, Doroana M, Lucas M, Antunes F, Victorino RM, Sousa AE.** 2011. Cell-associated viral burden provides evidence of ongoing viral replication in aviremic HIV-2-infected patients. *J Virol* **85**:2429-2438.
6. **Bouree P, Lamour P, Bisaro F, Didier E.** 1995. [Study of an HIV positive, tropical origin population in a refugee center in France]. *Bull Soc Pathol Exot* **88**:24-28.
7. **Menendez-Arias L, Alvarez M.** 2014. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antiviral Res* **102**:70-86.
8. **Rodes B, Holguin A, Soriano V, Dourana M, Mansinho K, Antunes F, Gonzalez-Lahoz J.** 2000. Emergence of drug resistance mutations in human immunodeficiency virus type 2-infected subjects undergoing antiretroviral therapy. *J Clin Microbiol* **38**:1370-1374.
9. **Gottlieb GS, Badiane NM, Hawes SE, Fortes L, Toure M, Ndour CT, Starling AK, Traore F, Sall F, Wong KG, Cherne SL, Anderson DJ, Dye SA, Smith RA, Mullins JI, Kiviat NB, Sow PS.** 2009. Emergence of multiclass drug-resistance in HIV-2 in antiretroviral-treated individuals in Senegal: implications for HIV-2 treatment in resource-limited West Africa. *Clin Infect Dis* **48**:476-483.
10. **Witvrouw M, Pannecouque C, Switzer WM, Folks TM, De Clercq E, Heneine W.** 2004. Susceptibility of HIV-2, SIV and SHIV to various anti-HIV-1 compounds: implications for treatment and postexposure prophylaxis. *Antivir Ther* **9**:57-65.
11. **Smith RA, Anderson DJ, Pyrak CL, Preston BD, Gottlieb GS.** 2009. Antiretroviral drug resistance in HIV-2: three amino acid changes are sufficient for classwide nucleoside analogue resistance. *J Infect Dis* **199**:1323-1326.
12. **MacNeil A, Sarr AD, Sankale JL, Meloni ST, Mboup S, Kanki P.** 2007. Direct evidence of lower viral replication rates in vivo in human

- immunodeficiency virus type 2 (HIV-2) infection than in HIV-1 infection. *J Virol* **81**:5325-5330.
13. **Lori F, Malykh AG, Foli A, Maserati R, De Antoni A, Minoli L, Padrini D, Degli Antoni A, Barchi E, Jessen H, Wainberg MA, Gallo RC, Lisziewicz J.** 1997. Combination of a drug targeting the cell with a drug targeting the virus controls human immunodeficiency virus type 1 resistance. *AIDS Res Hum Retroviruses* **13**:1403-1409.
 14. **Biron F, Lucht F, Peyramond D, Fresard A, Vallet T, Nugier F, Grange J, Malley S, Hamedi-Sangsari F, Vila J.** 1995. Anti-HIV activity of the combination of didanosine and hydroxyurea in HIV-1-infected individuals. *J Acquir Immune Defic Syndr Hum Retrovirol* **10**:36-40.
 15. **Frank I, Bosch RJ, Fiscus S, Valentine F, Flexner C, Segal Y, Ruan P, Gulick R, Wood K, Estep S, Fox L, Nevin T, Stevens M, Eron JJ, Jr., Team AP.** 2004. Activity, safety, and immunological effects of hydroxyurea added to didanosine in antiretroviral-naïve and experienced HIV type 1-infected subjects: a randomized, placebo-controlled trial, ACTG 307. *AIDS Res Hum Retroviruses* **20**:916-926.
 16. **Harris KS, Brabant W, Styrchak S, Gall A, Daifuku R.** 2005. KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. *Antiviral Res* **67**:1-9.
 17. **Mullins JI, Heath L, Hughes JP, Kicha J, Styrchak S, Wong KG, Rao U, Hansen A, Harris KS, Laurent JP, Li D, Simpson JH, Essigmann JM, Loeb LA, Parkins J.** 2011. Mutation of HIV-1 genomes in a clinical population treated with the mutagenic nucleoside KP1461. *PLoS One* **6**:e15135.
 18. **Dapp MJ, Patterson SE, Mansky LM.** 2013. Back to the future: revisiting HIV-1 lethal mutagenesis. *Trends Microbiol* **21**:56-62.
 19. **Dapp MJ, Clouser CL, Patterson S, Mansky LM.** 2009. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J Virol* **83**:11950-11958.
 20. **Clouser CL, Patterson SE, Mansky LM.** 2010. Exploiting drug repositioning for discovery of a novel HIV combination therapy. *J Virol* **84**:9301-9309.
 21. **Clouser CL, Holtz CM, Mullett M, Crankshaw DL, Briggs JE, O'Sullivan MG, Patterson SE, Mansky LM.** 2012. Activity of a novel combined antiretroviral therapy of gemcitabine and decitabine in a mouse model for HIV-1. *Antimicrob Agents Chemother* **56**:1942-1948.
 22. **Clouser CL, Holtz CM, Mullett M, Crankshaw DL, Briggs JE, Chauhan J, VanHoutan IM, Patterson SE, Mansky LM.** 2011. Analysis of the ex vivo and in vivo antiretroviral activity of gemcitabine. *PLoS One* **6**:e15840.
 23. **Rawson JM, Heineman RH, Beach LB, Martin JL, Schnettler EK, Dapp MJ, Patterson SE, Mansky LM.** 2013. 5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors. *Bioorg Med Chem* **21**:7222-7228.

24. **Balzarini J.** 2004. Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Curr Top Med Chem* **4**:921-944.
25. **Smith RA, Gottlieb GS, Anderson DJ, Pyrak CL, Preston BD.** 2008. Human immunodeficiency virus types 1 and 2 exhibit comparable sensitivities to Zidovudine and other nucleoside analog inhibitors in vitro. *Antimicrob Agents Chemother* **52**:329-332.
26. **Smith RA, Raugi DN, Kiviat NB, Hawes SE, Mullins JI, Sow PS, Gottlieb GS.** 2011. Phenotypic susceptibility of HIV-2 to raltegravir: integrase mutations Q148R and N155H confer raltegravir resistance. *AIDS* **25**:2235-2241.
27. **Clouser CL, Chauhan J, Bess MA, van Oploo JL, Zhou D, Dimick-Gray S, Mansky LM, Patterson SE.** 2012. Anti-HIV-1 activity of resveratrol derivatives and synergistic inhibition of HIV-1 by the combination of resveratrol and decitabine. *Bioorg Med Chem Lett* **22**:6642-6646.
28. **Roquebert B, Damond F, Collin G, Matheron S, Peytavin G, Benard A, Campa P, Chene G, Brun-Vezinet F, Descamps D.** 2008. HIV-2 integrase gene polymorphism and phenotypic susceptibility of HIV-2 clinical isolates to the integrase inhibitors raltegravir and elvitegravir in vitro. *J Antimicrob Chemother* **62**:914-920.
29. **Andreatta K, Miller MD, White KL.** 2013. HIV-2 antiviral potency and selection of drug resistance mutations by the integrase strand transfer inhibitor elvitegravir and NRTIs emtricitabine and tenofovir in vitro. *J Acquir Immune Defic Syndr* **62**:367-374.
30. **Daelemans D, Pauwels R, De Clercq E, Pannecouque C.** 2011. A time-of-drug addition approach to target identification of antiviral compounds. *Nat Protoc* **6**:925-933.
31. **Post K, Guo J, Howard KJ, Powell MD, Miller JT, Hizi A, Le Grice SF, Levin JG.** 2003. Human immunodeficiency virus type 2 reverse transcriptase activity in model systems that mimic steps in reverse transcription. *J Virol* **77**:7623-7634.
32. **Boyer PL, Clark PK, Hughes SH.** 2012. HIV-1 and HIV-2 reverse transcriptases: different mechanisms of resistance to nucleoside reverse transcriptase inhibitors. *J Virol* **86**:5885-5894.
33. **Clavel F, Guyader M, Guetard D, Salle M, Montagnier L, Alizon M.** 1986. Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* **324**:691-695.
34. **Ahn J, Hao C, Yan J, DeLucia M, Mehrens J, Wang C, Gronenborn AM, Skowronski J.** 2012. HIV/simian immunodeficiency virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex CRL4DCAF1. *J Biol Chem* **287**:12550-12558.

35. **Lahouassa H, Daddacha W, Hofmann H, Ayinde D, Logue EC, Dragin L, Bloch N, Maudet C, Bertrand M, Gramberg T, Pancino G, Priet S, Canard B, Laguette N, Benkirane M, Transy C, Landau NR, Kim B, Margottin-Goguet F.** 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* **13**:223-228.
36. **Baldauf HM, Pan X, Erikson E, Schmidt S, Daddacha W, Burggraf M, Schenkova K, Ambiel I, Wabnitz G, Gramberg T, Panitz S, Flory E, Landau NR, Sertel S, Rutsch F, Lasitschka F, Kim B, Konig R, Fackler OT, Keppler OT.** 2012. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat Med* **18**:1682-1687.
37. **St Gelais C, de Silva S, Amie SM, Coleman CM, Hoy H, Hollenbaugh JA, Kim B, Wu L.** 2012. SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons. *Retrovirology* **9**:105.
38. **Kim B, Nguyen LA, Daddacha W, Hollenbaugh JA.** 2012. Tight interplay among SAMHD1 protein level, cellular dNTP levels, and HIV-1 proviral DNA synthesis kinetics in human primary monocyte-derived macrophages. *J Biol Chem* **287**:21570-21574.
39. **Diamond TL, Roshal M, Jamburuthugoda VK, Reynolds HM, Merriam AR, Lee KY, Balakrishnan M, Bambara RA, Planelles V, Dewhurst S, Kim B.** 2004. Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* **279**:51545-51553.
40. **Nguyen LA, Kim DH, Daly MB, Allan KC, Kim B.** 2014. Host SAMHD1 Protein Promotes HIV-1 Recombination in Macrophages. *J Biol Chem* **289**:2489-2496.
41. **Boyer PL, Sarafianos SG, Clark PK, Arnold E, Hughes SH.** 2006. Why do HIV-1 and HIV-2 use different pathways to develop AZT resistance? *PLoS Pathog* **2**:e10.
42. **Ntemgwa ML, d'Aquin Toni T, Brenner BG, Camacho RJ, Wainberg MA.** 2009. Antiretroviral drug resistance in human immunodeficiency virus type 2. *Antimicrob Agents Chemother* **53**:3611-3619.
43. **Amie SM, Daly MB, Noble E, Schinazi RF, Bambara RA, Kim B.** 2013. Anti-HIV host factor SAMHD1 regulates viral sensitivity to nucleoside reverse transcriptase inhibitors via modulation of cellular deoxyribonucleoside triphosphate (dNTP) levels. *J Biol Chem* **288**:20683-20691.

APPENDIX V

DECLARATION OF CONTRIBUTIONS TO CO-AUTHORED PUBLICATION: CHARACTERIZATION OF CYTOPLASMIC GAG-GAG INTERACTIONS BY DUAL-COLOR Z-SCAN FLUORESCENCE FLUCTUATION SPECTROSCOPY

I am a co-author on the following publication:

Fogarty KH, Chen Y, Grigsby IF, Macdonald PJ, Smith EM, Johnson JL, **Rawson JM**, Mansky LM, Mueller JD. Characterization of cytoplasmic Gag-Gag interactions by dual-color z-scan fluorescence fluctuation spectroscopy. *Biophys J* (2011) 100/6, 1587-95. Reprinted with permission; copyright © 2011, Elsevier Ltd.

My contributions to this publication are as follows:

- Designed, constructed, and verified the HTLV-1 codon-optimized Gag-EYFP reporter construct
- Designed, constructed, and verified HIV-1 Gag-EYFP mutants

Abstract

Fluorescence fluctuation spectroscopy (FFS) quantifies the interactions of fluorescently-labeled proteins inside living cells by brightness analysis. However, the study of cytoplasmic proteins that interact with the plasma membrane is challenging with FFS. If the cytoplasmic section is thinner than the axial size of the observation volume, cytoplasmic and membrane-bound proteins are coexcited, which leads to brightness artifacts. This brightness bias, if not recognized, leads to erroneous interpretation of the data. We have overcome this challenge by introducing dual-color z-scan FFS and the addition of a distinctly colored reference protein. Here, we apply this technique to study the cytoplasmic interactions of the Gag proteins from human immunodeficiency virus type 1 (HIV-1) and human T-lymphotropic virus type 1 (HTLV-1). The Gag protein plays a crucial role in the assembly of retroviruses and is found both at the membrane and in the cytoplasm. Dual-color z-scans demonstrate that brightness artifacts are caused by a dim non-punctate membrane-bound fraction of Gag. We perform an unbiased brightness characterization of cytoplasmic Gag by avoiding the membrane bound fraction and reveal previously unknown differences in the behavior of the two retroviral Gag species. HIV-1 Gag exhibits concentration-dependent oligomerization in the cytoplasm, while HTLV-1 Gag lacks significant cytoplasmic Gag-Gag interactions.

Introduction

Two-photon fluorescence fluctuation spectroscopy (FFS) monitors the fluorescence fluctuations caused by single molecules that migrate in and out of a diffraction-limited observation volume. Analysis of the fluctuations provides information on the concentration, mobility, and brightness of fluorescent proteins (1-3). An exciting application of FFS lies in the characterization of protein-protein interactions in living cells by brightness analysis (1, 4). However, caution is required because the cellular environment is far more complex than that in an aqueous solution. We recently demonstrated that cytoplasmic brightness

analysis leads to artifacts if the observation volume exceeds the physical boundaries of the cell (5). Z-scan FCS was introduced to account for artifacts in the study of diffusion in both model and cellular lipid bilayers (6-8). Our group adapted this technique to account for the brightness artifacts observed in thin cells (5). Here we expand the repertoire of FFS by characterizing retroviral Gag proteins (9, 10), which are found both in the cytoplasm and on the plasma membrane. Gag molecules are retroviral structural proteins that have been shown to be sufficient for the production and release of viral-like particles (VLPs). VLPs exhibit similar size and morphology to immature infectious virions (11, 12). Hundreds to thousands of Gag proteins assemble and organize at the plasma membrane to form individual VLPs (13-15). It is believed, however, that the process of viral assembly is initiated in the cytoplasm, followed by Gag targeting to the membrane for further assembly (16-18). In this study, we investigated the cytoplasmic Gag-Gag interactions of two viruses (HIV-1 and HTLV-1) to characterize the initial steps of the virus particle assembly pathway.

Cells expressing Gag display bright puncta at the membrane, which are the putative assembly sites of VLPs. While it is easy to avoid puncta, FFS measurements resulted in artifacts that were ultimately traced to the existence of a non-punctate Gag-population at the membrane. The finite thickness of cytoplasmic sections leads to coexcitation of cytoplasmic and membrane-bound Gag, which obfuscates brightness analysis. In this work, we introduced a modified z-scan FFS technique to identify appropriate measurement positions for brightness analysis of cytoplasmic Gag proteins. The modified FFS technique relies on a dual-color z-scan, which was recently introduced into FCS (6).

We applied the technique to quantify cytoplasmic Gag-Gag interactions using brightness analysis for the first time. Our results demonstrate interesting differences in the behavior of HTLV-1 and HIV-1 Gag, which appear related to the myristic acid moiety known to be involved in Gag membrane targeting (16-18). We anticipate that the methodology described in this paper will prove useful

in the characterization of other cytoplasmic proteins that interact with the membrane.

Materials and Methods

FFS Instrumentation. A mode-locked Ti:sapphire laser (Tsunami, Spectra-Physics, Mountain View, CA) serves as a source for two-photon excitation of a modified Zeiss Axiovert 200 microscope (Thornwood, NY) as previously described (1). Cells reside in a chamber slide mounted on a motorized stage (PZ2000 piezo stage, ASI, Eugene, OR). The location and positioning of cells expressing both the Gag-EYFP construct and mCherry was achieved using fluorescence epi-illumination. The microscope was subsequently switched to bright-field illumination to verify cell-health as well as to locate a measurement position which appeared free from organelles. Finally, fluorescence epi-illumination was used to ensure that no puncta were present at the chosen measurement location.

Once a cell was selected, the excitation source was switched to the Ti:sapphire laser for FFS experiments. Excitation light is focused through a Zeiss 63x Plan Apochromat oil immersion objective (N.A. = 1.4). Control experiments (data not shown) confirm that 0.3 mW excitation power is sufficiently low to avoid saturation and photobleaching effects. Each FFS measurement lasts 82 s and uses two-photon excitation of the sample at 1000 nm. Intensity z-scans were obtained using a PZ2000 piezo stage (ASI, Eugene, OR) to move the sample in the axial direction. Scan voltages were controlled by an Agilent 33250A arbitrary waveform generator (Agilent Technologies, Santa Clara, CA) running a linear ramp signal with a frequency of 30.5 mHz and a peak-to-peak amplitude of 1.4 V. This voltage corresponds to an axial travel of 14.2 μm . The fluorescence emission of the FFS and z-scan measurements was separated into two different detection channels with a 580 nm dichroic mirror (585DCXR, Chroma Technology). Photon counts were detected with avalanche photodiodes (APD) (Perkin-Elmer, SPCM-AQ-14) and recorded by a data acquisition card (ISS,

Champaign, IL), which stores the complete sequence of photon counts using sampling frequencies ranging from 20 to 200 kHz. Two-photon imaging is described in the Supporting Material.

Sample Preparation and Data Analysis. Plasmid construction, transfection, and cell treatment are described in the Supporting Material. The photon counts were analyzed with programs written in IDL 6.0 (Research Systems, Boulder, CO). The determination of protein concentrations and brightness values from FFS experiments has been described previously (1, 2, 4). The z-scan fluorescent intensity traces $F(z; h)$ were fit using (5):

$$F(z; h) = A \int \text{RIPSF}(\xi) \cdot S(z + \xi) d\xi \quad (1)$$

The fluorescent intensity at position z for a slab of height h depends on the amplitude A and the spatial convolution of the radially-integrated point spread function RIPSF with a sample shape factor S . For cell measurements the shape factor is modeled as a slab of height h ,

$$S(z; h_0, h) = \begin{cases} 1, & h_0 < z < h_0 + h \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The parameter h_0 is required to describe the difference between the initial position of the z-scan and the start position of the slab. The RIPSF of the observation volume is modeled using a modified Gaussian-Lorentzian point spread function (PSF) as previously described (5). The beam parameters of the PSF were determined daily from fits of z-scan profiles to Eq. 1 and changed very little. The axial beam waist z_R is $\sim 0.7 \mu\text{m}$ and the y-factor is ~ 1.9 .

Results

Cells which express either the HIV-1 Gag-EYFP or the HTLV-1 Gag-EYFP construct exhibit both bright puncta and a diffuse, dim fluorescence from cytoplasmic regions of the cell (see Fig. S1 of Supporting Material). The puncta are generally believed to be plasma membrane-anchored sites of VLP assembly

containing hundreds to thousands of individual Gag copies (19-21), while the diffuse staining is believed to be the cytoplasmic fraction of Gag. To investigate the poorly understood cytoplasmic oligomerization-state of HTLV-1 and HIV-1 Gag, FFS brightness measurements of cytoplasmic Gag need to be conducted. To avoid contamination of the cytoplasmic Gag fluorescence by the puncta, FFS measurements are conducted at cellular locations that are free of puncta. The normalized brightness values acquired by this method are shown in Figure A5-1 as a function of concentration for HTLV-1 Gag-EYFP measured in HeLa cells. Each point in the graph represents a separate cytoplasmic measurement. Normalized brightness b is determined by calculating the ratio of the brightness of Gag-EYFP (λ_{app}) to the brightness of an EYFP monomer ($\lambda_{monomer}$) (1). Normalized brightness provides a measure of the average stoichiometry of the labeled proteins. For example, a normalized brightness of $b = 2$ corresponds to a dimer. The data has been screened for fluorescence intensity spikes that were found infrequently as described in the Supporting Material.

Despite avoiding puncta and the removal of intensity spikes, the brightness values in Fig. 1 exhibit an unusual feature. While the majority of the data fall between monomeric and dimeric brightness values, a substantial fraction of the data shows a normalized brightness significantly below a value of 1. Such low brightness values correspond to an unphysical Gag-EYFP stoichiometry of less than monomer. Despite the unphysical results, fluorescence autocorrelation data exhibited no unusual features. A discussion of the autocorrelation data is provided in the Supporting Materials. This unexpected result warrants a careful examination of potential artifacts that arise in FFS experiments.

One source of experimental artifacts in cytoplasmic brightness measurements is the cell thickness (5). FFS defines brightness as the average fluorescence intensity of a single molecule sampled at all locations of the observation volume. Focusing on the mid-section of a thin sample excludes access of the fluorophore to the outer edges of the observation volume, where

the intensity is lowest. As a consequence, cytoplasmic Gag-EYFP spends most of the time in the higher intensity, central areas of the excitation light, and the resulting spatial average skews the apparent brightness upwards. Z-scan FFS was introduced to correct this geometry related artifact (5). In this technique, the observation volume is scanned uniformly along the z-axis at a given cell location.

The FFS measurement is performed at the same location with the focus at the midsection of the cell. The intensity trace of the z-scan is determined by the convolution of the observation volume with the geometric profile of the cell at the location of the measurement. Based on previous work we modeled the observation volume of two-photon excitation by a modified Gaussian-Lorentzian PSF and the cell profile as a slab of thickness L with uniform fluorophore distribution (5). The fit of the intensity trace to Eq. 1 determines the shape factor γ_2 , which incorporates the geometry effect of sample thickness (5, 22). This shape factor is used to calculate the corrected brightness of the FFS measurement. Our earlier study found that without this correction the bias in brightness could be as high as a factor of two, but would always lead to an increase in brightness. Thus, z-scan FFS has the potential to reduce some of the high brightness scatter in the Gag-EYFP data, but cannot explain abnormally low brightness values.

Another factor potentially impacting FFS experiments is the presence of Gag-EYFP at the plasma membrane. While we selected measurement sites that appear free of puncta, we cannot rule out the presence of membrane-bound Gag-EYFP with certainty. The fluorescence contributions from a population of Gag-EYFP at the membrane could influence brightness determination and complicate positioning of the observation volume. Brightness measurements are typically conducted at the center of the cellular slab, which is identified by focusing the observation volume at the point of maximum fluorescent intensity along the z-profile of the cell. If plasma membrane-bound fluorescent protein is present, the maximum of the fluorescent intensity can shift towards the periphery of the cell, which leads to an off-center position of the observation volume. These

issues confound the interpretation of brightness measurements. However, it is difficult to experimentally identify the existence of a membrane fraction if its fluorescence intensity is of the same order as the cytoplasmic intensity. We found that autocorrelation analysis of the FFS data is not sufficient to unequivocally identify the presence of membrane-bound protein as discussed in the Supporting Materials.

These challenges prompted us to introduce an internal standard by coexpressing the red fluorescent protein mCherry together with Gag-EYFP. Because mCherry is a soluble protein that does not interact with membrane, it is a suitable protein for finding the center of the cell slab, determining cell thickness, and identifying Gag-EYFP at the membrane. The green channel records the fluorescence from Gag-EYFP, while the red channel records mCherry fluorescence as long as the crosstalk from Gag-EYFP is removed as described in the Supporting Materials. For simplicity, we refer to the red channel when specifying mCherry fluorescence without explicitly stating the correction procedure.

Figure A5-2 shows two-color z-scan intensity traces of HeLa cells expressing HTLV-1 Gag-EYFP and mCherry. It should be noted that HIV-1 Gag-EYFP expressing cells yielded data with the same features as presented in Figure A5-2. All intensity curves have been normalized to a maximum intensity of one to facilitate comparison. The z-scan profiles of Figure A5-2A illustrate perfect overlap between red and green channel data. Separate fits of the red and the green channel intensity profile to the slab model (Eq. 1) determined a thickness of $1.86 \pm 0.02 \mu\text{m}$ for both channels with reduced chi-square values of 1.4 and 1.5. Both the Gag-EYFP and mCherry fluorescence intensity profiles coincide and are well-described by a simple slab model as expected for purely cytoplasmic protein. We refer to z-scans with these attributes as membrane-negative, because of the absence of membrane-associated Gag-EYFP. Next, the observation volume is focused at the center of the slab using the red channel and the brightness of Gag-EYFP is determined. The cell thicknesses obtained from

the z-scan data is used to account for the impact of sample thickness on the brightness. Fig. A5-3A is a cartoon illustrating the hypothesized cellular conditions which gives rise to the z-scan data seen in Fig. A5-2A. The observation volume (gradient filled oval) is positioned in the center of the cytoplasm, with cytoplasmic Gag-EYFP being the sole Gag species present at the measurement position.

Figure A5-2B shows z-scan data from a different cell location than Figure A5-2A, but with comparable thickness. We notice immediately that the intensity profiles of the green and red channel are distinctly different. The intensity trace of the red channel (diamonds) is symmetric, has a single peak, and is well-approximated by a fit to a simple slab model (dashed line) with a thickness of $1.96 \pm 0.01 \mu\text{m}$. Thus, the z-scan data of the red channel demonstrate that mCherry is cytoplasmic. In contrast, the intensity profile of the green channel (solid line) has two peaks, is wider than the profile of mCherry, and cannot be reproduced by a simple slab model. The green channel data exhibits features that are consistent with a significant population of membrane-bound Gag-EYFP at both the top and bottom membrane giving rise to the two peaks as illustrated in Figure A5-3B.

The valley between the two peaks is formed by the signal from cytoplasmic Gag-EYFP with additional contributions from membrane-associated Gag-EYFP, which is excited at the periphery of the observation volume. We refer to z-scans that share the characteristics seen in Figure A5-3B as membrane positive, because the data reveal the presence of membrane-bound Gag-EYFP. While Figure A5-3B illustrates an example with Gag-EYFP at both the top and bottom membranes, we also found membrane-positive z-scans with Gag-EYFP only visible at one of the two membranes. It is important to note that the measurement site was selected in a region that appeared devoid of puncta. Thus, the membrane-bound Gag-EYFP (shown as gray boxes in Figure A5-3B) has to represent a population of Gag that is distinct from the puncta. Its intensity has to be much lower than that of puncta, which indicates a low degree of

oligomerization and a concentration comparable to cytoplasmic Gag. The population of membrane-bound Gag is further believed to be non-punctate as discussed at a later point. Immediately before taking the z-scan, an FFS measurement was performed at the center of the slab, which was identified using the red channel intensity. The normalized brightness of the measurement was, $b = 0.5$, which is significantly below one. This result strongly suggests that the unphysical low brightness values observed in Figure A5-1 are associated with the presence of membrane-bound Gag-EYFP.

Figure A5-3C is an example in which the Gag-EYFP's intensity profile of the green channel suggests the absence of membrane-bound Gag. However, comparison of the intensity profile of the green and red channel reveals a lack of overlap between them. The profile of the green channel (solid line) is broadened, which indicates the presence of a Gag membrane fraction. The mCherry fluorescence profile of the red channel (diamonds) is reproduced by a fit (dashed line) to a slab model with a cell thickness of $1.11 \pm 0.03 \mu\text{m}$. However, the profile of the green channel can also be approximated by a fit to a slab model, which results in a thickness of $1.75 \pm 0.02 \mu\text{m}$ and a reduced chi-square of 1.9. Thus, the data of the green channel on its own might lead to the false conclusion that the Gag protein is purely cytoplasmic. The z-scan traces of Figure A5-2C are an example in which the dual-color comparison of red and green channels is essential for the positive identification of Gag-EYFP at the membrane. The cartoon of Fig. A4-3C illustrates conditions that can give rise to the z-scan data of Fig. A5-2C. Just as in the case of Figure A5-3B, both cytoplasmic and membrane-associated Gag are present at the measurement site.

The crucial difference, which gives rise to the difference in the appearance of the two z-scans, is the cell thickness at the point of measurement. The spatial resolution of the observation volume is insufficient to identify the Gag fluorescence from the bottom and top membrane as independent peaks. The thin sample geometry effectively merges the membrane and cytoplasmic fluorescence contributions into a single, albeit slightly broadened peak. This

result implies that the relative broadening of the z-scan profile of the green channel to that of the red channel diminishes as the sample thickness decreases. For this reason, measurement locations with a thickness of $\sim 0.5 \mu\text{m}$ or less were rejected, because the spatial resolution is insufficient to unequivocally identify membrane-positive data. An FFS measurement was performed at the center of the cellular slab to determine the thickness-corrected brightness of Gag-EYFP. While the normalized brightness at this measurement position was higher, $b = 0.71$, than that measured at the position of Fig. A4-2B, it is still sufficiently low to be unphysical.

Thus far all measurements have been performed at locations free of puncta. We now purposefully target puncta in order to capture their fluorescence in a z-scan experiment. Figure A5-2D displays a successful z-scan measurement of a punctum. It should be noted that for purposes of clarity the green channel (solid line) was normalized using data outside of the large intensity spike present in the data. The normalization was performed in this manner to emphasize the spike's relative scale versus the red channel control (diamonds). The data clearly displays the profound impact of a Gag punctum positioned at the measurement site on the fluorescence intensity profile of the green channel. A cartoon of cellular conditions that give rise to the z-scan data seen in Fig. A5-2D is illustrated in Fig. A5-3D. The highly oligomerized Gag punctum resides at the top cellular membrane (filled circle). Though it is difficult to see due to the scale differences, the fluorescent intensity profile at the bottom membrane of the cell also exhibits a membrane-positive signature, which must be due to non-punctate membrane-bound Gag (gray box). This difference in intensity scale clearly establishes that there are two distinct pools of Gag at the membrane, bright puncta and dim non-punctate Gag.

All brightness experiments described in this paper have been performed on HeLa cells using z-scan FFS. In selected cases, multiple FFS measurements have been conducted in the same cell at different locations. We found both membrane-positive and membrane-negative z-scans within the same cell. In

addition, the experiments revealed that FFS measurements at membrane positive locations resulted in some but not all cases in an unphysical low brightness. These observations clearly spell out the need to discard all membrane positive FFS experiments to ensure that the measured brightness faithfully represents cytoplasmic Gag. Selection of membrane negative FFS experiments is in principle straightforward by examining the z-scan profiles. However, the z-scan and FFS measurement are not conducted simultaneously, but sequentially. If measurement locations switch from membrane positive to negative, or vice versa, during the timescale of the measurements, false negatives and positives are introduced, undermining the accuracy of the selection process. We performed a series of continuous z-scans and analyze the resulting kymographs to determine the temporal stability of membrane-positive and membrane-negative locations (see Supporting Materials). These experiments establish that the sequential z-scan and FFS experiments provide a robust method for identifying the presence or absence of Gag-EYFP at the membrane.

We now have the tools to revisit the brightness characterization of cytoplasmic Gag-EYFP. Dual-color z-scan FFS is used to select membrane negative experiments and determine their brightness, while discarding any measurements with membrane positive characteristics. Figure A5-4A shows the normalized brightness versus concentration for cytoplasmic HIV-1 Gag-EYFP (diamonds) measured at membrane negative locations. It is immediately apparent that dual-color z-scan FFS has been successful in removing the unphysical low brightness values seen in Figure A5-1. The lowest normalized brightness data are consistent with monomeric Gag-EYFP. Brightness values slightly below one reflect the statistical uncertainty in the experiment, which in our experience is ~10% for cellular brightness measurements (see Supplemental Materials for further details concerning brightness statistics).

The data reveal the existence of monomeric HIV-1 Gag at low concentrations and hint at a concentration dependent Gag oligomerization in the

cytoplasm, because the data trends towards slightly elevated brightness values at the upper end of the measured concentrations. This result is consistent with the myristoyl switch model (16, 17, 23, 24). This model predicts that Gag-Gag interactions in the cytoplasm directly lead to membrane-binding through the exposure of a myristic acid moiety at the N-terminus of the Gag molecules. Gag monomers, on the other hand, sequester the myristic acid moiety in a hydrophobic pocket of the matrix domain, and thus are restricted to the cytoplasm.

To facilitate observation of cytoplasmic Gag-Gag interactions, we measured the G2A Gag mutant, which lacks the myristoyl moiety and is therefore membrane-binding deficient (25, 26). In contrast to wild-type HIV-1 Gag, the HIV-1 G2A mutant exhibits an increased brightness and concentration range (Figure A5-4B). Brightness of the G2A mutant (asterisks) increases monotonically with concentrations and reaches a value of ~ 8 , which reflects an average complex size of eight Gag molecules. Thus, dimerization of Gag is followed by further increase in oligomerization. In addition, G2A Gag mutants reach cytoplasmic concentrations almost a degree of magnitude higher than wild-type HIV-1 Gag. Despite the contrast in the two results, the brightness of wild-type HIV-1 Gag (gray diamonds) and G2A mutant (asterisks) are identical over the concentration range accessible to the wild-type protein.

Figure A5-4C shows the normalized brightness versus concentration for cytoplasmic HTLV-1 Gag-EYFP (diamonds) measured at membrane negative locations. As with the HIV-1 Gag data, dual-color z-scan FFS has been successful in removing the unphysical low brightness values seen in Figure A5-1. The lowest and highest normalized brightness values are 1 and 1.6 if the experimental uncertainty in brightness determination is taken into account. Unlike HIV-1 Gag, there is no discernible concentration-dependence of the brightness. We notice instead a large but uniform scatter in the brightness throughout the concentration range. This result points to a potential difference between HIV-1 and HTLV-1 Gag behavior.

Analogous to our approach to HIV-1 Gag, we now turn to the HTLV-1 Gag G2A mutant for further investigation of cytoplasmic Gag-Gag interactions (Figure A5-4D). Again, the G2A mutant lacks the myristoyl post-translational modification, and so is deficient in membrane targeting. While the G2A Gag mutant (asterisks) reaches cytoplasmic concentrations almost a degree of magnitude higher than wild-type HTLV-1 Gag, its brightness behavior departs markedly from that of the HIV-1 G2A species. The normalized brightness remains close to one over the entire concentration range. The highest brightness values for the HTLV-1 G2A mutant, around 1.4, actually lie below the highest measured values for wild-type HTLV-1 Gag. Another striking difference between HTLV-1 mutant and wild-type is the degree of scatter in the G2A data (Figure A5-4D), which is considerably higher for the wild-type protein (gray diamonds). The FFS characterization of HTLV-1 G2A Gag did not follow the results expected by the myristoyl-switch model. Our data point to the near absence of HTLV-1 Gag oligomerization at all cytoplasmic concentrations.

Discussion

Retroviral Gag proteins are the structural proteins of retroviruses, assembling into the retroviral structural lattice consisting of hundreds to thousands of Gag molecules (13, 14, 27). It is believed that after Gag molecules initiate lattice assembly in the cytoplasm, the Gag-Gag interactions target small Gag complexes to the inner leaflet of the plasma membrane for further assembly (for reviews, see (9) and (10)). The degree of Gag oligomerization in the cytoplasm remains poorly understood. Techniques such as FRET have confirmed the presence of cytoplasmic Gag-Gag interactions, but quantification remains lacking (28, 29). FFS brightness analysis (1, 2, 4) has the potential to shed light on this area of Gag behavior, but requires the development and adoption of new FFS protocols.

Previous brightness studies conducted by our group were restricted to soluble proteins residing exclusively in the nucleoplasm or cytoplasm of cells (1,

2, 4). The study of Gag characterizes a protein with simultaneous presence in the cytoplasm and as puncta at the plasma membrane. Although puncta were avoided in FFS measurements, the brightness of Gag-EYFP revealed unusual characteristics, which we ultimately linked to the presence of non-punctate Gag-EYFP at the plasma membrane. Adding membrane-bound Gag-EYFP (Figure A5-3B) introduces a second species that is coexcited with cytoplasmic Gag. The exact brightness of a mixture of species depends on many details, but its value will always lie between the brightness of the dimmest and brightest species.

The membrane-bound species depicted in Figure A5-3B will contribute an unusually low brightness, because of its location at the periphery of the observation volume, where the excitation intensity is weak. This reasoning implies that membrane-bound Gag will have a lower brightness than cytoplasmic Gag. Consequently, the brightness of the mixture containing dim membrane-bound Gag is less than the brightness of cytoplasmic Gag. For example, if cytoplasmic Gag has a normalized brightness of one, the mixture exhibits a value of less than one. This scenario provides a qualitative explanation for the unphysical low brightness values observed in Figure A5-1. We previously demonstrated that the presence of immobile fluorophores decreases brightness (30).

Note that the reduction of brightness for the current case is not caused by immobile fluorophores, but is due to a geometric effect that restricts mobile fluorophores to the periphery of the observation volume. Thus, the influence of membrane-bound Gag on FFS experiments depends in general on geometric factors. As the cell thickness reduces, the membrane-bound species moves closer to the center of the observation volume (Figure A5-3C), thereby increasing its brightness. Under these circumstances the brightness at the membrane is not necessarily lower than for cytoplasmic Gag.

In fact, the small drop in brightness from the off-center positioning of the membrane can be sufficiently small that membrane-bound Gag is brighter than cytoplasmic Gag, if the degree of oligomerization at the membrane is higher than

in the cytoplasm. For example, if Gag at the membrane is predominantly dimeric, its brightness should be higher or at least comparable to the brightness of cytoplasmic Gag. In conclusion, adding membrane-bound Gag to a population of cytoplasmic Gag may lead to either an increase or a decrease in the measured brightness. Correcting the change in brightness is difficult, because the sample geometry, the oligomeric state, and the concentration of all species play a role. Because these parameters are not sufficiently well known, it is best to avoid the complications that arise from membrane-bound Gag.

Measuring in thick sections, where the observation volume is completely contained in the cytoplasm, offers a straightforward solution that prevents excitation of proteins at the membrane. Because the height of spreading HeLa cells decreases rapidly outside the nuclear region of the cell, the only location thick enough to accommodate the observation volume is the perinuclear region. Unfortunately, performing quantitative FFS measurements in this region is exceedingly difficult, because organelles and fluorescence from internal membranes introduce analytically intractable artifacts. For this reason, we are forced to measure thin cell regions. This approach requires a technique that accurately identifies fluorescence contributions from the plasma membrane to prevent brightness artifacts. We demonstrate in this paper that z-scan FFS provides a suitable solution.

Because the z-profile of a protein with a cytoplasmic and membrane-bound population is complex, we implemented dual-color z-scan FFS to facilitate its characterization. The green-channel detects the EYFP-labeled protein, while the red-channel monitors mCherry, which serves as a marker of the cytoplasmic volume. A membrane negative scan is defined by identical shapes for the green and red channel, which signals the absence of membrane-bound protein (Figure A5-2A). A membrane-positive scan is characterized by deviating shapes in both channels, which identify the presence of membrane-bound protein (Figure A5-2B). The z-profile of the green channel is often sufficient to identify membrane-positive scans (Figure A5-3B), but comparison with the red channel becomes

crucial once the cell section becomes thinner (Figure A5-2C). It is important to recognize that the method breaks down once the cell thickness falls below a critical value, because the optical resolution of the z-scan is insufficient to pick out the broadening of the green-channel profile. We discarded all measurements with a thickness of less than 0.5 μm , which is the critical thickness for our current setup.

The intensity contrast between the z-scan at a puncta-positive (Figure A5-2D) and experimental z-scans at puncta-negative (Figures A5-2B & C) positions clearly demonstrate the existence of a non-punctate population of Gag at the membrane. The non-punctate Gag population has to have a low degree of oligomerization, otherwise z-scan FFS measurements at thin cytoplasmic sections (Figure A5-2C) would yield large brightness values, which was not observed. In addition, kymographs confirm that the Gag membrane fraction persists for many minutes, while its fluorescence intensity fluctuates with time (data not shown). This implies that the non-punctate membrane fraction is mobile and spatially distributed.

The non-punctate fraction is not uniformly distributed across the plasma membrane, because we have found both membrane-negative and membrane-positive locations within a single cell. To the best of our knowledge, our results provide the first direct observation of a non-punctate Gag population at the plasma membrane of living cells. However, our finding is not unexpected, since it is well-known that low-order HIV-1 Gag complexes are capable of binding the plasma membrane (for reviews, see (9) and (10)). In addition, biochemical experiments indicate the interaction of HTLV-1 Gag with the plasma membranes of 293T cells (31).

Because the focus of this article is the characterization of cytoplasmic Gag-Gag interactions, we eliminated membrane-positive measurements with dual-color z-scan FFS and directly recorded cytoplasmic Gag-EYFP brightness values that appear free of artifacts. Brightness analysis of wild-type HIV-1 Gag is consistent with the myristoyl-switch model. The protein is monomeric at low

concentrations. The increase in brightness towards higher concentration (Figure A5-4A) indicates the onset of Gag-Gag interactions. Gag-Gag interactions initiate the myristoyl switch, which targets wild-type oligomers to the membrane, limiting Gag accumulation in the cytoplasm. By comparison, the HIV-1 G2A Gag mutant is membrane-binding deficient as confirmed by the absence of membrane-positive z-scans. This allows for the accumulation and consequent oligomerization of G2A Gag in the cytoplasm. Both proteins indicate the onset of detectable Gag-Gag interactions at ~500 nM, but only the mutant establishes the formation of dimer and higher order Gag oligomers in the cytoplasm, as membrane-targeting depletes the wild-type oligomers (Figure A5-4B).

HTLV-1 and HIV-1 Gag show distinct differences in their oligomerization. This difference is most clearly visible when comparing their G2A mutants. The HTLV-1 G2A Gag lacked the concentration-dependent oligomerization behavior exhibited by HIV-1 G2A Gag, although both proteins reach comparable concentrations in the cytoplasm (Figures A5-4B and D). While there is a slight trend upwards in the HTLV-1 G2A brightness data with concentration, the brightness data essentially indicate the absence of Gag oligomerization in the concentration range accessible to wild-type HTLV-1 Gag. Thus, according to the myristoyl switch model, HTLV-1 Gag should not interact with the membrane, which is clearly inconsistent with our experimental z-scan data. Based on the above results we would expect to observe monomeric HTLV-1 Gag in the cytoplasm. The experiment, however, detects a concentration independent normalized brightness with an unusual large scatter with values ranging from 1 to 1.6 (Figures A5-4C and D). The amplitude of the scatter is much larger than expected from the experimental uncertainty of brightness measurements (see Supplemental Materials). While the reason for this scatter is unknown, its presence indicates that parameters other than concentration play a role in determining HTLV-1 Gag oligomerization.

Our data suggest that the myristoyl moiety might play a different role in Gag oligomerization for HTLV-1 when compared to HIV-1 Gag. Our observations

are consistent with a model where HTLV-1 Gag is capable of targeting the membrane without the requirement of prior cytoplasmic interactions. This result implies myristoyl moiety exposure on HTLV-1 Gag monomers, which lies in contrast to the sequestered myristoyl moiety of HIV-1 Gag monomers. The evidence is consistent with previous biochemical work by Mamoun and coworkers which found no evidence for HTLV-1 G2A mutant interactions in cell extracts (31, 32). This difference in membrane targeting which HTLV-1 Gag exhibits may be a contributing factor to the concentration-independent brightness scatter observed for the wild-type species. More work needs to be done to account for the impact of such a model in our interpretation of observed cytoplasmic brightnesses for HTLV-1 Gag

Performing FFS measurements on cytoplasmic proteins with a membrane-bound population requires special precautions to avoid artifacts. While it is straightforward to identify and avoid bright puncta, the presence of a dispersed membrane-bound population is much more difficult to recognize. The introduction of dual-color z-scan FFS provides a convenient method to identify the presence of membrane-bound protein. Furthermore, we removed brightness artifacts by selecting membrane-negative z-scan FFS data and performed the first quantification of Gag oligomerization in the cytoplasm of living cells. Our data demonstrate that cytoplasmic HIV-1 and HTLV-1 Gag differ in their interactions, which suggests differences in the assembly pathway of both viruses. Further work is needed to investigate these differences.

Cytoplasmic proteins that interact with the membrane are ubiquitous. We expect that dual-color z-scan FFS will prove useful in the brightness characterization of such proteins. At this point, dual color z-scan FFS is used to reject membrane-positive data. A potentially interesting and useful extension of the technique would allow the simultaneous characterization of cytoplasmic and membrane-bound proteins. In the future, we hope to apply such a technique towards a better understanding of retroviral assembly.

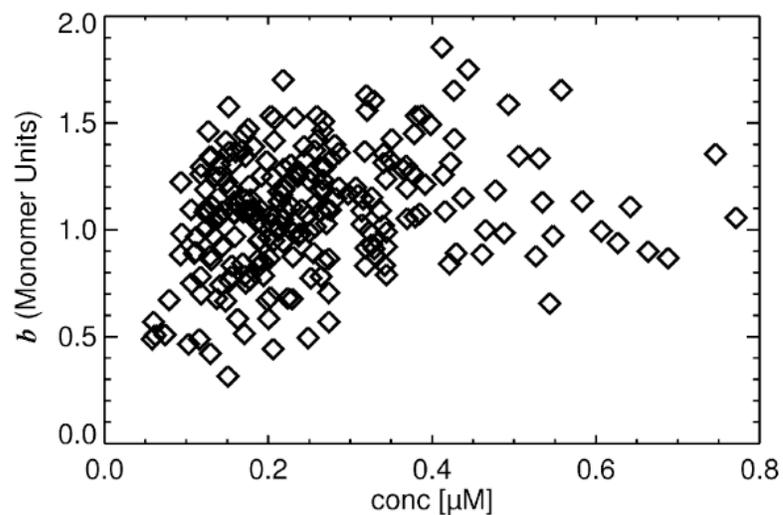


Figure A5-1. Normalized brightness of HTLV-1 Gag-EYFP expressed in HeLa cells measured at locations that appeared free of puncta. The data were screened to eliminate artifacts due to intensity spikes. Normalized brightness values of less than one indicate an unphysical stoichiometry of less than one for Gag-EYFP.

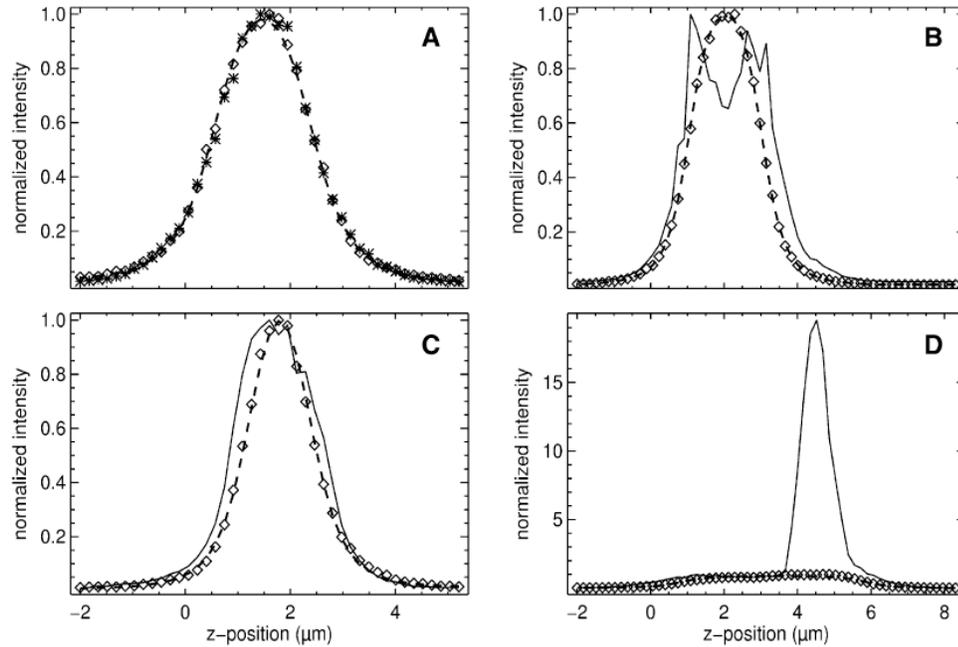


Figure A5-2. Dual-color z-scan intensity traces from HeLa cells expressing HTLV-1 Gag-EYFP and mCherry. Fluorescence intensity traces were obtained for Gag-EYFP (green channel) and mCherry (red channel) as detailed in the article. The intensity trace with amplitude normalized to one is graphed versus the axial beam position. (A) Example of a membrane-negative z-scan. The normalized intensity trace of the green (asterisks) and red (diamonds) channel overlap completely. The fit (solid line) of the trace to a slab model identifies a thickness of $1.86 \pm 0.02 \mu\text{m}$. Purely cytoplasmic Gag-EYFP yields a membrane-negative z-scan. (B) Example of a membrane-positive z-scan. The distinct shape of the green channel curve (solid line) indicates the presence of membrane-bound Gag-EYFP. For comparison, cytosolic mCherry is represented by the red-channel data (diamonds) and its fit (dashed line) to a slab model. (C) Example of a membrane-positive z-scan. Here the shape of the green-channel data (solid line) is not sufficient to recognize the presence of membrane-bound Gag-EYFP. Comparison with the red channel curve (diamonds) is needed to identify the existence of membrane-bound Gag. The dashed line denotes the fit of the red channel data to a slab model. (D) Example of a z-scan at a puncta location. Normalization of the amplitude of the green channel curve (solid line) was

performed outside the region with the intensity spike to illustrate the high intensity of the punctum. The mCherry fluorescence (diamonds) of the red channel and its fit (dashed line) to a slab model is shown for reference. Comparison of the green and red channel reveals in addition a dim non-punctate Gag-EYFP membrane fraction at the bottom membrane.

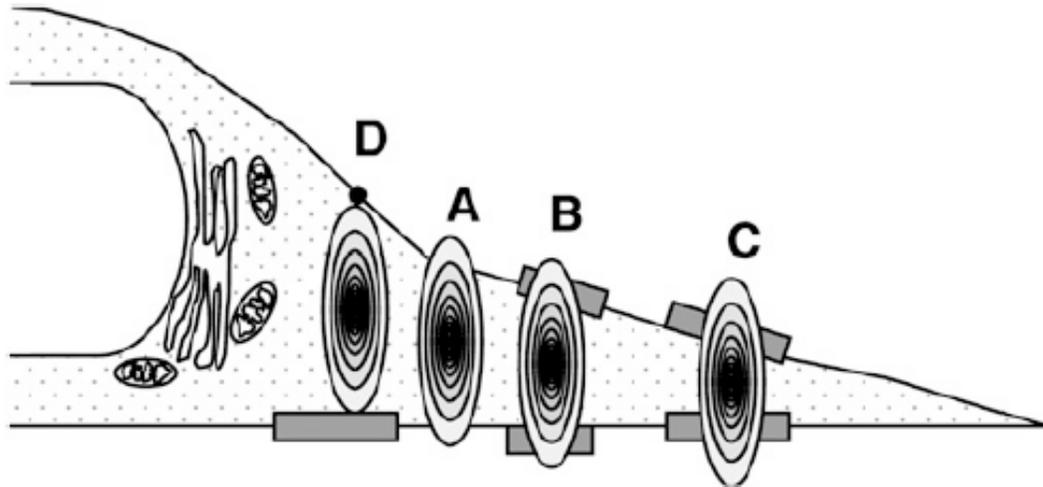


Figure A5-3. Depiction of the vertical cross section of a cell with the FFS observation volume (oval) placed at four different cytoplasm locations. The image includes cytoplasmic Gag (dotted fill pattern), non-punctate membrane-bound Gag (gray bar), and punctate Gag (solid circle). The four measurement locations (A, B, C, D) provide an interpretation of the z-scans of Figure A5-2. (A) Location with purely cytoplasmic Gag. (B) Thick cytoplasmic slab with non-punctate Gag at the top and bottom membrane. (C) Thin cytoplasmic slab with non-punctate Gag at the top and bottom membrane. (D) Cytoplasmic slab with punctum at the top membrane and non-punctate Gag at the bottom membrane.

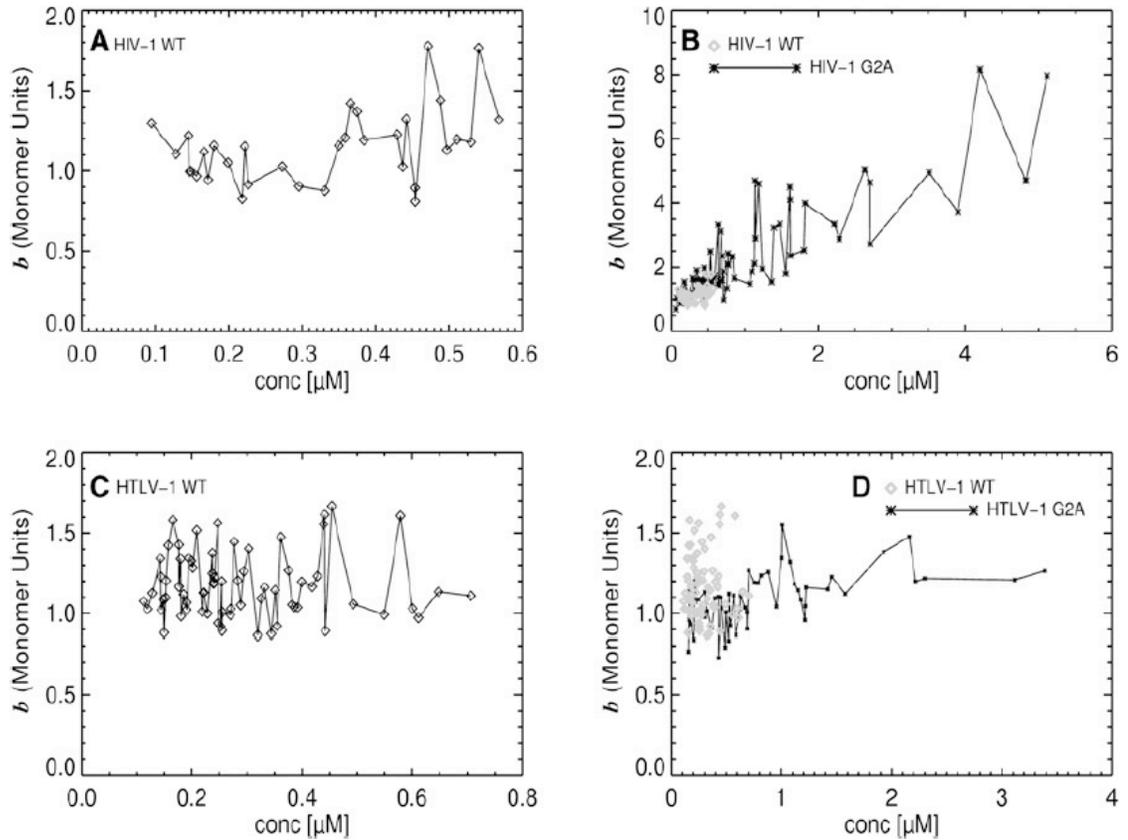


Figure A5-4. Normalized brightness of cytoplasmic Gag-EYFP versus concentration from membrane-negative dual-color z-scan FFS data taken in HeLa cells. (A) Normalized brightness of HIV-1 Gag-EYFP; (B) Normalized brightness of HIV-1 G2A Gag-EYFP (asterisks) and HIV-1 Gag-EYFP (gray diamonds); (C) Normalized brightness of HTLV-1 Gag-EYFP; (D) Normalized brightness of HTLV-1 G2A Gag-EYFP (asterisks) and HTLV-1 Gag-EYFP (gray diamonds). The data reveal distinct differences for the Gag-Gag interactions for both viruses as further discussed in the main article.

References

1. **Chen Y, Wei LN, Muller JD.** 2003. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *Proc Natl Acad Sci U S A* **100**:15492-15497.
2. **Chen Y, Muller JD, Ruan QQ, Gratton E.** 2002. Molecular brightness characterization of EGFP in vivo by fluorescence fluctuation spectroscopy. *Biophys J* **82**:133-144.
3. **Berland KM, So PTC, Gratton E.** 1995. 2-Photon Fluorescence Correlation Spectroscopy - Method and Application to the Intracellular Environment. *Biophys J* **68**:694-701.
4. **Chen Y, Wei LN, Muller JD.** 2005. Unraveling protein-protein interactions in living cells with fluorescence fluctuation brightness analysis. *Biophys J* **88**:4366-4377.
5. **Macdonald PJ, Chen Y, Wang X, Mueller JD.** 2010. Brightness Analysis by z-scan Fluorescence Fluctuation Spectroscopy for the Study of Protein Interactions within Living Cells. *Biophys J* **99**:979-988.
6. **Stefl M, Kulakowska A, Hof M.** 2009. Simultaneous Characterization of Lateral Lipid and Prothrombin Diffusion Coefficients by Z-Scan Fluorescence Correlation Spectroscopy. *Biophys J* **97**:LO1-LO3.
7. **Humpolickova J, Gielen E, Benda A, Fagulova V, Vercammen J, Vandeven M, Hof M, Ameloot M, Engelborghs Y.** 2006. Probing diffusion laws within cellular membranes by Z-scan fluorescence correlation spectroscopy. *Biophys J* **91**:L23-L25.
8. **Benda A, Benes M, Marecek V, Lhotsky A, Hermens WT, Hof M.** 2003. How to determine diffusion coefficients in planar phospholipid systems by confocal fluorescence correlation spectroscopy. *Langmuir* **19**:4120-4126.
9. **Swanstrom R, Wills JW.** 1997. Synthesis, assembly, and processing of viral proteins. *Retroviruses*:263-334.
10. **Klein KC, Reed JC, Lingappa JR.** 2007. Intracellular destinies: Degradation, targeting, assembly, and endocytosis of HIV gag. *Aids Rev* **9**:150-161.
11. **Gheysen D, Jacobs E, Deforesta F, Thiriart C, Francotte M, Thines D, Dewilde M.** 1989. Assembly and Release of Hiv-1 Precursor Pr55gag Virus-Like Particles from Recombinant Baculovirus Infected Insect Cells. *Cell* **59**:103-112.
12. **Karacostas V, Nagashima K, Gonda MA, Moss B.** 1989. Human Immunodeficiency Virus-Like Particles Produced by a Vaccinia Virus Expression Vector. *Proc Natl Acad Sci U S A* **86**:8964-8967.
13. **Chen Y, Wu B, Musier-Forsyth K, Mansky LM, Mueller JD.** 2009. Fluorescence Fluctuation Spectroscopy on Viral-Like Particles Reveals Variable Gag Stoichiometry. *Biophys J* **96**:1961-1969.
14. **Grigsby IF, Zhang W, Johnson JL, Fogarty KH, Chen Y, Rawson JM, Crosby AM, Mueller JD, Mansky LM.** 2010. Biophysical analysis of

- HTLV-1 Particles Reveals Novel Insights into Particle Morphology and Gag Stoichiometry. *Retrovirology* **7**:doi:10.1186/1742-4690-1187-1175.
15. **Carlson LA, Brigg JAG, Glass B, Riches JD, Simon MN, Johnson MC, Muller B, Grunwald K, Krausslich HG.** 2008. Three-Dimensional Analysis of Budding Sites and Released Virus Suggests a Revised Model for HIV-1 Morphogenesis. *Cell Host Microbe* **4**:592-599.
 16. **Zhou WJ, Resh MD.** 1996. Differential membrane binding of the human immunodeficiency virus type 1 matrix protein. *J Virol* **70**:8540-8548.
 17. **Resh MD.** 2004. A myristoyl switch regulates membrane binding of HIV-1 Gag. *Proc Natl Acad Sci U S A* **101**:417-418.
 18. **Tang C, Loeliger E, Luncsford P, Kinde I, Beckett D, Summers MF.** 2004. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proc Natl Acad Sci U S A* **101**:517-522.
 19. **Le Blanc I, Blot V, Bouchaert I, Salamero J, Goud B, Rosenberg AR, Dokhlar MC.** 2002. Intracellular distribution of human T-cell leukemia virus type 1 Gag proteins is independent of interaction with intracellular membranes. *J Virol* **76**:905-911.
 20. **Wang HT, Machesky NJ, Mansky LM.** 2004. Both the PPPY and PTAP motifs are involved in human T-cell leukemia virus type 1 particle release. **78**:1503-1512.
 21. **Jouvenet N, Bieniasz PD, Simon SM.** 2008. Imaging the biogenesis of individual HIV-1 virions in live cells. *Nature* **454**:236-240.
 22. **Thompson NL.** 1991. Fluorescence correlation spectroscopy. *Topics in Fluorescence Spectroscopy* **1**:337-378.
 23. **Resh MD.** 2005. Intracellular trafficking of HIV-1 Gag: How Gag interacts with cell membranes and makes viral particles. *Aids Rev* **7**:84-91.
 24. **Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF.** 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. **103**:11364-11369.
 25. **Gottlinger HG, Sodroski JG, Haseltine WA.** 1989. Role of Capsid Precursor Processing and Myristoylation in Morphogenesis and Infectivity of Human Immunodeficiency Virus Type-1. *Proc Natl Acad Sci U S A* **86**:5781-5785.
 26. **Lindwasser OW, Resh MD.** 2004. Human immunodeficiency virus type 1 Gag contains a dileucine-like motif that regulates association with multivesicular bodies. *J Virol* **78**:6013-6023.
 27. **Briggs JAG, Riches JD, Glass B, Bartonova V, Zanetti G, Krausslich HG.** 2009. Structure and assembly of immature HIV. *Proc Natl Acad Sci U S A* **106**:11090-11095.
 28. **Larson DR, Ma YM, Vogt VM, Webb WW.** 2003. Direct measurement of Gag-Gag interaction during retrovirus assembly with FRET and fluorescence correlation spectroscopy. *J Cell Biol* **162**:1233-1244.
 29. **Derdowski A, Ding LM, Spearman P.** 2004. A novel fluorescence resonance energy transfer assay demonstrates that the human

- immunodeficiency virus type 1 Pr55(Gag)I domain mediates Gag-Gag interactions. *J Virol* **78**:1230-1242.
30. **Skinner JP, Chen Y, Muller JD.** 2008. Fluorescence fluctuation spectroscopy in the presence of immobile fluorophores. *Biophys J* **94**:2349-2360.
 31. **Rayne F, Kajava AV, Lalanne J, Mamoun RZ.** 2004. In vivo homodimerisation of HTLV-1 Gag and MA gives clues to the retroviral capsid and TM envelope protein arrangement. *J Mol Biol* **343**:903-916.
 32. **Rayne F, Bouamr F, Lalanne J, Mamoun RZ.** 2001. The NH2-terminal domain of the human T-cell leukemia virus type 1 capsid protein is involved in particle formation. *J Virol* **75**:5277-5287.

APPENDIX VI

DECLARATION OF CONTRIBUTIONS TO CO-AUTHORED PUBLICATION: BIOPHYSICAL ANALYSIS OF HTLV-1 PARTICLES REVEALS NOVEL INSIGHTS INTO PARTICLE MORPHOLOGY AND GAG STOICHIOMETRY

I am a co-author on the following publication:

Grigsby IF, Zhang W, Johnson JL, Fogarty KH, Chen Y, **Rawson JM**, Crosby AJ, Mueller JD, Mansky LM. Biophysical analysis of HTLV-1 particles reveals novel insights into particle morphology and Gag stoichiometry. *Retrovirology* (2010) 7/75. This article is openly distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>).

My contributions to this publication are as follows:

- Designed, constructed, and verified the HTLV-1 codon-optimized Gag-EYFP reporter construct
- Assisted in writing the corresponding Materials and Methods section of the manuscript

Abstract

Human T-lymphotropic virus type 1 (HTLV-1) is an important human retrovirus that is a cause of adult T-cell leukemia/lymphoma. While an important human pathogen, the details regarding virus replication cycle, including the nature of HTLV-1 particles remain largely unknown due to the difficulties in propagating the virus in tissue culture. In this study, we created a codon-optimized HTLV-1 Gag fused to an *EYFP* reporter as a model system to quantitatively analyze HTLV-1 particles released from producer cells. The codon-optimized Gag led to a dramatic and highly robust level of Gag expression as well as virus-like particle (VLP) production. The robust level of particle production overcomes previous technical difficulties with authentic particles and allowed for detailed analysis of particle architecture using two novel methodologies. We quantitatively measured the diameter and morphology of HTLV-1 VLPs in their native, hydrated state using cryo-transmission electron microscopy (cryo-TEM). Furthermore, we were able to determine HTLV-1 Gag stoichiometry as well as particle size with the novel biophysical technique of fluorescence fluctuation spectroscopy (FFS). The average HTLV-1 particle diameter determined by cryo-TEM and FFS was 71 ± 20 nm and 75 ± 4 nm, respectively. These values are significantly smaller than previous estimates made of HTLV-1 particles by negative staining TEM. Furthermore, cryo-TEM reveals that the majority of HTLV-1 VLPs lacks an ordered structure of the Gag lattice, suggesting that the HTLV-1 Gag shell is very likely to be organized differently compared to that observed with HIV-1 Gag in immature particles. This conclusion is supported by our observation that the average copy number of HTLV-1 Gag per particle is estimated to be 510 based on FFS, which is significantly lower than that found for HIV-1 immature virions. In summary, our studies represent the first quantitative biophysical analysis of HTLV-1-like particles and reveal novel insights into particle morphology and Gag stoichiometry.

Introduction

There are approximately 15-20 million people infected by human T-lymphotropic virus type 1 (HTLV-1) worldwide (1). HTLV-1 infection can result in a number of severe disorders including adult T cell leukemia/lymphoma (ATLL) as well as HTLV-1 associated myelopathy/tropical paraparesis (HAM/TSP) (2, 3). Despite its association with cancer and significant impact on human health, many of the details regarding the replication, assembly and fundamental virus particle structure remain poorly understood.

The Gag polyprotein is the main retroviral structural protein and is sufficient, in the absence of other viral proteins, for the production and release of immature VLPs (4). The Gag polyprotein is composed of three functional domains: matrix (MA), capsid (CA), and nucleocapsid (NC). Typically upon budding or immediately after immature particle release, proteolytic cleavage of the Gag polyproteins takes place and results in virus particle core maturation. The Gag polyprotein is cleaved into MA, CA, and NC by the viral protease. The newly processed proteins reorganize into structurally distinct mature virions: MA remains associated with the viral membrane, CA undergoes conformational changes and reassembles into a viral core, which encapsulates a complex of NC, genomic RNA, and other important viral proteins (5-7).

Studies with many retroviruses, including human immunodeficiency virus type 1 (HIV-1), have shown that retroviral assembly is initiated by binding the myristoyl moiety of MA with lipid rafts at the plasma membrane (8-11). The MA-membrane interaction is thought to stimulate Gag oligomerization, the interaction between viral genomic RNA and NC, and the recruitment of a variety of host factors. Accumulation of Gag at the plasma membrane triggers the activation of the ESCRT machinery which creates the membrane curvature that results in the budding of immature virus particles (12). Analysis of Gag molecules in immature HIV-1 particles have revealed that the MA domain is located at the membrane with the CA and NC domains projecting towards the center of the particle (13).

Cryo-electron tomography (cryo-ET) combined with three-dimensional (3D) reconstructions have provided highly detailed structural information for HIV-1. Structural studies have revealed that HIV-1 Gag proteins form an incomplete paracrystalline lattice in immature particles (14, 15). This incomplete Gag lattice was observed to consist of a hexameric organization with 80-Å distance between neighboring ring-like structures (14, 15). While the myristoyl moiety of MA appeared to be associated with membrane, the hexameric ring structure in the 3D maps were attributed to CA, and the Gag-Gag interactions in the immature particles were proposed to be primarily stabilized by CA and SP1, rather than the affinity of membrane-binding via MA (15).

Despite limited amino acid sequence homology among different retroviruses, the atomic tertiary structures of individual Gag domains exhibit high similarity (16-18). Therefore, structural and assembly mechanisms of HIV-1 are generally used as a reference model for other retroviruses. However, structural evidence indicates that the conservation of Gag organization between HTLV-1 and HIV-1 is poorly understood. In this study, we have performed cryo-TEM on HTLV-1-like particles. Our study is the first to study HTLV-1 particles in their native, hydrated state. Our results demonstrate an average HTLV-1 particle diameter of ~ 73 nm, which is smaller than previously predicted based on conventional negative staining TEM (19). Using the novel biophysical technology of FFS, we further demonstrate that there are ~ 510 copies of Gag per HTLV-1 particle, a number that is significantly lower than what is typically found in HIV-1 particles. Finally, our cryo-TEM images analysis reveals a less ordered Gag structure compared to that reported for HIV-1, suggesting that the HTLV-1 Gag shell has a distinct architecture.

Results

Creation of a tractable and robust system for the production of HTLV-1-like particles. Previous molecular analyses of HTLV-1 replication have been severely hampered by the fragility of HTLV-1 proviral sequences as well as the low levels of viral replication in tissue culture. Given the technical and experimental limitations of working with HTLV-1, we first sought to create an experimental model system that would be amenable to successfully and efficiently analyze HTLV-1 Gag trafficking and virus particle assembly and release. It is well-established that retroviral Gag polyprotein is sufficient for the assembly and release of VLPs [reviewed by (20)]. Our previous studies indicated that HTLV-1 Gag constructs express Gag at low levels (Huating Wang and Louis Mansky, unpublished observations), presumably due to missing cis-elements on the RNA transcript required for efficient nuclear export.

In order to create a tractable and robust system for Gag expression and virus-like particle production, we designed and created a codon-optimized HTLV-1 Gag construct to improve HTLV-1 Gag expression. In order to readily detect Gag expression, trafficking, and incorporation into VLPs, we fused the EYFP to the C-terminal end of the Gag protein. Figure A6-1A shows the HTLV-1 Gag-EYFP expression construct. In this construct, the Gag-EYFP is expressed from a CMV promoter, and a Kozak consensus sequence was engineered upstream of the start codon to facilitate translation initiation as well as an in-frame insertion of the EYFP gene sequence just prior to the HTLV-1 Gag gene stop codon. The plasmid is quite stable and readily amplified in *E. coli* (data not shown).

To confirm expression of the fusion construct, 293T cells were transiently transfected with three independent clones of pEYFP-N3 HTLV-1 Gag in parallel experiments. Thirty-six hours post-transfection, HTLV-1 Gag-EYFP protein expression was examined from both cell culture supernatants (Fig. A5-1B, lane 1-3) and from cellular lysates (Fig. A5-1B, lane 4-6). The Gag precursor-EYFP fusion protein, with a molecular mass of approximately 80 kDa was very readily

observed, with each of the 3 clones analyzed expressing very high and comparable levels of HTLV-1 Gag-EYFP. The minor bands of smaller molecular mass likely represent partially degraded HTLV-1 Gag-EYFP and not cleavage products of the viral protease, since it is not present in the Gag expression construct. The Gag-EYFP observed in VLPs was primarily full length (Fig. A5-1B, lane 1-3), with undetectable levels of mature capsid (p24) protein.

To investigate the morphology of the particles produced from cells expressing pEYFP-N3 HTLV-1 Gag, transiently transfected 293T cells were harvested and examined by TEM. MT-2 cells, a T-cell line chronically infected by HTLV-1, were examined as a control. As shown in Figure A6-1C, VLPs can be observed from 293T cells transiently transfected with the pEYFP-N3 HTLV-1 Gag construct (Fig. A5-1C, left panel). In comparison to HTLV-1 produced from MT-2 cells (Fig. A6-1C, right panel), the VLPs produced from the fusion construct resemble immature particles. In particular, the intense electron density along the lipid bilayer of VLPs likely represents the accumulation of Gag-EYFP (Fig. A6-1C, left inset) in contrast to the mature viral cores observed with HTLV-1 particles from MT-2 cells (Fig. A6-1C, right inset).

We also examined the cellular localization of the Gag-EYFP compared to Gag produced from a HTLV-1 molecular clone. The pEYFP-N3 HTLV-1 Gag construct was transiently transfected into HeLa cells, and 36 h post transfection, cells were fixed and analyzed by confocal microscopy (Fig. A6-2A, B). Comparable punctuate localization of Gag was observed for both the Gag-EYFP and the Gag expressing from the full-length molecular clone. Our observations suggest that Gag-EYFP expression in cells results in an intracellular localization pattern like that of Gag produced from a HTLV-1 molecular clone. In total, our findings provide evidence this construct results in the robust expression of HTLV-1 Gag as well as the highly efficient production of HTLV-1-like particles.

Analysis of HTLV-1-like particle morphology by cryo-TEM. To further characterize the VLPs produced from the HTLV-1 Gag-EYFP expression construct, we examined the VLP morphology by cryo-TEM. Supernatants from

293T cells transiently-cotransfected with the HTLV-1 Gag-EYFP expression construct and a VSV-G construct were harvested, concentrated, and then subjected to a 10-40% linear sucrose gradient. The resulting VLPs were then used in cryo-TEM. As shown in Fig. A5-3A, the majority of the resulting VLPs were found to be spherical, with less than 20% of the population showing an elongated morphology. Another example of the particles we observed in our study is shown in Supplemental Figure A6-1. Interestingly, VLPs produced in the absence of an envelope protein resulted in VLPs with irregular shapes, suggesting that the envelope protein helped to stabilize the VLP membrane (data not shown). We used the cryo-TEM images to next measure the diameter of the VLPs, where the average diameter was based on two measurements (as illustrated in Fig. A5-3B), with a total of 1734 particles examined. Similar to other retroviruses, there was a range of particle size. For completeness, we counted all particles that were spherical in shape that appeared to have an electron dense interior. Using these criteria, a total of 1734 particles were examined, ranging from 30 to 237 nm. While the overall range of particles observed was quite wide, the smallest (i.e., under 40 nm) and largest (i.e., over 170 nm) particles represented less than 1% of the total number of particles observed, and their inclusion had little impact on the mean particle size (i.e., 71 +/- 20 nm versus 72 nm +/- 18). We observed that over 25% of the total population was in the 70-80 nm range, with a mean particle size of 71 +/- 20 nm.

Analysis of VLP radial profile. We next used the information obtained by cryo-TEM to examine the VLP radial profile. For the majority of VLPs, cryo-TEM revealed that the inner Gag structure was indistinguishable (Fig. A5-3A). The partially ordered Gag lattice can be observed (data not shown), although the structure is less obvious compared to that reported for HIV-1 immature particles (13). Furthermore, the inner density appears to vary among VLPs, with some exhibiting homogenous inner density, while others seem to have an uneven distribution of electron densities attributable to Gag (Fig. A5-3A arrow).

To further analyze the electron density of VLPs, we investigated the radial density profile of VLPs. First, the average radial density profile was determined for several particles whose diameters ranged between 70-80 nm. As shown in Figure A6-4, the average distance between the highest density peaks of inner and outer leaflets of viral membrane with MA domain is approximately 30-Å. The MA domain is indistinguishable from the inner layer of membrane. The electron density profile approaching the center of the particle is relatively flat, suggesting a homogenized inner density. Our observations indicate that the HTLV-1-like particles are quite distinct from those produced from HIV-1 Gag.

FFS measurement of VLP size and Gag copy number. FFS provides information about the size of a particle through the autocorrelation function and the brightness and concentration of the particles through the photon counting histogram (PCH). Recent advances have expanded this technique to allow for the examination of protein oligomerization of larger complexes, including our recent analysis of HIV-1 particles (21). In the current experiments, we performed measurements on the same cell culture supernatant from 293T cells transiently transfected with HTLV-1 Gag-EYFP and VSV-G expression constructs. The supernatant from these cells was clarified by a low-speed centrifugation to eliminate large cell debris, and then directly used for FFS analysis. Figure A6-5A shows a representative fluorescence intensity trace of a FFS experiment performed on the cell culture supernatant. The discrete fluorescence intensity spikes are produced by VLPs passing through the observation volume. This raw data was analyzed by fluorescence correlation spectroscopy to determine the average particle size from the autocorrelation function (Fig. A5-5B). A fit to a single species diffusion model accurately describes the correlation function and identifies a diffusion time of 5.2 ms. This diffusion time corresponds to an average hydrodynamic diameter of 74 nm as determined by the Stokes Einstein relation. Repeating the measurement ($n = 5$) on independently prepared samples resulted in a mean diameter of 75 ± 4 nm for the VLPs.

The same raw data was analyzed with PCH analysis to determine the average copy number and concentration of VLP samples. A model assuming a single VLP brightness species leads to poor fits of the experimental PCH data (reduced $\chi^2 \geq 10$). Including a second VLP brightness species into the fit model was required to reproduce the experimental data. A fit of the photon counting histogram to a 2-species model (reduced $\chi^2 = 1.5$) is shown in Fig. A5-5B. The presence of two brightness species indicates brightness heterogeneity in the VLP sample. In other words, the VLP particles passing through the laser excitation volume are not all of equal brightness, which gives rise to the additional brightness species. Each species i is characterized by its normalized brightness b_i and average particle number N_i in the observation volume. Note that the normalized brightness is the same as the Gag copy number of the VLP. It is illustrative to briefly ignore the brightness heterogeneity by calculating the average Gag copy number b_{avg} of the VLP sample according to (22).

Based on measurements of several HTLV-1-like particle samples ($n = 5$) we determined an average Gag copy number per VLP of 510 ± 50 (Figure A6-6). To put this number into perspective recall that a copy number of ~ 5000 Gag is required to completely fill the surface of a 140 nm HIV-1 VLP (5). Thus, a maximum Gag copy number of ~ 1300 is expected for the smaller (~ 73 nm) HTLV-1 VLP assuming that both Gag proteins occupy a comparable surface area at the membrane. The observation of an average Gag copy number of 510 indicates that, on average, Gag at the membrane only covers about half of the available surface area.

The average Gag copy number was determined from the two brightness species identified by PCH analysis. Repeated measurements of multiple independent sample preparations confirmed the presence of the two species. Their brightness values, which typically varied very little across experiments, correspond to Gag copy numbers of $b_1 = 300 \pm 60$ and $b_2 = 880 \pm 100$ (Figure A6-6). The concentrations N_1 and N_2 varied from sample to sample, reflecting that total VLP production was dependent on sample-dependent factors, such as

the initial cell density. However, the population fraction $f_2 = N_2 / (N_1 + N_2)$ remained approximately constant for all measured samples, $f_2 = 19 \pm 7\%$. Thus, a population of ~19% of the VLPs is associated with the higher Gag copy number. Note that a similar heterogeneity in Gag copy numbers has also been reported for HIV-1 VLPs (21).

Discussion

Recent progress in cryo-TEM, cryo-ET and 3D reconstruction has led to many major breakthroughs in our understanding of virus structure. For instance, the architecture of immature and mature HIV-1 (13, 23, 24), murine leukemia virus (MuLV) (25), and Rous sarcoma virus (RSV) (26) have been investigated in great detail. Although HTLV-1 was the first human retrovirus to be discovered (27, 28), very little is known about HTLV-1 morphology. Progress in this area of HTLV-1 biology has been hampered due to the fragile nature of HTLV-1 proviral sequences as well as limited levels of viral gene expression and viral replication in tissue culture. HTLV-1 pathogenesis is typically observed decades after infection with low viral loads. In fact, studies have shown that HTLV-1 restricts its own gene expression via viral regulatory factors (29, 30). HTLV-1 has likely evolved such a replication strategy for immune escape. Furthermore, high AU-content of the retroviral genome may lead to instability during nuclear transport of mRNAs (31), which also contributes to the overall low level of viral gene and protein expression. In this study, we have designed a model system to study HTLV-1 Gag trafficking in cells and VLP production and morphology. The basis for this model system is a codon-optimized HTLV-1 Gag-EYFP construct, which can be readily amplified as a plasmid, expresses high levels of HTLV-1 Gag in mammalian cells, and robustly produces VLPs. This is the first model system developed for HTLV-1 for the study of virus particle assembly, release, as well as virus particle morphology.

While our model system does not express Gag in the context of a proviral sequence (i.e., codon-optimized and EYFP-tagged), our results indicate that the

VLPs produced have the morphology of the authentic HTLV-1 immature particles. Furthermore, while the Gag trafficking pathways used by the HTLV-1 Gag in this model system may be different from that of Gag expressed from the provirus, the production of VLPs argues that the trafficking pathways are biological relevant since VLP production is the result of expression of the codon-optimized Gag-EYFP fusion. The altered Gag trafficking pathways could influence envelope incorporation into VLPs, though our cryo-TEM data revealed an abundance of VLPs with VSVG. The VLPs characterized in our study resemble immature HTLV-1 and can be readily observed in ultrathin sections of 293Ts transfected with pEYFP-N3 HTLV-1 Gag (Fig. A5-1C). In addition, cell culture supernatants from 293Ts transiently transfected with pEYFP-N3 HTLV-1 Gag contain high levels of Gag-EYFP fusion proteins (Fig. A5-1B), which provides second line of evidence for the production of VLPs. In the fraction of sucrose gradients containing the highly-fluorescent material, cryo-TEM reveals that these fractions are highly concentrated with VLPs (Fig. A5-3A). Expression of EYFP alone in cells did not lead to the release of fluorescence in the cell culture supernatant (data not shown), arguing that we were specifically detecting the Gag-EYFP fusion in the VLPs.

We found that the intracellular localization of HTLV-1 Gag-EYFP was comparable to that of authentic Gag in HeLa cells (Fig. A5-2). This implies, though does not formally prove, that there are similarities in the Gag trafficking pathway used by Gag-EYFP and authentic Gag. Among retroviruses, intracellular Gag polyproteins are thought to target and accumulate at membrane compartments prior to viral assembly. In the case of HIV-1, Gag is thought to primarily target specific domains of the plasma membrane where PI(4,5)P2 and cholesterol are enriched, though endosomal trafficking may also play a role. For HTLV-1, several studies have suggested the association of Gag with several markers found on the membranes of late endosomes and multivesicular bodies – these markers are also enriched at the plasma membrane (32-35).

Our cryo-TEM and FFS analysis determined that the average VLP diameter was 71 ± 20 nm and 75 ± 4 nm, respectively. As observed in other retroviruses, the size of HTLV-1-like particles varies greatly, ranging from 30 to 237 nm. According to the size distribution (Fig. A5-3B), over 25% of the population is between 70-80 nm in diameter, indicating that HTLV-1 is smaller, on average, than previously believed. The average diameter of HTLV-1 has been shown to be anywhere from 95.1 ± 19.0 nm to 110.0 ± 15.5 nm depending on different types of staining used for TEM (19). However, morphological details are lost with staining methods when the biological specimens are completely dehydrated. Examining frozen, hydrated samples via cryo-TEM reflects the native morphology of the viral particles. Moreover, FFS offers a unique way to determine the average hydrodynamic radius in the cell supernatant without any special treatment or preparation prior to FFS analysis. The use of two independent methods for determining VLP diameter provides a strong argument in favor of the relatively small particle diameter for the HTLV-1-like particles analyzed in our study.

We used FFS to also investigate Gag stoichiometry in the VLPs by performing brightness analysis of the FFS data. We determined that the average Gag copy number per VLP is ~ 510 , which implies that only half of the available membrane surface is covered by Gag. Brightness analysis further revealed heterogeneity of the Gag copy number by identifying two brightness species. The presence of heterogeneity in the Gag copy number has also been observed for HIV-1 Gag-based VLPs (21). Since FFS analysis involves an ensemble average over all measured VLPs, the information in the PCH curve only provides a rough approximation of the true Gag copy number distribution for the VLPs. Thus, the two brightness species identified by PCH analysis do not necessarily reflect two distinct populations of VLPs, but more likely reflect the analytical approximation of a broad distribution of Gag stoichiometries that approximately range from 300 to 880. PCH analysis also demonstrates that only $\sim 20\%$ of VLPs have high copy numbers.

Among the thousands of cryo-TEM images of VLPs examined in our study, we commonly observed particles that did not have electron density consistent with a Gag shell covering the entire membrane surface (Fig. A5-3A). These results suggest that the majority of HTLV-1 particles analyzed contain an incomplete shell of Gag lattice. In the case of HIV-1, previous 3D structural analyses revealed that most immature virions contain a continuous, but incomplete, hexameric arranged Gag shell, covering approximately 40-60% of the membrane surface (14, 15, 36). The average copy number of Gag per particle was calculated to be approximately $2,400 \pm 700$ per immature particle. The Gag number increased significantly, however, when defects were introduced during budding (36). In fact, the data is in agreement with our previous FFS study indicating that HIV-1 Gag stoichiometry ranges from 750 to 2,500 (21). Since the mature core consists of only 1,000-1,500 molecules of CA (23), it is reasonable to believe that an equivalent number of Gag molecules are needed to form an immature particle. Our current study is the first to provide insights into the structural details for HTLV-1.

In vitro studies suggest that the HTLV-1 Gag shell is very likely to be organized differently compared to that of HIV-1 Gag (16-18). When examining the cryo-TEM images of HTLV-1-like particles, we rarely observed a highly ordered Gag lattice next to the lipid bilayer (Fig. A5-3A), a feature frequently observed in immature HIV-1 particles. The HTLV-1 particles analyzed in our study were fairly uniform in their overall inner density. Furthermore, in contrast to HIV-1, no defined peaks representing the CA or NC domains were found in the HTLV-1 radial density profile. The two peaks representing the lipid bilayers could be clearly determined (Fig. A5-4), whereas the inner density profile appeared to be relatively flat. Since cryo-TEM images represent a two-dimensional projection of the virus particle, a more rigorous structural analysis, such as cryo-ET, is needed to further examine the protein organization in the HTLV-1-like particles.

In summary, we have developed the first efficient and robust model system for the analysis of HTLV-1 Gag cellular trafficking, virus particle

assembly, release and particle morphology. This system will allow for significant advancements in understanding of the basic mechanisms of HTLV-1 replication – which has been severely hampered due to the limitations in studying HTLV-1 in tissue culture. Our study also represents the first description of immature HTLV-1 particles as well as quantitative measurements of particle size, Gag copy number, and an initial analysis of the HTLV-1 Gag lattice. Future application of cryo-electron tomography will aid in gaining greater insight into HTLV-1 particle morphology. A deeper understanding of the basic mechanisms involved in HTLV-1 particle assembly and morphology should help to enhance our global understanding of the basis of HTLV-1 particle infectivity, transmission and pathogenesis.

Materials and Methods

Construction of codon-optimized HTLV-1 gag-yfp fusion. A codon-optimized HTLV-1 Gag gene was designed using the UpGene program (37) and synthesized by GenScript Co. (Piscataway, NJ). The synthetic HTLV-1 *gag* contains an optimal Kozak consensus sequence (38, 39) at the 5' end of the gene: GCCACCATGG (start codon underlined). Two restriction enzyme sites, Hind III and Bam HI, were also engineered into the 5' and 3' end of the gene, respectively, for sub-cloning purposes. For reporter gene construction, the artificial HTLV-1 *gag* was cloned into a pEYFP-N3 vector using the HindIII and BamHI restriction sites, creating pEYFP-N3 HTLV-1 Gag.

Immunoblotting. 293T cells were transiently transfected with the pEYFP-N3 HTLV-1 Gag construct using GenJet (SignaGen, Gaithersburg, MD) according to the manufacturer's instructions. Thirty-six hours post-transfection, cell pellets and supernatant were collected and lysates were prepared as previously described (40). Lysates were subjected to electrophoresis on 12.5% polyacrylamide gels and transferred to nitrocellulose (Bio-Rad, Hercules, CA). HTLV-1 Gag polyprotein was detected with a primary mouse anti-HTLV-1 p24 antiserum (Abcam, Cambridge, MA) at 1:1500 dilution followed by a horseradish

peroxidase-conjugated goat anti-mouse IgG (Thermo Fisher, Rockford, IL) at 1:5000 dilution. Gag polyprotein expression was detected with a ChemiDoc XRS system (Bio-Rad).

Immunofluorescence and fluorescence microscopy. HeLa cells were grown on Lab-Tek II chamber slides (Fisher Scientific) and transfected with either the pEYFP-N3 HTLV-1 Gag construct or a HTLV-1 proviral clone (a kind gift from Dr. Marie-Christine Dokhelar) (41). Thirty-six hours post-transfection, cells were washed twice with 1x PBS buffer and fixed with 4% paraformaldehyde for 20 min. For cells transfected with pEYFP-N3 HTLV-1, cells were washed three times after fixation, and stained for 5 min with 1 μ g/ml DAPI (Sigma-Aldrich, St. Louis, MO) in 1x PBS containing 0.05% Triton X-100 (Sigma-Aldrich), then preserved using ProLong Gold antifade mounting reagent (Invitrogen, Carlsbad, CA). For cells transfected with the HTLV-1 proviral clone, permeabilization was achieved by treating with 1x PBS containing 0.5% Triton X-100 for 2 min at room temperature following fixation. Cells were then washed three times and blocked with 1x PBS containing 5% normal donkey serum (Sigma-Aldrich) for 30 min. Primary mouse anti-HTLV-1 p24 antisera (Abcam) were diluted (1:150) in blocking solution and incubated with cells. After incubation for 2 hr at room temperature, cells were washed three times, followed by a second incubation for 1 hr at room temperature with diluted (1:250) Alexa Fluor 488-conjugated donkey anti-mouse IgG (Invitrogen). Prior to mounting, cells were washed five times and stained with DAPI as described above. Intracellular localization of Gag polyprotein was detected with an Olympus FV500 confocal laser scanning microscope. Optical sections of cells were collected with a Plan-Apo 60x /1.45 NA TIRFM objective at 1.5 zoom. The z-series were reconstructed using Olympus FluoView software.

VLPs purification for cryo-TEM. 293T cells were co-transfected with pEYFP-N3 HTLV-1 and a vesicular stomatitis virus G (VSV-G) protein (10:1) expression construct using GeneJet. Twenty-four hours post-transfection, the cell culture media was changed to a serum-free media and incubated for an

additional 12 hr. In order to harvest VLPs, tissue culture supernatant was centrifuged at 3000 x g for 5 min to remove large cellular debris, then the supernatant was passed through an Amicon Ultra- 15 Centrifugal Filter Unit (100 KDa) (Millipore, Billerica, MA) to concentrate samples. The concentrated samples were then subjected to a 10-40% linear sucrose gradient prepared with a Gradient Master (BioComp, Fredericton, NB, Canada). Samples were then ultracentrifuged at 35,000 rpm for 30 min at 4 °C using a SW55 Ti rotor. The VLP fraction was extracted and pelleted at 35,000 rpm, 4 °C for 1.5 hr using a SW55 Ti rotor (Beckman). After centrifugation, the pellet was resuspended in 1X STE buffer (10 mM Tris-Cl, pH 7.4, 100 mM NaCl, 1 mM EDTA) at 4 °C for 4 hr and then analyzed by cryo-TEM.

TEM of transfected cells. 293T cells were transfected with either pEYFP-N3 HTLV-1 or a HTLV-1 proviral clone as described above. Thirty-six hours post-transfection, cells were harvested and washed twice with 1x PBS followed by an additional wash in 0.1M sodium cacodylate. To prepare thin sections, cell pellets were first fixed with 2.5% glutaraldehyde for 40 min and then washed three times with 0.1 M sodium cacodylate. After washing, the samples were post-fixed with 1% OsO₄ for 30 min, followed by three rinses. The samples were then subjected to increasing concentrations of ethanol for dehydration. Immediately following the application of 70% ethanol, en bloc staining was added to the samples for 30 min before embedding in Epon 812 resin. Ultrathin sections (65 nm) were acquired and stained with uranyl acetate and lead citrate, then examined by electron microscopy using a JEOL 1200EX transmission electron microscope.

Cryo-TEM of HTLV-1 VLPs and calculation of radial profile. A 3ul aliquot of the purified and concentrated HTLV-1 VLP sample preparation was applied to a glow-discharged c-flat holey carbon grid (Ted Pella, Redding, CA) and used for plunge freezing into liquid ethane (42) with a FEI Vitrobot MarkIII system. The frozen grids were then transferred to a FEI TF30 field emission gun transmission electron microscope at liquid nitrogen temperature. Images were recorded at a

magnification of 40 k to 100 k at low-dose ($\sim 30 \text{ e} / \text{\AA}^2$) and 1 to 5 μm underfocus conditions using a Gatan 4 k by 4 k CCD camera.

In order to calculate the radial density profile, images of VLPs with spherical morphology were boxed using RobEM (<http://cryoem.ucsd.edu/programDocs/runRobem.txt>). The center of each particle was determined using the program EMCORORG that calculates cross-correlation of each image with its 180° rotational image (<http://cryoem.ucsd.edu/programDocs>). The radial profile for each particle was then calculated by computing the rotationally averaged density relative to the center of the particle. A group of 14 VLP images with a diameter in the range of 70-80 nm was used for calculation of the averaged radial profile. Each pixel in the image corresponds to a 3.0- \AA spacing in the VLP. The defocus levels of these images were between 1.5-3.7 μm , which allows for visualization of both membrane leaflets of the viral membrane. The radial profile of each particle was first calculated to obtain the highest density position of the outer membrane leaflet. The radial profile of each particle was then linearly interpreted to match the position of the outer membrane to the averaged position (367- \AA radius). The average radial profile and standard deviation were then calculated.

VLP size measurements. Cryo-TEM images were analyzed using ImageJ software (NIH, Bethesda, MD). For each VLP, two perpendicular diameters were used to calculate the average diameter. The histogram was generated using GraphPad Prism 5 software (GraphPad, La Jolla, CA).

VLP preparation and FFS experimental setup. 293T cells were co-transfected with pEYFP-N3 HTLV-1 and a VSV-G expression construct (10:1) as described earlier. Aliquots of the cell culture supernatants used for subsequent cryo-TEM analysis were removed for parallel analysis by FFS. Thirty-six hours post-transfection, VLPs were harvested and clarified of cellular debris by low-speed centrifugation at 3000 x g for 5 min as well as passing through a 0.22 μm filter. The resulting clarified supernatants were then used for FFS measurements.

A mode-locked Ti:sapphire laser (Tsunami, Spectra-Physics, Mountain View, CA) pumped by an intracavity doubled Nd:YVO₄ laser (Millenia, Spectra Physics) is the source of two-photon excitation. Experiments were performed on a modified Zeiss Axiovert 200 microscope (Thornwood, NY) as previously described (22). Each FFS measurement collects data at a sampling frequency of 20 kHz for a duration of 20-30 min at an excitation wavelength of 905 nm. The viral particles are measured using a 63x C-Apochromat water immersion objective (N.A.=1.2). The excitation power at the objective ranged from 0.1-0.4 mW. A volume of 200 μ l of VLP solution was added to an 8-well Nunc Lab-Tek Chamber Slide mounted on the microscope. To avoid evaporation unused wells were filled with water and the slide was closed with a lid. Measurements were taken 10 μ m above the bottom of the well.

FFS data analysis. A brief description of the analysis method is provided here. A detailed discussion of FFS analysis of VLP samples can be found elsewhere (21). The diffusion time was determined by fitting the calculated autocorrelation function to a single species diffusion model (43). The ratio of diffusion time for the two samples is equated, according to the Stokes-Einstein relation, to the ratio of the hydrodynamic radii of the diffusing particles $\tau_{D1}/\tau_{D2}=r_1/r_2$. The measured diffusion time of fluorescent spheres with a known radius of 50 nm serves as a reference to calculate the average diameter of the VLPs (44). The FFS data was also fit to a 3-species PCH model with deadtime and afterpulsing corrections (45). Each independent species in PCH is defined by its brightness ϵ and the average number N of particles in the optical observation volume. The particle number N is converted into a concentration after the observation volume is calibrated with a dye sample. One of the three species is required to take the auto-fluorescent background of the solution into account as discussed in a recent paper on HIV-1 VLPs (21). This background species, which has a vanishingly small brightness, is included in every PCH fit and will not be reported. The other two species of the PCH model define the VLP sample. The presence of two brightness species indicates the existence of

Gag copy number heterogeneity within the HTLV-1 VLP population as previously observed for HIV-1 VLPs (21). PCH fitting was carried out by programs written in IDL 6.4 (Research Systems, Boulder CO). Error analysis of FFS data was carried out as previously described (46).

FFS brightness calibration and experimental considerations. To avoid unwanted optical effects, all experiments are conducted in a power range where the fluorescence intensity of YFP scales quadratically with excitation power. We also confirmed that within this power range the average occupation number N remains constant, which establishes a constant optical observation volume. The brightness of a protein complex scales with the number of YFP-labels it contains (21). The YFP copy number of a complex is determined by the normalized brightness $b = \epsilon / \epsilon_{\text{YFP}}$, where ϵ_{YFP} is the brightness of the YFP monomer and ϵ is the brightness of the complex. A calibration measurement of YFP brightness, ϵ_{YFP} , is necessary to determine copy number. Because YFP brightness is difficult to determine at the low powers that the VLPs must be measured at to avoid saturation of the detector, the YFP brightness, ϵ_{high} , is measured at a higher excitation power, P_{high} . Conversion to the YFP brightness ϵ_{low} for the low power P_{low} of VLP experiments is achieved by using the relationship of power to brightness $\epsilon_{\text{low}} = \epsilon_{\text{high}} (P_{\text{low}}^2 / P_{\text{high}}^2)$. The data acquisition time for the VLP measurements was chosen such that at least 1000 VLPs passed through the observation volume, which is sufficient for statistical analysis of the data. All VLP measurements were performed at excitation powers that are free from saturation and bleaching artifacts (21). The FFS experiments identified two brightness species for the VLP sample. The average normalized brightness b_{avg} of the two species is determined by a non-linear relationship (22),

$$b_{\text{avg}} = \frac{b_1^2 N_1 + b_2^2 N_2}{b_1 N_1 + b_2 N_2},$$

where b_i and N_i are the normalized brightness and the number of particles in the observation volume of each species.

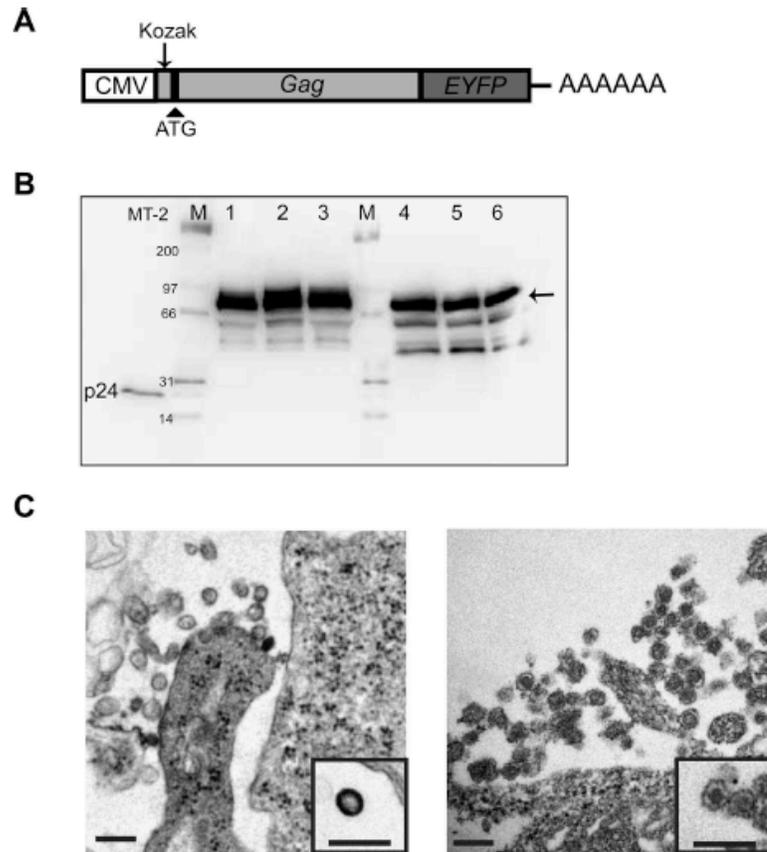


Figure A6-1. Development of a model system for the efficient expression of HTLV-1 Gag and robust production of VLPs. (A) HTLV-1 Gag expression construct. The HTLV-1 Gag gene was codon-optimized with the insertion of a Kozak consensus sequence (arrow) upstream of the ATG start codon (arrowhead). The *EYFP* gene was inserted in-frame prior to the Gag gene stop codon. The CMV promoter and 3'-end poly A are indicated. (B) Immunoblot analysis of HTLV-1 Gag. An anti-HTLV-1 p24 monoclonal was used to detect HTLV-1 Gag-EYFP (arrow). Cell culture supernatants were collected from MT-2 cells was used as a positive control. Lane 1-3 are cell culture supernatants from three independent experiments in which pEYFP-N3-HTLV-1 Gag was transiently transfected into 293T cells; lane 4-6 are the cellular lysates. Lane "M", molecular markers. (C) Transmission electron microscopy of VLPs. Left panel, VLPs produced from 293T cells transiently transfected with pEYFP-N3-HTLV-1 Gag; right panel are HTLV-1 particles from MT-2 cells. Scale bar = 200 nm.

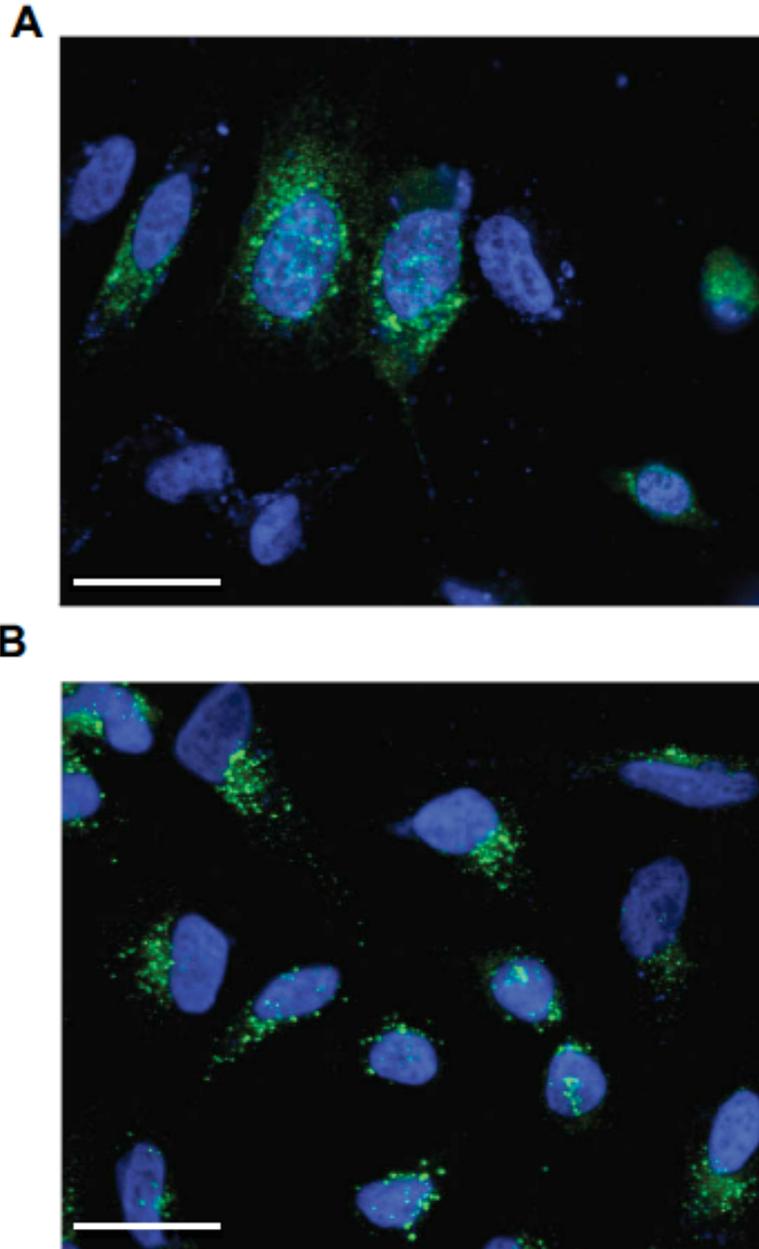


Figure A6-2. Cellular localization of HTLV-1 Gag-EYFP and HTLV-1 Gag. HeLa cells were transiently transfected with pEYFP-N3-HTLV-1 Gag (A) or a HTLV-1 molecular clone (B). The locations of nuclei were identified by DAPI staining (blue), HTLV-1 Gag (green). Scale bars= 28 μ m.

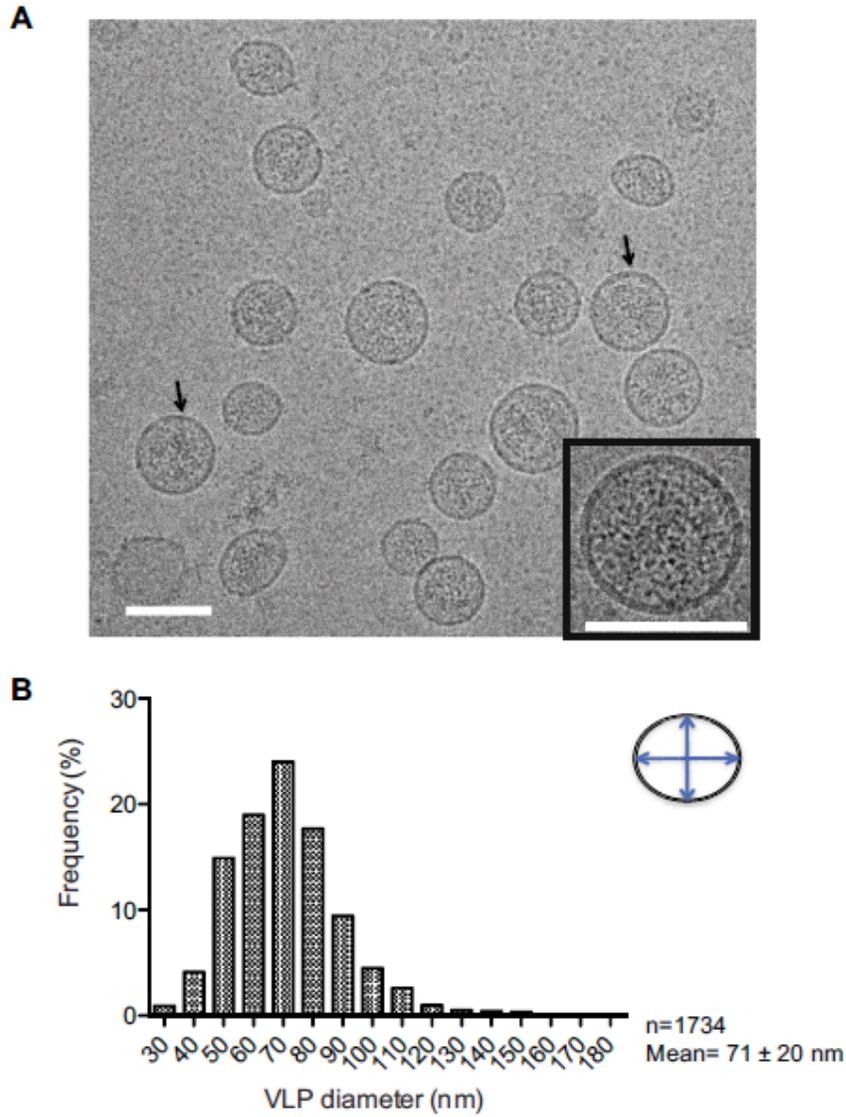


Figure A6-3. Cryo-TEM analysis of HTLV-1 Gag-based VLPs. (A) Cryo-TEM images of VLPs produced from 293T cells. Examples of VLPs that have partially occupied inner electron density are indicated with arrows. The inset shows a magnified view of a representative VLP. Scale bars = 100 nm. (B) Distribution of VLP diameter. Particle diameter was determined by averaging the longest and shortest measurements as indicated in the diagram at the top right corner using the ImageJ software. A total of 1734 VLPs were examined (mean = 71 +/- 20 nm).

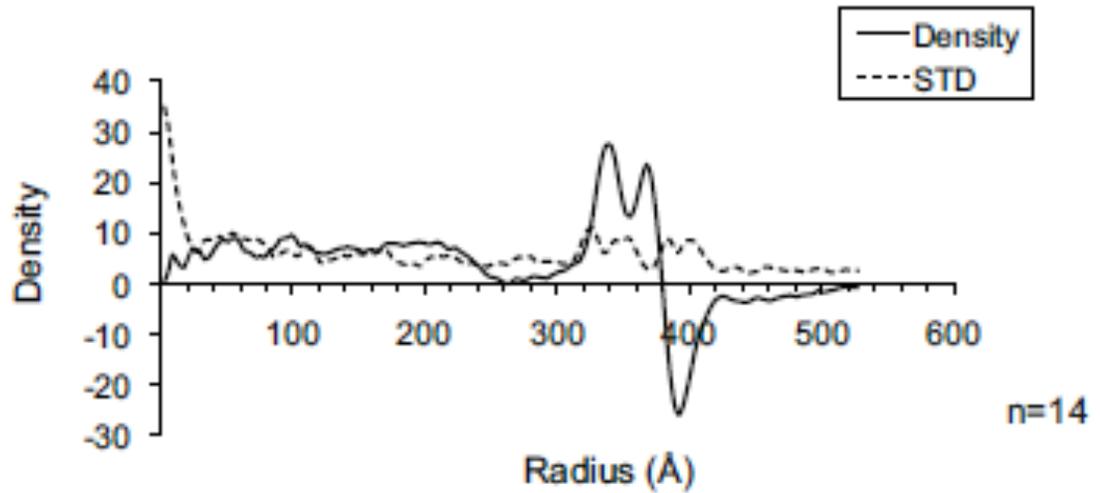


Figure A6-4. Radial density profile of the HTVL-1-like particles. The solid line represents the average density measured; dashed line indicates the standard deviation (n=14).

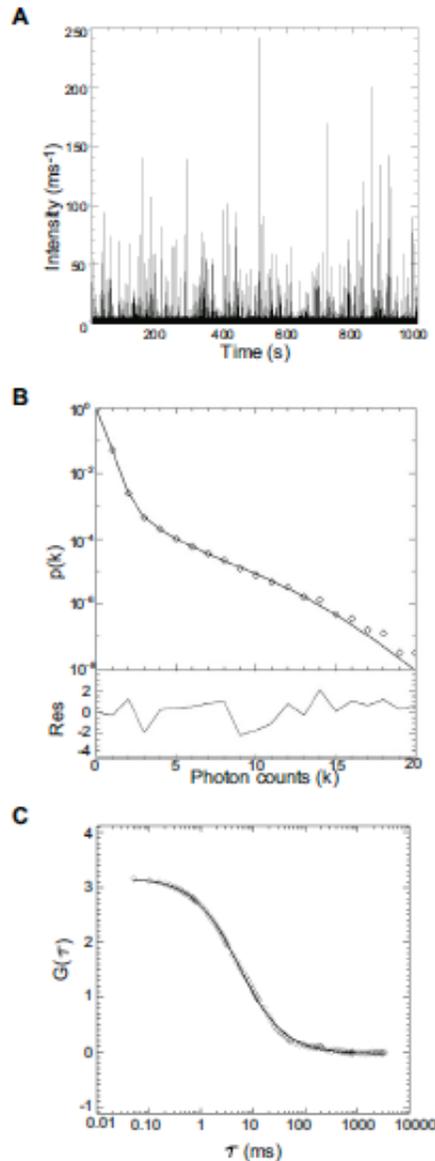


Figure A6-5. Fluorescence fluctuation spectroscopy analysis of HTLV-1 Gag-based VLPs. (A) The fluorescence intensity trace shows discrete peaks, which correspond to individual VLPs diffusing through the observation volume. (B) Experimental photon counting histogram (diamonds) of the VLP sample. A fit (solid line) of the histogram to a 2-species model with background identifies the concentration and Gag copy number of the VLPs. The presence of two species indicates the existence of heterogeneity in the Gag copy number of VLPs. A weighted average of the two species leads to an average Gag copy number of 530 per VLP. The first VLP species has a copy number of 270 and a

concentration of 20.5 pM. The second VLP species, which is brighter than the first, has a copy number of 800 and a concentration of 6.5 pM. (C) A fit (solid line) of the autocorrelation function (diamonds) to a diffusion model identifies an average hydrodynamic diameter of 74 nm for the VLPs.

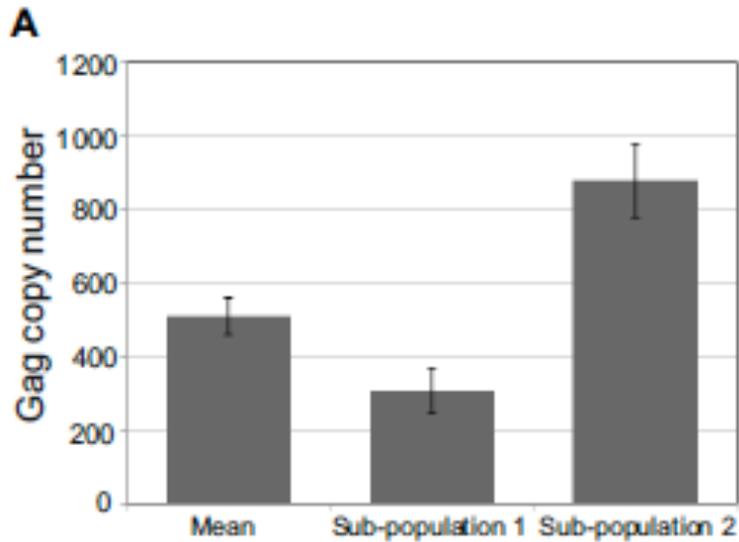


Figure A6-6. Gag copy number of HTLV-1 Gag based VLPs. The Gag copy number was determined by FFS analysis of several independent VLP samples ($n = 5$). The mean copy number per VLP is shown together with the corresponding copy number of the two subpopulations identified by FFS analysis. The error bars represent the standard deviation of the multiple measurements.

References

1. **Proietti FA, Carneiro-Proietti AB, Catalan-Soares BC, Murphy EL.** 2005. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* **24**:6058-6068.
2. **Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, de The G.** 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* **2**:407-410.
3. **Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M.** 1986. HTLV-I associated myelopathy, a new clinical entity. *Lancet* **1**:1031-1032.
4. **Gheysen D, Jacobs E, de Foresta F, Thiriart C, Francotte M, Thines D, De Wilde M.** 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* **59**:103-112.
5. **Briggs JA, Simon MN, Gross I, Krausslich HG, Fuller SD, Vogt VM, Johnson MC.** 2004. The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* **11**:672-675.
6. **Lanman J, Lam TT, Emmett MR, Marshall AG, Sakalian M, Prevelige PE, Jr.** 2004. Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat Struct Mol Biol* **11**:676-677.
7. **Benjamin J, Ganser-Pornillos BK, Tivol WF, Sundquist WI, Jensen GJ.** 2005. Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. *J Mol Biol* **346**:577-588.
8. **Ono A, Orenstein JM, Freed EO.** 2000. Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. *J Virol* **74**:2855-2866.
9. **Ono A, Ablan SD, Lockett SJ, Nagashima K, Freed EO.** 2004. Phosphatidylinositol (4,5) biphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc Natl Acad Sci U S A* **101**:14889-14894.
10. **Bryant M, Ratner L.** 1990. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc Natl Acad Sci U S A* **87**:523-527.
11. **Riffel N, Harlos K, Iourin O, Rao Z, Kingsman A, Stuart D, Fry E.** 2002. Atomic resolution structure of Moloney murine leukemia virus matrix protein and its relationship to other retroviral matrix proteins. *Structure* **10**:1627-1636.
12. **Usami Y, Popov S, Popova E, Inoue M, Weissenhorn W, H GG.** 2009. The ESCRT pathway and HIV-1 budding. *Biochem Soc Trans* **37**:181-184.
13. **Fuller SD, Wilk T, Gowen BE, Krausslich HG, Vogt VM.** 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr Biol* **7**:729-738.
14. **Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, Krausslich HG.** 2009. Structure and assembly of immature HIV. *Proc Natl Acad Sci U S A* **106**:11090-11095.

15. **Wright ER, Schooler JB, Ding HJ, Kieffer C, Fillmore C, Sundquist WI, Jensen GJ.** 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *Embo J* **26**:2218-2226.
16. **Bertola F, Manigand C, Picard P, Belghazi M, Precigoux G.** 2000. Human T-lymphotrophic virus type I nucleocapsid protein NCp15: structural study and stability of the N-terminal zinc-finger. *Biochem J* **352 Pt 2**:293-300.
17. **Christensen AM, Massiah MA, Turner BG, Sundquist WI, Summers MF.** 1996. Three-dimensional structure of the HTLV-II matrix protein and comparative analysis of matrix proteins from the different classes of pathogenic human retroviruses. *J Mol Biol* **264**:1117-1131.
18. **Cornilescu CC, Bouamr F, Yao X, Carter C, Tjandra N.** 2001. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. *J Mol Biol* **306**:783-797.
19. **Doultree JC, Kiernan RE, Lee JY, Bowden DS, McPhee DA, Tokuyasu KT, Marshall JA.** 1992. A new electron microscope positive staining method for viruses in suspension. *J Virol Methods* **37**:321-335.
20. **Wills JW, Craven RC.** 1991. Form, function, and use of retroviral gag proteins. *AIDS* **5**:639-654.
21. **Chen Y, Wu B, Musier-Forsyth K, Mansky LM, Mueller JD.** 2009. Fluorescence fluctuation spectroscopy on viral-like particles reveals variable gag stoichiometry. *Biophys J* **96**:1961-1969.
22. **Chen Y, Wei LN, Muller JD.** 2003. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *Proc Natl Acad Sci U S A* **100**:15492-15497.
23. **Briggs JA, Wilk T, Welker R, Krausslich HG, Fuller SD.** 2003. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J* **22**:1707-1715.
24. **Briggs JA, Grunewald K, Glass B, Forster F, Krausslich HG, Fuller SD.** 2006. The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure* **14**:15-20.
25. **Nermut MV, Mulloy B.** 2007. Consideration of the three-dimensional structure of core shells (capsids) in spherical retroviruses. *Micron* **38**:462-470.
26. **Briggs JA, Johnson MC, Simon MN, Fuller SD, Vogt VM.** 2006. Cryo-electron microscopy reveals conserved and divergent features of gag packing in immature particles of Rous sarcoma virus and human immunodeficiency virus. *J Mol Biol* **355**:157-168.
27. **Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC.** 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A* **77**:7415-7419.

28. **Poiesz BJ, Ruscetti FW, Reitz MS, Kalyanaraman VS, Gallo RC.** 1981. Isolation of a new type C retrovirus (HTLV) in primary uncultured cells of a patient with Sezary T-cell leukaemia. *Nature* **294**:268-271.
29. **Gaudray G, Gachon F, Basbous J, Biard-Piechaczyk M, Devaux C, Mesnard JM.** 2002. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. *J Virol* **76**:12813-12822.
30. **Clerc I, Polakowski N, Andre-Arpin C, Cook P, Barbeau B, Mesnard JM, Lemasson I.** 2008. An interaction between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ. *J Biol Chem* **283**:23903-23913.
31. **Schwartz S, Felber BK, Pavlakis GN.** 1992. Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. *J Virol* **66**:150-159.
32. **Blot V, Perugi F, Gay B, Prevost MC, Briant L, Tangy F, Abriel H, Staub O, Dokhlar MC, Pique C.** 2004. Nedd4.1-mediated ubiquitination and subsequent recruitment of Tsg101 ensure HTLV-1 Gag trafficking towards the multivesicular body pathway prior to virus budding. *J Cell Sci* **117**:2357-2367.
33. **Dorweiler IJ, Ruone SJ, Wang H, Burry RW, Mansky LM.** 2006. Role of the human T-cell leukemia virus type 1 PTAP motif in Gag targeting and particle release. *J Virol* **80**:3634-3643.
34. **Mazurov D, Heidecker G, Derse D.** 2006. HTLV-1 Gag protein associates with CD82 tetraspanin microdomains at the plasma membrane. *Virology* **346**:194-204.
35. **Mazurov D, Heidecker G, Derse D.** 2007. The inner loop of tetraspanins CD82 and CD81 mediates interactions with human T cell lymphotropic virus type 1 Gag protein. *J Biol Chem* **282**:3896-3903.
36. **Carlson LA, Briggs JA, Glass B, Riches JD, Simon MN, Johnson MC, Muller B, Grunewald K, Krausslich HG.** 2008. Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **4**:592-599.
37. **Gao W, Rzewski A, Sun H, Robbins PD, Gambotto A.** 2004. UpGene: Application of a web-based DNA codon optimization algorithm. *Biotechnol Prog* **20**:443-448.
38. **Kozak M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res* **12**:857-872.
39. **Kozak M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**:8125-8148.
40. **Wang H, Norris KM, Mansky LM.** 2003. Involvement of the matrix and nucleocapsid domains of the bovine leukemia virus Gag polyprotein precursor in viral RNA packaging. *J Virol* **77**:9431-9438.

41. **Le Blanc I, Rosenberg AR, Dokhelar MC.** 1999. Multiple functions for the basic amino acids of the human T-cell leukemia virus type 1 matrix protein in viral transmission. *J Virol* **73**:1860-1867.
42. **Baker TS, Olson NH, Fuller SD.** 1999. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol Rev* **63**:862-922, table of contents.
43. **Thompson NL, Lieto AM, Allen NW.** 2002. Recent advances in fluorescence correlation spectroscopy. *Current Opinion in Structural Biology* **12**:634-641.
44. **Wu B, Chen Y, Muller JD.** 2008. Fluorescence correlation spectroscopy of finite-sized particles. *Biophys J* **94**:2800-2808.
45. **Hillesheim LN, Müller JD.** 2003. The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy with Non-Ideal Photodetectors. **85**:1948-1958.
46. **Müller JD, Chen Y, Gratton E.** 2000. Resolving Heterogeneity on the Single Molecular Level with the Photon-Counting Histogram. **78**:474-486.

APPENDIX VII
COPYRIGHT PERMISSIONS

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Dec 28, 2015

This is a License Agreement between Jonathan Rawson ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Jonathan Rawson
Customer address	Moos Tower 18-242 MINNEAPOLIS, MN 55455
License number	3777731415164
License date	Dec 28, 2015
Licensed content publisher	Elsevier
Licensed content publication	Bioorganic & Medicinal Chemistry
Licensed content title	5,6-Dihydro-5-aza-2'-deoxycytidine potentiates the anti-HIV-1 activity of ribonucleotide reductase inhibitors
Licensed content author	Jonathan M. Rawson, Richard H. Heineman, Lauren B. Beach, Jessica L. Martin, Erica K. Schnettler, Michael J. Dapp, Steven E. Patterson, Louis M. Mansky
Licensed content date	15 November 2013
Licensed content volume number	21
Licensed content issue	22

Number of pages	7
Start Page	7222
End Page	7228
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	STUDIES ON THE DETERMINANTS OF HIV MUTAGENESIS AND STRATEGIES FOR ITS ENHANCEMENT
Expected completion date	Dec 2015
Estimated size (number of pages)	390
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

Confirmation Number: 11517811
Order Date: 12/28/2015

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

Jonathan Rawson
rawso018@umn.edu
+1 (616)6345299
Payment Method: n/a

Order Details

The Journal of general virology

Order detail ID: 69390818
Order License Id: 3777750768216
ISSN: 0022-1317
Publication Type: Journal
Volume:
Issue:
Start page:
Publisher: SOCIETY FOR GENERAL
MICROBIOLOGY
Author/Editor: SOCIETY FOR GENERAL
MICROBIOLOGY ; FEDERATION OF
EUROPEAN MICROBIOLOGICAL
SOCIETIES

Permission Status:  **Granted**

Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

[Hide details](#)

Requestor type	Academic institution
Format	Print, Electronic
Portion	chapter/article
Title or numeric reference of the portion(s)	Entire article
Title of the article or chapter the portion is from	Novel inhibitors of human immunodeficiency virus type 2 infectivity.
Editor of portion(s)	N/A
Author of portion(s)	Lauren Beach
Volume of serial or monograph	N/A
Page range of portion	
Publication date of portion	01/12/2014
Rights for	Main product

Duration of use	Current edition and up to 5 years
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Made available in the following markets	Professional
The requesting person/organization	Jonathan Rawson
Order reference number	
Author/Editor	Jonathan Rawson
The standard identifier	JR
Title	STUDIES ON THE DETERMINANTS OF HIV MUTAGENESIS AND STRATEGIES FOR ITS ENHANCEMENT
Publisher	University of Minnesota
Expected publication date	Dec 2015
Estimated size (pages)	390

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. [More info](#)

\$ 0.00

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Dec 29, 2015

This is a License Agreement between Jonathan Rawson ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Jonathan Rawson
Customer address	Moos Tower 18-242 MINNEAPOLIS, MN 55455
License number	3778350428593
License date	Dec 28, 2015
Licensed content publisher	Elsevier
Licensed content publication	Biophysical Journal
Licensed content title	Characterization of Cytoplasmic Gag-Gag Interactions by Dual-Color Z-Scan Fluorescence Fluctuation Spectroscopy
Licensed content author	Keir H. Fogarty, Yan Chen, Iwen F. Grigsby, Patrick J. Macdonald, Elizabeth M. Smith, Jolene L. Johnson, Jonathan M. Rawson, Louis M. Mansky, Joachim D. Mueller
Licensed content date	16 March 2011
Licensed content volume number	100
Licensed content issue number	6

Number of pages	9
Start Page	1587
End Page	1595
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	STUDIES ON THE DETERMINANTS OF HIV MUTAGENESIS AND STRATEGIES FOR ITS ENHANCEMENT
Expected completion date	Dec 2015
Estimated size (number of pages)	
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD