INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

MODULATION OF HIV-1 GENETIC RECOMBINATION

by

Wenfeng An

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in The University of Michigan
2003

Doctoral Committee:

Associate Professor Alice Telesnitsky, Chair
Assistant Professor Kathleen L. Collins
Professor Wesley Dunnick
Professor Michael J. Imperiale
Assistant Professor John V. Moran

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3106005

UMI®

UMI Microform 3106005
Copyright 2003 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To my parents, my wife, and my daughter

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Dr. Alice Telesnitsky, for her academic guidance, support, and commitment to my success over the past four years. I greatly admire her scientific vision and creative thinking. Her rather unique, critical yet encouraging nature has greatly promoted my scientific development.

I would like to acknowledge the former and present members from the Telesnitsky lab: Nam Hee, Julie, Nicole and Sharon for their training and help; Jessica for her putting up with me in the same bay and her help with the RNA capture assay; Adewunmi for her extraordinary contribution to the recombination experiments shown in Chapter three; Jen, Crystal, Steve, Jennifer, Mallika, Kelly, Nathan and Sarah for their help with various aspects of the project.

I wish to express my appreciation to the wonderful members of my doctoral committee, Drs. Kathleen Collins, Wesley Dunnick, Michael Imperiale, John Moran and a former committee member, Dr. Erle Robertson, for their scientific guidance, constructive advice and enthusiastic support of my thesis work.

I am obligated to the remaining faculty and staff members at the Department of Microbiology and Immunology for their academic advice, technical and administrative assistance.

I wish to express my extreme gratefulness to my parents, whose sacrifice and expectation have been served as consistent inspiration, from my childhood to this stage. I am greatly indebted for their care to my whole family during their visit.

Finally, my deepest gratitude is to my wife, Ping Ye, for her invaluable love, support and understanding, and to my daughter, Serena Jiani An, for all the joys she has

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

brought into my life. I wish I could have spent more time with them while I was writing this thesis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PREFACE

Human immunodeficiency virus type 1 (HIV-1) is characterized by its extraordinary genetic diversity, which can be largely attributed to its error-prone DNA replication and high frequency of genetic recombination. The goal of this thesis is to gain a better understanding of the factors that modulate the frequency of HIV-1 recombination.

Genetic recombination is suggested to occur much more frequently for HIV-1 than for murine leukemia virus (MLV). To more precisely compare their recombination frequencies, a series of single replication cycle assays were performed for both viruses using highly similar vectors and assay conditions. Results from two-vector recombination assays and reciprocal repeat deletion experiments confirmed the reported differences in intermolecular recombination. However, two different one-vector approaches, including direct repeat deletion assays and transduction-type recombination assays, indicated that template switching properties were very similar for HIV-1 and MLV. These results provide genetic evidence that coexpressed RNAs are less accessible to the recombination machinery during MLV replication than for HIV-1.

Effects of altering genetic distance on the frequency of template switching were examined by using HIV-1 vectors containing degenerate repeated sequences, which provided a wide range of sequence variation similar to that of intra-subtype, inter-subtype, and inter-group HIV-1 strains. Results from these experiments indicated that the frequency of template switching decreased dramatically as sequence divergence between repeated sequences increased. For sequences that differed by about 25% or more, HIV-1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

template switching directed by sequence homology appeared to be no more frequent than that which is homology independent.

Retroviral transduction of cellular genes is frequently observed for oncoretroviruses. The transduction potential of HIV-1 was investigated with transduction-type recombination assays using vectors reminiscent of retroviral transduction intermediates. It was demonstrated that readthrough transcripts arose at a rate up to 7% of the normal viral RNA. RNA sequences downstream of the 3' long terminal repeat (LTR) acquired by polyadenylation signal readthrough could efficiently serve as templates for HIV-1 genetic recombination. The readthrough sequences downstream of the LTR appeared to be preferentially used for HIV-1 recombination when compared to homologous sequences located between the 5' and 3' LTRs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGMENTS iii

PREFACE v

LIST OF FIGURES ix

LIST OF TABLES xi

CHAPTER ONE 1

Introduction

HIV-1 genetic variation 1
HIV-1 mutation 4
HIV-1 genetic recombination 8
HIV-1 RNA dimerization and recombination 29
Overview of the thesis 40
Bibliography 42

CHAPTER TWO 60

Frequency Of Direct Repeat Deletion And Effects Of Varying Sequence Similarity On Frequency Of Direct Repeat Deletion During HIV-1 Reverse Transcription

Introduction 60
Materials and methods 63
Results 66
Discussion 82
Bibliography 87

CHAPTER THREE 93

HIV-1 Genetic Recombination Is More Frequent Than That Of MLV Despite Similar Template Switching Rates

Introduction 93
Materials and methods 95
Results 101

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Discussion 116
Bibliography 120

CHAPTER FOUR 125

High Frequency Of Transduction-Type Recombination During HIV-1 Replication

Introduction 125
Materials and methods 127
Results 130
Discussion 149
Bibliography 157

CHAPTER FIVE 164

Conclusions And Future Directions

Conclusions 164
Discussion and future directions 165
Bibliography 183

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Figure

1-1 Evolutionary relationships among non-recombinant HIV-1 strains from the HIV-1/SIVcpz lineage 2

1-2 Two vector recombination assays 11

1-3 Model for direct repeat deletion 15

1-4 Minus strand exchange model for genetic recombination 18

1-5 A model for retroviral transduction 21

1-6 RNA structures in the dimerization/packaging region near the 5' end of HIV-1 RNAs and kissing-loop model for dimerization 32

1-7 Nucleotide alignment of DIS sequences from complete HIV-1 genomes 33

2-1A HIV-1 vector system 67

2-1B Protocol for direct repeat deletion assay 69

2-2 lacZ inactivation and repeat deletion frequencies for HIV-1 and MLV 74

2-3 HIV-1 transducing vectors that contain sequence degeneracy within lacZ 77

2-4 Repeat deletion rates for vectors with identical lacZ variant repeats 79

2-5 Effects of varying extents of sequence identity on repeat deletion frequencies 81

3-1 Inter- and intramolecular recombination assay vectors 103

3-2 Two-vector (intermolecular) genetic recombination assay 104

3-3 RNase protection assay to quantify virion RNAs 106

3-4 Southern blot of pooled integrated MLV vector products 109

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3-5 One-vector transduction-style recombination assay 110

3-6 Reciprocal repeat recombination assay 114

4-1 HIV-1 vector reminiscent of transduction intermediate and protocol for determining template switch junctions 132

4-2 Summary of template switch patterns from restriction/sequencing analysis 134

4-3 Proviral structure and template switch junctions for 7 white clones 137

4-4 Frequency of tail lacZ allele at different regions of lacZ gene in the characterized proviral clones 142

4-5 Distribution of template switch sites across the lacZ sequence in the characterized proviral clones 144

4-6 HIV-1 polyadenylation signal readthrough rate and readthrough RNA-related recombination 145

4-7 The 3' appended RNA from readthrough transcription as preferred recombinogenic template switch donor 148

5-1 Direct repeat deletion assay to determine the relative contribution of inter- and intramolecular template switch to repeat deletion 169

5-2 Two-vector assay to study the effect of DIS sequence variation on HIV-1 genetic recombination 176

5-3 HIV-1 recombination rates between viral RNAs bearing identical or different DIS sequences 179

5-4 RNase protection assay to quantify virion RNAs 181

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Table

2-1 Effect of HU or AZT treatment on *lacZ* inactivation rates of HIV-1 70

2-2 Deletion rates of direct repeats of different length of HIV-1 71

2-3 *LacZ* inactivation rates and deletion rates of direct repeats of different length of MLV in mATRC1/293 cells 73

2-4 Effect of HU or AZT treatment on deletion rates of 117-base direct repeats of HIV-1 76

4-1 Frequency of template switching in characterized proviral clones..... 140

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER ONE

INTRODUCTION

Retroviruses replicate their genomes through reverse transcription carried out by viral reverse transcriptase. The process of reverse transcription is highly error-prone. This review chapter starts with a brief introduction of HIV-1 genetic variation, focuses on two driving forces of retroviral genetic diversity—mutation and recombination, and concludes with a discussion of RNA dimerization and its role in genetic recombination.

HIV-1 GENETIC VARIATION

HIV-1 nomenclature and genetic diversity

Human immunodeficiency virus type 1 (HIV-1) is characterized by its extraordinary genetic diversity. Based on phylogenetic analysis, three distinct groups of HIV-1 (termed M, N, and O) have been identified for globally circulating strains, with group M further divided into at least nine subtypes (A-D, F-H, J, and K) (Figure 1-1) (161). Group M comprises the vast majority of circulating strains responsible for the worldwide AIDS epidemic. Group O strains represent a minority of the HIV-1 isolates (24, 58, 105, 187), and have been suggested to be endemic to Cameroon and its neighboring countries in West Central Africa (145), constituting 2-5% of HIV-1 positive samples. Group N has only recently been identified, and only a limited number of group

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

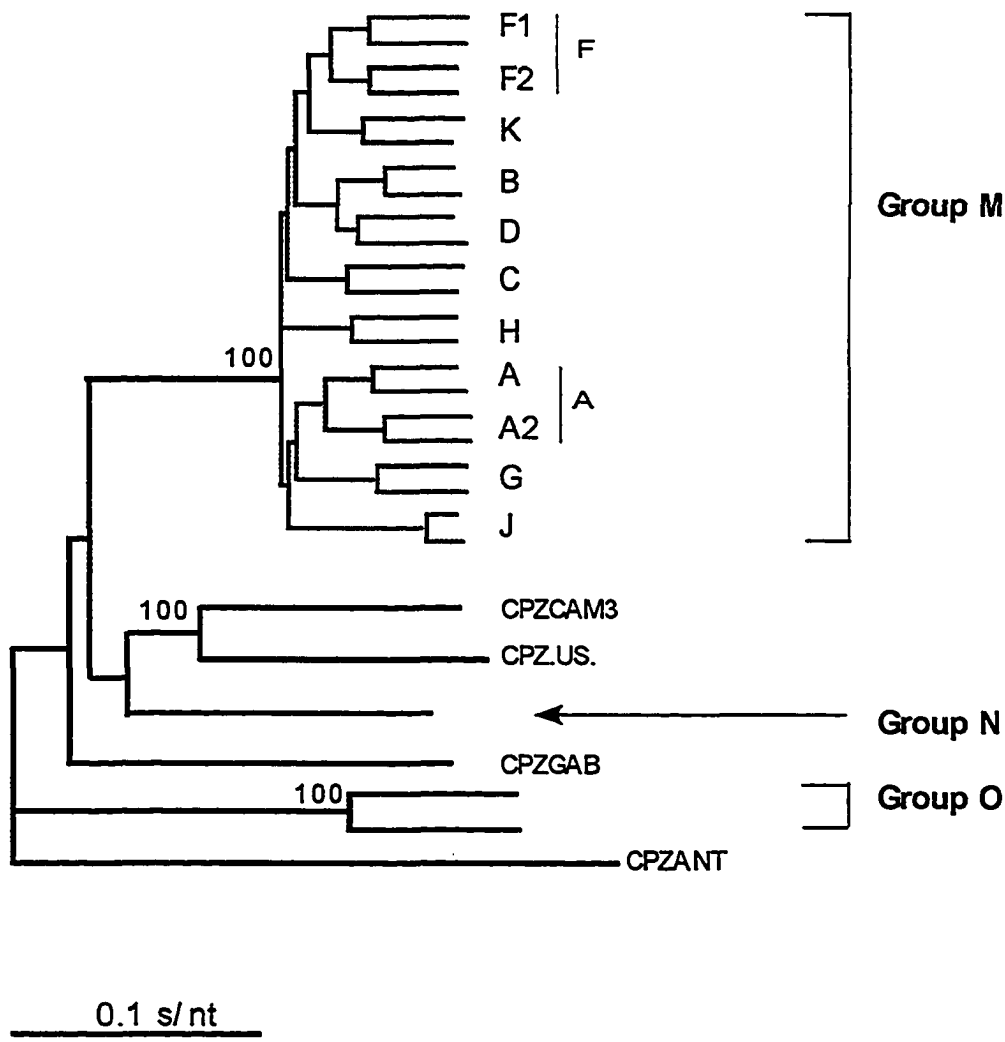


Figure 1-1. Evolutionary relationships among non-recombinant HIV-1 strains from the HIV-1/SIVcpz lineage, based on neighbor joining phylogenetic analysis of near full length genome sequences. The phylogenetic tree shows the different HIV-1 groups, subtypes, and sub-subtypes. Each of the internal branches defining a subtype or sub-subtype is supported by the maximum bootstrap value (marked by the number 100), which is an estimate of the statistical significance of a particular branching event. Distance between two sequences can be derived by summing the length of the connecting branches, using the scale (in nucleotide substitutions per nucleotide [s/nt]). Adapted from reference (143).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

N isolates have been reported so far (6, 175). HIV-1 strains from different tiers of categories generally reflect the genetic distance between them. Among isolates within a given subtype, extents of sequence variation are relatively modest (0-15% in *env* nucleotide sequence). Different subtypes differ from each other by roughly 15 to 30% in *env* nucleotide sequence (156). Significant sequence divergence exists between two groups: for example, group M and O viruses' homology in various genome regions ranges from 55%-76%, and is about 65% on average (182). The extent of genetic divergence among HIV-1 isolates has presented an unprecedented challenge to diagnosis, prevention and treatment of HIV-1 infection (145).

Factors contributing to HIV-1 genetic diversity

HIV-1 diversity can be attributed to multiple factors, some common to other retroviral species, others unique to HIV-1. These factors include: (i) low fidelity of HIV-1 DNA synthesis (approximately 3×10^{-5} mutations per base per replication cycle (112)); (ii) a high recombination frequency between copackaged RNAs (at least 2-3 recombination events per genome per replication cycle (71, 177, 206)); (iii) a high level of virus production (10^{8}-10^{10} virions per day (64, 148, 194)); (iv) a large number of viral replication cycles (140-300 cycles per year (31, 148)), and (v) a large pool of infected individuals due to the HIV-1 pandemic around the world. The relative contribution of these factors to HIV-1 genetic variation has not been evaluated. It has been suggested that recombination has the capability of shuffling the genetic information between two genomes, providing the potential for evolutionary leaps with far more dramatic genetic consequences than the incremental accumulation of individual mutations due to base misinsertion errors (109). Indeed, retroviruses evolve much faster than their host genomes, a notion supported by the observation that the retroviral v-*mos* oncogene of Moloney mouse sarcoma virus and its cellular counterpart c-*mos* accumulated mutations

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

at $1.3x10^{-3}$ and $1.7x10^{-9}$ substitutions per site per year, respectively (54). The contribution by two of the factors, retroviral mutation and recombination, will be detailed below.

HIV-1 MUTATION

Retroviral variation has been reported since early in the 20th century. Now retroviral variation is well documented in sequential isolates of HIV-1 and simian immunodeficiency virus (SIV) taken from individuals infected with a known source of virus (155). Sequence analyses reveal that the virus exists within its host as a pool of closely related genetic variants, known collectively as quasispecies. Within infected individuals, both genetic divergence (differences from the founder strain) and diversity (breadth of genetic variation within the viral population) increases by ~1% per year during the early phase of viral infection, with increases in divergence typically continuing for the duration of the infection (170). Such viral genetic fluctuation has also been observed in tissue culture upon continuous passage of retroviruses (169).

The composition of a retroviral swarm is influenced by many factors, including mutation rate per replication cycle, the number of replication cycles, fitness (the selective advantage or disadvantage possessed by the variant viruses), population size, competition and sampling. Several mathematical models have been proposed to describe the dynamics of retroviral variation (31, 49, 148). A simple steady-state model estimates the number of variant genomes in a large freely competing population of viruses as the number of pre-existing variants that replicate, plus the number of parental genomes that mutate, minus the number of variants that revert to parental type (31). A prediction of this model is that very small changes in the relative growth rate result in rapid and dramatic changes in the distribution of variants within a retroviral swarm (31). This might be complicated by the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

discovery that retroviruses evolve at different speeds in relatively isolated sub-compartments (122, 192). It has been argued that the overall composition of a large, rapidly replicating HIV-1 population is likely determined primarily by replicative fitness and not mutation rate (154).

Sources of retroviral mutation

Retroviral mutations can be introduced during any combination of the following three steps: either during transcription of proviral DNA by host RNA polymerase II (pol II), during minus- or plus-strand DNA synthesis by reverse transcriptase (RT), or during proviral DNA replication by cellular DNA polymerase. Due to the high fidelity of cellular DNA replication (10^{-9} to 10^{-11} substitutions per base pair per cell division) (46, 191), the contribution of cellular DNA polymerase to retroviral mutation is nearly negligible. Human pol II possesses an intrinsic 3'→ 5' exonuclease activity and can remove misincorporated nucleotides from nascent transcripts in the presence of transcription elongation factor SII *in vitro* (185). Although the error rate for eukaryotic pol II has not been successfully measured (171), the relative contribution to retroviral mutation by pol II and RT has been investigated (83, 131). A study on spleen necrosis virus (SNV) replication demonstrated that retroviral mutation rates during early (RNA transcription plus minus-strand synthesis) and late (plus-strand synthesis plus DNA repair) stages were equal or within two-fold of each other, implying an important role of RT in retroviral mutation (83). It was later confirmed by a similar study on HIV-1, which suggests that RT is likely the primary contributor to HIV-1 mutagenesis (131). In addition to polymerase errors, mutations may result from modifications of retroviral RNA or DNA templates. Double strand RNA adenosine deaminase has been implicated in avian retroviruses A→G transitions (48, 60, 83). Posttranscriptional RNA editing has been observed for viral messager RNAs that are engaged in viral gene expression, but

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

appears not to be involved in modification of either genomic viral RNA or proviral sequences (17).

Retroviral mutation rates

The fidelity of RT was initially measured *in vitro* utilizing purified RT and either synthetic or natural templates (81). The estimation of retroviral mutation rates (the number of mutations per nucleotide pre replication cycle) became more reliable due to the advance of single-replication-cycle assays in cultured cells (45). Such assays typically employ selectable or screenable reporter genes, including *lacZ* (77, 151) or its subunit *lacZα* (112, 113, 141, 142), herpes thymidine kinase gene (138, 189), neomycin phosphotransferase gene (45, 188), and green fluorescent protein (*gfp*) gene (214). The majority of the mutational studies identify mutations as loss of function or forward mutation assays (77, 112, 113, 138, 141, 142, 151, 189), while others detect the gain of function as the result of reversion of inactivating mutations, also termed reverse mutation assays (45, 188). One advantage of using a forward mutation assay system is to survey the types of mutations in addition to the estimation of the mutation rates. Forward mutation assay systems using longer target genes presumably have a better chance to capture a wider variety of mutations, especially gross genetic rearrangements (138, 189).

Using single-cycle assays, forward mutation rates for various types of viral species have been reported. From experiments in cultured cells, a wide spectrum of mutations is observed, including base substitutions, frame shift mutations, and genetic rearrangements (deletions/insertions) (112, 113, 138, 141, 142, 189). Among them, base substitution errors arise most rapidly, averaging 1.5×10^{-5} per nucleotide per replication cycle, or roughly once during the synthesis of one to ten genomes (155). Simple frameshifts and genetic rearrangements average 5×10^{-6} per nucleotide per replication cycle (155). Mutations rates for various retroviral species have been compared by using

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

vector viruses containing identical *lacZα* peptide gene, including SNV ($1.2x10^{-5}$) (141, 142), BLV ($4.8x10^{-6}$) (113), HTLV-1 ($6.7x10^{-6}$) (111), and HIV-1 ($3.4x10^{-5}$) (112). The mutation rate for HIV-1 was later measured to be 8.5×10^{-5} in LTRs in a different assay (131). The measured HIV-1 mutation rate is about 20-fold lower than the values obtained for purified HIV-1 RT with the same target sequence (19, 112). A similar observation was made when the fidelity of purified MLV RT was compared to the rate of base substitution errors during the replication of an MLV-based retroviral vector on an identical sequence template (188). This discrepancy in fidelity could occur as the result of any alteration of polymerases during purification and/or of cell-free conditions of DNA synthesis, and suggests that viral and/or cellular factors may significantly affect retroviral mutagenesis (112, 188). The mutation rates for BLV and HTLV-1, the type members of the BLV/HTLV-1 genus of the retroviridae family, are relatively lower than those of SNV and HIV-1. However, the low level of genetic diversity of HTLV-1 has been speculated to be due to oligoclonal expansion of infected cells and very low level of virus replication in infected individuals rather than their low mutation rates (111, 193).

Other factors that affect retroviral mutation rates

Other factors that affect retroviral mutation rates include perturbation of intracellular environment and treatment with antiviral nucleoside analogs. The composition of intracellular nucleotide pools can be manipulated by treatment with thymidine or hydroxyurea. Hydroxyurea treatment alters dNTP pools by inhibiting ribonucleotide reductase, resulting in depletion of all dNTPs, with the most significant reductions in the dATP pool. Both thymidine and hydroxyurea treatment induce significant dNTP pool imbalances. Hydroxyurea has been shown to modestly increase the *lacZ* inactivation rates for both SNV and MLV (77, 151, 180). These studies suggest that overall sizes of dNTP pools as well as the natural pool imbalances are likely to influence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

mutation rates and spectra (77, 151). By analogy, mutation rates in different cell lines might vary if dNTP levels or natural dNTP pool imbalances differ. The activation state of the cell has been shown to influence the levels of dNTPs (52). Therefore, the mutation rates of retroviruses may differ between proliferating and quiescent cells. Treatment with the antiviral nucleoside analog, 3'-azido-3'-deoxythymidine (AZT), increased mutation rates 2-fold for MLV, 10-fold for SNV when measured by *lacZ* inactivation, but 2-3 fold increase with *lacZα* (76). It has been shown that AZT treatment has little effect on dNTP pools. One implication of such experiment is that AZT treatment of HIV-1 infection may increase the degree of viral variation and alter virus evolution or pathogenesis. Similarly, 5-azacytidine increased the SNV mutation rate 13-fold in a single cycle of replication (140). This increase on mutation rates is suggested to occur due to its incorporation into RNA or DNA, or some secondary effects on nucleotide pools (140).

HIV-1 GENETIC RECOMBINATION

Retroviral genetic recombination

Genetic recombination in animal retroviruses has been known for decades. In fact, early work with avian RNA tumor viruses suggested that recombination was so frequent that genetic markers re-assorted essentially as if they were unlinked (202). This is surprising when one considers that the entire genetic content of retroviruses is contained on a single RNA (190). Because covalently linked genes cannot re-assort on separate RNAs as they do for multi-segmented viruses such as influenza, retroviruses must have evolved a mechanism to reassort physically linked genes at an unprecedentedly high frequency (33, 109).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Once sequencing full-length HIV-1 genomes became common practice, it became apparent that the ancestry of many viral isolates was not linear. Instead, these recombinant strains contained patches of sequence derived from more than one previously recognized subtype (143, 156). The frequent observation of natural recombinants suggests that recombination is a relatively common occurrence among different strains of HIV-1 (143, 156). The majority of HIV-1 recombinants characterized thus far contain sequences from members of different group M subtypes. Some of these are classified as circulating recombinant forms and have been found to contribute to the global HIV-1 pandemic (143, 161). Inter-subtype recombinants (ISR) have been identified in nearly every region of the world where two or more subtypes cocirculate, and they may account for over 10% of new HIV-1 infection (145). Intra-subtype recombinants have been identified within individual patients and may be more common than has been recognized due to the lower sequence divergence among parental strains. Recombination can apparently occur even among distantly related strains, because a limited number of inter-group M/O recombinants have been reported (144, 182).

Systems for examining retroviral recombination rates

In vivo observations and in vitro studies

Early work to determine whether or not HIV-1 and other primate lentiviruses could recombine demonstrated that genetic recombination was readily detectable. For example, Clavel and colleagues showed that an integrated provirus containing a stop codon could be rescued by recombination with a second defective virus that had a different mutation (25). Other groups found that recombination could lead to the co-segregation of drug resistance markers in tissue culture (82, 124), and that SIV recombination occurred in an experimentally coinfected rhesus monkey (200). However,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

because it is difficult to determine generation times in a replicating virus population, and because competitive advantage largely dictates population composition (31), it would be inaccurate to deduce the frequency of recombination from the prevalence of replication-competent recombinants. Recombination has also been studied with *in vitro* assays, which are performed with purified enzymes and model primer-templates in reconstituted reactions (130). The findings from such systems vary in the extents to which they reproduce events as they occur during viral replication, but they provide valuable insight into dissecting mechanistic features of template switching. In contrast, single-replication-cycle assays in cultured cells are the preferred way to study recombination, with the advantage of recapitulating retroviral recombination in a complex cellular environment in the absence of selection pressure (205). Two types of such assay are discussed below.

Two-vector recombination assay

Recombination studies between copackaged gammaretrovirus vectors were pioneered by Temin and colleagues (68, 69, 184). An overview of these approaches is presented in Figure 1-2. An initial step in these assays was to establish cell lines that coexpress virion proteins and two retroviral vectors. The two vectors were highly similar, but each was engineered to express only one of two different drug resistance markers (68). If recombination between marker-inactivating mutations occurred during reverse transcription of heterozygous virions, a provirus encoding resistance to both drugs could result. Provided that the multiplicity of infection was low enough that the chance of single cell coinfection was minimal, the number of target cells that expressed both markers was an indication of genetic recombination.

Recombination frequencies can be calculated from single- and double-resistant selectable marker titers after consideration of a number of parameters (205). One is to calculate the predicted "recombining population" size, or fraction of virions that are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

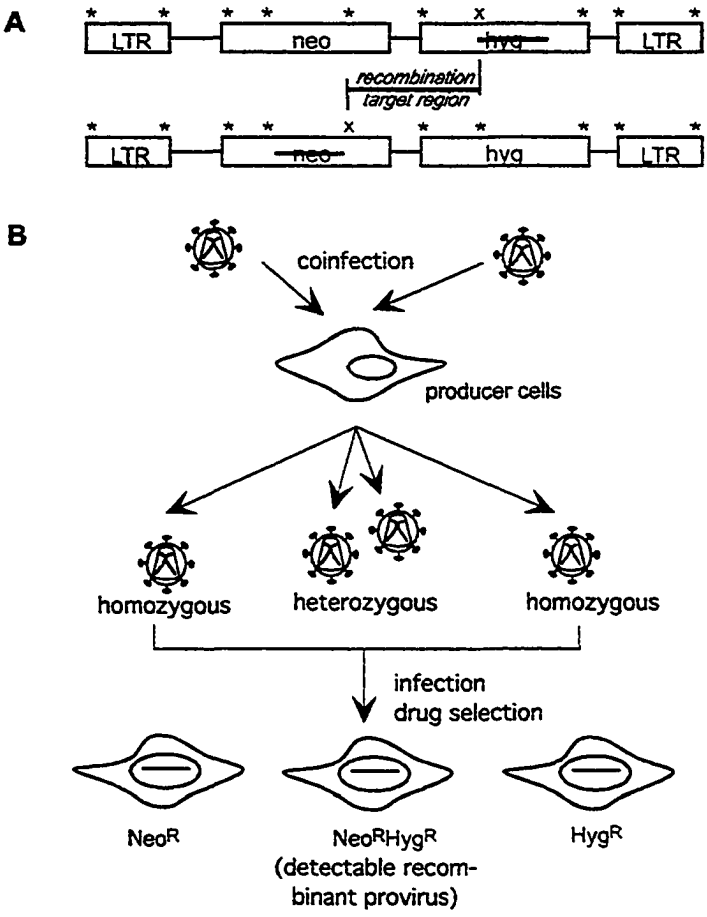


Figure 1-2. Two vector recombination assays. A) Schematic representation of gammaretrovirus recombination assay vectors used by Hu and Temin (69). * indicates position of genetic difference between templates; x indicates genetic difference responsible for inactivating marker gene; recombination target region is interval between inactivating mutations. LTR indicates long terminal repeat, which contains U3, R, and U5 sequences, neo and hyg indicate drug-selectable marker genes. Crossed-out text represents inactivated gene. B) Overview of experimental scheme for recombination assays that use two vectors. Two different proviruses are introduced into single producer cells: in the example here, by co-infection. Virions are then harvested from vector co-expressing producer cells. As described in the text, some virions are predicted to contain RNA heterodimers. Vector-containing virions are used to infect target cells. Progeny provirus-containing cells are identified by drug selection. Provided each parental vector conferred resistance to only one drug and infections were performed at a low multiplicity of infection, dual drug resistance is diagnostic of recombination. Note that in this figure, virions are represent with a capsid morphology more reminiscent of HIV-1 than of gammaretroviruses.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

capable of generating detectable recombinants because they have copackaged both vectors. It has been assumed that if two packagable RNAs are coexpressed, they will associate and dimerize randomly based on their intracellular proportions (33). This generates heterozygous and homozygous virions at frequencies predicted by the equation $A^2 + 2AB + B^2 = 1$, known in population genetics as the Hardy-Weinberg equation. If the concentration of RNA A is the same as that of RNA B, AB heterodimers should be present in half of the total virions, and AA and BB homodimers are predicted to be present in 25% of the total apiece. For recombination frequency calculations, the proportions of coexpressed RNAs in one dimer form or another have often been inferred from single drug resistance titers (69, 206).

To calculate recombination frequencies, dual drug resistance titers per calculated recombining population member are doubled to account for the fact that for every dual resistant provirus, the reciprocal recombinant (undetected, dual drug sensitive) should form (73, 212). Values are also adjusted to account for the fraction of the genome in which recombination was measured (the recombination target). For example, if the target interval is 0.8 kb and the entire vector is 8 kb, then target interval recombination values must be multiplied by 10 to yield a genome-wide rate. These approaches have yielded frequencies of approximately one recombination event during the synthesis of every 2 to 7 gammaretroviruses (3, 68, 69, 72).

Similar single-cycle two-vector recombination assays were also developed for HIV-1 (71, 206). Assay vectors were derived from two HIV-1 strains, HXB2 and BCSG3, respectively, which differ in nucleotide sequence by approximately 5%. Both vectors retained most of the native HIV-1 genomic sequences except that part of *env* was replaced by a different marker gene in each vector. Progeny proviral DNA was amplified by PCR and various non-overlapping segments of the vectors were subjected to heteroduplex tracking assay, homo- and hetero-duplex fragments yielded distinct gel migration patterns that were used to identify the parentage of each segment, and the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

genome intervals in which crossovers occurred were determined. The data yielded an average HIV-1 recombination rate of about three crossovers per genome per replication cycle, which was roughly 10-fold higher than those which had previously been reported for simple retroviruses, using similar experimental approaches (3, 69).

Potential weaknesses of two-vector recombination assays include the fact that they are technically challenging and that they are based, in part, on some theoretical assumptions. One assumption is that the copackaging of two coexpressed RNAs is fully random. If two different RNAs become copackaged less frequently than random copackaging would predict, the actual recombination rate for the recombining virion population would be much higher than that calculated based on the assumption of random copackaging. One technical challenge is to establish a situation where the investigator can be reasonably certain whether or not a particular proviral product was generated by a heterozygous virion. One way that this has been addressed is to coexpress vector RNAs at very different levels (137). If 99% are RNA A and only 1% are RNA B, then Hardy-Weinberg equilibrium predicts that the vast majority of proviral products harboring marker B were derived from AB heterozygotes.

Direct repeat deletion assay

It has been observed that repeated sequences are often precisely deleted from retroviral vectors, suggesting that their removal is a form of homologous recombination (142, 160, 204). In assays designed to exploit this, repeat deletion has been used to inactivate or reconstitute marker genes (75, 209). A repeat deletion vector that contains both a selectable marker (a puromycin resistance gene) and a screenable marker (*lacZ*) was developed for MLV (151). This vector was engineered to contain a coding region-internal sequence duplication that disrupts the *LacZ* translational reading frame, so that parental-form proviruses confer puromycin resistance but do not express functional β-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

galactosidase. A functional *lacZ* gene is generated if precise repeat deletion occurs during reverse transcription, and cells harboring a deleted provirus stain blue when incubated with the chromogenic substrate, X-gal (151). Because repeat deletion is presumed to be a recombination-like property, the blue to total (blue plus unstained) puromycin resistance titer is an indication of recombinogenic template switching frequency. Similar vectors that express *gfp* and can be scored by florescence activated cell sorting have also been developed (180, 209).

Direct repeat deletion rates are generally proportional to the repeat length (40, 41, 75, 151, 213). A study on MLV suggests that the frequency of direct repeat deletion increases as the linear distance between the repeats increases up to 1.5-kb (41). In the same study, the pattern of template switch locations was analyzed. It appeared that the frequency of template switch was greater within the 5' regions in comparison to the 3' regions (41). A model was proposed to explain the observation that distance between homologous sequences increases the frequency of direct repeat deletion (41). It was later modified to incorporate the role of RNase H in direct repeat deletion (Figure 1-3) (70, 180).

A major advantage of repeat deletion assays is that template switching can be monitored using a single vector. Because each vector contains both the sequence from which RT will switch (the donor template region) and the sequence to which RT will switch (the acceptor region), this approach obviates the need to determine template RNA ratios. Although repeat deletion is sometimes referred to as intramolecular recombination, a weakness of this experimental approach is that it is unclear whether or not repeat deletion is a good mechanistic mimic for genetic recombination. Possible concern is raised by findings from *E. coli*, where direct repeat deletion is RecA independent but other forms of homologous recombination are RecA dependent (106).

An obvious limitation of the recombination assays presented above is that they represent viral replication in cultured cells. Single cycle experiments are intended to be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

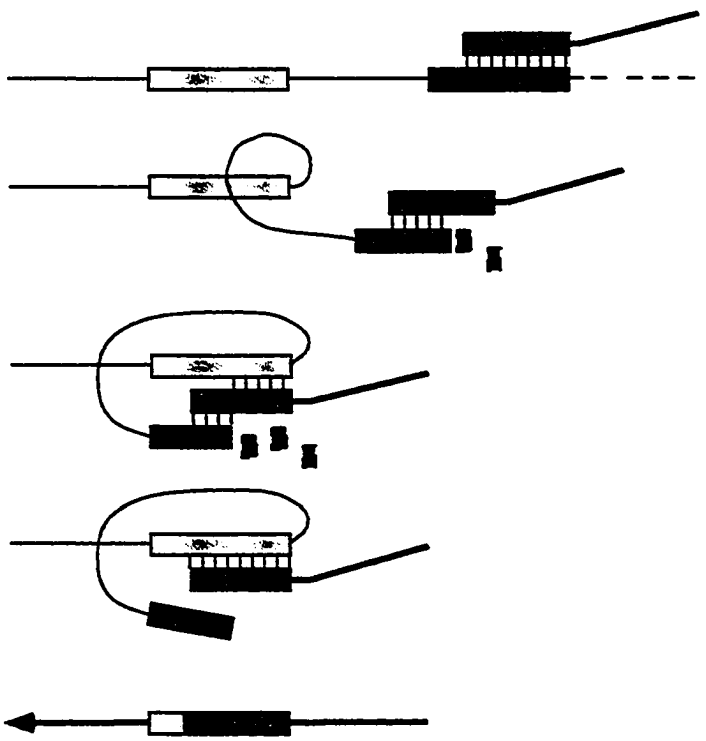


Figure 1-3. Model for direct repeat deletion. The thin black line represents the RNA template and the repeats are represented by two colored boxes. Reverse transcriptase is not included in the diagram. Nascent minus strand DNA is shown as a thick blue arrow, with the portion corresponding to the repeat region represented by a box in colors matching those of the RNA templates. The degradation of template RNA is illustrated by broken lines or small boxes outlined with dotted lines.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

largely free of selective pressures. Because small differences in selective advantage can have large effects on replicating populations (31), tissue culture-based results obviously do not predict outcomes of natural infection. What these experiments do provide is baseline rates of recombination that can be compared to clinical outcomes to help assess effects of immunologic and other selective forces *in vivo*.

Mechanisms for recombination

Intermolecular vs. intramolecular recombination

Retroviruses contain two genomic RNAs in their virions. Thus retroviral recombination can occur either between the two copackaged RNA molecules (intermolecular) or between two sequences within the same RNA molecule (intramolecular). For replicative template switches, minus-strand transfer can occur either intramolecularly or intermolecularly (4, 69, 73, 137, 186, 198, 206), and plus-strand transfer is mostly intramolecular (69, 73, 137, 206). The modes of recombinogenic template switches have been studied using vectors containing different genetic markers. By definition, two-vector recombination assays are designed to study the properties related to intermolecular recombination. In contrast, in single-vector recombination assays, for example, repeat deletion assays, the recombination can be either intramolecular or intermolecular. Studies suggest that high rates of recombination between two identical sequences within the same RNA molecule resulted primarily from intramolecular recombination (66, 208) (See Chapter five for discussion). Conflicting evidence exists as to whether such intramolecular recombination occurs during minus-strand DNA synthesis or plus-strand DNA synthesis (18, 97). An earlier study suggests that template switching rates during minus- or plus-strand DNA synthesis are nearly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

equal (18), but later it was disputed by the observation from another group that most intramolecular recombination occur during minus-strand DNA synthesis (97).

Models for intermolecular recombination

Several models have been proposed to explain retroviral recombination during either minus-strand DNA or plus-strand DNA synthesis. The forced copy-choice model proposes that recombination occurs during minus-strand DNA synthesis. When RT encounters a break in the RNA genome, it can switch to the copackaged RNA for continuous DNA synthesis (34). It was later suggested that breaks in viral RNA are not required for template switching (67, 75, 204). Another model for retroviral recombination during minus strand synthesis is the minus strand exchange model (Figure 1-4) (20, 32, 130, 147), a currently favored model supported by recent RT structure-function studies (20, 149, 180) (detailed later). A third model, the strand-displacement-assimilation model, proposes that recombination occurs during plus-strand DNA synthesis (79). It was based on the observation that plus-strand DNA synthesis is discontinuous and could be initiated at multiple locations of the viral genome. An internally initiated plus-strand DNA fragment can be displaced from one minus-strand DNA molecule and be assimilated with the other minus-strand DNA. A recombinant is generated after DNA mismatch repair of the DNA heterduplex using the assimilated DNA as a template.

The relative contribution of two modes of recombination was examined using single replication cycle assays in simple retroviruses. In one study, the recombination mechanism for observed recombinants was deduced based on their genotype—the number of crossovers: a recombinant that is generated during minus-strand DNA synthesis would appear to have one crossover, whereas a recombinant that is generated during plus-strand DNA synthesis would contain two crossovers (69). The composition of the recombinants containing either one crossover or more than one crossover was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

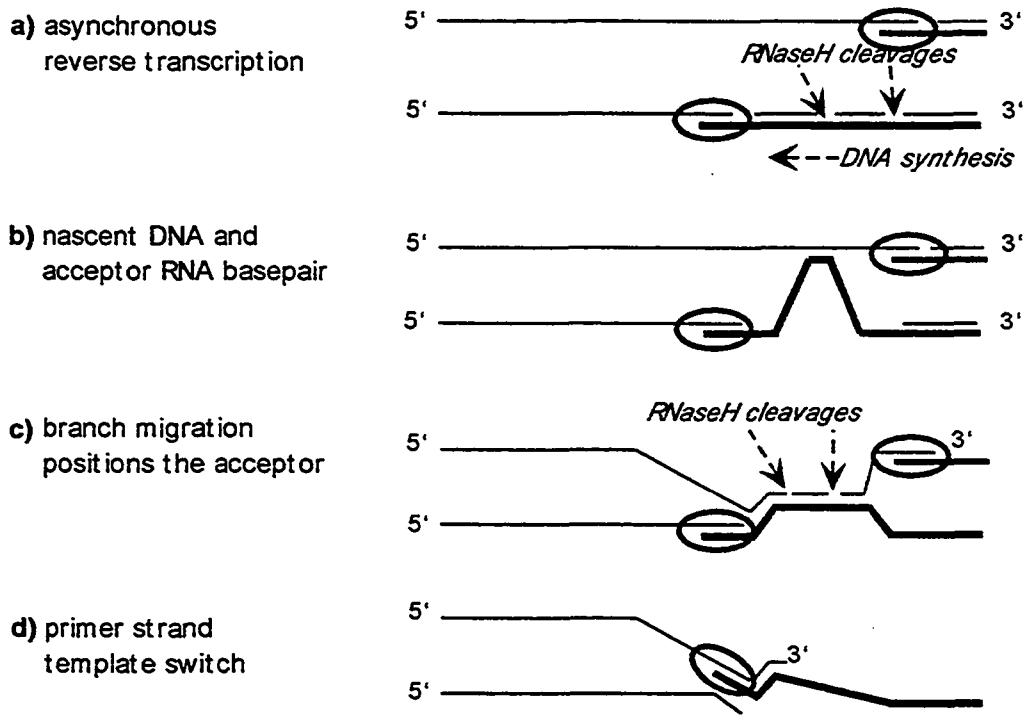


Figure 1-4. Minus strand exchange model for genetic recombination. a) DNA synthesis becomes asynchronous once RT advances further on one co-packaged RNA (the donor template) than on the other. b) Cleavage of the RNA template strand by RNaseH exposes single stranded minus sense DNA which can base pair with complementary sequences on the acceptor RNA. c) Branch migration of the nascent DNA: acceptor RNA duplex positions the donor template growing point near the acceptor template, and RNaseH degradation of the donor template causes the slower RT complex to abort. d) Template switch is accomplished by a repositioning of the primer strand onto the acceptor template, as has been demonstrated in reconstituted reactions. Thin lines represent RNA, thick lines represent DNA, ovals represent RT.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

compared to that from a system that only allowed recombination to occur during minus-strand DNA synthesis (4). Based on similar compositions of recombinant genotypes and close resemblance of recombinant phenotypes, it was claimed that the minus-strand recombination model is the major mechanism for retroviral recombination (4). Later, this topic was examined by a different approach using a system that allows both modes of recombination but is defective in DNA repair (210). The plus-strand DNA recombination model predicts the formation of a DNA heteroduplex which is then repaired by host cell repair machinery. In DNA repair deficient cells, the heteroduplex could not be repaired and will be resolved after cell division in forming a mixed colony. In such study, no mixed colonies were observed, suggesting that minus-strand recombination is the major mechanism of recombination (210).

Contribution of other enzymes

Single replication cycle assays used in studying retroviral recombination start with transcription of a provirus by cellular pol II in the producer cells, continue through reverse transcription of viral RNA by RT during infection, and end with replication of integrated proviral DNA by cellular DNA polymerase as an integral element of the host chromosome. All three polymerases could theoretically contribute to the overall rate of retroviral recombination. It has been suggested that pol II can utilize double-strand breaks and gaps in DNA to generate recombinant transcripts (80). Nevertheless, it has always been assumed that RT is mainly responsible for the frequent recombination observed in such assays. Only recently, the relative contribution of these three enzymes to retroviral recombination was explored in systems involving recombination between two identical sequences (96, 97). An MLV vector that contains two mutated *gfp* genes in tandem positions was used in the recombination assay. If recombination occurs between the two mutated *gfp* genes, a functional *gfp* gene will be created. Cells containing such a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

functional recombinant *gfp* appear green under fluorescence microscopy. Green cells that appear during proliferation of a clonal clear-cell population indicate recombination at proviral DNA level, and was measured to be approximately 10^{-5} events/cell division, which is extremely rare compared to the overall frequency of recombination of 51% (96). The contribution to retroviral recombination by host cell RNA polymerase II was examined by quantifying the green fluorescence intensity from clonal clear producer cells in a flow cytometric analysis. The frequency of recombinant transcript was measured to be 0.1%, indicating that intramolecular recombination between two identical sequences during transcription by host RNA polymerase II is minimal (97). These observations support the conclusion that most retroviral recombination occurs during minus-strand DNA synthesis by RT (97).

Factors affecting recombination

Sequence homology

Recombination between highly similar sequences (retroviral homologous recombination) is very efficient. Retroviruses also perform recombination between unrelated sequences (non-homologous recombination). For example, the junctions between host and viral sequences in retroviruses with host-derived oncogenes presumably arose via non-homologous recombination (Figure 1-5) (211). Gammaretrovirus assays designed to compare frequencies of homologous and non-homologous recombination concluded that homologous recombination is at least 10^3 times more frequent than non-homologous recombination (212).

Non-homologous recombinants may arise essentially by chance, may rely on residual homology or homology independent recombination triggers, or may be disproportionately common in short regions of high donor/acceptor template sequence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

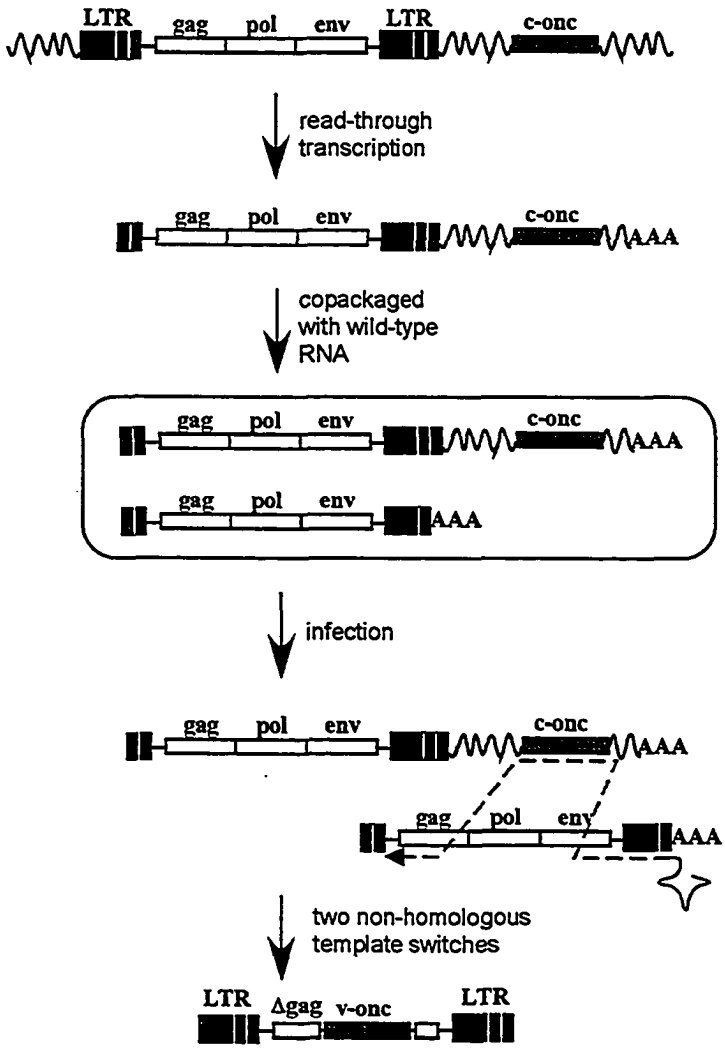


Figure 1-5. A model for retroviral transduction. Chance integration of a provirus upstream of a host oncogene allows occasional production of a read-through transcript that initiates in the upstream LTR and proceeds through the polyadenylation signal downstream of the oncogene. Once this RNA has been packaged and transmitted to a fresh cell, non-homologous template switching between sequences in the virus body (between LTRs) and sequences in appended downstream oncogene sequence can embed the oncogene in viral sequences. Note that other models for transduction have also been proposed, including ones in which rare host RNA co-packaging is envisioned to occur without physical linkage of host and viral RNAs..

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

identity. In gammaretrovirus systems, as little as 12 nucleotides of identity between recombination donor and acceptor templates appears sufficient to accurately target homologous recombination, albeit at a lower frequency than that observed for more extensive regions of homology (36, 150). Small patches of sequence identity are sometimes observed at otherwise non-homologous retroviral recombination junctions, suggesting that at least some non-homologous recombination is driven by highly localized regions of homology (211). Template features besides sequence homology, such as RNA secondary structure, spacing of sequences, homopolymeric stretches, and other less-well defined features of sequence context, appear to modulate recombination frequency (41, 98, 130, 141).

RT and its RNase H activity

One important factor that contributes to the high retroviral recombination rate is the unusual properties of RT. Retroviral RT contains two active sites: one for DNA polymerization and a second for RNase H, the nuclease that degrades RNA from RNA-DNA duplexes (183). Most DNA polymerases do not readily perform the sorts of template switches that are required for retroviral recombination. It has been postulated that the reason reverse transcription is prone to recombinogenic switching is because replicative template switches are required during the synthesis of every viral DNA (34, 184). Replicative strong stop template switching and genetic recombination are believed to be mechanistically similar processes (184). Structure-function studies have revealed that some alterations to gammaretrovirus RT affect template switching (149, 180), but these studies have not yielded a clear picture of RT structural determinants required for recombination.

The RNase H activity is an undisputed requirement for high frequency template switching (21, 23). Template switching frequencies are decreased several fold for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

gammaretroviruses that contain certain RNase H-defective forms of RT or that have reduced levels of active RNase H (20, 149, 180). These findings are consistent with "minus strand exchange" (20, 32, 130, 147), a currently favored model for most retroviral recombination.

Cell type

The type of cell a virus infects can affect reverse transcription, and single replication cycle assays have been used to address the effects of altering the intracellular environment on recombination frequencies. The immortalized tissue culture cells used in most experiments described above are metabolically highly active. Compared to tissue culture cells, HIV-1 reverse transcription takes longer and/or is disrupted in less active cells (152, 207). In gammaretrovirus recombination experiments where the time required to complete reverse transcription was tripled by serum starvation or by treating cells with hydroxyurea, the frequency of template switching increased 3-fold over the frequencies observed in untreated cells (151, 180). These findings suggest that frequencies of HIV-1 genetic recombination may differ significantly from one cellular context or host compartment to another during natural infection. Because HIV-1 DNA synthesis may take up to ten times as long to complete in resting cells than in actively dividing ones (152), such conditions may increase the frequency of recombination even more than the 3-fold achieved by experimental substrate limitation (151, 180). The numbers of crossovers per genome observed among 86 tested HIV-1 recombinants generated in immortalized cells ranged from one to ten, demonstrating that at least 10 crossovers can occur per HIV replication cycle in immortalized cells (71). Thus, because slowing reverse transcription can increase template switching at least 3-fold, it seems reasonable to postulate that within the range of recombinants generated in multiply infected patients,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

there may be some whose synthesis involves 30 or more crossovers during a single replication cycle.

Superinfection and coinfection

A fundamental prerequisite for recombination between two RNAs is coinfection on the cellular level. For the two vector assays described here, RNA coexpression is often achieved by unnatural means such as pseudotyping virions to permit infection of non-susceptible cells, or by experimental transfection (205). In the case of naturally arising inter-subtype and some intra-subtype recombinants, superinfection of a patient is an obvious prerequisite to cellular coinfection. The prevalence of recombinants and observations that serial infection can occur even in the presence of robust immune responses provide evidence that patient superinfection occurs more readily than had previously been speculated (16, 156, 158).

Cell coinfection for simple retroviruses is frequently limited by viral interference: the phenomenon whereby cell infection by one virus limits repeated infection by a second virus that shares the same receptor (195). Even though HIV-1 is known to down-regulate the surface expression of its receptor and viral interference has been observed *in vitro*, the extent to which cellular resistance to superinfection affects HIV-1 spread *in vivo* is unclear. Some reports suggest that cell coinfection *in vivo* is common (55, 78, 156).

Recombination between variants that arise within an individual can be detectable without superinfection of a patient, provided cell coinfection occurs, and it is likely that the extent of this has been under-appreciated because the similarity of parental strains masks recombinants (56). It has been argued that one competitive advantage of recombination is its potential to repair deleterious mutations that arise during error-prone replication or to promote phenotypic reversion (34). Such phenomena clearly function during HIV replication *in vivo*, as studies of the dynamics of drug resistance have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

revealed evidence of intrapatient recombination between circulating strains and archived proviruses (57). It is interesting to note that as integrated elements, defective proviruses can persist without replication. Some may even have a survival advantage if their defects shield cells that contain them from immune detection. Many patient-derived isolates contain mutations that likely render HIV replication defective (201), but because two different defective retroviruses can readily recombine to form infectious virus, even defective genomes can serve as functional recombination substrates.

RNA copackaging

Copackaging two complete RNAs in a single virion provides two templates to the reverse transcription machinery and is a critical factor in the high frequency of retroviral recombination. It has been proposed that viral heterozygotes, containing heterodimeric RNA complexes consisting of two genetically distinct subunits, are obligate intermediates in recombination (196, 203). Such observation was later supported by single-replication-cycle recombination assay, in which recombination was observed after target cells were infected with virions harvested from producer cells containing two genetically distinct proviruses, but not after target cells were coinfected with a mixture of two viruses harvested from two different producer cells containing only each of the two corresponding proviruses (68). The only difference between these two experimental conditions is the presence of heterozygotes in the virions harvested from producer cells containing two genetically distinct proviruses, indicating that recombinant proviruses are progeny of heterozygous virions (68).

The RNA content of retroviral particles is not tightly fixed. It remains unclear whether or not most retroviruses contain precisely two genomic RNAs, or if variation—from zero to four or more copackaged genomic RNAs—is normal. Both lentiviruses and gammaretroviruses are capable of packaging Ψ+ RNAs more than twice the native length

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(87, 174), and alpharetroviruses may copackage more than two RNAs when RNAs are short (164). Findings that host factors can modulate the molar ratio of RNA to virion protein in HIV-1 particles supports the notion that lentiviruses share the property of flexibility in packaged RNA composition (123).

If HIV-1 can copackage more than one dimer, speculatively, it may be possible for HIV-1 to copackage—and to recombine in a single cycle—more than two genetically distinct RNAs. It has been suggested that the appearance of sequences from more than two parental strains in many HIV-1 recombinants likely reflects recombination between pre-existing recombinants rather than triple or higher multiplicities of infection (156). However, recent *in situ* hybridization experiments with patient-derived single cells suggest it may not be unusual for individual cells to contain four or more different proviruses *in vivo* (55, 78). Taken together, these observations leave open the possibility that recombination between more than two parental strains may sometimes occur in a single round of reverse transcription.

Negative interference & recombination hotspots

It has been assumed that recombination occurs at all genomic positions at more-or-less uniform frequencies rather than principally at specific sites. Experimental work has confirmed that although recombination frequencies are not uniform—indicating the presence of local "hot" and/or "cold" spots for recombination, recombination is observed throughout the HIV-1 genome (2, 71, 199). Recombinogenic crossover frequencies are at least roughly proportional to recombination target length for gammaretroviruses (75, 213). In fact, the distribution of recombination breakpoints appears to be random along the HIV-1 genome when all the characterized natural HIV-1 recombinants were inspected (156).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The vast majority of gammaretroviruses show no genetic evidence of recombination. However, among recombinants, most display evidence of more than one crossover and some display several. These genetic observations have been argued to support a "high negative interference" model for retroviral recombination, which suggests that genetic recombination is not a routine process but instead biphasic, with crossovers during reverse transcription either frequent or else non-existent (34). It has been proposed that the performance of one recombination event predisposes the ordinarily recombination-proof virion to performing additional switches, or alternatively that only a subset of virions are prone to generating recombinant products (4, 66). It has further been suggested that the viral nucleoprotein complex architecture of one subset of particles is somehow aberrant, which results in a high frequency of recombination among products of virions with this postulated defect (73).

An alternate hypothesis for why only a subset of gammaretroviruses display evidence of recombination is that only a subset of particles copackage two different RNAs (33). Retroviral RNA trafficking—including which host factors are involved, which subset of RNAs are packagable, the intracellular compartment where commitment to packaging is made, and whether or not *cis*-acting packaging sequences are sufficient for RNA packaging—differs dramatically among retroviral species (9, 22, 35). As mentioned above, recombinogenic template switching between copackaged RNAs is only detectable if the copackaged RNAs differ. If two different RNAs become copackaged less frequently than random copackaging would predict, the following alternate hypothesis for gammaretrovirus recombination emerges. Recombinogenic template switching occurs at a uniform and high distribution of frequencies for all virions. However, recombination is observed for only the small subset of virions that copackage two different RNAs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The fidelity of retroviral recombination

In reconstituted *in vitro* reactions designed to mimic HIV-1 template switching, base substitutions are observed at a significant proportion of all template switching junctions (43, 146). These observations with purified RT led to the hypothesis that genetic recombination might cause many of the substitutions found in HIV-1 genomes (139, 147). It was suggested that transfer-associated errors could result from non-templated addition of an uncoded base followed by mispair extension upon growing point transfer to the second template. Nontemplated addition occurs at more than 50% of all template ends when studied with purified RT (139, 147). Potential support for this possibility came from observations that errors at strong stop switch junctions appear as frequently as in 5% of all replication products for some retroviruses and retroelements (86, 127). It has been suggested that host DNA recombination may be mutagenic (95), and the hypermutable recombination junctions that are observed for immune system gene rearrangements seem to confirm that some forms of recombination can be mutagenic (84).

In contrast to products from purified biochemical reactions, recombination junctions generated during replication in tissue culture appear fairly free of errors. Two-vector recombination assays have been used to study whether or not misincorporation errors occur more frequently at recombination junctions than during synthesis that proceeds without recombination (15, 150, 213). In studies that addressed gammaretrovirus recombination fidelity, the recombination target region was sequenced from 29 (213) or 182 (15) individual products of two vector recombination assays. The frequency with which substitutions were observed—no substitutions in a total of about 1.5 kb (213) or 65.6 kb (15) sequenced—suggests substitutions at recombination junctions are not significantly more frequent than reported whole genome substitution frequencies (45, 142, 188). Similar studies performed with HIV-1-based vectors

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

concluded that HIV-1 recombination also is not particularly error-prone (215). An alternate approach for studying gammaretrovirus fidelity involved a genetic system for forced recombinogenic template switching between the 5' end of one RNA and an acceptor target region on a copackaged RNA. If template switching were faithful, a restriction site would form at the transfer junction, but if a non-templated acceptor template non-complementary base were added prior to template switching, the restriction site would be destroyed. The findings demonstrated that errors at that position occurred below the <1% threshold of detection (150).

Interestingly, in experiments where gammaretrovirus template switching frequency was suppressed by RNase H limitation, the fidelity of DNA synthesis was drastically reduced (20). Some or all of the reduced fidelity may have resulted from RT defects unrelated to template switching (214). Nonetheless, template switching generally may be more of a fidelity factor than a mechanism of error introduction. During the reverse transcription of every retroviral DNA, RT must switch templates when it reaches its tRNA template's first modified base, which is perceived to be non-coding (181). By analogy, it is possible that recombinogenic template switching may provide an error escape mechanism when the reverse transcription machinery reaches a dysfunctional environmentally modified base (42). Error-avoidance—particularly from broken templates-- is part of "forced copy choice" models for genetic recombination (34), which remain useful even though it now appears that recombination rates are higher than broken RNA frequencies (204).

HIV-1 RNA DIMERIZATION AND RECOMBINATION

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Retroviral dimer linkage structure

Retroviral genome consists of two usually identical RNA molecules. The diploidity of retroviral genomes was initially suggested by sedimentation and electron microscopy (EM) studies (89, 110, 178). Initially, EM studies on dimers of genomic RNA extracted from RSV virions revealed that the two molecules are held together at multiple contact/interacting points (110). Later, the association near their 5' ends was shown to be the most stable contact between the two genomic RNAs, and was termed dimer linkage structure (DLS) (11, 89)). DLS has been confirmed in a wide variety of murine and avian retroviruses (10, 11, 88, 89, 108, 129). The general orientation of the two RNA strands appeared to be parallel (11, 89). The DLS was proposed to be stabilized by hydrogen bonds or other non-covalent interactions, since the dimers dissociated into monomer under relatively mild conditions (10, 88, 178). The native DLS of full-length HIV-1 genomic RNA was visualized in 1997 (65). It was shown that the HIV-1 RNA genome contains a central dimer linkage structure and additional loop structures within each monomer subunit. In contrast to the DLS of other retroviruses, the free 5' ends of the two HIV-1 RNA strands were not visualized, suggesting that the 5' end regions of HIV-1 RNAs are involved in interstrand complementary base pairing (65).

Localization of dimerization initiation site

Precise localization of the DLS began in the late 1980s when Darlix and coworkers showed that synthetic RNAs corresponding to the 5' end of retroviral genomes dimerize *in vitro* (14, 38, 153). It was shown that a subgenomic HIV-1 fragment encompassing the 5' leader region and the 5' part of the *gag* gene of HIV-1 Mal spontaneously dimerized *in vitro*, and that this dimerization process was stimulated by the nucleocapsid protein (38). Based on the cation dependence of the HIV-1 RNA dimerization formation, a conserved purine-rich motif downstream of the SD site was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

proposed to be involved in HIV-1 RNA dimerization through the formation of quartets involving both guanines and adenines (114). It was then disputed by reports suggesting that HIV-1 RNA dimerization is stabilized by quartets containing solely guanines (5, 179, 197). Later it was shown that sequences located upstream of the SD site are involved in the dimerization process and dimerization of RNA fragments lacking sequences upstream of the SD site differs from that of RNAs containing these sequences in its cation dependence, the kinetics of dimerization and the thermal stability of the dimers (115). Nucleotides that are essential for initiation of RNA dimerization were identified by chemical modification interference (176) and site-directed mutagenesis (134). This region (nucleotides 260-294 of HIV-1 Mal isolate) was shown to adopt a stem-loop structure with a self-complementary sequence in the loop (61, 176), and was termed the dimerization initiation site (DIS) of genomic RNA (176). Similar localization of the sequences required for dimerization of HIV-1 Lai RNA was later determined by studying dimerization of truncated and mutated RNA fragments (30, 90, 128), and inhibition of dimer formation by antisense DNA oligonucleotides (128). The contribution of DIS to HIV-1 dimerization has later been confirmed by cell-culture based studies of mutating this sequence (29, 59, 94, 167, 172, 173). HIV-1 DIS has also been designated kissing-loop domain (KLD) (91, 92, 94, 173), or stem-loop 1 (SL1) (29, 30), which is the first stem-loop structure among the proposed four stem-loop secondary structures in HIV-1 5' leader sequence based on computer modeling and biochemical/enzymatic analysis (Figure 1-6) (26, 61) and mutagenesis studies (29, 116, 117).

The DIS stem-loop structure is conserved among all HIV-1 isolates (85). Two major DIS loop variants are observed among natural viral isolates (Figure 1-7). The loop is comprised of either GCGCGC (mainly seen in subtype B and D strains) or GUGCAC (subtype A and a variety of other subtypes including group O and N strains) flanked by two 5' and one 3' purines. In subtype C and group O strains, the first and last nucleotides of DIS loop are complementary, thus actually reducing the loop length to 7-nt. Compared

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

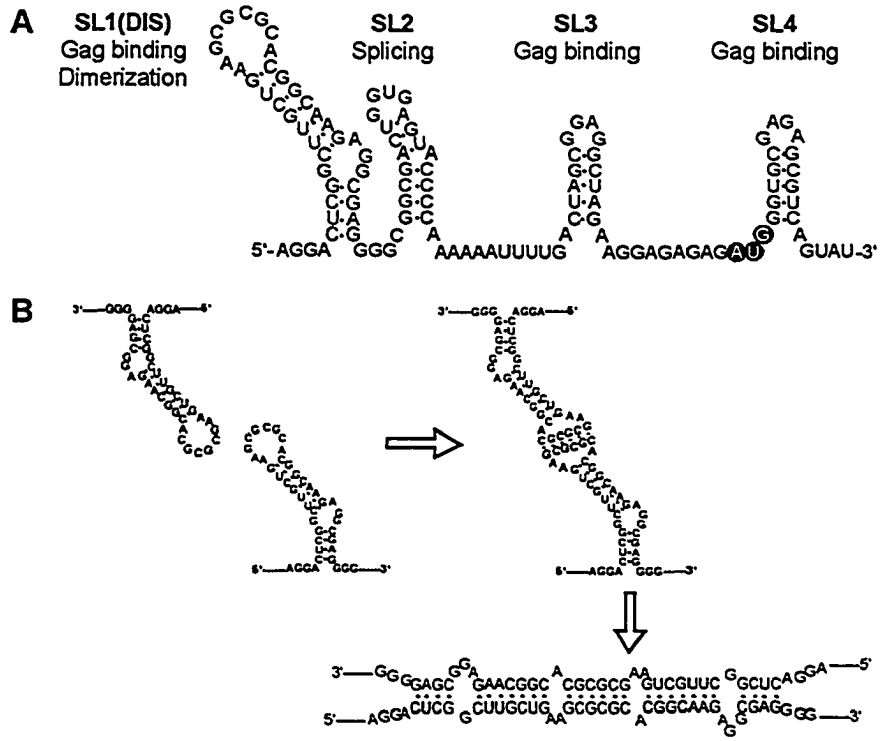


Figure 1-6. RNA structures, including the dimerization initiation sites, in the dimerization/packaging region near the 5' end of HIV-1 RNAs. A) Putative stem-loop structures of sequences in the vicinity of the dimerization initiation site (DIS, also known as stem-loop 1 or SL1) of subtype B HIV-1 genomic RNA, and their implicated functions. B) Schematic representation of the "kissing" interaction between the DIS palindromic loops of two RNAs and postulated isomerized structure. Adapted from (29).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

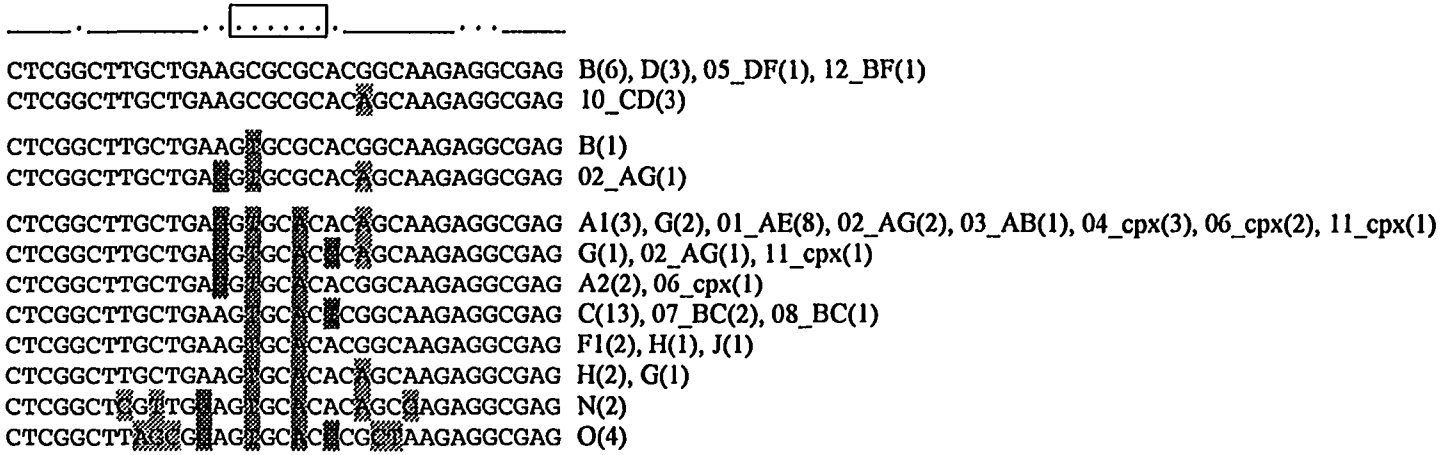


Figure 1-7. Nucleotide alignment of DIS sequences from complete HIV-1 genomes (85). The origin of the sequences is indicated to the right of the corresponding sequences according to HIV-1 nomenclature (160). Numbers in the parentheses indicate the number of strains of each HIV-1 category that contain the corresponding DIS sequences. Note not all the complete sequences for each HIV-1 category are included in this alignment. The proposed secondary structure of DIS is illustrated on top of the alignment with short lines (nucleotides in stems), dots (nucleotides located in bulges and the apical loop) and a box (the palindrome). Sequence divergence is highlighted.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to group M strains, strains from group O and N display sequence covariation in the DIS upper stem, and thus the secondary structure of the DIS stem-loop is maintained (Figure 1-7). Similar to the naturally existing DIS variants, *in vitro* evolution studies of HIV-1 kissing loop aptamers displayed marked sequence conservation of the DIS loop (103, 104). It was shown that adenines at positions 1 and 9, a purine at position 2, and the guanine and cytosine at positions 5 and 6 as the main constraints for optimal *in vitro* dimerization (103), which is consistent to the observation from a mutagenesis study (30).

In addition to the DIS motif, several other sequence elements have been implicated in viral RNA dimerization (163, 167, 173). Most of the early *in vitro* dimerization studies with synthesized RNAs suggest a role in dimerization of purine-rich tracts (5, 179, 197). In contrary, studies with full-length HIV-1 RNAs extracted from virions indicate that guanine quartets do not contribute significantly to the dimer stability (50, 162). Interestingly, other sequence motifs, such as stem-loop 3 (162), a GU-rich region immediately downstream of SL3 (162, 173), GU-rich sequences in U5 (163), and RNA sequence between PBS and DIS (167, 173) were implicated in HIV-1 dimerization. However, the relative contribution to genomic RNA dimerization by DIS and above elements still needs to be evaluated. For example, it is debatable as to whether or not the contribution by individual elements is additive (162, 172, 173).

Kissing-loop model of dimerization

The mechanism of retroviral dimerization has been mostly studied *in vitro* with RNAs containing parts of or the complete 5' non-coding region of the HIV-1 genome. It was shown that mutations that destroyed the self-complementarity in the DIS loop abolished dimerization (30, 59, 90, 93, 128, 134, 176), whereas introduction of compensatory mutations restored the process (134, 136). Certain complementary loop mutants that prevented homodimerization could complement each other and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

heterodimerize (30, 134, 136). It is now generally accepted that dimerization takes place by symmetric intermolecular interactions between the self-complementary sequences of the DIS loop, thus forming a "kissing-loop" complex (30, 59, 90, 93, 128, 134, 136, 176). It was initially suggested that the loop-loop interaction might be followed by formation of an extended duplex (90, 93, 128, 134, 176). However, using *trans*-complementary mutants that all pairs can form the loop-loop kissing complex but only two of them can potentially form the extended duplex, it was shown that no extended duplex is formed during dimerization of HIV-1 RNA *in vitro* (135).

NMR and crystallographic structures of HIV-1 kissing-loop dimers have been described (37, 47, 125). NMR data suggest that the dimer interface extends beyond the 6-mer interaction between the palindromes, and that it includes at least the three flanking bases from each DIS that stack to form a pair of interlocking triple-base arrays (125). Crystal structures of HIV-1 subtype A and B DIS indicate that they are virtually identical or have identical conformations, and the two interacting hairpins show a perfect coaxial alignment (47). The kissing-loop dimer gains stability by isomerization into the extended duplex form, of which the NMR structure is also known (53, 126). It is suggested that the disruption of the top base pair of the DIS stems in the kissing-loop dimer can lower the energy requirements for the transition of the kissing dimer to the extended duplex (125). Interestingly, some difference in the dimerization mechanism of subtype A and B RNAs was suggested by *in vitro* analyses of dimerization-competent DIS aptamers (74, 104). For example, unlike subtype A, subtype B DIS appears to require the binding of a divalent cation in the loop to promote RNA dimerization (74). In addition, the open or closed state of the last base pair in the stem seems to differ in the two subtypes (104).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Functions of DIS in viral life cycle

The function of DIS in the viral life cycle has been studied by examining properties of virions containing DIS mutations. Mutations introduced into the DIS sequence have been shown to affect multiple aspects of the HIV-1 replication cycle, including RNA dimerization (29, 59, 94, 167, 172, 173), RNA packaging (12, 29, 62, 101, 116, 117, 133, 167), viral infectivity (12, 29, 59, 62, 94, 101, 133), and other steps of HIV-1 replication (94, 99-101, 133, 172).

RNA Dimerization. An earlier study showed that the genomic RNA extracted from mutant virions was fully dimeric (12). However, the majority of the DIS mutageneis studies revealed a dimerization anomaly (29, 59, 94, 167, 172, 173). It was shown that a significant fraction of the genomic RNA inside the mutant virions was monomeric (29, 59), and that the proportion of dimeric genomic RNA was reduced to approximately 50% (94, 172, 173). However, the thermal stability of the dimerized RNAs was similar to that of the wild-type dimers (12, 29, 94, 168, 172, 173).

The roughly 2-fold reduction on viral RNA dimerization by DIS mutants observed from above cell-culture based studies is in stark contrast to the much more dramatic impairment of dimerization *in vitro* for subgenomic RNA fragments (30, 59, 90, 93, 94, 128, 134, 176). One explanation is that DIS sequence may not be absolutely required for *in vivo* RNA dimerization (12, 29, 59, 94, 167, 172, 173), and that other viral or cellular factors, such as some signals outside of the DIS motif (163, 167, 173), may compensate the role of DIS in viral RNA dimerization. Alternatively, this difference may reflect the dissimilarity between *in vitro* biochemical dimerization assays and cell-culture based experiments examining virion RNAs. For example, the latter examine the proportion of dimers among RNAs that had been encapsidated into virions rather than among the total RNA population. If dimerized RNAs had an encapsidation advantage

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

compared to monomeric RNAs, the difference on sampling between the two systems would result in the distinct effects of DIS mutations on dimerization.

RNA packaging. HIV-1 packaging signal (Ψ) is composed of multiple RNA elements near the 5' end of the genomic RNA that is necessary for preferential encapsidation of Ψ^+ RNAs ((13) and references therein), including SL1/DIS (12, 29, 62, 94, 117, 133, 167, 172), SL3 (29, 62, 117, 162), SL4 (27, 117), TAR (27, 63, 118), poly(A) stem-loop (27, 39, 118, 167), and elements from U5/PBS stem-loop (28, 118, 163, 167). The processes of RNA dimerization and packaging are intrinsically linked. Experimental evidence suggests that dimerization of retroviral RNA genomes occurs prior to or concomitant with viral encapsidation and budding (159). It appears that RNAs initially packaged into nascent virions are dimeric and the genomic RNA dimer undergoes a post-assembly, viral protease-dependent maturation process (50, 51, 178).

Due to the colocalization of the DIS domain with the encapsidation signal, it is difficult to discern whether dimerization and encapsidation are two independent processes. Mutagenesis studies of DIS revealed various degrees of RNA packaging defect. The level of genomic RNA packaging was found to be decreased 2- to 5-fold for DIS mutants (12, 29, 94, 133, 172). It appeared that the structures, but not the specific sequences, of the DIS stems are critical for RNA packaging (29, 62). The reduced genomic content of the mutant virions is associated with an increased content of spliced viral transcripts, suggesting DIS contributes to the discrimination between spliced and unspliced RNAs (29). Recently, a system was established in which the DIS/DLS region was duplicated in viral RNA. Such duplication of DIS/DLS signals appears not to affect RNA packaging and virion formation. Monomeric RNAs are preferentially packaged in virus particles, most likely due to intramolecular interaction between the two DIS/DLS regions on a single genomic RNA (165, 166). This system was used to map the location of DIS/DLS in the 5' region of the HIV-1 genome, presumably with minimum effect on RNA packaging. It was shown that the DIS lower stem was important for both

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dimerization and packaging, but deletion of the palindromic sequence at the top of the loop only partially affect dimerization (167). However, it is still not clear whether the observed interrelationship between dimerization and packaging is a mere consequence of the colocalization of the DIS domain with the encapsidation signal or these two processes are mechanistically coupled. It is important to note that when the DIS motif is crippled by mutagenesis, the augmented level of monomeric RNAs in virions (29, 59, 94, 172, 173) is usually accompanied by a significant decrease in the overall packaging efficiency of genomic RNAs (12, 29, 94, 133, 172). Such observation suggests that dimer formation is not absolutely required for RNA packaging but is required for *efficient* genomic RNA packaging.

Reverse transcription and other steps of viral replication. The second DNA strand transfer during reverse transcription was found to be severely impaired upon stem-loop deletion (133), and mildly reduced by 80% when the lower stem and the internal bulge of HIV-1 DIS were destroyed (172). Some studies suggested that the DIS deletion mutations moderately inhibit the level of viral gene expression (94, 99, 101). It was later demonstrated that infection with these DIS mutants had delayed processing of viral gag proteins and abnormal ultrastructure of progeny viral particles (100).

Viral Infectivity. Mutations in the DIS of infectious HIV-1 molecular clones dramatically reduce the replication rate (12, 29, 59, 133) and the infectivity of the resulting viruses was reduced 10- to 1000-fold (29, 94, 133). In general, deletion of DIS structure has a more profound effect on viral infectivity than loop substitutions (29, 94, 133, 177). These findings offer compelling evidence that genomic dimerization initiated by DIS is required for optimal infectivity of HIV-1 virions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Dimerization & recombination

The diploid genome of retroviruses provides the opportunity for switching from a damaged to a copackaged intact template during reverse transcription. However, whether copackaging is synonymous to dimer formation and how the physical linkage between dimerized RNAs contributes to template switching are not known.

Although viral RNAs are packaged as dimers, the ability of RNAs to heterodimerize is not absolutely required for copackaging. Simple retroviruses have provided several examples of recombinogenic patch repair by endogenous retroelements known to be incorporated into retroviral particles but which are unlikely to heterodimerize with virus genomes, and RNAs which presumably only occasionally and by chance become encapsidated can also recombine with viral genomes (121). Copackaged RNAs can recombine and form a functional provirus if at least one of the RNAs contains *cis*-acting sequences necessary for replication. It has been reported that encapsidated RNAs that lack retroviral *cis*-acting signals can be converted into DNA and integrated like proviruses in certain circumstances (102). However, although recombinants between RNAs that cannot dimerize have been observed, it has been assumed that these products arise only rarely.

One way by which DIS may contribute to retroviral recombination is to provide physical linkage between recombination donor and acceptor. DIS has been implicated in promoting recombination at the region flanking the dimerization initiation sequence in a forced recombination vector system for MLV (107, 119, 120), and *in vitro* strand transfer studies with HIV-1 subgenomic fragment (7, 8). Indirect evidence from melting profile of nicked genomic RNAs suggests that extensive RNA-RNA interactions exist at other regions of the RNA genome as well (132), and these may also contribute to intermolecular recombination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Inspection of the HIV sequence database indicates that most of the circulating recombinants have putative parents with an identical DIS palindrome (44). However, inter-subtype recombinants are detected in approximately 1% of the human peripheral blood mononuclear cells dually infected with pairs of HIV-1 isolates from different subtypes, some of which contain distinct DIS sequences (157). The capability of different HIV-1 isolates in heterodimerization has only been investigated using *in vitro* transcribed subgenomic fragments from different HIV-1 subtypes (1). This study suggests that complementarity between the palindromic sequences in the DIS hairpins is necessary and sufficient for *in vitro* dimerization of two subtype RNAs (1). DIS incompatibility has been suggested be one of the factors that limit inter-subtype recombination (44). However, recently, it was demonstrated in a cell-culture based recombination assay that the presence of non-homologous DIS elements from different HIV-1 subtypes is not a principal obstacle to inter-subtype recombination (177). Nevertheless, it is still not clear to what extent dimerization affects the efficiency of recombination. For example, it is not known whether recombination is more efficient between dimerized RNAs than that between two copackaged RNAs that are not dimerized.

OVERVIEW OF THE THESIS

In Chapter two, single-replication-cycle assays were established for HIV-1, and HIV-1 forward mutation rates and retroviral recombination-related properties of direct repeat deletion were examined by using HIV-1 vectors and assay conditions, which were highly similar to those previously employed for MLV. In agreement with previous observations from dissimilar assay systems for HIV-1 and MLV, the forward mutations for these two retroviral species were very similar using our *lacZ* inactivation assay.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

However, in contrast to reported large differences in intermolecular recombination frequencies between HIV-1 and MLV, we found that direct repeat deletion rates for HIV-1 and MLV were very similar.

The effects of systematic variation in the extents of sequence homology on recombination rates were also addressed in Chapter two using a repeat deletion system containing degenerate repeated sequences. A series of vectors in which the two repeats ranged from 100% identical to only 58% homologous were designed, which provided sequence variation similar to that of intra-subtype, inter-subtype, and inter-group HIV-1 strains. Results from these experiments indicated that the frequency of template switching decreased dramatically as sequence divergence between repeated sequences increased. For sequences that differed by about 25% or more, HIV-1 template switching directed by sequence homology appeared to be no more frequent than that which is homology independent.

The work in Chapter three was designed to directly compare recombination frequencies between HIV-1 and MLV. A series of single replication cycle assays were performed for both viruses using the same producer and recipient cell types for each virus, virus-specific replication proteins, and recombination assay vectors that differed only in virus-specific *cis*-acting sequences. Results from two-vector recombination assays and reciprocal repeat deletion experiments confirmed the reported differences in intermolecular recombination between HIV-1 and MLV. However, transduction-type recombination assays, which assured copackaging of recombination donor and acceptor sequences, indicated that template switching properties were very similar for HIV-1 and MLV. These results provide genetic evidence that coexpressed RNAs are less accessible to the recombination machinery during MLV replication than for HIV-1.

In Chapter four, the transduction potential of HIV-1 was investigated by transduction-type recombination assays using vectors reminiscent of retroviral transduction intermediates—readthrough RNAs that captured cellular sequence at their 3'

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ends. The acceptor sequence was genetically marked with scattered single point mutations that allowed the differentiation of donor from acceptor sequences by restriction analysis of individual proviral clones. Under such condition, readthrough transcripts arose at a rate up to 7% of the normal viral RNA. RNA sequences downstream of the 3' LTR acquired by polyadenylation signal readthrough could efficiently serve as templates for HIV-1 genetic recombination. The readthrough sequences downstream of the 3' LTR appeared to be preferentially used for HIV-1 recombination when compared to homologous sequences located between the 5' and 3' LTRs.

In Chapter five, the findings from this thesis work are summarized. Their implications and future directions are discussed.

BIBLIOGRAPHY

1. **Andersen, E. S., R. E. Jeeninga, C. K. Damgaard, B. Berkhout, and J. Kjems** 2003. Dimerization and template switching in the 5' untranslated region between various subtypes of human immunodeficiency virus type 1 J Virol. **77:**3020-30.

2. **Anderson, J. A., E. H. Bowman, and W. S. Hu** 1998. Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation J Virol. **72:**1195-202.

3. **Anderson, J. A., V. K. Pathak, and W. S. Hu** 2000. Effect of the murine leukemia virus extended packaging signal on the rates and locations of retroviral recombination J Virol. **74:**6953-63.

4. **Anderson, J. A., R. J. Teufel, 2nd, P. D. Yin, and W. S. Hu** 1998. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination J Virol. **72:**1186-94.

5. **Awang, G., and D. Sen** 1993. Mode of dimerization of HIV-1 genomic RNA Biochemistry. **32:**11453-7.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

6. **Ayouba, A., S. Souquieres, B. Njinku, P. M. Martin, M. C. Muller-Trutwin, P. Roques, F. Barre-Sinoussi, P. Mauclere, F. Simon, and E. Nerrienet** 2000. HIV-1 group N among HIV-1-seropositive individuals in Cameroon Aids. **14:**2623-5.

7. **Balakrishnan, M., P. J. Fay, and R. A. Bambara** 2001. The kissing hairpin sequence promotes recombination within the HIV-I 5' leader region J Biol Chem. **276:**36482-92.

8. **Balakrishnan, M., B. P. Roques, P. J. Fay, and R. A. Bambara** 2003. Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis J Virol. **77:**4710-21.

9. **Banks, J. D., B. O. Kealoha, and M. L. Linial** 1999. An Mpsi-containing heterologous RNA, but not env mRNA, is efficiently packaged into avian retroviral particles J Virol. **73:**8926-33.

10. **Bender, W., Y. H. Chien, S. Chattopadhyay, P. K. Vogt, M. B. Gardner, and N. Davidson** 1978. High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures J Virol. **25:**888-96.

11. **Bender, W., and N. Davidson** 1976. Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules Cell. **7:**595-607.

12. **Berkhout, B., and J. L. van Wamel** 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1 J Virol. **70:**6723-32.

13. **Berkowitz, R., J. Fisher, and S. P. Goff** 1996. RNA packaging Current Topics in Microbiology & Immunology. **214:**177-218.

14. **Bieth, E., C. Gabus, and J. L. Darlix** 1990. A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis in vitro Nucleic Acids Res. **18:**119-27.

15. **Bircher, L. A., J. C. Rigano, V. G. Ponferrada, and D. P. Wooley** 2002. High fidelity of homologous retroviral recombination in cell culture Arch Virol. **147:**1665-83.

16. **Blackard, J. T., D. E. Cohen, and K. H. Mayer** 2002. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences Clin Infect Dis. **34:**1108-14.

17. **Bourara, K., S. Litvak, and A. Araya** 2000. Generation of G-to-A and C-to-U changes in HIV-1 transcripts by RNA editing Science. **289:**1564-6.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

18. **Bowman, R. R., W. S. Hu, and V. K. Pathak** 1998. Relative rates of retroviral reverse transcriptase template switching during RNA- and DNA-dependent DNA synthesis J Virol. **72:**5198-206.

19. **Boyer, J. C., K. Bebenek, and T. A. Kunkel** 1992. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates Proc Natl Acad Sci U S A. **89:**6919-23.

20. **Brincat, J. L., J. K. Pfeiffer, and A. Telesnitsky** 2002. RNase H activity is required for high-frequency repeat deletion during Moloney murine leukemia virus replication J Virol. **76:**88-95.

21. **Buiser, R. G., J. J. DeStefano, L. M. Mallaber, P. J. Fay, and R. A. Bambara** 1991. Requirements for the catalysis of strand transfer synthesis by retroviral DNA polymerases J Biol Chem. **266:**13103-9.

22. **Butsch, M., and K. Boris-Lawrie** 2002. Destiny of unspliced retroviral RNA: ribosome and/or virion? J Virol. **76:**3089-94.

23. **Champoux, J. J.** 1993. Roles of ribonuclease H in reverse transcription, p. 103-117. *In* A. M. Skalka, and S. P. Goff (eds), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

24. **Charneau, P., A. M. Borman, C. Quillent, D. Guetard, S. Chamaret, J. Cohen, G. Remy, L. Montagnier, and F. Clavel** 1994. Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group Virology. **205:**247-53.

25. **Clavel, F., M. D. Hoggan, R. L. Willey, K. Strebel, M. A. Martin, and R. Repaske** 1989. Genetic recombination of human immunodeficiency virus J Virol. **63:**1455-9.

26. **Clever, J., C. Sassetti, and T. G. Parslow** 1995. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1 J Virol. **69:**2101-9.

27. **Clever, J. L., D. A. Eckstein, and T. G. Parslow** 1999. Genetic dissociation of the encapsidation and reverse transcription functions in the 5' R region of human immunodeficiency virus type 1 J Virol. **73:**101-9.

28. **Clever, J. L., D. Mirandar, Jr., and T. G. Parslow** 2002. RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome J Virol. **76:**12381-7.

29. **Clever, J. L., and T. G. Parslow** 1997. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation Journal of Virology. **71:**3407-14.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

30. **Clever, J. L., M. L. Wong, and T. G. Parslow** 1996. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA J Virol. **70:**5902-8.

31. **Coffin, J. M.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy [see comments] Science. **267:**483-9.

32. **Coffin, J. M.** 1996. Retroviridae: the viruses and their replication, p. 763-843. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (eds), Fundamental Virology, Third ed. Lippincott-Raven Publishers, Philadelphia, PA.

33. **Coffin, J. M.** 1982. Structure of retroviral genomes, p. 261-368. *In* R. A. Weiss, N. M. Teich, H. E. Varmus, and J. M. Coffin (eds), RNA Tumor Viruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

34. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses J Gen Virol. **42:**1-26.

35. **Cullen, B. R.** 2000. Nuclear RNA export pathways Mol Cell Biol. **20:**4181-7.

36. **Dang, Q., and W. S. Hu** 2001. Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication J Virol. **75:**809-20.

37. **Dardel, F., R. Marquet, C. Ehresmann, B. Ehresmann, and S. Blanquet** 1998. Solution studies of the dimerization initiation site of HIV-1 genomic RNA Nucleic Acids Res. **26:**3567-71.

38. **Darlix, J. L., C. Gabus, M. T. Nugeyre, F. Clavel, and F. Barre-Sinoussi** 1990. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1 J Mol Biol. **216:**689-99.

39. **Das, A. T., B. Klaver, B. I. Klasens, J. L. van Wamel, and B. Berkhout** 1997. A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication J Virol. **71:**2346-56.

40. **Delviks, K. A., W. S. Hu, and V. K. Pathak** 1997. Psi- vectors: murine leukemia virus-based self-inactivating and self- activating retroviral vectors J Virol. **71:**6218-24.

41. **Delviks, K. A., and V. K. Pathak** 1999. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching J Virol. **73:**7923-32.

42. **DeStefano, J. J., R. A. Bambara, and P. J. Fay** 1994. The mechanism of human immunodeficiency virus reverse transcriptase- catalyzed strand transfer from internal regions of heteropolymeric RNA templates J Biol Chem. **269:**161-8.

43. **DeStefano, J. J., A. Raja, and J. V. Cristofaro** 2000. In vitro strand transfer from broken RNAs results in mismatch but not frameshift mutations Virology. **276:**7-15.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

44. **Dirac, A. M., H. Huthoff, J. Kjems, and B. Berkhout** 2002. Requirements for RNA heterodimerization of the human immunodeficiency virus type 1 (HIV-1) and HIV-2 genomes J Gen Virol. **83:**2533-42.

45. **Dougherty, J. P., and H. M. Temin** 1988. Determination of the rate of base-pair substitution and insertion mutations in retrovirus replication J Virol. **62:**2817-22.

46. **Drake, J. W.** 1991. A constant rate of spontaneous mutation in DNA-based microbes Proc Natl Acad Sci U S A. **88:**7160-4.

47. **Ennifar, E., P. Walter, B. Ehresmann, C. Ehresmann, and P. Dumas** 2001. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site Nat Struct Biol. **8:**1064-8.

48. **Felder, M. P., D. Laugier, B. Yatsula, P. Dezelee, G. Calothy, and M. Marx** 1994. Functional and biological properties of an avian variant long terminal repeat containing multiple A to G conversions in the U3 sequence J Virol. **68:**4759-67.

49. **Frost, S. D., M. J. Dumaurier, S. Wain-Hobson, and A. J. Brown** 2001. Genetic drift and within-host metapopulation dynamics of HIV-1 infection Proc Natl Acad Sci U S A. **98:**6975-80.

50. **Fu, W., R. J. Gorelick, and A. Rein** 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions J Virol. **68:**5013-8.

51. **Fu, W., and A. Rein** 1993. Maturation of dimeric viral RNA of Moloney murine leukemia virus J Virol. **67:**5443-9.

52. **Gao, W. Y., A. Cara, R. C. Gallo, and F. Lori** 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:**8925-8.

53. **Girard, F., F. Barbault, C. Gouyette, T. Huynh-Dinh, J. Paoletti, and G. Lancelot** 1999. Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex J Biomol Struct Dyn. **16:**1145-57.

54. **Gojobori, T., and S. Yokoyama** 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues Proc Natl Acad Sci U S A. **82:**4198-201.

55. **Gratton, S., R. Cheynier, M. J. Dumaurier, E. Oksenhendler, and S. Wain-Hobson** 2000. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers Proc Natl Acad Sci U S A. **97:**14566-71.

56. **Groenink, M., A. C. Andeweg, R. A. Fouchier, S. Broersen, R. C. van der Jagt, H. Schuitemaker, R. E. de Goede, M. L. Bosch, H. G. Huisman, and M. Tersmette** 1992. Phenotype-associated env gene variation among eight related human immunodeficiency virus type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain J Virol. **66:**6175-80.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

57. **Gunthard, H. F., A. J. Leigh-Brown, R. T. D'Aquila, V. A. Johnson, D. R. Kuritzkes, D. D. Richman, J. K. Wong, and T. D'Aquila R** 1999. Higher selection pressure from antiretroviral drugs in vivo results in increased evolutionary distance in HIV-1 pol Virology. **259:**154-65.

58. **Gurtler, L. G., P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, and L. Kaptue** 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon J Virol. **68:**1581-5.

59. **Haddrick, M., A. L. Lear, A. J. Cann, and S. Heaphy** 1996. Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1 J Mol Biol. **259:**58-68.

60. **Hajjar, A. M., and M. L. Linial** 1995. Modification of retroviral RNA by double-stranded RNA adenosine deaminase J Virol. **69:**5878-82.

61. **Harrison, G. P., and A. M. Lever** 1992. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure J Virol. **66:**4144-53.

62. **Harrison, G. P., G. Miele, E. Hunter, and A. M. Lever** 1998. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line J Virol. **72:**5886-96.

63. **Helga-Maria, C., M. L. Hammarskjold, and D. Rekosh** 1999. An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA J Virol. **73:**4127-35.

64. **Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz** 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection Nature. **373:**123-6.

65. **Hoglund, S., A. Ohagen, J. Goncalves, A. T. Panganiban, and D. Gabuzda** 1997. Ultrastructure of HIV-1 genomic RNA Virology. **233:**271-9.

66. **Hu, W. S., E. H. Bowman, K. A. Delviks, and V. K. Pathak** 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference J Virol. **71:**6028-36.

67. **Hu, W. S., and H. M. Temin** 1992. Effect of gamma radiation on retroviral recombination J Virol. **66:**4457-63.

68. **Hu, W. S., and H. M. Temin** 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:**1556-60.

69. **Hu, W. S., and H. M. Temin** 1990. Retroviral recombination and reverse transcription Science. **250:**1227-33.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

70. **Hwang, C. K., E. S. Svarovskaia, and V. K. Pathak** 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:**12209-14.

71. **Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome J Virol. **74:**1234-40.

72. **Jones, J. S., R. W. Allan, and H. M. Temin** 1993. Alteration of location of dimer linkage sequence in retroviral RNA: little effect on replication or homologous recombination J Virol. **67:**3151-8.

73. **Jones, J. S., R. W. Allan, and H. M. Temin** 1994. One retroviral RNA is sufficient for synthesis of viral DNA J Virol. **68:**207-16.

74. **Jossinet, F., J. C. Paillart, E. Westhof, T. Hermann, E. Skripkin, J. S. Lodmell, C. Ehresmann, B. Ehresmann, and R. Marquet** 1999. Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding Rna. **5:**1222-34.

75. **Julias, J. G., D. Hash, and V. K. Pathak** 1995. E- vectors: development of novel self-inactivating and self-activating retroviral vectors for safer gene therapy J Virol. **69:**6839-46.

76. **Julias, J. G., T. Kim, G. Arnold, and V. K. Pathak** 1997. The antiretrovirus drug 3'-azido-3'-deoxythymidine increases the retrovirus mutation rate J Virol. **71:**4254-63.

77. **Julias, J. G., and V. K. Pathak** 1998. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate J Virol. **72:**7941-9.

78. **Jung, A., R. Maier, J. P. Vartanian, G. Bocharov, V. Jung, U. Fischer, E. Meese, S. Wain-Hobson, and A. Meyerhans** 2002. Multiply infected spleen cells in HIV patients Nature. **418:**144.

79. **Junghans, R. P., L. R. Boone, and A. M. Skalka** 1982. Retroviral DNA H structures: displacement-assimilation model of recombination Cell. **30:**53-62.

80. **Kandel, E. S., and E. Nudler** 2002. Template switching by RNA polymerase II in vivo. Evidence and implications from a retroviral system Mol Cell. **10:**1495-502.

81. **Katz, R. A., and A. M. Skalka** 1990. Generation of diversity in retroviruses Annu Rev Genet. **24:**409-45.

82. **Kellam, P., and B. A. Larder** 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance J Virol. **69:**669-74.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

83. **Kim, T., R. A. Mudry, Jr., C. A. Rexrode, 2nd, and V. K. Pathak** 1996. Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication J Virol. **70:**7594-602.

84. **Kinoshita, K., and T. Honjo** 2001. Linking class-switch recombination with somatic hypermutation Nat Rev Mol Cell Biol. **2:**493-503.

85. **Kuiken, C., B. Foley, B. Hahn, P. Marx, F. McCutchan, J. Mellors, S. Wolinksy, and B. Korber** (eds.) 2001 HIV Sequence Compendium. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

86. **Kulpa, D., R. Topping, and A. Telesnitsky** 1997. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors Embo J. **16:**856-65.

87. **Kumar, M., B. Keller, N. Makalou, and R. E. Sutton** 2001. Systematic determination of the packaging limit of lentiviral vectors Hum Gene Ther. **12:**1893-905.

88. **Kung, H. J., J. M. Bailey, N. Davidson, M. O. Nicolson, and R. M. McAllister** 1975. Structure, subunit composition, and molecular weight of RD-114 RNA J Virol. **16:**397-411.

89. **Kung, H. J., S. Hu, W. Bender, J. M. Bailey, N. Davidson, M. O. Nicolson, and R. M. McAllister** 1976. RD-114, baboon, and woolly monkey viral RNA's compared in size and structure Cell. **7:**609-20.

90. **Laughrea, M., and L. Jette** 1994. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA Biochemistry. **33:**13464-74.

91. **Laughrea, M., and L. Jette** 1996. HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain Biochemistry. **35:**9366-74.

92. **Laughrea, M., and L. Jette** 1997. HIV-1 genome dimerization: kissing-loop hairpin dictates whether nucleotides downstream of the 5' splice junction contribute to loose and tight dimerization of human immunodeficiency virus RNA Biochemistry. **36:**9501-8.

93. **Laughrea, M., and L. Jette** 1996. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation Biochemistry. **35:**1589-98.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

94. **Laughrea, M., L. Jette, J. Mak, L. Kleiman, C. Liang, and M. A. Wainberg** 1997. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization J Virol. **71:**3397-406.

95. **Lercher, M. J., and L. D. Hurst** 2002. Human SNP variability and mutation rate are higher in regions of high recombination Trends Genet. **18:**337-40.

96. **Li, T., and J. Zhang** 2000. Determination of the frequency of retroviral recombination between two identical sequences within a provirus J Virol. **74:**7646-50.

97. **Li, T., and J. Zhang** 2002. Intramolecular recombinations of Moloney murine leukemia virus occur during minus-strand DNA synthesis J Virol. **76:**9614-23.

98. **Li, Y., and S. Carpenter** 2001. Cis-acting sequences may contribute to size variation in the surface glycoprotein of bovine immunodeficiency virus J Gen Virol. **82:**2989-98.

99. **Liang, C., X. Li, Y. Quan, M. Laughrea, L. Kleiman, J. Hiscott, and M. A. Wainberg** 1997. Sequence elements downstream of the human immunodeficiency virus type 1 long terminal repeat are required for efficient viral gene transcription J Mol Biol. **272:**167-77.

100. **Liang, C., L. Rong, E. Cherry, L. Kleiman, M. Laughrea, and M. A. Wainberg** 1999. Deletion mutagenesis within the dimerization initiation site of human immunodeficiency virus type 1 results in delayed processing of the p2 peptide from precursor proteins J Virol. **73:**6147-51.

101. **Liang, C., L. Rong, M. Laughrea, L. Kleiman, and M. A. Wainberg** 1998. Compensatory point mutations in the human immunodeficiency virus type 1 Gag region that are distal from deletion mutations in the dimerization initiation site can restore viral replication J Virol. **72:**6629-36.

102. **Linial, M., E. Medeiros, and W. S. Hayward** 1978. An avian oncovirus mutant (SE 21Q1b) deficient in genomic RNA: biological and biochemical characterization Cell. **15:**1371-81.

103. **Lodmell, J. S., C. Ehresmann, B. Ehresmann, and R. Marquet** 2000. Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site Rna. **6:**1267-76.

104. **Lodmell, J. S., C. Ehresmann, B. Ehresmann, and R. Marquet** 2001. Structure and dimerization of HIV-1 kissing loop aptamers J Mol Biol. **311:**475-90.

105. **Loussert-Ajaka, I., M. L. Chaix, B. Korber, F. Letourneur, E. Gomas, E. Allen, T. D. Ly, F. Brun-Vezinet, F. Simon, and S. Saragosti** 1995. Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France J Virol. **69:**5640-9.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

106. **Lovett, S. T., P. T. Drapkin, V. A. Sutera, and T. J. Gluckman-Peskind** 1993. A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in Escherichia coli Genetics. **135:**631-42.

107. **Lund, A. H., J. G. Mikkelsen, J. Schmidt, M. Duch, and F. S. Pedersen** 1999. The kissing-loop motif is a preferred site of 5' leader recombination during replication of SL3-3 murine leukemia viruses in mice J Virol. **73:**9614-8.

108. **Maisel, J., W. Bender, S. Hu, P. H. Duesberg, and N. Davidson** 1978. Structure of 50 to 70S RNA from Moloney sarcoma viruses J Virol. **25:**384-94.

109. **Malim, M. H., and M. Emerman** 2001. HIV-1 sequence variation: drift, shift, and attenuation Cell. **104:**469-72.

110. **Mangel, W. F., H. Delius, and P. H. Duesberg** 1974. Structure and molecular weight of the 60-70S RNA and the 30-40S RNA of the Rous sarcoma virus Proc Natl Acad Sci U S A. **71:**4541-5.

111. **Mansky, L. M.** 2000. In vivo analysis of human T-cell leukemia virus type 1 reverse transcription accuracy J Virol. **74:**9525-31.

112. **Mansky, L. M., and H. M. Temin** 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:**5087-94.

113. **Mansky, L. M., and H. M. Temin** 1994. Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus J Virol. **68:**494-9.

114. **Marquet, R., F. Baudin, C. Gabus, J. L. Darlix, M. Mougel, C. Ehresmann, and B. Ehresmann** 1991. Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism Nucleic Acids Res. **19:**2349-57.

115. **Marquet, R., J. C. Paillart, E. Skripkin, C. Ehresmann, and B. Ehresmann** 1994. Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site Nucleic Acids Res. **22:**145-51.

116. **McBride, M. S., and A. T. Panganiban** 1996. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures J Virol. **70:**2963-73.

117. **McBride, M. S., and A. T. Panganiban** 1997. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo J Virol. **71:**2050-8.

118. **McBride, M. S., M. D. Schwartz, and A. T. Panganiban** 1997. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation J Virol. **71:**4544-54.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

119. **Mikkelsen, J. G., A. H. Lund, M. Duch, and F. S. Pedersen** 2000. Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo J Virol. **74:**600-10.

120. **Mikkelsen, J. G., A. H. Lund, M. Duch, and F. S. Pedersen** 1998. Recombination in the 5' leader of murine leukemia virus is accurate and influenced by sequence identity with a strong bias toward the kissing- loop dimerization region J Virol. **72:**6967-78.

121. **Mikkelsen, J. G., and F. S. Pedersen** 2000. Genetic reassortment and patch repair by recombination in retroviruses J Biomed Sci. **7:**77-99.

122. **Morris, A., M. Marsden, K. Halcrow, E. S. Hughes, R. P. Brettle, J. E. Bell, and P. Simmonds** 1999. Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain: evidence for frequent in vivo recombination events in the evolution of regional populations J Virol. **73:**8720-31.

123. **Mouland, A. J., J. Mercier, M. Luo, L. Bernier, L. DesGroseillers, and E. A. Cohen** 2000. The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation J Virol. **74:**5441-51.

124. **Moutouh, L., J. Corbeil, and D. D. Richman** 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:**6106-11.

125. **Mujeeb, A., J. L. Clever, T. M. Billeci, T. L. James, and T. G. Parslow** 1998. Structure of the dimer initiation complex of HIV-1 genomic RNA Nat Struct Biol. **5:**432-6.

126. **Mujeeb, A., T. G. Parslow, A. Zarrinpar, C. Das, and T. L. James** 1999. NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA FEBS Lett. **458:**387-92.

127. **Mules, E. H., O. Uzun, and A. Gabriel** 1998. Replication errors during in vivo Ty1 transposition are linked to heterogeneous RNase H cleavage sites Mol Cell Biol. **18:**1094-104.

128. **Muriaux, D., P. M. Girard, B. Bonnet-Mathoniere, and J. Paoletti** 1995. Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide J Biol Chem. **270:**8209-16.

129. **Murti, K. G., M. Bondurant, and A. Tereba** 1981. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy J Virol. **37:**411-19.

130. **Negroni, M., and H. Buc** 2001. Retroviral recombination: what drives the switch? Nat Rev Mol Cell Biol. **2:**151-5.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

131. **O'Neil, P. K., G. Sun, H. Yu, Y. Ron, J. P. Dougherty, and B. D. Preston** 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:**38053-61.

132. **Ortiz-Conde, B. A., and S. H. Hughes** 1999. Studies of the genomic RNA of leukosis viruses: implications for RNA dimerization J Virol. **73:**7165-74.

133. **Paillart, J. C., L. Berthoux, M. Ottmann, J. L. Darlix, R. Marquet, B. Ehresmann, and C. Ehresmann** 1996. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis J Virol. **70:**8348-54.

134. **Paillart, J. C., R. Marquet, E. Skripkin, B. Ehresmann, and C. Ehresmann** 1994. Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA J Biol Chem. **269:**27486-93.

135. **Paillart, J. C., E. Skripkin, B. Ehresmann, C. Ehresmann, and R. Marquet** 1996. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA Proc Natl Acad Sci U S A. **93:**5572-7.

136. **Paillart, J. C., E. Westhof, C. Ehresmann, B. Ehresmann, and R. Marquet** 1997. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA J Mol Biol. **270:**36-49.

137. **Panganiban, A. T., and D. Fiore** 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription Science. **241:**1064-9.

138. **Parthasarathi, S., A. Varela-Echavarria, Y. Ron, B. D. Preston, and J. P. Dougherty** 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications J Virol. **69:**7991-8000.

139. **Patel, P. H., and B. D. Preston** 1994. Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends Proc Natl Acad Sci U S A. **91:**549-53.

140. **Pathak, V. K., and H. M. Temin** 1992. 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate J Virol. **66:**3093-100.

141. **Pathak, V. K., and H. M. Temin** 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:**6024-8.

142. **Pathak, V. K., and H. M. Temin** 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:**6019-23.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

143. **Peeters, M.** 2000. Recombinant HIV sequences: Their role in the global epidemic, p. I-39-54. *In* C. L. Kuiken, B. Foley, B. Hahn, B. Korber, F. McCutchan, P. A. Marx, J. W. Mellors, J. I. Mullins, J. Sodroski, and S. Wolinksy (eds), Human Retroviruses and AIDS 2000. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

144. **Peeters, M., F. Liegeois, N. Torimiro, A. Bourgeois, E. Mpoudi, L. Vergne, E. Saman, E. Delaporte, and S. Saragosti** 1999. Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient J Virol. **73:**7368-75.

145. **Peeters, M., and P. M. Sharp** 2000. Genetic diversity of HIV-1: the moving target Aids. **14:**S129-40.

146. **Peliska, J. A., and S. J. Benkovic** 1994. Fidelity of in vitro DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase Biochemistry. **33:**3890-5.

147. **Peliska, J. A., and S. J. Benkovic** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase Science. **258:**1112-8.

148. **Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time Science. **271:**1582-6.

149. **Pfeiffer, J. K., M. M. Georgiadis, and A. Telesnitsky** 2000. Structure-based moloney murine leukemia virus reverse transcriptase mutants with altered intracellular direct-repeat deletion frequencies J Virol. **74:**9629-36.

150. **Pfeiffer, J. K., and A. Telesnitsky** 2001. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication J Virol. **75:**11263-74.

151. **Pfeiffer, J. K., R. S. Topping, N. H. Shin, and A. Telesnitsky** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription J Virol. **73:**8441-7.

152. **Pierson, T. C., Y. Zhou, T. L. Kieffer, C. T. Ruff, C. Buck, and R. F. Siliciano** 2002. Molecular characterization of preintegration latency in human immunodeficiency virus type 1 infection J Virol. **76:**8518-31.

153. **Prats, A. C., C. Roy, P. A. Wang, M. Erard, V. Housset, C. Gabus, C. Paoletti, and J. L. Darlix** 1990. cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA J Virol. **64:**774-83.

154. **Preston, B. D.** 1997. Reverse transcriptase fidelity and HIV-1 variation Science. **275:**228-9; discussion 230-1.

155. **Preston, B. D., and J. P. Dougherty** 1996. Mechanisms of retroviral mutation Trends Microbiol. **4:**16-21.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

156. **Quinones-Mateu, M. E., and E. J. Arts** 1999. Recombination in HIV-1: update and implications AIDS Rev. **1:**89-100.

157. **Quinones-Mateu, M. E., S. C. Ball, A. J. Marozsan, V. S. Torre, J. L. Albright, G. Vanham, G. van Der Groen, R. L. Colebunders, and E. J. Arts** 2000. A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression J Virol. **74:**9222-33.

158. **Ramos, A., D. J. Hu, L. Nguyen, K. O. Phan, S. Vanichseni, N. Promadej, K. Choopanya, M. Callahan, N. L. Young, J. McNicholl, T. D. Mastro, T. M. Folks, and S. Subbarao** 2002. Intersubtype human immunodeficiency virus type 1 superinfection following seroconversion to primary infection in two injection drug users J Virol. **76:**7444-52.

159. **Rein, A.** 1994. Retroviral RNA packaging: a review Arch Virol Suppl. **9:**513-22.

160. **Rhode, B. W., M. Emerman, and H. M. Temin** 1987. Instability of large direct repeats in retrovirus vectors J Virol. **61:**925-7.

161. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber** 1999. HIV-1 Nomenclature Proposal, p. 492-505. *In* C. L. Kuiken, B. Foley, B. Hahn, B. Korber, F. McCutchan, P. A. Marx, J. W. Mellors, J. I. Mullins, J. Sodroski, and S. Wolinksy (eds), Human Retroviruses and AIDS 1999. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

162. **Russell, R. S., J. Hu, V. Beriault, A. J. Mouland, M. Laughrea, L. Kleiman, M. A. Wainberg, and C. Liang** 2003. Sequences downstream of the 5' splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA J Virol. **77:**84-96.

163. **Russell, R. S., J. Hu, M. Laughrea, M. A. Wainberg, and C. Liang** 2002. Deficient dimerization of human immunodeficiency virus type 1 RNA caused by mutations of the u5 RNA sequences Virology. **303:**152-63.

164. **Sakalian, M., J. W. Wills, and V. M. Vogt** 1994. Efficiency and selectivity of RNA packaging by Rous sarcoma virus Gag deletion mutants J Virol. **68:**5969-81.

165. **Sakuragi, J., A. Iwamoto, and T. Shioda** 2002. Dissociation of genome dimerization from packaging functions and virion maturation of human immunodeficiency virus type 1 J Virol. **76:**959-67.

166. **Sakuragi, J., T. Shioda, and A. T. Panganiban** 2001. Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions J Virol. **75:**2557-65.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

167. **Sakuragi, J., S. Ueda, A. Iwamoto, and T. Shioda** 2003. Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging J Virol. 77:4060-9.

168. **Sakuragi, J. I., and A. T. Panganiban** 1997. Human immunodeficiency virus type 1 RNA outside the primary encapsidation and dimer linkage region affects RNA dimer stability in vivo Journal of Virology. 71:3250-4.

169. **Seillier-Moiseiwitsch, F., B. H. Margolin, and R. Swanstrom** 1994. Genetic variability of the human immunodeficiency virus: statistical and biological issues Annu Rev Genet. 28:559-96.

170. **Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X. L. Huang, and J. I. Mullins** 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection J Virol. 73:10489-502.

171. **Shaw, R. J., N. D. Bonawitz, and D. Reines** 2002. Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast J Biol Chem. 277:24420-6.

172. **Shen, N., L. Jette, C. Liang, M. A. Wainberg, and M. Laughrea** 2000. Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging J Virol. 74:5729-35.

173. **Shen, N., L. Jette, M. A. Wainberg, and M. Laughrea** 2001. Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization J Virol. 75:10543-9.

174. **Shin, N. H., D. Hartigan-O'Connor, J. K. Pfeiffer, and A. Telesnitsky** 2000. Replication of lengthened Moloney murine leukemia virus genomes is impaired at multiple stages J Virol. 74:2694-702.

175. **Simon, F., P. Mauclere, P. Roques, I. Loussert-Ajaka, M. C. Muller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barre-Sinoussi, and F. Brun-Vezinet** 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O Nat Med. 4:1032-7.

176. **Skripkin, E., J. C. Paillart, R. Marquet, B. Ehresmann, and C. Ehresmann** 1994. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro Proc Natl Acad Sci U S A. 91:4945-9.

177. **St Louis, D. C., D. Gotte, E. Sanders-Buell, D. W. Ritchey, M. O. Salminen, J. K. Carr, and F. E. McCutchan** 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. 72:3991-8.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

178. **Stoltzfus, C. M., and P. N. Snyder** 1975. Structure of B77 sarcoma virus RNA: stabilization of RNA after packaging J Virol. **16:**1161-70.

179. **Sundquist, W. I., and S. Heaphy** 1993. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA Proc Natl Acad Sci U S A. **90:**3393-7.

180. **Svarovskaia, E. S., K. A. Delviks, C. K. Hwang, and V. K. Pathak** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching J Virol. **74:**7171-8.

181. **Swanstrom, R., J. M. Bishop, and H. E. Varmus** 1982. Structure of a replication intermediate in the synthesis of Rous sarcoma virus DNA in vivo J Virol. **42:**337-41.

182. **Takehisa, J., L. Zekeng, E. Ido, Y. Yamaguchi-Kabata, I. Mboudjeka, Y. Harada, T. Miura, L. Kaptu, and M. Hayami** 1999. Human immunodeficiency virus type 1 intergroup (M/O) recombination in cameroon J Virol. **73:**6810-20.

183. **Telesnitsky, A., and S. P. Goff** 1997. Reverse transcriptase and the generation of retroviral DNA, p. 121-160. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

184. **Temin, H. M.** 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation Proc Natl Acad Sci U S A. **90:**6900-3.

185. **Thomas, M. J., A. A. Platas, and D. K. Hawley** 1998. Transcriptional fidelity and proofreading by RNA polymerase II Cell. **93:**627-37.

186. **van Wamel, J. L., and B. Berkhout** 1998. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly Virology. **244:**245-51.

187. **Vanden Haesevelde, M., J. L. Decourt, R. J. De Leys, B. Vanderborght, G. van der Groen, H. van Heuverswijn, and E. Saman** 1994. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate J Virol. **68:**1586-96.

188. **Varela-Echavarria, A., N. Garvey, B. D. Preston, and J. P. Dougherty** 1992. Comparison of Moloney murine leukemia virus mutation rate with the fidelity of its reverse transcriptase in vitro J Biol Chem. **267:**24681-8.

189. **Varela-Echavarria, A., C. M. Prorock, Y. Ron, and J. P. Dougherty** 1993. High rate of genetic rearrangement during replication of a Moloney murine leukemia virus-based vector J Virol. **67:**6357-64.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

190. **Vogt, V. M.** 1997. Retroviral virions and genomes, p. 27-69. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

191. **Wabl, M., P. D. Burrows, A. von Gabain, and C. Steinberg** 1985. Hypermutation at the immunoglobulin heavy chain locus in a pre-B-cell line Proc Natl Acad Sci U S A. **82:**479-82.

192. **Wang, T. H., Y. K. Donaldson, R. P. Brettle, J. E. Bell, and P. Simmonds** 2001. Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system J Virol. **75:**11686-99.

193. **Wattel, E., J. P. Vartanian, C. Pannetier, and S. Wain-Hobson** 1995. Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy J Virol. **69:**2863-8.

194. **Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, and et al.** 1995. Viral dynamics in human immunodeficiency virus type 1 infection Nature. **373:**117-22.

195. **Weiss, R. A.** 1982. Experimental biology and assay of RNA tumor viruses, p. 209-260. *In* R. A. Weiss, N. M. Teich, H. E. Varmus, and J. M. Coffin (eds), RNA Tumor Viruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

196. **Weiss, R. A., W. S. Mason, and P. K. Vogt** 1973. Genetic recombinants and heterozygotes derived from endogenous and exogenous avian RNA tumor viruses Virology. **52:**535-52.

197. **Weiss, S., G. Hausl, M. Famulok, and B. Konig** 1993. The multimerization state of retroviral RNA is modulated by ammonium ions and affects HIV-1 full-length cDNA synthesis in vitro Nucleic Acids Res. **21:**4879-85.

198. **Wilhelm, M., M. Boutabout, T. Heyman, and F. X. Wilhelm** 1999. Reverse transcription of the yeast Ty1 retrotransposon: the mode of first strand transfer is either intermolecular or intramolecular J Mol Biol. **288:**505-10.

199. **Wooley, D. P., L. A. Bircher, and R. A. Smith** 1998. Retroviral recombination is nonrandom and sequence dependent Virology. **243:**229-34.

200. **Wooley, D. P., R. A. Smith, S. Czajak, and R. C. Desrosiers** 1997. Direct demonstration of retroviral recombination in a rhesus monkey J Virol. **71:**9650-3.

201. **Wrobel, J. A., S. F. Chao, M. J. Conrad, J. D. Merker, R. Swanstrom, G. J. Pielak, and C. A. Hutchison, 3rd** 1998. A genetic approach for identifying critical residues in the fingers and palm subdomains of HIV-1 reverse transcriptase Proc Natl Acad Sci U S A. **95:**638-45.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

202. **Wyke, J. A., and J. A. Beamand** 1979. Genetic recombination in Rous sarcoma virus: the genesis of recombinants and lack of evidence for linkage between pol, env and src genes in three factor crosses J Gen Virol. **43:**349-64.

203. **Wyke, J. A., J. G. Bell, and J. A. Beamand** 1975. Genetic recombination among temperature-sensitive mutnats of Rous sarcoma virus Cold Spring Harb Symp Quant Biol. **39:**897-905.

204. **Xu, H., and J. D. Boeke** 1987. High-frequency deletion between homologous sequences during retrotransposition of Ty elements in Saccharomyces cerevisiae Proc Natl Acad Sci U S A. **84:**8553-7.

205. **Yu, H., A. E. Jetzt, and J. P. Dougherty** 1997. Use of single-cycle analysis to study rates and mechanisms of retroviral mutation Methods. **12:**325-36.

206. **Yu, H., A. E. Jetzt, Y. Ron, B. D. Preston, and J. P. Dougherty** 1998. The nature of human immunodeficiency virus type 1 strand transfers J Biol Chem. **273:**28384-91.

207. **Zack, J. A., S. J. Arrigo, S. R. Weitsman, A. S. Go, A. Haislip, and I. S. Chen** 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure Cell. **61:**213-22.

208. **Zhang, J., and Y. Ma** 2001. Evidence for retroviral intramolecular recombinations J Virol. **75:**6348-58.

209. **Zhang, J., and C. M. Sapp** 1999. Recombination between two identical sequences within the same retroviral RNA molecule J Virol. **73:**5912-7.

210. **Zhang, J., L. Y. Tang, T. Li, Y. Ma, and C. M. Sapp** 2000. Most retroviral recombinations occur during minus-strand DNA synthesis J Virol. **74:**2313-22.

211. **Zhang, J., and H. M. Temin** 1993. 3' junctions of oncogene-virus sequences and the mechanisms for formation of highly oncogenic retroviruses J Virol. **67:**1747-51.

212. **Zhang, J., and H. M. Temin** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication Science. **259:**234-8.

213. **Zhang, J., and H. M. Temin** 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone J Virol. **68:**2409-14.

214. **Zhang, W. H., E. S. Svarovskaia, R. Barr, and V. K. Pathak** 2002. Y586F mutation in murine leukemia virus reverse transcriptase decreases fidelity of DNA synthesis in regions associated with adenine-thymine tracts Proc Natl Acad Sci U S A. **99:**10090-5.

215. **Zhuang, J., A. E. Jetzt, G. Sun, H. Yu, G. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty** 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots J Virol. **76:**11273-82.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER TWO

FREQUENCY OF DIRECT REPEAT DELETION AND EFFECTS OF VARYING SEQUENCE SIMILARITY ON FREQUENCY OF DIRECT REPEAT DELETION DURING HIV-1 REVERSE TRANSCRIPTION

INTRODUCTION

Cumulative evidence suggests that, like other members of the *Retroviridae* family, HIV-1 reverse transcription is prone to both nucleotide misincorporation (mutations) and genetic recombination. Although the steady state composition of retroviral populations is determined more by selection than by error frequency, the high rate of errors associated with HIV-1 replication is an important factor in HIV-1 disease because it provides the genetic diversity from which variants can be selected.

Many studies with purified enzymes have demonstrated that HIV-1 reverse transcriptase base misincorporation rates are high, and that this kind of mutation also arises frequently during viral replication (44). Although the rate of misincorporation during HIV-1 replication is much higher than that of the cellular replication machinery, it is about 20-fold lower than values obtained for purified HIV-1 RT (27). Most studies suggest that misincorporation rates during lentivirus replication are similar to or somewhat higher than those for simple retroviruses (25, 28, 38, 39). For example, in a study that used *lacZα* as a reporter, a forward mutation rate of 3.4×10^{-5} mutations per

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

base per HIV-1 replication cycle was observed (27). This is 2-fold higher than values obtained during spleen necrosis virus (SNV) replication (38, 39).

Some studies suggest that genetic recombination may be a more significant source of retroviral phenotypic variation than base misincorporation, especially for HIV-1 (25, 36). Retroviral genetic recombination results from template switching between copackaged RNAs (12, 13, 51). One approach to the study of retroviral recombination is to introduce genetic markers into two different retroviral vectors and to examine the progeny proviral DNA from virions containing these coexpressed RNAs (58). By this approach, the recombination frequency of HIV-1 was determined to be 2-3 recombination events per genome per replication cycle (14). This is roughly 10- fold higher than rates determined using similar approaches for SNV (12) and murine leukemia virus (MLV; (1)). Frequencies of homologous (between identical donor and acceptor sequences) and non-homologous recombination (between unrelated sequences) have been compared in this way for gammaretroviruses, and it was determined that homologous recombination is at least 10^3 times more frequent than non-homologous recombination (62).

Another approach for studying retroviral recombination-related properties is to measure direct repeat deletion. Repeated sequences are deleted frequently during retroviral replication (46), and deletion rates have been measured previously for SNV and MLV (7, 8, 16, 39, 43). Though not an absolute rule, MLV and SNV deletion frequencies correlate roughly with repeat length (7, 8, 16, 43, 63). Repeat deletion rates have not been reported for HIV-1.

Among global isolates, the sequence divergence of HIV-1 viruses is great. Based on phylogenetic analysis, three distinct groups of HIV-1 (termed M, N, and O) have been identified, with group M further divided into at least nine subtypes (A-D, F-H, J, and K) (48). Among isolates within a given subtype, extents of sequence variation are relatively modest (0-15% in *env* nucleotide sequence). Different subtypes differ from each other by roughly 15 to 30% in *env* nucleotide sequence (45). Significant sequence divergence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

exists between two groups: for example, group M and O viruses' homology in various genome regions ranges from 55%-76%, and is about 65% on average (53).

Despite this variation, the frequent observation of natural recombinants suggests that recombination is a relatively common occurrence among different strains of HIV-1 (41, 45). The majority of HIV-1 recombinants characterized thus far contain sequences from members of different group M subtypes. Some of these are classified as circulating recombinant forms and have been found to contribute to the global HIV-1 pandemic (41, 48). Intra-subtype recombinants have been identified within individual patients and may be more common than has been recognized due to the lower sequence divergence among parental strains. Recombination can apparently occur even among distantly related strains, because a limited number of inter-group M/O recombinants have been reported (40, 53). However, whether the generation of viable in-frame M/O recombinants is a chance event detectable only after further virus replication, or how the frequency of recombination is affected by sequence divergence such as that among natural HIV-1 isolates, has not been determined.

Here, we established a system for analyzing HIV-1 mutation and the recombination-related property of direct repeat deletion during single cycles of viral replication. In contrast to reported large differences in intermolecular recombination frequencies between HIV-1 and certain simple retroviruses, we found that mutation rates and direct repeat deletion rates for HIV-1 and MLV were very similar. We also addressed the effects of systematic variation in the extents of sequence homology on recombination rates, using a repeat deletion system. We designed a series of vectors in which the two repeats ranged from 100% identical to only 58% homologous, which provided sequence variation similar to that of intra-subtype, inter-subtype, and inter-group HIV-1 strains. We determined that the frequency of template switching decreased dramatically as sequence divergence increased between repeated sequences.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

MATERIALS AND METHODS

Plasmids and cells. The helper plasmid used, pCMVΔR8.2 (33), encodes all HIV-1 *trans*-acting proteins including Vpu but not envelope, driven by a CMV immediate early promoter. The HIV-1 transducing vector, pHIVlac, was a derivative of pHR'CMVlacZ (34) that contains *cis*-acting sequences necessary for reverse transcription, packaging, and integration. To construct pHIVlac, the Mfe I-Nhe I plasmid backbone of pHR'CMVlacZ was replaced with an Eco RI-Xba I fragment from pUC18. Then a 1-kb Asp 718-Sal I fragment containing an SV40 early promoter and a puromycin-resistance gene from an MLV-based vector (20) were inserted downstream of *lacZ*. pHIVlaac117, pHIVlaac284, and pHIVlaac971 were constructed by replacing the Fsp I-Eco RI *lacZ* fragment of pHIVlac with corresponding fragments from MLV-based vectors, pLaac-117, pLaac-284, and pLaac-971 (43).

pHIVlacWT was a derivative of pHIVlac with silent mutations that generated Eco RI and Xho I restriction sites flanking a 156-bp region of *lacZ*. Derivatives of pHIVlacWT in which the156-bp sequence was 95%, 91%, 82%, 73%, 63%, or 58% homologous to the wild type sequence were then constructed. For each derivative, a pair of lengthy sense and antisense oligonucleotides with 3' terminal complementary were annealed, and standard procedures for *E. coli* DNA polymerase I Klenow fragment were used to mutually prime and fill out the oligonucleotides. Eco RI and Xho I restriction sites engineered into the 5' ends of the oligonucleotides were used to generate replacement fragments, forming pHIVlac95%, pHIVlac91%, pHIVlac82%, pHIVlac73%, pHIVlac63%, and pHIVlac58%.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Repeat vectors were then constructed by joining a plasmid half that contained the 156-bp portion of *lacZ* and upstream sequences from one pHIVlac derivative with a plasmid half that included the 156-bp segment and downstream sequences from a second vector, using the Xho I and EcoR1 sites that flanked the varied portion of *lacZ* and the plasmid backbone site, Bcl I. These two larger fragments were ligated with a 1.5-kb Xho I-Mfe I linker fragment that contained murine leukemia virus coding sequence from pNCA (6). In this way, repeat vectors with various combinations of 5' and 3' repeats were constructed. For example, pHIVlaac95%-WT contained a 156-bp segment that differed by 5% from wild type, followed by the linker segment and then a wild type 156-bp segment.

ET cells are a 293T derivative that constitutively expresses ecotropic envelope (43). mATRC1/293 cells are derived from the human 293 cell line, and stably express the receptor for ecotropic Env (24). Both of these cell lines and their derivatives were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Gemini).

Transfection and virus harvesting. HIVlac and HIVlaac vector viruses were generated by transiently cotransfecting ET cells with pCMVΔR8.2 and the corresponding transducing vector plasmid by calcium phosphate precipitation (31). Virus was harvested by replacing culture medium 16-24 h after transfection and collecting medium 36-48 h post-transfection. Residual cells were removed from the medium by 4°C low speed centrifugation for 5 min at 3000 g. Virus was stored at –70°C prior to use. For each HIV-1 transducing vector, three viral stocks were prepared from separate cotransfections.

MLV vector viruses were produced using the same approaches, except that ET cells were cotransfected with the MLV helper construct, pMLV Ψ^-, and the MLV transducing constructs pLaac-117, pLaac-284, and pLaac-971 (43), respectively. Virus stocks were prepared and stored as described above and used to infect mATRC1/293 cells as described below for the HIV-1 based vectors.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Infection and functional assay for LacZ activity. mATRC1/293 cells in 6-cm-diameter dishes were infected 1 h in the presence of 8 μg/ml of hexadimethrine bromide (polybrene; Sigma). Puromycin (Sigma) selection, cell fixation, and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal; Sigma) staining were performed as previously reported (43) except that infected cells in each 6-cm-diameter dish were subcultured into one or two 10-cm-diameter dishes before puromycin selection. Infection with each virus stock was performed in parallel in three or six dishes as indicated. The numbers of blue and white colonies were summed for each infection experiment, and were tabulated in tables 2-1, 2-2, 2-3, and 2-4.

To determine *lacZ* inactivation rate, ET cells were cotransfected with pHIVlac and pCMVΔR8.2 helper, virus was harvested, and mATRC1/293 cells were infected, selected in puromycin, and stained as above. The inactivation rate was calculated as the ratio of white colonies per total colonies. To determine repeat deletion rates, pHIVlaac117, pHIVlaac284, or pHIVlaac971 was used in cotransfections with the helper plasmid. Where indicated, infected cells were maintained in media supplemented with 0.2 mM hydroxyurea (HU; Sigma) or 1.0 μM 3'-azido-3'deoxythymidine (AZT; Sigma) until puromycin selection was initiated. Results reported in tables 2-1 and 2-4 are from experiments in which HU or AZT treatment was started at the end of the 1 h infection period. Experimental repetitions were also performed in which HU or AZT treatment was initiated 4 h before infection and continued through the infection period. Under the latter condition, HU and AZT treatment resulted in increases in *lacZ* inactivation and deletion rates that were within the range of 1.3 to 1.8-fold: values which are similar to those obtained under conditions used to generate the data reported in the tables (Data not shown).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RESULTS

Vector and cell system to study HIV-1 direct repeat deletion during a single cycle of replication. An assay was established to examine direct repeat deletion rates during single cycles of HIV-1 replication. A single cycle is defined as starting with transcription and virus production in the producer cell, and then continuing through infection, reverse transcription, and proviral DNA integration into the target cell. In our system, replication was limited to a single cycle because vector virus was pseudotyped with ecotropic MLV envelope that was constitutively expressed by the producer cells (43), and 293-derived cells that express ecotropic receptor were used as the human target cell line (24). Vector viruses generated in the producer cells could not infect producer cells because they lack ecotropic receptors, and no infectious viruses were produced by infected target cells since the target cells were devoid of helper function. The titer of vector virus was typically 10^4 transducing units/ml, which is about one order of magnitude lower than those reported for similar HIV-1 vectors pseudotyped with amphotropic envelope or VSV-G (34).

For the assays employed here, the lentivirus helper construct, pCMVΔR8.2 (33), provided all HIV-1 proteins except envelope *in trans* for HIV-1 based transducing vectors (Figure 2-1A). The wild type transducing vector, pHIVlac, was a derivative of pHR'CMVlacZ (34), engineered to include a puromycin resistance gene driven by the SV40 early promoter. pHIVlaac117, pHIVlaac284, and pHIVlaac971 differed from pHIVlac in that each contained a 117, 284, or 971 base pair direct repeat within *lacZ*, respectively (Figure 2-1A).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

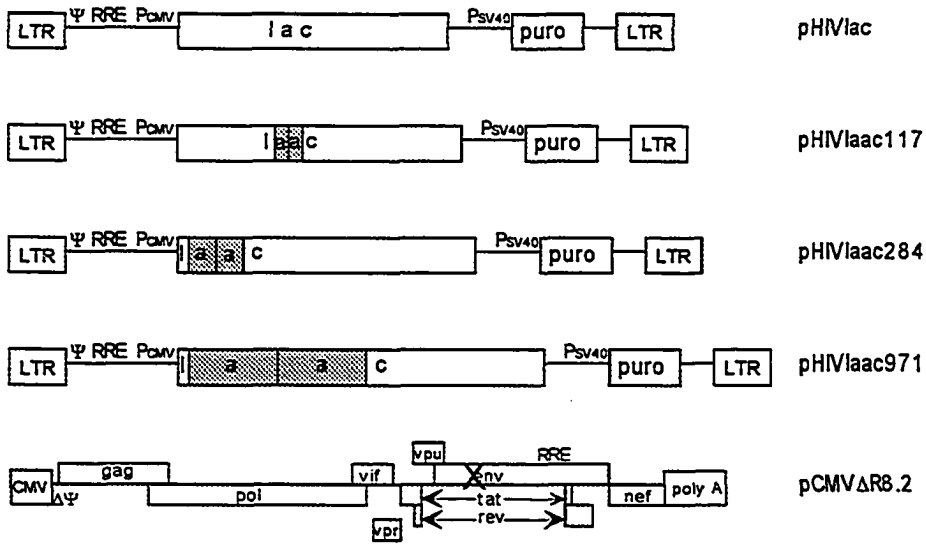


Figure 2-1A. HIV-1 vector system. Schematic drawings of HIV-1 transducing vectors (pHIVlac, pHIVlaac117, pHIVlaac284, and pHIVlaac971) and helper function vector (pCMVΔR8.2; (33). LTR, long terminal repeat; Ψ, packaging signal; RRE, Rev response element; PCMV, cytomegalovirus immediate-early promoter; PSV40, simian virus 40 promoter; puro, puromycin resistance gene. The locations of direct repeats are indicated by shaded boxes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To produce viral particles, individual transducing vector and helper constructs were cotransfected into producer cells, and virus was harvested and used to infect susceptible human cells (24) (Figure 2-1B). After 2 weeks of selection, puromycin-resistant colonies were X-gal stained. If precise repeat deletion occurred during reverse transcription, cells should stain blue; if the repeats were maintained, colonies would not stain. Deletion rates were calculated as ratios of blue colonies to the total number of colonies. For each transducing vector, virus was prepared from three separate cotransfections, and two independent infections were performed in triplicate on each virus stock to assess deletion rates during single replication cycles.

***lacZ* inactivation during HIV-1 replication.** Due to HIV-1's high mutation rate, the reconstituted *lacZ* gene in some proviral DNAs may be inactivated even if precise deletion takes place. To determine the inactivation rate of *lacZ* in our system, pHIVlac (which contains uninterrupted *lacZ*) was cotransfected into producer cells with pCMVΔR8.2, target cells were infected, and puromycin resistant colonies were counted after X-gal staining. The inactivation rate of *lacZ* was calculated as the ratio of white colonies to total colonies (Table 2-1). The results suggested that *lacZ* was inactivated at a rate of 6.4% for HIV-1 during single replication cycles, which is comparable to frequencies measured for MLV (17, 18, 43). The roughly 1.6-fold difference between our measurement and previously reported HIV-1 error frequencies (27) may not be significant, or it may reflect the different approaches used.

Deletion rates for different length direct repeats. To determine repeat deletion values for HIV-1, the cotransfection, infection, selection and staining protocol outlined above was repeated using pHIVlaac117, pHIVlaac284, and pHIVlaac971 as transducing vectors and pCMVΔR8.2 as helper (Table 2-2). The deletion rates measured by blue/total colony ratios for 117, 284, and 971 base direct repeats were 6.4%, 18.6%, and 81.4%, respectively. Factoring in the 6.4% *lacZ* inactivation rate determined above, calculated deletion frequencies were 6.8%, 19.9%, and 87%, respectively.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

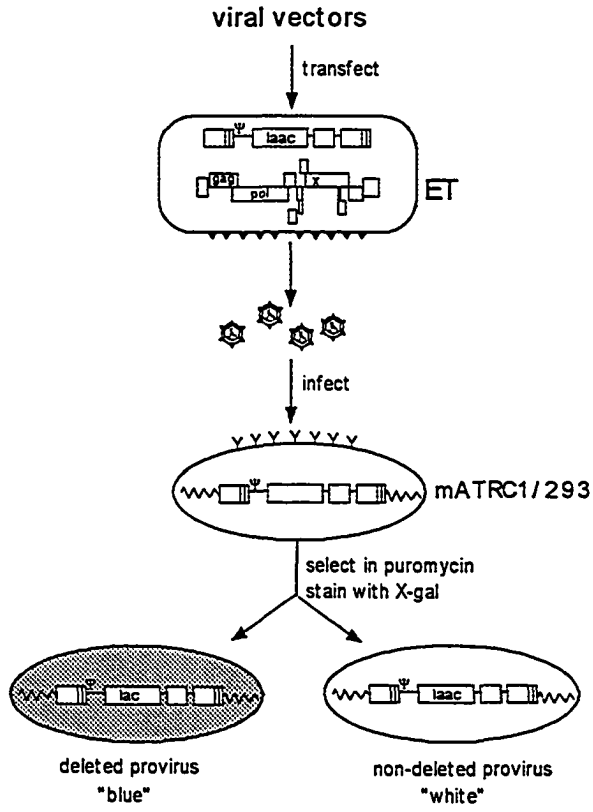


Figure 2-1B. Protocol for direct repeat deletion assay. To generate vector virus, individual transducing vector was co-transfected with pCMVΔR8.2 into ET cells. mATRC1/293 cells were infected with viruses, selected with puromycin, and were subsequently stained with X-gal. The numbers of blue and white colonies were counted and used to calculate the rates of direct repeat deletion.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2-1. Effect of HU or AZT treatment on *lacZ* inactivation rates of HIV-1

Experiment no.[a]	No treatment			HU treatment			AZT treatment		
	No. of blue colonies	No. of white colonies	Inactivation rate (%)[b]	No. of blue colonies	No. of white colonies	Inactivation rate (%)[b]	No. of blue colonies	No. of white colonies	Inactivation rate (%)[b]
1.1	245	27	9.9	180	19	9.5	196	17	8.0
1.2	3313	240	6.8	2207	258	10.5	2104	213	9.2
2.1	726	36	4.7	194	15	7.2	256	20	7.2
2.2	3034	218	6.7	1621	144	8.2	602	39	6.1
3.1	1637	85	4.9	682	64	8.6	628	42	6.3
3.2	3240	173	5.1	1216	87	6.7	596	45	7.0
Avg	-	-	6.4	-	-	8.5	-	-	7.3
SD	-	-	2.0	-	-	1.4	-	-	1.2

[a] The first number indicates separate co-transfections from which viruses were produced, and the second number indicates separate infections performed in triplicate. Data provided are sums of values from the three dishes.

[b] The *lacZ* inactivation rate was calculated as the ratio of the number of white colonies to the total number of colonies.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2-2. Deletion rates of direct repeats of different length of HIV-1

Experiment no.[a]	117-base repeat			284-base repeat			971-base repeat		
	No. of blue colonies	No. of white colonies	Deletion rate (%)[b]	No. of blue colonies	No. of white colonies	Deletion rate (%)[b]	No. of blue colonies	No. of white colonies	Deletion rate (%)[b]
1.1	99	1437	6.4	254	1005	20.2	2429	578	80.8
1.2	500	5546	8.3	228	1119	16.9	2192	547	80.0
2.1	289	3757	7.1	339	1401	19.5	2125	462	82.1
2.2	185	2324	7.4	840	3419	19.7	1236	236	84.0
3.1	61	1348	4.3	317	1405	18.4	637	134	82.6
3.2	30	620	4.6	106	514	17.1	1914	522	78.6
Avg	-	-	6.4	-	-	18.6	-	-	81.4
SD	-	-	1.6	-	-	1.4	-	-	1.9

[a] The first number indicates separate co-transfections from which viruses were produced, and the second number indicates separate infections performed in triplicate. Data provided are sums of values from the three dishes.

[b] The deletion rate was calculated as the ratio of the number of blue colonies to the total number of colonies.

mATRC1/293 cells were also infected with MLV-based vectors that contained the same *lacZ* repeats to compare deletion frequencies for the MLV and HIV-1 replication machinery. The values determined here for MLV infection of mATRC1/293 cells are shown in Table 2-3. In our previous work, murine NIH-3T3 cells were infected with these same vectors, and MLV repeat deletion values ranged from 4.8×10^{-2} per base per cycle of replication (5.6% deletion of a 117 base repeat) to 10.6×10^{-2} per base per cycle (30% deletion of a 284 base repeat), with an intermediate value of 6.8×10^{-2} per base per cycle (66% deletion of a 971 base repeat) for the longest repeat studied (43). As indicated in Figure 2-2, the rates of repeat deletion for MLV vectors reverse-transcribed in human mATRC1/293 cells were very similar to the MLV deletion frequencies previously determined for products generated in murine NIH-3T3 cells.

Effect of HU and AZT treatment on lacZ inactivation and repeat deletion rates. HU inhibits cellular ribonucleotide reductase and can induce dNTP pool imbalance. HU treatment has been shown to modestly increase direct repeat deletion rates for MLV and *lacZ* inactivation rates in SNV and MLV (18, 43, 52). AZT, the RT nucleoside inhibitor widely used in clinical therapy of HIV disease, also affects mutation rates (17, 26).

To address effects of HU and AZT treatment on *lacZ* inactivation in our system, mATRC1/293 cells were infected with HIVlac vector virus and incubated in culture media supplemented with 0.2mM HU or 1.0μM AZT, with untreated cells used as a control (Table 2-1). The HU treatment resulted in 8.5% *lacZ* inactivation, or a 1.3-fold increase compared to untreated cells. This slight increase for HIV-1 is lower than reported 1.8- to 2.7-fold increases for MLV and a 2.1-fold increase reported for SNV undergoing reverse transcription in HU-treated cells (18, 43). When target cells were infected in the presence of AZT, the *lacZ* inactivation rate was 7.3%, or 1.4-fold higher than in untreated cells. This magnitude of effect is lower than those reported in other studies, where treatment with a similar concentration of AZT increased the MLV *lacZ*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2-3. *LacZ* inactivation rates and deletion rates of direct repeats of different length of MLV in mATRC1/293 cells

Experiment no.[a]	Wild type			117-base repeat			284-base repeat			971-base repeat		
	No. of blue colonies	No. of white colonies	Inactivation rate (%)[b]	No. of blue colonies	No. of white colonies	Deletion rate (%)[c]	No. of blue colonies	No. of white colonies	Deletion rate (%)[c]	No. of blue colonies	No. of white colonies	Deletion rate (%)[c]
1	3639	337	8.5	89	2947	2.9	1215	2303	34.5	928	300	75.6
2	1908	168	8.1	65	1909	3.3	703	1457	32.5	1100	479	69.7
Avg	-	-	8.3	-	-	3.1	-	-	33.5	-	-	72.6
SD	-	-	0.3	-	-	0.3	-	-	1.4	-	-	4.2

[a] Each experiment involves independent infection of six dishes. Data provided are sums of values from the six dishes.
[b] The *lacZ* inactivation rate was calculated as the ratio of the number of white colonies to the total number of colonies.
[c] The deletion rate was calculated as the ratio of the number of blue colonies to the total number of colonies.

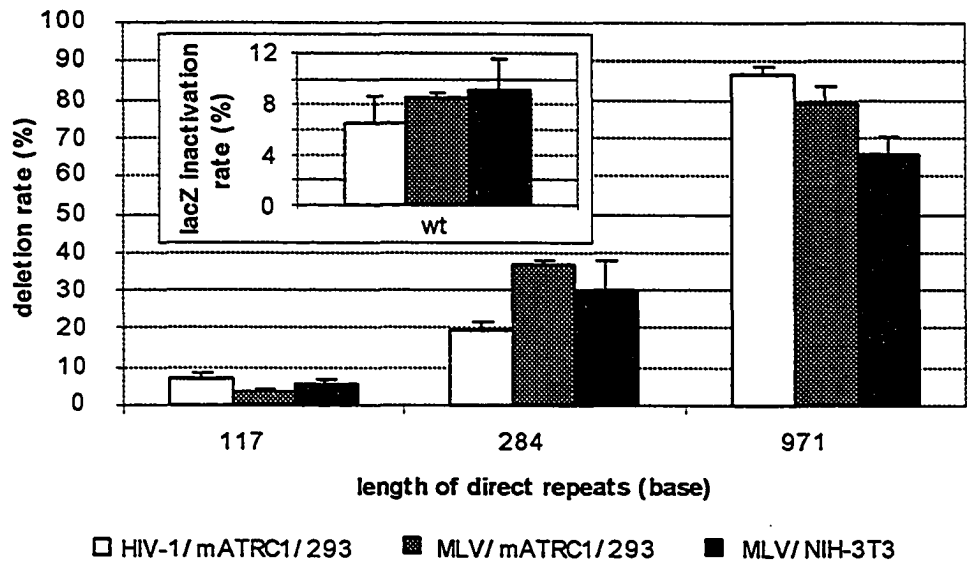


Figure 2-2. *lacZ* inactivation and repeat deletion frequencies for HIV-1 and MLV. Comparison of lacZ inactivation rates is shown in the insert and repeat deletion frequencies are indicated in the main figure. Values for HIV-1 vectors assayed in mATRC1/293 cells (white bars), MLV in mATRC1/293 cells (grey bars) and MLV assayed in NIH-3T3 cells (black bars; from (43). Repeat deletion values were corrected to account for *lacZ* inactivation frequencies.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

inactivation rate three-fold (17). Even lower concentrations of AZT were found to increase *lacZ* inactivation for SNV by 7 to 10-fold (17), and to increase forward mutation of *lacZα* 7-fold for HIV-1 (26). The cause of the differences in magnitudes of AZT effects between values determined here and those reported previously was not determined, but differences may reflect cell-line specific differences in AZT processing or investigator-specific differences in methodology.

To study the effect of HU and AZT treatment on HIV-1 direct repeat deletion, target cells were infected with HIVlaac117 virus and incubated in media containing 0.2mM HU or 1.0μM AZT. Deletion rates were determined as above (Table 2-4). When treated with 0.2mM HU, the deletion rate increased to 11.4%. This 1.8-fold increase upon HU treatment is similar in magnitude to that which we observed for the identical *lacZ* repeat in MLV vectors (43), and is very similar to the 1.7-fold increase observed by another group that subsequently measured HU effects on MLV repeat deletion (52). When treated with 1.0μM AZT, the deletion rate increased to 10.3%, a 1.6-fold increase over the deletion rate in untreated target cells.

Establishing assay vectors that contain sequence degeneracy within lacZ. In order to determine the effects of varying sequence homology on template switching, pHIVlacWT had been engineered to contain Eco RI and Xho I restriction sites 156 nt apart within *lacZ* coding sequences (Figure 2-3A). The region between Eco RI and Xho I sites was subsequently modified by synonymous substitution to make *lacZ* derivatives for use in the experiments below (Figure 2-3B). Sequence modifications introduced into this region followed three criteria: (i) the amino acid coding was not altered; (ii) the mutations were distributed evenly over the region; and (iii) mutations in the variants were introduced based on the codon frequency of the parental *lacZ* gene (to minimize any effect of codon usage on translational efficiency) wherever possible. The integrity of the *lacZ* variants was confirmed by transfecting individual vectors into ET cells and examining *lacZ* expression with X-gal staining (not shown).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2-4. Effect of HU or AZT treatment on deletion rates of 117-base direct repeats of HIV-1

	No treatment			HU treatment			AZT treatment		
Experiment no.[a]	No. of blue colonies	No. of white colonies	Deletion rate (%)[b]	No. of blue colonies	No. of white colonies	Deletion rate (%)[b]	No. of blue colonies	No. of white colonies	Deletion rate (%)[b]
1.1	99	1437	6.4	44	360	10.9	42	395	9.6
1.2	500	5546	8.3	533	3188	14.3	185	1708	9.8
2.1	289	3757	7.1	135	1177	10.3	159	1277	11.1
2.2	185	2324	7.4	101	747	11.9	48	330	12.7
3.1	61	1348	4.3	63	526	10.7	50	542	8.4
3.2	30	620	4.6	20	175	10.3	13	115	10.2
Avg	-	-	6.4	-	-	11.4	-	-	10.3
SD	-	-	1.6	-	-	1.5	-	-	1.5

[a] The first number indicates separate co-transfections from which viruses were produced, and the second number indicates separate infections performed in triplicate. Data provided are sums of values from the three dishes.

[b] The *lacZ* inactivation rate was calculated as the ratio of the number of white colonies to the total number of colonies.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

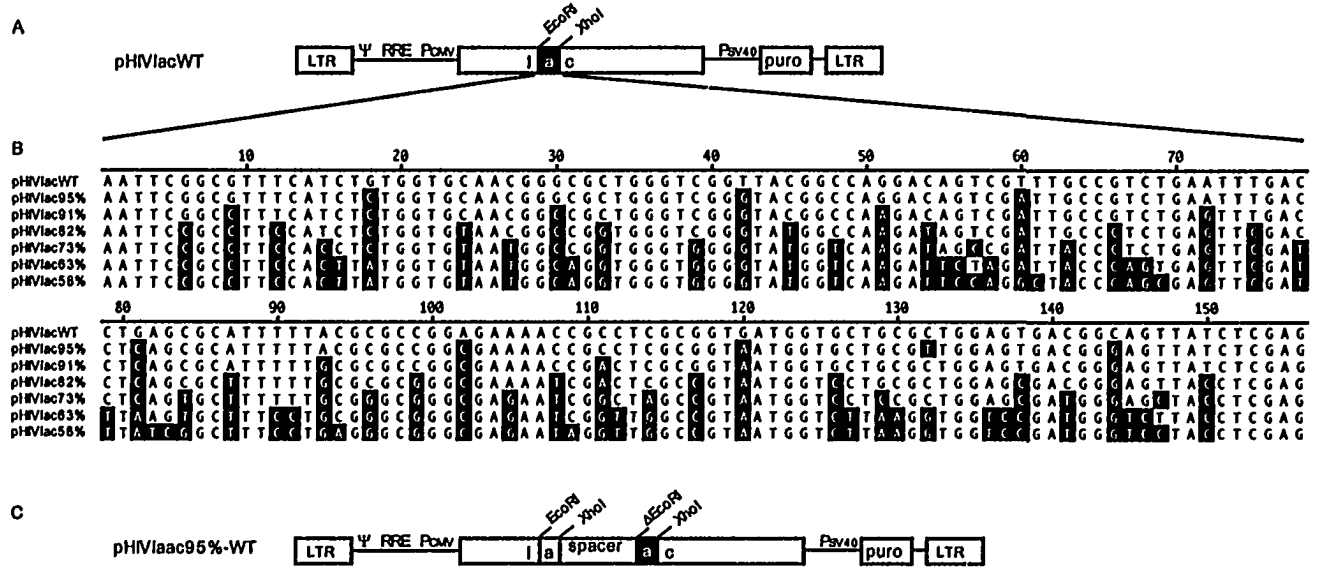


Figure 2-3. HIV-1 transducing vectors that contain sequence degeneracy within *lacZ*. **(A)** Structure of pHIVlacWT. **(B)** Nucleotide sequence alignment of the 156-nt EcoRI-XhoI lacZ fragments from pHIVlacWT, pHIVlac95%, pHIVlac91%, pHIVlac82%, pHIVlac73%, pHIVlac63%, and pHIVlac58%. Nucleotides of the variants that differed from wild type sequence are highlighted. **(C)** pHIVlaac95%-WT, a representative repeat vector. Its 5' repeat was derived from the 156-nt *lacZ* fragment of pHIVlac95%, and its 3' repeat was derived from pHIVlacWT. Abbreviations are as defined for figure 2-1.

Repeat vectors containing two copies of this 156-nt fragment within *lacZ* were then constructed (Figure 2-3C). It has been shown for MLV that deletion rates increase when repeated sequences are separated by a spacer sequence, presumably because the distance between repeats increases the chance of template switching (8). Accordingly, all repeat vectors used here contained the same 1.5-kb nucleotide spacer between the 156-nt repeats (Figure 2-3C).

Deletion rates between different identical repeats. It has been suggested that reverse transcriptase template switches do not occur at uniform frequencies and are highly context-dependent (1, 56). For example, sequences that potentially form secondary structures have been shown to be hot spots for both mutation and recombination (14, 15, 37). Thus, to address potential sequence-specific effects of the introduced variation on template switching rates, the initial deletion vectors tested contained each of the 156-nt region sequence variants in tandem with an identical repeat of itself (Figure 2-4). Because the repeats in each of these vectors were identical to one another, differences among vectors' deletion rates would reveal whether or not particular 156-nt variants displayed significant differences in intrinsic template switching properties.

HIV-1 vector virions were generated by transiently cotransfecting ET cells—a 293T derivative that constitutively expresses ecotropic envelope (43)—with the HIV-1 helper plasmid, pCMVΔR8.2 (33), and each of the identical repeat vectors (Figure 2-4). The resulting virions were harvested, and mATRC1/293 cells (24) were infected, selected in puromycin, and assayed for β-galactosidase by blue-white screening as previously reported (43). For each vector, three independent infections were performed in triplicate to assess deletion rates during single replication cycles. The numbers of blue and white colonies were summed for each infection experiment, and deletion rates were calculated as ratios of blue colonies to the total number of puromycin resistant colonies. To correct for the *lacZ* inactivation rate (5.0% for pHIVlacWT (not shown)), the above calculated deletion rates were divided by 0.95 (i.e. 1 – 0.05). The corrected deletion rates were then

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

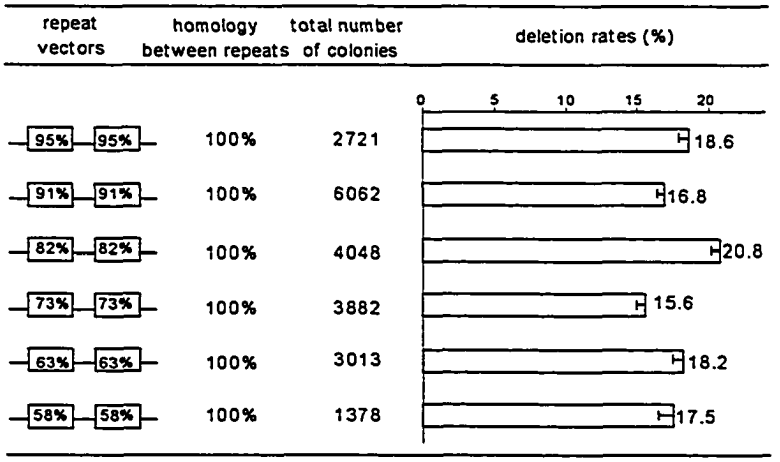


Figure 2-4. Repeat deletion rates for vectors with identical *lacZ* variant repeats. The repeat vector structures are illustrated with boxed percentages. For instance, —95%—95%— represents pHIVlaac95%-95%. Homology between repeats was 100% for each vector in this series. The repeat deletion rates were calculated as described in the text (means [bars] minus standard errors of the means [error bars] are shown).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

averaged from three independent infections for each vector. For the 156-nt wild type tandem repeat (wt-wt), the measured deletion rate was 15.6% (averaged from Figures 2-5). Deletion rates for vectors containing identical repeats of 95%, 91%, 82%, 73%, 63%, or 58% homologies to the wild type sequence were all similar in value, and were observed to be 18.6%, 16.8%, 20.8%, 15.6%, 18.2%, 17.5%, respectively (Figure 2-4). Note that pairwise χ^2 tests suggest that deletion rates for certain pairs of vectors, such as the 82%-82% and 73%-73% vectors, may differ significantly ($P < 0.05$). However, most pairwise comparisons do not differ statistically, and thus observed deletion rate variation among identical repeat vectors should not affect interpretation of the trends reported here.

Deletion rates between non-identical repeats. Vectors with two different repeats were constructed to test the effects of sequence divergence on repeat deletion (Figures 2-5). It is believed that repeat deletion takes place mostly during retroviral DNA minus strand synthesis when the growing point of the newly synthesized minus strand DNA dissociates from the 3' repeat template RNA (template switch donor) and re-associates with the homologous 5' repeat (template switch acceptor) (61). Thus, deletion rates were first determined for vector viruses containing wild type lacZ sequence as their 3' repeat, because identical sequences would serve as the template switch donor for each vector under these conditions, provided that template switching occurred during minus strand recombination. As shown in Figure 2-5A, deletion rates were 9.6%, 3.9%, 0.77%, < 0.04%, < 0.05%, and < 0.08% for repeats of 95%, 91%, 82%, 73%, 63%, and 58% homologies, respectively.

An additional set of vectors was constructed which differed from the first in that the orientation of the repeats was inverted. Thus, the new vectors contained different variants as template switching donors and identical wild type sequences as template switching acceptors (Figure 2-5B). When assayed, deletion rates were determined to be 7.6%, 4.9%, 0.2%, < 0.3%, < 0.2%, and < 0.1% for repeats of 95%, 91%, 82%, 73%, 63%, and 58% homologies, respectively (Figure 2-5B). Although the total number of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

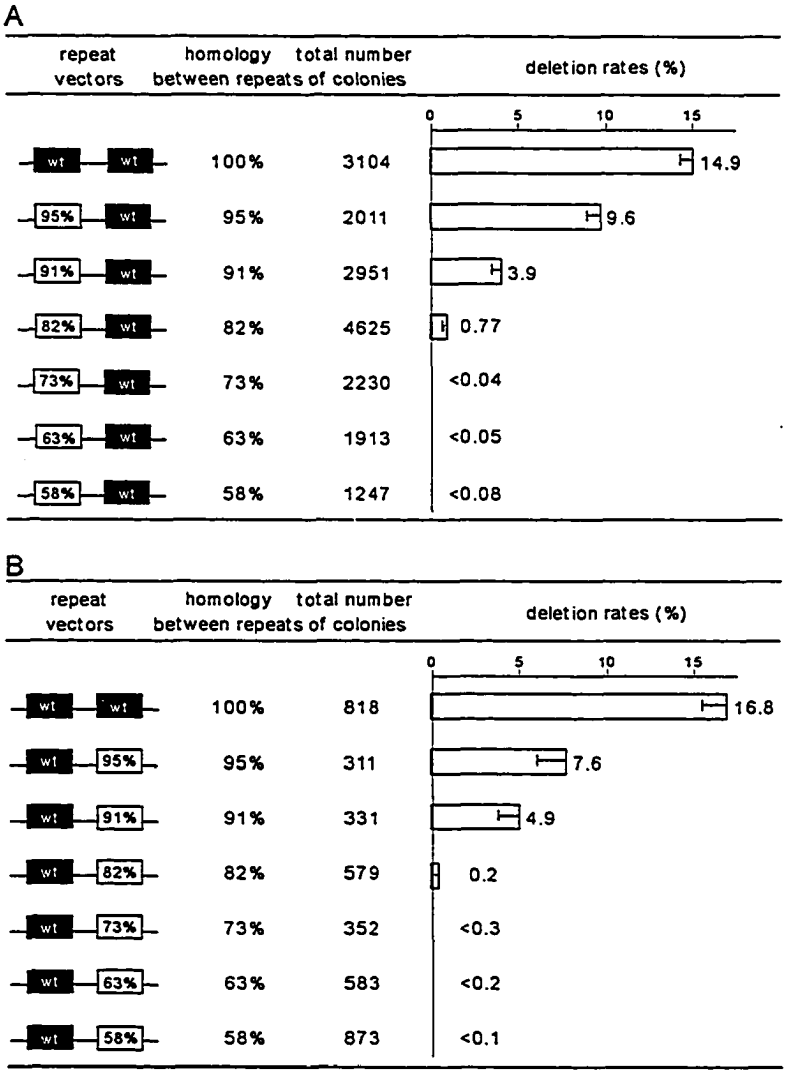


Figure 2-5. Effects of varying extents of sequence identity on repeat deletion frequencies. **(A)** Repeat deletion rates for vectors with wild type *lacZ* sequence in the 3' repeat. The repeat vector structures are illustrated with boxed percentages (the 5' and 3' repeats are as indicated in the left and right boxes, respectively). For instance, —95%— wt — represents pHIVlaac95%-WT in which the homology between repeats was 95%. Deletion rates were calculated as described in the text (means [bars] minus standard errors of the means [error bars] are shown). **(B)** Repeat deletion rates for vectors with wild type *lacZ* sequences in the 5' repeat. The methods for illustrating repeat vector structures and calculation of deletion rates are as described for panel A.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

colonies examined for each of the vectors in this second set was lower than that of the previous set, deletion rate trends associated with sequence modulation were consistent with those observed for the set of vectors analyzed in Figure 2-5A. Statistical analysis revealed no significant difference of deletion rates between the two sets of vectors (χ^2 test or Fisher exact test, $p > 0.05$).

DISCUSSION

We established a system for examining HIV-1 direct repeat deletion rates during single cycles of viral replication. Direct repeat deletion, measured by the frequency of marker gene reconstitution, has been used as a measure of homologous recombination frequency during both cellular and viral DNA synthesis (4, 7, 8, 16, 42, 43, 52, 57, 60). An important feature of this new HIV-1-based assay was that the *lacZ* expression unit and the repeats within *lacZ* used to measure deletion frequencies were identical to those previously used to assess MLV template switching (43). This was important for allowing a comparison of deletion frequencies by MLV and HIV-1 because reverse transcriptase errors do not occur at uniform frequencies and are highly context-dependent (1, 47, 56). For example, sequences that potentially form secondary structures have been shown to be hot spots for both mutation and recombination (14, 15, 37). These and other poorly understood template features that affect error frequencies are not present at uniform frequencies throughout templates. The system used here eliminated template structure and sequence context as variables and allowed direct comparison of MLV replication properties in murine cells to those of HIV-1 in human cells.

Findings reported here demonstrated that direct repeat deletion frequencies of MLV and HIV-1, when both viruses were tested with identical repeats in the same human

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

cell type, were very similar. The measured values agreed well with deletion rates previously determined by others for MLV in murine cells and SNV in canine cells (7, 8, 16, 43, 52). Thus, we conclude that HIV-1 deletes repeats at a frequency similar to simple retroviruses like MLV and SNV. This finding was unexpected because previous measurements of intermolecular template switching rates between copackaged RNAs have suggested that HIV-1 template switching is about 10-fold more frequent than that of SNV or MLV (14).

Possible explanations for this new finding of similar repeat deletion rates for HIV-1 and MLV—despite vast reported recombination differences—range from technical differences of no biological importance to potentially significant differences in replication properties. For example, whereas HIV-1 intermolecular switching in a single-cycle assay occurred 2 or 3 times per cycle (14), data with replication-competent virus from a different research group suggests only one crossover occurs per replication cycle (50). It is similarly possible that if another research group using different approaches addressed simple retrovirus recombination rates, they might be revised upwards. Thus, technical differences between labs might erode the reported 10-fold difference between HIV-1 and simple retrovirus recombination frequencies.

Another possibility is that although the highly mosaic nature of natural HIV-1 recombinants appears consistent with high recombination frequencies (45), these mosaic genomes may have resulted from negative interference, which occurs during the replication of simple retroviruses (10). Negative interference is the phenomenon in which retroviral genomes that show any evidence of genetic recombination tend to display more crossovers than would be predicted based on calculated average recombination rates. If previously described HIV-1 recombinants are not typical HIV-1 reverse transcription products but instead products of a rare "recombining population" (1), then HIV-1 and simple retrovirus recombination rates may be indistinguishable over all.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A third possibility is that direct repeat deletion and intermolecular recombination may differ mechanistically, so one cannot be used as a predictor for the other. A possible precedent for this notion lies in the observation that direct repeat deletion in *E.coli* is independent of the recombination protein RecA, whereas other forms of homologous recombination are RecA-dependent (22).

A fourth possibility is that template switching frequencies may be the same for the studied retroviruses but that their genome organizations may differ, thus leading to the appearance of differences where none exist. Retroviral recombination is detectable only when it occurs between genetically distinct templates (11). Although the possibility that RNA copackaging may not be random has been discussed (10), a general assumption has been that when two similar retroviral RNAs are coexpressed in single cells, they will dimerize and thus copackage randomly, and hence numbers of 'heterozygous' particles can be calculated from the total amount of each RNA in the population. However, how RNAs are recruited for packaging is not well understood, and there is some evidence that these processes may differ among retroviruses (19). If MLV RNA homodimers were packaged preferentially but dimerization and copackaging were random for HIV-1, HIV-1 recombination would appear to be more frequent than that of MLV even if template switching rates were indistinguishable.

Introducing sequence heterogeneity allowed us to test effects of altering the genetic distance between repeated sequences on HIV-1 template switching. HIV-1 deletion frequency between repeats of 95% homology was about 65% as high as values for identical repeats. The frequency decreased to 26% of the identical repeat rate for repeats with 91% homology, and was about 5% for repeats with 82% homology. Failure to detect any recombinants among thousands of products that resulted from vectors with repeats that differed by 27% (73% homology) suggests that this level of divergence decreased recombination frequency more than 300-fold. Considering that non-homologous recombination reportedly occurs roughly 1000-fold less frequently than

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

homologous retroviral recombination (62), this level of recombination suppression appears similar in magnitude to that predicted for two unrelated sequences. The effects of sequence context on repeat deletion rates were tested by placing each variant repeat in tandem with itself. Whether or not there were hotspots for template switching in our 156-nt sequences was not addressed directly. However, the level of similarity in deletion rates observed for the different identical repeat vectors suggests that the magnitude of any possible hotspots was too low to alter the above conclusion.

These results suggest that retroviral recombination is more permissive of genetic variation than is bacterial or mammalian DNA recombination. It has previously been established that homologous recombination is far more frequent during retroviral replication than during cellular replication (11, 21): our new findings suggest that greater levels of sequence divergence are tolerated during retroviral recombination as well. Studies on cellular homology requirements during recombination usually take one of two approaches: either to determine the rate of recombination as a function of the length of shared uninterrupted homology, or to test the effects of base pair mismatch on the rate of recombination ((54) and citations therein). The system used here could be envisioned as following the second approach whereby mismatch was introduced into the 156-base lacZ gene fragment, resulting in variants with differing degrees of sequence homology. In bacteria, 16% mismatch between sequences reduces phage-plasmid recombination rates by a factor of 100 (49), and in mammalian cells, 19% mismatch reduced intra-chromosomal recombination over 1000-fold (55). In comparison, our results show18% mismatch reduced HIV-1 repeat deletion by roughly 20-fold (comparing the observed deletion rate of 0.77% reported here for 82% homologous repeats with the 14.9% deletion rate determined for identical repeats). We realized that this comparison has its limitation because the corresponding rate of cellular recombination was not measured for the 156-bp sequence used in this study. Nevertheless, the findings presented here suggest that HIV-1 recombination may be less sensitive to sequence variation than is the cellular

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

replication machinery. Recent research has demonstrated that DNA mismatch repair systems play critical roles in inhibiting recombination between non-identical sequences in both bacteria and eukaryotes (9). Mutational inactivation of components of the mismatch repair system can significantly reduce the block to 'homeologous' recombination between similar but non-identical sequences (32). Whether or not such factors as the host cells' mismatch repair machinery affected our read-out of apparent HIV-1 deletion frequencies was not addressed here.

Here, repeat deletion was used to measure the effect of sequence divergence on HIV-1 template switching, whereas genetic recombination as commonly defined occurs between sequences on different copackaged RNAs. Repeat deletion appears to occur primarily through intramolecular template switching, at least for gammaretroviruses (10, 59). In contrast, recombination between different strains of HIV-1 is an intermolecular event, and our assays did not address parameters that differ between inter- and intramolecular template switching. These parameters include the physical interaction that has been implicated in promoting recombination at the region flanking the dimerization initiation sequence (2, 3, 23, 29, 30). Indirect evidence from melting profile of nicked genomic RNAs suggests that extensive RNA-RNA interactions exist at other regions of the RNA genome as well (35), and these may also contribute to intermolecular recombination.

Another way our approach differed from intermolecular recombination is that tandem repeats provide a total of four possible templates for the assayed region—two on each copackaged RNA—instead of the two present for intermolecular recombination. Thus, in our system, competitive effects of the homologous repeat on the copackaged RNA may have affected deletion rates. However, since such competition exists for both non-identical and identical repeat vectors, the observed decreases in deletion clearly do reflect effects of decreasing homology on template switching, but may misrepresent the magnitude of such effects during intermolecular switching.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Not surprisingly, our results suggest that regions with high sequence similarity are more likely to serve as HIV-1 recombination substrates than sequences that differ more significantly. However, as evidenced by the natural occurrence of M/O recombinants, whose parent strains differ by an average of roughly 35%, recombination between highly divergent sequences can take place, even though scattered 27% sequence differences reduced recombination frequencies greater than 300-fold in our assays. Selective pressures lead to differences in extents of sequence variation among isolates in different portions of the HIV-1 genome, with some regions more conserved than others, and it is possible that local regions of similarity serve disproportionately as recombination crossover sites. However, due to the very high replication rate of HIV-1, even recombination products that arise at frequencies no higher than those between unrelated sequences likely contribute to viral populations if they confer a selective advantage (5).

BIBLIOGRAPHY

1. **Anderson, J. A., E. H. Bowman, and W. S. Hu** 1998. Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation J Virol. **72:**1195-202.

2. **Balakrishnan, M., P. J. Fay, and R. A. Bambara** 2001. The kissing hairpin sequence promotes recombination within the HIV-I 5' leader region J Biol Chem. **276:**36482-92.

3. **Balakrishnan, M., B. P. Roques, P. J. Fay, and R. A. Bambara** 2003. Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis J Virol. **77:**4710-21.

4. **Bierne, H., D. Vilette, S. D. Ehrlich, and B. Michel** 1997. Isolation of a dnaE mutation which enhances RecA-independent homologous recombination in the Escherichia coli chromosome Mol Microbiol. **24:**1225-34.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5. **Coffin, J. M.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy [see comments] Science. **267:**483-9.

6. **Colicelli, J., and S. P. Goff** 1988. Sequence and spacing requirements of a retrovirus integration site J Mol Biol. **199:**47-59.

7. **Delviks, K. A., W. S. Hu, and V. K. Pathak** 1997. Psi- vectors: murine leukemia virus-based self-inactivating and self- activating retroviral vectors J Virol. **71:**6218-24.

8. **Delviks, K. A., and V. K. Pathak** 1999. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching J Virol. **73:**7923-32.

9. **Harfe, B. D., and S. Jinks-Robertson** 2000. DNA mismatch repair and genetic instability Annu Rev Genet. **34:**359-399.

10. **Hu, W. S., E. H. Bowman, K. A. Delviks, and V. K. Pathak** 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference J Virol. **71:**6028-36.

11. **Hu, W. S., V. K. Pathak, and H. M. Temin** 1993. Role of reverse transcriptase in retroviral recombination, p. 251-74, Reverse transcriptase. Cold spring harbor laboratory press, Plainview, NY.

12. **Hu, W. S., and H. M. Temin** 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:**1556-60.

13. **Hunter, E.** 1978. The mechanism for genetic recombination in the avian retroviruses Curr Top Microbiol Immunol. **79:**295-309.

14. **Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome J Virol. **74:**1234-40.

15. **Jones, J. S., R. W. Allan, and H. M. Temin** 1994. One retroviral RNA is sufficient for synthesis of viral DNA J Virol. **68:**207-16.

16. **Julias, J. G., D. Hash, and V. K. Pathak** 1995. E- vectors: development of novel self-inactivating and self-activating retroviral vectors for safer gene therapy J Virol. **69:**6839-46.

17. **Julias, J. G., T. Kim, G. Arnold, and V. K. Pathak** 1997. The antiretrovirus drug 3'-azido-3'-deoxythymidine increases the retrovirus mutation rate J Virol. **71:**4254-63.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

18. **Julias, J. G., and V. K. Pathak** 1998. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate J Virol. **72:**7941-9.

19. **Kaye, J. F., and A. M. Lever** 1999. Human immunodeficiency virus types 1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation J Virol. **73:**3023-31.

20. **Kulpa, D., R. Topping, and A. Telesnitsky** 1997. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors Embo J. **16:**856-65.

21. **Li, T., and J. Zhang** 2000. Determination of the frequency of retroviral recombination between two identical sequences within a provirus J Virol. **74:**7646-50.

22. **Lovett, S. T., P. T. Drapkin, V. A. Sutera, and T. J. Gluckman-Peskind** 1993. A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in Escherichia coli Genetics. **135:**631-42.

23. **Lund, A. H., J. G. Mikkelsen, J. Schmidt, M. Duch, and F. S. Pedersen** 1999. The kissing-loop motif is a preferred site of 5' leader recombination during replication of SL3-3 murine leukemia viruses in mice J Virol. **73:**9614-8.

24. **Malhotra, S., A. G. Scott, T. Zavorotinskaya, and L. M. Albritton** 1996. Analysis of the murine ecotropic leukemia virus receptor reveals a common biochemical determinant on diverse cell surface receptors that is essential to retrovirus entry J Virol. **70:**321-6.

25. **Mansky, L. M.** 2000. In vivo analysis of human T-cell leukemia virus type 1 reverse transcription accuracy J Virol. **74:**9525-31.

26. **Mansky, L. M., and L. C. Bernard** 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:**9532-9.

27. **Mansky, L. M., and H. M. Temin** 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:**5087-94.

28. **Mansky, L. M., and H. M. Temin** 1994. Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus J Virol. **68:**494-9.

29. **Mikkelsen, J. G., A. H. Lund, M. Duch, and F. S. Pedersen** 2000. Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo J Virol. **74:**600-10.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

30. **Mikkelsen, J. G., A. H. Lund, M. Duch, and F. S. Pedersen** 1998. Recombination in the 5' leader of murine leukemia virus is accurate and influenced by sequence identity with a strong bias toward the kissing- loop dimerization region J Virol. **72:**6967-78.

31. **Miller, A. D., D. G. Miller, J. V. Garcia, and C. M. Lynch** 1993. Use of retroviral vectors for gene transfer and expression Methods Enzymol. **217:**581-99.

32. **Myung, K., A. Datta, C. Chen, and R. D. Kolodner** 2001. SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination Nat Genet. **27:**113-6.

33. **Naldini, L., U. Blomer, F. H. Gage, D. Trono, and I. M. Verma** 1996. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector Proc Natl Acad Sci U S A. **93:**11382-8.

34. **Naldini, L., U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector [see comments] Science. **272:**263-7.

35. **Ortiz-Conde, B. A., and S. H. Hughes** 1999. Studies of the genomic RNA of leukosis viruses: implications for RNA dimerization J Virol. **73:**7165-74.

36. **Parthasarathi, S., A. Varela-Echavarria, Y. Ron, B. D. Preston, and J. P. Dougherty** 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications J Virol. **69:**7991-8000.

37. **Pathak, V. K., and H. M. Temin** 1992. 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate J Virol. **66:**3093-100.

38. **Pathak, V. K., and H. M. Temin** 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:**6024-8.

39. **Pathak, V. K., and H. M. Temin** 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:**6019-23.

40. **Peeters, M., F. Liegeois, N. Torimiro, A. Bourgeois, E. Mpoudi, L. Vergne, E. Saman, E. Delaporte, and S. Saragosti** 1999. Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient J Virol. **73:**7368-75.

41. **Peeters, M., and P. M. Sharp** 2000. Genetic diversity of HIV-1: the moving target Aids. **14:**S129-40.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

42. **Pfeiffer, J. K., M. M. Georgiadis, and A. Telesnitsky** 2000. Structure-based moloney murine leukemia virus reverse transcriptase mutants with altered intracellular direct-repeat deletion frequencies J Virol. **74:**9629-36.

43. **Pfeiffer, J. K., R. S. Topping, N. H. Shin, and A. Telesnitsky** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription J Virol. **73:**8441-7.

44. **Preston, B. D., and J. P. Dougherty** 1996. Mechanisms of retroviral mutation Trends Microbiol. **4:**16-21.

45. **Quinones-Mateu, M. E., and E. J. Arts** 1999. Recombination in HIV-1: update and implications AIDS Rev. **1:**89-100.

46. **Rhode, B. W., M. Emerman, and H. M. Temin** 1987. Instability of large direct repeats in retrovirus vectors J Virol. **61:**925-7.

47. **Ricchetti, M., and H. Buc** 1990. Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent Embo J. **9:**1583-93.

48. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber** 1999. HIV-1 Nomenclature Proposal, p. 492-505. *In* C. L. Kuiken, B. Foley, B. Hahn, B. Korber, F. McCutchan, P. A. Marx, J. W. Mellors, J. I. Mullins, J. Sodroski, and S. Wolinksy (eds), Human Retroviruses and AIDS 1999. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

49. **Shen, P., and H. V. Huang** 1986. Homologous recombination in Escherichia coli: dependence on substrate length and homology Genetics. **112:**441-57.

50. **St Louis, D. C., D. Gotte, E. Sanders-Buell, D. W. Ritchey, M. O. Salminen, J. K. Carr, and F. E. McCutchan** 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:**3991-8.

51. **Stuhlmann, H., and P. Berg** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription J Virol. **66:**2378-88.

52. **Svarovskaia, E. S., K. A. Delviks, C. K. Hwang, and V. K. Pathak** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching J Virol. **74:**7171-8.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

53. **Takehisa, J., L. Zekeng, E. Ido, Y. Yamaguchi-Kabata, I. Mboudjeka, Y. Harada, T. Miura, L. Kaptu, and M. Hayami** 1999. Human immunodeficiency virus type 1 intergroup (M/O) recombination in cameroon J Virol. **73:**6810-20.

54. **Waldman, A. S., and R. M. Liskay** 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology Mol Cell Biol. **8:**5350-7.

55. **Waldman, A. S., and R. M. Liskay** 1987. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells Proc Natl Acad Sci U S A. **84:**5340-4.

56. **Wooley, D. P., L. A. Bircher, and R. A. Smith** 1998. Retroviral recombination is nonrandom and sequence dependent Virology. **243:**229-34.

57. **Xu, H., and J. D. Boeke** 1987. High-frequency deletion between homologous sequences during retrotransposition of Ty elements in Saccharomyces cerevisiae Proc Natl Acad Sci U S A. **84:**8553-7.

58. **Yu, H., A. E. Jetzt, and J. P. Dougherty** 1997. Use of single-cycle analysis to study rates and mechanisms of retroviral mutation Methods. **12:**325-36.

59. **Zhang, J., and Y. Ma** 2001. Evidence for retroviral intramolecular recombinations J Virol. **75:**6348-58.

60. **Zhang, J., and C. M. Sapp** 1999. Recombination between two identical sequences within the same retroviral RNA molecule J Virol. **73:**5912-7.

61. **Zhang, J., L. Y. Tang, T. Li, Y. Ma, and C. M. Sapp** 2000. Most retroviral recombinations occur during minus-strand DNA synthesis J Virol. **74:**2313-22.

62. **Zhang, J., and H. M. Temin** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication Science. **259:**234-8.

63. **Zhang, J., and H. M. Temin** 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone J Virol. **68:**2409-14.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER THREE

HIV-1 GENETIC RECOMBINATION IS MORE FREQUENT THAN THAT OF MLV DESPITE SIMILAR TEMPLATE SWITCHING RATES

INTRODUCTION

Genetic recombination contributes significantly to the genetic variability observed in populations of retroviruses such as HIV-1 and Moloney murine leukemia virus (MLV) (6, 26, 29, 36). Retroviral recombination occurs shortly after cell infection and results from template switching between RNA or DNA replication intermediates during reverse transcription (18, 26, 49). One factor that contributes to the frequency of recombination is retroviruses' property of copackaging two intact genomes, which provides unusual access to a genetic recombination substrate (29). Another contributing factor is reverse transcriptase's propensity to perform recombinogenic switching, which may reflect selective pressure to maintain a polymerase capable of performing the strong stop template switches that are required during retroviral DNA synthesis (8, 47).

How frequently retroviral recombination occurs is difficult to assess in replicating virus populations due to the influence of selective pressures. Recombination frequencies can be examined with single cycle recombination assays that monitor marker co-segregation during individual rounds of DNA synthesis (50). Such assays have yielded estimated spleen necrosis virus (SNV) recombination frequencies of about one

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

recombination event during the synthesis of every three to seven viral DNAs (19, 20, 23) and similar rates have been reported for the related gammaretrovirus, MLV (4). Surprisingly, recombination occurs once to three times on average during the synthesis of every HIV-1 DNA (22, 43, 51), which is perhaps 10-fold more frequent than calculated SNV or MLV rates.

What accounts for this dramatic difference between HIV-1 and gammaretrovirus recombination rates? One possibility is that retroviruses' reverse transcriptases may possess different enzymatic properties and that template switching rates vary among viruses (22). Possible support for this notion comes from the observation that although average missense mutation rates for HIV-1 and gammaretroviruses appear similar, retroviruses differ in how frequently they generate errors in response to specific template features (2, 39)

Alternately, the untested assumption that gammaretroviruses and HIV-1 both randomly select coexpressed RNAs for copackaging may be incorrect. Reported recombination frequencies have been calculated by dividing the frequency of scorable recombinants by the "recombining population" size (19). The size of this population, which consists of virions that contain two different RNAs, has not been measured directly but has been inferred by assuming random co-segregation of coexpressed RNAs into nascent virions (19, 51). Models for random copackaging of genomic RNAs are based on the assumption that among unspliced viral RNAs that reach the cytoplasm, some will become mRNAs and others will become genomes, and that RNA copackaging results from random pairing among packaging-competent cytoplasmic RNAs. However, there is strong evidence that gammaretrovirus unspliced RNAs reside in two functionally separate pools (12, 27). Additionally, how retroviruses export unspliced RNAs to the cytoplasm, how RNAs are partitioned between progeny genome and mRNA functions, and what intracellular processes precede dimeric RNA packaging all appear to differ among retroviruses (5, 7, 9, 12, 13, 27, 41). If the randomness of RNA copackaging also varied

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

among retroviruses, then species-specific differences in genetic marker re-assortment might reflect differences in the fractions of virions that had copackaged two different RNAs rather than differences in template switching rates. This hypothesis suggests that recombinogenic template switching may occur more frequently during the replication of certain retroviruses than is evident from an examination of their reverse transcription products. Furthermore, the frequencies of genetic recombination (evident from the re-assortment of genetic markers) and of biochemical recombination (the occurrence of recombinogenic template switches) might be more similar for some retroviruses than for others.

Here, the frequencies of marker re-assortment for HIV-1 and MLV were compared directly, using the same producer and recipient cell types for each virus, virus-specific replication proteins, and recombination assay vectors that differed only in virus-specific *cis*-acting sequences. For both HIV-1 and MLV, the frequencies of recombination between two different coexpressed RNAs were compared to template switching rates observed when donor and acceptor sequences were parts of a single RNA and thus uniformly copackaged.

MATERIALS AND METHODS

Plasmids and cells. The helper plasmids used were pCMVΔR8.2 (30), which encodes all HIV-1 *trans*-acting proteins including Vpu but not envelope, and pMLV Ψ- for MLV (35). In control experiments to test whether or not transfection-related recombination contributed to results, the mammalian *lacZ* expression plasmid, pCH110 (16), was used.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

One recombination assay vector was pHR'CMVlacZ (31). The other vectors were derivatives of the HIV-1 vector, pHIVlac (2), or the MLV vector, pLacPuro (35), which have previously been described. The previously described (2) pHIVlac contains a CMV promoter-driven *lacZ* gene followed by an SV40 early promoter and puromycin-resistance gene in an HIV-1 based retroviral vector. pLacPuro is an MLV-based vector with the same structure as pHIVlac, except that *lacZ* expression is driven by the MLV LTR and the HIV-1 *cis*-acting sequences necessary for genome metabolism are replaced with the cognate sequences from MLV. To reflect the relationships among vectors, pHR'CMVlacZ (31) is referred to as HIV Lac in this report, and pHIVlac (2) is called HIV LacPuro. MLV L^cPuro and HIV L^cPuro were derivatives of pLacPuro and pHIVlac, respectively, and contain a 117 base pair deletion between the Eco RV and Ssp I sites in *lacZ*. The entire puromycin expression cassette was removed from pLacPuro to make MLV Lac. To make MLV L^cPuro-ac and HIV L^cPuro-ac, MLV and HIV L^cPuro were modified in their downstream LTR and flanking regions by deletion of polyadenylation sequences in R. For both viruses, this was achieved by truncating the downstream R at a restriction site (Sac I for HIV and Asp 718 for MLV) upstream of the polyadenylation addition signal and about 40 nucleotides downstream of the U3/R junction. A 2.6 kilobase pair Fsp I-Bsi WI fragment of *lacZ*, flanked on both ends with polylinker and adjacent sequences derived from passage through Litmus 38 (New England Biolabs) was appended downstream of the polyadenylation-inactivated R regions. These fragments of *lacZ* were flanked at their 3' ends with SV40 polyadenylation sequences derived from pRep8 (Invitrogen). The repeat deletion vectors were derivatives of pHIVlac that was first engineered to contain Eco R1 and Xho I restriction sites flanking the 156 base pair coding segment that was mutagenized (1). The degenerate replacement fragment was synthesized using two lengthy 3' terminal complementary oligonucleotides, which were filled out with Klenow and digested with Xho I and Eco RI. The tandem repeat vectors contained a 1.6 kilobase Xho I to Mfe I

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

linker fragment derived from MLV coding sequences between the two coding repeats: spacers have previously been implicated as increasing repeat deletion without altering high negative interference for simple retroviruses (11). Construction details for all plasmids are available upon request.

ET cells, which are a 293T derivative that constituitively expresses ecotropic envelope (35), were transfected to produce virus. mATRC1/293 cells, which are derived from human 293 cells and stably express ecotropic receptor (28) were used in infections. These and derivative cell lines were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Gemini). NIH-3T3 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% calf serum (Gibco).

Transfection and virus harvesting. Vector-containing virions were generated by transiently cotransfecting ET cells with pCMVΔR8.2 (30) or pMLV Ψ- (35) helper plasmids and the appropriate assay vector plasmid(s) by calcium phosphate precipitation (40). When two vectors were cotransfected, similar molar equivalents of each plasmid DNA were used, with a total of 8 μg DNA per 6-cm-diameter plate. As a result, all virions, whether HIV-1 or MLV-derived, possessed ecotropic envelope and could infect cells that express ecotropic receptor. Virus was harvested by replacing culture medium 16-24 hours after calcium phosphate exposure and collecting medium 36-48 hours post-transfection. Residual cells were removed by filtration. Virus was stored at –70°C prior to use.

Infection and functional assay for LacZ activity. mATRC1/293 cells and NIH 3T3 cells in 6-cm-diameter dishes were infected 1 h in the presence of 8 μg/ml of hexadimethrine bromide (polybrene; Sigma) at low multiplicities of infection (<0.01). Puromycin (Sigma) selection, cell fixation, and 5-bromo-4-chloro-β-D-galactopyranoside (X-gal; Sigma) staining were performed as previously reported (35).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Southern blot. At least 200 independent puromycin resistant colonies were pooled for each vector, and genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega). DNA was digested with Cla I and Mlu I, separated on an agarose gel, blotted and probed with a random primed radiolabeled probe generated from a 331-bp PCR fragment that overlapped with the Cla I site and was upstream to the Eco RV site in the *lacZ* gene, using standard approaches (34). The probe hybridized to both the lengthy (5.2 kb) and the shorter (478 bp and 316 bp) bands visible in the blot. Note that the length of probe that hybridized to the 316-bp band was the same as that which hybridized to the 478-bp band.

Virus RNA quantification. Virion RNA was purified from filtered culture media from transfected cells as previously described (42), and used in RNase protection assays. Vector RNAs were differentiated in RNase protection assays and their ratios quantified by PhosphorImager (Molecular Dynamics). To allow differentiation of both vectors with a single probe, the probe was complementary to the regions between Cla I and Ssp I in *lacZ* and thus included sequences complementary to the region deleted in pL^cPuro and related vectors. To quantify vector RNAs, RNase protection assay background values were subtracted and molarities were normalized by dividing protected product values by the number of radiolabeled nucleoside monophosphate residues in each.

Calculating the recombining population size and number of crossovers for two-vector recombination assays. To calculate the proportion of virions predicted to contain one each of coexpressed vector RNAs, the equation $D^2+2DA+A^2=1$, which assumes random copackaging, was used as previously described (20). D stands for the molar proportion of total virion RNA comprised of the donor *lacZ* vector (MLV or HIV Lac) as determined by RNase protection assay, and A for the corresponding amount of puromycin resistance-conferring acceptor vector (MLV or HIV L^cPuro). Only one of each heterozygous RNA pair contained a puromycin resistance cassette. Thus, assuming that whether a product DNA contained 3' end sequences from the puromycin vector or

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from the other vector was random, the portion of heterozygous virions capable of generating puromycin colonies was half of 2DA. $DA + A^2$ represented the total number of puromycin resistant colonies. One half the 2DA term was the recombining population, whose portion of the total puromycin resistant colony population could be calculated from $DA/(DA+ A^2)$, which equals D: the molar proportion of total virion RNA that was the donor.

Generating LacZ-positive puromycin resistant proviruses required at least one recombination event in the recombination target region, which was the 1.8-kb 3' portion of *lacZ*. For the calculations in Figure 3-3C, the observed frequency of blue colonies (25.7%) for the HIV-1 based vectors was divided by 0.94 to correct for colonies that would have scored blue but became white due to *lacZ* mutational inactivation (in this case, 27.4%. Inactivation frequency was from control data in experiments here, not shown, and (2)). Based on donor and acceptor RNA proportions, the recombining population for this example was calculated to be about 57%. The corrected blue colony frequency per recombining population (47.4%) reflected the percent single plus other odd-numbered crossovers in the target region minus the percentages of double and other even-numbered crossovers in this region, since the latter did not contribute to the blue colony population because colonies would be white. Thus, the following calculation was used to determine crossover frequencies as in Figure 3-3:

$$\sum_{n=1}^{\infty} X^{2n-1} - \sum_{n=1}^{\infty} X^{2n} = b$$

Where b was the corrected blue colony frequency (47.4% in the example in Figure 3-3). In the case of the HIV-1 vectors example in Figure 3-3, this yielded a target region crossover frequency (X) of roughly 0.86. Note, however, that as a result of the calculation methods used, values close to one become somewhat inaccurate and thus these values are approximations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Genome-wide recombination frequency calculations. Recombination events per genome were calculated based on transduction-type recombination assays. Assumptions included: 1) Two template switching events were required to generate functional *lacZ*. 2) The region between Fsp I and Bsi WI of the acceptor template (2633 bases of the sequence between the LTRs) was the recombination target. 3) Native HIV-1 genomes are ~3.5 times this length and switching could occur at similar frequencies per distance for all genome segments. 4) Some *lacZ* genes that would have been functional were inactivated by mutation (2). 5) Only half the double crossovers were detected because whether donor or acceptor RNA templated the deletion region was random. 6) The frequency of each recombination event was independent of others. 7) The inability to score abortive products that would have resulted from switch to the 3' donor and failure to switch back before reaching acceptor non-complementary sequences between R and the internal *lacZ* fragment would not appreciably affect values, nor would omission of calculated crossings-over at frequencies other than two. 8) Crossings-over between the two RNAs that involved only sequences between the two R's of each RNA, which were not detectable because they did not change the *lacZ* allele, did not occur at significant frequency.

Using these assumptions to calculate recombination frequency with transduction-type assay values, the average frequency of blue colony formation (36.4%, using the MLV values) was divided by 0.92 to correct for *lacZ* inactivation by the MLV replication machinery as assayed in mATRC1/293 cells (2), and doubled to account for the possibility of either allele in the deletion region (yielding 0.79). Because this is the frequency of two crossovers in 2.6 kb and each is assumed independent, the frequency of one crossover in this interval is the square root of that value (0.89). This value is then multiplied by 3.5 to account for genome length, to yield a final value of 3.1 crossovers per genome per replication cycle. Note, however, that we believe that some of these parameters, most notably assumptions 7 and 8, are most likely incorrect. Because these

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

events probably did occur and did reduce apparent recombination rates, we believe these calculations likely underestimate template switching frequencies.

Recombination frequency calculations based on HIV-1 two-vector assays. RNA copackaging was assumed to be random. The averaged HIV-1 crossover frequency value of 0.85 for the 1.8-kb interval can be multiplied by 5.1 to account for the whole genome's length, to yield a genome-wide frequency of approximately 4.3 crossovers per replication cycle.

Calculating predicted deletion frequencies for reciprocal repeat vectors. Assuming equal expression levels when reciprocal vectors were coexpressed, and random copackaging of RNAs, the recombining population in these experiments would be half the total population. For each vector (L*a*acPuro and La*a*cPuro), any attempt at intramolecular repeat deletion would be unsuccessful (1). The assumption was made that for reasons such as the limited time spent copying the downstream repeat, the replication machinery would have an opportunity to consider only one potential acceptor region. If the replication machinery were equally likely to seek an acceptor template region on the same vector as on the copackaged vector, then only the half that sought an acceptor in the copackaged vector would have the potential of generating a deletion product. Therefore, the predicted repeat deletion frequency for coexpression of two reciprocal vectors would be half of one half, or 25% of the frequency of single, identical repeat vector deletion.

RESULTS

Genetic recombination frequencies for HIV-1 and for MLV. The evidence that HIV-1 recombines more frequently than gammaretroviruses is derived from a comparison of reports about MLV and SNV to results for HIV-1 (4, 19, 20, 22, 23, 43, 51, 52).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

However, each of these previous studies involved a single type of retrovirus, and because the work was performed in several separate laboratories which each used different assay vectors, it remained possible that some components of the reported differences were assay-specific or due to variation in experimental approaches among researchers. Thus, our initial studies were aimed at determining whether or not HIV-1's higher rate of recombination would remain evident when both viruses were assayed under parallel conditions.

Virus-specific recombination assay vectors were developed to allow a direct comparison of recombination frequencies during HIV-1 replication to those for MLV. Two different vectors, each derived from previously described LacPuro vectors (Figure 3-1A (2, 35)) were used in each virus system (Figure 3-1B and C). The first contained a 117-base coding region deletion within a *lacZ* expression cassette, followed by a puromycin resistance cassette (MLV L^cPuro and HIV L^cPuro for MLV and HIV, respectively; see Figure 3-1B). The second vector contained intact *lacZ* alone (MLV Lac or HIV Lac; Figure 3-1C). The rationale in using these vectors was as follows: Reverse transcription of the first vector would yield a puromycin resistant provirus that did not express functional β-galactosidase; the second vector alone would yield a *lacZ* provirus that did not confer puromycin resistance. During reverse transcription of virions that had copackaged one of each vector RNA, template switching within regions of sequence identity could yield recombinant viral DNAs, some of which would contain functional versions of both *lacZ* and puromycin resistance genes (Figure 3-2A). The frequency of such recombinants could be scored as that fraction of the total puromycin resistant colonies that stained blue when incubated with X-gal. Because they provided the scored genetic material, MLV Lac and HIV Lac were referred to as the donor template vectors; L^cPuro vectors with disrupted *lacZ* sequences were the acceptor vectors. This system allowed us to compare recombination rates for HIV-1 and MLV using the same producer and recipient cell types for each virus, virus-specific replication proteins, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

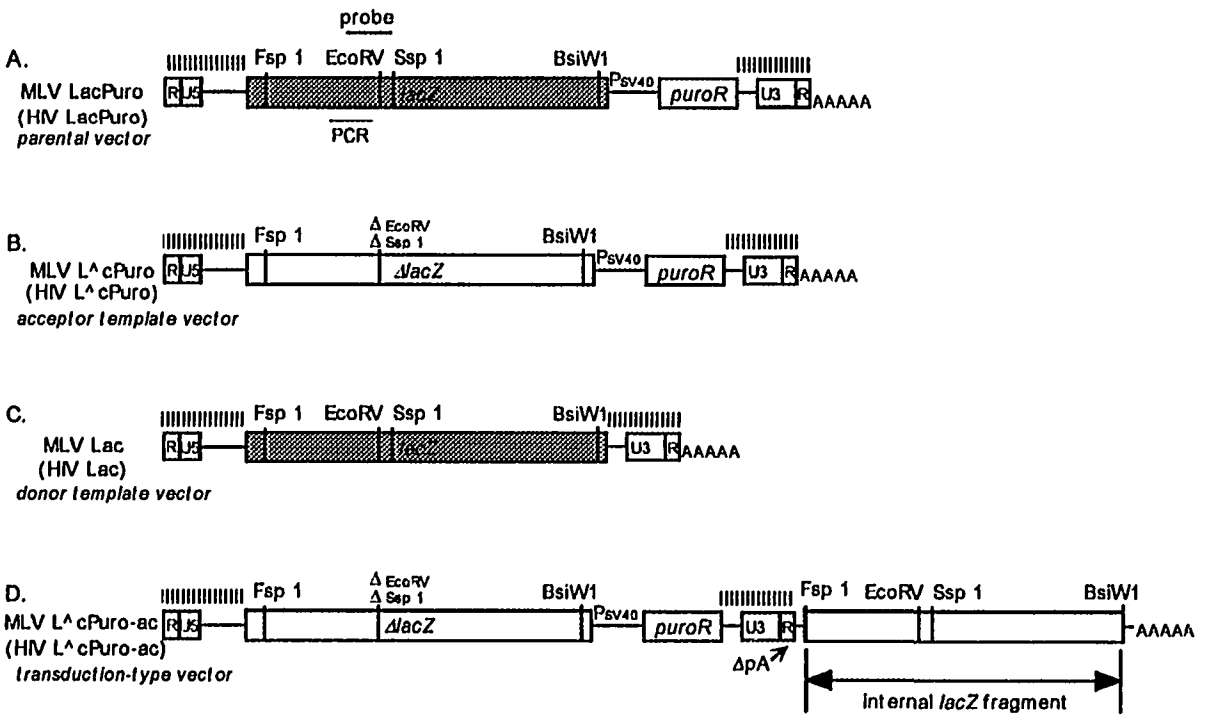


Figure 3-1. Inter- and intramolecular recombination assay vectors. The vectors as drawn were for MLV. Shaded boxes represent intact functional *lacZ* genes. PSV40 indicates the SV40 promoter; puroR refers to the gene that confers puromycin resistance. Positions of restriction sites used in construction and analysis are indicated. MLV based vector MLV LacPuro **(A)**, MLV L^cPuro **(B)**, MLV Lac **(C)**, and MLV L^cPuro-ac **(D)** are indicated at the left, with the name of the corresponding HIV-1 based vector in parentheses. The regions containing virus-specific *cis* acting sequences, which differ between the vectors for MLV and those for HIV-1, are indicated by hatched bars over the diagrams. Note that the vector called HIV Lac here is pHR'CMVlacZ. The bar marked "probe" indicates the location of sequences complementary to the RNA probe used for the RNase protection assays presented in Figure 3-3. The bar labeled "PCR" marked the position of the PCR product templating random primed probes for the Southern blot in Figure 3-4.

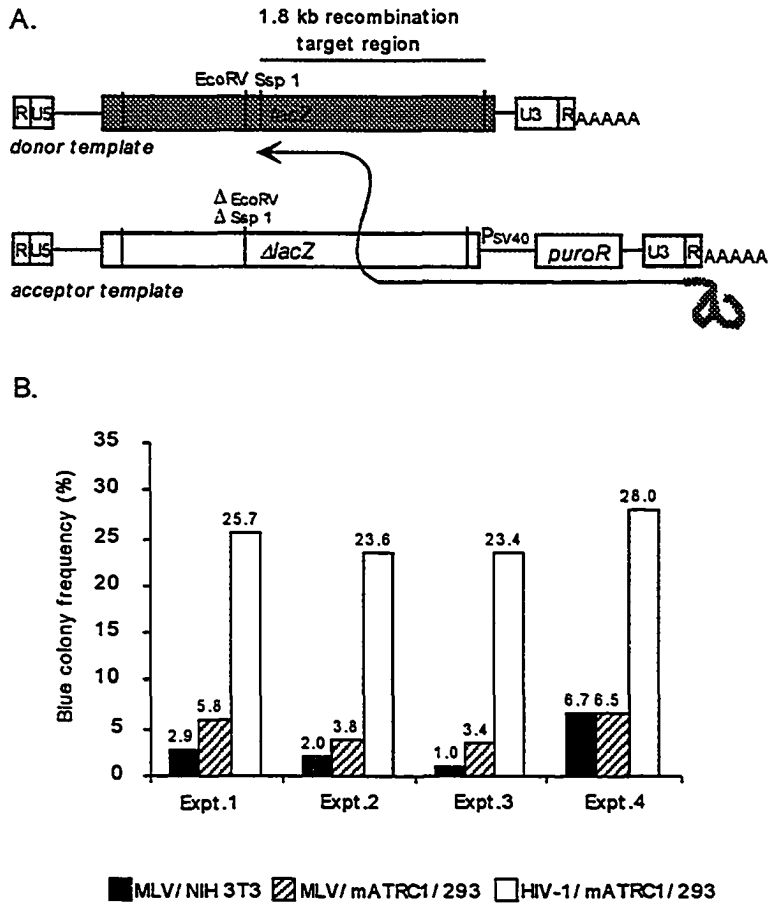


Figure 3-2. Two-vector (intermolecular) genetic recombination assay. **(A)** Schematic representation of how template switching during reverse transcription could yield a puromycin-resistant vector with functional lacZ. Line with arrowhead represents minus strand DNA synthesis after minus strand strong stop transfer. **(B)** A comparison of MLV and HIV-1 based two vector assay results, presented as blue per total puromycin resistant colony values. Hatched boxes, values for MLV vectors in human cells; open boxes, HIV-1 results; black boxes, values for MLV vectors assayed in murine cells. All values shown are results from four separate cotransfection and infection experiments, and at least 400 puromycin-resistant colonies were tallied to yield each value. Note that although the four MLV vector titers from murine cells are grouped with data for both MLV and HIV in human cells, the murine cell data shown were collected from experiments performed after the human cell experiments. The data shown for human cells is from the same experiments as those used to generate the data in Figure 3-3. Although performed later, the reported murine cell experiments were controlled with and normalized to human cell experiments that were performed in parallel.

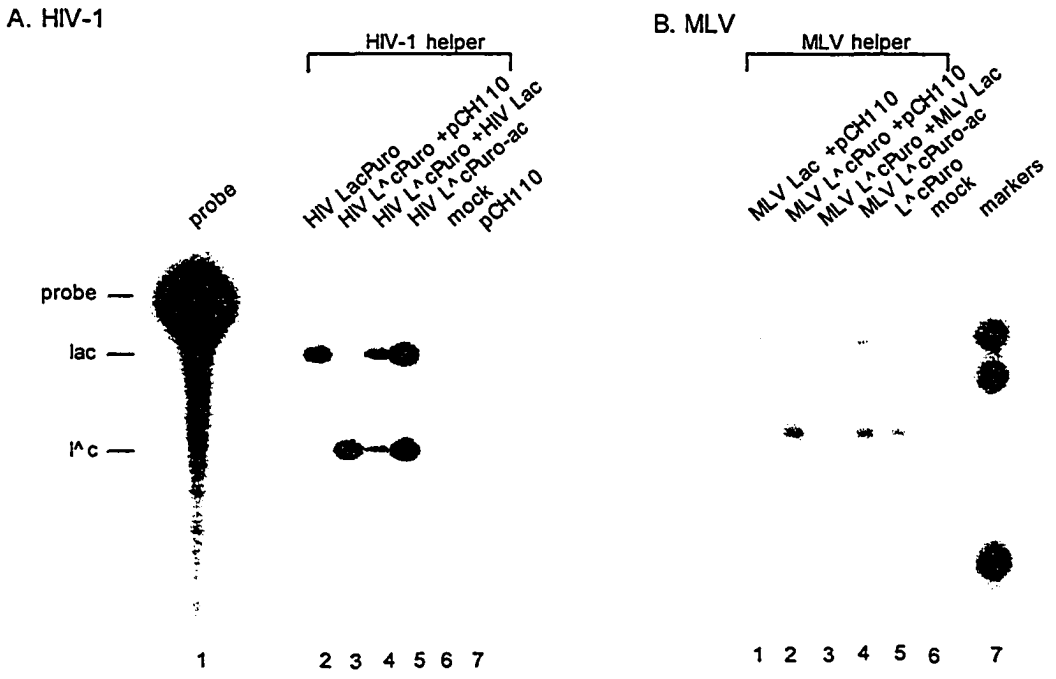
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

recombination assay vectors that differed only in virus-specific *cis*-acting sequences such as those in the long terminal repeats and packaging regions.

To perform these experiments, producer cells were cotransfected with helper plasmids and approximately equimolar amounts of plasmids encoding the two recombination vectors. Fresh cells were infected, and puromycin resistant colonies were selected and stained with X-gal. Results from four independent cotransfection experiments for each virus, followed by infection experiments that generated more than 200 colonies on each of two or three plates for each virus stock, are shown in Figure 3-2B. In each experiment, four to seven fold more of the total puromycin resistant colonies stained blue—indicative of recombination—for the HIV-1 based vectors (open bars) than for the MLV products (hatched bars).

To more accurately quantify recombination rates, the portion of colonies produced by virions predicted to contain two different RNAs was calculated by first determining virion RNA ratios and then using these ratios to deduce the size of the heterozygous virion "recombining population". RNA ratios within the population of virion RNA were determined by RNase protection assays, using RNA prepared from aliquots of the same virion stocks used in the infection experiments above (Figure 3-3A and B). The crossover frequency in the assayed 1.8-kb interval (final column in Figure 3-3C) was calculated using observed blue colony frequencies and virion RNA ratios as follows. First, the recombining population, which was the proportion of virions predicted to contain one of each vector, was calculated assuming random RNA copackaging (see materials and methods). Next, the fraction of this recombining population that generated functional *lacZ* was calculated. Finally, values were adjusted to reflect the fact that puromycin-resistant products with odd numbers of crossovers in the assayed interval would yield blue colonies, while those with even numbers of crossovers would yield white colonies (see materials and methods). The results are summarized in Figure 3-3C. Comparing the averages of the values calculated for two independent assays for MLV to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



C.

		blue total	vector proportions acceptor	donor	recombining population	corrected blue per recombining population	crossover frequency per 1.8 kb
MLV	expt. 1	5.8%	.54	.46	46%	13.5%	16%
	expt. 2	3.8%	.57	.43	43%	9.6%	11%
HIV-1	expt. 1	25.7%	.44	.56	57%	47.4%	86%
	expt. 2	23.6%	.44	.56	56%	45.8%	84%

Figure 3-3. RNase protection assay to quantify virion RNAs. Positions corresponding to the mobilities of undigested probe, probe protecting a fragment of undisrupted *lacZ*, and probe protecting a fragment of the disrupted *lacZ* allele are indicated by probe, lac, and l^c, respectively, at the left. **(A)** RNase protection assay of virion RNAs harvested from ET producer cells that had been cotransfected with the indicated HIV-1 based vector, the HIV-1 virion protein helper construct (pCMVΔR8.2), and/or control plasmids. Lane 1, undigested RNA probe used for both panels. Lane 2, digested products of virion RNA from cells cotransfected with the HIV-1 helper plasmid and HIV LacPuro. Lane 3, products from cells with helper, HIV L^cPuro, and the non-retroviral lacZ mammalian expression plasmid, pCH110 (16). Lane 4, helper, HIV L^c,Puro and HIV Lac. Lane 5, helper and HIV L^cPuro-ac, Lane 6, helper alone. Lane 7, helper and pCH110. **(B)** MLV-based vector products analyzed by RNase protection assay. The same probe as in panel A was used. Lane 1, virion RNA from cells cotransfected with the MLV helper, pMLVΨ-, MLV Lac, and the non-retroviral lacZ mammalian expression plasmid, pCH110 (16). Lane 2, helper, MLV L^cPuro and pCH110. Lane 3, helper, MLV L^cPuro, and MLV Lac. Lane 4, helper and MLV L^cPuro-ac. Lane 5, helper and MLV L^cPuro, Lane 6, helper alone. Lane 7, size standards. **(C)** Comparison of crossover frequencies for MLV and HIV-1 based vectors. Values presented are from two independent experiments for each virus and were calculated as described in the text.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

those for HIV-1 suggested that the total number of crossovers per presumptive recombining population member for HIV-1 was roughly 6.3-fold higher than for MLV.

Several controls specific to each set of experiments were performed to address the possibility that recombination of plasmid DNA during transfection—which is known to occur and which might generate products that inaccurately resembled virus replication-associated recombinants (10, 32)—might contribute to frequencies of recombination reported in this paper. The first of these, which was performed as a parallel control for every repetition of every transfection experiment reported in this paper, was to cotransfect assay vectors with a large molar excess of the non-retroviral mammalian *lacZ* expression plasmid, pCH110 (16). All assays described here score retroviral recombination by screening for LacZ function. Thus, the rationale for this control was that although pCH110 RNAs are Ψ^- and should not be encapsidated, the extended region of homology to vector genes suggests pCH110 plasmid DNA should be at least as capable of generating plasmid DNA-level recombinants as any of the utilized vector plasmids. Virions harvested from these cotransfections were used to infect fresh cells and puromycin resistant colonies were screened for β-galactosidase activity. Although blue-staining colonies were detectable in some control repetitions, these arose at a frequency significantly less than 1% of the white colony titer, which suggests DNA-level recombination did not affect outcomes measurably.

Another control also addressed the potential impact of DNA-level recombination on this assay's results. If plasmid DNAs had recombined during transfection to generate a transcription template for a product that would erroneously score as a retroviral recombinant, a recombinant RNA in virions would result. The experiment that suggests such products did not arise at a frequency high enough to affect assay outcomes is shown in Figure 3-3 lane 3A. If DNA-level recombination had generated recombinants to a detectable level, this RNase protection assay lane would contain a protected product with the mobility indicated by 'lac' at the panel's left. However, this autoradiogram and darker

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

exposures (not shown) indicate clearly that if any such product existed, it was present at a level significantly lower than 1% of the non-recombinant product with mobility indicated by 'l^c' in lane 3A.

Other controls were performed to verify the differences between HIV-1 and MLV intermolecular recombination frequencies. To address whether or not the lower frequency of recombination observed for MLV could be attributed to the lack of host species-specific factors in human cells, MLV recombination frequencies were measured in murine cells (black boxes in Figure 3-2B) as well as in human cells (hatched boxes in Figure 3-2B). As in the human cells, blue-white colony ratios observed in murine cells were low, thus confirming that the higher recombination rate of HIV was not due to host effects on MLV recombination. To address the possibility that recombination was as high for MLV as it was for HIV but that most colonies with MLV recombination products aberrantly remained unstained, the structures of pooled MLV reverse transcription products were examined by Southern blot (Figure 3-4). The absence of detectable recombination-diagnostic Lac band among products of virions from vector co-producing cells whereas parental L^cPuro vector product was clearly visible (lane 4), is consistent with expectations if MLV recombination were much lower than that of HIV.

Template switching independent of copackaging variables. The above results confirmed that marker segregation was more frequent for HIV-1 than for MLV, but calculated recombination frequencies relied heavily on the assumption that RNA copackaging was random. Thus, a second assay, which relied on a single vector with both the disrupted *lacZ* acceptor and the source of recombinational repair on a single RNA, was developed (Figure 3-1D and Figure 3-5). Every virion that packaged this single vector RNA copackaged template switch donor and acceptor sequences in equal amounts.

Vector design was based on putative retroviral transduction intermediates (44). It has been postulated that retroviral transduction of host sequences results from a series of rare events that start when a lengthened RNA, which results from read-through of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

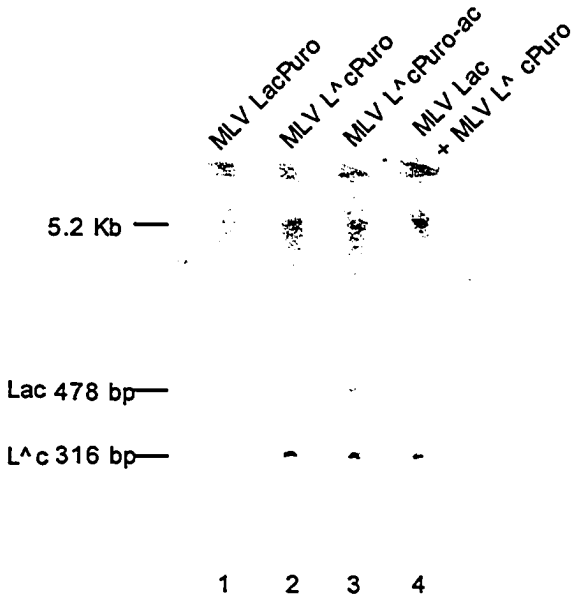


Figure 3-4. Southern blot of pooled integrated MLV vector products. Genomic DNA was collected from pooled MLV provirus-containing colonies for each of the vectors or vector pairs indicated at the top of the autoradiogram. This DNA was digested with ClaI and MluI, which produced a 5.6-kb band for all vectors as well as a shorter band diagnostic of either the intact (478-bp from Lac) or the corresponding deleted (316-bp from L^c) ClaI to MluI region of *lacZ*. The position of the PCR fragment that was used in the random priming reaction is indicated in Figure 3-1A. Lengths of predicted products are indicated at the left. The products in lane 3 are from transduction-type vectors used in the experiments described in Figure 3-5.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

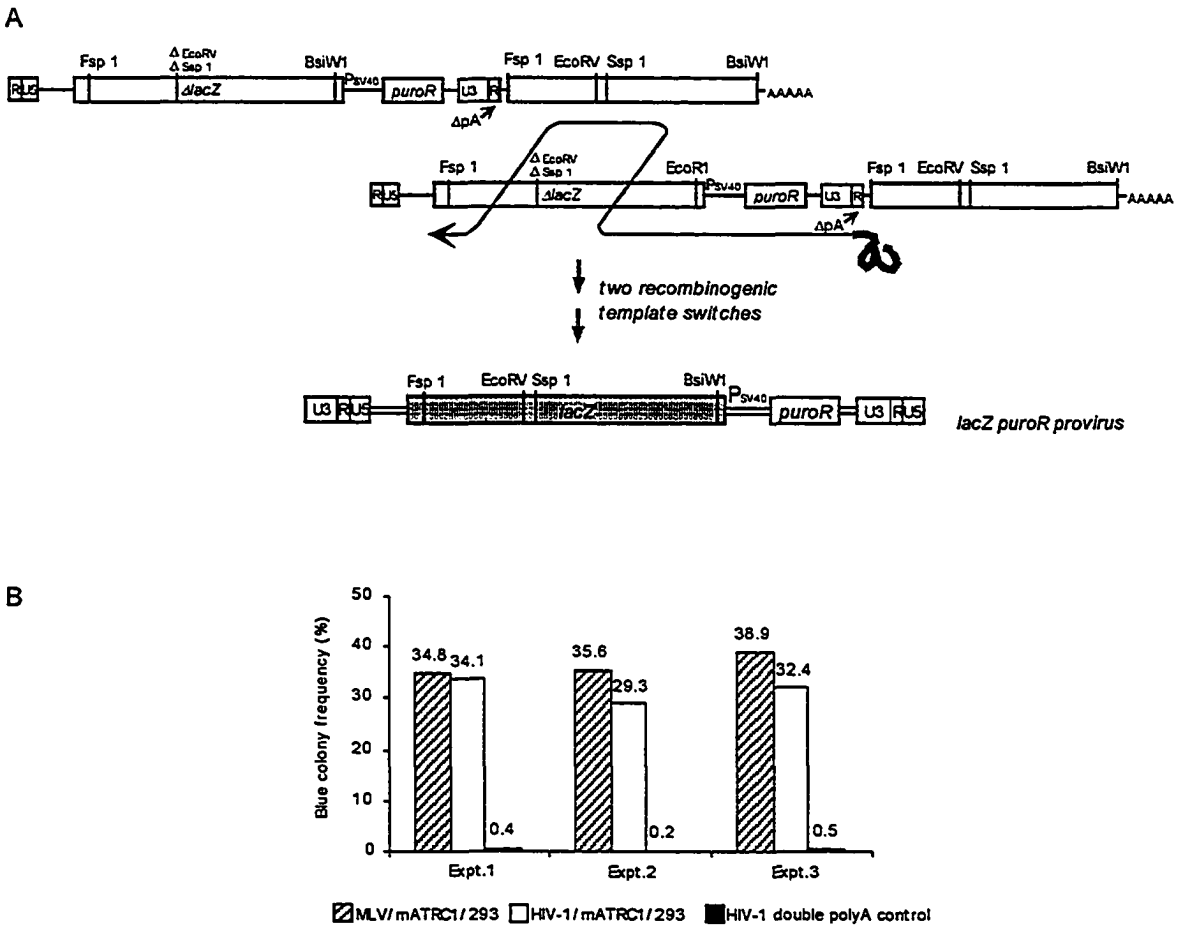


Figure 3-5. One-vector transduction-style recombination assay. **(A)** Transduction/patch repair process that would generate recombination-repaired *lacZ* (indicated by shaded box) during reverse transcription of L^cPuro-ac transduction-type vectors (see Figure 3-1D). Only those steps of minus strand synthesis that follow minus strong stop template switch are shown. **(B)** Comparison of MLV and HIV-1 based transduction-type assay results, presented as blue per total puromycin resistant colony values. Each value shown is the results of a separate transfection and infection experiment, and at least 400 puromycin resistant colonies were tallied to yield each value. Shaded boxes, values for MLV; hatched boxes, HIV-1 results; black boxes, results for the HIV-1 double polyA transduction vector, as described in the text.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

retroviral polyadenylation signals into adjacent cellular sequences, is packaged into virions. The cellular sequences in these RNAs' 3' ends are believed to subsequently become patched into the retroviral genome via rare non-homologous recombination events (44, 54).

In this report's transduction-type vectors, polyadenylation signals of MLV and HIV L^cPuro (Figure 3-1B) were removed to allow constitutive 3' extension into an appended internal portion of the *lacZ* coding sequence (indicated by ΔpA in the L^cPuro-ac vectors, Figure 3-1D). RNase protection assays confirmed that both disrupted (from the body of the vectors) and non-disrupted sequences (from the region downstream of the inactivated polyadenylation signals) were present at equimolar amounts in virions that contained L^cPuro-ac RNAs (Figure 3-3 lanes A5 and B4). Despite the unusual structures of these vectors, they were not excessively lengthy: at ~9.7 kb, the HIV-1 based transduction vector was only about 5% longer than native HIV-1, and the MLV version was actually shorter than MLV.

It has been demonstrated that reverse transcription of a polyadenylation read-through transcript can proceed normally despite the RNA's elongated 3' end (45). Thus, our expectation was that reverse transcription of the transduction vectors ordinarily would involve copying only those sequences that were between the two R repeats. This would generate puromycin resistant transductants with *lacZ* genes containing the engineered disruption. However, each encapsidated vector RNA would bring into virions a source of recombinational patch repair in its extended 3' end. Thus, if two template switches occurred between the disrupted *lacZ* gene in a vector's body (the acceptor template region) and the fragment of *lacZ* at the 3' end of either the same or the copackaged RNA (the recombination donor sequences), a functional *lacZ* gene could be generated (Figure 5A).

Using these vectors, template switching frequencies for HIV-1 and MLV could be compared by a direct determination of the portion of puromycin-resistant colonies that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

stained blue with X-gal. These experiments were performed by cotransfecting vector plasmid with the cognate helper plasmid into producer cells, infecting fresh target cells with the resulting virus, and determining X-gal stained to total (stained and unstained) puroymycin resistant colony ratios. The results of three independent transfection and infection experimental repetitions for each virus are shown in Figure 3-5C. Surprisingly, blue colony formation was frequent: about 32% for HIV-1 and a similar average frequency of 36% for the MLV vector. The ratios of Lac to L^c bands observed in Figure 3-4 blot of transduction vector products (lane 3) appear roughly consistent with these values.

Controls were performed in which a molar excess of pCH110, a non-retroviral mammalian *lacZ* expression plasmid (16), was cotransfected along with helper and the deletion-containing *lacZ* expression vectors that lacked 3' extended regions (MLV L^cPuro and HIV L^cPuro, see Figure 3-1B). The resulting virions generated no X-gal staining colonies (<0.01% blue colonies) for the HIV-1 based vector, and very low levels of blue colonies ($\leq$0.5%) for the MLV-based vector, thus suggesting that DNA-level recombination during transfection did not contribute appreciably to the observed findings. A second control provided additional support for the assertion that plasmid-level recombination did not cause transduction-related patch repair. For this control, a second HIV transduction-type vector was constructed, which contained both the native HIV polyadenylation signal and a second polyadenylation signal derived from SV40, in the downstream LTR between the vector body and the appended 3' end *lacZ* sequences (not shown). Data on the fraction of puromycin-resistant products of this "double polyA" HIV transduction-type vector that stained blue are presented in Figure 3-5C (black bars). Because of the similarity of the plasmid encoding this vector to that for the assayed HIV L^cPuro-ac vector, these values likely represent the maximum approximate contribution of plasmid DNA-level recombination to the reported transduction values.

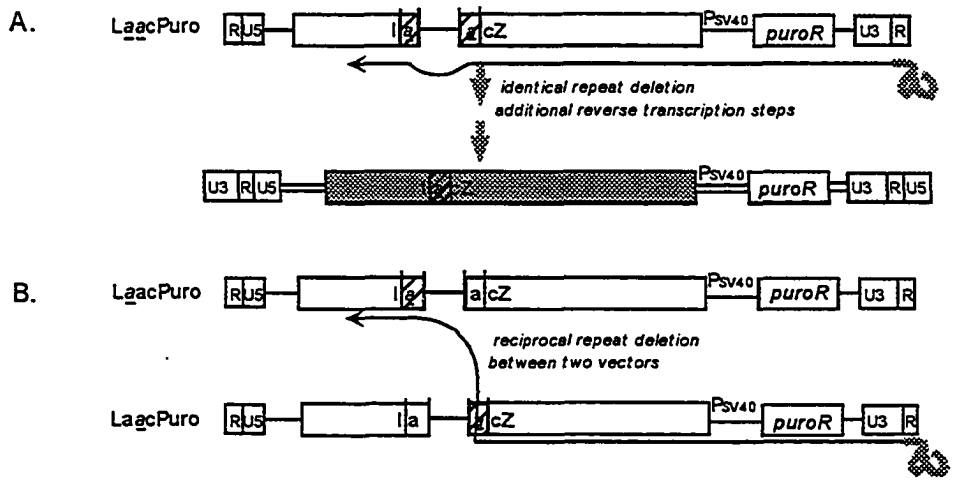
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Intermolecular recombination using reciprocal repeat vectors. By demonstrating that recombination rates differed when acceptor and donor were on separate RNAs but were similar when acceptor and donor were covalently joined, the above results suggested that what differed between HIV-1 and MLV was the accessibility of coexpressed acceptor and donor vector sequences to the viral recombination machinery. A prediction of this hypothesis was that if both switching between two RNAs and switching between two positions on a single RNA were possible, more intermolecular switching would be detectable for HIV-1 than for MLV. Here, this notion was tested with reciprocal repeat vectors. Repeated sequences are deleted frequently during reverse transcription, in a process presumed to be related to recombinogenic template switching (25, 38, 53). Because the coexpressed vectors used in this set of experiments were highly similar to one another, these experiments also addressed whether or not the differences in intermolecular template switching frequencies above resulted because the coexpressed vectors differed significantly. As in the work described above, experiments to monitor possible DNA-level recombination using the pCH110 control were included with these experiments, and the results suggested that DNA-level recombination did not significantly affect the following assays' results.

Aided by the degeneracy of the genetic code, a *lacZ* substitution variant of HIV LacPuro (Figure 3-1A) was constructed. This variant contained a 156 base pair gene segment that was identical to the parental allele in coding capacity but that differed by 27% in nucleic acid sequence due to the substitution of about one nucleotide per codon throughout the 156 base pair region. As assayed by the production of X-gal staining puromycin resistant colonies, the degenerate gene segment served as well as wild type as a functional part of *lacZ* (data not shown).

Coding region-internal duplications were incorporated into *lacZ* to generate tandem repeat derivatives (Figure 3-6). Repeat deletion during reverse transcription of these vectors was assayed by the restoration of *lacZ* function (35) (Figure 3-6A). Vectors

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



C.

	vector(s)	repeats	HIV-1 blue/total	MLV blue/total
	LaacPuro	(wt-wt)	15.5±0.6%	16.6±0.6%
	LaacPuro	(73%-73%)	14.3±0.6%	24.9±0.8%
	LaacPuro	(73%-wt)	<0.04%	<0.05%
	LaacPuro	(wt-73%)	<0.28%	<0.06%
LaacPuro &	LaacPuro	(73%-wt and wt-73%)	4.1±0.2%	0.6±0.1%

Figure 3-6. Reciprocal repeat recombination assay. **(A)** Schematic representation of LaacPuro type repeat deletion assay vector as described in the text. Template switching between repeated sequences would result in a single copy of the repeat, which would yield a colony that expresses functional β-galactosidase (functional gene indicated with shaded box). Backwards slashes and the underscore and italics (as indicated by *a*) indicate that the repeated region in this vector was engineered to differ 27% from the wild type sequence (see text.) **(B)** Schematic representation of template switching between co-packaged RNAs that would result in the generation of a functional *lacZ* gene. 'a' box without backwards slashes is the wild type sequence. **(C)** Observed frequencies of functional *lacZ* among vector products. Values presented are averages from 3 to 5 independent experimental repetitions for each vector. Vector designations reflect the repeats in vectors with structures shown in panels A and B. For the first vector, both repeats contained the parental wild type (wt) *lacZ* sequence. The repeat vector for which both repeats contained silent mutations that resulted in a 27% difference from the parental allele was designated L*aa*cPuro. Where La*a*cPuro + L*a*acPuro is indicated, the value presented was for products of virions that resulted from cells co-transfected with helper and both these vector plasmids. The values presented for both La*a*cPuro and L*a*acPuro alone indicate that no blue colonies were observed for either virus using any of these vectors in any experimental repetitions and thus there are no differences among these; the values as presented differ slightly due to the different numbers of colonies scored.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were generated that contained exact duplications of the wild type or of the degenerate gene segment (HIV LaacPuro and HIV L*aa*cPuro; the mutant repeat allele is indicated by an italic underlined a (*a*) in Figure 3-6C). As observed previously, when repeats were identical to one another, repeat deletion occurred at similar frequencies regardless of whether the wild type or the degenerate repeat was studied (Figure 3-6C, (1)). When the two different synonymous repeats were placed in tandem in a single vector, this introduced sufficient genetic distance between repeats that homologous recombination—as measured by repeat deletion-- did not occur (HIV L*a*acPuro vector; Figure 3-6C). The reciprocal vector (HIV La*a*cPuro), which differed from HIV L*a*acPuro only in the orientation of the two repeats, also showed no evidence of deletion among more than 2000 colonies examined. However, when HIV L*a*acPuro and HIV La*a*cPuro were coexpressed and cells were infected with the resulting virions, blue-staining colonies were observed (Figure 3-6C). These findings suggest that the two different vector RNAs were copackaged and template switching occurred between them (Figure 3-6B). The frequency with which reciprocal, intermolecular repeat-deletion recombinants arose was similar to values predicted by models in which repeat deletion between randomly copackaged RNAs or within a single RNA were equally likely (see materials and methods and discussion).

These repeats were also incorporated into MLV-based vectors (final column, Figure 3-6C). As was the case for HIV-1 based vectors, identical repeats were efficiently deleted. Also similar to results with HIV-1 vectors, no repeat deletion was detected when the two different repeats were assayed in single vectors. However, reciprocal repeat deletion for coexpressed MLV vectors was about seven-fold lower than that observed for HIV-1. Overall, these MLV findings were consistent with previous genetic studies of gammaretrovirus repeat deletion (11, 52), but demonstrate that HIV-1 and gammaretroviruses differ in that template switching events which genetically appear principally intramolecular for MLV can readily occur intermolecularly for HIV-1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DISCUSSION

HIV-1 and MLV virus-specific assay vectors were developed to directly compare frequencies of genetic recombination during replication by these viruses. Despite the fact that the experiments presented here described results for virions generated by transient transfection, the results mirrored findings implicit in a comparison of previous reports performed with virions generated by stable vector producing cells (20, 22). Namely, recombination between two viral genomes occurs significantly more frequently for HIV-1 than for the gammaretrovirus, MLV. However, when transduction-type assays that assured copackaging of recombination donor and acceptor sequences were employed, template switching was as frequent for MLV as it was for HIV-1. Results with reciprocal repeat vectors confirmed that template switching events which others have shown are principally intramolecular for MLV (11, 52) can occur intermolecularly for HIV-1, and showed that the differences in recombination frequency observed with our initial set of recombination vectors could not be ascribed to sequence dissimilarity of donor and acceptor vectors. These findings demonstrate that when RNA availability is a constant, MLV and HIV-1 template switching occur at the same frequency, and provide genetic evidence that coexpressed vectors are less accessible to the recombination machinery during MLV replication than is the case for HIV-1.

The cause of this difference in RNA accessibility was not studied directly in this report, but the speculation we provide here outlines a few of many possible causes of this difference. The simplest explanation for our findings may be that recruitment of genomic RNA for encapsidation differs between MLV and HIV-1, with HIV-1 apparently selecting RNAs for copackaging from among coexpressed RNAs in a more random

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

fashion than that employed by MLV. It has generally been assumed that the way coexpressed RNAs become copackaged is as partners in mixed RNA heterodimers. Whether or not two different RNAs can heterodimerize has not been measured directly, and attempts to visualize heterozygote RNA dimers by electron microscopy have been unsuccessful (21). Our results raise the untested possibility that coexpressed MLV RNAs may not readily heterodimerize. Although on first consideration this seems unlikely, the literature provides examples of differences between HIV-1 and MLV on which speculative hypotheses that support this notion can be based. RNA trafficking and processing for lentiviruses such as HIV-1 are known to differ in several ways from the analogous processes for simple retroviruses (37), and the findings presented here may suggest that MLV and HIV-1 also differ in how RNAs are selected for copackaging. Our data might indicate a preferential affinity of one dimerizable MLV RNA for another RNA templated by the same provirus in coinfected cells. Speculatively, MLV might accomplish this by such means as a co-transcriptional commitment to dimerization or a more stable dimer precursor, in contrast to an apparent more random selection of dimerization partner at a later replication phase or a dimerization intermediate that is more susceptible to RNA exchange for HIV-1.

The current findings cannot rule out the alternate possibility that HIV-1 and MLV differ when recombination occurs between two vectors but are the same when recombination occurs between two regions on a single vector. In other words, MLV may be less capable of switching between copackaged RNAs than is HIV-1. However, because extended retroviral genome-length RNAs are more than 50 times as long as virions are wide, virion RNA must be highly condensed. Thus, there seems no *a priori* reason to anticipate that the acceptor and donor sequences in the transduction-style vectors, which are separated by half a genome length, would be in closer proximity during reverse transcription than would two sequences on two different RNAs. It is also noteworthy that genomic RNA is degraded by ribonuclease H during retroviral DNA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

synthesis (46). As a result, the acceptor and donor regions of our transduction-type vectors were probably parts of separate RNA molecules at the time of the assayed template switches. However, other virus-specific differences may prove to be the cause of our apparent copackaged RNA accessibility differences. For example, HIV-1 and MLV are known to differ in viral components that are retained in post-entry nucleoprotein complexes (14, 15). It is possible that our observed recombination properties reflect differences in how copackaged HIV-1 and MLV RNAs are related to one another within intracellular reverse transcription complexes.

One attractive feature of the model that HIV-1 and simple retroviruses differ in accessibility of coexpressed RNAs is that it provides a plausible explanation for the genetic phenomenon of "high negative interference". High negative interference, which has been observed for both replication-competent retroviruses (8) and among gammaretroviral vector products (3, 4), is the appearance in simple retroviruses of multiple non-selected crossovers at a far greater frequency than would be predicted if each recombination event were independent. Most gammaretrovirus DNAs produced by virions from coinfected cells are non-recombinant, but the relatively rare recombinants often contain evidence of multiple template switches. To explain negative interference, it has been suggested that only a subset of virions are recombination-competent or else that although recombination generally does not occur, when it does, one recombination event predisposes the replication machinery to performing additional template switches (17, 33). Our findings are consistent with an alternate explanation for negative interference: namely, for gammaretroviruses the subset of viral particles that are produced by cells coinfected by two different viruses and that copackage two different RNAs is smaller than that predicted if the copackaging of two coexpressed RNAs is random. HIV-1 does not appear to display negative interference (22).

In addition to explaining negative interference, it seems likely that differences in virion RNA availability account for reported differences between HIV-1 and simple

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

retroviruses in whether or not strong stop switching can occur between two RNAs (20, 24, 48, 51). For HIV-1, evidence of intermolecular strong stop switching is common (48, 51). However, for gammaretroviruses, genetic evidence suggests that strong stop switching generally only occurs between the two ends of a single kind of RNA (20, 24). The findings presented here suggest that these differences between HIV and simple retroviruses may reflect differences in the accessibility of genetically distinct RNAs rather than enzymological differences.

Although calculations still require some untested assumptions—most of which likely lead to underestimates—the transduction vector data here provide new information on which estimated retroviral recombination rate calculations can be based (see materials and methods). Factoring in the measured frequencies of marker assortment, *lacZ* inactivating mutation rates (2), and the assumption that template switching can occur at a similar frequency at any genome position (55), our transduction vector data suggest that recombinogenic template switching occurs approximately three times during the synthesis of every retroviral DNA. Applying similar assumptions to the two-vector data yields an estimated crossover frequency of about four per replication cycle (see materials and methods). These values represent a significantly higher recombination frequency than those previously reported for gammaretroviruses (4, 19, 20, 24), but they are similar to some previous values for HIV-1 (22, 51). Based on our findings that HIV-1 and MLV template switching rates are indistinguishable when assayed with transduction-type vectors, we hypothesize that the biochemical act of recombinogenic template switching is as frequent for MLV as it is for HIV-1. However, the genetic consequences of recombinogenic switching are greater for HIV-1 than for MLV due to differences in which coexpressed RNAs can serve as templates during reverse transcription.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

BIBLIOGRAPHY

1. **An, W., and A. Telesnitsky** 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:**7897-902.

2. **An, W., and A. Telesnitsky** 2001. Frequency of direct repeat deletion in a human immunodeficiency virus type 1 vector during reverse transcription in human cells Virology. **286:**475-82.

3. **Anderson, J. A., E. H. Bowman, and W. S. Hu** 1998. Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation J Virol. **72:**1195-202.

4. **Anderson, J. A., V. K. Pathak, and W. S. Hu** 2000. Effect of the murine leukemia virus extended packaging signal on the rates and locations of retroviral recombination J Virol. **74:**6953-63.

5. **Bray, M., S. Prasad, J. W. Dubay, E. Hunter, K. T. Jeang, D. Rekosh, and M. L. Hammarskjold** 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev- independent Proc Natl Acad Sci U S A. **91:**1256-60.

6. **Burke, D. S.** 1997. Recombination in HIV: an important viral evolutionary strategy Emerg Infect Dis. **3:**253-9.

7. **Butsch, M., and K. Boris-Lawrie** 2002. Destiny of unspliced retroviral RNA: ribosome and/or virion? J Virol. **76:**3089-94.

8. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses J Gen Virol. **42:**1-26.

9. **Cullen, B. R.** 2000. Nuclear RNA export pathways Mol Cell Biol. **20:**4181-7.

10. **de Saint Vincent, B. R., and G. M. Wahl** 1983. Homologous recombination in mammalian cells mediates formation of a functional gene from two overlapping gene fragments Proc Natl Acad Sci U S A. **80:**2002-6.

11. **Delviks, K. A., and V. K. Pathak** 1999. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching J Virol. **73:**7923-32.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

12. **Dorman, N., and A. Lever** 2000. Comparison of viral genomic RNA sorting mechanisms in human immunodeficiency virus type 1 (HIV-1), HIV-2, and Moloney murine leukemia virus J Virol. **74:**11413-7.

13. **Dupont, S., N. Sharova, C. DeHoratius, C. M. Virbasius, X. Zhu, A. G. Bukrinskaya, M. Stevenson, and M. R. Green** 1999. A novel nuclear export activity in HIV-1 matrix protein required for viral replication Nature. **402:**681-5.

14. **Fassati, A., and S. P. Goff** 2001. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1 J Virol. **75:**3626-35.

15. **Fassati, A., and S. P. Goff** 1999. Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus J Virol. **73:**8919-25.

16. **Hall, C. V., P. E. Jacob, G. M. Ringold, and F. Lee** 1983. Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells J Mol Appl Genet. **2:**101-9.

17. **Hu, W. S., E. H. Bowman, K. A. Delviks, and V. K. Pathak** 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference J Virol. **71:**6028-36.

18. **Hu, W. S., V. K. Pathak, and H. M. Temin** 1993. Role of reverse transcriptase in retroviral recombination, p. 251-74, Reverse transcriptase. Cold spring harbor laboratory press, Plainview, NY.

19. **Hu, W. S., and H. M. Temin** 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:**1556-60.

20. **Hu, W. S., and H. M. Temin** 1990. Retroviral recombination and reverse transcription Science. **250:**1227-33.

21. **Hunter, E.** 1978. The mechanism for genetic recombination in the avian retroviruses Curr Top Microbiol Immunol. **79:**295-309.

22. **Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome J Virol. **74:**1234-40.

23. **Jones, J. S., R. W. Allan, and H. M. Temin** 1993. Alteration of location of dimer linkage sequence in retroviral RNA: little effect on replication or homologous recombination J Virol. **67:**3151-8.

24. **Jones, J. S., R. W. Allan, and H. M. Temin** 1994. One retroviral RNA is sufficient for synthesis of viral DNA J Virol. **68:**207-16.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

25. **Julias, J. G., D. Hash, and V. K. Pathak** 1995. E- vectors: development of novel self-inactivating and self-activating retroviral vectors for safer gene therapy J Virol. **69:**6839-46.

26. **Katz, R. A., and A. M. Skalka** 1990. Generation of diversity in retroviruses Annu Rev Genet. **24:**409-45.

27. **Levin, J. G., P. M. Grimley, J. M. Ramseur, and I. K. Berezesky** 1974. Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with actinomycin D J Virol. **14:**152-61.

28. **Malhotra, S., A. G. Scott, T. Zavorotinskaya, and L. M. Albritton** 1996. Analysis of the murine ecotropic leukemia virus receptor reveals a common biochemical determinant on diverse cell surface receptors that is essential to retrovirus entry J Virol. **70:**321-6.

29. **Malim, M. H., and M. Emerman** 2001. HIV-1 sequence variation: drift, shift, and attenuation Cell. **104:**469-72.

30. **Naldini, L., U. Blomer, F. H. Gage, D. Trono, and I. M. Verma** 1996. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector Proc Natl Acad Sci U S A. **93:**11382-8.

31. **Naldini, L., U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector [see comments] Science. **272:**263-7.

32. **Olson, P., H. M. Temin, and R. Dornburg** 1992. Unusually high frequency of reconstitution of long terminal repeats in U3-minus retrovirus vectors by DNA recombination or gene conversion J Virol. **66:**1336-43.

33. **Pathak, V. K., and W. S. Hu** 1997. "Might as well jump!" Template switching by retroviral reverse transcriptase, defective genome formation, and recombination Seminars in Virology. **8:**141-50.

34. **Pfeiffer, J. K., and A. Telesnitsky** 2001. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication J Virol. **75:**11263-74.

35. **Pfeiffer, J. K., R. S. Topping, N. H. Shin, and A. Telesnitsky** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription J Virol. **73:**8441-7.

36. **Preston, B. D., and J. P. Dougherty** 1996. Mechanisms of retroviral mutation Trends Microbiol. **4:**16-21.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

37. **Rabson, A. B., and B. J. Graves** 1997. Synthesis and processing of viral RNA, p. 205-262. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

38. **Rhode, B. W., M. Emerman, and H. M. Temin** 1987. Instability of large direct repeats in retrovirus vectors J Virol. **61:**925-7.

39. **Ricchetti, M., and H. Buc** 1990. Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent Embo J. **9:**1583-93.

40. **Robson, N. D., and A. Telesnitsky** 2000. Selection of optimal polypurine tract region sequences during moloney murine leukemia virus replication J Virol. **74:**10293-303.

41. **Scheifele, L. Z., R. A. Garbitt, J. D. Rhoads, and L. J. Parent** 2002. Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein Proc Natl Acad Sci U S A. **99:**3944-9.

42. **Shin, N. H., D. Hartigan-O'Connor, J. K. Pfeiffer, and A. Telesnitsky** 2000. Replication of lengthened Moloney murine leukemia virus genomes is impaired at multiple stages J Virol. **74:**2694-702.

43. **St Louis, D. C., D. Gotte, E. Sanders-Buell, D. W. Ritchey, M. O. Salminen, J. K. Carr, and F. E. McCutchan** 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:**3991-8.

44. **Swain, A., and J. M. Coffin** 1992. Mechanism of transduction by retroviruses Science. **255:**841-5.

45. **Swain, A., and J. M. Coffin** 1989. Polyadenylation at correct sites in genome RNA is not required for retrovirus replication or genome encapsidation J Virol. **63:**3301-6.

46. **Telesnitsky, A., and S. P. Goff** 1997. Reverse transcriptase and the generation of retroviral DNA, p. 121-160. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

47. **Temin, H. M.** 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation Proc Natl Acad Sci U S A. **90:**6900-3.

48. **van Wamel, J. L., and B. Berkhout** 1998. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly Virology. **244:**245-51.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

49. **Vogt, P. K.** 1973. The genome of avian RNA tumor viruses: A discussion of four models, p. 35-41. *In* l. Sylvestri (ed.), Possible episomes in eukaryotes, Proceedngs of the Fourth Lepetit Colloquium. North Holland, Amsterdam.

50. **Yu, H., A. E. Jetzt, and J. P. Dougherty** 1997. Use of single-cycle analysis to study rates and mechanisms of retroviral mutation Methods. **12:**325-36.

51. **Yu, H., A. E. Jetzt, Y. Ron, B. D. Preston, and J. P. Dougherty** 1998. The nature of human immunodeficiency virus type 1 strand transfers J Biol Chem. **273:**28384-91.

52. **Zhang, J., and Y. Ma** 2001. Evidence for retroviral intramolecular recombinations J Virol. **75:**6348-58.

53. **Zhang, J., and C. M. Sapp** 1999. Recombination between two identical sequences within the same retroviral RNA molecule J Virol. **73:**5912-7.

54. **Zhang, J., and H. M. Temin** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication Science. **259:**234-8.

55. **Zhang, J., and H. M. Temin** 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone J Virol. **68:**2409-14.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER FOUR

HIGH FREQUENCY OF TRANSDUCTION-TYPE RECOMBINATION DURING HIV-1 REPLICATION

INTRODUCTION

Retroviral transduction of cellular genes has been well documented for oncoretroviral species (6, 69), and was initially evidenced by discovering the cellular counterpart (c-*onc* genes) of viral oncogenes (v-*onc*) in acute transforming retroviruses (59). The v-*onc* genes were found to be embedded in viral genomes, replacing part of the viral sequences and rendering most of the acute transforming retroviruses replication defective. Several models have been proposed for oncogene acquisition by retroviruses, diverging in how the 5' and 3' viral/cellular junctions are formed (5, 18, 19, 23, 50, 53, 56, 60, 63, 65, 73).

A currently favored model postulates that the retroviral transduction of host sequences results from a series of rare events that are intrinsic to the replicative strategy of retroviruses (61, 67). The model is summarized as follows. During the retroviral replication cycle, viral RNA is converted by virally coded reverse transcriptase into a double-stranded DNA form, and the latter is integrated into the host cell chromosome, serving as the template for new viral RNA synthesis by host RNA polymerase II. Retroviral polyadenylation signals are leaky, and it has been shown that readthrough

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

transcripts amount to approximately 15% of the total viral RNA in avian leucosis virus (ALV) (22). These readthrough transcripts can be efficiently encapsidated into virions and serve as templates for reverse transcription (23, 64). The cellular sequences at the 3' ends of these RNAs are thought to subsequently become patched into the retroviral genome via rare non-homologous recombination events (63, 74). Following oncogene capture, it is likely that some oncogenes acquired mutations as the result of error-prone reverse transcription before they became oncogenic (51, 66).

In contrast to simple retroviruses, retroviral transduction of cellular genes by lentiviruses such as HIV-1 is largely unknown. Scattered experimental evidence suggests the possibility of transduction by HIV-1 (54, 62). For example, a 147-bp cellular sequence has been identified in the proviral 5' long terminal repeat (LTR) from cultured cells infected with HIV-1 vector viruses, although the transduced cellular sequence was not linked to the site of the viral integration (62). In a separate study, an HIV-1 readthrough transcript was reported in chronically infected promonocytes (15). The efficiency of the HIV-1 polyadenylation signal has been measured in studies on the mechanism of polyadenylation site selection in HIV-1. The level of readthrough RNAs ranged from being undetectable to as high as 9% (4, 7, 8, 10, 13, 14, 31, 57). In addition, results from a recent study suggest that the murine leukemia virus (MLV) polyadenylation site is significantly weaker than the equivalent HIV-1 processing site (16), a finding that was later confirmed by another group (72).

Recently, we developed a transduction-type recombination assay to circumvent the copackaging variable in typical recombination studies (42). The vector used contained both a recombination donor and acceptor, with the acceptor sequence located downstream to the 3' LTR in which the viral polyadenylation signal was inactivated. RNA transcribed from this vector was reminiscent of a transduction intermediate—a readthrough RNA that captured cellular sequence at its 3' end. Not surprisingly, due to the extended homology between the donor and acceptor sequences, recombination was observed at a very high

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

frequency (42). To better understand the nature of this transduction-type recombination, we sought to characterize the template switch junctions. The acceptor sequence was genetically marked with scattered single point mutations that allowed the differentiation of donor from acceptor sequences by restriction analysis of individual proviral clones. Our finding confirmed that the appended 3' sequence could be frequently used as recombination acceptor in the presence of significant homology between the donor and acceptor. Surprisingly, non-homologous recombination was also detected in about 3% of the proviral clones. In addition, using a vector containing a wild-type polyadenylation signal in the 3' LTR, the HIV-1 polyadenylation signal readthrough rate was determined to be as high as 7%, a rate consistent to the level of the transduction-type recombination detected for the vector.

MATERIALS AND METHODS

Plasmids and cells. The helper plasmid used, pCMVΔR8.2 (39), encodes all HIV-1 *trans*-acting proteins except envelope, driven by a CMV immediate early promoter. The constitutive readthrough vector, HIV L^cPuro-ac*, has previously been described as HIV L^cPuro-ac (42). It is an HIV-1 based retroviral vector reminiscent of a retroviral transduction intermediate, and is composed of two parts: a body and a tail. The body refers to the genetic elements located between the 5' and 3' LTRs, including a CMV immediate early promoter-driven *lacZ* gene disrupted by a 117-bp deletion between the EcoRV and SspI sites in *lacZ*, and an SV40 early promoter-driven puromycin-resistance gene. The downstream LTR and flanking regions were modified by deletion of the HIV-1 polyadenylation sequence in R. The tail refers to the genetic sequence downstream of the polyadenylation-inactivated R regions, including a 2.6-kb

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

FspI-BsiWI fragment of *lacZ* followed by an SV40 polyadenylation signal sequence (indicated by the character * in the vector name). The transduction-intermediate-type vector for determining the template switch junctions, HIV L^cPuro-äc*, was derived from HIV L^cPuro-ac*. The tail *lacZ* allele was genetically marked with 10 point mutations (the marked *lacZ* allele is represented by character ä in the vector name), creating 10 restriction sites for PstI, XbaI, NheI, HindIII, KpnI, PvuII, NcoI, MfeI, AclI, and BclI, respectively. The natural readthrough vector, HIV L^cPuro*-äc* was derived from HIV L^cPuro-äc* by restoring the wild-type HIV-1 polyadenylation signal in 3' LTR (the restored HIV-1 polyadenylation signal is represented by the first * character in its name). The double polyadenylation signal control vector, HIV L^cPuro**-ac*, was a derivative of HIV L^cPuro*-äc* with the insertion of an SV40 polyadenylation signal immediately downstream to the 3' LTR (the additional polyadenylation signal is represented by the second character * in HIV L^cPuro**-ac*). The wild-type *lacZ* vector, HIV LacPuro, has been previously described (42). It is an HIV-1 vector containing a *lacZ* gene driven by a CMV immediate early promoter and a puromycin resistance gene driven by an SV40 early promoter. The vector HIV ac-ac* contained two copies of a 1.8-kb FspI-SacI fragment of *lacZ* gene. The two *lacZ* alleles could be differentiated from each other by single synonymous point mutation that was introduced into the region corresponding to the 117-bp deletion in HIV L^cPuro*. Construction details for all plasmids are available upon request.

ET cells are a 293T derivative that constitutively express ecotropic envelope (48). mATRC1/293 cells are derived from a human 293 cell line and stably express the receptor for ecotropic Env (35). These cell lines and their derivatives were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Gemini).

Transfection and virus harvesting. Vector viruses were generated by transiently transfecting ET cells with pCMVΔR8.2 (39) and the appropriate transducing vector

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

plasmid by calcium phosphate precipitation (9). Virus was harvested by replacing culture medium 16-24 hours after calcium phosphate exposure and collecting medium 36-48 hours post-transfection. Residual cells were removed by 4°C low speed centrifugation for 5min at 3000g. Virus was stored at –70°C prior to use. At least two viral stocks were prepared from separate transfections for each transducing vector.

Infection and functional assay for *lacZ* activity. mATRC1/293 cells in 60-millimeter-diameter dishes were infected 1 h in the presence of 8 μg/ml of hexadimethrine bromide (polybrene; Sigma). Puromycin (Sigma) selection, cell fixation, and 5-bromo-4-chloro-β-D-galactopyranoside (X-gal; Sigma) staining were performed as previously reported (48) except that infected cells in each 60-millimeter-diameter dish were subcultured into two 100-millimeter-diameter dishes before puromycin selection. Infection with each virus stock was performed in duplicate or triplicate dishes. The numbers of blue and white colonies from all dishes were summed for each infection experiment, and were used to calculate the recombination rates.

Cell cloning, PCR and restriction analysis. For isolation of individual proviral clones, infection with HIV L^cPuro-äc* vector virus was performed at a very low multiplicity of infection (MOI) so that well-dispersed cell colonies were formed after puromycin selection. For this particular infection experiment, puromycin selection was performed without the prior subculture step as stated in the previous section. A total of 159 puromycin-resistant provial clones were isolated from three 100-millimeter-diameter cell culture dishes. Individual clones were transferred to 24-well plates, and expanded into 100-millimeter-diameter dishes after several passages. The *lacZ* expression was determined for every proviral clone by X-gal staining of cells derived from the same clone in a separate dish. Cells harvested from additional parallel dishes were frozen and stored in liquid nitrogen. Genomic DNA for use in PCR was isolated using a Wizard Genomic DNA Purification Kit (Promega). Four pairs of primers were used to amplify the *lacZ* gene for restriction analysis. These and other primers specific to various genetic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

elements from the vectors were used to map the sequence structure of the seven clones. All primer sequences are available upon request. PCR was performed in appropriate buffer (10 mM Tris, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 250 μM dNTPs, 5% DMSO, and 50 pmol each primer) for 30 cycles with an annealing temperature of 58°C. Expand Long Template System (Roche) was used for PCRs on long templates. *LacZ* PCR products were digested with appropriate restriction enzymes, resolved on 1% agarose gels, and recorded using Kodak Digital Science 1D Gel Documentation System.

Virus RNA quantification. Virion RNA was purified from filtered culture media from transfected cells as previously described (58) and used in RNase protection assays. Vector RNAs were differentiated in RNase protection assays and their ratios quantified by PhosphorImager (Molecular Dynamics). To allow differentiation of both vectors with a single probe, the probe was complementary to the regions between ClaI and SspI in *lacZ* and thus included sequences complementary to the region deleted in the body *lacZ* allele of transduction-intermediate type vectors. To quantify vector RNAs, RNase protection assay background values were subtracted and molarities were normalized by dividing protected product values by the number of radiolabeled nucleoside monophosphate residues in each.

RESULTS

Examining template switch junctions using a vector reminiscent of a transduction intermediate. Previously we developed an HIV-1 recombination assay that relied on a single vector with both the disrupted *lacZ* acceptor and the source of recombinational repair on a single RNA (42). Vector design was based on putative retroviral transduction intermediates. In this transduction-intermediate type vector, viral

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

polyadenylation signals located in the 3' LTR were mutationally inactivated to allow constitutive 3' extension into an appended internal portion of the *lacZ* coding sequence (Figure 4-1A). Reverse transcription of such readthrough RNAs would ordinarily involve copying only those sequences that were between the two R repeats. This would generate puromycin resistant transductants with *lacZ* genes containing the engineered disruption. However, each encapsidated vector RNA would bring into virions a source of recombinational patch repair in its extended 3' end. Thus, if two template switches occurred between the disrupted *lacZ* gene in a vector's body (the recombination acceptor region) and the fragment of *lacZ* at the 3' end of either the same or the copackaged RNA (the recombination donor sequences), a functional *lacZ* gene could be generated (Figure 4-1B). Using these vectors, template switch frequencies for HIV-1 could be assessed as the portion of puromycin-resistant colonies that stained blue with X-gal. These experiments were performed by cotransfecting vector plasmid with the cognate helper plasmid into producer cells, infecting fresh target cells with the resulting virus, and determining X-gal stained to total (stained and unstained) puromycin-resistant colony ratios. Blue colony formation was determined to be about 32% for HIV-1 (42).

To characterize the template switch patterns in this transduction-type recombination assay, we modified the vector used in the previous study by genetically marking the tail *lacZ* allele with 10 synonymous point mutations (Figure 4-1A, indicated by the 10 asterisks in the tail *lacZ* allele). These mutations carried the following properties. (1) They maintained the correct amino acid coding and were tested to be fully functional in the context of the full-length *lacZ* gene in a transfection assay (data not shown). (2) They were single nucleotide mutations, which introduced minimal sequence heterogeneity compared to wild-type *lacZ* gene. In fact, the mutated tail *lacZ* allele shared 99.6% identity with its wild-type counterpart. (3) The mutations were relatively evenly spaced in the tail *lacZ* allele, and defined 11 sequence intervals ranging in size from 175 bp to 287 bp (239 bp on average). (4) The mutations created ten new restriction

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

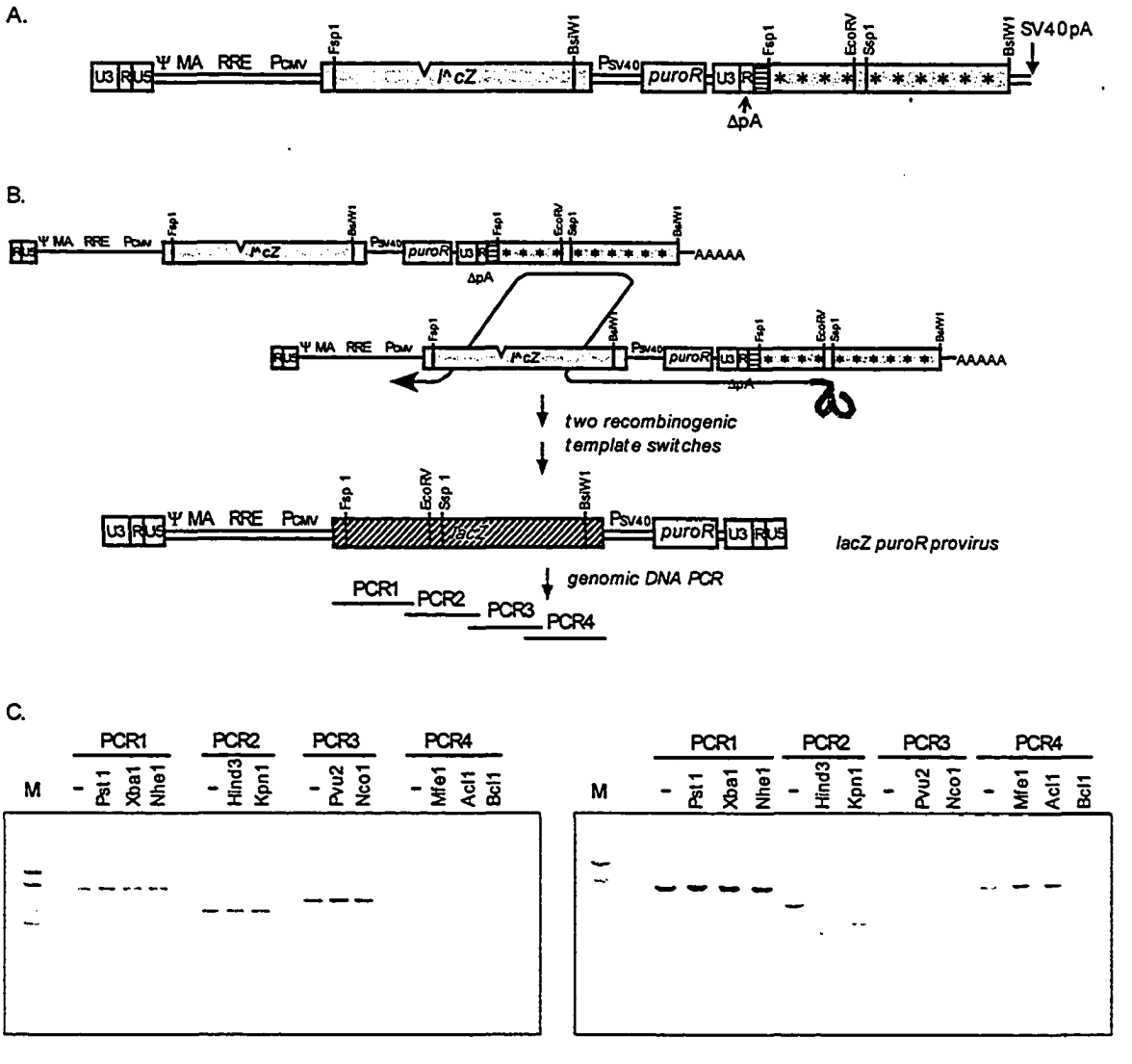


Figure 4-1. HIV-1 vector reminiscent of transduction intermediate and protocol for determining template switch junctions. **(A)** HIV-1 constitutive readthrough vector. Illustrated is HIV L^c(ΔpA)äc*, the vector with downstream *lacZ* allele genetically marked. It was identical to its parental vector, HIV L^c(ΔpA)ac*(42), except for the ten mutations introduced in the tail *lacZ* allele (indicated by the asterisks in the diagram). Ψ, packaging signal; MA, matrix protein coding sequence with frameshift mutation; RRE, Rev-responsive elements; PCMV, CMV immediate early promoter; *lacZ*, β-galactosidase gene; PSV40, SV40 early promoter; *puroR*, puromycin resistance gene. **(B)** Transduction-type recombination. Predicted template switch pattern for generating a blue provirus (*lacZpuroR*) is shown. Genomic DNA was extracted from individual puromycin-resistant proviral clones, and *lacZ* gene was amplified into four overlapping segments (PCR1 to PCR4). **(C)** Restriction pattern for a white proviral clone (left panel) and a blue proviral clone (right panel). The four PCR products for each clone were digested with appropriate restriction enzymes as indicated. Undigested DNA was loaded as size control (lane '-'). M, DNA size marker pUC19/ApaLI+HaeIII. Inverted images are shown.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sites in the tail *lacZ* allele, which enabled us to determine template switch junctions in recombinant proviral clones by restriction analysis.

The new vector, HIV L^cPuro-äc*, was used to transfect ET cells to produce virus. The frequency of recombination was comparable to that of the parental vector without genetic markers (data not shown). To isolate individual proviral clones, infection with HIV L^cPuro-äc* vector virus was performed at low MOI so that well-separated cell colonies were formed after puromycin selection. A total of 159 puromycin-resistant proviral clones were isolated, and the *lacZ* expression was determined for every proviral clone by X-gal staining. Thirty-two blue clones and 71 white clones were randomly selected for further analysis. Genomic DNA was amplified using 4 pairs of *lacZ* primers that would generate 4 consecutive end-overlapping PCR products (PCR1-4, Figure 4-1B). Note that the sense primer for PCR1 and the antisense primer for PCR4 were located in the regions of *lacZ* that were absent from the tail *lacZ* allele. Therefore, successful amplification of PCR1 or PCR4 product was indicative of recombinogenic template switch between the body and tail *lacZ* alleles. PCR products were digested with individual restriction enzymes, and separated on 1% agarose gels. Sample gels for one blue proviral clone and one white clone were shown in Figure 4-1C.

General characteristics of template switch patterns for blue and white proviral clones. The template switch pattern for each proviral clone was deduced from restriction analysis. Those for the 32 blue clones analyzed are schematically presented in Figure 4-2A, and shared the following genotypic characteristics. All four PCR reactions (PCR1 to PCR4) targeted to the *lacZ* gene were productive and were of sizes comparable to those templated by a wild-type *lacZ* gene. This suggested that for all blue clones at least two template switches took place between the body and tail *lacZ* alleles in regions flanking the 117-bp internal deletion located on the body *lacZ* allele. Indeed, restriction analysis revealed that all blue clones could be cleaved with KpnI whose recognition site was located only 21-bp downstream of the 117-bp deletion in the body *lacZ* allele,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

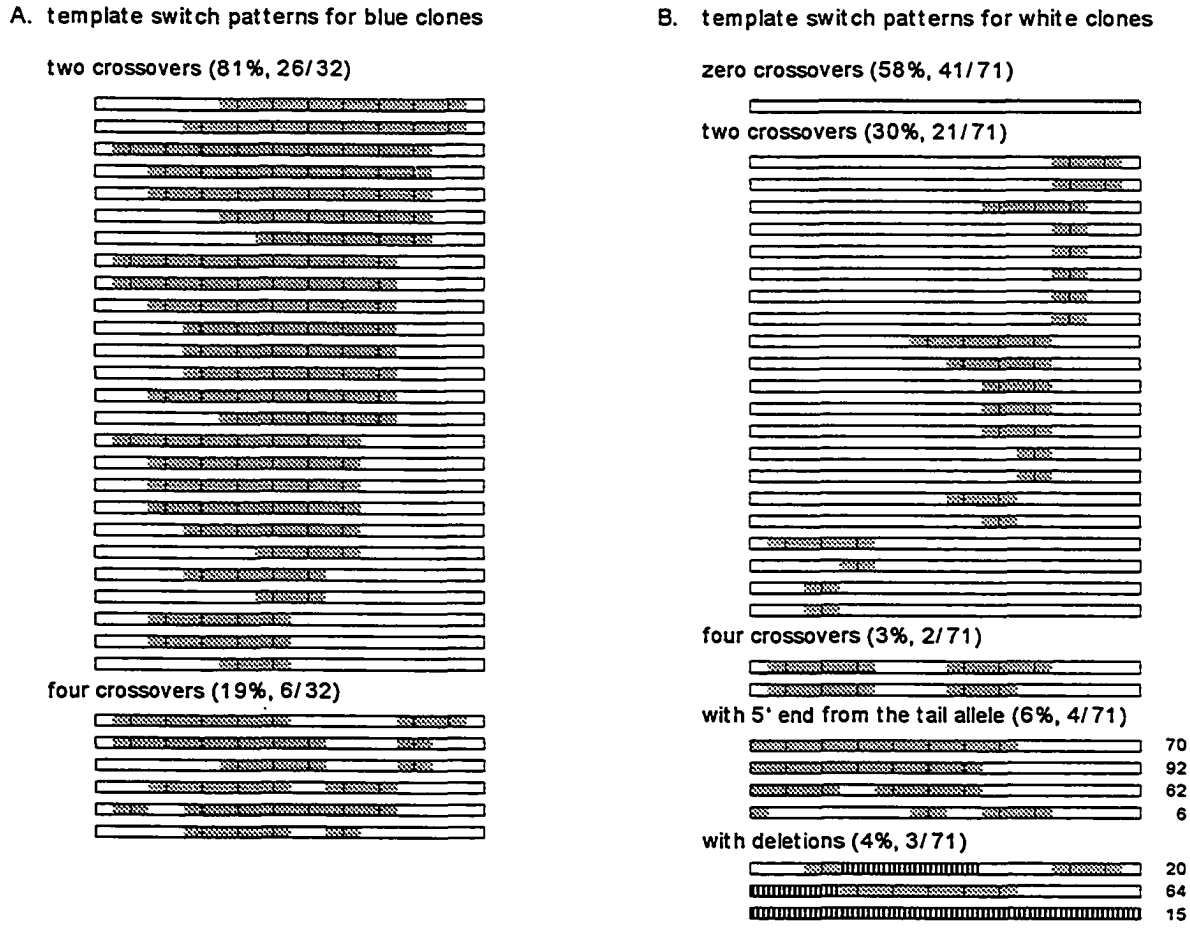


Figure 4-2. Summary of template switch patterns from restriction/sequencing analysis. The template switch pattern for each proviral clone was deduced from the restriction /sequencing analysis results, and is schematically represented. **(A)** Template switch profiles for 32 blue clones. **(B)** Template switch profiles for 71 white clones. Each horizontal bar represents the region of *lacZ* from an individual proviral clone that was characterized by restriction analysis, and for some, by sequencing. The origins of the sequence were differentiated by shading. The uncolored regions represent regions derived from the body *lacZ* allele, and the shaded regions represent those derived from the tail *lacZ* allele, with the short vertical lines marking the positions of the restriction sites. Deletions were shown as hatched regions. The clone numbers are indicated to the right of the corresponding clones for the 7 white clones shown in Figure 4-3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

confirming the occurrence of the patch repair event. Among the 32 blue clones, 81% (26/32) had two template switches, and 19% (6/32) had four crossovers (Figure 4-2A).

Among the 71 white clones characterized, seven clones failed at least one of the four PCR reactions, but the remaining 64 clones were positive for all four PCRs. The template switch patterns were deduced from restriction analysis on positive PCR products, as well as from sequencing results for those 7 clones that were further mapped with additional PCR reactions using primers that were specific to other regions of the provirus. The proviral structures of the aforementioned 7 clones are detailed in Figure 4-3. The majority of the white clones (64/71) had a 700-bp PCR2 product, equivalent to the size of a PCR2 product that was templated by the body *lacZ* allele, suggesting that these proviral clones still maintained the 117-bp deletion in the *lacZ* sequence. This observation was confirmed by restriction analysis that showed no cleavage by KpnI in any of these clones. Five clones had 817-bp PCR2 products but failed to acquire the 5' end of the *lacZ* gene, and the remaining 2 clones were negative for PCR2 (Figure 4-3). Among the 71 white clones, 58% (41/71) had no template switches, 30% (21/71) had two, 3% (2/71) had four, 6% (4/71) had an odd number of template switches, and 4% (3/71) had sequence deletions in this region (Figure 4-2B). Taken together, the above template switch pattern analysis unambiguously verified the experimental design of using transduction-intermediate type vector for studying recombination and template switching.

Template switch patterns for seven white clones. Seven white clones failed in one or more of the four *lacZ* PCRs, including clones 6, 15, 20, 62, 64, 70, and 92 (Figure 4-2B). Interestingly, five of them (clones 6, 62, 64, 70, and 92) produced a PCR2 product of 817 bp, suggesting that, like all the blue clones, a patch repair event occurred during reverse transcription. However, the PCR amplification failure suggested possible sequence alteration in the corresponding proviral *lacZ* gene, most likely due to aberrant template switch events during reverse transcription. Retroviral DNA replication is known to be an error-prone process, and a wide spectrum of genetic variations has been

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

identified (36, 37, 43, 45, 46, 68). To better understand HIV-1 reverse transcriptase's template switch properties, we set out to examine the overall sequence structure of these proviral clones.

We reasoned that there were at least two possible events that could generate proviruses with such genotypes. One was the occurrence of a second, nonhomologous template switch after the first homologous template switch as described below. Reverse transcriptase could first switch templates into the tail *lacZ* allele and copy the 117-bp patch that was deleted in the body *lacZ* allele. It might then switch to a non-homologous region that was upstream to the *lacZ* coding sequence in the body *lacZ* allele. Thus, the resultant provirus would lack the 5' end of the *lacZ* coding sequence and render PCR1 unfruitful. Alternatively, reverse transcriptase might keep going and copy all the way toward the 5' end of the new template RNA until it was forced to switch templates at the end of the current template. A prediction from the second possibility was that such clones would contain the sequence information that was located between the mutated 3' LTR and the tail *lacZ* allele. To explore these possibilities, primers specific to other sequence elements of the vector were used to map the approximate location of the 5' template switch points for those clones missing the 5' end of *lacZ*. PCR products were gel purified and subjected to automatic sequencing. Template switch junctions were deduced from the sequence information, and overall proviral structures for these clones are illustrated in Figure 4-3.

Indeed, sequencing results for one clone (clone 64) and four clones (clones 6, 62, 70, and 92) conformed to the first and second hypothesis, respectively. Clone 64 had a non-homologous template switch from a 5' internal position of the *lacZ* gene to a position upstream of the CMV promoter sequence, with 1.7-kb of sequence deleted in the resulting provirus. Sequence alignment indicated a 20-bp insertion at the template switch junction (Figure 4-3), which had perfect homology with the sequence from the U5/PBS junction. Interestingly, such a long distance duplication of genetic material from one

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

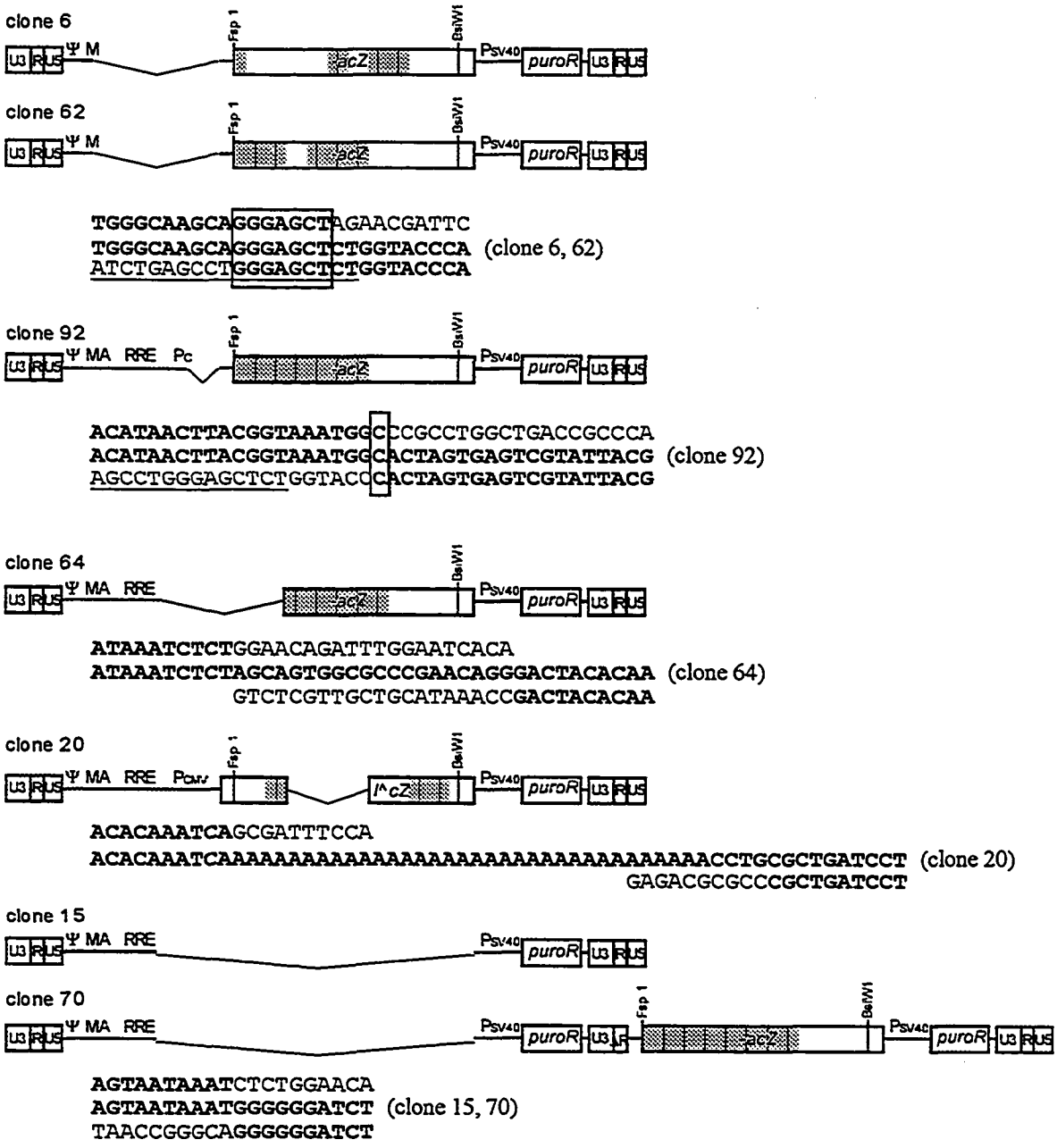


Figure 4-3. Proviral structure and template switch junctions for 7 white clones. Structures are drawn approximately to scale. Deletions are represented with downward lines of varied lengths. Sequence elements affected by deletions are marked by part of the identities of the corresponding elements. For example, M and PC indicate that the only the 5' portion of the matrix protein coding gene (MA) and CMV promoter (P_{CMV}) were left, respectively. Sequence from 3' R was underlined. The origins of the *lacZ* sequence were differentiated by shading. The uncolored regions represent regions derived from the body *lacZ* allele, and the shaded regions represent those derived from the tail *lacZ* allele, with the short vertical lines marking the positions of the restriction sites. Refer to Figure 4-1 legend for explanation of other symbols used. Template switch junctions that were involved in sequence deletions are shown under each proviral structure in the form of sequence alignment of the donor sequence (bottom sequence), the actual sequence (center), and the acceptor sequence (top sequence). The actual sequence and the homologous sequence in the donor and acceptor are highlighted in boldface font. See text for detailed description of individual clones.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

portion of the genome to another has previously been reported in another study in which the first 15 bases of the HIV-1 envelope gene were inserted in the RT coding region (34).

Sequence information indicated that clone 92 failed to switch back to the body lacZ allele. It switched templates after it copied through the tail *lacZ* allele, including the 22-base spacer sequence between the partially deleted R region and the tail *lacZ* allele, to a nonhomologous position located in the 5' portion of the CMV promoter. Sequence alignment of the template switch junctions suggested a single nucleotide sequence homology between the template switch donor and acceptor sequences (Figure 4-3).

Clones 6 and 62 had precisely identical template switch junctions by sequencing mapping (Figure 4-3). Similar to clone 92, both copied through the tail *lacZ* allele and even further copied 4 bases into the R region before switching templates. They switched to the matrix (MA) coding sequence at the 5' end of the genomic RNA, resulting a 1.9-kb deletion.

Clone 70 was a unique case in which the internal disrupted LTR structure was maintained in the provirus. It appeared that reverse transcriptase first switched to the tail allele of the copackaged long RNA that had not been used for minus strand DNA synthesis beyond the synthesis of minus strand strong stop DNA. Otherwise, most likely, the tail *lacZ* allele would be cleaved away from the body allele and the 3' U3-R sequence. The structure of this clone provided direct evidence that reverse transcriptase could use the tail *lacZ* allele from the copackaged RNA as a template switch acceptor.

Clone 20 had a nonhomologous template switch between the body *lacZ* allele and the tail *lacZ* allele that resulted in a 0.9-kb deletion (Figure 4-3). Interestingly, a stretch of homopolymeric A's was found between the junctions. It was speculated that two sequential non-homologous template switches were involved in its formation: viral RNA synthesis was first switched onto a poly(A) template, and then reverse transcriptase slipped away from the poly(A) and switched back to an upstream portion of the tail *lacZ* allele. In total, this clone had at least 5 crossovers in this region.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Clone 15 had a 4.2-kb deletion including the CMV promoter and the full length of the *lacZ* gene. The SV40 promoter and the puromycin resistance gene were intact which rendered the clone puromycin-resistant. Surprisingly, no sequence homology was involved in the template switch that caused the above deletion (Figure 4-3).

Frequency of template switch and genetic recombination based on genotypic analysis. Using the transduction-type vector HIV L^cPuro-äc*, the HIV-1 recombination rate could be assessed in two ways. Based on blue/white phenotypic screening, the recombination rate, defined as the ratio of the number of blue colonies to the total number of puromycin-resistant colonies, was calculated to be 24.6% (after corrected with the *lacZ* gene inactivation rate). Alternatively, the recombination rate could be evaluated based on proviral genotype. A proviral clone would be considered to be a recombinant if a template switch could be identified in the inspected genetic sequence. In other words, the detection of any genetic markers derived from the tail *lacZ* allele in the proviral *lacZ* sequence targeted by the four *lacZ* PCR reactions would place the proviral clone into the recombinant category.

To calculate the frequency of template switch and/or recombination, we did the following statistical analysis. In total, we characterized 32 blue clones and 71 white clones. The 32 blue clones consisted of all the blue clones from the infection, and, therefore, represented an unbiased sampling of blue clones. However, the 71 white clones were only a random subset of 126 white clones recovered. We could not simply combine the data from white clones with those from blue clones, since fewer white clones were analyzed than blue clones compared to the blue/white ratio in a large population. To reconstitute properties of the total population, we assigned a weighing factor to the data from blue clones before they were combined with those from the 71 white clones. The ratio of blue colonies to white colonies was 0.305 in large-scale infection experiments. If all proviral clones were picked randomly from such experiments, only 22 blue clones would have been chosen along with the randomly picked 71 white clones. To derive the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

number of template switches for 22 blue clones, the number of template switches from the 32 blue clones was multiplied by a weighing factor 0.68, i.e. 71*0.305/32 = 0.68).

Table 4-1. Frequency of template switch in characterized proviral clones

Categories	No. of clones	No. of crossovers per clone	Recombination rate
Blue	32	2.4 (76/32)	100% (32/32)
White	71	0.9 (67/71)	42% (30/71)
Total *	93	1.3 (118/93)	56% (52/93)

* Values in the total category have been standardized according to the blue/white ratio from large-scale infections (0.305, see text for methods of calculation).

According to this criterion, the genotype-based recombination rate was calculated to be 56% (Table 4-1), which was about two-fold higher than the recombination rate estimated based on phenotypic screening. Apparently, the genotype-based rate was more accurate because the phenotype screening would inevitably fail to differentiate recombinant white colonies from non-recombinant white colonies. In fact, our genotypic screening indicated that 42% of the white colonies were recombinants (Table 4-1). Nevertheless, the phenotypic screening method has proven to be a simple and valuable tool for studying retroviral recombination, especially for a two-vector assay in which a single template switch is sufficient in creating a viable recombinant provirus.

With the restriction analysis data, the frequency of template switch in the *lacZ* gene could also be evaluated. This value is more informative in comparing recombination rates among different studies. On average, RT switched templates 1.3 times during replication of the 2.6-kb *lacZ* target sequence (Table 4-1). This was equivalent to at least 4 crossovers per wild-type length of HIV-1 genome (9.2 kb), a value consistent to that estimated by two-vector recombination assays (42).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The tail allele frequency was examined by calculating the frequency of cutting by the 10 restriction enzymes. To generate Figure 4-4, the frequency of cutting was first summed at each position for blue clones and white clones, respectively. Then data from white and blue clones were combined by adjusting the ratio of blue clones among total clones to reconstruct a balanced sample. The combined data were then divided by the standardized sample size to derive the frequency of cutting. The distance from one site to another varies in the range from 165 bp to 282 bp. To be able to compare among different locations, the target genetic sequence was divided into 10 contiguous regions, with each region centered at the restriction site, extending 5' and 3' to the midpoints between two adjacent sites. The size of defined regions ranged 171~263 bp (225 bp on average). The above frequency data were then divided by 225 (the average size) and multiplied by 100 to derive the relative frequency of being cleaved at each position per 225 bases (Figure 4-4). The data varied from 4.6-33.4%, averaging 22.4%. It suggested that for every 225 bases, about 22% of them were copied from the tail *lacZ* allele, i.e. almost every 1 kilobase contains sequences from the other RNA.

Distribution of template switch sites. To look at whether there were hotspots for template switch within the target region, the sites of crossovers were examined. These crossover sites could be classified as either entering points or leaving points, with entering points defined as crossover sites where reverse transcriptase switched to the tail *lacZ* allele and leaving points defined as those where reverse transcriptase switched back to the body *lacZ* allele. The frequency of template switch was evaluated in the 11 regions delimited by the 10 restriction sites and the 5' and 3' ends of the tail *lacZ* allele. First, the number of entering and leaving points were tabulated for each region for all blue clones and all white clones, respectively, and they were summed after the data for the blue clones were normalized based on the number of white clones characterized. The combined data were then divided by the normalized sample size to derive the average probability of a template switch in each region for any single clone. Since the 11 regions

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

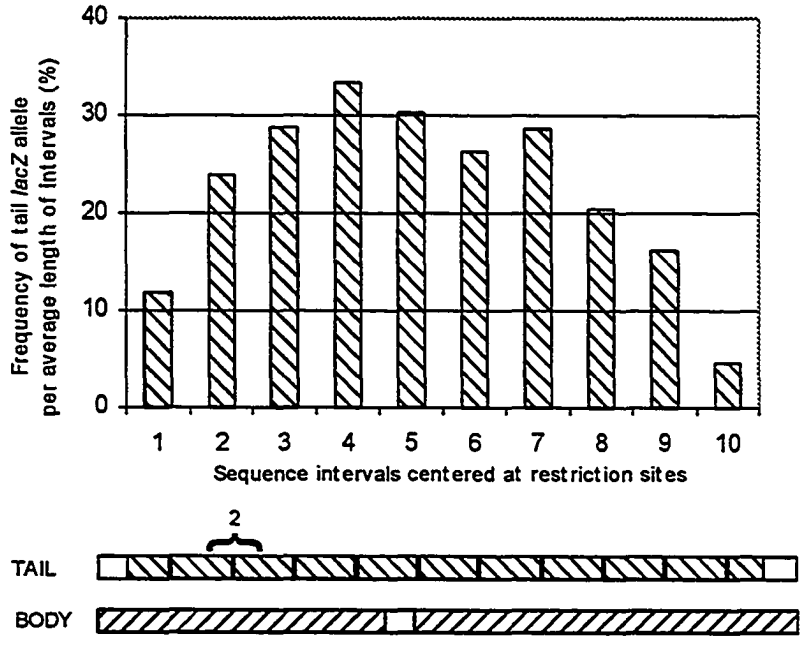


Figure 4-4. Frequency of tail *lacZ* allele at different regions of *lacZ* gene in the characterized proviral clones. The tail *lacZ* allele frequency is represented as the frequency of cutting by the 10 restriction enzymes (see text for methods of calculation). The positions of the 10 sequence intervals relative to the body and tail *lacZ* alleles are exemplified by the interval 2. The deleted regions in the tail and body *lacZ* alleles are shown as open boxes. The short vertical lines in the tail *lacZ* allele indicate the locations of individual diagnostic restriction sites. .

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

varied in length, the above values were then divided by the sequence size of an individual region to get the probability of a template switch per 100 bases of sequence in each region. Among the 11 regions, the probability of reverse transcriptase performing a template switch ranged from 1.8-11.3%, 4.8% on average (Figure 4-5). Most of the regions had a template switch probability close to the average value, with exceptions at three regions. The probabilities for regions 5 (2.4%) and 11 (1.8%) were two-fold below the average. The low value for region 5 was likely due to the disrupted nature of this region, which hosted the 117-base deletion in the body *lacZ* allele. Region 11, located at the 3' end of the *lacZ* gene, lent less chance for reverse transcriptase to switch its template. Interestingly, the template switch probability on region 9 was 2.4-fold higher than the average value, suggesting that it was a template switch hot spot.

Rate of HIV-1 readthrough transcription. The above results unequivocally showed that recombination between the genetic elements from the vector body and those located downstream of the 3' LTR could occur quite efficiently if significant homology existed between them. In the case of retroviral transduction of cellular sequences, non-homologous recombination is thought to be instrumental in bringing the downstream cellular sequence into the vector body of the progeny viruses (63, 73). Nevertheless, the great homology in our vector enabled us to address questions such as the frequency of HIV-1 readthrough transcription and HIV-1's transduction potential. To look at HIV-1 readthrough transcription, we made another vector which was similar to the above vector, but with a wild-type polyadenylation site in the 3' LTR (Figure 4-6A, HIV L^cPuro*-äc*). Viral RNA transcripts from HIV L^cPuro*-äc* would normally polyadenylate at the end of the 3' R region, forming a wild-type viral RNA transcript. Readthrough transcripts could be formed if the cellular transcription machinery bypassed the viral polyadenylation signal and instead polyadenylated at the downstream SV40 polyadynelation signal. Since the readthrough transcript would have a packaging signal identical to a wild-type viral RNA transcript, it was conceivable that some of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

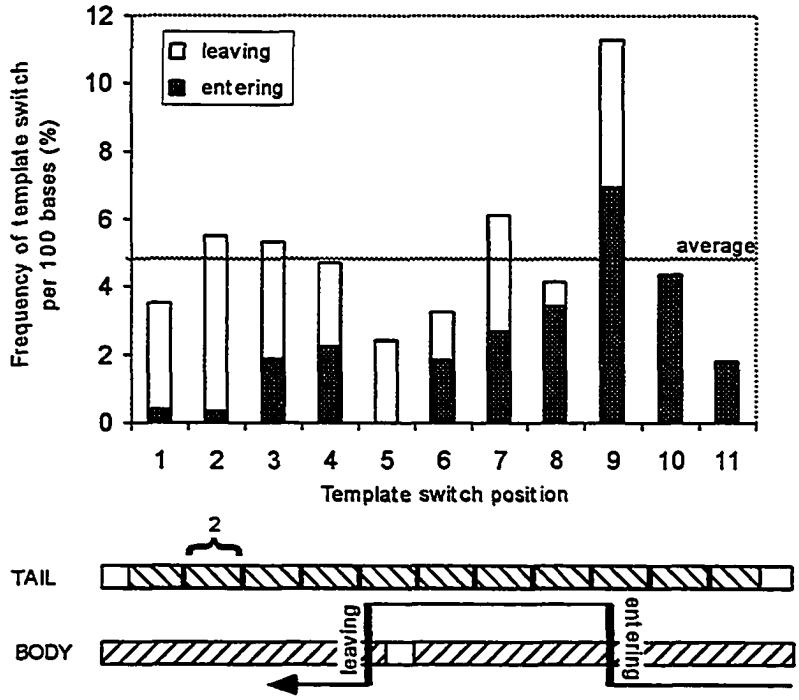


Figure 4-5. Distribution of template switch sites across the *lacZ* sequence in the characterized proviral clones. Template switch sites are categorized into either entering or leaving points in relation to which *lacZ* allele was used as template switch acceptor (see text for details). The positions of the 11 sequence intervals relative to the body and tail *lacZ* alleles are exemplified by interval 2. The deleted regions in the tail and body *lacZ* alleles are shown as open boxes. The short vertical lines in the tail *lacZ* allele indicate the locations of individual diagnostic restriction sites. The average frequency of template switch per 100 bases is indicated.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

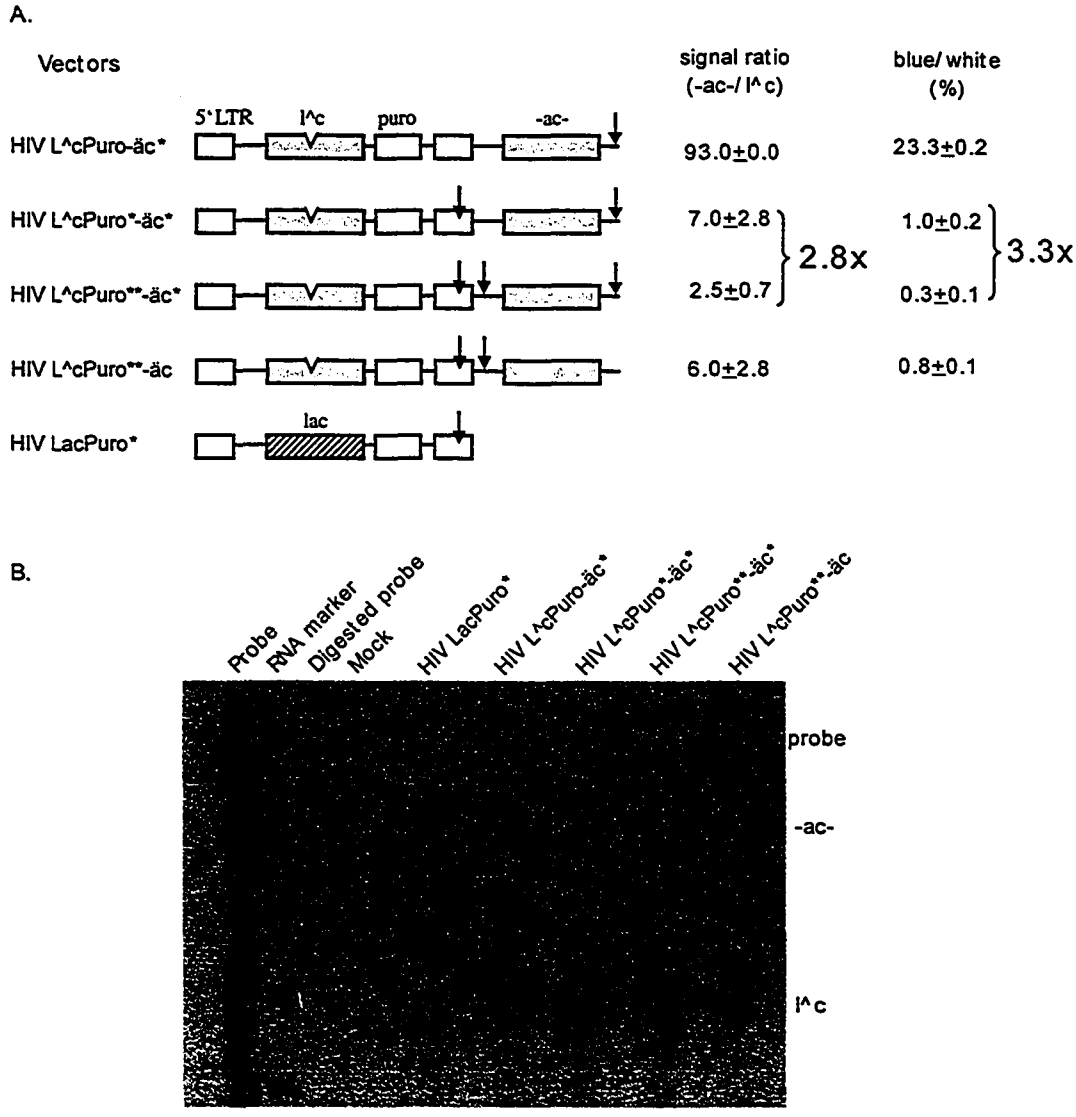


Figure 4-6. HIV-1 polyadenylation signal readthrough rate and readthrough RNA-related recombination. **(A)** Proviral structure, *lacZ* allele signal ratio from RNase protection assay, and blue/white ratio for vectors used to determine the HIV-1 readthrough transcription rate. Listed from the top downward were the constitutive readthrough vector (HLV L^cPuro-ac*), natural readthrough vector (HIV L^cPuro*-ac*), double polyA control vector (HIV L^cPuro**-ac*), double polyA vector without a polyA signal at its 3' end (HIV L^cPuro**-ac), and wild-type *lacZ* vector (HIV LacPuro*). Downward arrows indicated the position of polyA signals within the vector. The RPA signal ratio of tail *lacZ* allele to body *lacZ* allele and the recombination rate were listed to the right of the corresponding vectors. **(B)** RNase protection assay with viral RNAs produced using vectors shown in panel A. Signals corresponding to the protected body and tail *lacZ* alleles were marked with 'l^c' and '-lac-', respectively. Mock, culture supernatant from cells transfected with helper plasmid only.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

readthrough transcripts would be copackaged with wild-type transcripts into virions. Upon infection, recombination between the vector body and vector tail could generate blue colonies after selection and staining. We set to examine the viral RNA contents and recombination rates as a way to investigate the readthrough transcription rates and HIV-1 transduction potential.

Viral RNA contents were examined by RNase protection assay (Figure 4-6B). Several vectors were used as controls (Figure 4-6A). One control vector was HIV L^cPuro**-äc* (Figure 4-6A), which contained a second copy of the SV40 polyadenylation signal immediately downstream of 3' LTR. HIV L^cPuro**-äc was a similar control but with the far-downsteam SV40 polyA signal removed (Figure 4-6A). With this vector, a readthrough transcript could be made only if the RNA polymerase II was able to locate a cryptic polyadenylation signal in the vector backbone or to use the viral polyadenylation signal in the 5' LTR. HIV LacPuro* served a size marker for the recombinant signal (Figure 4-6A). The riboprobe was complimentary to the region of *lacZ* that contained the 117-bp deletion from the body allele (42). Therefore, a longer protected band signified the presence of the tail allele and the short band that of the body allele (Figure 4-6B). To compare values between vectors, a relative value for the long signal was derived while the short signal was set to 100 for each vector (Figure 4-6A). From duplicate experiments, the readthrough signal was about 2.8-fold higher than the background control vector HIV L^cPuro**-lac*, suggesting that the majority of the long signal resulted from readthrough transcription (Figure 4-6A). For the HIV L^cPuro*-lac* vector, the frequency of readthrough transcription was about 7%.

Transduction-type recombination frequency is consistent with the level of readthrough transcription. Although a relatively low but detectable level of readthrough transcription was observed for HIV-1 using the vector HIV L^cPuro*-lac*, it was not known whether these readthrough transcripts could make their way into virions and eventually complete the cycle of transduction by patch pair recombination. To test

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

this, we infected fresh target cells with viruses produced from ET cells cotransfected with individual transducing vector and the helper plasmid, and performed the recombination assay as previously described. The frequency of blue colony formation for HIV L^cPuro*-lac* was 1.0%, a value that was about 3.3-fold higher than that for HIV L^cPuro**-lac* vector viruses (Figure 4-6A). The observed recombination for the latter was presumably due to residual readthrough transcription of both HIV-1 and SV40 polyadenylation signals rather than DNA-level recombination during transfection, as the level of blue colony formation was much lower (< 0.1%) in a two-vector recombination assay when the recombination donor sequence was placed on a non-retroviral expression vector (42). Thus, the observation that blue colony formation was more frequent for HIV L^cPuro*-lac* than that for HIV L^cPuro**-lac* corroborated our finding on HIV-1 readthrough transcription.

The 'cellular RNA' from readthrough transcription is highly accessible to recombination machinery. The fairly high frequency of transduction-type recombination observed above suggested that the cellular RNA incorporated into virions as a chimeric RNA could efficiently serve as template switch acceptor. Since this "extra" RNA appended downstream to the viral RNA was usually not reverse transcribed, one relevant question was whether the status of not being reverse transcribed confers any advantage for the cellular RNA to serve as template switch acceptor compared to the viral sequence in the vector body. To address this question, we designed the vector HIV ac-ac*, which contained two copies of the internal portion of *lacZ* gene (Figure 4-7A). One copy (the body allele) was located between two LTRs, and the other (the tail allele) was placed downstream to the 3' LTR. They were identical except that a synonymous point mutation was introduced into the region corresponding to the 117-bp deletion in HIV L^cPuro*. They shared 974-bp homology to the *lacZ* sequence in HIV L^cPuro* upstream to its internal deletion, and 711-bp homology downstream to its internal deletion. Therefore, in a two-vector recombination assay using HIV L^cPuro* and HIV

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

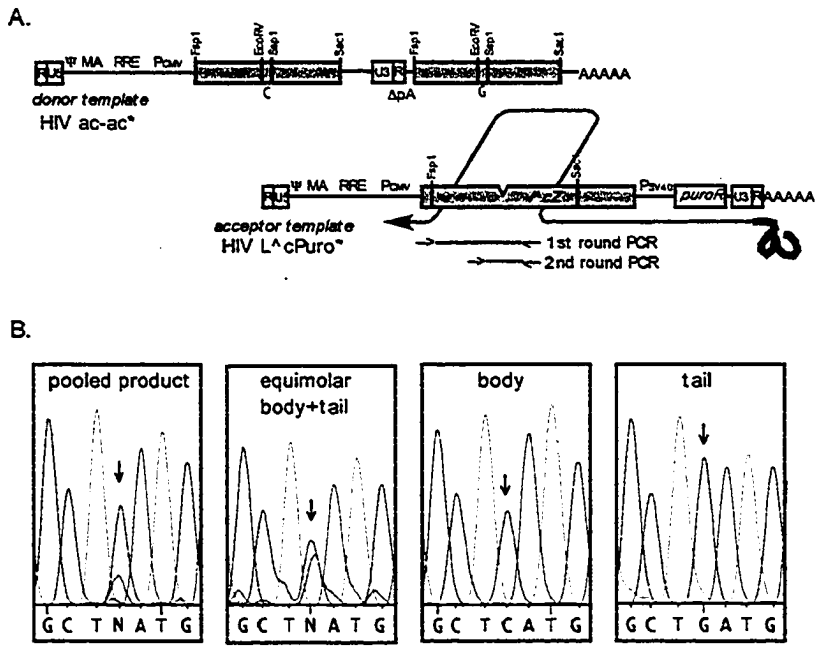


Figure 4-7. The 3' appended RNA from readthrough transcription as preferred recombinogenic template switch donor. **(A)** Schematic representation of how template switching during reverse transcription could yield a puromycin-resistant vector with functional *lacZ*. Note that the donor RNA template contains two copies of the internal portion of the *lacZ* gene both of which could potentially serve as patch repair donor. Only the patch repair event with the tail *lacZ* allele is shown. The approximate location of the 1st and 2nd round PCRs for amplifying pooled proviral DNAs is shown beneath the acceptor template. **(B)** Chromographs from sequencing of PCR products from pooled genomic DNA (pooled product), from plasmid containing the body *lacZ* allele only (body), from plamid containing the tail *lacZ* allele only (tail), and from HIV ac-ac* plasmid DNA (equimolar body + tail). Arrows indicate the nucleotide position that differed between the body and tail *lacZ* alleles.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ac-ac*, either *lacZ* allele could theoretically serve as the patch repair donor in forming a functional *lacZ* gene (Figure 4-7A).

Virus was produced by cotransfecting HIV L^cPuro*, HIV ac-ac* and helper plasmid into producer cells, and used to infect target cells. Genomic DNA was extracted from over 3000 pooled puromycin-resistant cell clones, amplified first by PCR with a pair of primers specific to proviruses generated through two crossover events (1st round PCR, illustrated in Figure 4-7A), and then reamplified with a second pair of nested primers annealing to the inner region of the 1st round PCR products. The second pair of primers was identical to those used for PCR2 shown in Figure 4-1B. Two specific bands were expected from the 2nd round PCR if a significant portion of the pooled proviral DNA was from recombinant proviral DNA: a 700-bp band templated by non-recombinant proviruses and an 817-bp band templated by recombinant proviruses. Since only 1-3% of the puromycin-resistant colonies were recombinant as indicated by X-gal staining in parallel infection experiments, the 817-bp band was very weak. Thus, it was then purified from the gel and reamplified by PCR (not shown). Purified final PCR products were used for automatic sequencing to estimate the relative contribution of the two *lacZ* alleles as patch repair donor templates (Figure 4-7B). Although not an absolute value, the signal representing the tail *lacZ* allele (nucleotide G) was significantly greater than that representing the body *lacZ* allele (nucleotide C), suggesting that the tail *lacZ* allele was preferentially used in recombinogenic template switch over the body *lacZ* allele (Figure 4-7B).

DISCUSSION

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We have established an experimental system to examine transduction-related properties by using HIV-1 vectors reminiscent of transduction intermediates. Extensive homology was present between recombination donor and acceptor, thus high rates of recombination were observed. This system allowed us to examine the individual steps involved in retroviral transduction.

According to the prevailing model for retroviral transduction of cellular genes, cellular RNA is joined 3' to the viral RNA as a result of readthrough transcription and is subsequently encapsidated into virions (23, 63, 73). This suggests that the frequency of readthrough transcription could be one of the rate-limiting steps for retroviral transduction. Using our system, we determined that the frequency of readthrough transcription is as high as 7% for HIV-1. This value was lower than the reported value for ALV (22) but was within the range of those measured by studies utilizing one or two HIV-1 LTRs (4, 7, 8, 10, 13, 14, 31, 57). The relative contribution of the different readthrough transcription rates to transduction-type recombination among different retroviral species is unknown. However, it could be tested if similar assay systems as we presented in this study were developed for other viral species. Due to the existence of extensive homology between the 'viral sequence' (the body *lacZ* allele) and the 'cellular sequence' (the tail *lacZ* allele), viral transduction could be readily detected in such systems. In fact, for HIV-1, this mode of transduction occurred during up to 1% of the viral replication in our system (Figure 4-6A), or approximately 3-fold higher than the baseline level of recombination representing the maximum that could be attributed to DNA-level recombination (this study and (42)).

Restriction analysis verified the transduction-type recombination, i.e. recombination between 'viral' sequence in the vector body and the 'cellular' sequence at the vector tail. In contrast to nonhomologous recombination that has been postulated to be responsible for the retroviral transduction of cellular genes, homologous recombination was responsible for the generation of most of the recombinant proviruses

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in our assay due to the existence of extensive sequence homology between the body *lacZ* allele and the tail *lacZ* allele located downstream to the mutated 3' LTR. The transduction-type recombination was extraordinarily efficient in our assay, as was manifested by two observations from our study. First, restriction analysis indicated that the normalized transduction-type recombination rate was 56% in the proviral clones examined. In terms of the phenotype, all blue proviral clones underwent two or more template switches that were necessary to reconstitute a functional *lacZ* gene, and 40% of the white clones were recombinants. Secondly, about 10% of the proviral clones displayed more than two template switch events within the *lacZ* region examined. Two clones even had five template switch events. The high recombination rates obtained with the constitutive readthrough vectors made it possible for us to study the transduction-type recombination with the natural readthrough vectors. In addition, the high level of functional recombinants generated from our system suggests that it can potentially be utilized for gene shuffling (21). For example, one could replace the body and tail *lacZ* sequences with two naturally existing variants of a gene of interest in which function can be tested experimentally. Functional recombinants can be experimentally identified and compared. Therefore, this system could serve a useful tool for generating chimeric gene products from natural variants of desired genes that share a certain level of sequence identity.

Interestingly, in addition to the homologous recombination that accounted for the majority of template switches, nonhomologous recombination was found in the formation of seven proviral clones. These nonhomologous template switch junctions were identified by sequencing analysis, and could be classified into three groups: 3/7 utilized short stretches of sequence identity at template switch junctions (seven nucleotides in clones 6 and 62, and one nucleotide in clone 92), 2/7 lacked any sequence homology at the template switch junctions (clones 15 and 70), and 2/7 contained insertions (clones 64 and 20). They resembled the characteristic template switch junctions for nonhomologous

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

recombination observed at the 3' junctions of oncogene-virus sequences (11, 17, 20, 33, 73). In our assay, the frequency of nonhomologous template switch was 2.5% (3/118) relative to homologous template switch, a value consistent with the range of 100-1000 fold difference suggested by other studies (60, 74) (Note that the numerator of the fraction 3/118 includes one crossover in clone 64 and two crossovers in clone 20, and the denominator 118 has been normalized). Certainly, the spacing of the genetic heterogeneity between the two *lacZ* alleles in our vector limited its capability of resolving homologous template switches that might have occurred between two adjacent marker positions. However, the probability for such template switches was presumably low based on the overall rate of template switch in the 2.6-kb target *lacZ* region (1.3 crossovers per clone). Taken together, our results suggest that retroviral transduction through nonhomologous recombination between readthrough RNA and wild-type RNA may not be as improbable as it has been thought.

Several interesting features were observed from proviral clones with non-homologous template switch events. Clone 20 had a stretch of over 30 bases of A's at the junctions of an internal *lacZ* deletion. Theoretically, this polyhomomeric adenine sequence could be derived from either polymerase slippage or template switch events between viral RNA and poly(A) sequences. In the latter case, the poly(A) template could presumably originate from the poly(A) tails of the viral genomic RNA, readthrough RNA, or a copackaged cellular mRNA. Interestingly, the incorporation of short stretches of A's has been reported at the 3' junctions of some v-*onc* genes from naturally occurring highly oncogenic viruses (26) and in some recombinants from several other experimental systems (50, 52, 63, 74). Speculatively, the formation of this junction would involve two non-homologous template switch events, one from the body *lacZ* allele switching into poly(A) sequence and the other from the poly(A) sequence into the tail *lacZ* allele. Additionally, there were three homologous template switch events during the synthesis of the *lacZ* sequence in clone 20. Although clone 20 was the sole example of such from our

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

study, the above observation suggests that reverse transcriptase has the potential to utilize all sorts of available genetic information during its synthesis of one viral DNA.

In sharp contrast to the large number of recombinant clones that acquired an intact 5' end of *lacZ* from the body *lacZ* allele, four proviral clones maintained the 5' end of the tail *lacZ* allele (clone 6, 62, 70, and 92). The length of the 5' flanking sequence to the tail *lacZ* allele in these clones varied. Clone 92 copied only 22 of the 28 bases between the shortened R and the tail *lacZ* allele, and then switched to the CMV promoter region. The subsequent deletion was 0.7 kb compared to typical white clones. Clone 6 and 62, which had identical template switch junctions at the deletion but differed at other *lacZ* regions, copied 9 nucleotides of the 3' end of shortened R in addition to the 28 bases between the shortened R and the tail *lacZ* allele and then switched to the MA coding sequence at the 5' portion of the genome, resulting in 1.9-kb of deletion compared to typical white clones. Clone 70 was unique among all proviral clones in that, in addition to the 28 bases between the shortened RNA and the tail *lacZ* allele, the original mutated viral 3' LTR was present between the two terminal LTRs, thus constituting a three-LTR provirus. Restriction and sequencing mapping results suggests the following scenario in its formation: it initially involved a template switch of the growing point of minus strand DNA from the body *lacZ* allele of one genomic RNA template to the tail *lacZ* allele on the copackaged RNA. Then the minus strand DNA synthesis proceeded through the partially deleted 3' LTR, puromycin resistance expression cassette, and up to the 3' end of the body *lacZ* allele before it switched templates to the 5' portion of the genomic RNA. Consequently, as confirmed by PCR and sequencing analysis, this provirus acquired three LTRs, two complete copies of the puromycin resistance expression cassette, and a chimeric *lacZ* allele. Interestingly, evidence of a 3-LTR structure has also been described among products of an ASV-based transduction assay vector (Doshi and Coffin, personal communication). As a side note, clone 70 may not be the only proviral clone that contained a 3-LTR structure. Another clone, clone 63 (not shown), had both

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the 700-bp and 817-bp products in PCR2 reaction, and displayed partial digestion by seven restriction enzymes. Such a genotype was consistent with its phenotype—mixed white and blue cells after X-gal staining, suggesting that it may contain a similar 3-LTR structure. However, as described before, no internal LTR structures were found in three other proviral clones which had copied through the 5' end of the tail *lacZ* allele (clone 6, 62, and 92). The frequent occurrence of various lengths of sequence deletion among our characterized proviral clones was not unusual, as high rates of genetic rearrangement have been reported for several simple retroviruses (43, 49, 68) including occasional transduction of cellular RNAs between deletion junctions (49, 68).

The structure of these 4 proviral clones shed new light on the mechanism of the formation of 5' viral/c-*onc* junction. At least three different pathways have been proposed for the formation of the 5' junction: (i) it has been suggested that these 5' junctions might be created by DNA level deletion, based on the observation of intron sequences from c-*onc* genes at the 5' viral/c-*onc* junctions of some oncogenic viruses (30, 32, 50, 55, 65); (ii) RNA splicing across the viral and cellular sequences may serve as an alternative pathway for viral/c-*onc* fusion. This was supported by the observation that the 5' recombination point coincides with an RNA splice junction (27, 38); and (iii) the ability of reverse transcriptase to perform template switch independent of sequence similarity at the crossover sites suggests the formation of the 5' junction can be achieved by a nonhomologous template switch by reverse transcriptase as has been implicated in the generation of the 3' junctions (63, 74). The proviral structures of clones 6, 62, and 92 provided further evidence supporting the third model. Our transduction-type recombination assay was initially designed to compare the intermolecular recombination rates between MLV and HIV-1 independent of the assumption of random copackaging of the coexpressed two viral RNAs (42). It was reasoned, and has been verified by this study, that a function *lacZ* gene could be generated if two template switches occurred between the disrupted *lacZ* gene in the vector's body and the tail *lacZ* allele. The tail *lacZ*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

allele at the 3' end of either the same or the copackaged RNA can serve as the acceptor sequence for the first template switch. In the case of an 'intramolecular' template switch, the 3' LTR of this genomic RNA presumably has been degraded by RNaseH activity and the cellular RNA downstream to the 3'LTR has been severed from the remaining 5' sequences. If the growing point of minus strand DNA switches to this 3' RNA originally from the same RNA molecule, it will eventually be forced to switch templates again when the DNA synthesis proceeds to the 5' end of the new template. This second template switch to the 5' part of the viral RNA may rescue the viral DNA synthesis and generate a provirus with an acquired cellular gene, as is illustrated by clone 6, 62, and 92. In contrast, the formation of clone 70 likely involved a first template switch from the vector body to the tail *lacZ* of the copackaged RNA, the 3' LTR of which had not been degraded and was later copied into provirus.

The characterization of proviral clones from our transduction-type recombination assay also provides us a unique opportunity to reevaluate the so-called "high negative interference" model for retroviral recombination. This model proposes that the performance of one recombination event predisposes the ordinarily recombination-proof virion to performing additional switches, or alternatively that only a subset of virions are prone to generating recombinant products (3, 12, 25). In contrast to the two-vector recombination assays, the transduction-type recombination assay is a one-vector approach for studying genetic recombination (42). Viral genomes consist of two homozygous genomic RNAs containing both recombination donor and acceptor sequences, and thus are capable of performing genetic recombination. Since in our assay, even numbers of template switches were more or less selected over odd numbers, the comparison below focused on template switches of even numbers. In general, approximately 43% of the proviral clones displayed two template switches over the inspected *lacZ* gene. Assuming template switches occur independently, the expected number of proviral clones that had four template switches was 17%. In fact, only about

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

10% of proviral clones had more than two template switches. Therefore, we found no evidence for high negative interference in our transduction-type recombination experiments. Besides the difference in experimental designs, further investigation is warranted to reconcile the discrepancy between our results and pervious findings.

It has been suggested that reverse transcriptase template switches do not occur at uniform frequencies and are highly context-dependent (2, 71). For example, sequences that potentially form secondary structures have been shown to be hot spots for both mutation and recombination (28, 29, 44). Several lines of evidence from our experiments suggest the existence of local hot spots among otherwise relatively evenly distributed regions for template switch: (i) Sequence analysis between restriction site 8 and 9 showed at least a two-fold increase of template switch over other regions; (ii) Among the 7 white clones that contained deletions and were mapped by sequencing, two pairs had completely identical template switch junctions at the deletion (clone 6 and clone 62, clone 15 and clone 70, respectively). Interestingly, one pair involved micro-homology and the other pair showed no sequence homology. RNA folding by M-fold (75) did not yield any conclusive support for the role of RNA structure on the frequency of template switch in these regions.

Using our transduction-type recombination assay, we have shown that readthrough transcripts are frequently synthesized and could be efficiently used as template switch acceptors. However, besides some anecdotal evidence derived from tissue culture experiments (62), transduction of cellular genes by HIV-1 has not been reported in humans, although HIV-1 has been implicated in lymphomagenesis through insertional activation (41). In addition to the low frequency of every step leading to retroviral transduction, this rarity for HIV-1 may also be restricted by the short half-life of infected cells or by the complex regulation of HIV-1 expression. The average half-life for the HIV-1 infected cells has been estimated to be on the order of days (24, 47, 70). Such rapid turnover of infected cells lends little chance for passing transduced cellular

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

genes to the next generation. On the other hand, HIV-1 has greater complexity in its genome than oncoviruses that efficiently replicate without virally coded regulatory proteins. The replication of HIV-1 depends on certain regulatory proteins that are generated from multiply spliced viral RNAs. It has been suggested that generation of these multiply spliced RNAs and their *cis*-acting targets would be destroyed by a cellular genetic insert, essentially inactivating the virus (40). In this regard, the long latency of AIDS and prolonged life span of AIDS patients due to the antiviral treatments may increase the likelihood of HIV-1 transduction. Coincidently, in a recent study, an in-frame 33-nt insertion in the reverse transcriptase coding region was found to augment HIV-1 multiple-drug resistance in an HIV-1 infected patient (54). A sequence homology search by BLAST (1) revealed that one of the clones had 96% sequence identity to a human cDNA located on chromosome 17, suggesting its cellular origin.

BIBLIOGRAPHY

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs Nucleic Acids Res. **25**:3389-402.

2. **Anderson, J. A., E. H. Bowman, and W. S. Hu** 1998. Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation J Virol. **72**:1195-202.

3. **Anderson, J. A., R. J. Teufel, 2nd, P. D. Yin, and W. S. Hu** 1998. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination J Virol. **72**:1186-94.

4. **Ashe, M. P., P. Griffin, W. James, and N. J. Proudfoot** 1995. Poly(A) site selection in the HIV-1 provirus: inhibition of promoter- proximal polyadenylation by the downstream major splice donor site Genes Dev. **9**:3008-25.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5. **Besmer, P., J. E. Murphy, P. C. George, F. H. Qiu, P. J. Bergold, L. Lederman, H. W. Snyder, Jr., D. Brodeur, E. E. Zuckerman, and W. D. Hardy** 1986. A new acute transforming feline retrovirus and relationship of its oncogene v-kit with the protein kinase gene family Nature. **320:**415-21.

6. **Bishop, J. M.** 1983. Cellular oncogenes and retroviruses Annu Rev Biochem. **52:**301-54.

7. **Bohnlein, S., J. Hauber, and B. R. Cullen** 1989. Identification of a U5-specific sequence required for efficient polyadenylation within the human immunodeficiency virus long terminal repeat J Virol. **63:**421-4.

8. **Brown, P. H., L. S. Tiley, and B. R. Cullen** 1991. Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences J Virol. **65:**3340-3.

9. **Cepko, C., and W. Pear** 2003. Transduction of genes using retrovirus vectors. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (eds), Current protocols in molecular biology online. John Wiley & Sons, Inc.

10. **Cherrington, J., and D. Ganem** 1992. Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences Embo J. **11:**1513-24.

11. **Chesters, P. M., K. Howes, J. C. McKay, L. N. Payne, and K. Venugopal** 2001. Acutely transforming avian leukosis virus subgroup J strain 966: defective genome encodes a 72-kilodalton Gag-Myc fusion protein J Virol. **75:**4219-25.

12. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses J Gen Virol. **42:**1-26.

13. **DeZazzo, J. D., J. E. Kilpatrick, and M. J. Imperiale** 1991. Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation Mol Cell Biol. **11:**1624-30.

14. **DeZazzo, J. D., J. M. Scott, and M. J. Imperiale** 1992. Relative roles of signals upstream of AAUAAA and promoter proximity in regulation of human immunodeficiency virus type 1 mRNA 3' end formation Mol Cell Biol. **12:**5555-62.

15. **Dron, M., L. Hameau, L. Benboudjema, J. Guymarho, C. Cajean-Feroldi, P. Rizza, C. Godard, C. Jasmin, M. G. Tovey, and M. C. Lang** 1999. Cloning of a long HIV-1 readthrough transcript and detection of an increased level of early growth response protein-1 (Egr-1) mRNA in chronically infected U937 cells Arch Virol. **144:**19-28.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

16. **Furger, A., J. Monks, and N. J. Proudfoot** 2001. The retroviruses human immunodeficiency virus type 1 and Moloney murine leukemia virus adopt radically different strategies to regulate promoter-proximal polyadenylation J Virol. **75:**11735-46.

17. **Girod, A., A. Drynda, F. L. Cosset, G. Verdier, and C. Ronfort** 1996. Homologous and nonhomologous retroviral recombinations are both involved in the transfer by infectious particles of defective avian leukosis virus-derived transcomplementing genomes J Virol. **70:**5651-7.

18. **Goldfarb, M. P., and R. A. Weinberg** 1981. Generation of novel, biologically active Harvey sarcoma viruses via apparent illegitimate recombination J Virol. **38:**136-50.

19. **Goodrich, D. W., and P. H. Duesberg** 1990. Evidence that retroviral transduction is mediated by DNA not by RNA Proc Natl Acad Sci U S A. **87:**3604-8.

20. **Hajjar, A. M., and M. L. Linial** 1993. A model system for nonhomologous recombination between retroviral and cellular RNA J Virol. **67:**3845-53.

21. **Harayama, S.** 1998. Artificial evolution by DNA shuffling Trends Biotechnol. **16:**76-82.

22. **Herman, S. A., and J. M. Coffin** 1986. Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA J Virol. **60:**497-505.

23. **Herman, S. A., and J. M. Coffin** 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction Science. **236:**845-8.

24. **Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz** 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection Nature. **373:**123-6.

25. **Hu, W. S., E. H. Bowman, K. A. Delviks, and V. K. Pathak** 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference J Virol. **71:**6028-36.

26. **Huang, C. C., N. Hay, and J. M. Bishop** 1986. The role of RNA molecules in transduction of the proto-oncogene c-fps Cell. **44:**935-40.

27. **Ikawa, S., K. Hagino-Yamagishi, S. Kawai, T. Yamamoto, and K. Toyoshima** 1986. Activation of the cellular src gene by transducing retrovirus Mol Cell Biol. **6:**2420-8.

28. **Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome J Virol. **74:**1234-40.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

29. **Jones, J. S., R. W. Allan, and H. M. Temin** 1994. One retroviral RNA is sufficient for synthesis of viral DNA J Virol. **68:**207-16.

30. **Kan, N. C., C. S. Flordellis, G. E. Mark, P. H. Duesberg, and T. S. Papas** 1984. Nucleotide sequence of avian carcinoma virus MH2: two potential onc genes, one related to avian virus MC29 and the other related to murine sarcoma virus 3611 Proc Natl Acad Sci U S A. **81:**3000-4.

31. **Klasens, B. I., A. T. Das, and B. Berkhout** 1998. Inhibition of polyadenylation by stable RNA secondary structure Nucleic Acids Res. **26:**1870-6.

32. **Klempnauer, K. H., T. J. Gonda, and J. M. Bishop** 1982. Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene Cell. **31:**453-63.

33. **Lee, S. Y., T. M. Howard, and S. Rasheed** 1998. Genetic analysis of the rat leukemia virus: influence of viral sequences in transduction of the c-ras proto-oncogene and expression of its transforming activity J Virol. **72:**9906-17.

34. **Lobato, R. L., E. Y. Kim, R. M. Kagan, and T. C. Merigan** 2002. Genotypic and phenotypic analysis of a novel 15-base insertion occurring between codons 69 and 70 of HIV type 1 reverse transcriptase AIDS Res Hum Retroviruses. **18:**733-6.

35. **Malhotra, S., A. G. Scott, T. Zavorotinskaya, and L. M. Albritton** 1996. Analysis of the murine ecotropic leukemia virus receptor reveals a common biochemical determinant on diverse cell surface receptors that is essential to retrovirus entry J Virol. **70:**321-6.

36. **Mansky, L. M., and H. M. Temin** 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:**5087-94.

37. **Mansky, L. M., and H. M. Temin** 1994. Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus J Virol. **68:**494-9.

38. **Martin, P., C. Henry, F. Ferre, C. Bechade, A. Begue, C. Calothy, B. Debuire, D. Stehelin, and S. Saule** 1986. Characterization of a myc-containing retrovirus generated by propagation of an MH2 viral subgenomic RNA J Virol. **57:**1191-4.

39. **Naldini, L., U. Blomer, F. H. Gage, D. Trono, and I. M. Verma** 1996. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector Proc Natl Acad Sci U S A. **93:**11382-8.

40. **Nevins, J. R., and P. K. Vogt** 1996. Cell transformation by viruses, p. 267-310. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (eds), Fundamental Virology, Third ed. Lippincott-Raven Publishers, Philadelphia, PA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

41. **Ng, V. L., and M. S. McGrath** 1998. The immunology of AIDS-associated lymphomas Immunol Rev. **162:**293-8.

42. **Onafuwa, A., W. An, N. D. Robson, and A. Telesnitsky** 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. 77:4577-87.

43. **Parthasarathi, S., A. Varela-Echavarria, Y. Ron, B. D. Preston, and J. P. Dougherty** 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications J Virol. **69:**7991-8000.

44. **Pathak, V. K., and H. M. Temin** 1992. 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate J Virol. **66:**3093-100.

45. **Pathak, V. K., and H. M. Temin** 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:**6024-8.

46. **Pathak, V. K., and H. M. Temin** 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:**6019-23.

47. **Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time Science. **271:**1582-6.

48. **Pfeiffer, J. K., R. S. Topping, N. H. Shin, and A. Telesnitsky** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription J Virol. **73:**8441-7.

49. **Pulsinelli, G. A., and H. M. Temin** 1991. Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer J Virol. **65:**4786-97.

50. **Raines, M. A., N. J. Maihle, C. Moscovici, L. Crittenden, and H. J. Kung** 1988. Mechanism of c-erbB transduction: newly released transducing viruses retain poly(A) tracts of erbB transcripts and encode C-terminally intact erbB proteins J Virol. **62:**2437-43.

51. **Reddy, E. P., R. K. Reynolds, E. Santos, and M. Barbacid** 1982. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene Nature. **300:**149-52.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

52. **Robson, N. D., and A. Telesnitsky** 1999. Effects of 3' untranslated region mutations on plus-strand priming during moloney murine leukemia virus replication J Virol. **73:**948-57.

53. **Roebroek, A. J., J. A. Schalken, C. Onnekink, H. P. Bloemers, and W. J. Van de Ven** 1987. Structure of the feline c-fes/fps proto-oncogene: genesis of a retroviral oncogene J Virol. **61:**2009-16.

54. **Sato, H., Y. Tomita, K. Ebisawa, A. Hachiya, K. Shibamura, T. Shiino, R. Yang, M. Tatsumi, K. Gushi, H. Umeyama, S. Oka, Y. Takebe, and Y. Nagai** 2001. Augmentation of human immunodeficiency virus type 1 subtype E (CRF01_AE) multiple-drug resistance by insertion of a foreign 11-amino- acid fragment into the reverse transcriptase J Virol. **75:**5604-13.

55. **Schwartz, D. E., R. Tizard, and W. Gilbert** 1983. Nucleotide sequence of Rous sarcoma virus Cell. **32:**853-69.

56. **Schwartz, J. R., S. Duesberg, and P. H. Duesberg** 1995. DNA recombination is sufficient for retroviral transduction Proc Natl Acad Sci U S A. **92:**2460-4.

57. **Scott, J. M., and M. J. Imperiale** 1996. Reciprocal effects of splicing and polyadenylation on human immunodeficiency virus type 1 pre-mRNA processing Virology. **224:**498-509.

58. **Shin, N. H., D. Hartigan-O'Connor, J. K. Pfeiffer, and A. Telesnitsky** 2000. Replication of lengthened Moloney murine leukemia virus genomes is impaired at multiple stages J Virol. **74:**2694-702.

59. **Stehelin, D., H. E. Varmus, J. M. Bishop, and P. K. Vogt** 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA Nature. **260:**170-3.

60. **Stuhlmann, H., M. Dieckmann, and P. Berg** 1990. Transduction of cellular neo mRNA by retrovirus-mediated recombination J Virol. **64:**5783-96.

61. **Sugden, B.** 1993. How some retroviruses got their oncogenes Trends Biochem Sci. **18:**233-5.

62. **Sun, G., P. K. O'Neil, H. Yu, Y. Ron, B. D. Preston, and J. P. Dougherty** 2001. Transduction of cellular sequence by a human immunodeficiency virus type 1-derived vector J Virol. **75:**11902-6.

63. **Swain, A., and J. M. Coffin** 1992. Mechanism of transduction by retroviruses Science. **255:**841-5.

64. **Swain, A., and J. M. Coffin** 1989. Polyadenylation at correct sites in genome RNA is not required for retrovirus replication or genome encapsidation J Virol. **63:**3301-6.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

65. **Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop** 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus Proc Natl Acad Sci U S A. **80:**2519-23.

66. **Tabin, C. J., S. M. Bradley, C. I. Bargmann, R. A. Weinberg, A. G. Papageorge, E. M. Scolnick, R. Dhar, D. R. Lowy, and E. H. Chang** 1982. Mechanism of activation of a human oncogene Nature. **300:**143-9.

67. **Telesnitsky, A., and S. P. Goff** 1997. Reverse transcriptase and the generation of retroviral DNA, p. 121-160. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

68. **Varela-Echavarria, A., C. M. Prorock, Y. Ron, and J. P. Dougherty** 1993. High rate of genetic rearrangement during replication of a Moloney murine leukemia virus-based vector J Virol. **67:**6357-64.

69. **Varmus, H. E.** 1984. The molecular genetics of cellular oncogenes Annu Rev Genet. **18:**553-612.

70. **Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, and et al.** 1995. Viral dynamics in human immunodeficiency virus type 1 infection Nature. **373:**117-22.

71. **Wooley, D. P., L. A. Bircher, and R. A. Smith** 1998. Retroviral recombination is nonrandom and sequence dependent Virology. **243:**229-34.

72. **Zaiss, A. K., S. Son, and L. J. Chang** 2002. RNA 3' readthrough of oncoretrovirus and lentivirus: implications for vector safety and efficacy J Virol. **76:**7209-19.

73. **Zhang, J., and H. M. Temin** 1993. 3' junctions of oncogene-virus sequences and the mechanisms for formation of highly oncogenic retroviruses J Virol. **67:**1747-51.

74. **Zhang, J., and H. M. Temin** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication Science. **259:**234-8.

75. **Zuker, M., D. H. Mathews, and D. H. Turner** 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a pratical guide, p. 11-43. *In* J. Barciszewski, and B. F. C. Clark (eds), RNA Biochemistry and Biotechnology. Kluwer Academic Publishers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER FIVE

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

In this thesis, a series of single replication cycle assays were performed for both HIV-1 and MLV using highly similar vectors and assay conditions. Forward mutation rates, repeat deletion rates, and intermolecular template switching/recombination frequencies between HIV-1 and MLV were compared. The transduction potential of HIV-1 was investigated with transduction-type recombination assays using vectors reminiscent of retroviral transduction intermediates. The findings from these experiments were as follows:

(i) No significant difference was observed for the forward mutation rates between HIV-1 and MLV, suggesting that HIV-1 RT is not necessarily more prone to replication errors than MLV RT.

(ii) The frequency of direct repeat deletion and the intermolecular recombination rate as measured by transduction-type vectors were similar for HIV-1 and MLV, indicating the template switching properties of HIV RT resembles those of MLV RT.

(iii) The frequency of template switching as measured by repeat deletion decreased when varied levels of sequence similarity were introduced into the repeated sequences. More dramatic effects on the reduction of repeat deletion rates were observed with vectors containing higher degree of sequence heterogeneity between the repeats.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(iv) The HIV-1 intermolecular recombination was six- to seven-fold higher than MLV as determined by two-vector assays, in contrast to the one-vector assays (ii), in which the detection of recombinants relied on RNA copackaging,

(v) The combined data from (ii) and (iv) provide genetic evidence that coexpressed RNAs are less accessible to the recombination machinery during MLV replication than for HIV-1.

(vii) Readthrough of HIV-1 polyadenylation signals located at the 3' R occurred at a rate up to 7% of the normal viral transcription.

(vi) RNA sequences downstream of the 3' LTR acquired by readthrough transcription could be incorporated into the proviral DNA through nonhomologous recombination. However, such incorporation through homologous recombination was much more efficient.

(vii) The readthrough sequences downstream of the 3' LTR appeared to be preferentially used as recombination donor templates for HIV-1 intermolecular recombination as compared to homologous sequences located between the 5' and 3' LTRs.

Discussion and future directions

Are direct repeat deletion and recombination mechanistically related?

Direct repeat deletion has been widely used as a model system to study template switch properties of reverse transcriptase (4, 5, 12, 13, 24, 25, 27, 47, 48, 59, 63). In this thesis work, direct repeat deletion systems were employed to compare the template switch frequency between HIV-1 and MLV (Chapter two and three). One advantage of using direct repeat deletion system for studying retroviral template switching is that it constitutes a one-vector approach which circumvents the technical difficulty in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

coexpressing two viral RNAs as in two-vector assays (1). In addition, the frequency of direct repeat deletion is high, a short target sequence can be used in constructing the repeats, and the repeat deletion can be easily monitored by phenotypic assays (5, 12, 25, 27, 47, 48, 59, 63), endonuclease restriction (4, 24), Southern blot (4, 12, 13, 47), or sequencing analysis. Observations from direct repeat deletion assays have often been extrapolated to explain the mechanism for retroviral recombination. However, it was not clear whether direct repeat deletion mechanistically resembles intermolecular recombination between two RNA molecules.

One of the concerns is that, unlike intermolecular recombination between two RNA molecules, it has been argued that direct repeat deletion is mainly through an intramolecular template switch mechanism (24) (62). However, this argument is debatable after careful examination and reinterpretation of the data in light of our findings on the different recombination rates for HIV-1 and MLV. In the study by Hu and colleagues, different restriction sites flanking the direct repeat sequence were introduced into two spleen necrosis virus (SNV) direct repeat vectors, and the mode of template switch during repeat deletion could be determined by examining the pattern of flanking restriction sites in individual proviral clones (24). When single-drug-resistant clones were examined, all clones (27/27) that underwent repeat deletion had flanking restriction sites identical to either of the parental vectors. Based on this observation, the authors concluded that direct repeat deletion occurs primarily through intramolecular template switch (24). However, in the same study, 21% (4/19) of the double-drug-resistant clones that underwent repeat deletion contained flanking restriction sites different from any of the two parental vectors, suggesting that they had undergone intermolecular template switch. In fact, these vectors were so designed that single-drug-resistant clones could be derived from infection with homozygous virions containing two identical copies of either of the two direct repeat vector RNAs, or from infection with heterozygous virions containing one of each of the two vector RNAs, while the double-drug resistant clones

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

could only be derived from infection with heterozygous virions. The authors overlooked the data from double-drug-resistant clones likely because of the following two reasons: (i) the number of total single-drug-resistant clones characterized (90; among them, 27 had repeat deletion) was much higher than that of double-drug-resistant clones (53; among them, 19 had repeat deletion), and (ii) the authors assumed that half of the 90 single-drug-resistant clones were derived from infection with heterozygous virions (24). The latter assumption would have been valid if the two SNV viral RNAs expressed equally and their copackaging into virions were a random process. However, in light of our findings on the difference on recombination rates between HIV-1 and MLV, if SNV behaves similarly to MLV in terms of the process of dimerization, it is possible that the majority of the single-drug-resistant clones characterized in the above study (24) were derived from infection by homozygous virions. In this case, any intermolecular template switch would have been ignored because both intermolecular and intramolecular template switch would have parental restriction site patterns.

A second study employed a similar two-vector approach with MLV (Reference (62) Figure 5). The repeat deletion rate was 48.5% for single-drug-resistant clones and 36% for double-drug-resistant clones (Reference (62) Table 3). The intra- or intermolecular nature of the template switch that caused each deletion event was not directly examined in this study. Instead, the authors deduced the rate of intermolecular repeat deletion (0.5% = 36% * 1.3%) by multiplying the deletion rate for the double-drug-resistant clones (36%) by the intermolecular recombination rate (1.3%), the latter of which was calculated as the ratio of the double-drug-resistant titer to the sum of two single-drug-resistant titers. Since the above calculated rate of intermolecular repeat deletion (0.5%) was much less than the combined inter- and intramolecular repeat deletion rate (48.5%) from single-drug-resistant clones, the authors claimed that the recombination between two identical sequences within the same RNA molecule resulted

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

mostly from intramolecular template switch (62). However, the validity of such calculation remained to be evaluated by a direct experimental approach.

Results from our experiments with reciprocal direct repeat vectors suggest that intra- and intermolecular template switching may contribute equally to retroviral direct repeat deletion (Figure 3-6). In fact, the relative contribution of intra- and intermolecular template switching to direct repeat deletion can be directly tested with our HIV-1 direct repeat vector assay system after incorporating flanking restriction site polymorphism. Such an experiment is outlined in Figure 5-1. My prediction is that among all the proviral clones that have undergone direct repeat deletion, approximately 25% will be through intermolecular template switch. The reason for using an HIV-1 vector system is based on the observed genetic difference in recombinogenic template switching between HIV-1 and MLV, which suggests the possibility that HIV-1 selects RNAs for copackaging from among coexpressed RNAs in a more random fashion than that employed by MLV (Chapter three). With a similar assay using MLV vectors, I would predict that the frequency of direct repeat deletion resulting from intermolecular template switch would be much lower than that for HIV-1, as the deletion rate in experiments with MLV reciprocal vectors was 7-fold lower than that with HIV-1 reciprocal vectors (Figure 3-6).

Direct repeat vectors with degenerate repeat sequences can also be adapted to address a variety of other questions regarding retroviral template switch properties. For example, it is an intriguing question whether the distribution of template switch points occurs more frequently in the 5' side of the repeat (13). The sequence heterogeneity between repeated sequences can be used as markers to map the template switch positions and to check their distribution. Such experiment will provide useful information for understanding the sequence requirement in template switching. With careful design, certain sequence features can be incorporated into such vectors, such as stem-loop secondary structures, and their effects on direct repeat deletion can be tested.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

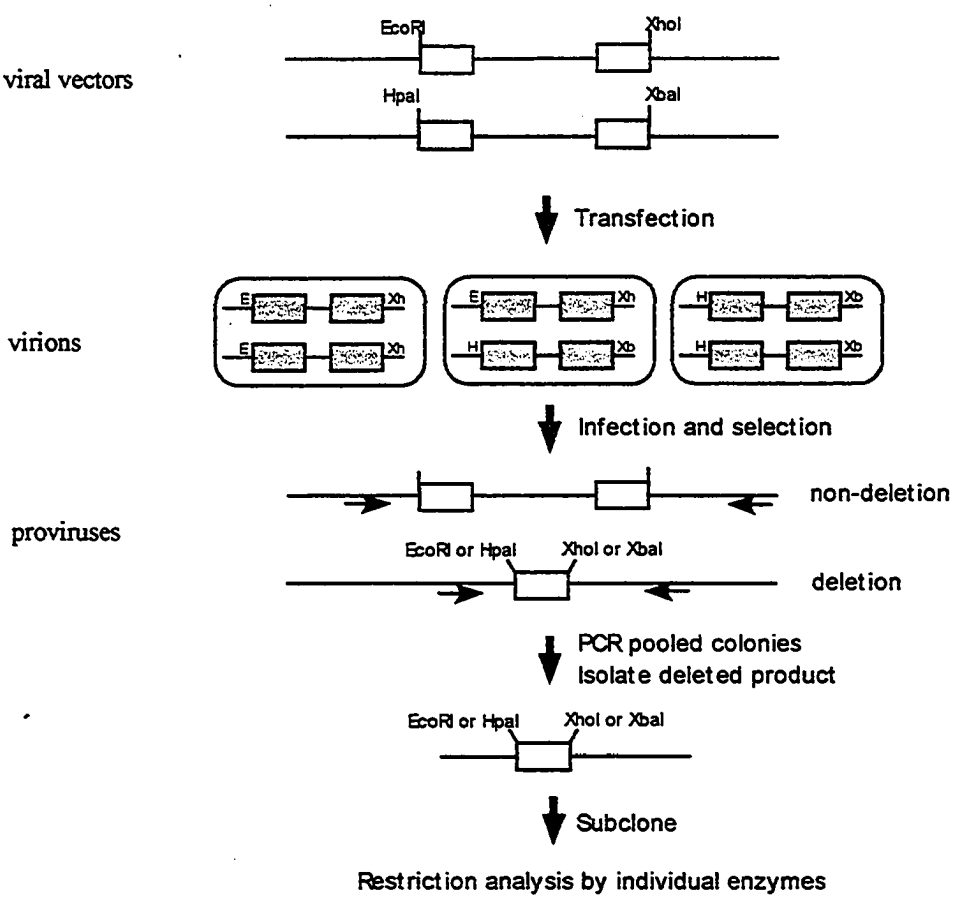


Figure 5-1. Direct repeat deletion assay to determine the relative contribution of inter- and intramolecular template switch to repeat deletion. To generate vector virus, two direct repeat vectors which are identical except the restriction sites flanking the repeated sequence are cotransfected with the helper plasmid into producer cells. The three types of virions generated are shown. Target cells are infected with the heterogeneous population of virions, and selected for puromycin resistance. Puromycin-resistant proviral clones are pooled and extracted for genomic DNA. The region that encompasses the repeated sequence is PCR amplified with sequence-specific primers (indicated by arrows). PCR product corresponding to the deleted product is isolated and subcloned into a plasmid. Individual plasmid clones are analyzed by the four restriction enzymes. The repeat deletion is considered to be an intermolecular template switch if the restriction pattern differs from both of the parental clones.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Alternative hypothesis for differences in recombination rates between HIV-1 and MLV—potential contribution by the accessory proteins and/or architecture and composition of the reverse transcription complex

The work described in this thesis provides convincing genetic evidence that coexpressed RNAs are less accessible to the recombination machinery during MLV replication than for HIV-1. The mechanism for the difference in accessibility of coexpressed RNAs has been speculative until recent studies from our laboratory on RNA dimerization. Using an RNA capture assay, it was found that coexpressed RNAs were copackaged less randomly for MLV than for HIV-1 (Flynn and Telesnitsky, manuscript in preparation). This difference in the randomness of viral RNA copackaging is likely at least partially responsible for the observed differences in recombination rates between these two retroviral species. The mechanisms that affect the randomness of HIV-1 and MLV RNA copackaging are not clear. It is noteworthy that the cytoplasmic trafficking of viral RNAs seems to differ between HIV-1 and MLV (8). In the cytoplasm, the unspliced viral RNAs serve two distinct roles: either as mRNA templates for translation of viral proteins or as genomic RNAs to be packaged into virions (52). The interrelationship between translation and packaging of genomic RNAs for various retroviral species has been investigated by using chemical inhibitors to modulate host cell transcription or translation (8). Results from such experiments indicate that two distinct populations of unspliced MLV RNAs serve as either mRNA for translation or genomic RNA for packaging, while a single population of full-length HIV-1 RNAs functions as both mRNA and genomic RNA (8).

Contribution by HIV-1 accessory proteins. Another possibility is that the differences in recombination rates between HIV-1 and MLV may be attributed to the function of some of the HIV-1 regulatory/accessory proteins. As is common to all retroviral species, both HIV-1 and MLV have genes encoding Gag and Env structural proteins and Pol enzymes. The four Gag proteins (MA, CA, NC, and HIV-1 p6 or MLV

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

p12) and the two Env proteins (SU and TM) are structural components that make up the virion core and the outer membrane envelope. The three Pol proteins (PR, RT, and IN) provide essential enzymatic functions (60). However, unlike MLV, HIV-1 also encodes two regulatory proteins (Tat and Rev) and four accessory proteins (Vif, Vpr, Vpu and Nef), three of which (Vif, Vpr and Nef) are found in viral particles (14, 17).

It has been shown that the four accessory genes of HIV-1 (*vif*, *vpr*, *vpu*, and *nef*) can be deleted without affecting viral replication in some *in vitro* models (40). For example, although Vpr is required for efficient transduction of macrophages at a low MOI and quiescent peripheral blood T lymphocytes, its function in the nuclear import is fully dispensable in other cell types, such as HeLa cells (18, 40, 50). Tat can also be eliminated from the helper construct by using strong constitutive promoters (40). The function of these proteins in HIV-1 recombination has not been studied. The data presented in this thesis was produced with a helper construct that could express all HIV-1 regulatory/accessory proteins. To explore the contribution of these regulatory/accessory proteins to HIV-1 recombination, similar recombination assays can be performed with helper constructs that are deficient in the expression of each of these genes.

Rev may be an excellent candidate for exploring the potential role of HIV-1 accessory proteins in genetic recombination. Full-length and singly spliced HIV-1 RNAs are actively transported into cytoplasm via a Rev-mediated nuclear export pathway, in which Rev interacts a *cis*-acting sequence, RRE (49). In contrast, simple retroviruses adopt quite different strategies for the nuclear export of their viral RNAs. For example, Mason-Pfizer monkey virus (MPMV) possesses a constitutive transport element (CTE) in the 3' untranslated region of its genome. The MPMV CTE interacts with cellular factors to promote RNA export from the nucleus (49). Similar CTEs have been reported for several other simple retroviruses, including MLV, simian retrovirus-1 (SRV-1), RSV, and murine intracisternal-A particles ((29) and reference therein). For MLV, two additional *cis*-elements were also implicated in promoting the expression of unspliced RNAs (23,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

42). However, the mechanism by which MLV RNAs are exported has not been defined (29).

One assay that can be used to dissect the role of these different nuclear export pathways on genetic recombination is to make Rev/RRE-independent HIV-1 vectors. In fact, attempts have been made to substitute MPMV CTE for the function of RRE in the *helper* construct (19). However, in this system, the nuclear export of the *transducing* vector RNA was still Rev/RRE dependent since it contained an RRE sequence (19). All the transducing vectors used in this thesis for testing HIV-1 genetic recombination contained an RRE sequence between the 5' leader sequence and the *lacZ* gene (i.e. RRE+ transducing vectors), and our helper plasmid could express Rev proteins (i.e. Rev+ helper). Our HIV-1 vectors can be modified to make the helper plasmid Rev-deficient and/or the transducing vectors devoid of RRE sequences. Genetic recombination can then be tested under following conditions: (i) Rev- helper plasmid and RRE- transducing vector; (ii) Rev+ helper and RRE- transducing vector; (iii) Rev- helper and RRE+ transducing vector. It is expected that the effects on recombination of these three conditions will be very similar since under each condition Rev/RRE interaction will be abolished. Results from these experiments can be compared to those obtained in this thesis. The observation of a reduced recombination rate for HIV-1 will suggest a role of Rev/RRE-dependent nuclear export pathway in the difference in recombination rates between HIV-1 and MLV.

Architecture and composition of viral reverse transcription complexes. A third possibility is that our observed recombination properties reflect differences in the architecture/composition of reverse transcription complex (RTC) between HIV-1 and MLV. Retroviral reverse transcription occurs within a high-molecular-weight ribonucleoprotein complex—the RTC, which contains all the viral functions necessary for viral DNA synthesis, nuclear transport, and integration (58). The properties and components of the RTC from newly infected cells have been characterized by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

biochemical analysis, and are highly dependent on the conditions used for its isolation (20). Data from such studies suggest that the constituents of RTCs differ between MLV and HIV-1 (15, 16, 41, 58). HIV-1 RTC comprises at least MA, Vpr, IN and RT (15, 41, 58), while MLV RTC consists of at least CA, IN and RT (16, 58). It has been suggested that HIV-1 MA, Vpr and IN promote the nuclear translocation of viral RTC (58). However, little is known about whether these proteins function during reverse transcription. Only recently, it was shown that both MLV and HIV-1 IN proteins augmented viral DNA synthesis in infected cells (31, 61). Currently, it is not known whether the difference in the composition of RTCs between MLV and HIV-1 contributes to the distinct difference on their recombination rates.

Current data suggest that the composition of RTCs changes over time after formation, which may represent discrete steps of uncoating that take place in the cell cytoplasm and nucleus (15, 16). There are probably a number of discrete RTCs capable of performing reverse transcription in the cytoplasm. At least three distinct species of MLV RTCs are observed shortly after viral entry into cells (16), and two for HIV-1 (15). Although both MLV and HIV-1 have virion protein CA, CA is often found in MLV RTCs (15, 16, 36, 58). It has been speculated that HIV-1 CA may dissociate from HIV-1 RTCs very early after infection (15). It is not known whether the infrequent detection of CA in HIV-1 RTCs reflects the difference of HIV-1 and MLV RTCs on their stability and/or dynamics. Similarly, there is little information on how the process of reverse transcription is initiated after formation of RTCs within the target cell. Recently, it was demonstrated that HIV-1 RTCs associate with microtubules (6, 36) and that efficient reverse transcription is dependent on the presence of actin microfilaments (6). An EM analysis revealed that HIV-1 RTCs were large nucleoprotein structures consisting of packed filaments (41), suggesting that reverse transcription occurs in a relatively loose nucleoprotein complex.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Implications of the high rate of HIV-1 genetic recombination

The observation that HIV-1 recombines more frequently than MLV despite similar template switch rates has a great impact on retroviral research in general. In contrast to point mutations that accumulate gradually, genetic recombination can reassort the genetic information between two genomes, providing the potential for evolutionary leaps with far more dramatic genetic consequences. The high recombination rate may be an important evolutionary strategy for HIV-1 to maximize viability and to increase the fitness or the chance of survival of the population when encountering various environmental challenges (7). For example, it has been shown that recombination can bring together mutations resistant to single antiretroviral drugs and make the virus multi-drug resistant and/or with enhanced resistance (28, 38). The mechanism by which HIV-1 RT is resistant to AZT has been studied (9, 37). However, the recombination properties of such mutant RT have not been investigated.

Various groups, subtypes and recombinant forms of HIV-1 have been identified (53). There has been some evidence indicating the existence of biological differences among them (46). For instance, it has been suggested that CRF01-AE stains (recombinant strains between subtype A and the now extinct subtype E) can be transmitted more efficiently by heterosexual intercourse than subtype B strains and this may correlate with the heightened infectability of Langerhans' cells by CRF01-AE strains (56). The emergence of CRF01-AE recombinant in Southeast Asia may have simply been a chance event. However, the faster spreading of CRF01-AE than subtype B in this region suggests that some recombinants may emerge through natural selection based on their transmission efficiency (7).

The extraordinary genetic divergence among HIV-1 isolates has presented an unprecedented challenge to the diagnosis, prevention and treatment of HIV-1 infection (35, 39, 46, 51). For instance, enzyme-linked immunoassays and Western blot assays are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the most common antibody test systems used for the diagnosis of HIV-1, which are initially developed based on subtype B viruses. Although group O antigens were later incorporated into such assays, current assays need to be frequently reevaluated or improved in order to keep up with the rapid emergence of recombinant viruses (46).

Effect of dimerization initiation sequence on HIV-1 genetic recombination

Retroviral genomes consist of two usually identical RNA molecules that are noncovalently linked to each other through a dimer linkage structure (DLS) (21). The HIV-1 sequence responsible for dimerization has been localized to a short segment in the 5' leader sequence of viral genomic RNA, and was termed the dimerization initiation site (DIS) or stem-loop 1 (SL1) (21). It is now generally accepted that dimerization takes place by symmetric intermolecular interactions between the self-complementary sequences of the DIS loop, thus forming a "kissing-loop" complex (21). Among natural HIV-1 isolates, the palindromic sequences that cap DIS stem-loops vary (30). Recently, it has been shown that complementarity between the palindromic sequences in the DIS hairpins is necessary and sufficient for *in vitro* dimerization of two subtype RNAs (2). Furthermore, data from a cell-culture based recombination assay suggest that the presence of non-homologous DIS elements from different HIV-1 subtypes is not a principal obstacle to intersubtype recombination (57). The biology behind these findings remains unclear, partially due to the limited measurement precision of these experiments which were performed with replication-competent virus (57).

Vector and cell system to study the effect of DIS sequence variation on HIV-1 genetic recombination. To study the effects of DIS sequence variation on HIV-1 intermolecular recombination, we established an HIV-1 single-replication-cycle assay system (Figure 5-2). Two proviral vectors were used in the recombination assay: the recombination acceptor vector contained a *lacZ* gene disrupted by a 117-bp internal

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

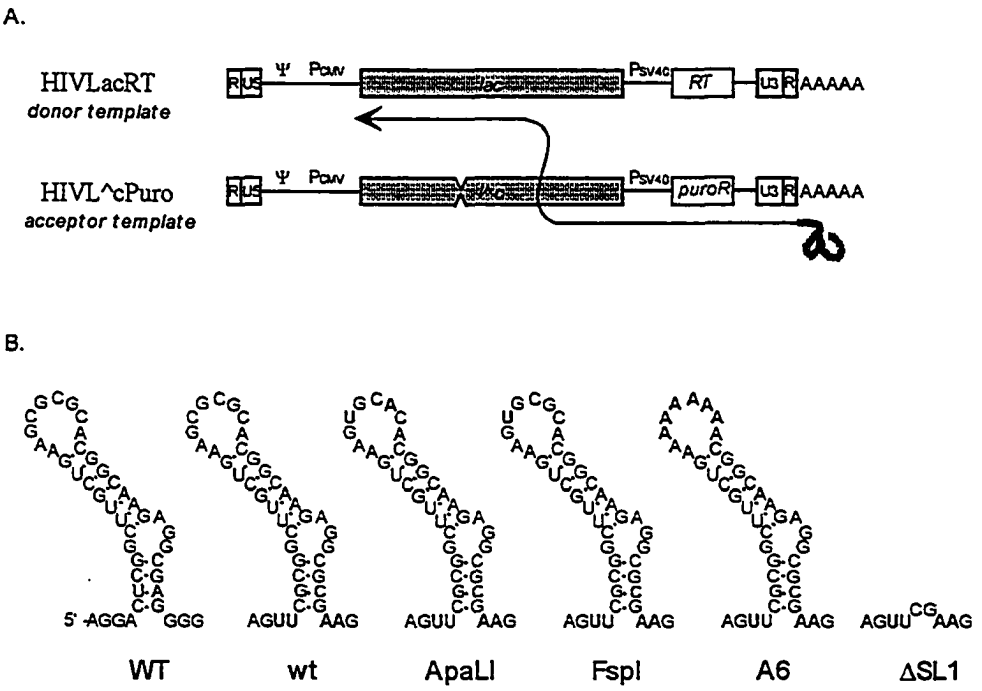


Figure 5-2. Two-vector assay to study the effect of DIS sequence variation on HIV-1 genetic recombination. **(A)** Schematic representation of how template switching during reverse transcription could yield a puromycin-resistant provirus with functional *lacZ*. The line with the arrowhead represents minus-strand DNA synthesis after minus-strand strong stop transfer. **(B)** Predicted secondary structures for DIS variants that were tested as part of the 5' leader sequences present in the recombination donor and acceptor RNAs shown in panel A. WT, natually occuring subtype B DIS; wt, WT with base substitutions at DIS base; ApaLI, naturally occuring subtype A DIS; FspI, naturally occuring DIS variant, seen in some subype B and D isolates; A6, a DIS loop substitution mutant; ΔSL1, a DIS deletion mutant.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

deletion followed by a puromycin-resistance gene, and the recombination donor vector had an intact *lacZ* gene and a 1.2-kb MLV RT coding sequence (Figure 5-2A). The rationale for using these vectors is as follows. Reverse transcription of the first vector would yield a puromycin-resistant provirus that did not express functional β-galactosidase; the second vector alone would yield a *lacZ* provirus that did not confer puromycin resistance. During reverse transcription of virions that had copackaged one of each vector RNA, template switching within regions of sequence identity could yield recombinant viral DNAs, some of which would contain functional versions of both *lacZ* and puromycin resistance genes (Figure 5-2A). The frequency of such recombinants can be scored as the ratio of the puromycin-resistant colonies that stained blue by X-gal to the total number of puromycin-resistant colonies. The MLV RT coding sequence in the recombination donor vector can serve as the sequence target to which a complementary synthetic oligonucleotide could specifically anneal in the RNA capture assay, which was developed to assess the proportion of the RNA heterodimers among total virion RNAs (Flynn and Telesnitsky, manuscript in preparation). The DIS sequences in the recombination donor and acceptor vectors can be replaced with other DIS variants to be tested. Pairs of recombination donor and acceptors with either identical or different DIS domains can be used in the recombination assay and the RNA capture assay, from which the effect of the DIS sequence variation on RNA dimerization and recombination can be deduced. Therefore, this system allows us to correlate the capability of heterodimer formation by different DIS variants and the observed recombination rates for HIV-1 in single replication cycles.

Several DIS variants were tested, including three naturally occurring DIS loop variants (WT/wt, ApaLI, and FspI) (54, 57), a DIS loop substitution mutant (A6) (57), and a DIS deletion mutant (ΔSL1) (10, 43). Producer cells were cotransfected with helper plasmid and approximately equimolar amounts of plasmids encoding the recombination donor and acceptor vectors. Fresh cells were infected, and puromycin resistant colonies

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were selected and stained with X-gal. Results from one or three independent cotransfection experiments for each virus, followed by infection experiments that generated more than 200 colonies on each of two plates for each virus stock, are shown in Figure 5-3. Several interesting scenarios were observed from the recombination assay. First, the recombination rates were very similar to each other for infections performed with vectors containing identical DIS sequences, ranging from 12.4 to 14.8%, including recombination donor and acceptor pairs that contained naturally occurring DIS variants (WT + WT, wt + wt, FspI + FspI, and ApaLI + ApaLI) and two mutant pairs (A6 + A6 and ΔSL1 + ΔSL1). Second, when recombination was tested between two different naturally occurring DIS variants (either wt + ApaLI or ApaLI + wt), it was reduced approximately by 3-fold no matter which vector was used for recombination donor or acceptor. Third, when recombination was tested between a naturally occurring DIS variant (wt) and each of the two DIS mutants (A6 or ΔSL1), the recombination rates were decreased by 3- to 6-fold when the recombination acceptor vector contained the wt DIS sequence and the recombination donor vector contained either of the mutant DIS sequences (Figure 5-3, wt + A6 and wt +ApaLI). Interestingly, the recombination rates showed no reduction when reciprocal vectors were used, i.e. the recombination acceptor vector had either of the mutant DIS sequences and the recombination donor vector had the wt DIS sequence (Figure 5-3, A6 + wt and ApaLI + wt).

Recombination rates after normalized with recombining population size. The reduction in recombination rates for experiments with vectors containing non-identical DIS variants could be potentially attributed to several factors. First, the recombination rates in Figure 5-3 were presented as blue colony frequency to the total number of puromycin-resistant colonies. To more accurately represent the recombination rates, the blue colony frequency needs to be normalized with the proportion of the heterozygous virions in the total virion population. If the relative abundance of the two virion RNAs is known, the proportion of the heterozygous virions among total virion population could be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

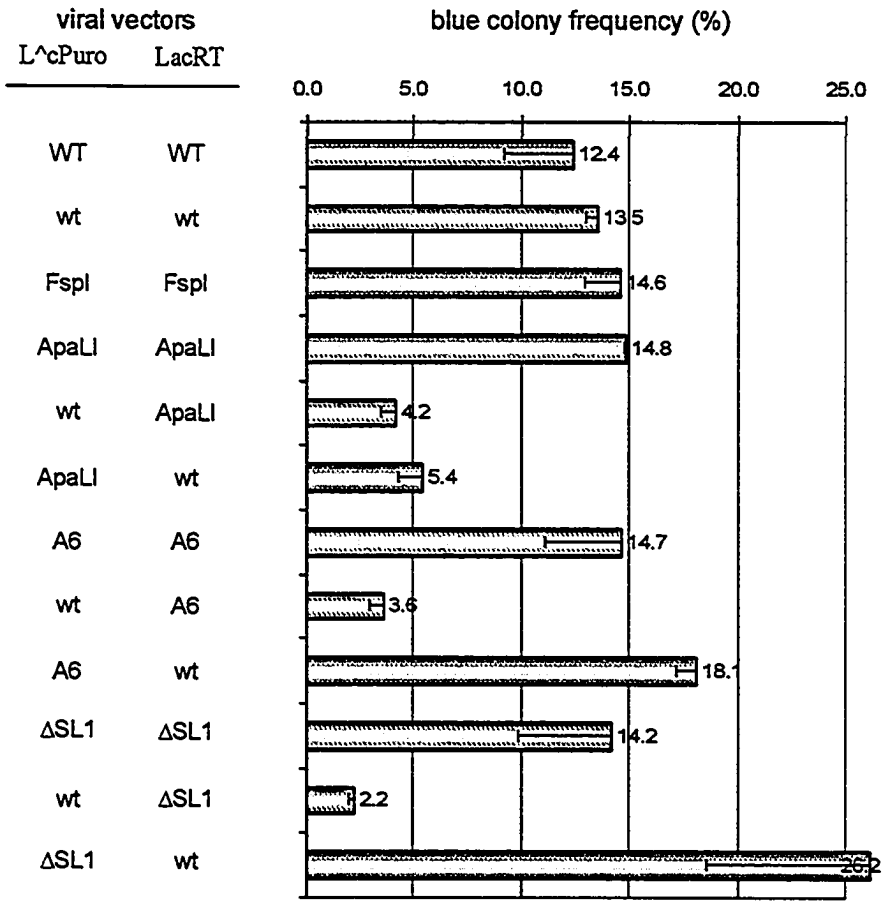


Figure 5-3. HIV-1 recombination rates between viral RNAs bearing identical or different DIS sequences. The recombination donor and acceptor proviral vectors for each virus were identified by the corresponding DIS variants contained in their 5' leader sequences, respectively. The recombination rates were represented as blue colonies per total puromycin-resistant colonies. All values shown were results from four separate cotransfection and infection experiments, and at least 400 puromycin-resistant colonies were tallied to yield each value.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

calculated based on Hardy-Weinberg equation assuming that two coexpressed viral RNAs are randomly copackaged into virions.

To determine virion RNA ratios, RNase protection assays were performed with virion RNAs prepared from aliquots of the same virion stocks that were used in the recombination experiment (Figure 5-4). Although equimolar vector DNAs were used to transfect producer cells, relatively fewer RNAs containing DIS mutations were packaged into virions than their wt counterpart (Figure 5-4, lanes wt + A6, A6 + wt, wt + ApaLI, and ApaLI + wt), consistent with the packaging defect reported for the same DIS mutants previously (10, 43, 57). The recombining population was then deduced from virion RNA ratios and the blue colony frequency was adjusted with the recombination population size to derive the recombination rates per recombining population. According to the calculated recombination rates per recombining population, the pattern of the difference between different recombination donor and acceptor pairs remained although the magnitude of difference slightly changed (data not shown). This result suggests that some other factors besides RNA expression levels mainly contribute to the reduction on recombination rates observed for experiments with non-identical DIS domains. One of such factors may the level of heterodimerization, which can be measured by RNA capture assays.

Dissecting the complex interrelationship between RNA dimerization and copackaging, and their effects on genetic recombination

If the above measurement of heterodimerization by vectors containing non-identical DIS sequences proves to be successful, such system can be further utilized to explore the complex interrelationship between RNA dimerization and copackaging, and their effects on genetic recombination. A partial list of questions relevant to this interaction network includes: Is copackaging synonymous to dimerization? When and where does dimerization take place? If two RNAs are copackaged but not dimerized, will

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

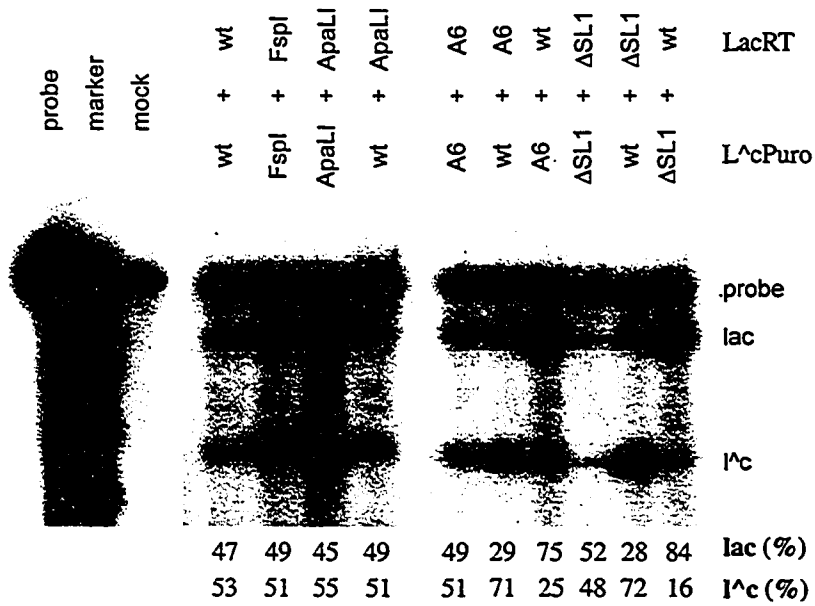


Figure 5-4. RNase protection assay to quantify virion RNAs. Virion RNAs were harvested from ET producer cells that had been cotransfected with the indicated pair of recombination donor and acceptor vectors and the HIV-1 helper construct. Sample that was harvested from cells transfected with the helper construct only was used control (mock). Positions corresponding to the mobility of undigested probe (probe), probe protecting a fragment of undisrupted *lacZ* (lac), and probe protecting a fragment of the disrupted *lacZ* (l^c) are indicated at the right. Recombination donor and acceptor proviral vectors for each virus were identified by the corresponding DIS variants contained in their 5' leader sequences, respectively. The relative abundance of the lac and l^c signals in each sample was labeled in percentage.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

recombination be less efficient than that between two copackaged and dimerized RNAs? Is the level of RNA copackaging affected by the inability of the two RNAs to heterodimerize? Is dimerization required for efficient RNA packaging or copackaging of two different RNAs?

One way to approach some of these questions is to utilize RNAs that are defective in either of the two functions: packaging and dimerization. Both naturally occurring DIS variants and other engineered DIS mutants can be tested for their packaging efficiency, dimerization and recombination. However, as shown by other studies, even subtle base substitutions caused various degree of deficiency in viral RNA encapsidation (3, 10, 22, 32, 43, 54, 55). Whether this is due to a direct effect on encapsidation through a disruption of the encapsidation signal or due to an indirect effect on encapsidation as the result of less efficient dimerization is not known. A large selection of DIS variants from *in vitro* dimerization studies have not been evaluated in cultured cell systems (26, 34). It is likely that some of those DIS variants are capable of being packaged to wild type level. If this is the case, they provide a valuable addition to the dimerization study toolbox.

In *in vitro* studies, some DIS variants in which the palindromic symmetry is disrupted by base substitutions are not capable of forming homodimers, but selected pairs of them could efficiently heterodimerize when the disrupted palindromes in the two RNAs are complimentary to each other (11, 33, 34, 44, 45). If such variants retain the same characteristic behavior *in vivo* as *in vitro*, they can be used for at least two purposes. First, since heterodimers predominate in the virion populations, intermolecular recombination studies using these vectors would have the advantage of not relying on any assumptions on copackaging randomness. Secondly, they could provide direct evidence as to whether or not dimerization affects genetic recombination. Furthermore, it is possible that a spectrum of dimerization levels could be established ranging from minimal heterodimerization to exclusive heterodimerization. These proposed experiments may benefit our exploration of the complex interrelationship among RNA dimerization,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

copackaging and recombination, and further advance our understanding of HIV-1 genetic recombination in general.

BIBLIOGRAPHY

1. **An, W., and A. Telesnitsky** 2002. HIV-1 genetic recombination: experimental approaches and observations AIDS Rev. **4:**195-212.

2. **Andersen, E. S., R. E. Jeeninga, C. K. Damgaard, B. Berkhout, and J. Kjems** 2003. Dimerization and template switching in the 5' untranslated region between various subtypes of human immunodeficiency virus type 1 J Virol. **77:**3020-30.

3. **Berkhout, B., and J. L. van Wamel** 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1 J Virol. **70:**6723-32.

4. **Bowman, R. R., W. S. Hu, and V. K. Pathak** 1998. Relative rates of retroviral reverse transcriptase template switching during RNA- and DNA-dependent DNA synthesis J Virol. **72:**5198-206.

5. **Brincat, J. L., J. K. Pfeiffer, and A. Telesnitsky** 2002. RNase H activity is required for high-frequency repeat deletion during Moloney murine leukemia virus replication J Virol. **76:**88-95.

6. **Bukrinskaya, A., B. Brichacek, A. Mann, and M. Stevenson** 1998. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton J Exp Med. **188:**2113-25.

7. **Burke, D. S.** 1997. Recombination in HIV: an important viral evolutionary strategy Emerg Infect Dis. **3:**253-9.

8. **Butsch, M., and K. Boris-Lawrie** 2002. Destiny of unspliced retroviral RNA: ribosome and/or virion? J Virol. **76:**3089-94.

9. **Caliendo, A. M., A. Savara, D. An, K. DeVore, J. C. Kaplan, and R. T. D'Aquila** 1996. Effects of zidovudine-selected human immunodeficiency virus type 1 reverse transcriptase amino acid substitutions on processive DNA synthesis and viral replication J Virol. **70:**2146-53.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

10. **Clever, J. L., and T. G. Parslow** 1997. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation Journal of Virology. **71:**3407-14.

11. **Clever, J. L., M. L. Wong, and T. G. Parslow** 1996. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA J Virol. **70:**5902-8.

12. **Delviks, K. A., W. S. Hu, and V. K. Pathak** 1997. Psi- vectors: murine leukemia virus-based self-inactivating and self- activating retroviral vectors J Virol. **71:**6218-24.

13. **Delviks, K. A., and V. K. Pathak** 1999. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching J Virol. **73:**7923-32.

14. **Emerman, M., and M. H. Malim** 1998. HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology Science. **280:**1880-4.

15. **Fassati, A., and S. P. Goff** 2001. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1 J Virol. **75:**3626-35.

16. **Fassati, A., and S. P. Goff** 1999. Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus J Virol. **73:**8919-25.

17. **Frankel, A. D., and J. A. Young** 1998. HIV-1: fifteen proteins and an RNA Annual Review of Biochemistry. **67:**1-25.

18. **Gallay, P., T. Hope, D. Chin, and D. Trono** 1997. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway Proc Natl Acad Sci U S A. **94:**9825-30.

19. **Gasmi, M., J. Glynn, M. J. Jin, D. J. Jolly, J. K. Yee, and S. T. Chen** 1999. Requirements for efficient production and transduction of human immunodeficiency virus type 1-based vectors J Virol. **73:**1828-34.

20. **Goff, S. P.** 2001. Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways J Gene Med. **3:**517-28.

21. **Greatorex, J., and A. Lever** 1998. Retroviral RNA dimer linkage J Gen Virol. **79:**2877-82.

22. **Haddrick, M., A. L. Lear, A. J. Cann, and S. Heaphy** 1996. Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1 J Mol Biol. **259:**58-68.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

23. **Hoshi, S., T. Odawara, M. Oshima, Y. Kitamura, H. Takizawa, and H. Yoshikura** 2002. cis-Elements involved in expression of unspliced RNA in Moloney murine leukemia virus Biochem Biophys Res Commun. **290:**1139-44.

24. **Hu, W. S., E. H. Bowman, K. A. Delviks, and V. K. Pathak** 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference J Virol. **71:**6028-36.

25. **Hwang, C. K., E. S. Svarovskaia, and V. K. Pathak** 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:**12209-14.

26. **Jossinet, F., J. C. Paillart, E. Westhof, T. Hermann, E. Skripkin, J. S. Lodmell, C. Ehresmann, B. Ehresmann, and R. Marquet** 1999. Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding Rna. **5:**1222-34.

27. **Julias, J. G., D. Hash, and V. K. Pathak** 1995. E- vectors: development of novel self-inactivating and self-activating retroviral vectors for safer gene therapy J Virol. **69:**6839-46.

28. **Kellam, P., and B. A. Larder** 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance J Virol. **69:**669-74.

29. **King, J. A., J. M. Bridger, F. Gounari, P. Lichter, T. F. Schulz, V. Schirrmacher, and K. Khazaie** 1998. The extended packaging sequence of MoMLV contains a constitutive mRNA nuclear export function FEBS Lett. **434:**367-71.

30. **Kuiken, C., B. Foley, B. Hahn, P. Marx, F. McCutchan, J. Mellors, S. Wolinksy, and B. Korber** (eds.) 2001 HIV Sequence Compendium. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

31. **Lai, L., H. Liu, X. Wu, and J. C. Kappes** 2001. Moloney murine leukemia virus integrase protein augments viral DNA synthesis in infected cells J Virol. **75:**11365-72.

32. **Laughrea, M., L. Jette, J. Mak, L. Kleiman, C. Liang, and M. A. Wainberg** 1997. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization J Virol. **71:**3397-406.

33. **Lodmell, J. S., C. Ehresmann, B. Ehresmann, and R. Marquet** 2000. Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site Rna. **6:**1267-76.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

34. **Lodmell, J. S., C. Ehresmann, B. Ehresmann, and R. Marquet** 2001. Structure and dimerization of HIV-1 kissing loop aptamers J Mol Biol. **311:**475-90.

35. **McCutchan, F. E.** 2000. Understanding the genetic diversity of HIV-1 Aids. **14:**S31-44.

36. **McDonald, D., M. A. Vodicka, G. Lucero, T. M. Svitkina, G. G. Borisy, M. Emerman, and T. J. Hope** 2002. Visualization of the intracellular behavior of HIV in living cells J Cell Biol. **159:**441-52.

37. **Meyer, P. R., S. E. Matsuura, A. M. Mian, A. G. So, and W. A. Scott** 1999. A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase Mol Cell. **4:**35-43.

38. **Moutouh, L., J. Corbeil, and D. D. Richman** 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:**6106-11.

39. **Najera, R., E. Delgado, L. Perez-Alvarez, and M. M. Thomson** 2002. Genetic recombination and its role in the development of the HIV-1 pandemic Aids. **16:**S3-16.

40. **Naldini, L.** 1998. Lentiviruses as gene transfer agents for delivery to non-dividing cells Curr Opin Biotechnol. **9:**457-63.

41. **Nermut, M. V., and A. Fassati** 2003. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes J Virol. 77:8196-206.

42. **Odawara, T., M. Oshima, K. Doi, A. Iwamoto, and H. Yoshikura** 1998. Threshold number of provirus copies required per cell for efficient virus production and interference in moloney murine leukemia virus-infected NIH 3T3 cells J Virol. **72:**5414-24.

43. **Paillart, J. C., L. Berthoux, M. Ottmann, J. L. Darlix, R. Marquet, B. Ehresmann, and C. Ehresmann** 1996. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis J Virol. **70:**8348-54.

44. **Paillart, J. C., R. Marquet, E. Skripkin, B. Ehresmann, and C. Ehresmann** 1994. Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA J Biol Chem. **269:**27486-93.

45. **Paillart, J. C., E. Skripkin, B. Ehresmann, C. Ehresmann, and R. Marquet** 1996. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA Proc Natl Acad Sci U S A. **93:**5572-7.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

46. **Peeters, M., and P. M. Sharp** 2000. Genetic diversity of HIV-1: the moving target Aids. **14:**S129-40.

47. **Pfeiffer, J. K., M. M. Georgiadis, and A. Telesnitsky** 2000. Structure-based moloney murine leukemia virus reverse transcriptase mutants with altered intracellular direct-repeat deletion frequencies J Virol. **74:**9629-36.

48. **Pfeiffer, J. K., R. S. Topping, N. H. Shin, and A. Telesnitsky** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription J Virol. **73:**8441-7.

49. **Pollard, V. W., and M. H. Malim** 1998. The HIV-1 Rev protein Annual Review of Microbiology. **52:**491-532.

50. **Popov, S., M. Rexach, G. Zybarth, N. Reiling, M. A. Lee, L. Ratner, C. M. Lane, M. S. Moore, G. Blobel, and M. Bukrinsky** 1998. Viral protein R regulates nuclear import of the HIV-1 pre-integration complex Embo J. **17:**909-17.

51. **Quinones-Mateu, M. E., and E. J. Arts** 1999. Recombination in HIV-1: update and implications AIDS Rev. **1:**89-100.

52. **Rabson, A. B., and B. J. Graves** 1997. Synthesis and processing of viral RNA, p. 205-262. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

53. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber** 1999. HIV-1 Nomenclature Proposal, p. 492-505. *In* C. L. Kuiken, B. Foley, B. Hahn, B. Korber, F. McCutchan, P. A. Marx, J. W. Mellors, J. I. Mullins, J. Sodroski, and S. Wolinksy (eds), Human Retroviruses and AIDS 1999. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

54. **Shen, N., L. Jette, C. Liang, M. A. Wainberg, and M. Laughrea** 2000. Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging J Virol. **74:**5729-35.

55. **Shen, N., L. Jette, M. A. Wainberg, and M. Laughrea** 2001. Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization J Virol. **75:**10543-9.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

56. **Soto-Ramirez, L. E., B. Renjifo, M. F. McLane, R. Marlink, C. O'Hara, R. Sutthent, C. Wasi, P. Vithayasai, V. Vithayasai, C. Apichartpiyakul, P. Auewarakul, V. Pena Cruz, D. S. Chui, R. Osathanondh, K. Mayer, T. H. Lee, and M. Essex** 1996. HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV Science. **271:**1291-3.

57. **St Louis, D. C., D. Gotte, E. Sanders-Buell, D. W. Ritchey, M. O. Salminen, J. K. Carr, and F. E. McCutchan** 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:**3991-8.

58. **Stevenson, M.** 1996. Portals of entry: uncovering HIV nuclear transport pathways Trends in Cell Biology. **6:**9-15.

59. **Svarovskaia, E. S., K. A. Delviks, C. K. Hwang, and V. K. Pathak** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching J Virol. **74:**7171-8.

60. **Vogt, V. M.** 1997. Retroviral virions and genomes, p. 27-69. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (eds), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

61. **Wu, X., H. Liu, H. Xiao, J. A. Conway, E. Hehl, G. V. Kalpana, V. Prasad, and J. C. Kappes** 1999. Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex J Virol. **73:**2126-35.

62. **Zhang, J., and Y. Ma** 2001. Evidence for retroviral intramolecular recombinations J Virol. **75:**6348-58.

63. **Zhang, J., and C. M. Sapp** 1999. Recombination between two identical sequences within the same retroviral RNA molecule J Virol. **73:**5912-7.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.