

2004

Mechanisms of retroviral reverse transcription and assembly

Sara Kirsten Rasmussen
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Rasmussen, Sara Kirsten, "Mechanisms of retroviral reverse transcription and assembly" (2004). *Graduate Theses, Dissertations, and Problem Reports*. 2090.
<https://researchrepository.wvu.edu/etd/2090>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

MECHANISMS OF RETROVIRAL REVERSE TRANSCRIPTION AND ASSEMBLY

DISSERTATION

Submitted to the School of Medicine
at
West Virginia University
In Partial Fulfillment of the Requirements for
The Degree of Doctor of Philosophy
In Microbiology and Immunology

Morgantown
West Virginia
2004

Committee Members:

Dr. Nyles Charon
Dr. Daniel Flynn
Dr. Vinay K. Pathak
Dr. Michael Miller
Dr. Wei-Shau Hu, Committee Chair

Keywords: Retrovirus, Reverse transcription, retroviral assembly, complementation, reverse transcriptase, integrase, capsid, gene reconstitution, gene therapy.

Copyright 2004 by Sara K. Rasmussen

ABSTRACT

Mechanisms of Retroviral Reverse Transcription and Assembly

Sara Kirsten Rasmussen

The work described in these dissertation studies examines three aspects of retrovirus replication and dissects the uncovered phenomena to increase understanding of how retroviruses thrive in complex host environments. The first aspect of replication discussed is the mechanisms behind minus-strand DNA transfer, a crucial step of reverse transcription. We proved for the first time that non-viral sequences can be utilized to carry out this step of the replication process, and that the minus-strand DNA transfer step can be exploited to reconstitute a functional gene in the infected host cell. This finding serves as a proof-of-principle for a new class of self-activating gene therapy vectors. In the second study, mechanisms of trans-complementation were studied between two distantly related retroviruses, murine leukemia virus (MLV) and spleen necrosis virus. We showed that two different species of pol proteins can be incorporated into virions and cooperate to complete the reverse transcription and integration steps of replication. This finding highlights how different species of retroviruses could possibly interact and combine to form new and possibly more virulent or pathogenic species. Finally, in the third work, a previously uninvestigated region of MLV capsid protein was examined. Specifically, the contribution of the 33 carboxyl-terminal amino acids of MLV CA were found to consist of highly charged residues. The contribution of these residues to virus assembly was investigated via deletion analysis. A new phenotype of MLV CA assembly defect was demonstrated, and a structural model was proposed to explain the possible role of this region, termed the charged assembly helix, in MLV assembly. In summary, the work presented in this dissertation has revealed new properties of retroviral replication as well as further clarified previous knowledge of retroviral biology.

SARA KIRSTEN RASMUSSEN

EDUCATION

1996 - present West Virginia University Morgantown, WV
M.D./Ph.D. candidate

1994 - 1996 University of Virginia Charlottesville, VA
B.A./ Chemistry

1992 - 1994 Hollins College Roanoke, VA
Biology and Chemistry Major

PROFESSIONAL EXPERIENCE

2004 Virginia Commonwealth University/Medical College of Virginia
Health Systems. Residency in General Surgery. Richmond, VA

1999 – 2002 HIV Drug Resistance Program, NCI Frederick, MD
Predocctoral Fellow, Cancer Research Training Award

a. Predocctoral Fellowship to study retroviral replication with Wei-Shau Hu, Ph.D., chief, Viral Recombination Section

1997 Department of Biochemistry, WVU Morgantown, WV
Van Liere Research Program

b. \$2,500 summer stipend to study Ca-ATPase in the lab of Jim Mahaney, Ph.D.

1994 - 1996 Mylan Pharmaceuticals Morgantown, WV
Chemist, Research and Development

c. Full-time summer chemist with Mylan Pharmaceuticals R&D department, to assist in methods development and quality control.

1992 - 1994 Hollins College Roanoke, VA
Research Assistant

d. Research assistant in the Comparative Morphology lab of Bruce Young, Ph.D, to study behavior in spitting cobras and skull anatomy of giraffes and okapis.

1993 Department of Chemistry, WVU Morgantown, WV
Summer Internship

- e. Internship in the Physical Chemistry lab of Paul Jagodzinski, Ph.D., to study Raman Spectroscopy of organic compounds

1990 – 1992 National Institutes of Occupational Safety and Health

Morgantown, WV

High School Internship

- f. Internship in the Pharmacology lab of Jeff Fedan, Ph.D., to study pharmacology of nucleotide analogs in smooth muscle tissue.

MEETINGS AND PUBLICATIONS

Second HIV DRP Antiviral Symposium: Understanding Antiviral Drug Resistance. Dec 9-12, 2002. Chantilly, VA. Cheslock, SR, Poon, DTK, Fu, W, Rhodes, TR, Henderson, LH, McGrath, C, and W.-S. Hu. Identification and Characterization of a Novel “Electric Wire” Motif in the CA of Gammaretroviruses.

Cold Spring Harbor Laboratory International Retroviruses Meeting, Poster Sessions. May 23-28, 2002. Cheslock, SR, Poon, DTK, Fu, W, Rhodes, TR, Henderson, LH, McGrath, C, and W.-S. Hu. Identification and Characterization of a Novel “Electric Wire” Motif in the CA of Gammaretroviruses.

Cold Spring Harbor Laboratory International Retroviruses Meeting, Poster Sessions. May 23-28, 2002. Cheslock, SR, Zhang, W, Hu, W.-S., and VK Pathak. Two Species of Pol from Distantly Related Retroviruses Can Interact to Complete Viral Replication.

HIV Drug Resistance Program Think Tank meeting April 2002. Cheslock, SR, Poon, DTK, Fu, W, Rhodes, TR, Henderson, LH, McGrath, C, and W.-S. Hu. Identification and Characterization of a Novel “Electric Wire” Motif in the CA of Gammaretroviruses.

First HIV DRP Symposium: Understanding Antiviral Drug Resistance. Dec 3-6, 2000. Chantilly, VA. Cheslock, SR, Anderson, JA, Hwang, CK, Pathak, VK, and W.-S. Hu. Utilization of Non-Viral Sequences for Minus-Strand DNA Transfer and Gene Reconstitution.

Cold Spring Harbor Laboratory International Retroviruses meeting. May 23-28 2000. Cheslock, SR, Anderson, JA, Hwang, CK, Pathak, VK, and W.-S. Hu. Utilization of Non-Viral Sequences for Minus-Strand DNA Transfer and Gene Reconstitution.

HIV Drug Resistance Program Think Tank meeting April 6, 2000. Cheslock, SR, Anderson, JA, Hwang, CK, Pathak, VK, and W.-S. Hu. Utilization of Non-Viral Sequences for Minus-Strand DNA Transfer and Gene Reconstitution.

WVU Health Sciences Research Day, March 1998. Rasmussen, SK, and

Mahaney, J. Studies of Ca-ATPase in cardiac muscle.

In Preparation: Zaslau's Review for Step 2, Blackwell Publishing. Contributor to Psychiatry section. Due in press September 2004.

Publications

Cheslock, S.R., D.T.K. Poon, W. Fu, T.D. Rhodes, L.E. Henderson, K. Nagashima, C. McGrath, and W.-S. Hu. Charged Assembly Helix: An important domain for virus assembly and particle size determination. Accepted to *J. of Virology*, April 2003.

Svarovskaia, E. S., Cheslock, S. R., Zhang, W. H., Hu, W. S., Pathak, V. K. Retroviral mutation rates and reverse transcriptase fidelity. *Frontiers in Biosciences*, 2003 8:D117-134

Cheslock, S.R., J.A. Anderson, C.K. Hwang, V.K. Pathak, and W.S. Hu. Utilization of non-viral sequences for minus-strand DNA transfer and gene reconstitution in retroviral replication. *Journal of Virology*. 2000, 74(20):9571-9579.

*Rasmussen, S., Krimm, H., and B.A. Young. On the spitting behavior of cobras. *Journal of the Zoological Society of London*, 1995.

AWARDS RECEIVED

- g. International Health Program Scholarship recipient, 2003, for a one-month elective in Uganda
- h. Scholarship to attend the American Medical Women's Association Meeting in Atlanta, GA. Jan 24-26, 2003.
- i. Van Liere Reseach Stipend, Summer 1997
- j. Scholar, half tuition award 1992-1994
- k. Natural Sciences Student of the Year, Honorable Mention 1994

Hobbies

- l. Scuba Diving, Piano, Skiing, photography, biking, quilting, watercolor painting, gardening

TABLE OF CONTENTS

Abstract.....	ii
Curriculum Vitae	iii
Table of Contents.....	vi
List of Tables and Illustrations.....	viii
Acknowledgments.....	xii
Chapter 1. Introduction.....	1
Figure Legends.....	23
References.....	24
Figures.....	32
Chapter 2. Utilization of Non-Viral Sequences for Minus-strand DNA	
Transfer and Gene Reconstitution During Retrovirus Replication.....	36
Figure Legends and Figures.....	63
Tables.....	71
References.....	73
Chapter 3. Pol Gene Products From Two Distantly Related Retroviruses	
Can Co-Package And Cooperate To Complete Reverse Transcription	80
Figure Legends.....	108
Tables and figures.....	109
References.....	119
Chapter 4. Charged assembly helix motif in murine leukemia virus capsid:	
An important region for virus assembly and particle size determination	126

Tables.....	148
Figure Legends and Figures.....	150
References.....	162
Chapter 5. Discussion.....	178
Figure Legends.....	181
Figures.....	182
References.....	183
Appendix.....	190

LIST OF TABLES AND ILLUSTRATIONS

Chapter 1. Introduction.....	1
Figure 1: Retroviral Life Cycle.....	32
Figure 2: Nucleic Acid Structures of Retroviral Genomes.....	33
Figure 3: Reverse Transcription	34
Figure 4: Crystal Structure of HIV-1 RT.....	35
Chapter 2. Utilization of Non-Viral Sequences for Minus-strand DNA Transfer and Gene Reconstitution During Retrovirus Replication	36
Table 1: Primers used for Vector Construction, PCR Amplification, and Sequencing	71
Table 2: Viral titers generated by PG13 cells transfected with vector plasmids	72
Figure 1: MLV-based vectors and strategy used to study the requirement of sequence context in mediating minus-strand DNA transfer.....	65
Figure 2: Experimental protocol	66
Figure 3: Flow cytometry analyses of infected cells	67
Figure 4: Strategy for PCR amplification of proviral LTRs and DNA sequencing	68
Figure 5: Two models for the generation of proviruses with U3-GF-R-U5 in the LTRs.....	69
Figure 6: Model for generating proviruses with U3-FP-R-U5	

by using read-in RNA transcripts	70
Chapter 3. Pol Gene Products From Two Distantly Related	
Retroviruses Can Co-Package And Cooperate To Complete	
Reverse Transcription.	80
Table 1: Primers used to construct Pol	
mutants.....	109
Table 2: Replication Properties of SNV and	
MLV <i>Pol</i> Mutants.....	114
Table 3: Quantification of mature vs. immature	
virion particles in IN HHCC insertion mutants.....	112
Table 4: Complementation of SNV and MLV <i>Pol</i>	
Phenotypically Mixed Virions Derived	
from the SNV- <i>gag</i> D150E Cell Line.....	115
Table 5: Complementation of SNV <i>Pol</i> Mutants in the	
SNV <i>gag</i> -YVNN Cell Line	116
Figure 1: A: Vectors used in study	
B: Experimental Protocols.....	110
Figure 2: A: MLV RT ^{IN+} cell line	
B: Protocol for Complementation.....	111
Figure 3: Western Blot profile of Viruses harboring IN mutations.....	113
Figure 4: A:SNV RT ^{IN+} cell line	
B: Protocol for complementation of the	

SNV RT ⁻ IN ⁺ with RT ⁺ IN ⁻ vectors.....	117
Appendix, Figure 1	118
Chapter 4. Charged assembly helix motif in murine leukemia virus	
capsid: An important region for virus assembly	
and particle size determination	126
Table 1: Virus titers generated by <i>gag-pol</i> expression	
constructs with wild-type or mutant motifs	148
Table 2: Distribution of particle sizes from pWZH30-	
and pSR33-derived virions	149
Figure 1: Location and sequence of the motif rich in charged residues	153
Figure 2: Amino acid sequences of the motif rich in	
charged residues in wild-type and mutant <i>gag-pol</i>	
expression constructs	154
Figure 3: Characterization of pWZH30- and	
pSR33-transfected cells and virions derived from these cells	155
Figure 4. Electron micrographs of pWZH30-derived	
virions (A) and pSR33-derived virions.....	156
Figure 5. Comparison of the RNA contents in pWZH30-	
and pSR33-derived virions.	157
Figure 6. Western analyses of the wild type and deletion mutants.....	158
Figure 7: Western analyses of mutants generated to test the	
proposed α -helical structure	159

Figure 8. Space-filling models of the predicted helical structure of the wild type and mutant MLV charged assembly helix	
of the wild type and deletion mutants	160
Chapter 5. Discussion.....	178
Figure 1: Multiple sequence alignment of SNV and MLV Integrase.....	183
Figure 2: Chou-Fasman structure prediction values for SNV CA.....	182

ACKNOWLEDGEMENTS

When I started medical school, I would have described myself as an enthusiastic but uncertain student. The decision to enter the M.D./Ph.D. program was the best decision I have ever made. I owe a deep debt of gratitude to my advisors Wei-Shau Hu and Vinay Pathak, for their deep investment in my education and growth. Through their training I developed into a much more focused person and acquired a precious experience at the NCI in Frederick. Under their guidance at NCI I was immersed in a world that challenged me intellectually. Their encouragement of my academic development has helped me transform into the person I always wanted to be and find rewards in the scientific process and sharing of ideas.

I also would like to thank my committee, Dr. Vinay Pathak, Dr. Mike Miller, Dr. Nyles Charon, Dr. Dan Flynn, and Dr. Herbert Thompson, for their continued support and interest in my progress. Their comments and encouragement have improved my critical thinking skills tremendously. I also have been privileged to enjoy the support of the MD/Ph.D. program, in particular the support of Dr. Charles Craig and Dr. John Traubert who, along with Dr. Jim Mahaney, originally encouraged me to join the program. I appreciate the continued support of the program from Dr. Anne Cather and Dr. Norman Ferrari.

The efforts of any student are very self-involved, and if one is lucky enough to have the support of family and friends, it makes the journey much easier. The largest debt of love I owe is to my husband and life-long friend Julian, who has unrelentingly been my biggest supporter of the last several years. I also would like to thank my

parents, Barbara and Norval Rasmussen and Sharon and Julian A. “Tiny” Cheslock, and brother and sister-in-law Christopher and Renee for their belief in my abilities. I especially acknowledge my mother who has been my life-long example of what a professional woman can be. I also would like to thank the many friends and lab-mates who took the time to help me learn and grow, as well as made the journey a welcome camaraderie.

Lastly, I indulge myself to mention the names of two of my most precious friends that could not be here to share in the culmination of this step in my education, Christopher Wasson and Julian A. “Tiny” Cheslock. My development as a scientist and a person would be incomplete without their influence and belief in me, and I dedicate this work in part to their memory.

CHAPTER ONE

INTRODUCTION

Worldwide, retroviruses affect a startling degree of human ecology and biology. Most notable of these viruses is human immunodeficiency virus type 1 (HIV-1), the retrovirus which causes Acquired Immune Deficiency Syndrome, or AIDS. Currently 40 million people are infected with HIV-1, a global epidemic that is devastating populations, especially in developing countries (21). Available treatments are known as highly active antiretroviral therapy (HAART), and consist of combinations of antiretroviral drugs directed at different stages of the virus life cycle. Treatments are often complicated by high toxicity and severe side effects. HAART is able to block a certain percentage of viral replication but cannot eradicate HIV-1 from the infected host. Eventually, HIV-1 escapes every possible treatment regimen by its ability to rapidly mutate and recombine. Furthermore, the treatments that are available are only affordable in developed countries, such as the United States and Europe, where less than 1% (900,000 people in the United States) of the populations are infected. HAART is not available in third world countries, where the infection rates in some places approach 40% or higher. HAART is not available in the developing nations in Asia and Africa. If no effective treatments for HIV-1 are found, everyone who is infected with this virus will eventually die of complications caused by it. Research of HIV-1 is necessary to develop more viable therapies, and most importantly, a vaccine to halt the spread of HIV-1 in the population.

AIDS alone is enough to mandate a call to retrovirology research. However, retroviruses also cause cancer in humans and animals. They can cause cancer by direct infection of cells, as in the case of human T-cell leukemia virus (HTLV) or murine

leukemia virus (MLV) (13, 18, 26, 33, 39, 40). Retroviruses have also been demonstrated to capture and package oncogenes from the genome of the infected host and cause cancer by activating or altering these genes (49). Increasingly, more and more cancers and other diseases have been thought to arise from infection with viruses, though not necessarily retroviruses, via oncogene capture and activation, and by direct infection.

Finally, retroviruses are uniquely suited as viable gene therapy vectors. Because retroviruses permanently insert a copy of their genetic material into the infected host genome, they are ideal candidates for gene delivery vehicles. Retroviral vectors can be engineered to carry therapeutic genes to replace host cell deficiencies, or they may be modified to carry “suicide” genes that target cells such as virus-infected cells or cancer cells. Such vectors have been used with success to treat Adenosine Deaminase (ADA) deficiency and Severe Combined Immune Deficiency (SCID) (4, 12).

Life Cycle. The retroviridae are divided into 7 genera, largely based on sequence similarity of the *pol* gene. All retroviruses are single-stranded RNA viruses that replicate through a DNA intermediate via an enzyme called reverse transcriptase. HIV-1 is a member of the lentivirus genus. MLV and spleen necrosis virus (SNV) are members of the gammaretrovirus genus. The three studies discussed in this work were carried out with either MLV- or SNV- based systems. The life cycle of a retrovirus is shown in Figure 1.

Retroviruses require the synthesis of a double-stranded DNA intermediate of their RNA genome to successfully complete their replication cycle. A retrovirus first infects a cell by attaching to a surface receptor of a host cell and entering the cell either by direct fusion or endocytosis (Fig 1A). In most retroviruses, reverse transcription occurs in the

cytoplasm of the infected host cell. Reverse transcription is the process by which the viral enzyme reverse transcriptase (RT) copies the genomic viral RNA into a double-stranded DNA (Fig 1B). This viral DNA can integrate into the host genome to form a provirus. Integration of the provirus into host cell DNA is mediated by the viral enzyme integrase (IN). In HIV-1, this double-stranded DNA is imported to the nucleus and integrated into the genome of the host cell to form a provirus (Fig 1C). Unlike HIV-1, the pre-integration complexes of SNV and MLV lack the ability to traverse the nuclear membrane; so therefore, MLV and SNV viral DNA can only integrate into the chromosomes of mitotically active cells because the nuclear membrane is dissolved temporarily during mitosis. Once inserted, the cellular enzyme RNA Pol II can transcribe the provirus to make RNA transcripts, which are modified by cellular enzymes so each transcript contains 3' poly-A tail and a 5' cap (Fig 1D)(51). Some of the viral RNA transcripts are spliced, and both the full-length viral RNA and spliced transcripts can be used as a template for translation to generate viral proteins. The full-length viral RNA transcripts are specifically packaged into the virion. The viral Gag polyproteins and viral RNA are transported to the membrane for virus assembly. Viral proteins and two copies of viral RNA are assembled into a virion (Fig 1E), which are released from the cell surface by budding.

Virus Genome Structure. The process of reverse transcription requires that retroviruses possess a mechanism by which the viral RNA can be synthesized into a DNA copy. The two forms of a retroviral genome, RNA and DNA, are shown in figure 2. The terminal regions of the retrovirus RNA genome differ from that of the DNA form, and therefore reverse transcription of the RNA into DNA involves the duplication of the

terminal ends of the viral RNA in the DNA copy. The viral sequence in the RNA transcript starts at the beginning of the 5' R and concludes at the end of the 3' R (Fig 2A). As a result there is one copy of the U3 at the 3' end, one copy of U5 at the 5' end, and two copies of R in the viral RNA. During the process of reverse transcription, the U3 and U5 are duplicated to form the long terminal repeat (LTR) at both ends of the DNA genome (Fig 2B). LTRs flank the internal part of the proviral genome and contain many *cis*-acting sequences necessary for viral replication, including a U3 region, a U5 region, an R region, and two attachment (*att*) sites, attL and attR. The U3 contains promoter and enhancer sequences important for RNA transcription (22, 30). The R plays a critical role in the minus-strand DNA transfer of reverse transcription, and may contain other regulatory sequences such as a transcription termination signal and *cis*-acting elements that enhance transcription (23). The *att* sites are recognized by the IN to specifically integrate the viral DNA into the host cell chromosome (9, 41). The U5 region plays an important role in reverse transcription and carries the transcription termination signal for some retroviruses. The region between the two LTRs (Figure 3) also contains *cis*-acting sequences, such as the primer binding site (PBS), packaging or encapsidation signals (denoted Ψ for MLV and E for SNV), and the polypurine tract (PPT). Ψ and E are required for the incorporation of viral RNA into virions. The PBS is the site of initiation of reverse transcription, and the PPT is the first site of initiation of plus-strand DNA synthesis. All retroviruses have at least four genes, termed *gag*, *pro*, *pol*, and *env*. *Gag* (short for group-specific antigen) encodes the proteins that form the viral structures, *pro* (short for protease) encodes the viral protease, and *pol* encodes RT, and IN. *Env*

encodes the envelope glycoproteins, which mediate entry into the host cell during infection. These proteins are further discussed below.

Retroviruses as research tools. Genetically modified retroviral systems can be used to study aspects of the retroviral life cycle. A retroviral vector is a retrovirus that has had some or all of its protein-coding genes deleted so that only the *cis*-acting sequences remain intact, an aspect that is exploitable to develop a system for studying virus replication. Replication of the viral RNA into DNA requires the *cis*-acting sequences discussed above, but can occur in the absence of the genes encoding the viral proteins. The proteins necessary for viral replication can be provided in *trans* by “helper” constructs that express the viral enzymes and structural proteins (37). A retroviral vector can be introduced into cells containing helper constructs; since the vector RNA contains the packaging signal the RNA will be specifically packaged by viral structural proteins to produce virus particles. Produced viruses can infect permissive target cells, and the viral proteins in the virions can carry out reverse transcription and integration of the viral DNA. However, viruses generated from the helper cells cannot replicate, because the vector RNA packaged into virions do not encode viral proteins; once integrated into the host these vectors cannot produce more infectious virions. Therefore vectors and helper cell lines make it possible to investigate retrovirus biology during one round of replication.

Reverse Transcription. The process of reverse transcription requires an uncharged cellular tRNA that codes for a specific amino acid to anneal to the PBS, which is located near the 5' end of the retrovirus RNA (Figure 3). RT uses the tRNA as a primer and synthesizes a DNA copy of the U5 and R that is termed minus-strand strong-

stop DNA (Fig 3A). A component of the RT, known as ribonuclease H (RNase H), degrades the viral RNA in RNA: DNA hybrids and leaves the minus-strand strong-stop DNA as single stranded DNA. The R region of the single-stranded minus-strand strong-stop DNA is complementary to the R region at the 3' end of the viral RNA. Using this complementarity, the minus-strand strong-stop DNA is transferred to the 3' end of the retroviral RNA, an event termed minus-strand DNA transfer (Fig 3B). Once this event occurs, minus-strand DNA synthesis can resume and start copying the RNA near the 3' end of the retroviral genome. Frequently during reverse transcription, RT can dissociate and re-associate from the template. If the RT dissociates from the viral template while it is copying the 5' R region into minus-strand DNA, only a portion of this R is copied into minus-strand DNA, and is termed minus-strand weak-stop DNA. It is estimated that murine leukemia virus (MLV) and spleen necrosis virus (SNV) use minus-strand weak-stop DNA to carry out minus-strand DNA transfer 10-20% of the time (34, 42) .

Once minus-strand DNA synthesis proceeds past the PPT, RNase H makes a specific cleavage in this sequence, resistant to degradation by RNase H, which allows it to serve as the primer for the initiation of plus-strand DNA synthesis. Plus-strand DNA synthesis occurs in a 5'-to-3' direction, and copies the minus-strand DNA and 18 nt of the tRNA primer (47)(Fig 3C). After the plus-strand DNA synthesis reaches the 5' end of the minus-strand DNA, a second transfer step, termed plus-strand DNA transfer, must occur to complete reverse transcription. Plus-strand DNA transfer is mediated by the complementarity between the PBS sequences in the minus-strand DNA and the 18 nucleotides of the 3' end of the tRNA that is complementary to the PBS, on the 5' end of the minus-strand DNA. This second DNA strand transfer event does not occur until the

plus-strand strong-stop DNA has been synthesized fully and the minus-strand DNA synthesis has proceeded through the PBS. Most plus-strand DNA synthesis occurs in a continuous fashion, but in some retroviruses additional priming sites for plus-strand DNA synthesis allow for discontinuous synthesis of plus-strand DNA (43). Once plus-strand DNA transfer has occurred, plus-strand DNA synthesis continues until the entire minus-strand DNA is copied. The result is a double-stranded DNA copy of the retrovirus genome, ready for insertion into the host cell's genome (Fig 3D).

Integration of Viral DNA. IN carries out the insertion of the viral cDNA into the host genome to form the provirus. Integration is a two-step process. The first step is to process the 3'OH on each end of the viral *att* sites in the cDNA, and the second step is a strand-transfer reaction that involves a nucleophilic attack of the 3'OH to the acceptor DNA in the cellular genome. Following these two reactions, cellular ligases repair the nicks left in the viral-cellular DNA junction. Selection of sites in the host genome for integration is not site-specific. However, there may be more preferred sites in the genome. For example, areas where DNA is coiled around histones seem to be favored and not transcriptionally active DNA that is unwound, as was previously thought (50). Once integrated into the host genome, the provirus is permanently ensconced in the genetic material of the cell.

Retroviral Assembly. Retroviral assembly is a complex process that requires an ordered interplay among viral proteins, viral RNA, and cellular components. Once a retrovirus has its DNA integrated into the infected cell, it must successfully transport transcripts of its RNA to the cytoplasm of the infected cell, synthesize its major structural proteins and enzymes, and transport these components to a designated location in the

infected cell so that a highly organized virus particle can form. During virus assembly parameters such as virion size, content, and morphology must all be within narrow constraints for formation of viable virus particles. The roles of many of these components of viral assembly are still being dissected.

Retroviral assembly is directed by the group-specific antigen gene (*gag*) that is both necessary and sufficient to direct assembly of virus-like particles *in vitro* (17). In murine leukemia virus (MLV), the Gag polyproteins consist of four protein domains: matrix (MA), p12 (a protein of 12 kD in size), capsid (CA), and nucleocapsid (NC). MA contains a myristolation signal, and is crucial for Gag targeting to the membrane of the infected cell for budding. The p12 proteins contain a conserved late domain motif, PPPY, which is critical for release of the particle from the cell. NC is a nucleic-acid binding protein that binds to viral RNA to stabilize it, and although NC has non-specific nucleic acid binding properties, it is responsible for the specific incorporation of viral genomic RNA into the virions (3, 5). NC also has been shown to play a role in the processivity of reverse transcription (55), as well as in strand-transfer events. CA is the major structural protein of retroviruses. It forms the virus core, which encapsulates the viral enzymes and nucleic acid. The Gag polyproteins are synthesized together as a large precursor protein and remain so throughout the assembly and budding process. Shortly after virus release from the infected cell the Gag polyproteins are cleaved into their smaller components by the virus protease. The protease is a domain of the larger Gag-Pol precursor protein that contains the virus enzymes for reverse transcription, integration, and proteolytic processing. With the exception of human foamy virus (HFV), the Pol polyproteins are always translated as part of a Gag-Pol precursor, at

roughly 10% the rate of the Gag polyproteins. This Gag-Pol polyproteins are translated either by an in-frame suppression of a stop codon between the Gag and PR sequences in the mRNA, or by a -1 frame shift after the NC sequences (25, 46, 53, 54). The Gag-Pol and Gag polyproteins associate together to form the budding virus particles.

In MLV, the first step of assembly is trafficking the Gag and Gag-Pol polyproteins to the plasma membrane, from which they bud as immature virions. Once at the membrane, the Gag and Gag-Pol proteins associate together in an organized fashion and begin their egress from the infected cell. During this process, viral RNA must be incorporated into the budding particle. The genomic RNA acts as a structural scaffold on which the Gag and Gag-Pol molecules associate (38). Once the virus has completely budded from the cell, it is released from the outer surface. Shortly after this, viral protease in the Gag-Pol cleaves all the polyproteins to their smaller components. Proteolytic processing leads to maturation of the virus core to make a virus that is capable of infecting another cell. In this process, several parameters are crucial: the specific incorporation of viral RNA, exclusion of cellular components, and maintenance of a well-defined particle size and morphology.

The maturation of viruses is evident by electron micrograph (EM) studies as a transition from a hollow-looking sphere to spheres that contain an electron-dense region in their centers (19). Mature, infectious virus particles are typically 100 nm in diameter and contain a condensed virus core in the center.

CA has been studied in several different retroviruses, and a brief overview of what is known about this protein in different retroviruses is discussed below. The role of the CA protein in retrovirus assembly is not clearly understood. It is well known that the

CA protein forms the condensed cores of the mature virus particles, but mutational analyses of this protein indicate it plays a role in assembly as well. CA proteins consist of two domains, an N-terminal domain (NTD) and a C-terminal domain (CTD). Cross-linking studies of the N terminal and C-terminal CA domains in HIV-1 revealed that both domains participate in CA dimerization (36) that is required for formation of the highly organized core structure. HIV-1 CA has also been shown to bind human cyclophilin A (Cyp A) at its amino terminal end, possibly to help promote virus core disassembly following infection (20). Mutational analyses of the NTD of HTLV-1 were found to be critical for particle production, while mutations in the CTD were not (44).

In the central region of the CA protein is a universally conserved stretch of 20 amino acids termed the Major Homology Region (MHR) (10). The structure of the MHR has been examined by nuclear resonance magnetic (NMR) spectroscopy and circular dichroism spectroscopy, and a helical structure to the MHR was proposed (6, 7). Several roles for the MHR during virus assembly have been proposed. Mutational analyses of the MHR in HIV-1 have suggested this region has a role in morphogenesis, and it has also been suggested that the MHR may be involved in CypA binding (8). In addition, mutation of the conserved residues of the MHR led to abolishment of viral replication and significant reduction of particle formation in vivo (35). Lastly, the MHR may be involved in promoting ordered interactions between Gag and Gag-Pol molecules (48). CA association properties have been demonstrated to be pH dependent, suggesting that the milieu surrounding the CA proteins during assembly influences its oligomerization properties (16).

CA proteins have also been recognized as a genetic determinant of size in Rous Sarcoma Virus (RSV) particles (28). Specifically, a region of CA as well as the spacer peptide (SP) mediated the size of RSV particles. Deletions of the CA-SP region resulted in low-density particles of various sizes. Furthermore, it was also proposed that the SP of RSV Gag was involved in oligomerization and/or folding of the RSV CA subunits during assembly (11). HIV-1 also contains a spacer peptide, p2, and mutations in the CA-p2 region abrogated HIV-1 core formation (1, 27). Furthermore, it was suggested that this CA-p2 region forms a putative alpha-helix structure critical for formation of mature virus cores (1). Thus, two genetically distinct retroviruses, RSV and HIV-1, have been demonstrated to possess sequences in the C-terminal portion of CA that work in conjunction with a small peptide to ensure proper particle formation.

Few studies have looked at the C-terminal domain of CA in other retroviruses to investigate a possible role for this region in virus assembly. Although some other retroviral families, such as the gammaretrovirus family, do not possess a spacer peptide between CA and NC like RSV and HIV-1, they must possess a mechanism by which they form mature particles. It is possible that a region in the C-terminal end of CA or the N-terminal end of NC also contributes the function of the spacer peptide in other retroviruses. Extensive mutational analysis of murine leukemia virus (MLV) NC has been done, and this region does not seem to play a role in virus capsid formation, or in virus particle size determination. (45, 55-57). However, there is a sequence at the end of MLV CA, which has not been characterized and could possibly play the same role as the CA-SP domain in RSV and the CA-p2 domain in HIV-1 in virus core formation. In aim three, mutational analysis of this region will be discussed, as well as its impact on virus

assembly, morphology, and replication. A structural model will be proposed to explain the impact of this region on MLV virus assembly.

Viral Enzymes. The PR domain of the Pol protein cleaves the individual viral proteins from their polyproteins precursors late in the assembly process. The budding process is poorly understood, but it involves ordered Gag-Gag and Gag-GagPol interactions among the Gag and Gag-Pol precursor proteins to regulate the size and content of the new virions. The budding and assembly process is regulated such that new virus particles are of a defined size (100 nm in diameter) and morphology (circular with a dense core).

The products of the *Pol* gene, namely RT, RNase H (RH), and integrase (IN) are the main participants in reverse transcription and integration, respectively. RT is a RNA- and DNA-dependent DNA polymerase, and is required for polymerization during reverse transcription. IN is the viral enzyme responsible for integration of viral DNA into the host cell, but it has also been shown to play a non-enzymatic role in reverse transcription in MLV and HIV-1 as well (29, 32, 52).

The primary sequence of RT encodes its polymerase function in the amino terminal two-thirds and the RH function in the 3' end (Figure 4). It is possible to mutate the polymerase region of RT without affecting RH function, and vice versa.

HIV-1 RT is found as a heterodimer solution, consisting as two subunits- a p51 subunit and a p66 subunit (31). The difference between the p51 and p66 subunit is that the p66 subunit contains the RH domain. Although HIV-1 RT has two polymerase subunits, all enzymatic activity is located in the p66 subunit. In contrast, it has been hypothesized that MLV RT is a monomer that dimerizes when presented with a template

(REF), and the oligomerization properties of SNV RT have not been published. Crystal structures of HIV-1 RT show that the active p66 subunit is shaped like a right hand. The domains of RT have been correspondingly called the “fingers”, “palm”, and “thumb” domains, and there is also a connecting region and the RNase H domain. The palm domain contains the catalytic core of RT. The catalytic core domain of RT contains three conserved aspartate residues, a triad required for polymerase activity. Two of the three conserved aspartate residues are located in the YVDD motif common to all retroviral RTs. The fingers form the template-primer grip, and the thumb helps to position the template as it proceeds through the RT complex. The fingers and thumb of RT are thought to be flexible, a property that helps to position the template and primer in the enzyme complex. The RH domain of RT contains a catalytic core that binds metal ions for its enzymatic activity. By itself, HIV-1 RNase H cannot bind its nucleic acid substrate; hence it requires physical association with the RT to be active in vivo.

IN carries out the integration reaction to form the provirus during viral infection. Retroviral INs range from 280 to 450 amino acids in length and contain three conserved regions: an amino terminal HHCC motif, a catalytic core, and a less conserved carboxyl terminus. The HHCC motif is thought to bind Zinc to form a zinc finger, the central catalytic core has a D,D(35)E motif (a conserved feature of endonucleases), and the 3' end has DNA binding properties. IN proteins must oligomerize on each end of the virus cDNA to carry out these steps; therefore a minimum of four IN molecules per infectious event are needed to integrate one viral DNA.

Biological Impact of Reverse Transcription. The process of reverse transcription is important to retrovirus biology in many ways. RT has a low affinity for

the viral template during viral DNA synthesis compared with eukaryotic polymerases, rendering the enzyme likely to dissociate from the template frequently during polymerization. Low affinity leads to the high recombination rates previously measured in retroviruses and contributes to retroviral evolution. Mutations and recombinations occurring during reverse transcription generate viral variants that increase host immune system evasion and resistance to antiviral therapy. Vaccines designed to prevent or treat HIV-1 infection will have to confer protection from a diverse population of viruses. Increased viral diversity also poses a treatment problem, as HIV-1 is able to acquire resistance to chemotherapeutic agents. Therefore, understanding the mechanisms of reverse transcription and recombination may have direct implications for management of the global AIDS epidemic. In addition, retroviral vectors have been used in gene therapy trials to treat human diseases. Understanding reverse transcription and recombination can help develop better and safer vectors for gene delivery. Additionally, new targets are also needed for antiretroviral therapy; possible targets could be developed to disrupt viral assembly.

Aim 1. In MLV, the R region is primarily involved in the minus-strand DNA transfer reaction, and conflicting reports have tested whether the transfer reaction is driven by specific RNA sequences in the RNA or by the sequence homology between the two R regions. Some labs have shown that other viruses' R sequences can mediate minus-strand DNA transfers while mediation of minus-strand DNA transfers by non-viral sequences was not detected (2). Substitution of different virus R sequences suggests that homology plays a role in minus-strand DNA transfer.

In the first chapter, the importance of the ‘R’ sequence context on reverse transcription was investigated. Specifically, a set of retroviral vectors containing non-viral sequences in their LTRs was constructed to see if these sequences could be used instead of the native viral R sequence to mediate minus-strand DNA transfer. Furthermore, these sequences were engineered such that their use for minus-strand DNA transfer would result in functional reconstitution and activation of a gene in the provirus.

By providing alternate sequences for homology during minus-strand transfer we directly determined the requirement of sequence in the R region. The R region’s role in minus-strand transfer could either be through the homology contained in the two copies of the R region or by specific sequences in the R region. We proposed that the mechanisms by which the R region mediated minus-strand DNA transfer was through the homology it provides between the 5’ and 3’ end of the viral genome, and not from specific sequences contained in the R region. If this was true, then alternate sequences could be inserted into the LTRs of a retroviral vector and homology between these regions would mediate minus-strand DNA transfer. Homology plays a role in other template-switching events in reverse transcription. Direct repeat deletion is a mechanism by which the RT dissociates from the RNA template during reverse transcription and re-associates with the template at a different place that contains sequence identity to the newly synthesized DNA. The result is that all the intervening sequences between the points of disassociation and re-association are deleted from the transcribed DNA. This phenomenon has been successfully exploited to reconstitute functional genes in vivo and a model has been proposed to explain it (15). Retrovirus vectors that contained directly repeated sequences underwent deletion at a high rate (>90%), a process called direct

repeat deletion. The sequences of the direct repeat did not appear to be important, as several different sequences have been used to demonstrate this phenomenon. Unless RT has different properties when copying the R region of the RNA, it is reasonable to propose that nonviral sequences of homology could mediate minus-strand DNA transfer as well as the native viral R region. Some labs have suggested that R regions cannot be substituted, but these reports were generated from in vitro systems that only roughly approximate conditions in viruses (2). Recently, it was shown that R sequences could mediate minus-strand DNA transfer even when they are incomplete, with as little as 12 nt in length (14). Furthermore, if non-viral sequences could mediate minus-strand DNA transfer, sequences could be inserted into the viral vector such that successful minus-strand DNA transfer using the non-viral sequence would reconstitute a functional gene in the LTR of the provirus. This would be a novel strategy for a self-activating gene therapy vector.

Aim 2. The highly flexible nature of retroviruses suggests that two species of retroviruses could interact. If a single cell were infected with two different retroviruses, there is a potential that the Gag and Gag-Pol transcripts could interact and form a phenotypically mixed virus. Normally, this would not happen because the Gag polyproteins are the necessary and sufficient element for virus particle production, and it has been suggested that two different species of Gag, particularly of MLV and SNV, cannot co-assemble to form infectious viruses (Beasley and Hu, unpublished data). However, since Pol proteins are always expressed in the context of a Gag-Pol fusion protein (with the exception of HFV), it is not known if different species of Pol could be assembled into the same virus particle. A chimeric Gag-Pol expression vector could be

used to determine if this mixing of Pol species into the virus were possible. If it were possible and two different species of Pol could co-assemble into the same virion this would expose another mechanism by which two different species of retrovirus Pol could interact to increase viral diversity.

In order to determine whether different species of Pol can interact and co-assemble into the same virus particle, a system that makes it possible to determine whether two different species of Pol are in the virus must be designed. The best way to devise such a system would be trans-complementation, which has been a powerful tool for dissecting the roles of the individual Pol proteins in the retrovirus (24). To successfully utilize a trans-complementation system to determine co-packaging properties of different Pols, mutations were introduced into the Pol domains of an SNV Gag-Pol and an SNV Gag-MLV Pol expression vectors. A cell line was constructed that stably expressed both species of Pol, and viruses produced from this cell line were tested to see if they contained both species of Pol.

More recently, evidence has been provided that RT and IN are both necessary for efficient reverse transcription. HIV-1 IN could be provided in trans by a Vpr-IN fusion protein to rescue the virus carrying the IN mutations that disrupted the HHCC domain. HIV-1 vectors containing various mutations that disrupted the HHCC domain were unable to synthesize early (minus-strand strong-stop), intermediate (U3-U5), or late (R-gag) minus-strand DNA. However, early and late viral DNA synthesis could be rescued by providing IN in trans (52). In addition, only Vpr fused with HIV-1 IN was able to effectively rescue viral DNA synthesis. Vpr fused with HIV-2 IN only increased viral DNA synthesis 2-3 fold compared to IN mutants alone, although HIV-2 IN increased

integration frequency almost 100-fold, indicating that HIV-2 IN was in the virus particles and reverse transcription complex. This study indicated that IN was critical for successful reverse transcription in HIV-1 and that there was a direct physical interaction between the RT and IN. Furthermore, the roles of IN were somewhat species specific- HIV-1 IN could rescue reverse transcription and integration, but HIV-2 IN could only rescue integration. A similar approach was used to establish an IN role in reverse transcription for MLV. A Gag Δ NC-IN fusion protein was constructed to provide IN to MLV viruses in trans. It was found that without the functional IN provided from the Gag Δ NC-IN, reverse transcription in viruses carrying deletions of partial deletions of IN certain mutant forms of MLV IN was not detectable. Once the Gag Δ NC-IN fusion construct was introduced, reverse transcription was restored to wild type levels. Specifically, the C-terminal thirty-four amino acids of MLV IN was found to be crucial for reverse transcription in this study (29).

In the second chapter, the functional complementation of RT and IN proteins from two distantly related simple retroviruses was examined. Specifically, whether heterologous RT and IN from spleen necrosis virus (SNV) and MLV could complement each other *in vivo* was investigated in order to establish whether two species of Pol could co-assemble into the same virion. Trans-complementation is a powerful tool for dissecting the individual contributions of different retroviral proteins to the virus life cycle. Although RT alone can reverse transcribe RNA into DNA in a test tube, current studies indicated that more proteins were involved in the viral DNA synthesis *in vivo*. The second chapter investigated whether RT and IN from distantly related viruses could functionally complement each other. This provided insight into the mechanisms of

interaction between these two enzymes. Furthermore, this also delineated the interactions between distantly related retroviruses. Pol mutants in SNV and MLV *Pol* were constructed and used to establish a trans-complementation system for studying mutations in IN and RT and their effects on reverse transcription and integration.

Aim 3. Retroviral assembly is a complex process and many insights still are needed before it can be properly understood. During the assembly process, a retrovirus must regulate its size, contents and morphology within narrow constraints to ensure virus viability. It is known that Gag is sufficient for particle formation and release, and that individual protein components each have specific roles in particle assembly. Some retroviruses, such as Rous Sarcoma Virus (RSV) and HIV-1, encode proteins between CA and NC that are short peptides, the spacer peptide (SP) for RSV, and p2 for HIV-1. The spacer peptide and the p2 domain were found to be critical for proper assembly in RSV and HIV-1. Specifically, the spacer peptide and p2 each worked with CA protein to form the proper particle size and morphology. Even though not all retroviruses encode short peptides between the CA and NC to carry out this function, it was reasoned that they must have a mechanism by which they carry out these assembly functions. It was hypothesized that gammaretroviruses utilized a distinct motif at the 3' end of the CA domain to carry out this function.

In the third chapter, mechanisms of retrovirus assembly were studied, specifically the contribution of a unique motif in the 3' end of MLV CA to virus assembly. This novel motif has been observed in several gammaretroviruses, and its role in gammaretrovirus assembly has not been explored. Additionally, the role of CA as a structural protein in retroviruses has been well established, but its role in the assembly

process is not well understood. The impact of this domain on the budding and assembly process was investigated, particularly whether it contributed the same function to assembly in MLV as the spacer peptide in RSV and p2 in HIV-1. Several deletion mutants of this region were constructed and tested to determine the features of this motif that are required for its function, including its role in virus size determination, content, and morphology.

This page intended to be left blank.

FIGURE LEGENDS

Figure 1: Life Cycle of a retrovirus. Red circles denote enzymes used in reverse transcription and viral transcription. The yellow circle represents the cytoplasm of the infected cell. The blue circle is the nucleus. The wavy lines indicate cellular chromosomes. Blue and red rectangles represent the LTRs during replication.

Figure 2: Two nucleic acid forms of retroviral genomes. RNA is denoted with a dotted line, DNA with a heavy solid line. AttL, left attachment site, attR, right attachment site.

Figure 3: The steps of reverse transcription. RNA is denoted as a thin dark line. Minus-strand DNA is denoted as a dotted line. Plus-strand DNA is a thick solid line.

Figure 4: Structure of Retroviral *Pol* gene, with crystal structure of HIV-1 RT. From Retroviruses, edited by J.M. Coffin, et al.

REFERENCES

1. **Accola, M. A., S. Hoglund, and H. G. Gottlinger.** 1998. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J Virol* **72**:2072-8.
2. **Allain, B., J. B. Rascle, H. de Rocquigny, B. Roques, and J. L. Darlix.** 1998. CIS elements and trans-acting factors required for minus strand DNA transfer during reverse transcription of the genomic RNA of murine leukemia virus. *J Mol Biol* **277**:225-35.
3. **Berkowitz, R. D., A. Ohagen, S. Hoglund, and S. P. Goff.** 1995. Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric Gag polyproteins during RNA packaging in vivo. *J Virol* **69**:6445-56.
4. **Blaese, R. M., K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, and et al.** 1995. T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science* **270**:475-80.
5. **Certo, J. L., T. O. Kabdulov, M. L. Paulson, J. A. Anderson, and W. S. Hu.** 1999. The nucleocapsid domain is responsible for the ability of spleen necrosis virus (SNV) Gag polyprotein to package both SNV and murine leukemia virus RNA. *J Virol* **73**:9170-7.
6. **Clish, C. B., D. H. Peyton, and E. Barklis.** 1998. Solution structures of human immunodeficiency virus type 1 (HIV-1) and moloney murine leukemia virus

- (MoMLV) capsid protein major-homology-region peptide analogs by NMR spectroscopy. *Eur J Biochem* **257**:69-77.
7. **Clish, C. B., D. H. Peyton, and E. Barklis.** 1996. Spectroscopic study of an HIV-1 capsid protein major homology region peptide analog. *FEBS Lett* **378**:43-7.
 8. **Colgan, J., H. E. Yuan, E. K. Franke, and J. Luban.** 1996. Binding of the human immunodeficiency virus type 1 Gag polyprotein to cyclophilin A is mediated by the central region of capsid and requires Gag dimerization. *J Virol* **70**:4299-310.
 9. **Colicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* **42**:573-80.
 10. **Craven, R. a. J. W.** 1991. Form, Function and Use of Retroviral Gag Proteins. *AIDS* **5**:639-654.
 11. **Craven, R. C., A. E. Leure-duPree, C. R. Erdie, C. B. Wilson, and J. W. Wills.** 1993. Necessity of the spacer peptide between CA and NC in the Rous sarcoma virus gag protein. *J Virol* **67**:6246-52.
 12. **Culver, K. W., W. R. Osborne, A. D. Miller, T. A. Fleisher, M. Berger, W. F. Anderson, and R. M. Blaese.** 1991. Correction of ADA deficiency in human T lymphocytes using retroviral-mediated gene transfer. *Transplant Proc* **23**:170-1.
 13. **D'Andrea, A. D., J. F. Moreau, and M. O. Showers.** 1992. Molecular mimicry of erythropoietin by the spleen focus-forming virus gp55 glycoprotein: the first stage of Friend virus-induced erythroleukemia. *Biochim Biophys Acta* **1114**:31-41.

14. **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.
15. **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.
16. **Ehrlich, L. S., Liu, T., Scarlata S., Chu B., and Carter, C.A.** 2001. HIV-1 Capsid Protein Forms Spherical (Immature-Like) and Tubular (Mature-Like) Particles in Vitro: Structure Switching by pH-induced Conformational Changes. *Biophys J* **81**:586-594.
17. **Felsenstein, K. M., and S. P. Goff.** 1988. Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. *J Virol* **62**:2179-82.
18. **Feuer, G., and I. S. Chen.** 1992. Mechanisms of human T-cell leukemia virus-induced leukemogenesis. *Biochim Biophys Acta* **1114**:223-33.
19. **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.
20. **Gamble, T. R., F. F. Vajdos, S. Yoo, D. K. Worthylake, M. Houseweart, W. I. Sundquist, and C. P. Hill.** 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**:1285-94.
21. **Garrett, L.** 1994. *The Coming Plague: Newly Emerging Diseases in a World Out of Balance.* Penguin Books, New York.

22. **Graves, B. J., R.N. Eisenmand, and S.L. McKnight.** 1985. Delineation of transcriptional control signals within the Moloney murine sarcoma virus long terminal repeat. *Mol Cell Biol* **5**:1948-1958.
23. **Guntaka, R. V.** 1993. Transcription termination and polyadenylation in retroviruses. *Microbiol Rev* **57**:511-21.
24. **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.
25. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* **230**:1237-42.
26. **Johnson, P., and S. Benchimol.** 1992. Friend virus induced murine erythroleukaemia: the p53 locus. *Cancer Surv* **12**:137-51.
27. **Krausslich, H. G., M. Facke, A. M. Heuser, J. Konvalinka, and H. Zentgraf.** 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol* **69**:3407-19.
28. **Krishna, N. K., S. Campbell, V. M. Vogt, and J. W. Wills.** 1998. Genetic determinants of Rous sarcoma virus particle size. *J Virol* **72**:564-77.
29. **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.

30. **Laimins, L. A., P. Gruss, R. Pozzatti, and G. Khoury.** 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. *J Virol* **49**:183-9.
31. **Le Grice, S. F., T. Naas, B. Wohlgensinger, and O. Schatz.** 1991. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *Embo J* **10**:3905-11.
32. **Leavitt, A. D., L. Shiue, and H. E. Varmus.** 1993. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J Biol Chem* **268**:2113-9.
33. **Li, J. P., A. D. D'Andrea, H. F. Lodish, and D. Baltimore.** 1990. Activation of cell growth by binding of Friend spleen focus-forming virus gp55 glycoprotein to the erythropoietin receptor. *Nature* **343**:762-4.
34. **Lobel, L. I., and S. P. Goff.** 1985. Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. *J Virol* **53**:447-55.
35. **Mammano, F., A. Ohagen, S. Hoglund, and H. G. Gottlinger.** 1994. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J Virol* **68**:4927-36.
36. **McDermott, J., S. Karanjia, Z. Love, and E. Barklis.** 2000. Crosslink analysis of N-terminal, C-terminal, and N/B determining regions of the Moloney murine leukemia virus capsid protein. *Virology* **269**:190-200.
37. **Miller, A. D., J. V. Garcia, N. von Suhr, C. M. Lynch, C. Wilson, and M. V. Eiden.** 1991. Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J Virol* **65**:2220-4.

38. **Muriaux, D., J. Mirro, D. Harvin, and A. Rein.** 2001. RNA is a structural element in retrovirus particles. *PNAS* **98**:5246-5251.
39. **Palmarini, M., and H. Fan.** 2001. Retrovirus-induced ovine pulmonary adenocarcinoma, an animal model for lung cancer. *J Natl Cancer Inst* **93**:1603-14.
40. **Palmarini, M., J. M. Sharp, M. de las Heras, and H. Fan.** 1999. Jaagsiekte sheep retrovirus is necessary and sufficient to induce a contagious lung cancer in sheep. *J Virol* **73**:6964-72.
41. **Panganiban, A. T., and H. M. Temin.** 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature* **306**:155-60.
42. **Ramsey, C. A., and A. T. Panganiban.** 1993. Replication of the retroviral terminal repeat sequence during in vivo reverse transcription. *J Virol* **67**:4114-21.
43. **Ratray, A. J., and J. J. Champoux.** 1987. The role of Moloney murine leukemia virus RNase H activity in the formation of plus-strand primers. *J Virol* **61**:2843-51.
44. **Rayne, F., Bouamr, F, Lalanne, J., and R. Z. Mamoun.** 2001. The NH₂-terminal Domain of hte Human T-Cell Leukemia Virus TYpe 1 Capsid Protein is Involved in Particle Formation. *J Virol* **75**:5277-5287.
45. **Rein, A., D. P. Harvin, J. Mirro, S. M. Ernst, and R. J. Gorelick.** 1994. Evidence that a central domain of nucleocapsid protein is required for RNA packaging in murine leukemia virus. *J Virol* **68**:6124-9.
46. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. *Nature* **293**:543-8.

47. **Smith, J. K., A. Cywinski, and J. M. Taylor.** 1984. Initiation of plus-strand DNA synthesis during reverse transcription of an avian retrovirus genome. *J Virol* **49**:200-4.
48. **Srinivasakumar, N., M. Hammarskjold, and D. Rekosh.** 1995. Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that affect particle formation and Gag- Pol precursor incorporation. *J. Virol.* **69**:6106-6114.
49. **Varmus, H.** 1988. Retroviruses. *Science* **240**:1427-35.
50. **Weidhaas, J. B., E. L. Angelichio, S. Fenner, and J. M. Coffin.** 2000. Relationship between retroviral DNA integration and gene expression. *J Virol* **74**:8382-9.
51. **Wickens, M.** 1990. In the beginning is the end: regulation of poly(A) addition and removal during early development. *Trends Biochem Sci* **15**:320-4.
52. **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.
53. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proc Natl Acad Sci U S A* **82**:1618-22.
54. **Yoshinaka, Y., R. B. Shames, R. B. Luftig, G. W. Smythers, and S. Oroszlan.** 1985. In vitro cleavage of Pr65gag by the Moloney murine leukaemia virus

- proteolytic activity yields p30 whose NH₂-terminal sequence is identical to virion p30. *J Gen Virol* **66 (Pt 2)**:379-83.
55. **Zhang, W. H., C. K. Hwang, W. S. Hu, R. J. Gorelick, and V. K. Pathak.** 2002. Zinc Finger Domain of Murine Leukemia Virus Nucleocapsid Protein Enhances the Rate of Viral DNA Synthesis in Vivo. *J Virol* **76**:7473-7484.
56. **Zhang, Y., and E. Barklis.** 1997. Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation. *J Virol* **71**:6765-76.
57. **Zhang, Y., and E. Barklis.** 1995. Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J Virol* **69**:5716-22.

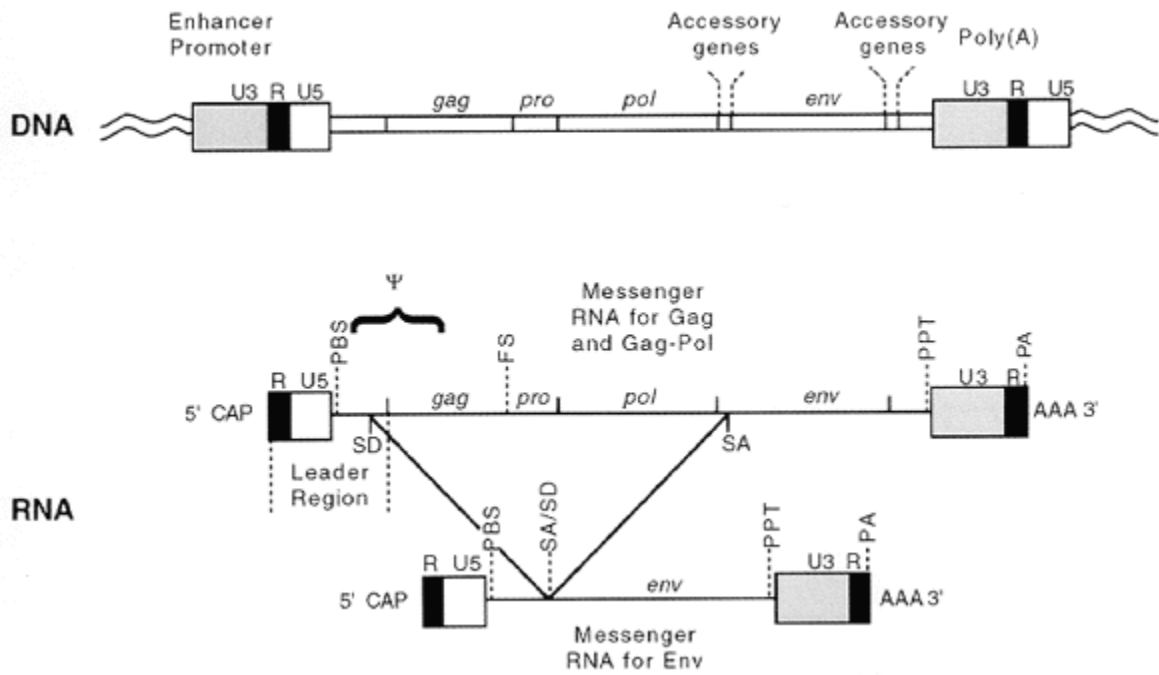


Figure 2. Nucleic Acid Structure of Retroviral Genomes

From Retroviruses, Coffin et al.

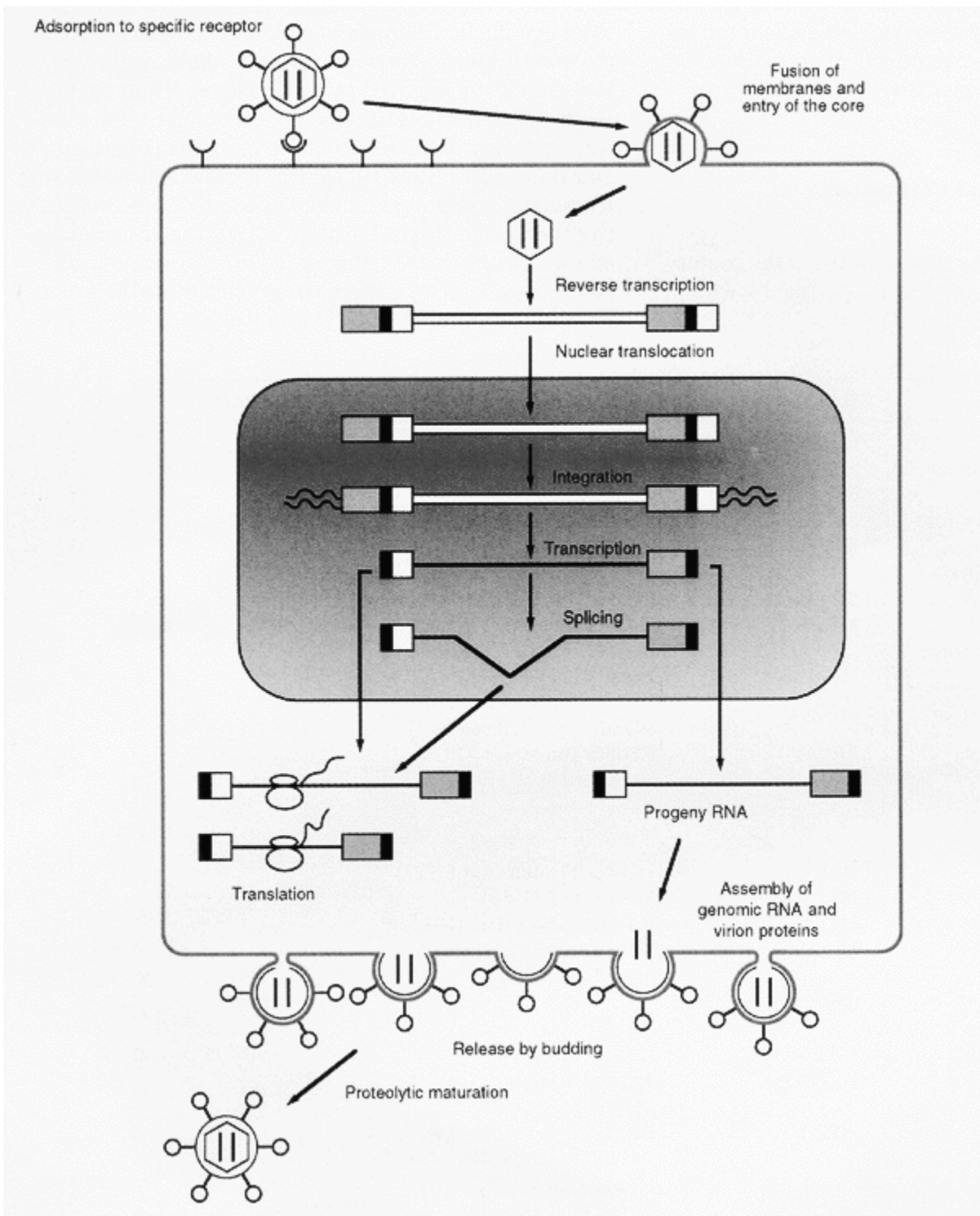


Figure 1. Retroviral Life Cycle. From Retroviruses, Coffin et al.

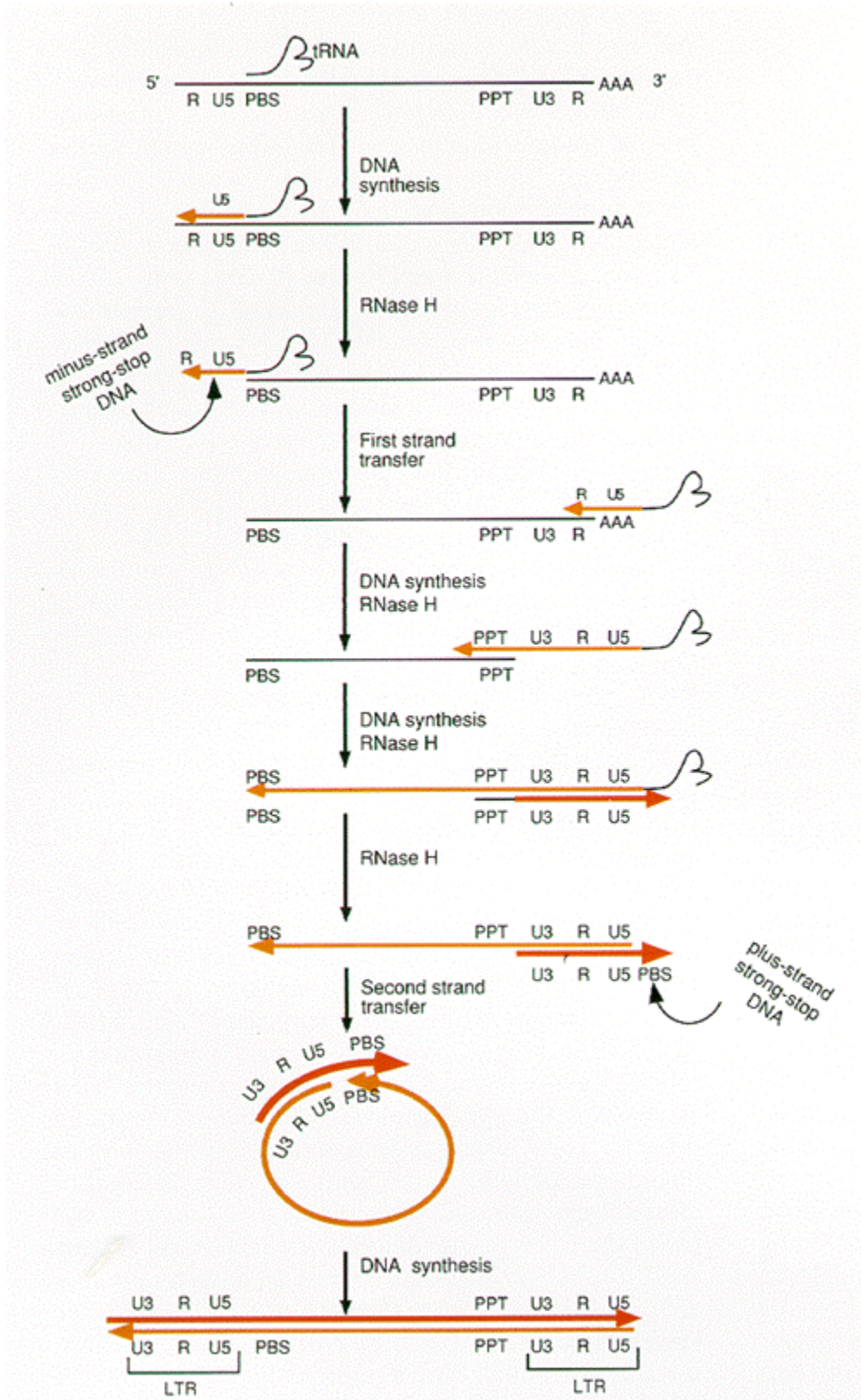


Figure 3. Reverse Transcription. From Retroviruses, Coffin et al.

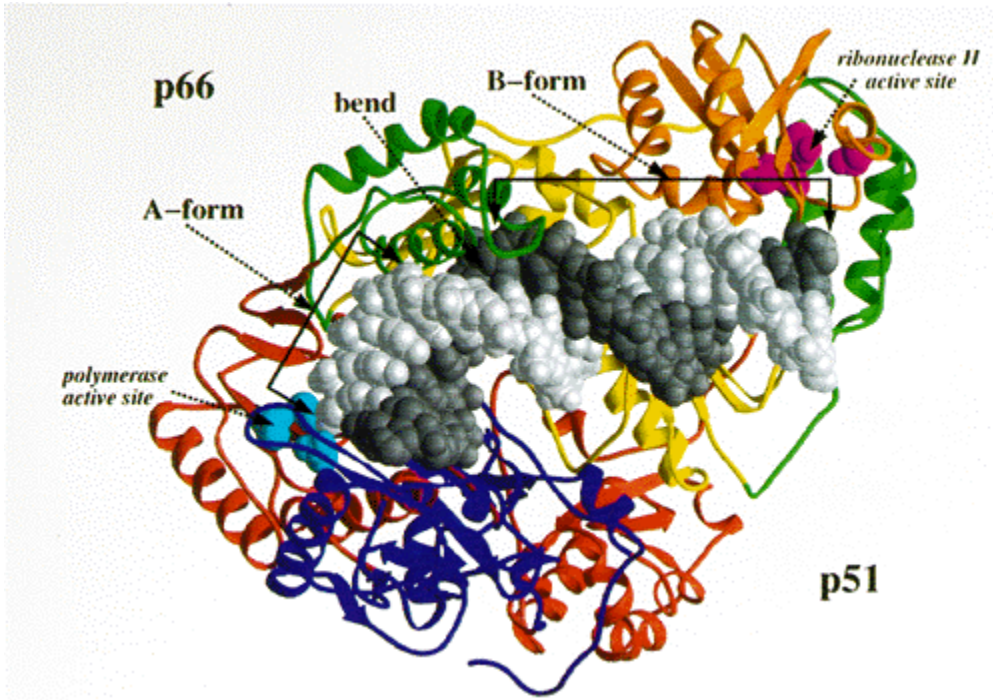


Figure 4. HIV-1 Reverse Transcriptase. From Retroviruses, Coffin et al.

CHAPTER TWO

This page intended to be left blank.

**Utilization of Nonviral Sequences for Minus-Strand DNA Transfer and Gene
Reconstitution during Retroviral Replication**

Sara Rasmussen Cheslock,^{1,2} Jeffrey A. Anderson,¹ Carey K. Hwang,^{1,2} Vinay K.
Pathak,² and Wei-Shau Hu^{2*}

Department of Microbiology and Immunology, West Virginia University, Morgantown,
West Virginia, 26506,¹ and HIV Drug Resistance Program, National Cancer Institute,
FCRDC, Frederick, Maryland 21702²

* Corresponding author. Mailing address: Rm. 336, Building 535, HIV Drug Resistance
Program, National Cancer Institute, FCRDC, Frederick, MD 21702. Phone: (301) 846-
1250. Fax: (301) 846-6013. E-mail: whu@mail.ncifcrf.gov.

Running title: Minus-strand DNA transfer and gene reconstitution.

ABSTRACT

Minus-strand DNA transfer, an essential step in retroviral reverse transcription, is mediated by the two repeat (R) regions in the viral genome. It is unclear whether R simply serves as a homologous sequence to mediate the strand transfer or contains specific sequences to promote strand transfer. To test the hypothesis that the molecular mechanism by which R mediates strand transfer is based on homology rather than specific sequences, we examined whether nonviral sequences can be used to facilitate minus-strand DNA transfer. The green fluorescent protein gene (GFP) was divided into two fragments, GF and FP, containing the 5' and 3' portions of GFP, respectively. An overlap of 85 bp of homology, designated F, existed between GF and FP. A murine leukemia virus-based vector was modified by insertion of the FP and GF fragments into the 5' and 3' long terminal repeats, respectively. Viral RNA generated from this vector contained FP-R at the 5' end and GF-R near the 3' end. Utilization of the F fragment to mediate minus-strand DNA transfer was expected to reconstitute GFP during reverse transcription. Flow cytometry analyses demonstrated that GFP was expressed in 73 to 92% of the infected cells, depending on the structure of the viral construct. This indicated that GFP was reconstituted at a high frequency; molecular characterization further confirmed the accurate reconstitution of GFP. These data indicated that nonviral sequences could be used to efficiently mediate minus-strand DNA transfer. Therefore, placement and homology, not specific sequence context, are the important elements in R for minus-strand DNA transfer. In addition, these experiments demonstrate that minus-strand DNA transfer can be used to efficiently reconstitute genes for gene therapy applications.

INTRODUCTION

All retroviruses replicate their genome using an RNA form to generate a DNA form in a process called reverse transcription (43). The viral RNA is characterized by short repeat (R) regions at the 5' and 3' ends (7, 15). The R region at the 5' end is immediately followed by a unique 5' sequence named U5. Because the viral RNA is the mRNA or the plus strand, the first strand of DNA synthesized is complementary to the viral RNA and is referred to as the minus-strand DNA. Viral DNA synthesis initiates near the 5' end of the viral RNA, using a tRNA primer that binds to the primer-binding site (PBS) in the viral RNA (7). Reverse transcriptase (RT) copies R and U5 and quickly reaches the 5' end of the RNA template. This short stretch of DNA that contains R and U5 is referred to as minus-strand strong-stop DNA. It is thought that the RNase H activity of RT degrades the RNA template in the RNA/DNA hybrid and exposes the strong-stop DNA. The newly synthesized R in the viral DNA is complementary to the R near the 3' end of the viral RNA. Presumably, the complementarity facilitates alignment and hybridization of the two nucleic acids and allows RT to continue DNA synthesis, using sequences near the 3' end of the viral RNA as a template. This switching of the RT-complex from the 5' end to near the 3' end of the viral RNA, known as minus-strand DNA transfer, is an essential step in reverse transcription (7, 15). Minus-strand DNA transfer is primarily mediated by the strong-stop DNA (7, 15); however, it has been observed that DNA containing U5 and only a portion of R, referred to as minus-strand weak-stop DNA, can also mediate minus-strand DNA transfer, although at a lower frequency (28, 29, 37, 45).

Minus-strand DNA transfer is accomplished through complex interactions between the viral proteins and nucleic acids. At least two viral proteins, RT and nucleocapsid (NC), play important roles in this transfer. The RNase H activity of RT is essential for minus-strand DNA transfer (5, 30, 35). Through various in vitro assays, NC was also clearly demonstrated to facilitate the efficiency of minus-strand DNA transfer (2, 9, 13, 17, 21, 26, 34, 36, 39, 40, 44, 46).

Less is known about the requirement of *cis*-acting sequences for minus-strand DNA transfer. The R regions vary significantly in length and sequence among different viruses. The length of R can vary up to 25-fold (from the shortest R, 12 nt, in mouse mammary tumor virus to the longest R, 249 nt, in human T cell leukemia virus type II) (4, 7). Furthermore, it was shown that in some viruses, such as murine leukemia virus (MLV), spleen necrosis virus, and human immunodeficiency virus type 1 (HIV-1), only a portion of R is needed to mediate minus-strand DNA transfer, because weak-stop DNA transfer events were observed (27-29, 37, 45). In a recent study using a viral vector-cell culture system, we defined the relationship between the length of homology and the efficiency of minus-strand DNA transfer (8). We found that 12 nt of homology is sufficient to mediate efficient minus-strand DNA transfer.

Although homology length can play a role in the efficiency of minus-strand transfer, it was not clear whether there is a requirement for specific sequences in the R region for this transfer. The R regions of various retroviruses do not have apparent conserved motifs. However, in an in vitro assay using purified proteins and RNA, it was demonstrated that R sequences from MLV, Rous sarcoma virus, and HIV-1 could mediate minus-strand transfer, whereas a nonviral sequence failed to mediate strand

transfer (3). This finding led to the hypothesis that the sequence context in the viral R region promotes strand transfer and viral sequences are required for efficient minus-strand DNA transfer.

Recent data from our laboratory indicated that efficient minus-strand DNA transfer can be mediated by short stretches of homology (8, 45). Therefore, we hypothesized that homology in the R region, rather than sequence context, is key to promoting this transfer. To test this hypothesis, we examined whether nonviral sequences can mediate efficient minus-strand DNA transfer.

MATERIALS AND METHODS

Construction of vectors. Plasmids pSR2-2GFP, pSR5-FP-GF, pSR6-2wtLTR, and pMS2-FP-GF-no3R were derived from pAR2 (45), an MLV-based vector that contains the hygromycin phosphotransferase B gene (*hygro*) (16). For the nomenclature used here, plasmid names begin with p, whereas the names of viruses derived from the plasmids do not (e.g., pSR5-FP-GF refers to the plasmid construct; SR5-FP-GF refers to the virus derived from this plasmid).

pAR2 was digested to completion with *AatII* and self-ligated to generate pTR1, a plasmid that contained a portion of *hygro* and the downstream long terminal repeat (LTR). The green fluorescent protein gene (GFP) (6) from pGreen-Lantern-1 (Gibco) was amplified by PCR using primers GFp1-*AscI* and PFG2p-csA. Sequences of the primers are shown in Table 1. The resulting DNA was digested with *AscI* and inserted into the *AscI* site between U3 and R in the upstream LTR of pAR2 to generate pCM1. GFP was amplified by PCR using primers GFp3-*EheI* and PFG4p-ehEI; the amplified

product was digested with *EheI* and inserted into the *EheI* site between U3 and R in the pTR1 LTR to generate pCM2. pCM1 and pCM2 were digested with *ScaI*, and the pCM1 DNA fragment containing the upstream LTR with GFP was ligated to the pCM2 DNA fragment containing the downstream LTR. The resulting plasmid, pSR1, contained *hygro*, and both LTRs had one copy of GFP. pSR1 was digested with *BstEII* and *ClaI* to excise *hygro*, which was replaced with the SV40 promoter-*hygro* fragment from pMSM2 to generate pSR2-2GFP.

The portion of GFP containing the 3' 462-bp fragment, termed FP, was amplified by PCR using primers FPAscp1 and PFG2-*pcsA* (Table 1). The PCR product was digested with *AscI* and inserted into the *AscI* site in the 5' LTR of pAR2 to generate pTR4. The portion of GFP containing the 5' 350-bp fragment, termed GF, was amplified using primers FG-*ehE*-4p and GFp3-*EheI* (Table 1). The PCR product was digested with *EheI* and inserted into the *EheI* site of pTR1 to generate pTR3. The pTR4 DNA fragment containing the upstream LTR with FP was isolated and the DNA fragment containing the downstream LTR with GF was isolated from pTR3. These two DNA fragments were ligated to form pTR5. pTR5 was digested with *BstEII* and *ClaI* to excise *hygro*, which was replaced with the 1.7-kb DNA fragment containing SV40 promoter-*hygro* to generate pSR3. DNA sequencing of pSR3 revealed the presence of an inactivating mutation in GF (data not shown). To remove this mutation from the plasmid, pCM2 was digested with *ClaI* and *MscI* to excise the 3' *hygro*, U3, and GF fragments, which were then ligated into the eluted 6.1-kb backbone of pSR3 that was generated by digestion with *ClaI* and partial digestion with *MscI*. In the resulting plasmid, pSR5-FP-

GF, the GF region contained the 5' 350-bp of GFP and the FP region contained the 3' 462-bp of GFP with a 85-bp F region shared by both GF and FP fragments.

pSR6-2wtLTR was generated by replacing the 1.4-kb *ClaI*-*BstEII* fragment of pAR2 between the Ψ packaging signal and the 3' end of *hygro* with the 1.7-kb *BstEII*-*ClaI* fragment containing SV40-*hygro* from pMSM2.

pJD220SVhy (14) was digested with *ClaI* and *BstZ17I* to isolate a 409-bp DNA fragment containing the SV40 termination signal. pCM2 was digested with *BstBI* and *BstZ17I* to delete the 3' 351-bp fragment containing a portion of GFP, R, and U5, then was treated with the Klenow fragment of *Escherichia coli* DNA polymerase I and ligated with the 409-bp *ClaI*-*BstZ17I* DNA fragment from JD220SVhy. The resulting plasmid, pMS1, was digested with *ScaI*, and the fragment containing the 3' LTR was isolated and ligated to *ScaI*-digested pSR5-FP-GF to generate pMS2-FP-GF-no3R. pMP1 was derived from pWH390 (10) by the insertion of GFP upstream of the internal ribosomal entry site (IRES) from encephalomyocarditis virus (22, 23).

Standard cloning techniques were used to construct all of the vectors (31). Plasmid structures were analyzed by restriction enzyme mapping. All PCR-amplified DNA fragments that were cloned into plasmids were further analyzed by DNA sequencing to detect inadvertent mutations generated during the PCR procedures.

Cell culture, DNA transfection, and virus infection. PG13 cells (American Type Culture Collection) were derived from NIH3T3 cells expressing MLV *gag-pol* and gibbon ape leukemia virus (GaLV) *env* (32). D17 (American Type Culture Collection) is a dog osteosarcoma cell line permissive for MLV infection(38). PG13 and D17 cells were maintained at 37°C in Dulbecco's modified Eagle's medium supplemented with

penicillin (50 U/ml; Gibco), streptomycin (50 µg/ml; Gibco), and bovine calf serum (10% for PG13 and 6% for D17). G418, a neomycin analog, was used for selection at a final concentration of 600 µg/ml for PG13 cells and 400 µg/ml for D17 cells. Hygromycin was used at a final concentration of 300 µg/ml in PG13 cells and 120 µg/ml in D17 cells.

Transfections were performed using the calcium phosphate precipitation method as previously described (31) or Transfast™ transfection reagents from Promega as recommended by the manufacturer. PG13 cells were plated at a density of 1×10^5 cells per 60-mm-diameter dish; 5 or 10 µg of vector DNA was used per dish for the Transfast™ or calcium phosphate transfection, respectively. Transfected cells were placed on the appropriate drug selection; drug-resistant colonies were pooled, expanded, and plated at a density of 5×10^6 cells per 100-mm-diameter dish. Virus was harvested from each transfected cell pool 48 h later and centrifuged at $3000 \times g$ for 10 min to remove cellular debris. Serial dilutions of the supernatants were used to infect D17 cells that were plated at 2×10^5 cells per 60-mm dish. Infected D17 cells were then placed on appropriate drug selection. Viral titers were calculated based on the number of drug-resistant colonies and standardized to RT activities.

RT assay. A portion of the virus harvested from transfected PG13 cells was subjected to RT assays as previously described (18). Briefly, harvested virus was centrifuged in a SW 28 rotor (Beckman) or a Surespin 630 rotor (Sorvall) at 25,000 rpm for 90 min. Viral pellets were resuspended in serum-free medium and stored at -80°C . Exogenous RT activities were determined by incubating 10 µl of virus with 50 µg/ml of 20-mer oligo T (Integrated DNA Technologies), 100 µg/ml of poly(A) (Pharmacia), 60 mM NaCl, 50 mM Tris (pH 8.0), 1 U/µl RNase inhibitor, 10 mM dithiothreitol, 0.6 mM

MnCl₂, 80 μM dTTP, 0.5% IGE Pal (Sigma), and 10 μCi of [³H]dTTP (72 Ci/mmol, ICN). The samples were incubated at 37°C for 90 min. The reaction mixtures were precipitated with 10% trichloroacetic acid (Sigma) and filtered through 0.45-μm-pore-diameter Metrical membranes (Gelman Sciences, Inc.); the amount of [³H] incorporated was determined using a scintillation counter.

Detection of GFP expression by flow cytometry and fluorescence microscopy.

The number of cells in the transfected pools that expressed GFP was measured using flow cytometry (FACScan, Becton Dickinson); results were analyzed using CellQuest software (Becton Dickinson). GFP expression in infected, drug-resistant D17 cells was analyzed by two methods. Drug-resistant cell colonies from plates containing 10⁰, 10⁻¹, and 10⁻² viral dilutions were separately pooled for flow cytometry analyses. Drug-resistant colonies from 10⁻³, 10⁻⁴, and 10⁻⁵ viral dilution plates were analyzed by fluorescence microscopy (Axiovert inverted fluorescence microscope, Zeiss). GFP expression in individual cell clones was also analyzed by flow cytometry and microscopy.

Analyses of proviral structure by PCR and DNA sequencing. Hygromycin-resistant cell clones were isolated and lysed for use as PCR substrates. The upstream LTR of the proviruses was amplified using primers in the U3 (MLVU3) and 5' Ψ regions (ispVLM5') (Table 1). PCR products were analyzed by DNA sequencing with an automated sequencer (PE Biosystem) using one or more of the following primers: GF101, 5UVLM, R/3UM-629, PFG695, and PFG464 (Table 1).

RESULTS

Vectors used to determine the primary sequence requirement for minus-strand DNA transfer. A series of MLV-based retroviral vectors was used to test the hypothesis that nonviral sequences can mediate minus-strand DNA transfer and that this process can reconstitute genes; the structure of these vectors are shown in Fig. 1A. All of these vectors contained the *cis*-acting sequences necessary for retroviral replication, such as the PBS, packaging signal (Ψ^+), polypurine tract, and attachment sites. In addition, most of *gag*- and the entire *pol*- and *env*-coding regions were deleted from these vectors. pSR2-2GFP, pSR5-FP-GF, pSR6-2wtLTR, and pMS2-FP-GF-no3R each contained an SV40 promoter upstream of *hygro* between the two LTRs. pSR6-2wtLTR contained two unmodified LTRs, whereas pSR2-2GFP had two modified LTRs, each containing the full-length GFP between U3 and R. pSR5-FP-GF also had two modified LTRs; the upstream LTR contained the 3' 462-bp FP fragment of GFP between U3 and R, whereas the downstream LTR contained the 5' 350-bp GF fragment of GFP between U3 and R. FP and GF share a homology stretch of 85 bp (the F region). pMS2-FP-GF-no3R also had two modified LTRs similar to those in pSR5-FP-GF, except that the R and U5 regions in the downstream LTR were replaced by a DNA fragment containing the SV40 termination signal. pMP1 contained two unmodified LTRs with GFP, IRES, and the neomycin phosphotransferase gene (*neo*) (24) between the LTRs.

pSR6-2wtLTR and pMP1 contained unmodified LTRs and were expected to produce RNAs with R at the two ends of the viral sequences that could mediate minus-

strand DNA transfer. The LTRs in pSR5-FP-GF were modified; as a result, the viral RNA produced was expected to contain FP-R-U5 at the 5' end and U3-GF-R at the 3' end of the viral sequences (Fig. 1B). During reverse transcription, DNA synthesis was expected to copy U5, R, and FP to form minus-strand strong-stop DNA (Fig. 1B). We hypothesized that if nonviral sequences can mediate minus-strand DNA strand transfer, then the 85-nt F region in the strong-stop DNA can hybridize to the F region at the 3' end of viral RNA, reverse transcription can continue to copy the G region of the RNA, and GFP can be reconstituted. On the other hand, if nonviral sequences cannot mediate minus-strand DNA transfer, then GFP cannot be reconstituted. It is also known that at a lower frequency (~1-10%), minus-strand DNA synthesis terminates early to form weak-stop DNA (28, 29, 37, 45). These weak-stop DNAs might contain only U5 and R and would be expected to transfer to the 3' end of viral RNA and form viral DNA containing GF between U3 and R. Therefore, minus-strand transfer of these weak-stop DNAs was not expected to reconstitute GFP.

The LTRs in pMS2-FP-GF-no3R were modified in a manner similar to those in pSR5-FP-GF except that R and U5 were deleted from the 3' LTR. The MS2-FP-GF-no3R viral RNA was also expected to have FP-R at the 5' end and U3-GF at the 3' end of the viral sequences; however, the viral RNA did not contain R at the 3' end. If the F region could mediate minus-strand DNA transfer, then the strong-stop DNA would transfer to the 3' end of the viral RNA sequence. However, unlike SR5-FP-GF, the weak-stop DNA containing only U5 and R was not complementary to the sequences at the 3' end of the viral RNA and should not be able to transfer efficiently. Therefore, if the F region could not mediate efficient minus-strand DNA transfer, it was expected that

MS2-FP-GF-no3R would not be able to replicate well and would have a severely reduced viral titer.

Each LTR in pSR2-2GFP contained a copy of GFP between U3 and R. The RNA produced from these constructs was expected to contain GFP-R-U5 at the 5' end and U3-GFP-R at the 3' end of the viral sequences. Regardless of whether the virus used the R region or the nonviral sequences to mediate minus-strand DNA transfers, the resulting viral DNA would have LTRs with GFP between the U3 and R sequences.

Experimental protocol. The outline of the experimental protocol is shown in Fig. 2. These vectors were separately transfected into PG13 helper cells that expressed MLV *gag-pol* and GaLV *env*. Transfected cells were subjected to appropriate drug selection, and resistant cell colonies were pooled. All of the pools contained at least 250 colonies. Viruses were harvested from these pools; for each sample, a portion of the virus was used to measure the RT activity and another portion was serially diluted and used to infect D17 cells. After the infected D17 cells were subjected to appropriate drug selection, the numbers of drug-resistant colonies were determined and were used to calculate the viral titers generated by these vectors. Within each set of experiments, viral titers were standardized to the RT activities. GFP expression from these drug-resistant colonies was analyzed by flow cytometry or fluorescence microscopy. Infected, drug-resistant D17 cell clones were isolated; GFP expression of these cell clones was determined by flow cytometry and fluorescence microscopy. To examine the molecular nature of the minus-strand DNA transfer, a portion of the proviruses containing minus-strand DNA transfer junctions was amplified from the infected cell clones by PCR and characterized by DNA sequencing.

In this system, viral titers and minus-strand DNA transfers were examined in a single viral replication cycle, from the viruses produced in the PG13 cells to the proviruses generated in the D17 cells. Viruses produced from the murine-derived PG13 cells contained GaLV Env, which could not efficiently infect murine cells (32). Therefore, reinfection had little opportunity to occur in PG13 cells. D17 target cells did not express *gag-pol* and *env* needed to produce virions containing these vectors to go through another round of viral replication. Therefore, only one round of retroviral replication was allowed in this system.

Viral titers after one round of retroviral replication. The titers for vectors MP1, SR2-2GFP, SR5-FP-GF, SR6-2wtLTR, and MS2-FP-GF-no3R are listed in Table 2. These titers were standardized to the RT activities. The SR6-2wtLTR titer varied between 8.3 and 130×10^4 CFU/ml. The SR5-FP-GF titer varied between 1.0 and 24×10^4 CFU/ml. The average difference between the SR6-2wtLTR titers and SR5-FP-GF titers was 6.8-fold. The reductions in viral titers could have been caused by inserting sequences in the R region, similar to the 6- to 8-fold decreases in titers previously observed by another laboratory (1). Alternatively, it was possible that the F region could not mediate the minus-strand DNA transfer and caused the decrease in the viral titer. This would suggest that most of the strong-stop DNA could not perform strand transfer and did not produce viral DNA capable of integrating into the host genome. Most of the viral titer would be generated from viral DNA produced by the transfer of minus-strand weak-stop DNA. Because the weak-stop DNA was expected to be generated at a frequency of ~1-10% (28, 29, 37, 45), we could not rule out the possibility that the 6.8-

fold decrease in viral titers was caused by the inability of the strong-stop DNA to perform minus-strand DNA transfer.

The SR2-2GFP and MS2-FP-GF-no3R titers varied from 4.1 to 14×10^4 CFU/ml and 1.3 to 21×10^4 CFU/ml, respectively. There was no significant difference among the SR2-2GFP, SR5-FP-GF, and MS2-FP-GF-no3R titers (Two sample T test; $P > 0.8$ for SR5-FP-GF and MS2-FP-GF-no3R, $P > 0.9$ for SR5-FP-GF and SR2-2GFP). MS2-FP-GF-no3R RNA only contained the 5' R and thus could not use R to mediate minus-strand DNA transfer. Therefore, in order for MS2-FP-GF-no3R to generate viral titers similar to those of SR2-2GFP and SR5-FP-GF, the F region had to be able to mediate minus-strand DNA transfer. The proviruses generated by F region-mediated transfer would have reconstituted GFP, which could be confirmed by GFP expression or structural analyses.

GF and FP fragments cannot confer positive GFP expression. GFP expression in all of the transfected PG13 helper cell pools was examined by flow cytometry analyses. As expected, pMP1- or pSR2-2GFP-transfected, drug-resistant cell pools contained significant numbers of GFP-expressing cells, generally between 45 and 70%. As expected, cell pools transfected with pSR6-2wtLTR did not contain a significant percentage of fluorescent cells (<2%), since these plasmids lacked GFP. Cell pools transfected with pSR5-FP-GF or pMS2-FP-GF-no3R also did not contain a significant percentage of fluorescent cells (<2%), indicating that neither FP nor GF could express functional fluorescent proteins.

GFP expression of infected cells. Flow cytometry analyses were also performed on D17 cells infected with virus produced by different transfected cell pools. Drug-

resistant D17 cells were pooled and analyzed by flow cytometry; a representative set of flow cytometry analyses are shown in Fig. 3A and the data from 5 independent experiments are summarized in Fig. 3B. MP1 and SR2-2GFP both contained intact GFP; most of the D17 cells infected with these viruses were positive for GFP expression, with an average of $79.6\% \pm 2.7\%$ (standard error [SE]) and $97.7\% \pm 0.2\%$ (SE), respectively. In contrast, very few SR6-2wtLTR-infected D17 cells ($0.3\% \pm 0.1\%$ [SE]) were positive for GFP expression since this virus did not contain GFP. In all experiments, a high proportion of cells infected with SR5-FP-GF were positive in GFP expression, ranging from 65% to 78.1% with an average of $72.9\% \pm 2.4\%$ (SE) (Fig. 3). Most of the MS2-FP-GF-no3R -infected cells were positive for GFP expression (88.4 to 95.2%, with an average of $91.7\% \pm 1.2\%$ [SE]). These experiments demonstrated that, most of the time, GFP was reconstituted during reverse transcription of SR5-FP-GF and MS2-FP-GF-no3R RNA. Therefore, the F region was used to mediate minus-strand DNA transfer during reverse transcription of these proviruses.

In addition to the flow cytometry analyses, GFP expression of the cell colonies was also examined using fluorescent microscopy. As expected, GFP expression was found in 0 of 369 (<0.1%) SR6-2wtLTR -infected colonies. GFP expression was found in 667 of 828 (80.6%) MP1-infected colonies, 305 of 341 (89.4%) SR2-2GFP-infected colonies, 629 of 856 (73.5%) SR5-FP-GF-infected colonies, and 1002 of 1092 (91.8%) MS2-FP-GF-no3R-infected colonies. The frequencies of GFP expression detected by microscopy were similar to those observed by flow cytometry analyses.

Molecular characterization of minus-strand DNA transfer in SR5-FP-GF and MS2-FP-GF-no3R. To directly analyze the molecular nature of minus-strand DNA

transfer, we isolated drug-resistant cell clones. From these clones, a portion of the proviral genome containing the upstream LTR was amplified and sequenced to characterize the molecular nature of the transfer events (Fig. 4).

Partial proviral structures from 13 SR5-FP-GF-infected cell clones were characterized; of the 13 clones, 4 were positive and 9 were negative for GFP expression as measured by fluorescence microscopy and flow cytometry. All of the 4 proviruses that expressed GFP contained the expected structures and had U3-reconstituted GFP-R-U5 in the LTRs. None of the 9 proviruses that were negative for GFP expression had the reconstituted GFP in their LTRs. Six of the 9 proviruses had U3-GF-R-U5 in the LTRs, and the other 3 contained U3-FP-R-U5 in their LTRs.

Viruses with U3-GF-R-U5 appeared to be the predicted structures for proviruses generated by weak-stop DNA mediated minus-strand transfers (Fig. 1B). However, upon further analyses, these proviruses were probably generated by a different mechanism. These vectors were derived from the pLN series plasmids (33). With the exception of pMS2-FP-GF-no3R, all of these vectors had an additional PBS directly downstream of the 3' LTR (Fig. 5A), as well as sequence variation between the 5' and 3' R regions (Fig. 5B). In these vectors, there were two possible mechanisms to generate viral DNA with LTRs containing U3-GF-R-U5: weak-stop minus-strand DNA transfer, and readthrough RNA transcripts with DNA synthesis initiated from the downstream PBS. Termination of retroviral RNA transcripts is known to be relatively inefficient; 15% of the total transcripts are read through the termination signal and contain sequences downstream of LTR (19). In addition, it has been shown that these readthrough transcripts can be efficiently packaged into viral particles (20, 41, 42). In SR5-FP-GF, these readthrough

transcripts would contain the entire downstream LTR and PBS (Fig. 5A). If DNA synthesis initiated from the downstream PBS, RT would copy the entire downstream LTR and proceed to copy the rest of the viral RNA. In this situation, reverse transcription would entirely bypass the step of minus-strand DNA transfer. The resulting viral DNA would have U3-GF-R-U5 in the LTRs (Fig. 5A), similar to the viral DNA generated from the weak-stop DNA transfer. The difference between these two viral DNAs was the origin of R in the LTR. With the weak-stop DNA transfer, a portion of R would be from the upstream R and another portion would be from the downstream R (Fig. 5A). With readthrough RNA transcripts and downstream PBS initiation, however, the entire R sequence would be derived from the downstream LTR. When we compared the R sequences of the 6 proviruses, all of them were entirely derived from the downstream R, indicating that these proviruses were generated by readthrough RNA transcripts and downstream PBS initiation rather than by weak-stop minus-strand DNA transfer.

Three of the 9 proviruses had U3-FP-R-U5 in the LTRs; these were likely to be generated from RNA transcripts initiating from the promoter upstream of U3, termed read-in transcripts (Fig. 6). We hypothesized that in the virus-producing cells, some transfected pSR5-FP-GF might integrate close to the promoters. Read-in transcripts would therefore contain the upstream U3 sequences along with FP, R, and U5. During reverse transcription, minus-strand DNA synthesis would copy U5, R, FP, and a portion of U3. The complementarity between the newly synthesized U3 DNA and U3 near the 3' end of RNA could be used to mediate minus-strand DNA transfer. The resulting provirus would have two LTRs containing U3, FP, R, and U5.

The molecular nature of minus-strand DNA transfer was also examined in 18 MS2-FP-GF-no3R-infected cell clones. Of these clones, 9 were positive and 9 were negative for GFP expression by both fluorescence microscopy and flow cytometry analyses. Similar to the SR5-FP-GF-infected clones, all of the 9 MS2-FP-GF-no3R proviruses that expressed GFP had precisely reconstituted GFP in the LTRs. Of the 9 MS2-FP-GF-no3R proviruses that did not express GFP, two had mutations in GFP: One had a reconstituted GFP that contained a G-to-C substitution mutation in the G region that converted a glycine to an arginine and thereby inactivated GFP; the other contained a mutant GFP with a 120-bp deletion in the P region. The structure of the LTR from the latter provirus contained U3, G, F, 52 bp of 5' P, a 120-bp deletion within P, 205 bp of 3' P, R, and U5. Because the junction between F and P as well as the junction between P and R remained intact, the deletion was probably independent of minus-strand DNA transfer. The other 7 MS2-FP-GF-no3R proviruses had a U3-FP-R-U5 structure in the LTR similar to those in SR5-FP-GF, which were presumably generated from read-in RNA transcripts.

Because the 3' R, U5, and PBS were deleted in pMS2-FP-GF-no3R, proviruses with U3-GF-R-U5 structure in the LTRs were not observed. This impacted the frequency of the GFP-expressing cells. Approximately 8% and 27% of the cells infected with MS2-FP-GF-no3R and SR5-FP-GF did not express GFP, respectively (Fig. 3). The 19% difference between the cells infected with MS2-FP-GF-no3R and SR5-FP-GF is close to the percentage of SR5-FP-GF proviruses containing the U3-GF-R-U5 structure in LTRs. Thus, the increase in the frequency of gene reconstitution in MS2-FP-GF-no3R was

directly correlated to the lack of readthrough transcription and initiation from the downstream PBS which generated proviruses with U3-GF-R-U5 LTRs.

Efficient minus-strand DNA transfers mediated by nonviral sequences.

Analyses of GFP expression in cell populations infected with SR5-FP-GF and MS2-FP-GF-no3R indicated that GFP was reconstituted at approximately 72.9% and 91.7% efficiencies, respectively. Molecular characterization of 13 GFP-expressing proviruses generated from SR5-FP-GF or MS2-FP-GF-no3R demonstrated that these proviruses contained precisely reconstituted GFP, indicating that the F region was used to mediate minus-strand DNA transfer in these viruses. In one round of replication, MS2-FP-GF-no3R generated titers similar to SR5-FP-GF and SR2-2GFP, which could use either the R or F region to mediate minus-strand DNA transfer. Taken together, these data established that the F region could be used to mediate minus-strand DNA transfer in an efficient manner. In addition, these data demonstrated that minus-strand DNA transfer could be used as an effective means to reconstitute genes during virus replication.

DISCUSSION

In this study, we intended to define the requirement of sequence context of the R region in mediating minus-strand DNA transfers. The current model postulates that the role of the R regions is to provide complementarity between the newly synthesized minus-strand DNA and the 3' R of the viral RNA to align the reverse transcription complex for accurate switching of the RNA template. This view suggests that the two R regions allow the hybridization of the nascent DNA and the 3' RNA. In this model, it is likely that R is required to exceed a certain length to allow precise and efficient

DNA/RNA alignment, but it is unlikely that R must contain specific sequences for the hybridization of nascent DNA and viral RNA. This prediction, however, was contradicted by an in vitro study suggesting that nonviral sequences could not mediate minus-strand DNA transfer (3).

Using a series of MLV vectors with modified LTRs, we demonstrated that a nonviral sequence, a portion of GFP, can efficiently mediate minus-strand DNA transfer. This finding is in contrast with the in vitro study in which the nonviral sequence was unable to mediate minus-strand DNA transfer. It is possible that this difference reflects the experimental systems employed in the two studies. In the in vitro system, strand transfer depended solely on the interactions between RT, NC, and nucleic acids. In our system, viruses containing the vector RNA were used to infect target cells; most of the DNA synthesis, including the strand transfer steps, was conducted within the reverse transcription complex in the cells. Many factors present in the in vivo but not in the in vitro system, such as the configurations of the RNAs and reverse transcription complexes, were very likely to be important in minus-strand DNA transfer. Therefore, the contrast of the two studies emphasizes the important roles of the elements missing in the in vitro assays.

Readthrough RNA transcripts and downstream PBS initiation. During the analyses of SR5-FP-GF-generated proviruses, we observed LTRs with U3-GF-R-U5 structures. Despite the resemblance of the predicted structures, these proviruses were unlikely to be generated by weak-stop minus-strand DNA transfer. The two R regions in the vectors we used contained sequence variation scattered throughout the length of the two R regions. During minus-strand DNA synthesis, the first difference in the two R

sequences that RT would encounter was at nt 6 at the 3' end of R (T for upstream R and G for downstream R). In order for a provirus to contain R regions with all of the markers from the downstream R, DNA synthesis would have to stop before nt 6 of the upstream R and use less than 6-nt complementarity between nascent DNA and 3' RNA to mediate minus-strand DNA transfer. However, we have observed in other studies that a 6-nt homology is not sufficient to mediate accurate and efficient minus-strand DNA transfer (8). Therefore, these proviruses were likely to be generated by other mechanisms.

We propose that these proviruses were generated from readthrough RNA transcription with DNA synthesis initiating from the downstream PBS. We have previously observed efficient initiation of DNA synthesis from the downstream PBS (V. K. Pathak, P. D. Yin, R. J. Teufel II, and W.-S. Hu, unpublished data). Approximately 27% of the SR5-FP-GF-infected cells did not express GFP (Fig. 3). Of the 9 proviruses analyzed, 6 had the U3-GF-R-U5 structure. Therefore, approximately 18% of the proviruses had the U3-GF-R-U5 structure, close to the observed 15% efficiency of readthrough RNA transcription (19). This result also suggested that DNA synthesis of these readthrough RNA transcripts mostly initiated from the downstream PBS and bypassed the minus-strand DNA transfer event. This was possible because the downstream PBS in these vectors contained a large portion of the sequences proposed to form a secondary structure that was important for efficient initiation of DNA synthesis. In addition, the GF fragment is 350 bp in length, whereas FP is 462 bp. If DNA synthesis initiated from both upstream and downstream PBS simultaneously, there would be a race for the 3' template usage. By the time DNA synthesis that was initiated from the upstream PBS reached the end of FP, DNA synthesis that was initiated from the

downstream PBS would have copied through GF and a portion of U3. Consequently, the RNA templates at the 3' end of the viral sequences would be degraded, including the F region of the RNA that could be used for minus-strand DNA transfer. Therefore, it was likely that the minus-strand DNA that was initiated from the upstream PBS would lose the “race” during reverse transcription, which may account for the proportion of proviruses that contained sequences derived from the downstream R.

Alternatively, it is possible that in readthrough RNA transcripts, minus-strand DNA synthesis initiated from the upstream PBS, copied a portion of U5, and used the U5 sequences to transfer to near the 3' end of the viral RNA. RT could then copy the 3' R, GF, and U3 sequences. In order for this mechanism to account for most of the proviruses containing U3-GF-R-U5 in the LTR, minus-strand DNA transfer had to occur at a high frequency during the copying of the 72-nt U5 region. Because minus-strand DNA transfer mediated by weak-stop DNA is known to occur at a low frequency while copying R, which is 68 nt in length, this mechanism is unlikely to be responsible for generating all of these proviruses.

Read-in RNA transcripts and minus-strand DNA transfer using the homology region including a portion of U3. In this study, we also observed proviruses containing U3-FP-R-U5 LTR structures. We hypothesized that these proviruses were generated from RNA transcripts that were initiated from upstream promoters. In another study from our laboratory, modified MLV vectors lacking the downstream U3 were used to examine minus-strand DNA transfer (8). Some of the proviruses resulting from that study contained portions of U3 sequences, indicating that these proviruses were also the products of read-in RNA transcripts. In addition, RNA species containing the upstream

U3 were identified in cellular RNA samples by RNase protection assays, further confirming our hypothesis.

This observation raised the question of whether U3 sequences are occasionally used in minus-strand DNA transfer during the replication of wild-type viruses. If an active promoter is located upstream of a provirus, RNA transcripts containing the upstream U3 could be easily generated. However, most of these transcripts would terminate at the 5' R; only 15% of the transcripts would read through the 5' R and terminate at the 3' R. Similar to the events observed in our experiments, RT could copy a portion of the upstream U3 and use the U3 sequences to align the nascent DNA and 3' RNA for minus-strand DNA transfer. Because these events depend on both the location of the proviral integration and the RNA transcript extending through 5' R, if U3 were used to mediate minus-strand DNA transfer in wild-type viruses, it would be likely to occur at a low frequency.

Application of minus-strand DNA transfer-mediated gene reconstitution to gene therapy and inducible gene expression systems. Previously, direct repeat deletion has been utilized to reconstitute various genes with a high efficiency; these genes included the drug-resistance gene *neo* and a suicide gene encoding the herpes simplex virus thymidine kinase that has been used for cancer therapy (10-12, 25). This strategy uses high frequency template switching events that occur during reverse transcription to reconstitute genes and delete sequences from the portions of the viral genome that are internal to the LTRs. In this report, we describe gene reconstitution using a different strategy utilizing the obligatory minus-strand DNA transfer step in the reverse transcription process. GFP was reconstituted at approximately 92% efficiency in the

vector pMS2-FP-GF-no3R. This high-frequency reconstitution could be exploited for gene delivery in gene therapy applications or as an inducible gene expression system. In the plasmid vector, a gene of interest or a cassette containing a gene expression unit could be divided into two portions with a small stretch of overlapping sequences in which neither portion could express the gene of interest. During reverse transcription and virus replication, this gene of interest could be reconstituted and expressed in the infected cells. Such an approach could be very useful for the delivery of potentially toxic genes in gene therapy applications in which the toxicity of the gene hampers the production of the viral vector-containing virions. This approach could also be very useful as an inducible gene expression system in which viral replication activates gene expression. Because the reporter gene is inactive prior to virus replication, there would be little background expression in the system. For example, this strategy could be used to detect the presence of replication-competent retroviruses (RCR). Test samples could be applied to target cells transfected with pMS2-FP-GF-no3R; if RCR is present in the sample, then MS2-FP-GF-no3R would be mobilized and the infected cells would express GFP. This could be performed in a one-step assay to directly detect RCR rather than a lengthy co-culture or infection assay.

In summary, we have established that homology, rather than sequence context, is important in the mediation of efficient minus-strand DNA synthesis. During this study, we also demonstrated that genes can be reconstituted during minus-strand DNA transfer efficiently with a short stretch of homology, a strategy that could also be exploited for the design of gene therapy vectors.

ACKNOWLEDGMENTS

We thank Carrie McBee, Terence Rhodes, Michelle Paulson, and Melanie Sal for the construction of some of the plasmids. We also thank Que Dang and Krista Delviks for critical reading of the manuscript, Anne Arthur for expert editorial revisions of the manuscript, Drs. John Coffin, Alan Rein, and Vineet KewalRamani for discussions, intellectual input, and suggestions regarding the manuscript, and VKR for help with the flow cytometry. This work was supported in part by research grants from NIH and ACS and also by the HIV Drug Resistance Program, National Cancer Institute. J.A.A. was fully supported and S.R.C. and C.K.H. were partially supported by the Medical Scientist Training Program at West Virginia University.

FIGURE LEGENDS

FIG. 1. MLV-based vectors and strategy used to study the requirement of sequence context in mediating minus-strand DNA transfer. (A) Structures of the MLV-based vectors. SV, SV40 promoter; *hygro*, hygromycin phosphotransferase B gene; GFP, green fluorescent protein gene; GF, the 5' 350-bp fragment of GFP; FP, the 3' 462-bp fragment of GFP; SV-ter, SV40 termination signal; Ψ^+ , the extended MLV packaging signal; IRES, internal ribosomal entry site; *neo*, neomycin phosphotransferase gene. (B) Strategy used to test the ability of nonviral sequences to mediate minus-strand DNA transfer. The structure of pSR5-FP-GF is illustrated at the top. Either minus-strand DNA transfer would be mediated by strong-stop DNA and the GFP would therefore be reconstituted, or minus-strand DNA transfer would be mediated by weak-stop DNA and the GFP would not be reconstituted.

FIG. 2. Experimental protocol. Virus was harvested from transfected PG13 helper cells and used to infect D17 target cells. Viral titers were determined by the number of drug-resistant colonies and standardized to the RT activities. GFP expression in infected target cells was analyzed by flow cytometry and fluorescence microscopy. Minus-strand DNA transfer junctions were analyzed in infected cell clones.

FIG. 3. Flow cytometry analyses of infected cells. (A) A representative set of flow cytometry analyses using cells infected with MP1, SR6-2wtLTR, SR2-2GFP, SR5-FP-GF, and MS2-FP-GF-no3R. In each plot, the Y axis is the number of events scored, which is interpreted as the number of cells, and the X axis is the intensity of the fluorescence. (B) The proportion of fluorescent cells infected with various vectors from 5 independent sets of infections. Error bars represent the standard error of the average.

FIG. 4. Strategy for PCR amplification of proviral LTRs and DNA sequencing. All abbreviations are the same as in Fig. 1. Zigzag lines, host DNA sequences; large arrows, PCR primers; small horizontal arrows, sequencing primers.

FIG. 5. Two models for the generation of proviruses with U3-GF-R-U5 in the LTRs. (A) Illustration of the two models. All abbreviations are the same as in Fig. 1. Sequences derived from upstream R are shown in gray, whereas sequences from downstream R are shown in white. (B) Sequence comparisons of the upstream and downstream R regions.

FIG. 6. Model for generating proviruses with U3-FP-R-U5 by using read-in RNA transcripts. All abbreviations are the same as in Figs. 1 and 4; pro, upstream promoter.

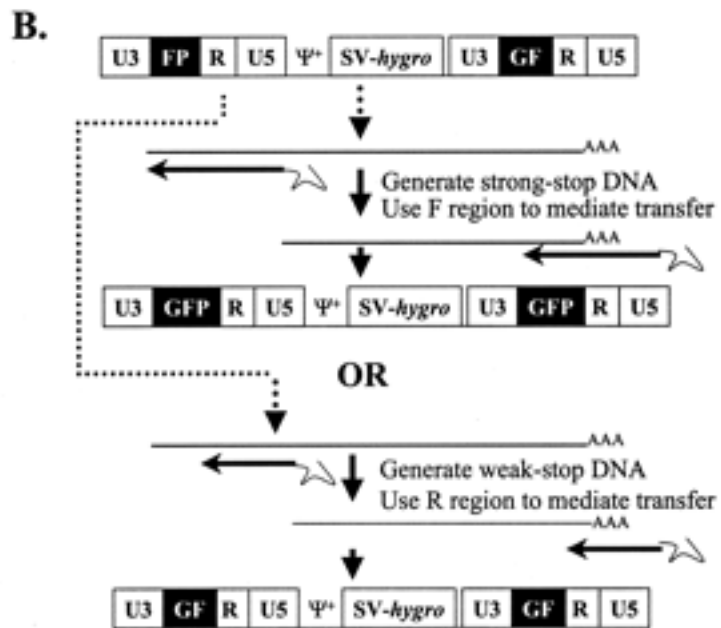
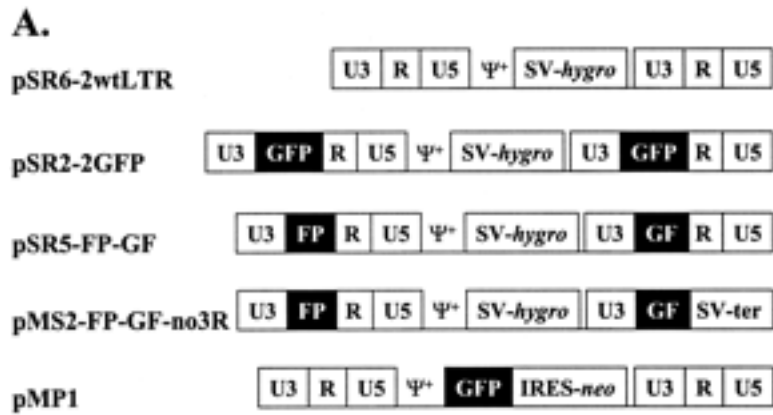


Figure 1.

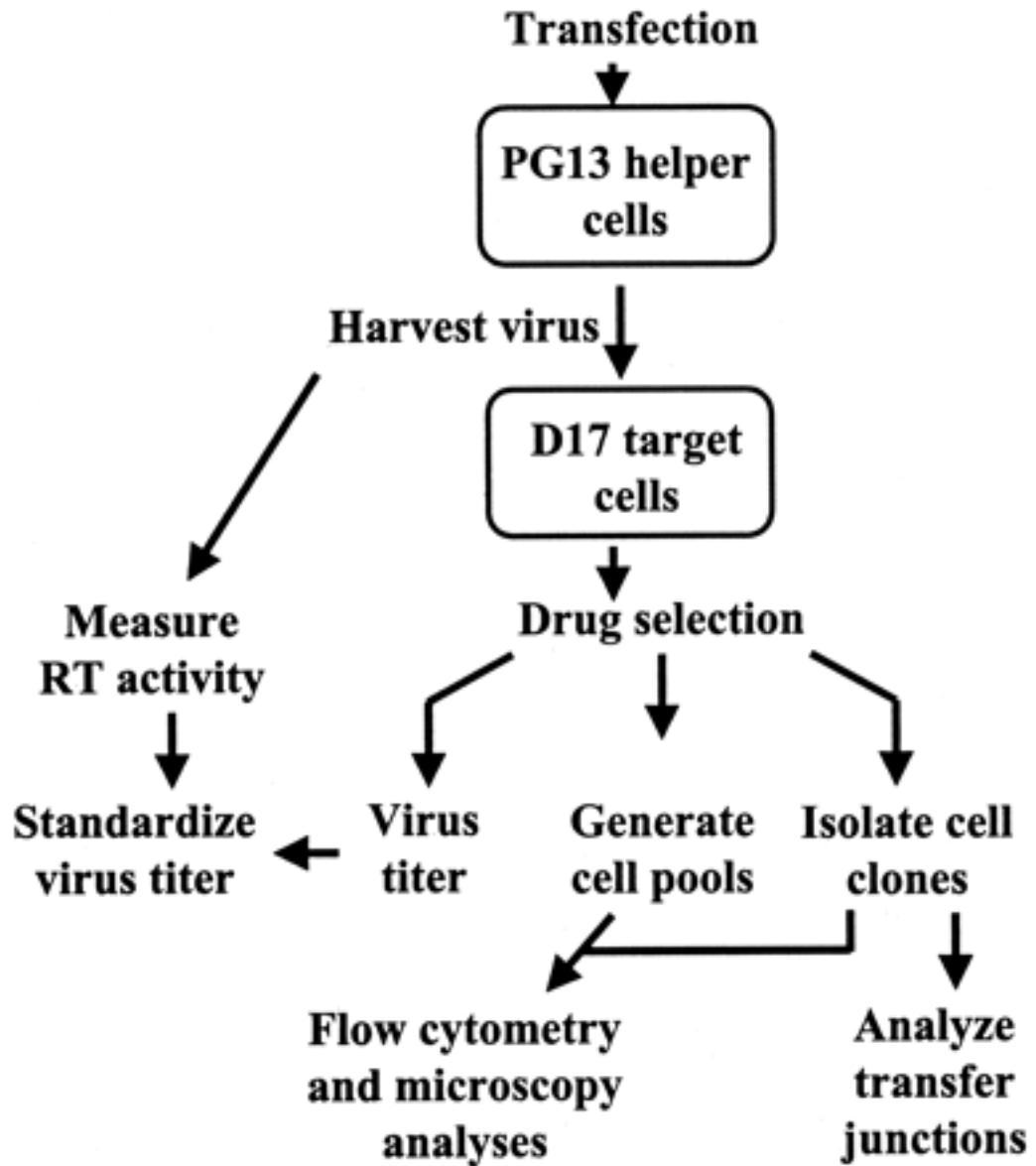


Figure 2.

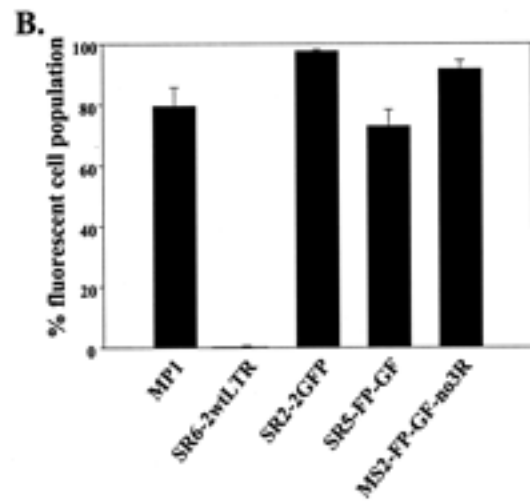
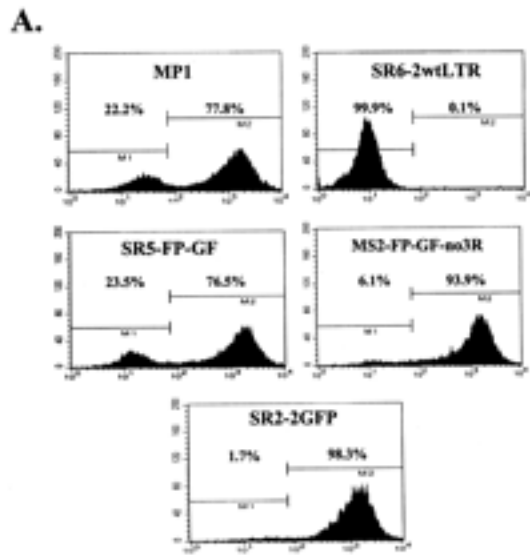


Figure 3.

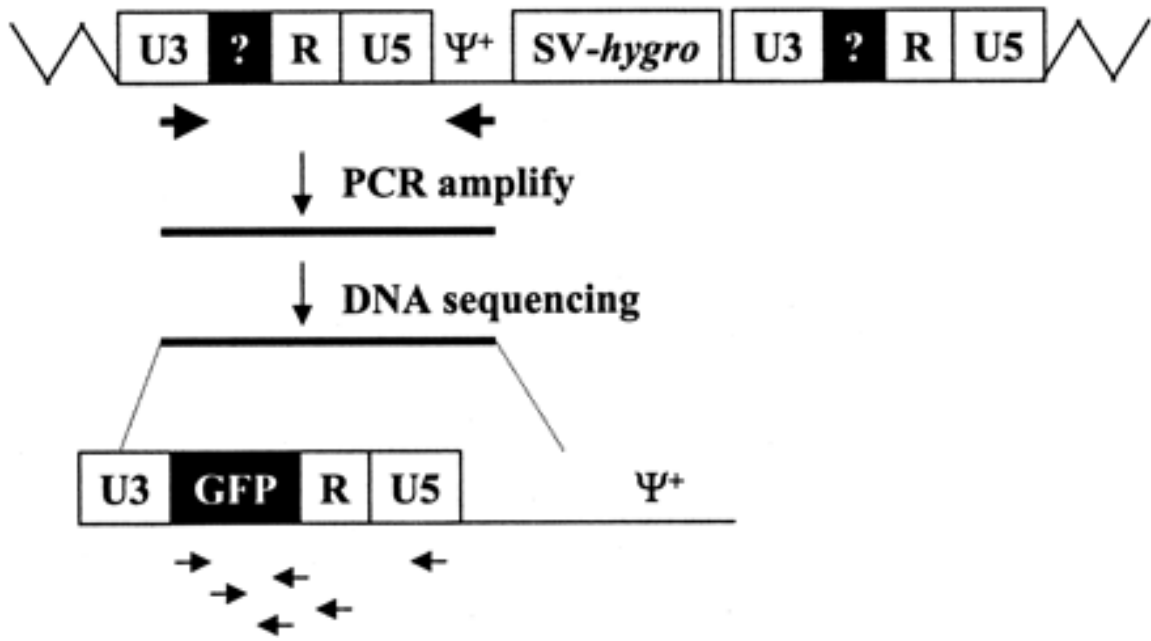
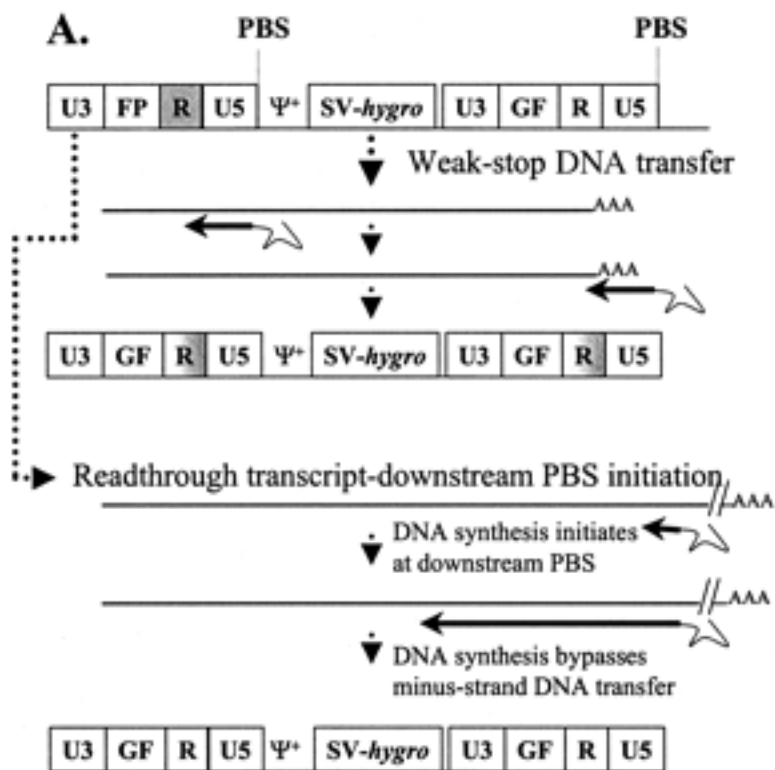


Figure 4.



B. Upstream R 5' GCGCCAGTCTTCCGATAGACTGCGTCG
 Downstream R 5' GCGCCAGTCCTCCGATTGACTGAGTCG

CCCGGGTACCCGTATTCCCAATAAAGCCTCTTGCTGTTTGCA 3'
 CCCGGGTACCCGTGTATCCAATAAACCTCTTGCAAG-TTGCA 3'

Figure 5.

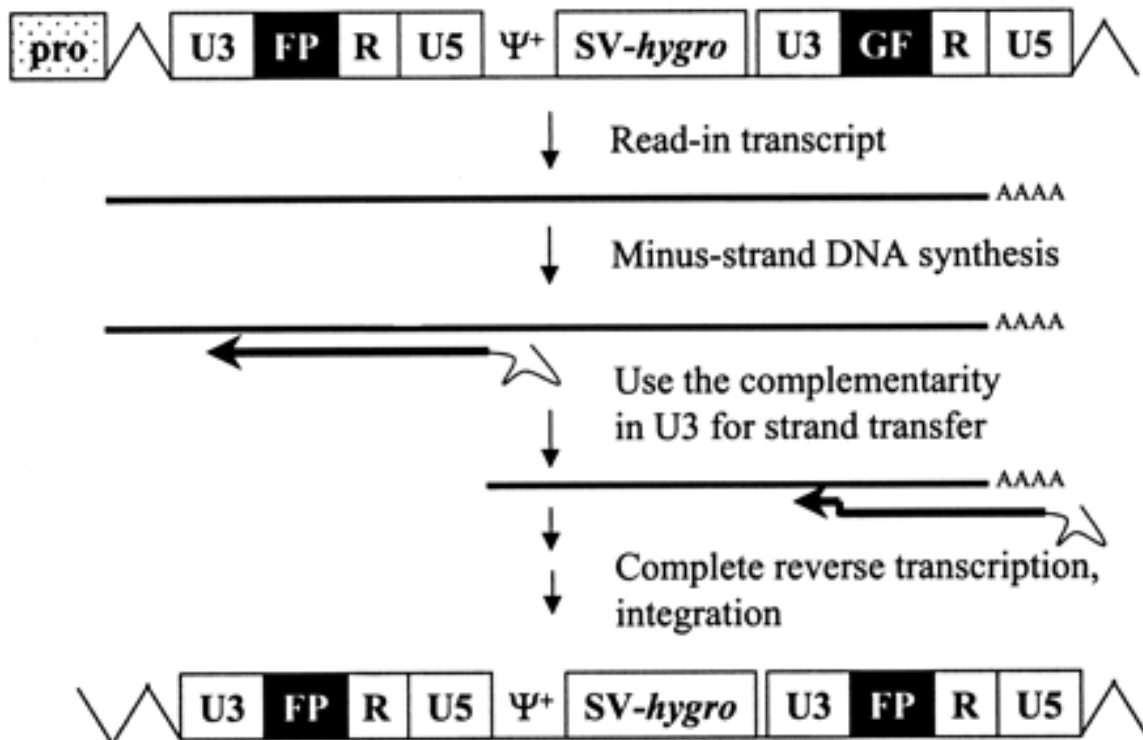


Figure 6.

TABLE 1. Primers used for vector construction, PCR amplification, and DNA sequencing

Primer name	Sequence
GFPp1-AscI	5'-TCTCCGAATGGCGCGCCGCCACCATGAGCAAGGGC-3'
PFG2p-csA	5'-CCTATCTCGAGGCGCGCCTCACTTGTACAGCTCGTCCATG-3'
GFPp3-EheI	5'-ATCCCGAATCGGCGCGCCGCCACCATGAGCAAGGGC-3'
PFG4p-ehEI	5'-GTCATGTCAAGGCGCCTCACTTGTACAGCTCGTCCATG-3'
FPAscp1	5'-ATATAGGCCTGGCGCGCCATGCCCGAGGGCTATG-3'
FG-ehE-4p	5'-CGGTAGCAATGGCGCCTCACTTGTACTTCAGCGCGGGTTC-3'
MLVU3	5'-ATGTTTCCAGGGTGCCCCAAGGACC-3'
IspVLM5'	5'-TCAAACATAGACACTAGACAATCGG-3'
GF101	5'-CAGCGGAGAGGGTGAAGGTG-3'
5UVLM	5'-TCGTGGGTAGTCAATCACTCAGAGG-3'
R/3UM-629	5'-CGACGCAGTCTATCGGAAGACT-3'
PFG-695	5'-CATGCCATGTGTGATCCCAGC-3'
PFG-464	5'-TTGTGCGCCATGATGTACAC-3'

TABLE 2. Viral titers generated by PG13 cells transfected with vector plasmids^a

Exp.	Viral Titers ($\times 10^4$ CFU/ml)				
	MP1	SR6-2wtLTR	SR2-2GFP	SR5-FP-GF	MS2-FP-GF-no3R
A	72	12	ND ^b	2.6	1.3
B	44	19	4.9	1.0	4.7
C	16	8.3	4.1 ^c	13	11
D	43	17	5.9	4.0	3.2
E	130	130	14	24	21

^a Within each set of experiments, the viral titers were standardized to the RT activities.

In general, within each set of experiments, RT activities were within 20% of the differences among the samples.

^bNot determined.

^cThis titer measurement was not standardized to RT activity.

REFERENCES

1. **Adam, M. A., W. R. Osborne, and A. D. Miller.** 1995. R-region cDNA inserts in retroviral vectors are compatible with virus replication and high-level protein synthesis from the insert. *Hum Gene Ther.* **6**:1169-76.
2. **Allain, B., M. Lapadat-Tapolsky, C. Berlioz, and J. L. Darlix.** 1994. Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome. *Embo J.* **13**:973-81.
3. **Allain, B., J. B. Rascle, H. de Rocquigny, B. Roques, and J. L. Darlix.** 1998. CIS elements and trans-acting factors required for minus strand DNA transfer during reverse transcription of the genomic RNA of murine leukemia virus. *J Mol Biol.* **277**:225-35.
4. **Berkhout, B., J. van Wamel, and B. Klaver.** 1995. Requirements for DNA strand transfer during reverse transcription in mutant HIV-1 virions. *J Mol Biol.* **252**:59-69.
5. **Blain, S. W., and S. P. Goff.** 1995. Effects on DNA synthesis and translocation caused by mutations in the RNase H domain of Moloney murine leukemia virus reverse transcriptase. *J Virol.* **69**:4440-52.
6. **Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. *Science.* **263**:802-5.
7. **Coffin, J. M.** 1996. *Retroviridae: the viruses and their replication*, vol. 3. Raven Press, New York, NY.
8. **Dang, Q., and W.-S. Hu.** manuscript in preparation. .

9. **Darlix, J. L., A. Vincent, C. Gabus, H. de Rocquigny, and B. Roques.** 1993. Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *C R Acad Sci III.* **316**:763-71.
10. **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.
11. **Delviks, K. A., and V. K. Pathak.** 1999. Development of murine leukemia virus-based self-activating vectors that efficiently delete the selectable drug resistance gene during reverse transcription. *J Virol.* **73**:8837-42.
12. **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.
13. **DeStefano, J. J.** 1996. Interaction of human immunodeficiency virus nucleocapsid protein with a structure mimicking a replication intermediate. Effects on stability, reverse transcriptase binding, and strand transfer. *J Biol Chem.* **271**:16350-6.
14. **Dougherty, J. P., and H. M. Temin.** 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3' RNA processing. *Proc Natl Acad Sci U S A.* **84**:1197-201.
15. **Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell.* **18**:93-100.

16. **Gritz, L., and J. Davies.** 1983. Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene*. **25**:179-88.
17. **Guo, J., L. E. Henderson, J. Bess, B. Kane, and J. G. Levin.** 1997. Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J Virol*. **71**:5178-88.
18. **Halvas, E. K., E. S. Svarovskaia, and V. K. Pathak.** 2000. Development of an in vivo assay to identify structural determinants in murine leukemia virus reverse transcriptase important for fidelity. *J Virol*. **74**:312-9.
19. **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.
20. **Herman, S. A., and J. M. Coffin.** 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science*. **236**:845-8.
21. **Hsu, M., L. Rong, H. Rocquigny, B. P. Roques, and M. A. Wainberg.** 2000. The effect of mutations in the HIV-1 nucleocapsid protein on strand transfer in cell-free reverse transcription reactions. *Nucleic Acids Res*. **28**:1724-1729.
22. **Jang, S. K., M. V. Davies, R. J. Kaufman, and E. Wimmer.** 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vivo. *J Virol*. **63**:1651-60.
23. **Jang, S. K., H. G. Krausslich, M. J. Nicklin, G. M. Duke, A. C. Palmenberg, and E. Wimmer.** 1988. A segment of the 5' nontranslated region of

- encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol.* **62**:2636-43.
24. **Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Mol Gen Genet.* **177**:65-72.
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. **Kim, J. K., C. Palaniappan, W. Wu, P. J. Fay, and R. A. Bambara.** 1997. Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J Biol Chem.* **272**:16769-77.
27. **Klaver, B., and B. Berkhout.** 1994. Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res.* **22**:137-44.
28. **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.
29. **Lobel, L. I., and S. P. Goff.** 1985. Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. *J Virol.* **53**:447-55.
30. **Luo, G. X., and J. Taylor.** 1990. Template switching by reverse transcriptase during DNA synthesis. *J Virol.* **64**:4321-8.

31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
32. **Miller, A. D., J. V. Garcia, N. von Suhr, C. M. Lynch, C. Wilson, and M. V. Eiden.** 1991. Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J Virol.* **65**:2220-4.
33. **Miller, A. D., and G. J. Rosman.** 1989. Improved retroviral vectors for gene transfer and expression. *Biotechniques.* **7**:980-2, 984-6, 989-90.
34. **Peliska, J. A., S. Balasubramanian, D. P. Giedroc, and S. J. Benkovic.** 1994. Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry.* **33**:13817-23.
35. **Peliska, J. A., and S. J. Benkovic.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science.* **258**:1112-8.
36. **Raja, A., and J. J. DeStefano.** 1999. Kinetic analysis of the effect of HIV nucleocapsid protein (NCp) on internal strand transfer reactions. *Biochemistry.* **38**:5178-84.
37. **Ramsey, C. A., and A. T. Panganiban.** 1993. Replication of the retroviral terminal repeat sequence during in vivo reverse transcription. *J Virol.* **67**:4114-21.
38. **Riggs, J. L., R. M. McAllister, and E. H. Lennette.** 1974. Immunofluorescent studies of RD-114 virus replication in cell culture. *J Gen Virol.* **25**:21-9.
39. **Rodriguez-Rodriguez, L., Z. Tsuchihashi, G. M. Fuentes, R. A. Bambara, and P. J. Fay.** 1995. Influence of human immunodeficiency virus nucleocapsid

- protein on synthesis and strand transfer by the reverse transcriptase in vitro. *J Biol Chem.* **270**:15005-11.
40. **Rong, L., C. Liang, M. Hsu, L. Kleiman, P. Petitjean, H. de Rocquigny, B. P. Roques, and M. A. Wainberg.** 1998. Roles of the human immunodeficiency virus type 1 nucleocapsid protein in annealing and initiation versus elongation in reverse transcription of viral negative-strand strong-stop DNA. *J Virol.* **72**:9353-8.
41. **Swain, A., and J. M. Coffin.** 1992. Mechanism of transduction by retroviruses. *Science.* **255**:841-5.
42. **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.
43. **Temin, H. M.** 1976. The DNA provirus hypothesis. *Science.* **192**:1075-80.
44. **Weiss, S., B. Konig, Y. Morikawa, and I. Jones.** 1992. Recombinant HIV-1 nucleocapsid protein p15 produced as a fusion protein with glutathione S-transferase in *Escherichia coli* mediates dimerization and enhances reverse transcription of retroviral RNA. *Gene.* **121**:203-12.
45. **Yin, P. D., V. K. Pathak, A. E. Rowan, R. J. Teufel, 2nd, and W. S. Hu.** 1997. Utilization of nonhomologous minus-strand DNA transfer to generate recombinant retroviruses. *J Virol.* **71**:2487-94.
46. **You, J. C., and C. S. McHenry.** 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem.* **269**:31491-5.

This page intended to be left blank.

CHAPTER THREE

This page intended to be left blank.

**Pol Gene Products From Two Distantly Related Retroviruses Can Co-Package And
Cooperate To Complete Reverse Transcription.**

Sara Rasmussen Cheslock^{1,2}, Wenhui Zhang¹, Wei-Shau Hu¹, and Vinay Pathak¹. ¹HIV
Drug Resistance Program, NCI-Frederick Frederick, MD, 21702. ²Department of
Microbiology, Immunology, and Cell Biology, West Virginia University, Morgantown,
WV 26506.

ABSTRACT

We synthesized several murine leukemia virus (MLV) and spleen necrosis virus (SNV) *pol* mutants in the reverse transcriptase (RT), RNase H and integrase (IN) domains in order to investigate whether Pol polyproteins from MLV could co-package and cooperate with Pol polyproteins from SNV to complete reverse transcription. A previously published chimeric SNV/MLV virus *gag-pol* expression vector system was utilized to investigate the co-assembly properties of these two viruses (Certo et. al, J.Virol., 72:73, 1998).

Western blotting, RT assay, and real-time PCR of reverse transcription products in infected cells examined the phenotypes of several SNV and MLV *pol* mutants. In both SNV and MLV IN, mutations of the Zn²⁺ finger caused a *gag*-processing defect, while IN catalytic core mutants were processed with efficiency similar to that of wild type IN. When early and intermediate reverse transcription products were compared to the total RT activity, the Zn²⁺ finger and catalytic core mutants did not display a defect in initiation of reverse transcription or minus-strand transfer. A mutation in SNV RNase H exhibited no processing defects and was found to be deficient in minus-strand transfer.

We also attempted to rescue the IN and RNase H mutants by complementation with an MLV *pol* vector that contained a mutation catalytic site of RT (D150E). Rescue of the SNV IN, SNV RNase H and MLV IN mutations was observed at a low rate, indicating that it was possible for the *pol* gene products from two different retroviruses to assemble into one virion. Because the MLV Pol could be weakly complemented with either SNV or MLV RT at roughly equal efficiencies, there was apparently no preference for the homologous RT. Furthermore, whether the Zn²⁺ finger or catalytic core of IN was

mutated seemed to have little effect on the efficiency of complementation. In summary, despite being genetically distinct viruses with little sequence homology (42%) in their *pol* genes, SNV and MLV Pol proteins could co-assemble into virions and cooperate to complete reverse transcription, which indicates that crucial protein-protein contacts in the reverse transcription complex are conserved between these two viruses.

INTRODUCTION

All retroviruses contain at least four functional genes- *gag*, *pro*, *pol*, and *env*. The *gag* gene encodes viral structural proteins, while the *pol* encodes the viral enzymatic machinery for replication. The *gag* of murine leukemia virus (MLV) is comprised of matrix (MA), p10, capsid (CA), and nucleocapsid (NC) proteins. The *pol* of retroviruses encodes the reverse transcriptase (RT), ribonuclease H (RNase H), and integrase (IN) proteins. The *pro* encodes the viral protease, which cleaves the polyproteins to their smaller proteins. The Pol polyproteins are always synthesized from a Gag-Pol precursor protein, and Gag is synthesized about 10 to 20 times more often than Gag-Pol (11, 13, 21). The incorporation of Pol into viruses is therefore dependent on Gag expression. Following translation of these proteins in gammaretroviruses, they traffic to the plasma cell membrane, where budding occurs. The Gag and Gag-Pol polyproteins bud from the cell membrane, and during this assembly process, the viral protease is activated and the polyproteins are cleaved to their smaller products, which allow maturation of the virus cores. This maturation process can be seen by electron microscopy; a mature virus has an electron-dense staining core in the center of the virus, while an immature particle lacks the electron-dense core (8, 15, 20, 22, 25, 28, 30).

It has been shown that Gag proteins from two different viruses cannot co-assemble into the same virus particle. For MLV and Spleen Necrosis Virus (SNV), it has been shown that the two Gags cannot co-assemble and that their ability to package RNA is non-reciprocal. MLV Gag can only package MLV genomic RNA, whereas SNV Gag can package SNV and MLV genomic RNA (5). Use of an SNV*gag*MLV*pol* expression

vector in a cell line that expressed both MLV vector RNA and SNV vector RNA demonstrated that the SNV Gag proteins were capable of recognizing both SNV and MLV viral RNA with equal efficiency.

It is possible to use *trans*-complementation systems to study virus replication (12, 17, 27, 29). Use of such systems allows determination of the roles of different proteins during reverse transcription *in vivo*. It was previously shown that RT and RNase H from two different Gag-Pol precursors could complement extremely well *in vivo*. One construct carries a mutation in the RH domain of *pol*, the other in the RT domain. When each is transfected into cells expressing an MLV vector RNA, viruses produced from the cell line carry both mutants. Furthermore, these Pol-mixed virions are able to replicate. Other labs have shown that RT and IN can complement each other *in trans* quite efficiently (17, 27, 29) either by use of two different Gag-Pol expression constructs in a cell line or by use of a vpr-IN fusion protein for HIV or Gag Δ NC-IN fusion protein for MLV.

It was of interest to determine whether the RT and IN proteins of SNV and MLV could functionally complement each other because pleiotropic effects on virus life cycle from IN mutations have been widely described in the literature (for review, see reference (6). For HIV-1, mutating the conserved HHCC motif (which putatively forms a Zn²⁺-finger) of the IN protein abrogated viral DNA synthesis and integration activity in infected cells. However, mutating any of the conserved central catalytic core residues by substitution caused only a loss of integration- not reverse transcription (29). Somewhat differing results were found for MLV IN; mutation of either the catalytic core or the HHCC motif did not cause a loss of viral DNA synthesis (17). In contrast, the C-terminal

domain of MLV IN appeared to be required for viral DNA synthesis. Mutations in the HHCC motif of IN have been shown to cause defects in virus protein processing (23), and deletion of the IN-coding domain from HIV-1 can cause a loss of processing altogether (3). Because of these pleiotropic effects of IN mutations, the IN mutations generated to establish the RT/IN complementation system were chosen carefully.

SNV and MLV are two very distantly related members of the gammaretrovirus family. Although they contain little homology in their *pol* genes (53%), the SNV RT and IN are able to mediate reverse transcription of MLV RNA as well as the MLV Pol proteins. Therefore, it should be possible to establish an RT/IN complementation system using SNV and MLV Pol proteins and MLV vector RNA. Complementation of the RT and IN of SNV with the RT or IN of MLV would demonstrate that although there is little sequence homology between these two viruses, crucial protein contacts in the reverse transcription and pre-integration complex are conserved

We seek to determine if the Pol proteins from MLV and SNV could also co-assemble and form infectious virions. In order to do this, it will be necessary to establish a complementation system that requires assembly of different species of Pol into the same virus for production of infectious virions. We wished to see if a chimeric RT/IN complementation system from SNV and MLV could be used to demonstrate that SNV and MLV Pol proteins could co-assemble into the same virus and if the RT from SNV or MLV could complement the MLV or SNV IN during reverse transcription and integration.

Two Gag-Pol expression constructs were used to study complementation between RT and IN in these experiments. The first is RD136, an SNV *gag-pol* construct, and the

second is SNVgagMLVpol. A panel of Pol mutants was constructed in both of these vectors in order to determine if Pol proteins from SNV could co-assemble with Pol proteins of MLV. Furthermore, the replication properties of these mutants were determined, and they were used in a trans-complementation system to study the efficiency with which Pol proteins from SNV could functionally replace Pol proteins from MLV.

MATERIALS AND METHODS

Vector Construction. Cloning was carried out using standard methods (24). All SNV gag-pol expression vectors were derived from pRD136, and all SNV-gag MLV-pol chimeric vectors were derived from pSNVgag. All primers used for cloning are shown in Table 1, and the vectors are shown in Fig. 1A. Vector pSR7-D150E was subcloned from an MLV D150E RT mutant by excising the fragment between two *BstEII* sites and inserting the fragment from the MLV-D150E vector that had been created by a similar *BstEII* digestion and treated with Calf Intestinal Phosphatase (Roche). The SR7-D150E clone contained a new *Bst1107I* site and mutated the Aspartate (D) residue at position 150 to a Glutamate (E). The LC2 vector contained a 4-amino acid insertion between the two Cysteine residues of the HHCC motif in IN. This insertion was made by annealing two oligonucleotides to create the linker 5'-AGC TGG TAT ACC-3', which was inserted between the GCA and AGC residues of the HHCC motif. pSNVgag was partially digested with *HindIII*, and the linker was inserted at nt 5165, which generated a unique *Bst1107I* site. The SR13 vector encoded a substitution mutation in the D125 residue of IN to an alanine (A), which was obtained by PCR mutagenesis. Briefly, the primer

SNVgag-4421 and D125AMLVIn for were used to PCR amplify a fragment from nt 4421 to 5253 of pSNVgag. Primers D125AMLV In rev and SNVgag6060C were used to PCR amplify a fragment from nt 5253 to 6060; these two fragments were joined together in a subsequent PCR reaction with primers SNVgag4421 and SNVgag6060C. The PCR product was digested with *Bpu1102I* and *PmII*. The product also contained an additional *BsmI* site that did not change the coding region and was used to identify mutants containing the new mutation. The fragment containing the substitution was cloned into pSNVgag, which had been digested with *Bpu1102I* and *PmII*.

The RD136YVNN vector was constructed via the Stratagene Quickchange mutagenesis kit. This vector incorporated two substitution mutations in the YVDD motif of RT- the two aspartic residues were changed to asparagine residues. Briefly, the mutagenic primer contained the substitution mutations and another primer knocked out a *HindIII* site at nt 5684 in pRD136. The PCR product contained an *HpaI* site at 3495 for screening. The fragment was digested with *Asp718* and *Bst1107I* and ligated into pRD136, which had been digested with *Asp718* and *Bst1107I* and treated with CIP.

The SR8 vector was constructed by partial digestion of RD136 with *SphI*, and insertion of a linker between the two C residues of the HHCC motif that inserted 4 amino acids (tryptophan, methionine, histidine, and alanine). The linker was formed by annealing IN-minus-5' to IN-minus-3'. Insertion of the linker added a *BamHI* site to the vector.

The SR12 integrase mutant incorporated a substitution at D109 to A. This substitution was generated by PCR mutagenesis that simultaneously introduced a *BsmI* site into the vector. The primer Nerd136-4372 and primer D109ASNVIfor were used to

amplify a DNA fragment from bp 4372 to 5209 of pRD136. Similarly, primers Nerd136-5475C and D109ASNVI_{rev} were used to amplify another DNA fragment from nt 5209 to 5475 of pRD136. The two PCR products were joined together in a subsequent PCR reaction with Nerd136-4372 and Nerd136-5474C. This fragment was digested with *SgrAI* and *AgeI*. The digested product was cloned into RD136 that had been digested with *SgrAI* and *AgeI*. DNA sequence analysis and restriction enzyme mapping confirmed all mutations and ensured that no other mutations had been introduced during cloning procedures.

Tissue Culture. The A3-4 clone of the GFFP cell line (31) was used for all transfections. These cells were maintained in Dulbecco's Modified Eagle's Media (DMEM) supplemented with 6% calf serum and 50 µg/mL streptomycin and 50 U/mL of penicillin. G418 selection was done at a concentration of 600 µg/mL and hygromycin selection was done at 120 µg/mL. Ouabain selection was performed at 73 µg/mL. Briefly, the GFFP cell line contains a copy of p-SV-A-MLV, which expresses amphotropic MLV envelope, and an MLV vector that carries the gene for neomycin resistance and the gene for green fluorescent protein (GFP) that has been divided into two fragments which contain overlapping homology to measure direct repeat deletion. The SNVgagD150E cell line was derived from the A3-4 GFFP cell line by transfection of SR7-D150E into the GFFP cells by DMSO-Polybrene transfection (2). SR7-D150E was co-transfected at a 10:1 ratio into the cell line with pSV α 3.6, which encodes the gene for the Na⁺/K⁺ pump and confers resistance to ouabain selection (ref). Cells were selected with ouabain 48 hours following transfection and pools of resistant cells were obtained. These cells were known as the SNVgagD150E cell line.

All vectors were co-transfected either into the GFFP cell line or the SNVgag-D150E cell line by DMSO-Polybrene method with the SV-hygro plasmid, which carries the gene for hygromycin phosphotransferase B (9), at a 10:1 (gag-pol vector: pSVhygro) ratio. Cells were selected for hygromycin resistance 24 hours following transfection, and pools of these cells were obtained for mutant and complementation analysis.

A portion of the A3-4 GFFP cells that were transfected with RD136-YVNN were maintained as the YVNN cell line, in order to test for SNV homologous complementation. These cells were co-transfected with each pol mutant vector and pSV α 3.6 and selected for ouabain resistance. All cells containing gag-pol expression vector constructs were maintained in AZT to prevent reinfection of producer cells.

Virus harvest and analysis. Virus-producing cells were plated to five 100-mm tissue culture plates, at a density of 5×10^6 cells per plate. Each plate of cells was fluid changed 24 hours following the plating with 8 mls of fresh media, and D17 cells were plated to 60 mm dishes, at a density of 1×10^5 cells per dish. Media containing the viruses were harvested 24 hours following this fluid change from the virus producing cells. The virus-containing supernatants were clarified by low-speed centrifugation (4000 RPM for 10 minutes) or by filtering through a 0.45 μ m filter (Corning). An aliquot of the virus was serially diluted in the presence of 50 μ g/mL polybrene and used to infect the D17 cells for four hours. Following infection, virus-containing media was aspirated from the D17 cells and replaced with fresh media. Twenty-four hours following infection the infected target cells were subjected to G418 selection to determine virus titers.

The remaining virus was concentrated in a Discovery 100S Preparative Ultracentrifuge (Sorvall) at 25,000 RPM at 4 degrees Celsius for 90 minutes, using a

Sorvall Surespin 630 rotor. The pelleted virus was resuspended in media and used for Western blotting and RT assay.

Real-Time PCR. D17 cells were plated at a density of 5×10^6 cells per 100-mm dish and infected with 2 mL of virus in the presence of 50 $\mu\text{g}/\text{mL}$ polybrene. One hour following application of virus, the cells were fluid-changed with fresh media. Six hours following infection, cells were lysed and total DNA was harvested using the Qiagen Blood Mini Kit, according to manufacturer's specifications. DNA was used in a Prism ABI 7700 real-time PCR to detect products of reverse transcription.

Table of primer sequences here. Primer-probe sets homologous to the R-U5 region of the virus vector DNA were used to detect initiation of reverse transcription, and primers homologous to the 3' of neo were used to detect minus-strand transfer and later products of reverse transcription. All PCR reactions were carried out using the TaqMan PCR kits according to the manufacturer's instructions. PCR conditions were as follows: denaturation at 95 degrees Celsius for 10 minutes, followed by 40 cycles of 95 degrees for 15 seconds, and annealing and extension at 60 degrees for 60 seconds. Standard curves were performed with pYV11, which contains one copy of the MLV LTR and the neomycin resistance gene. All real-time PCR analysis was done with the ABI 7700 software package.

Reverse Transcriptase Assay. Exogenous RT activities were determined by incubating 10 μl of virus with 50 $\mu\text{g}/\text{ml}$ of 20-mer oligo T (Integrated DNA Technologies), 100 $\mu\text{g}/\text{ml}$ of poly(A) (Pharmacia), 60 mM NaCl, 50 mM Tris (pH 8.0), 1 U/ μl RNase inhibitor, 10 mM dithiothreitol, 0.6 mM MnCl_2 , 80 μM dTTP, 0.5% IGE Pal (Sigma), and 10 μCi of [^3H]dTTP (72 Ci/mmol, ICN). The samples were incubated at

37°C for 60 min. The reaction mixtures were precipitated with ice-cold 10% trichloroacetic acid (Sigma) and filtered through 0.45-µm-pore-diameter Metrical membranes (Gelman Sciences, Inc.); the amount of [³H] incorporated was determined using a Tricarb 1600 TR Liquid Scintillation Analyzer (Packard).

Western Blotting. Western Blotting was used to determine the amount of virus released by detection with antibodies to reticuloendotheliosis virus (REV) CA protein (a kind gift from Nancy Rice). REV is a very close relative to SNV. Briefly, a portion of the harvested virus was diluted 1:2 with Laemlli's sample buffer (Biorad), heated to 99 degrees Celsius for 5 minutes, and electrophoresed on a 1.5-mm thick 12% Tris-Glycine gel (Invitrogen). Electrophoresis was performed for 90 minutes at 125 mV. Following electrophoresis, the gel was transferred to PVDF blotting membrane using the Owl semi-dry electroblotter for 2 hours at 125 mA. Following transfer, the PVDF membrane was blocked in 5% milk in Tris-buffered Saline (TBS) for 12 hours, probed with 1:10,000 rabbit α-REV CA antibody for 4 hours, washed 3 times with TBS-0.1% Tween-20, and incubated with 1:10,000 goat α-rabbit horse radish peroxidase for 60 minutes. The membrane was washed 3 times with TBS-0.1% Tween-20 for 5 minutes per wash and incubated with ECL-Plus Western Labeling Kit from Amersham Biosciences for 3 minutes. The blot was exposed to Kodak X-omat film. The amounts of unprocessed Gag protein and processed Capsid protein were compared.

Electron Microscopy. To determine the morphology of the integrase mutants that had processing defects, electron microscopy (EM) was done on virus-producing cells and the relative number of immature vs. mature virus particles for each sample was determined. For each sample, at least 100 viruses were examined. Briefly, 1×10^7 transfected cells

were collected, pelleted by low-speed centrifugation (1200 RPM for 2 minutes) and fixed by applying a solution of 2% glutaraldehyde in PBS to the top of each pellet. Samples were paraffin embedded and sliced into thin sections for EM analysis. The EM was performed by the Science Applications International Corporation core facility at NCI-Frederick.

RESULTS

Protocol for Characterization of Mutants. A panel of vectors was constructed to study RT, RNase H, and IN complementation in both SNV*gag-pol* and chimeric SNV*gag-MLVpol* expression vectors (Fig. 1A). Briefly, RD136 is an SNV *gag-pol* expression construct. YVNN is an RT mutant of RD136 in which the YVDD motif has been mutated to YVNN. SR8 is an SNV (IN) mutant that contains an insertion of four amino acids (W, M, H, and A) between the two C residues of the HHCC motif in SNV IN. SR12 is an SNV IN catalytic core mutant that contains a substitution of residue D109 to A. The D109 residue constitutes the first catalytic core residue of the highly conserved D, D, 35E motif of SNV IN. WZH73 is an SNV RNase H mutant that contains a substitution of residue D524 to N.

MP5 is a chimeric SNV*gagMLVpol* expression vector (5). SR7-D150E is an RT mutant that contains a substitution of residue D150 to E. This substitution has been shown to render the virus non-infectious (10). LC2 is an MLV IN mutant that contains an insertion of 4 amino acids (A, G, I, P) between the coordinating C residues in the HHCC motif. This mutation has been shown to have a processing defect with α -CA antibody by western blotting analysis, as well as reduced RT activity in vitro (23). SR13

is an MLV IN catalytic core mutant which contains a substitution in the conserved D, D, 35E motif of MLV IN catalytic core residue D125 to A. Substitution mutations of the catalytic core of IN have been shown to eliminate virus integration activity, while having no effect on processing or RT activity (1, 4, 7, 16-19).

The approach utilized to analyze the replication properties of the vectors is outlined in Fig. 1B. First, the ability of the mutants to support one cycle of replication was examined. The phenotype of the D150E mutation was previously described (10). The mutants were separately co-transfected into the A3-4 GFP cell line (26) with pSV-hygro. Pools of cells containing the stably expressed *gag-pol* expression vectors were obtained, and some of the cells were analyzed by EM to determine virion morphology. Virus was harvested from each pool, and infection of D17 cells, RT assay, western blotting, and real-time PCR were performed to examine the mutants' infectivity, ability to produce virus, and the ability to carry out reverse transcription. A portion of the infected D17 cells were selected with G418 to detect productive infection, and a second portion of the infected D17 cells was lysed 6 hours post-infection to obtain total cellular DNA for quantitative real-time PCR. Real time PCR was used to measure the ability of these mutants to carry out reverse transcription.

Protocol for Complementation Assays of RT and IN mutants. Following the phenotype determination of these mutants, the SNV*gag*-D150E cell line was used for complementation experiments (Fig. 1C). The SNV-*gag*-D150E cell line expresses a defective MLV RT and wild type MLV IN from the SR7-D150E mutant. For complementation analysis, the IN and RNase H pol mutants were each co-transfected with pSV*hygro* into the SNV*gag*-D150E cell line separately (Fig 1C). Pools of these

cells were obtained for each mutant. Viruses harvested from these cell pools should contain a mix of the Gag-Pol containing the D150E mutation, and the gag-pol containing the RNase H mutation or the IN mutations. These mixed viruses were used to infect D17 cells to determine titer by G418 selection and virion RT activity.

Titers of *Pol* Mutants. The titers of the viruses carrying the SNV and MLV Pol mutations are shown in Table 2. Titers were determined by three to six independent experiments. The titer for RD136, the SNV wild type gag-pol expression construct, was 2.3×10^3 cfu/mL. The titer of the SNV RNase H mutant, WZH73, was 2.3, or 1000-fold less than wild type RD136. The SR8 SNV HHCC IN mutant and the SR12 SNV IN catalytic core mutant had a titer of <1 cfu/mL. Similarly, while the wild type SNVgag-MLVpol vector MP5 had a titer of 6.6×10^3 , both the LC2 IN HHCC insertion mutant and the SR13 IN catalytic core mutant had titers of <1 cfu/mL. The titer assays confirmed that these Pol mutations were lethal to virus replication.

RT Activity of *Pol* Mutants. Virus preparations derived from Pol mutants were subjected to RT Assays. The results of the RT assays, determined in three to six independent experiments, are presented in Table 2. The values of the RT assay for each mutant are reported relative to SNV wild-type RD136 (set to 100%) RT activity. The RT activity of the YVNN RT mutant was 4.9% of wild type. The WZH73 SNV RNase H mutant RT activity was 272% of wild type. The RT activity of SR8 SNV IN HHCC mutant was 10% of wild type. The RT activity of SR12, the SNV IN catalytic core mutant, was 425% of wild-type activity. For the two mutants that showed RT activity significantly higher than wild type, the western blot profiles demonstrated that more virus was being produced for each set because of differences among transfection efficiencies

(see appendix). The RT activities for the MLV Pol mutants are also shown in Table 2. The RT activity of the wild-type MP5 has been normalized to 100%. The LC2 IN HHCC insertion mutant virus had an RT activity of 24%, and the SR13 IN catalytic core substitution mutant had an RT activity of 35% of wild type.

For all the RT activities, only the YVNN and SR8 SNV IN HHCC insertion mutant had an RT activity that was significantly lower than wild type (4.9%, and 10% of wild type, respectively). The diminished RT activity of the YVNN mutant was not surprising, as this mutation disrupted the catalytic core of the RT protein. This result suggested that the IN mutation in SR8 led to a reduction in either the production of RT or the enzymatic activity of RT.

Mutations in the HHCC domain of IN result in processing defects. The western blotting analysis for the MLV mutants is shown in Fig. 3A, and the analysis for SNV mutants is shown in Fig. 2B. The analysis of MP5 virus showed no detectable processing defects, and a band corresponding to p30 CA is evident (Fig 3A, lane 1). In contrast, analysis of the LC2 IN insertion mutant indicated that the p30 CA band was diminished, and a band corresponding to unprocessed Gag at p55 was evident (Fig 3A, lane 2). This suggested that the IN mutation present in LC2 resulted in a processing defect. Fig 3A, lane 3 shows the analysis of SR13, the IN catalytic core mutant. This mutant had a processing profile similar to wild-type.

The analysis of RD136 wild type virus showed no detectable processing defects, and a band corresponding to p30 CA is evident (Fig 3B, lane 1). In contrast, analysis of the SR8 HHCC IN insertion mutant indicated that the p30 CA band was diminished, and a band corresponding to unprocessed Gag at p55 was evident (Fig 3B, lane 2). This

suggested that the IN mutation present in SR8 resulted in a processing defect. Fig 3B, lane 3 shows the analysis of SR12, the IN catalytic core mutant. This mutant had a processing profile similar to wild type. Therefore, the HHCC motif mutants (LC2 and SR8) and the catalytic core mutants (SR13 and SR12) for SNV *pol* and MLV *pol* have the same processing phenotype. The western blot profile of WZH73 was also determined; it has processing activity equivalent to wild type (data not shown).

Virion Morphology. In order to see if the processing defects observed for SR8 and LC2 resulted in altered virion morphology, transfected cells producing the IN mutant viruses were examined by EM (Table 3). A total of 100 particles each for the MP5 wild type, SR8 SNV HHCC mutant, and the LC2 MLV HHCC mutant were visualized and the numbers of immature and mature particles in each set were determined. The MP5 virus had 28/100 immature particles, the SR8 mutant had 57/100 immature particles, and the LC2 mutant had 38/100 immature particles. Therefore, the proteolytic processing defect observed by western blotting analysis did not result in any apparent morphological changes or shifts in maturation in the IN-mutant viruses.

Real-Time PCR of Pol Mutants. In order to determine if the *pol* mutations had an effect on reverse transcription, D17 cells were infected with viruses containing the different Pol mutations; and 6 hours post-infection the infected cells were lysed, total cellular DNA was collected, and the amount of DNA was determined by a spectrophotometer. One microgram of each of these DNA samples were subjected to real-time PCR analysis to detect products of reverse transcription. A primer-probe set homologous to the R-U5 region was used to detect minus-strand strong-stop DNA, and a primer-probe set homologous to the 3' end of the *neo* gene was used to detect post-minus-strand DNA

transfer products of reverse transcription. The data are presented in Table 4. The amounts of PCR products for the SNV wild type RD136 were set to 100%. The RT-deficient YVNN mutant yielded very few reverse transcription products (0.2% and 0.1% of wild type R-U5 and neo products, respectively). SR8, the SNV IN HHCC insertion mutant, exhibited 8.7% of wild-type R-U5 copy number and 9.4% of the wild-type neo copy number. SR12, the IN catalytic core mutant, had 310% of wild-type R-U5 copy number and 240% of wild-type neo copy number. Finally, the SNV RNase H mutant WZH73 had 23% of the wild-type R-U5 copy number, and 7% of the wild-type neo copy number.

For the MLV Pol mutants, the amount of PCR products for wild type SNVgagMLVpol MP5 were set to 100%. LC2, the HHCC insertion IN mutant, was shown to have 18% of wild-type MP5 R-U5 copy number, and 21% of MP5 wild-type neo copy number. The SR13 IN catalytic core mutant was shown to have 19% of wild-type R-U5 copy number, and 15% of wild-type neo copy number.

Because of differences between transfection efficiencies and differences in RT activities among the *pol* mutant viruses, all values for real-time PCR were normalized to virion RT activity to account for differences in the amount of processed RT in each virus and the amount of virus released from each producer cell line. The results of the normalized amounts of reverse transcription products are also shown in Table 4. Following correction for RT activity, compared to the amounts of PCR products obtained from the wild type, SR8 HHCC insertion IN mutation had 80% of SNV wild type R-U5 copy number, and 93% of SNV wild type neo copy number; and SR12 catalytic core IN mutation had 100% of SNV wild-type R-U5 copy number, and 56% of SNV wild type

neo copy number. WZH73 had 8.5% of SNV wild type R-U5 copy number, and 2.6% of SNV wild-type neo copy number. For the MLV Pol mutant viruses, LC2 HHCC insertion IN mutation had 76% of MLV wild type R-U5 copy number and 89% of MLV wild type neo copy number, and SR13 catalytic core IN mutation had 49% of MLV wild type R-U5 copy number and 36% of MLV wild type neo copy number, following normalization to RT assay values. Therefore, once the amount of RT activity from each sample was accounted for, the IN mutants were shown to be competent in both the initiation of minus-strand DNA synthesis and minus-strand DNA transfer.

Complementation with SR7 D150E in Phenotypically mixed virions. In order to determine the ability of the MLV RT-IN⁺ construct to co-assemble with and complement an RT⁺IN⁻ *gag-pol* precursor, the SNV *gag-D150E* construct MLV (RT⁻IN⁺) was used to establish a cell line. The cell line also expressed amphotropic MLV envelope and contained an integrated copy of an MLV vector (Fig. 1C). These cells were co-transfected with IN⁻RT⁺ *gag-pol* constructs with pSV-hygro. Pools of hygro resistant cells for each mutant were obtained, and virus harvested from these pools was subjected to RT and titer assays. The results are shown in Table 5. For the SNV *gag-pol* expression vector RD136, the average titer in three independent sets of experiments was 223×10^2 cfu/mL. The SR8 HHCC IN insertion mutant produced a titer of 1.2×10^2 cfu/mL (0.5% of wild-type), and the SR12 catalytic core IN⁻ mutant produced a titer of 1.2×10^2 cfu/mL (0.5% of wild-type). The WZH73 RNase H mutant was produced a complemented titer of 5.3×10^2 cfu/mL, or 2.4% of SNV wild-type titer. In order to correct for the amount of processed RT available in the complemented virions, and to account for differences in transfection efficiencies, the complemented titers were all

normalized to RT activity. The normalized titers are shown in Table 5. The SNV wild type vector RD136 titer did not change with normalization. The SR8 HHCC IN insertion mutant had a normalized titer of 15×10^2 , or 7% of wild-type titer. The SR12 catalytic core IN- mutant had a normalized titer of 1.2×10^2 cfu/mL, 0.5% of wild type titer. WZH73 produced a normalized titer of 5.4×10^2 , or 2.4% of wild type titer. The SR12 catalytic core IN- mutant and WZH73 RNase H mutant titers did not change significantly with normalization to RT activity, while the SR8 HHCC IN mutant titer increased 10-fold. The level of complementation relative to RT activity was better for the SR8 HHCC IN insertion mutant than in then SR12 catalytic core IN mutant, probably because of the processing defect seen in the SR8 HHCC IN mutant reduced the RT activity generated by the construct.

For the SNV*gag-MLVpol* mutants, the MP5 wild type construct had a titer of 256×10^2 cfu/mL, and the LC2 HHCC IN insertion mutant (RT^+IN^-) produced a complementation titer of 0.8×10^2 cfu/mL (0.3% of wild-type). SR13 catalytic core IN mutant produced a titer of 0.1×10^2 cfu/mL (0.5% of wild-type). As with the SNV Pol mutant viruses, these titers were normalized for RT activity. The wild type MP5 titer did not change with normalization, and the LC2 HHCC IN (RT^+IN^-) insertion mutant had a normalized titer of 150 cfu/mL; this raised its relative wild type titer to 0.5%. The SR13 (RT^+IN^-) catalytic core mutant had a normalized titer of 6 cfu/mL; this lowered its relative wild type titer to 0.3%.

These results show for the first time that SNV*gagSNVpol* and SNV*gagMLVpol* proteins can co-assemble and replicate together. Complementation by MLV IN produced by IN^+RT^- constructs and MLV RT produced by the RT^+IN^- construct was as efficient as

complementation by SNV IN produced by IN⁺RT⁻ constructs. This suggests that the protein-protein contacts between the SNV and MLV reverse transcription complexes and the pre-integration complexes are similar in these two viruses.

Complementation with SNV YVNN mutant in Phenotypically Mixed Virions.

In order to determine the ability of an SNV RT⁺IN⁺ construct to co-assemble with and complement an RT⁺IN⁻ *gag-pol* precursor, the SNV*gag*-YVNN construct (SNV RT⁺IN⁺) was used to establish a cell line. The cell line also expressed amphotropic MLV envelope and contained an integrated copy of an MLV vector (Fig. 4A). These cells were co-transfected with IN⁺RT⁺ *gag-pol* constructs with pSV α 3.6. Pools of ouabain-resistant cells for each mutant were obtained, and virus harvested from these pools was subjected to RT and titer assays. The results are shown in Table 6. The SNV wild type RD136 vector produced a titer of 120×10^2 cfu/mL. The SR8 HHCC IN insertion mutant had no detectable titer, and it had an RT activity of 2% of wild type. This suggests that the YVNN had a dominant negative effect on the SR8 vector, since this vector was able to complement the MLV RT (Table 5). The SR12 catalytic core IN mutant had a titer of 30 cfu/mL (0.2%) and had 88% wild-type RT activity. The WZH73 RNase H mutant, which produced an RT activity of 68% of wild-type, produced a titer of 150 cfu/ml, or 1.3% of wild type titer. When normalized to RT activity, the SR12 catalytic core IN mutant had a titer of 34 cfu/mL, or 0.3% wild type titer, and the WZH73 RNase H mutant had a titer of 220 cfu/mL, or 1.8% wild type titer.

The results of the YVNN complementation with the MLV *pol* mutants are shown in Table 6. The MP5 wild-type SNV*gag*MLV*pol* vector had a titer of 130 cfu/mL. The LC2 HHCC IN insertion mutant had an RT activity that was 30% of wild type and

produced no detectable titer. The SR13 IN substitution mutant had a RT activity that was 140% of wild type and produced a titer of 60 cfu/mL, or 47% of the wild type. Following correction for RT activity, the SR13 IN HHCC mutant produced a complementation titer of 120 cfu/mL, or 100% of wild type.

First of all these results demonstrate that SNV*gag*SNV*pol* and SNV*gag*MLV*pol* proteins can complement each other and restore titer. Second of all, the LC2 and SR8 IN HHCC insertion mutants were unable to complement, despite the fact that the LC2 mutant had RT activity. It was noted that the YVNN mutant, in addition to its RT defect, had a processing defect on western blot (data not shown). The SR8 and LC2 HHCC insertion mutants also had processing defects; it is possible that because both Gag-Pol precursors had processing defects, they were unable to complement each other to restore virus titer.

DISCUSSION

Previously, it was shown that a chimeric SNV*gag*MLV*pol* vector could replicate with wild type efficiency as long as the vector RNA used was derived from MLV sequences (5). This chimeric vector was utilized to investigate whether Pol polyproteins from two different species of virus, SNV and MLV, could be incorporated into the same virion and cooperate to complete one round of virus replication. The results presented in this paper show for the first time that SNV and MLV Pol proteins can co-assemble together in the same virion, as long as the two polyproteins are synthesized with an SNV*gag* in the Gag-Pol molecule. Furthermore, it was shown that two different species of Pol polyproteins could interact to complete reverse transcription and integration. This

was shown definitively by the ability of an RT-deficient SNVgagMLVpol (RT⁻IN⁺) construct to be complemented by an IN-deficient SNVgagSNVpol (RT⁺IN⁻) construct to complete viral replication. Similarly, an RT-deficient SNVgagSNVpol (RT⁻IN⁺) was be rescued by an IN-deficient SNVgagMLVpol (RT⁺IN⁻).

The level of complementation seen between the SNV RT⁺IN⁻ constructs and the MLV RT⁻IN⁺ constructs was noted to be similar to the level of complementation seen between MLV RT⁺IN⁻ and MLV RT⁻IN⁺ in these experiments (Tables 5 and 6). These results showed that level of complementation of the MLV RT⁻IN⁺ vector were the same for SNV RT⁺IN⁻ and MLV RT⁺IN⁻, indicating that SNV and MLV RT proteins can interact equally well with MLV IN.

Thirdly, the nature of the mutation used to knock out IN function impacted the level of complementation seen in this system. The HHCC IN insertion mutations used in this study resulted in processing defects that accounted for a decrease in the amount of RT available for reverse transcription. The processing phenotype seen in both the MLV IN and SNV IN mutant viruses agrees with the phenotype seen by others for similar mutations in MLV IN (23). The level of complementation was greater for the HHCC mutation than for the catalytic core substitution mutant following normalization for RT activity. Because the HHCC IN insertion mutant had a processing defect, it produced less RT, and therefore its level of complementation was higher when corrected for RT activity.

Complementation systems have been widely utilized to explore mechanisms and properties of reverse transcription. Efficient complementation has been observed between MLV RT and MLV RH (12); this system showed wild-type interactions between

MLV RT⁻RH⁺ and MLV RT⁺RH⁻ constructs. In the system described here, complementation between the MLV RT⁻RH⁺ and the SNV RT⁺IN⁻ was as efficient as between the MLV RT⁻IN⁺ and the SNV RT⁺RH⁻, indicating that complementation between IN and RT is as efficient as RT and RH. The system utilized in this work differed from the previous investigation because the RT⁻IN⁺ cell lines were derived from pools of transfected cells rather than individual cell clones; this could have impacted the maximum amount of complementation that was observed as the study discussed above used an MLV RT⁻RH⁺ cell line that was derived from a single cell clone.

Many aspects of integrase function have been investigated using trans-complementation systems in HIV-1 and MLV. Initially, these studies proved a direct role for IN during reverse transcription for both HIV-1 and MLV. In HIV-1, the catalytic core of IN was dispensable for reverse transcription, while the HHCC motif was not (17, 29); surprisingly, different results were found for MLV. An MLV IN truncation mutant lacking the C-terminal domain of IN was shown to be defective in reverse transcription (11). Our real-time PCR data indicate that MLV HHCC IN insertion mutants are able to initiate reverse transcription and mediate minus-strand DNA transfer as well as wild-type IN, once normalized to RT activity, which is in agreement with the data reported by others for MLV (17). Our results showed no limitation of the SNV or MLV IN catalytic core mutants in RT activity or initiation of reverse transcription. However, the SNV RNase H mutant WZH73 was somewhat deficient in the initiation of minus-strand DNA synthesis, as well as deficient in minus-strand DNA transfer. This finding is similar to that found with MLV RNase H (14). The SNV and MLV IN HHCC insertion mutants were deficient in RT activity, which correlated with a processing defect seen in these

mutants. Therefore, our results do not definitively demonstrate a role for SNV or MLV IN during reverse transcription of MLV RNA; however, participation of these proteins in reverse transcription cannot be ruled out on the basis of these results.

A second important finding of IN trans-complementation studies has been to demonstrate an interaction between different species of IN proteins. In an HIV-1 system, an HIV-2 IN was not able to rescue HIV-1 reverse transcription; however, the HIV-2 IN was able to integrate HIV-1 cDNA once reverse transcription was completed (29). Our results showed no difference in the ability of MLV IN to rescue reverse transcription or integration in an SNV or MLV RT⁺IN⁻ construct. This strongly suggests that several crucial RT-IN interactions are conserved between SNV and MLV. These results indicate that protein interactions in the reverse transcription and pre-integration complexes for MLV and SNV are conserved, and that the individual proteins can interact with other distantly related retroviral proteins. SNV and MLV Pol proteins contain have relatively low sequence homology (53%). Despite the fact that SNV and MLV IN are very divergent in sequence, data by others suggest that the C-terminal portion of IN is involved the protein-protein interactions. The C-terminal regions of IN encompassing residues 345-365 of MLV has nearly 100% sequence identity with residues 329-349 of SNV IN; this could comprise a region of IN that contains the crucial contacts that allow for the interchangeability of SNV and MLV IN. This is of significance because it highlights a possible mechanism for evolution of new species of retroviruses, and perhaps viruses with increased pathogenicity or immune evasion for the host.

ACKNOWLEDGEMENTS

We thank Kunio Nagashima for EM analysis, and members of the Hu and Pathak labs for thoughtful discussion of experiments.

FIGURE LEGENDS

Figure 1. A: Vectors used in study. MA, matrix, pX, protein X, CA, Capsid, NC, nucleocapsid, Pr, protease, RT, reverse transcriptase, IN, integrase. Open box denotes SNV-derived sequences, while gray shading denotes MLV-derived sequences. B: Experimental Protocols. Gray shading denotes MLV-derived sequences. EM, electron microscopy, RT, reverse transcriptase.

Figure 2. A: MLV RT⁻IN⁺ cell line. Gray shading denotes MLV-derived sequences, open boxes indicate SNV-derived sequences. B: Protocol for Complementation.

Figure 3. Western Blot profile of Viruses harboring IN mutations. A: MP5, SNVgagMLVpol wild type virus. LC2, SNVgagMLVpol IN HHCC insertion mutant. SR13, SNVgagMLVpol IN D125A catalytic core mutant. B: RD136, SNVgagSNVpol wild type virus. SR8, SNVgagSNVpol IN HHCC insertion mutant, SR12, SNVgagSNVpol IN D109A catalytic core mutant. Two bands are visible in both blots, p55-Gag, which represents unprocessed Gag protein, and p30 CA, which represents the processed capsid protein.

Figure 4. A. SNV RT⁻IN⁺ cell line. Gray shading denotes MLV-derived sequences and open boxes denote SNV-derived sequences. B: Protocol for complementation of the SNV RT⁻IN⁺ with RT⁺IN⁻ vectors.

Appendix, Figure 1. RD136, SNVgag-pol wild type. SR12, SNVgag-pol IN D109A catalytic core mutant.

Table 1. Primers used to construct Pol mutants.

Primer Name	Primer Sequence
Nerd136-4372	5'-CTCAAGCAGAAGCCACCTTGTT-3'
D109A-SNVInfor	5'-GGAACACTGGGAGGTGGCATT <u>CACAGAAATGGTAAC</u> AGC-3'
D109A-SNVInrev	5'-GGCTGTTACCATTTCTGTGAATGCCACCTCCCAGTGTTTC C-3'
Nerd136-5475C	5'-AGGTCGGTACGCACAGTGGAGTT-3'
SNVgag-4421	5'-CCACACCTGGTACACGGATGGAA-3'
D125A-MLVInfor	5'-GGCACTCATTGGGAGATCGCATT <u>CACCGAGATAAAGCCC</u> GG-3'
D125A-MLVInrev	5'-CCGGGCTTTATCTCGGTGAATGCGATCTCCCAATGAGTGCC- 3'
SNVgag-6060C	5'-GTTGAACGCGCCATGTCAGTCTA-3'
IN-minus-5'	5'-GATCCATGCATC-3'
IN-minus 3'	5'-CATGGATCCATG-3'
MLV IN-linker5'	5'-AGCTGGTATAACC-3'
MLV IN-linker3'	5'-AGCTGGTATAACC-3'

A.

RD136	MA p12 CA NC	PR RT RH IN
YVNN	MA p12 CA NC	PR RT* RH IN
SR8	MA p12 CA NC	PR RT RH IN [∇]
SR12	MA p12 CA NC	PR RT RH IN*
WZH73	MA p12 CA NC	PR RT RH* IN
MP5	MA p12 CA NC	PR RT RH IN
D150E	MA p12 CA NC	PR RT* RH IN
LC2	MA p12 CA NC	PR RT RH IN [∇]
SR13	MA p12 CA NC	PR RT RH IN*

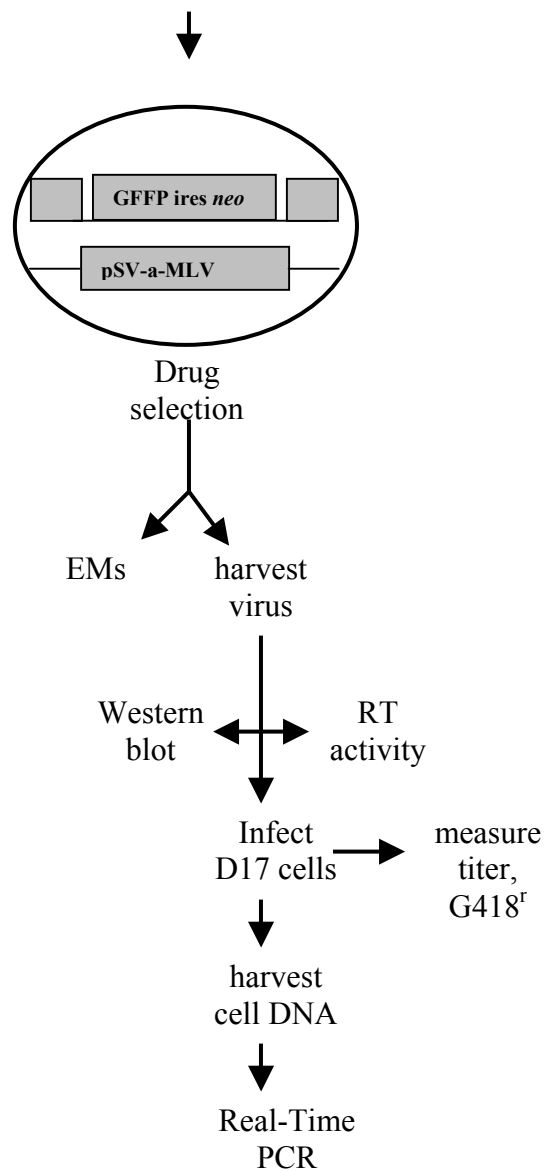
B.*gag-pol* vector + pSVhygro

Figure 1.

Figure 2.

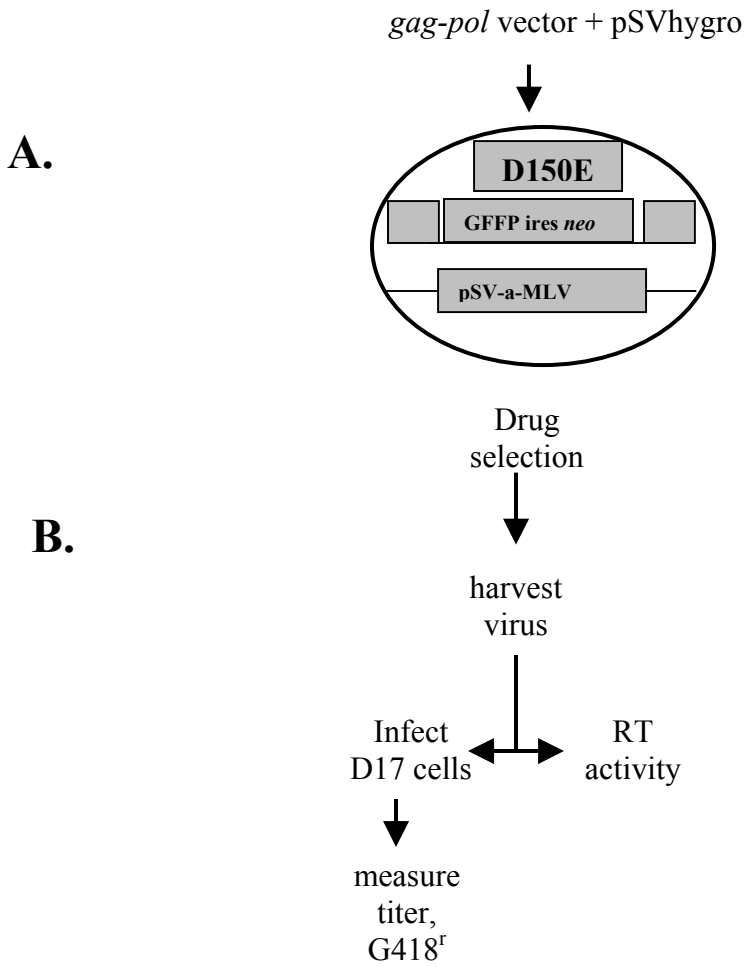


Table 3. Quantification of mature vs. immature virion particles in IN HHCC insertion mutants.

Virus	# immature particles/ # mature particles
MP5	28/72
LC2	38/62
SR8	57/43

Figure 3.

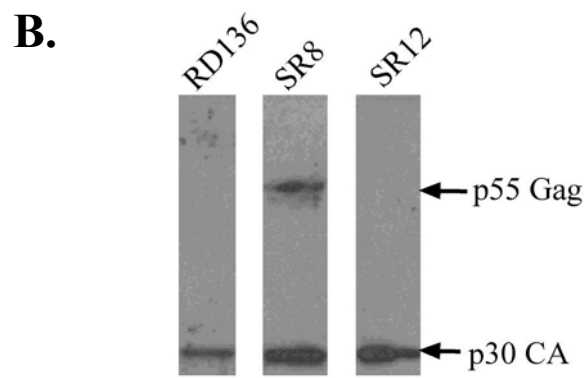
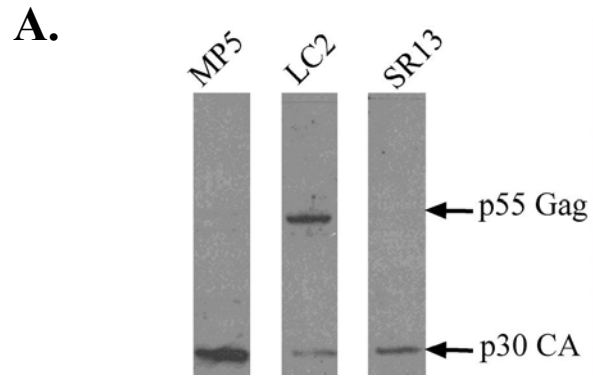


Table 2. Replication Properties of SNV and MLV *Pol* Mutants

Vector	Titer, cfu/mL ^a ±S.E.	Rel. RT Activity ^a ±S.E. % wt	Real Time PCR of Reverse Transcription Products			
			# R-U5 (%)±S.E.	# neo (%)±S.E.	% R-U5 normalized for RT Activity±S.E.	% Neo normalized for RT Activity±S.E.
SNV						
RD136	2.3± ^d 1.3×10 ³	100	100 ^b	100 ^b	100	100
YVNN	<1	4.9±4.8	0.2 ^c	0.1 ^c	4	2
WZH73	2.3±2	272±333 ^e	23±15 ^b	7±0.8 ^b	8±5	2.6±0.4
SR8	<1	10±6.6	8.7±4.5 ^b	9.4±5 ^b	80±45	93±49
SR12	<1	425±340	310±80 ^a	240±44 ^a	72±19	56±10
MLV						
MP5	6.6± 0.6×10 ³	100	100 ^a	100	100	100
LC2	<1	23±9	18±5 ^f	21±12	76±22	89±52
SR13	<1	35±18	17±7 ^a	13±9	49±19	36±25

^a Average of three independent experiments

^b Average of two independent experiments. Each reaction was performed three times.

^c One experiment. Each reaction was performed three times.

^d Six independent experiments.

^e One measurement out of three was significantly higher than the wild type. The other two measurements were equivalent to wild type. See Appendix for table of individual measurements of all mutants.

^f Average of two independent experiments

Table 4. Complementation of SNV and MLV *Pol* Phenotypically Mixed Virions Derived from the SNV-*gag* D150E Cell Line^a

Vector ^{b,c}	RT Activity, % wt ±S.E.	Titer, cfu/mL ±S.E.	% WT Titer ±S.E.	Titer Normalized for RT Activity ±S.E.	Normalized titer, % WT ±S.E.
SNV					
RD136	100	223±0.25×10 ²	100	223×10 ²	100
WZH73	99±21	5.3±4 × 10 ²	2.4	5.4×10 ²	2.4
SR8	8±4.5	1.2±0.6 × 10 ²	0.5	1.5×10 ³	7
SR12	101±24	1.2±0.4 × 10 ²	0.5	1.2×10 ²	0.5
MLV					
MP5	100	256±0.6 × 10 ²	100	256×10 ²	100
LC2	55±9	0.8±0.3 × 10 ²	0.3	1.5×10 ²	0.5
MP5	100	19±13 × 10 ²	100	19×10 ²	100
SR13	170±56	0.1±.07 × 10 ²	0.5	.06×10 ²	0.3

^a Averages of 3 independent experiments are shown.

^b RD136, SNV*gag-pol* wild type. WZH73, SNV*gag-pol* RNase H D524N mutant. SR8, SNV*gag-pol* IN HHCC insertion mutant. SR12, SNV*gag-pol* IN catalytic core mutant.

^c MP5, SNV*gagMLVpol* wild type. LC2, SNV*gagMLVpol* IN HHCC insertion mutant. SR13, SNV*gagMLVpol* IN D125A catalytic core mutant.

Table 5. Complementation of SNV *Pol* Mutants in the SNV*gag*-YVNN Cell Line

Vector	RT Activity, % wt	Titer, cfu/mL	% wt titer	Titer normalized for RT activity	% Wild Type titer, normalized for RT
RD136	100	120×10^2	100	120×10^2	100
WZH73	68	1.5×10^2	1.3	2.2×10^2	1.8
SR8	2	<1	-	-	-
SR12	88	0.3×10^2	0.2	0.34×10^2	0.3
MP5	100	1.3×10^2	100	130	100
LC2	30	<1	-	-	-
SR13	140	0.6×10^2	47	130	100

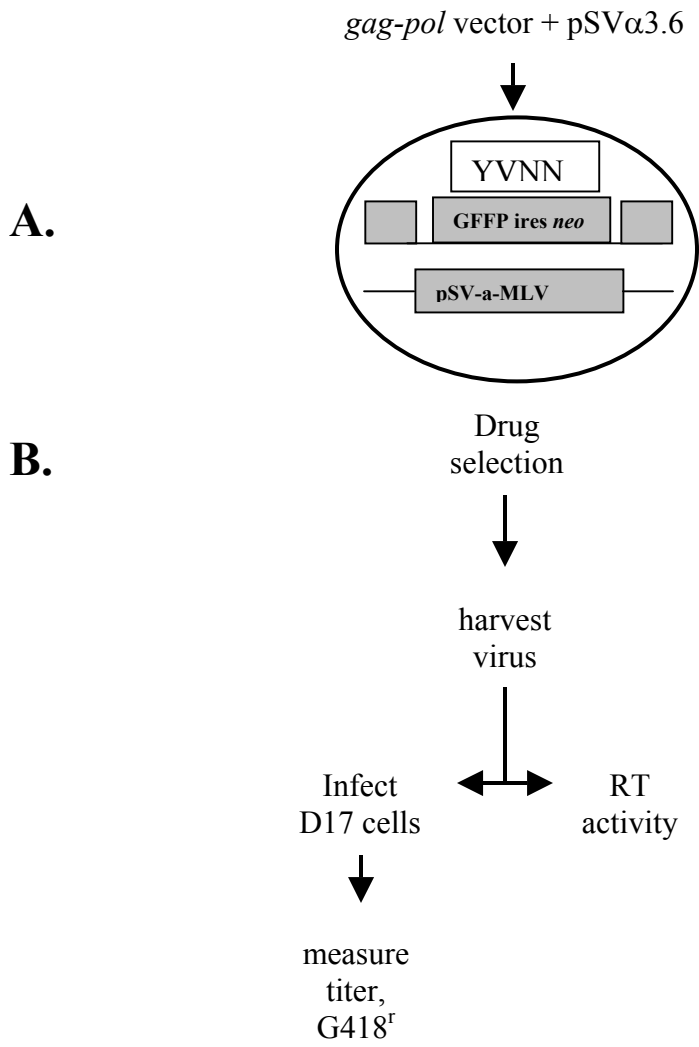
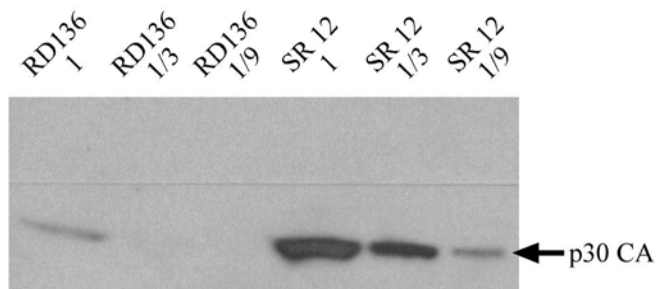


Figure 4.



Appendix Figure 1.

REFERENCES

1. **Ansari-Lari, M. A., L. A. Donehower, and R. A. Gibbs.** 1995. Analysis of human immunodeficiency virus type 1 integrase mutants. *Virology* **213**:680.
2. **Aubin, R. A., M. Weinfeld, R. Mirzayans, and M. C. Paterson.** 1994. Polybrene/DMSO-assisted gene transfer. Generating stable transfectants with nanogram amounts of DNA. *Mol. Biotechnol.* **1**:29-48.
3. **Bukovsky, A., and H. Gottlinger.** 1996. Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J. Virol.* **70**:6820-5.
4. **Cannon, P. M., W. Wilson, E. Byles, S. M. Kingsman, and A. J. Kingsman.** 1994. Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. *J. Virol.* **68**:4768-75.
5. **Certo, J. L., B. F. Shook, P. D. Yin, J. T. Snider, and W. S. Hu.** 1998. Nonreciprocal pseudotyping: murine leukemia virus proteins cannot efficiently package spleen necrosis virus-based vector RNA. *J. Virol.* **72**:5408-13.
6. **Engelman, A.** 1999. In vivo analysis of retroviral integrase structure and function. *Adv Virus Res* **52**:411-26.
7. **Engelman, A., and R. Craigie.** 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.* **66**:6361-9.

8. **Gottlinger, H. G., J. G. Sodroski, and W. A. Haseltine.** 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **86**:5781-5.
9. **Gritz, L., and J. Davies.** 1983. Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**:179-88.
10. **Halvas, E. K., E. S. Svarovskaia, and V. K. Pathak.** 2000. Role of murine leukemia virus reverse transcriptase deoxyribonucleoside triphosphate-binding site in retroviral replication and in vivo fidelity. *J. Virol.* **74**:10349-58.
11. **Hayman, M.** 1978. Synthesis and processing of avian sarcoma virus glycoproteins. *Virology* **85**:475-86.
12. **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.
13. **Jamjoom, G. A., R. B. Naso, and R. B. Arlinghaus.** 1977. Further characterization of intracellular precursor polyproteins of Rauscher leukemia virus. *Virology* **78**:11-34.
14. **Julias, J. G., A. L. Ferris, P. L. Boyer, and S. H. Hughes.** 2001. Replication of phenotypically mixed human immunodeficiency virus type 1 virions containing catalytically active and catalytically inactive reverse transcriptase. *J. Virol.* **75**:6537-46.

15. **Katoh, I., Y. Yoshinaka, A. Rein, M. Shibuya, T. Odaka, and S. Oroszlan.** 1985. Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145**:280-92.
16. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* **12**:2331-8.
17. **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.
18. **Leavitt, A. D., G. Robles, N. Alesandro, and H. E. Varmus.** 1996. Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. *J. Virol.* **70**:721-8.
19. **Leavitt, A. D., L. Shiue, and H. E. Varmus.** 1993. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J. Biol. Chem.* **268**:2113-9.
20. **Luftig, R. B., and Y. Yoshinaka.** 1978. Rauscher leukemia virus populations enriched for "immature" virions contain increased amounts of P70, the gag gene product. *J. Virol.* **25**:416-21.
21. **Oppermann, H., J. M. Bishop, H. E. Varmus, and L. Levintow.** 1977. A joint produce of the genes gag and pol of avian sarcoma virus: a possible precursor of reverse transcriptase. *Cell* **12**:993-1005.

22. **Peng, C., B. K. Ho, T. W. Chang, and N. T. Chang.** 1989. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J. Virol.* **63**:2550-6.
23. **Roth, M. J., P. Schwartzberg, N. Tanese, and S. P. Goff.** 1990. Analysis of mutations in the integration function of Moloney murine leukemia virus: effects on DNA binding and cutting. *J. Virol.* **64**:4709-17.
24. **Sambrook, J., E.F. Fritsch, and T. Maniatis.** 1989. *Molecular Cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. **Stewart, L., G. Schatz, and V. M. Vogt.** 1990. Properties of avian retrovirus particles defective in viral protease. *J. Virol.* **64**:5076-92.
26. **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.
27. **Telesnitsky, A., and S. P. Goff.** 1993. Two defective forms of reverse transcriptase can complement to restore retroviral infectivity. *Embo J.* **12**:4433-8.
28. **Voynow, S. L., and J. M. Coffin.** 1985. Truncated gag-related proteins are produced by large deletion mutants of Rous sarcoma virus and form virus particles. *J. Virol.* **55**:79-85.
29. **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.

30. **Yoshinaka, Y., and R. B. Luftig.** 1978. Morphological conversion of 'immature' Rauscher leukaemia virus cores to a 'mature' form after addition of the P65-70 (gag gene product) proteolytic factor. *J. Gen Virol.* **40**:151-60.
31. **Zhang, W. H., C. K. Hwang, W. S. Hu, R. J. Gorelick, and V. K. Pathak.** 2002. Zinc Finger Domain of Murine Leukemia Virus Nucleocapsid Protein Enhances the Rate of Viral DNA Synthesis in Vivo. *J. Virol.* **76**:7473-7484.

This page intended to be left blank.

THIS PAGE INTENDED TO BE LEFT BLANK.

CHAPTER FOUR

Charged assembly helix motif in murine leukemia virus capsid: An important region for virus assembly and particle size determination

Sara Rasmussen Cheslock^{1,2,#}, Dexter T.K. Poon^{1,#}, William Fu¹, Terence D. Rhodes^{1,2}, Louis E. Henderson³, Kunio Nagashima⁴, Connor F. McGrath⁵, and Wei-Shau Hu^{1,*}

¹HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702,

²Department of Microbiology, Immunology, and Cell Biology, School of Medicine, West

Virginia University, Morgantown, WV, 26506; and ³AIDS Vaccine Program, ⁴Image

Analysis Laboratory, and ⁵Structure-Based Drug Discovery Group, SAIC-Frederick,

Frederick, MD 21702

SRC and DTKP contributed equally to this work.

Running Title: charged assembly helix motif of MLV CA

* Corresponding author. Mailing address: HIV Drug Resistance Program, National Cancer Institute-Frederick, Bldg.# 535, Rm. 336, Frederick, MD 21702. Phone: (301) 846-1250. Fax: (301) 846-6013. E-mail: whu@ncifcrf.gov

ABSTRACT

We have identified a region near the C terminus of capsid (CA) of murine leukemia virus (MLV) that contains many charged residues. This motif is conserved in various lengths in most MLV-like viruses. One exception is that spleen necrosis virus (SNV) does not contain a well-defined domain of charged residues. When 33 amino acids from the MLV motif were deleted to mimic SNV CA, the resulting mutant produced drastically reduced amounts of virions and the virions were non-infectious. Furthermore, these viruses had abnormal sizes, often contained punctate structures resembling those in the cell cytoplasm, and packaged both ribosomal and viral RNA.

When 11 or 15 amino acids were deleted to modify the MLV CA to resemble those from other gammaretroviruses, the deletion mutants produced virions at levels comparable to the wild type virus and were able to complete one round of virus replication without detectable defects. We generated 10 more mutants that displayed either the wild-type or mutant phenotype. The distribution of the wild type or mutant phenotype did not directly correlate with the number of amino acids deleted, suggesting that the function of the motif is not simply determined by its length but by its structure. Structural modeling of the wild-type and mutant proteins suggested that this region forms α -helices; thus, we termed this motif charged assembly helix. This is the first description of the charged assembly helix motif in MLV CA and the demonstration of its role in virus budding and assembly.

INTRODUCTION

All retroviruses encode at least the *gag*, *pro*, *pol*, and *env* genes, which are translated into polyproteins and subsequently cleaved into individual virion proteins (38). The *gag* gene products form the structure of the virions, *pro* gene products process the viral proteins, *pol* gene products carry out important enzymatic reactions, and *env* gene products serve as the viral surface envelope (38). Although all four gene products are essential in the generation of wild-type, infectious virions, the *gag* gene product, Gag polyprotein, drives the formation of the virus particles. When Gag polyproteins are expressed in host cells, they can form virus-like particles without other retroviral components such as *pro*, *pol*, or *env* gene products or viral RNA containing the retroviral packaging signals (35, 41). Purified Gag polyproteins, or portions of the polyproteins, can also form virus-like particles *in vitro* (6, 7, 13, 19). Together, these data indicate that Gag polyproteins are necessary and sufficient for the formation of the virus-like particles.

Gag polyproteins from all retroviruses contain at least three domains that are later cleaved by viral protease into mature proteins: matrix (MA), capsid (CA), and nucleocapsid (NC) (38). Most retroviruses also contain other cleavage products that vary in size and in the location of the domain on the polyprotein. For example, murine leukemia virus (MLV) Gag has an additional 12-kDa cleavage product termed p12; the p12 domain is located between MA and CA in the Gag polyprotein. In human immunodeficiency virus type 1 (HIV-1), Gag polyproteins have three additional cleavage products, p2, p1, and p6. The p2 domain is located between CA and NC whereas p1 and p6 are located C-terminal to NC in the Gag polyprotein (32).

Different domains of Gag polyproteins contain distinct signals to direct virus assembly and budding. Signals in MA target Gag to the cell membrane, motifs in the NC domains mediate RNA binding, and motifs in CA and NC are important in the multimerization of Gag polyproteins (35, 38). After formation of the virus-like particles, host cell factors recruited by the late domains, which are located at various positions in Gag of different viruses, interact with the cell membrane to allow the release of the viral particles (16). During or soon after virus budding, viral proteases are activated and cleave the polyproteins into smaller proteins (35). Upon protease cleavage, the virions undergo morphological changes and become mature particles (35). The mature particle, and not immature particle, contains an electron-dense core, which can be identified during electron microscopy (EM) (38).

CA is involved in multiple stages in the retroviral replication cycle. As a cleaved, mature viral protein, CA forms the condensed core of the mature virus particle and constitutes part of the reverse transcription complex in certain viruses (4, 14, 38). As a part of the Gag polyprotein, CA plays a role in virus assembly. This is supported by the genetic studies, in which mutations in CA often lead to assembly defects (15, 20, 35). However, the exact mechanism of CA's involvement in virus assembly remains unclear. Structural analyses of CA from several viruses illustrate that these proteins consist of two α -helix-rich domains, N-terminal and C-terminal domain, connected by a flexible linker region (3, 18, 23, 24, 39). The primary sequences of the CA from different retroviruses are divergent; the most conserved region in CA is located in the C-terminal domain (QX₃EX₇R), and is termed the major homology region (MHR) (41). Mutations in this region often lead to defects in the viral replication cycle (2, 5, 10, 28).

The amino acid sequence at the C-terminal region of MLV CA has a stretch of 41 amino acids rich in charged residues (Fig. 1A). Of these 41 amino acids, 33 are charged residues with similar numbers of acidic and basic residues (16 and 17, respectively). The function of this motif rich in charged residues is unknown. In this report, we examined the effects of mutations on this motif, and found that disruption of this motif could result in assembly defects. Furthermore, mutants lacking functional motifs generated virions with highly varied particle sizes and packaged both ribosomal and viral RNAs. These results indicated that this motif rich in charged residues is an important region in the assembly of MLV particles.

MATERIALS AND METHODS

Plasmids and construction of mutant *gag-pol* expression constructs. All vectors were constructed using standard cloning procedures (34). Mutations were generated using overlapping PCR and cloning of the resulting DNA fragments into the MLV-based *gag-pol* expression construct pWZH30 (42), a generous gift from Vinay Pathak. Plasmid pWZH30 is derived from pLGPS (30), which expresses *gag-pol* from a truncated 5' long terminal repeat (LTR) and does not contain the packaging signal or the 3' LTR. The overall structures of the mutant *gag-pol* constructs were determined by restriction enzyme analyses; the portions of DNA generated by PCR were further characterized by DNA sequencing. Detailed cloning procedures are available upon request. Plasmids pEW22U and pEW22D also contained an additional H114R mutation in the capsid region; this mutation was located 108 amino acids N-terminal to the motif described in this report.

Plasmid pSV-a-MLV-env expresses amphotropic MLV *env* (27). MLV-based vector pSR2 (8) contains all the *cis*-acting elements essential for replication; pSR2 does not express viral proteins, but expresses hygromycin phosphotransferase B gene (*hygro*), which confers resistance to hygromycin (22). pSR2 also has two modified LTRs, both of which contain a copy of the green fluorescent protein gene.

Cell culture, DNA transfection, virus propagation, and virus infection. D17 cells are dog osteosarcoma cells permissive to amphotropic MLV infection (33); 293T cells are human embryonic kidney cells (12). Both cell lines were maintained at 37°C with 5% CO₂ in Dulbecco's modified Eagle's medium supplemented with serum and antibiotics (8, 12). DNA transfection was performed by the calcium phosphate precipitation method (34) using the CalPhos transfection kit (ClonTech) or MBS transfection kit (Stratagene) with a mixture of DNA containing the *gag-pol* expression construct, pSR2, and pSV-a-MLV-env at a 2:2:1 weight ratio.

Ten-fold dilutions of the viruses were used to infect D17 cells in the presence of Polybrene (50 µg/ml, final concentration). Infected cells were placed on hygromycin selection at a final concentration of 120 µg/ml, and virus titers were determined by the numbers of hygromycin-resistant colonies. In general, viral titers measured in these assays agreed with the reverse transcriptase (RT) activities. The virus titers presented were not standardized to RT activities because some of the mutants had low RT activities, which were not significantly higher than background.

RT Assay, western analyses, and EM analyses. Viruses were harvested and concentrated by centrifugation at 25,000 rpm for 90 min at 4°C using a Surespin 630

rotor (Sorvall) in a Discovery 100S ultracentrifuge (Sorvall). RT activities of the viral samples were determined using standard procedures (8, 36). Proteins from cell lysates or virus lysates were examined by western blotting analyses performed using standard procedures (34). Polyclonal rabbit anti-MLV-CA and anti-MLV-MA antibodies were kind gifts from the AIDS Vaccine Program (SAIC, NCI-Frederick).

For EM analyses, transfected 293T cells were washed with phosphate-buffered saline, pelleted by low-speed centrifugation, and fixed in 2% glutaraldehyde in cacodylate buffer (0.1 M, pH 7.4). EM analyses were performed as previously described (37) (Hitachi H7000) and images were captured by a digital camera system (Gatan).

RNA analyses. To examine the RNA encapsidated in wild type and mutant viruses, *gag-pol* expression constructs were used to separately transfect SR2-293T cells. SR2-293T cells are 293T cells containing integrated SR2 proviruses. These cells were generated by infecting 293T cells with SR2-containing virions, subjecting the infected cells to hygromycin selection, and pooling the resulting hygromycin-resistant cell colonies. Cell-free viruses were harvested from transfected SR2-293T cells, virion RNA were isolated using previously described protocols (17). The amounts of encapsidated SR2 RNA and ribosomal RNA were determined using quantitative real-time RT-PCR (ABI 7700 sequence detector). Primers and probe sequences derived from *hygro* were used to detect SR2 vector RNA. The primers and probe have the following sequences: 5' ACGAGGTCGCCAACATCTTC 3' (primer), 5' AGCGCGTCTGCTGCTCC 3' (primer), and 5' ^{FAM}TCTGGAGGCCGTGGTTGGCTTGTA^{TAMRA} 3' (probe). Primers and probe derived from 18S ribosomal RNA gene were used to detect 18S ribosomal RNA. The primers and probe have the following sequences: 5' GCCGCTAGAGGTGAAATTCTTG

3' (primer), 5' CATTCTTGGCAAATGCTTTTCG 3' (primer), and 5'
FAM^{FAM}ACCGGCGCAAGACGGACCAGA^{TAMRA} 3' (probe). RNA transcripts containing
hygro sequences and purified rRNAs were used as templates to generate standard curves
in the real-time RT-PCR for SR2 and rRNA detection, respectively. RNA isolated from
mock-transfected samples was used as negative control.

RESULTS

A motif rich in charged residues is conserved in the CA domain of most gammaretroviruses. To determine whether the motif rich in charged residues is unique to MLV CA, the primary amino acid sequences of CA from multiple retroviruses were compared. Most of the gammaretroviruses have similar motifs, albeit with different lengths. A notable exception is spleen necrosis virus, which does not have a region with many charged residues in its CA (Fig. 1B). Several viruses, including gibbon ape leukemia virus (GaLV), have motifs that are 11 amino acids shorter than that of MLV. Rat leukemia virus (RaLV) and feline leukemia virus (FeLV) both have motifs that are 15 amino acids shorter than that of MLV (Fig. 1B). All of the motifs rich in charged residues have nearly balanced numbers of acidic and basic residues.

We also examined the primary sequences of several other groups of more distant retroviruses and found that the CA proteins of these viruses do not have motifs rich in charged residues at the C-terminal region of their CA and sequences between CA-NC (not shown). These viruses include Rous sarcoma virus (RSV), an alpharetrovirus; Mason-Pfizer monkey virus, a betaretrovirus; human T cell leukemia virus type 1, a deltaretrovirus; and human immunodeficiency virus type 1 (HIV-1), a lentivirus.

The motif rich in charged residues is essential for efficient MLV replication and assembly. To test whether this motif is essential for MLV replication and to examine its possible roles, we constructed and characterized a mutant containing a large deletion that encompassed almost the entire motif. Plasmid pSR33 was derived from the MLV *gag-pol* expression construct pWZH30 (42) and contained a 33-amino-acid deletion in the motif (Fig. 2A). To examine the phenotype of pSR33, we used the following protocols. The MLV *gag-pol* expression construct pWZH30 or pSR33 was introduced into 293T cells by transfection along with pSV-a-MLV-env (27) and pSR2 (8). pSV-a-MLV-env expresses amphotropic MLV *env*. pSR2 is an MLV-based vector that does not express *gag-pol* or *env*, but instead expresses *hygro* (22), which confers resistance to hygromycin. Supernatants were harvested from the transfected 293T cells, a portion of the sample was used to infect target D17 cells to determine the virus titer, and the other portion was subjected to biochemical analyses. Infected D17 cells were subjected to hygromycin selection and the SR2 virus titers were determined by the number of hygromycin-resistant cell colonies. The amounts and properties of the viral protein released in the cell-free supernatants were determined by western blotting analyses and RT assays, and the morphology of the released virions was examined by EM.

Viral titers generated from six sets of independent experiments are summarized in Table 1. Supernatants from pWZH30-transfected cells on average generated a titer of 6.0×10^3 hygromycin-resistant colony forming units (CFU)/ml. However, supernatants harvested from pSR33-transfected cells did not generate detectable titers in any of the six experiments.

Representative analyses of viral proteins from transfected cells and harvested supernatants are shown in Fig. 3. Virions generated from pWZH30-transfected cells are referred to as wild type. Virions were released from pSR33-transfected cells at an approximately 30- to 100-fold reduced level compared with that of wild type (Fig. 3A and data not shown). Western analyses also showed that Gag polyproteins expressed by pSR33 were mostly processed; CA from pSR33-derived virions migrated faster than CA from wild-type viruses, consistent with the 33-amino-acid deletion, whereas MA from both viruses had the same mobility. Western analyses from transfected cell lysates showed that Gag polyproteins were expressed in pWZH30- and pSR33-transfected cells at similar levels (Fig. 3B), indicating that the 33-amino-acid deletion did not significantly affect the amount of Gag polyprotein expressed in the transfected cells. Therefore, viruses generated from pSR33-transfected cells had a defect in virus assembly and/or release.

To examine the morphology of the viruses generated from the pWZH30- or pSR33-transfected cells, EM analyses were performed. These analyses indicated that pSR33-derived viruses differed from wild-type viruses in both size and morphology. Representative images of the wild-type and pSR33-derived virions are shown in Fig. 5A-D; all of these images are shown with the same magnification. Wild-type virions were close to 100 nm in diameter, and both immature particles and mature particles with condensed cores were observed (Fig. 4A and data not shown). In contrast, pSR33-derived virions were often much larger than 100 nm, and all appeared to be immature (Fig. 4 B-D and data not shown). In addition, punctate structures similar to those in the cell cytoplasm were observed in these virions. Virions with a wide range of particle sizes

containing punctate structures were observed only in pSR33-transfected cells, not in untransfected cells or pWZH30-transfected cells (data not shown).

To further quantify the distributions of virion particle sizes, we measured the sizes of 100 wild-type virions and 100 pSR33-derived virions (Table 2). The particle sizes of the wild-type virions were tightly clustered; of the 100 virions, 98 were 90-130 nm in diameter. In contrast, the particle sizes of pSR33-derived virions were distributed through a wide range. Of the 100 pSR33-derived virions, only 31 were 90-130 nm in diameter; the other 69 viral particles were smaller or larger than this size range, with some viruses more than 400 nm in diameter.

The electron micrographs in Fig. 4C and D captured the images of pSR33-derived viruses in the process of budding out of the cells, and the contents of the mutant virions bear a striking resemblance to the cytoplasm. This observation led us to examine the RNA content of pSR33-derived virions. To avoid the potential complication caused by contamination of transfected MLV vector pSR2 DNA, we used 293T cells containing integrated SR2 proviruses (referred to as SR2-293T cells) in these experiments. SR2-293T cells were transiently transfected with pWZH30 or pSR33, and cell-free viruses were harvested. For each virus sample, an aliquot was used to isolate RNA and the other aliquot was used to generate lysates for western analyses. The amounts of viruses in these samples were determined by measuring the quantity of MA in cell-free viral lysates using western analyses on serially diluted lysates. The amounts of SR2 RNA and 18S ribosomal RNA in each sample were determined using quantitative real-time RT-PCR with probes and primers derived from *hygro* and 18S ribosomal RNA sequences, respectively. Data summarized from three sets of independent experiments are shown in

Fig. 5A and B. Virions derived from pSR33 encapsidated SR2 RNA at an efficiency of 79% of the wild-type virions. More strikingly, pSR33-derived virions packaged 3100% of the 18S ribosomal RNA relative to wild-type virions. To ensure that the RNA measured in these experiments was derived from virions rather than from cellular debris, we have also measured the amounts of SR2 RNA and 18S ribosomal RNA in the supernatants from mock-transfected cells, and compared with the amounts of RNA detected in the same volume of supernatant from pWZH30-transfected cells. In two independent experiments, we were not able to detect SR2 RNA (less than 0.01% of the signals from wild-type virus) and we detected low amounts of ribosomal RNA (an average of 1.6% of the amount detected from wild-type virus controls) in the supernatant from mock-transfected. These results confirmed that pSR33-derived virions contained ribosomal RNA; therefore, the punctate structures inside the virions observed by EM were most likely ribosomes.

Taken together, these data indicate that deletion of the motif rich in charged residues could result in defects in virus assembly, which would affect both the numbers of virions released from the cells and the properties of the released viral particles.

The MLV motif rich in charged residues can tolerate relatively large deletions: phenotypes of mutants with motifs mimicking those of other gammaretroviruses. As shown in Fig. 1B, many gammaretroviruses have shorter motifs compared with that of MLV. To test whether a shorter motif can function in the context of MLV, we generated three MLV-based mutants containing motifs that were the same lengths as those from GaLV, RaLV, and FeLV. These mutants (pSR11, pSR15, and pFeEW) contained 11-, 15-, and 15-amino-acid deletions, respectively (Fig. 2B). The

phenotypes of these mutants were examined using the same protocol as described above. Western analyses of the supernatants from cells transfected with these plasmids indicated that the levels of virions released and the processing of Gag polyproteins were similar to those of the wild type (Fig. 6A). Virions from these deletion mutants were infectious, generating titers similar to those from wild-type virus (Table 1). EM analyses indicated that these virions had the wild-type virion morphology (data not shown). These data established that these deletion mutants were phenotypically similar to wild-type viruses. Therefore, the functional requirement of the motif rich in charged residues could be fulfilled by shorter-than-wild-type sequences.

The function of the motif rich in charged residues is not solely determined by its length and primary sequences. Because the virions generated from pSR11-, pSR15-, and pFeEW-transfected cells were similar to wild-type viruses in all aspects examined, we questioned whether the MLV motif could tolerate even larger deletions. To investigate this question, we generated mutants that contained 17-, 19-, and 21-amino-acid deletions in the motif rich in charged residues (pEW17, pEW19, and pEW21, respectively) (Fig. 2C). To maintain the balance in acidic and basic residues, we used pSR15 as a template and deleted equal numbers of acidic and basic residues in these mutants. The phenotypes of these constructs were examined using the protocols described above. Western analyses indicated that pEW17- and pEW21-transfected cells released very little virus in their supernatants (Fig. 6B). In contrast, pEW19-transfected cells released amounts of virus similar to those of the wild type into their supernatants (Fig. 6B); furthermore, the Gag polyproteins from pEW19-generated viruses were processed similarly to the wild-type viruses (Fig. 6B). The reduction of the virion

production in pEW17- and pEW21-transfected cells was not due to lack of Gag expression, as indicated by western analyses of the cell lysates (Fig. 7C). EM studies of these samples revealed that viruses generated from pEW17 resembled those from pSR33, whereas pEW19 generated virions with morphology similar to that of wild-type viruses (data not shown). In four to six sets of independent infection experiments, pEW17 and pEW21 did not generate detectable titers; however, pEW19 and pWZH30 generated similar titers (Table 1). Phenotypically wild-type virions were generated by pEW19 but not by pEW17, which contained a smaller deletion. This result indicates that the length of the primary sequence is not a sufficient criterion for predicting the function of the mutant motif and suggests that a secondary or tertiary structure is important to the function of the motif rich in charged residues.

Testing the hypothesis that the MLV motif rich in charged residues forms a functionally important α -helical structure. Deletion analyses implied a periodicity to the correlation between the numbers of amino acids deleted and the phenotypes of the resulting mutants. For example, deletion of 11, 15, or 19 amino acids resulted in functional proteins, but deletion of 17 or 21 amino acids did not. The differences in the numbers of amino acids among pSR11, pSR15, and pEW19 are four and eight, which approximate one and two complete turns in α helices, respectively. In addition, analyses of the primary sequences of the MLV motif indicate that this region may form an α -helical structure (Chou-Fasman prediction, PepPlot, and PeptideStructure, GCG program, University of Wisconsin). We hypothesized that the MLV motif forms an α -helical structure and that maintenance of the overall structure and preservation of the helical phases are important to the function of this motif. We also hypothesized that deletion of

one or more complete turns of the α -helices can allow the preservation of the structure and can lead to the production of phenotypically wild-type virions, whereas deletions of incomplete turns could drastically alter the properties of the α -helices and could result in phenotypically altered virions similar to those generated by pSR33. To test this hypothesis, we generated additional mutants and examined their phenotypes by using the same protocols as described above.

A 13-amino-acid deletion should delete approximately 3.6 turns of the α -helices and should result in the formation of phenotypically altered viruses. Similarly, deletion of 20 amino acids (5.6 turns) should also result in an altered phenotype. Therefore, we generated pEW13 and pEW20 containing 13- and 20-amino-acid deletions in the motif, respectively (Fig. 2D), and we examined their properties. As predicted, western analyses indicated that these two mutants released drastically reduced levels of virions (Fig. 7A), and these virions generated titers less than 1 CFU/ml in two to five independent experiments (Table 1). Western analyses of cell lysates indicated that Gag polyproteins were expressed in pEW13- and pEW20-transfected cells (Fig. 7B).

Deletion of 18 and 22 amino acids of the motif should remove 5 and 6.1 turns of the α -helices, respectively, which should result in functional motifs. Two 18-amino-acid deletion mutants were generated based on pEW17: one (pEW18E) had an acidic residue deleted; the other (pEW18R), a basic residue deleted (Fig. 2D). Two 22-amino-acid deletion mutants were also generated based on pEW21: one (pEW22U) had a glutamic acid deleted and the other (pEW22D) had a histidine deleted (Fig. 2D). Our results indicated that all four of these mutants had wild-type phenotypes in the amounts of virus released (Fig. 7A) and generated infectious viruses with titers similar to those from wild-

type virus (Table 1). These data also demonstrated that we were able to restore the infectivity of the virions by deleting one amino acid in the motif from the two parental constructs pEW17 and pEW21, which had generated non-infectious virions with mutant phenotypes. Different numbers of acidic and basic residues were deleted in pEW18E and EW18R. In pEW18R, 8 acidic and 8 basic residues were deleted; whereas in pEW18E, 7 acidic and 9 basic residues were deleted. Since both constructs generated infectious virions, these data suggested that the motif could tolerate minor fluctuations in the numbers of charged residues. pEW22U and pEW22D also contained an additional mutation elsewhere in the capsid region (H114R). This mutation probably did not affect capsid function, since viruses derived from these two plasmids had the wild-type phenotype.

We also hypothesized that the individual amino acids may not affect the function of the motif as much as the maintenance of the helical phases. We postulated that the mutant phenotype of pEW17-derived virions was caused by alteration of the helical phase and not by the loss of charged residues. To test this hypothesis, we generated pEW15-res, in which two alanines were inserted into pEW17 to restore the phase of the helices (Fig. 2D). Western analyses and viral titers data indicated that pEW15-res-derived viruses had a wild-type phenotype (Fig. 7A, Table 1).

Taken together, these data strongly support the idea that the MLV motif rich in charged residues forms an α -helical structure; this structure and the phases of the helices play important roles during virus assembly.

DISCUSSION

In this report, we describe a novel motif in MLV CA that is rich in charged residues, affects virus assembly, and most likely forms a helical structure. Thus, we termed this motif the charged assembly helix. Many deletions in this motif cause an assembly defect with a new phenotype, i.e., virions with a wide range of particle sizes and contents that morphologically resemble the cell cytoplasm; in addition, these virions package both ribosomal and viral RNAs. Virus with the charged assembly helix motif mutant phenotype exhibits drastically reduced particle production and even more greatly reduced infectivity. We were able to capture EM images with viruses budding out of the cells (Fig. 4C and D). However, the charged assembly helix mutant phenotype is distinct from the late-domain mutants that have virions tethered on the cell surface or viral particles connected to one another via uncleaved membranes (16, 21, 40).

In a previous study, MLV NC mutants that could not efficiently package viral RNA were shown to package ribosomes (31). The phenotype of the charged assembly helix mutants differs from that of the NC mutants in two aspects: the packaging of both ribosomal and viral RNA, and the widely varied particle sizes. At least two possibilities can explain why charged assembly helix mutants but not NC mutants can package both ribosomal and viral RNA. One possibility is that particle sizes play a critical role in the encapsidation of both types of RNA. Compared with a 100-nm-diameter virion, virions with a 200-, 300-, or 400-nm diameter have 8, 27, and 64 times more volume in the particle, respectively. The larger particles in charged assembly helix mutants might simply have more volume in which to accommodate both ribosomal and viral RNAs, whereas the 100-nm-diameter NC mutants or wild-type viruses do not. The other

possibility is that the charged assembly helix mutants have lost a function(s) that prevents the incorporation of ribosomal RNA into the virion. It was previously proposed that one of the roles of NC in wild-type viruses is to exclude ribosomes (31). It is possible that Gag-RNA interaction has two distinct functions: to specifically package viral RNA, and to exclude ribosomes and possibly other cytoplasmic elements. NC mutants might have lost both functions, whereas the charged assembly helix mutants have lost only the ability to exclude ribosomes and other elements. Further studies will be needed to distinguish these possibilities.

Wild-type MLV virions are fairly uniform in size (Table 2), indicating that Gag multimerization is a regulated event. One of the most striking features of the charged assembly helix mutant phenotype is the wide range of virion particle size distribution, indicating defects in this regulation. In the MLV Gag polyprotein, CA and NC are immediately adjacent to each other and the charged assembly helix motif is located near their junction. Substitution mutations in the CA-NC junction in MLV can cause defects in virus assembly (A. Rein, personal communication). In some retroviruses, CA and NC are separated by other cleavage products, e.g., spacer peptides (SP) for RSV and p2 for HIV-1 (32). Extensive mutational analyses revealed that the RSV CA-SP region is critical in the regulation of virion particle sizes; deletions in this region caused the formation of virions with heterogeneous sizes (26). In HIV-1, mutations in p2 have been shown to cause assembly defects that generate virions with heterogeneous sizes and aberrant morphology (25). Interestingly, the CA-p2 junction was proposed to be an α -helical structure that is crucial for virus assembly (1). Together with data from this report, these results suggest that the regions near CA-NC junctions are important in virus

assembly and particle size regulation in many retroviruses. An intriguing possibility is that the charged assembly helix motif, SP, and p2 fulfill similar functions during retroviral assembly.

Structural studies of the C-terminal domain of HIV-1 CA revealed the presence of four α -helices; these helices play critical roles in maintaining the overall globular structure of the region as well as promoting CA-CA dimerization (18). However, the structure of the last 11 amino acids of HIV-1 CA and the putative α -helix in p2 is currently unknown, and their roles in Gag-Gag interactions during assembly are not determined. Although NMR studies were performed on peptides derived from MLV MHR (9); similar to HIV-1, the structure of the very C-terminus of MLV CA, including the charged assembly helix motif, remains unknown. We performed molecular modeling to simulate the putative structure of the MLV charged assembly helix motif (Fig. 8). The charge distribution in the wild-type motif was such that positively and negatively charged bands (shown in blue and red, respectively) were formed; in addition, a basic-residue cluster and a neutral cluster (shown in white) were located near the C terminus of CA. We observed that mutants with deletions of complete α -helical turns had the wild-type phenotypes, whereas mutants with deletions of incomplete turns, which drastically altered the helical phases, had the mutant phenotype. The effects of deleting incomplete turns can be visualized in the modeled structures. For example, in pSR11 and pEW18E, the basic clusters and neutral clusters near the C terminus of CA were oriented in a direction similar to that of the wild type. In contrast, in pEW13 and pEW17, the basic clusters and neutral clusters were oriented drastically differently than that of the wild type. The

alteration of the helical phases probably changed the property of the overall structure and thus interfered with the charged assembly helix function.

Currently, we do not know how this putative α -helical structure affects MLV assembly. It is possible that protein-protein interactions mediated by the charged assembly helix motif facilitates and regulates virus assembly. The charged assembly helix may interact with other domain(s) in Gag, a cellular factor(s), or other charged assembly helix motifs in other Gag molecules. It is also possible that the charged assembly helix serves as a structural element to orient NC during Gag-Gag interactions, although we consider this hypothesis less likely. Analyses of HIV-1 NC structure indicate that the regions flanking the Cys-His boxes, including the region immediately adjacent to the α -helical structure, are very flexible (11). Given this flexibility, it is likely that NC could compensate for some structural variations and would not have such a stringent requirement on the phases of the α -helices in this motif. Of the other three aforementioned possibilities on the protein-protein interactions involving the charged assembly helix motif, we favor the hypothesis that intermolecular interactions among the charged assembly helix motifs facilitate the Gag multimerization process. We find this hypothesis attractive because of the alternating positively and negatively charged bands on the proposed α -helices. In addition, cross-linking experiments indicated that the very C-termini of the MLV CA domains are in close proximity with one another in immature virions (29). This observation is in agreement with the possibility that charged assembly helix motifs from different Gag polyproteins interact with one another. Studies are underway to probe the structure of the proposed α -helices and their relation to the function of the charged assembly helix motif in MLV CA.

ACKNOWLEDGEMENTS

We thank Vinay K. Pathak for many helpful discussions, continuous intellectual input in this project, critical reading of the manuscript, and coining the term charged assembly helix; Alan Rein for communicating unpublished results, encouragement, and critical reading of the manuscript; John Coffin for encouragement, discussions, and critical reading of the manuscript; Anne Arthur for her expert editorial help on this manuscript; Sook-Kyung Lee for generating SR2-293T cells; Delphine Muriaux for ribosomal RNA used as a control in RNA analyses; Jane Mirro for help with western analyses of cell lysates; and time provided by the NCI's Frederick Biomedical Supercomputing Center.

This work is supported by the HIV Drug Resistance Program, NCI, NIH, and in part by NIH under SAIC contract #N01-CO-12400.

Table 1. Virus titers generated by *gag-pol* expression constructs with wild-type or mutant motifs

Construct	Viral titer ($\times 10^3$)*	
pWZH30	6.0 \pm 2.1	7.5 \pm 1.9
pSR33	<0.001	ND
pSR11	2.5 \pm 1.7	ND
pSR15	10.0 \pm 8.0	ND
pFeEW	12.3 \pm 11.4	ND
pEW17	<0.001	ND
pEW19	7.3 \pm 6.4	ND
pEW21	<0.001	ND
pEW13	ND	<0.001
pEW15-res	ND	10.8 \pm 5.1
pEW18E	ND	7.0 \pm 3.9
pEW18R	ND	3.9 \pm 2.8
pEW20	ND	<0.001
pEW22U	ND	4.5 \pm 2.4
pEW22D	ND	6.8 \pm 1.6

*Viral titer is shown as CFU/ml (mean \pm SE). The data from the left- and right-hand columns were generated from four to six and two to five sets of independent experiments, respectively. ND, not determined.

Table 2. Distribution of particle sizes from pWZH30- and pSR33-derived virions

Construct	Particle diameter (nm)										
	<100		100-150					150-200	200-250	250-300	300-350
	<90	90-100	100-110	110-120	120-130	130-140	140-150				
pWZH30	0	10	32	41	15	1	1	0	0	0	
pSR33	5	8	5	10	8	7	6	17	21	6	

FIGURE LEGENDS

Fig. 1. Location and sequence of the motif rich in charged residues. (A) Location and sequence of the motif in MLV. Amino acids between the motif and NC are underlined. MHR: major homology region. (B) Comparison of sequences in the motif among several gammaretroviruses. BaEV, baboon endogenous virus; GaLV, gibbon ape leukemia virus; KoRV, *Phascolarctos cinereus* (koala) retrovirus; MDEV, *Mus dunni* endogenous virus; PERV-A, porcine endogenous virus class A; PERV-B, porcine endogenous virus class B; RaLV, rat leukemia virus; FeLV-A, feline leukemia virus subgroup A; SNV, spleen necrosis virus. Acidic and basic residues are shown in red and blue, respectively.

Fig. 2. Amino acid sequences of the motif rich in charged residues in wild-type and mutant *gag-pol* expression constructs. Δ EW, the numbers of amino acids deleted.

Fig. 3. Characterization of pWZH30- and pSR33-transfected cells and virions derived from these cells. (A) Western analyses of cell-free virions. (B) Western analyses of cell lysates. Lanes labeled SR2, supernatants or cell lysates from cells transfected with pSR2 and pSV-a-MLV-env; other lane labels, supernatants or cell lysates from cells transfected with the indicated *gag-pol* expression construct, plus pSR2 vector and pSV-a-MLV-env. Both anti-MLV CA and anti-MLV MA antibodies were used for the detection of virion lysates, whereas only anti-MLV CA antibody was used to detect cell lysates. The larger molecular weight bands (near the 79 kDa marker) detected in the SR33 virion lysates are probably caused by incomplete processing. A wide smear (between 30 to 40 kDa) is present in the cell lysate analyses, which is presumably background created by

nonspecific antibody binding. The migration of molecular size standards (in kilodaltons) and Gag, CA and MA is shown.

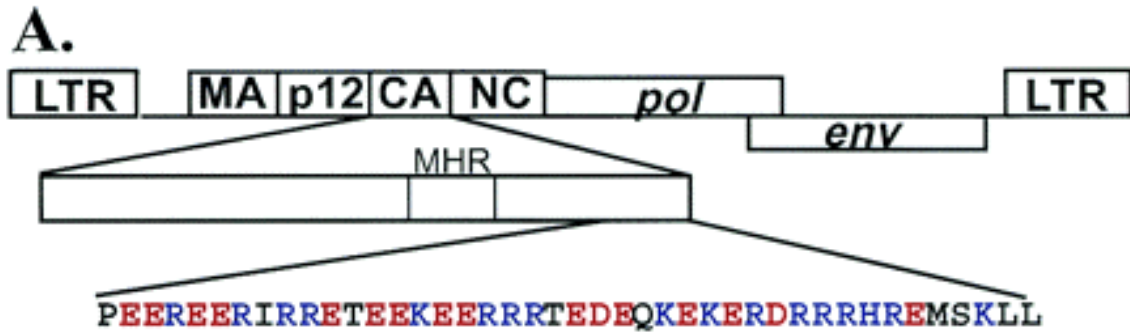
Fig. 4. Electron micrographs of pWZH30-derived virions (A) and pSR33-derived virions (B, C, and D). All four micrographs are shown in the same magnification, with the 100-nm bar indicated at the bottom of the graphs.

Fig. 5. Comparison of the RNA contents in pWZH30- and pSR33-derived virions. (A) Virion associated SR2 RNA. (B) Virion associated 18S ribosomal RNA. Western analyses were performed to estimate the amounts of virions, and RNAs from the same amount of pWZH30- and pSR33-derived virions were compared. RNAs from pWZH30-derived virions were set as 100%. Data from three independent experiments are shown as mean \pm SE.

Fig. 6. Western analyses of the wild type and deletion mutants. (A and B) Western analysis of cell-free virions. (C) Western analysis of cell lysates. The lane labels, antibodies used, and explanations are the same as in Fig. 3. The migration of molecular size standards (in kilodaltons) and Gag, CA, and MA is shown.

Fig. 7. Western analyses of mutants generated to test the proposed α -helical structure. (A) Western analysis of cell-free virions. (B) Western analysis of cell lysates. The lane labels, antibodies used, and explanations are the same as in Fig. 3. The migration of molecular size standards (in kilodaltons) and Gag, CA, and MA is shown.

Fig. 8. Space-filling models of the predicted helical structure of the wild type and mutant MLV charged assembly helix. N termini are located on the left. Amino acids are colored by charge: blue, positive; red, negative; and white, neutral. The wild-type charged assembly helix (labeled MLV) is shown in the center, two functional mutants (labeled pSR11 and pEW18E) are located at the left-hand side, and the two nonfunctional mutants (labeled pEW13 and pEW17) are located at the right-hand side. The basic clusters and neutral clusters near the C terminus of CA were oriented in the same direction in wild type and the two functional mutants. In contrast, the basic clusters and neutral clusters were oriented drastically differently than that of the wild type in the two nonfunctional mutants.



B.

MLV	PEEEREERIRRETEEKEERRRTEDEQKEKERDRRRHREMSKLL
BaEV	PEEERARLIKEQEEREDR-----RDRKRDKHLTKIL
GaLV	EEERQEREKKEAEEKERR-----RDRPKKKNLTKIL
KoRV	EEKQEREKKEETEERERR-----RDRRQEKNLTKIL
MDEV	EEERQEREKKEVEERENR-----RDRRQERNLSKIL
PERV-A	EEEREQRKEREREEREER-----RNKRQEKNLTKIL
PERV-B	EEEKEQRKEKEREEREER-----RDRRQEKNLTKIL
RaLV	QEEKEDRVRREKDE-----RDRKRNRRELSRIL
FeLV-A	PEEEREERL-----WQRQEERDKKRHKEMTKVL
SNV	PA-----KATHELTQKMAKVL

Figure 1.

A.

Construct		Δ EW
pWZH30	PEEREERIRRETEEKEERRRTEDEQKEKERDRRRHREMSKLL	0
pSR33	P-----HREMSKLL	33

B.

pSR11	PEEREERIRRETEEKEER-----RDRRRHREMSKLL	11
pSR15	PEEREERIRRETEE-----RDRRRHREMSKLL	15
pFeEW	PEEREERI-----EQKEKERDRRRHREMSKLL	15

C.

pEW17	PEEREERIRRETEE-----RRRHREMSKLL	17
pEW19	PEEREERIRRETE-----RRHREMSKLL	19
pEW21	PEEREERIRRET-----RHREMSKLL	21

D.

pEW13	PEEREERIRRETE--EER-----RDRRRHREMSKLL	13
pEW15res	PEEREERIRRETEE-----RAARRHREMSKLL	15
pEW18R	PEEREERIRRETEE-----RRHREMSKLL	18
pEW18E	PEEREERIRRETE-----RRRHREMSKLL	18
pEW20	PEEREERIRRETE-----RR-REMSKLL	20
pEW22U	PEEREERIRRET-----HREMSKLL	22
pEW22D	PEEREERIRRETE-----REMSKLL	22

Figure 2.

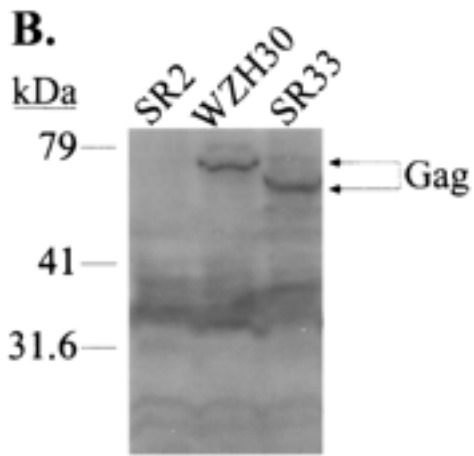
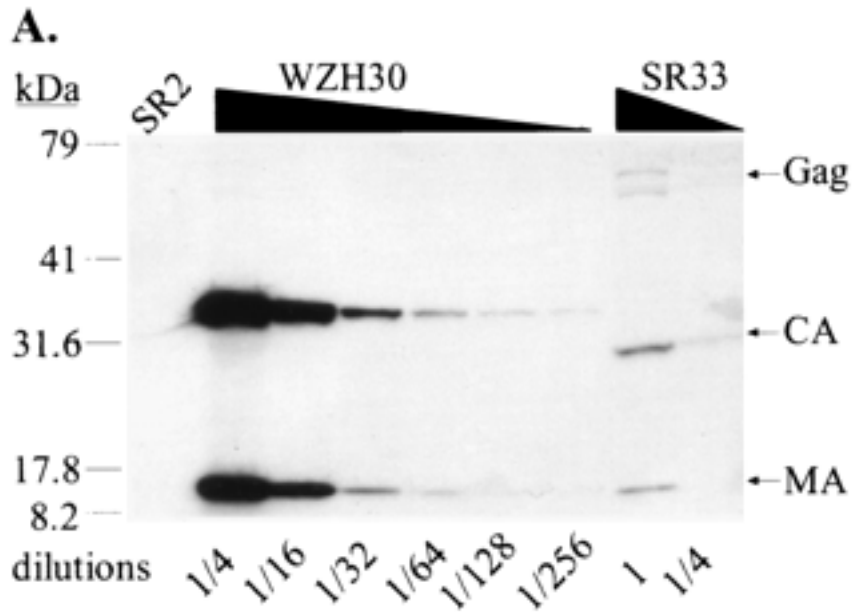
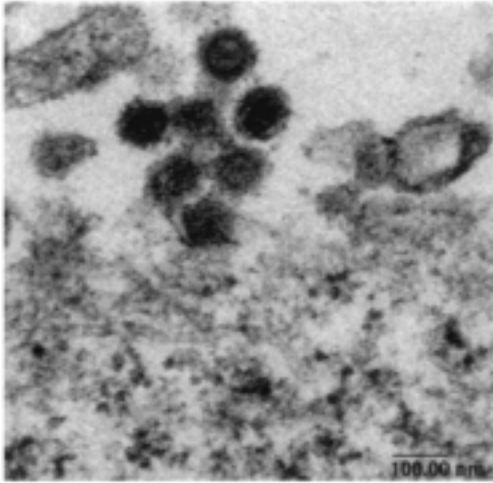
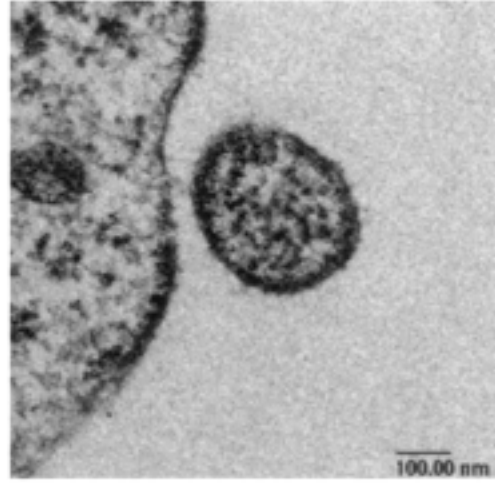


Figure 3.

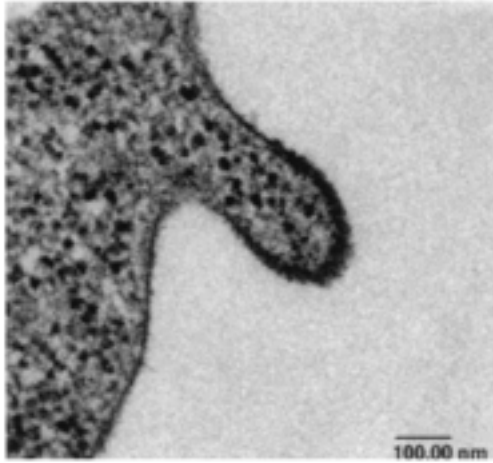
A.



B.



C.



D.

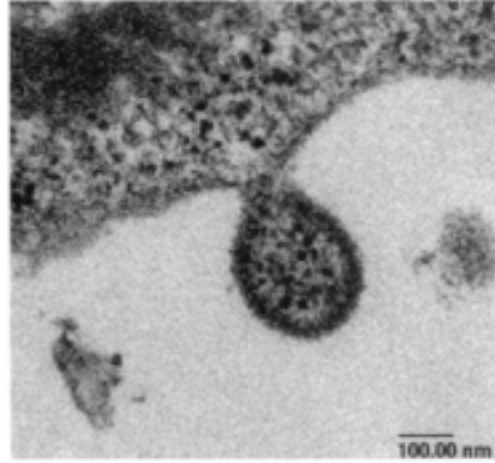


Figure 4.

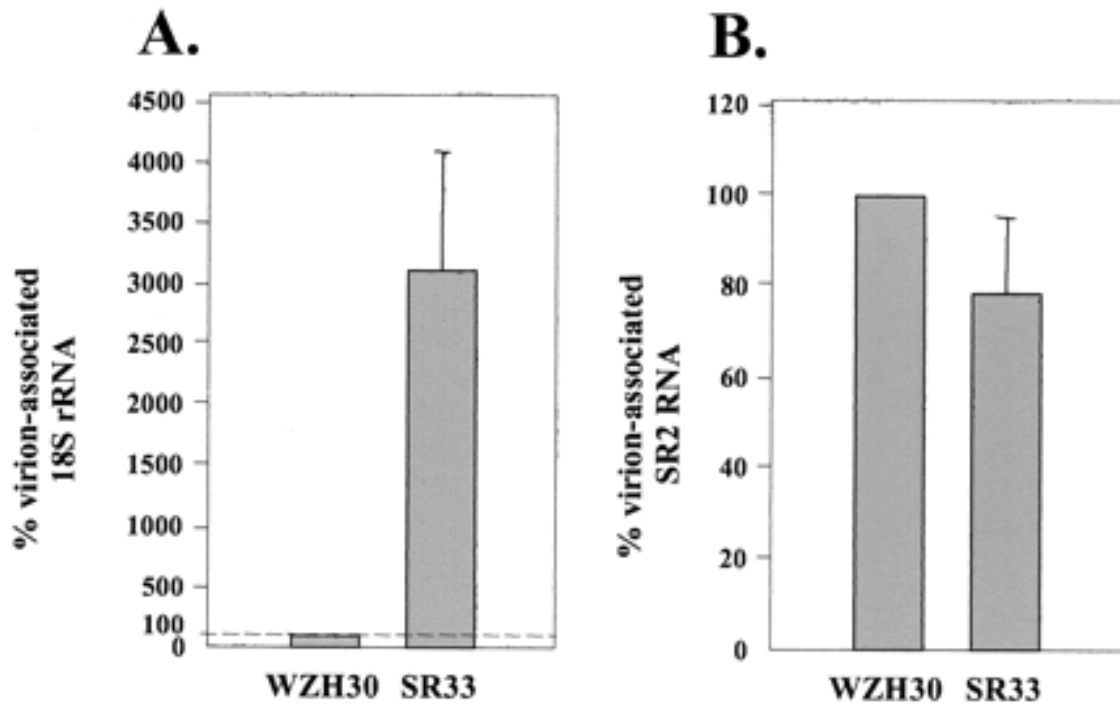


Figure 5.

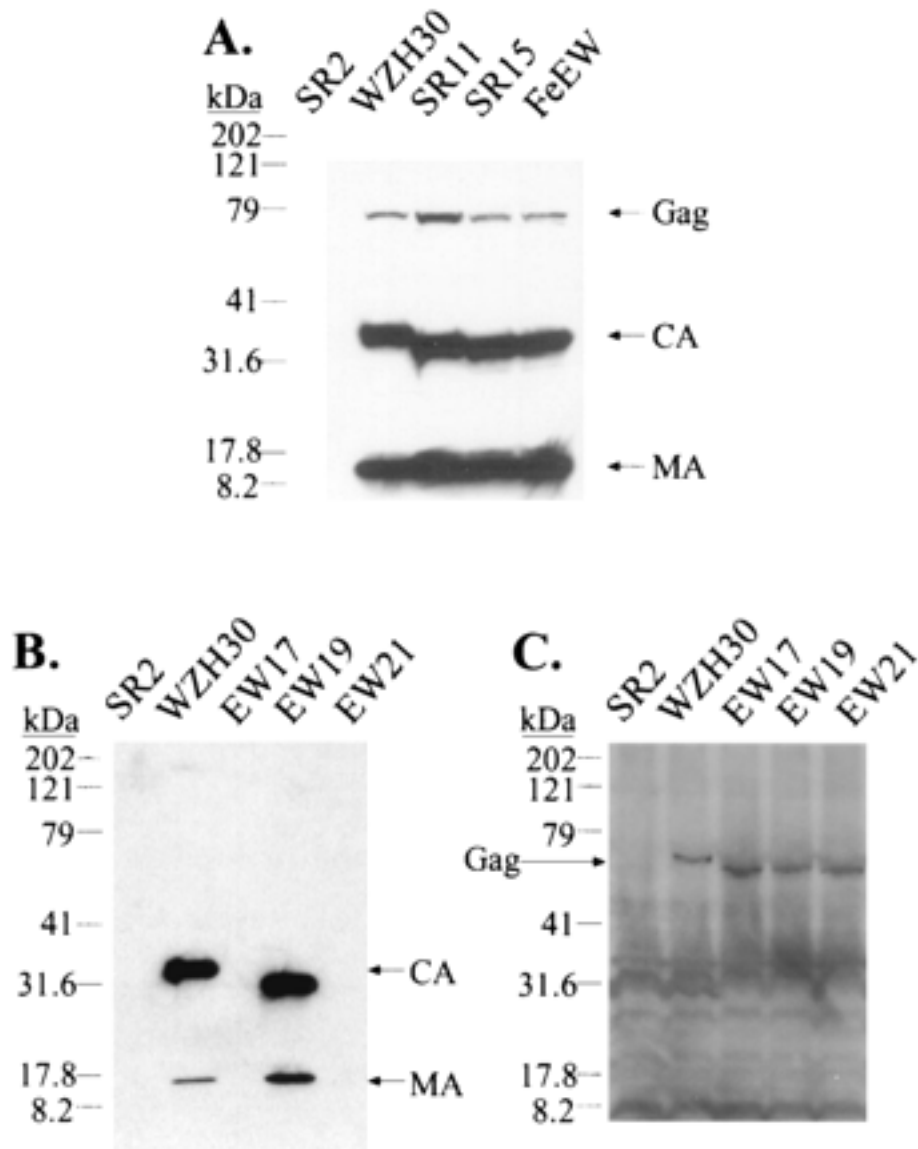
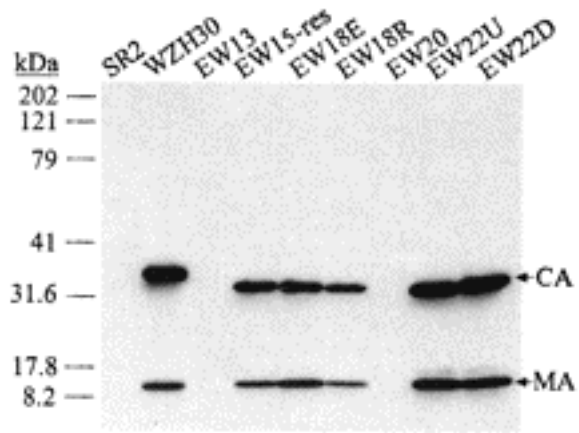


Figure 6.

A.



B.

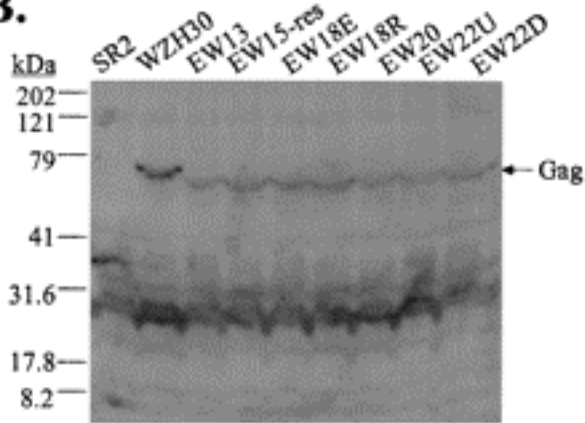


Figure 7.

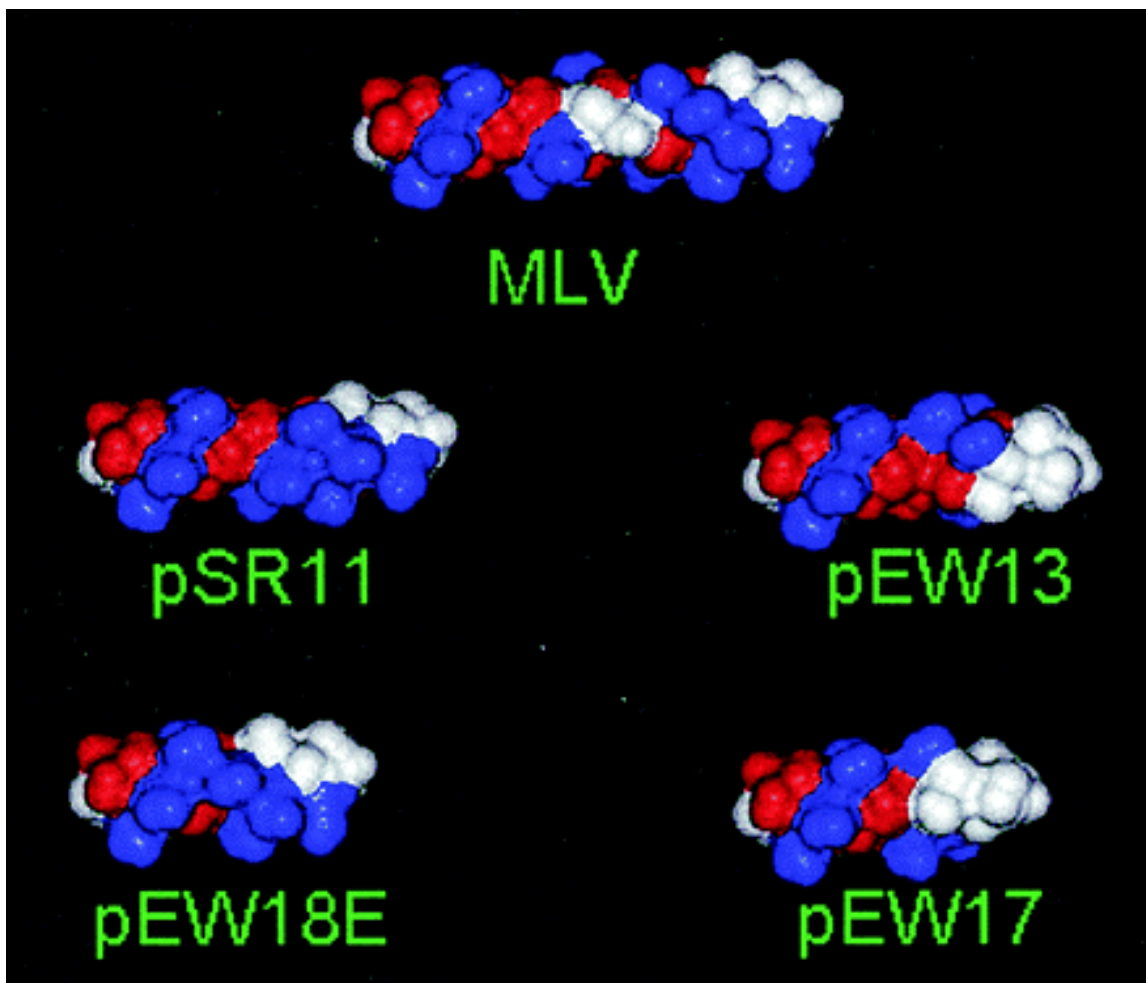


Figure 8.

REFERENCES

1. **Accola, M. A., S. Hoglund, and H. G. Gottlinger.** 1998. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J Virol* **72**:2072-2078.
2. **Alin, K., and S. P. Goff.** 1996. Amino acid substitutions in the CA protein of Moloney murine leukemia virus that block early events in infection. *Virology* **222**:339-51.
3. **Birkett, A. J., B. Yelamos, I. Rodriguez-Crespo, F. Gavilanes, and D. L. Peterson.** 1997. Cloning, expression, purification, and characterization of the major core protein (p26) from equine infectious anemia virus. *Biochim Biophys Acta* **1339**:62-72.
4. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* **3**:469-78.
5. **Cairns, T. M., and R. C. Craven.** 2001. Viral DNA synthesis defects in assembly-competent Rous sarcoma virus CA mutants. *J Virol* **75**:242-50.
6. **Campbell, S., and V. M. Vogt.** 1997. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J Virol* **71**:4425-4435.

7. **Campbell, S., and V. M. Vogt.** 1995. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J Virol* **69**:6487-6497.
8. **Cheslock, S. R., J. A. Anderson, C. K. Hwang, V. K. Pathak, and W. S. Hu.** 2000. Utilization of nonviral sequences for minus-strand DNA transfer and gene reconstitution during retroviral replication. *J Virol* **74**:9571-9579.
9. **Clish, C. B., D. H. Peyton, and E. Barklis.** 1998. Solution structures of human immunodeficiency virus type 1 (HIV-1) and moloney murine leukemia virus (MoMLV) capsid protein major-homology-region peptide analogs by NMR spectroscopy. *Eur J Biochem* **257**:69-77.
10. **Craven, R. C., A. E. Leure-duPree, R. A. Weldon, Jr., and J. W. Wills.** 1995. Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein. *J Virol* **69**:4213-27.
11. **De Guzman, R. N., Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, and M. F. Summers.** 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**:384-388.
12. **DuBridge, R. B., P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller, and M. P. Calos.** 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* **7**:379-387.
13. **Ehrlich, L. S., B. E. Agresta, and C. A. Carter.** 1992. Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *J Virol* **66**:4874-83.
14. **Fassati, A., and S. P. Goff.** 1999. Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. *J Virol* **73**:8919-25.

15. **Freed, E. O.** 1998. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* **251**:1-15.
16. **Freed, E. O.** 2002. Viral late domains. *J Virol* **76**:4679-4687.
17. **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**:5013-5018.
18. **Gamble, T. R., S. Yoo, F. F. Vajdos, U. K. von Schwedler, D. K. Worthylake, H. Wang, J. P. McCutcheon, W. I. Sundquist, and C. P. Hill.** 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **278**:849-853.
19. **Ganser, B. K., S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist.** 1999. Assembly and analysis of conical models for the HIV-1 core. *Science* **283**:80-3.
20. **Gottlinger, H. G.** 2001. The HIV-1 assembly machine. *Aids* **15 Suppl 5**:S13-20.
21. **Gottlinger, H. G., T. Dorfman, J. G. Sodroski, and W. A. Haseltine.** 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* **88**:3195-3199.
22. **Gritz, L., and J. Davies.** 1983. Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**:179-188.
23. **Khorasanizadeh, S., R. Campos-Olivas, and M. F. Summers.** 1999. Solution structure of the capsid protein from the human T-cell leukemia virus type-I. *J Mol Biol* **291**:491-505.

24. **Kingston, R. L., T. Fitzon-Ostendorp, E. Z. Eisenmesser, G. W. Schatz, V. M. Vogt, C. B. Post, and M. G. Rossmann.** 2000. Structure and self-association of the Rous sarcoma virus capsid protein. *Structure Fold Des* **8**:617-28.
25. **Krausslich, H. G., M. Facke, A. M. Heuser, J. Konvalinka, and H. Zentgraf.** 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol* **69**:3407-3419.
26. **Krishna, N. K., S. Campbell, V. M. Vogt, and J. W. Wills.** 1998. Genetic determinants of Rous sarcoma virus particle size. *J Virol* **72**:564-577.
27. **Landau, N. R., K. A. Page, and D. R. Littman.** 1991. Pseudotyping with human T-cell leukemia virus type I broadens the human immunodeficiency virus host range. *J Virol* **65**:162-169.
28. **Mammano, F., A. Ohagen, S. Hoglund, and H. G. Gottlinger.** 1994. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J Virol* **68**:4927-36.
29. **McDermott, J., S. Karanjia, Z. Love, and E. Barklis.** 2000. Crosslink analysis of N-terminal, C-terminal, and N/B determining regions of the Moloney murine leukemia virus capsid protein. *Virology* **269**:190-200.
30. **Miller, A. D., and C. Buttimore.** 1986. Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. *Mol Cell Biol* **6**:2895-2902.

31. **Muriaux, D., J. Mirro, K. Nagashima, D. Harvin, and A. Rein.** 2002. Murine leukemia virus nucleocapsid mutant particles lacking viral RNA encapsidate ribosomes. *J Virol* **76**:11405-11413.
32. **Petropoulos, C.** 1997. *Retroviral Taxonomy, Protein Structures, Sequences, and Genetic Maps.* Cold Spring Harbor Press, Cold Spring Harbor, NY.
33. **Riggs, J. L., R. M. McAllister, and E. H. Lennette.** 1974. Immunofluorescent studies of RD-114 virus replication in cell culture. *J Gen Virol* **25**:21-29.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular Cloning: A Laboratory Manual*, 2 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35. **Swanstrom, R., and J. W. Wills.** 1997. *Synthesis, Assembly, and Processing of Viral Proteins.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
36. **Temin, H. M., and S. Mizutani.** 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**:1211-1213.
37. **Tobin, G. J., K. Nagashima, and M. A. Gonda.** 1996. Immunologic and Ultrastructural Characterization of HIV Pseudovirions Containing Gag and Env Precursor Proteins Engineered in Insect Cells. *Methods* **10**:208-218.
38. **Vogt, V. M.** 1997. *Retroviral Virions and Genomes.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
39. **von Schwedler, U. K., T. L. Stemmler, V. Y. Klishko, S. Li, K. H. Albertine, D. R. Davis, and W. I. Sundquist.** 1998. Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *Embo J* **17**:1555-68.

40. **Wills, J. W., C. E. Cameron, C. B. Wilson, Y. Xiang, R. P. Bennett, and J. Leis.** 1994. An assembly domain of the Rous sarcoma virus Gag protein required late in budding. *J Virol* **68**:6605-6618.
41. **Wills, J. W., and R. C. Craven.** 1991. Form, function, and use of retroviral gag proteins [editorial]. *AIDS* **5**:639-54.
42. **Zhang, W. H., C. K. Hwang, W. S. Hu, R. J. Gorelick, and V. K. Pathak.** 2002. Zinc finger domain of murine leukemia virus nucleocapsid protein enhances the rate of viral DNA synthesis in vivo. *J Virol* **76**:7473-7484.

THIS PAGE INTENDED TO BE LEFT BLANK.

CHAPTER FIVE

DISCUSSION

The retroviridae are a complex family of viruses that exert a significant impact on global ecology and human health. Through the process of reverse transcription, retroviruses gain permanent residence in the genome of any host they infect (28). Once established in the infected cell, they can produce progeny virions that can infect other cells. During this process, retroviruses capture and activate cellular oncogenes, which can cause cancer and leukemia in the host, or they can themselves cause outright disease. The retrovirus with the biggest impact on human health and biology is HIV-1. HIV-1 infections started emerging in Africa and Haiti in the 1970's, and in the United States in the early 1980's; in less than half a century, HIV-1 has swept the world in all continents, causing a global health crisis (12). Africa's public health infrastructure is poorly equipped to contain its epidemic (9) and other nations are not faring much better. Even in the United States and Europe where the infection prevalence is the lowest in the world, infection rates are rising (30), and current therapies are not able to cure the disease. There is increasing viral resistance to current HAART therapy (22, 26) that will at some point render the drugs currently available obsolete.

Several advances in our understanding of retroviruses have led to therapies that preserve a high quality of life for people afflicted with AIDS. Elucidation of the mechanisms of reverse transcription allowed development of nucleoside analogs to effectively halt HIV-1 replication (24). Our understanding of retroviruses' high mutation rate (27) allowed us to understand why resistance to the NRTI's arose so quickly, and further fueled efforts to uncover other vulnerabilities in the HIV-1 replication cycle that could be exploited for new treatments. Inhibitors of the HIV-1 protease, reverse

transcriptase, envelope fusion inhibitors, and integrase inhibitors have developed from our increased understanding of the mechanisms of the retrovirus life cycle (10, 13, 14).

However, ultimately a vaccine will be needed to eradicate AIDS. Currently, there are theories that potent enough antiretroviral therapy could eradicate HIV-1 infection from a single human host (20); however the cost of curing even one infection is likely too expensive to be a viable option where cures are needed the most, such as Africa. Vaccine research is fraught with scientific and ethical difficulty as we lack a reliable animal model for testing vaccines.

In addition to HIV-1, which has had a huge impact on human ecology, other retroviruses are also important etiologic agents for cancers. MLV, Jaagsietke Sheep Retrovirus (JSRV), HTLV-1, and many others are known to cause cancer in human and animal hosts (7, 11, 17, 25). Understanding of how these viruses interact with and exploit weaknesses in hosts is crucial to developing strategies for preventing disease. One promising potential might be to develop vaccines against the viruses that cause cancers. This idea has been proposed before, however no effective vaccine against a cancer caused by a retrovirus has been developed.

The ability of simple retroviruses to rapidly infect mitotically active cells can also be exploited to treat disease (6). Use of retroviruses as gene therapy vectors has not been without controversy, however, and further research is needed to develop retroviruses into an effective and widely affordable gene therapy modality.

Clearly, the importance of retrovirus research cannot be understated. Given the huge impact retroviruses have on biology from the microcosm of the cell to human and animal ecology, the scientific community has a huge task set before it to understand

retroviruses fully enough to counteract diseases caused by infection, and hopefully, one day harness their potential as curative agents.

The work presented in these three chapters has investigated three aspects of retrovirus biology. All studies have utilized a helper cell line/retrovirus vector system to study mechanisms of replication in one complete life cycle of either MLV or SNV (23). All studies looked at the effects of modifying certain properties of either viral proteins or viral vectors to examine their effects on virus biology. Therefore, even though three divergent aspects of viral biology were probed in this work, strategies employed for these investigations were quite similar.

In the first aim, the specific sequence requirements for the R region in retrovirus minus strand DNA transfer were examined. It had been debated whether or not the specific sequence content of the 'R' region was necessary for successful minus strand DNA transfer (2, 3, 33). By substituting the native of viral 'R' sequence with gene fragments, we proved that not only was homology a sufficient factor for minus strand DNA transfer in MLV-based retrovirus vectors, but that minus strand DNA transfer could be used to reconstitute functional genes in the DNA proviruses (5).

In addition to establishing the ability of non-viral sequences to mediate minus-strand DNA transfer, the experimental system described in the first chapter supported mechanistic findings of others in reverse transcription. The first mechanism is that of read-through transcription. This phenomenon, first described in 1987, is the occasional suppression of the termination signal in the LTR of retroviruses, which can lead to incorporation of host sequences into the retrovirus genome (15). In previous studies, this was discovered to be a mechanism for oncogene transduction and activation. In our

system, read-through transcription lead to incorporation of an additional downstream PBS sequence that was contained in the original vector used to transfect virus-producing cells (29). This additional PBS, located downstream of the 3' U5 of the retroviral vector, allowed reverse transcription to initiate on the 3' end of the vector instead of the 5' LTR, bypassing minus-strand DNA transfer altogether, leading to a specific sequence finding on analysis of the provirus' 5' LTRs. LTRs from proviruses that had resulted from reverse transcription initiating from a 3' PBS contained 100% homology with the 3' LTR in the viral vector, which could only occur if minus-strand DNA transfer was bypassed in this manner. The frequency with which this phenomenon occurred (18%) agreed with the frequency found by others.

The second mechanism of retrovirus reverse transcription found in our system was that of read-in transcription. First described by Dang and Hu (8), this phenomenon arises when the transfected retroviral vector is integrated into the producer cells downstream of a cellular promoter that is stronger than the retroviral U3 promoter. This leads to incorporation of extra sequences upstream of the retroviral R region that begins the viral RNA transcripts. In the minus-strand gene reconstitution vectors, this generated viral transcripts that carried an extra region of homology (a 5' U3), which allowed the virus to use the U3 region to mediate minus-strand DNA transfer. Sequence analysis of proviruses that resulted from read-in transcription revealed LTRs that contained 100% homology with the 5' LTR of the viral vector.

The minus-strand transfer gene reconstitution vectors derived to establish the role of non-viral sequences in minus strand DNA transfer could be adapted as a new self-activating gene therapy vector. Ultimately, a suicide gene (18) could be divided into two

fragments that contain overlapping homology, and inserted into the MLV vector. A virus carrying the vector with the suicide gene could be used to infect cells in a mitotically active tumor, and once infected, reverse transcription would reconstitute the lethal gene in the infected tumor cells and cause cell death, hopefully leading to eradication of the cancer. A significant number of malignant tumors contain highly mitotically active cells; presumably an MLV vector would be ideal as a cancer therapy vector because MLV vectors can only infect actively dividing cells.

In addition to applications in gene therapy, this system could be used to test for mobilization of retroviruses in cell culture systems. Retroviral enzymes in the cell population would mobilize vectors carrying gene fragments and effective activity of these retroviruses could be monitored by expression of the reconstituted gene in a one-round replication assay, rather than traditional use of lengthy co-culture assays.

The significance of these results is threefold. First a new gene therapy principle was proven. Secondly, the nature of the mechanism of minus strand DNA transfer was more fully elucidated, and finally, a vector to test for the mobilization of retroviruses in cells was established. In addition to these findings, variations in viral RNA transcription were observed that accounted for some of the provirus forms observed in these experiments.

In the second aim, mechanisms of reverse transcription were investigated. Specifically, the genetic relationships between two distantly related retroviruses- MLV and SNV, were investigated by determining complementation relationships between various domains of the *Pol* gene. Multiple roles of IN have been elucidated for MLV and HIV replication; IN is vital for particle maturation, reverse transcription, and integration

of the viral DNA into the host chromosome. It has been proven that the HHCC domain of IN specifically for HIV-1 is necessary for HIV-1 reverse transcription (31), and it has been proposed that it is the C-terminal end of MLV IN is required for reverse transcription in MLV (21). A functional role of SNV IN during reverse transcription has not been described. The experimental system in the second aim was designed to determine if MLV IN proteins could complement either SNV or MLV RT to complete replication. Chimera SNVgagMLVpol vectors were expressed and the ability of these chimera gag-pol molecules to assemble into virions with SNVgag-pol molecules was proven. The system was based on vectors designed to examine co-packaging of SNV and MLV RNA (4).

Specific mutations were introduced into the *Pol* gene to make SNVRT⁺IN⁻ and SNV RT⁻IN⁺ vectors. MLV RT⁺IN⁻ and MLV RT⁻IN⁺ vectors were synthesized by introducing specific mutations into the *pol* of a chimeric SNVgagMLVpol expression construct. To make the RT⁻IN⁺ constructs, the residues in the catalytic core were substituted to inactivate the RT polymerase (SR7-D150E for MLV and YVNN for SNV). Two RT⁺IN⁻ vectors were constructed for SNV and MLV. A four-amino acid insertion in the HHCC motif was made to disrupt the zinc finger of the SNV and MLV IN for two of the vectors (SR8 for SNV and LC2 for MLV), and a catalytic core substitution was made to disrupt the enzymatic activity of IN for the other two vectors (SR12 and SR13 for SNV and MLV, respectively). Therefore, the vectors with inactivated IN should contain a biologically active RT, and the vectors with inactivated RT should contain biologically active IN.

The ability of SNV Pol proteins to interact and co-assemble with MLV Pol molecules was established. These results showed that level of complementation of the MLV RT⁻IN⁺ vector were the same for SNV RT⁺IN⁻ and MLV RT⁺IN⁻, indicating that SNV and MLV RT proteins can interact equally well with MLV IN.

Furthermore, two different functions of IN have recently been elucidated; these functions impacted the ability of SNV and MLV IN to interact with MLV RT in this system. The level of complementation was greater for the RT⁺IN⁻ HHCC mutation than for the RT⁺IN⁻ catalytic core substitution mutant following normalization for RT activity. Because the RT⁺IN⁻ HHCC IN insertion mutant had a processing defect, it produced less RT, and therefore its level of complementation was higher when corrected for RT activity than the RT⁺IN⁻ catalytic core mutants, which had no processing defect.

The SNV and MLV IN HHCC insertion mutants were deficient in RT activity, which correlated with a processing defect seen in these mutants. Therefore, our results do not definitively demonstrate a role for SNV or MLV IN during reverse transcription of MLV RNA; however, participation of these proteins in reverse transcription cannot be ruled out on the basis of these results. Our results showed no difference in the ability of MLV IN to rescue reverse transcription or integration in an SNV or MLV RT⁺IN⁻ construct. This strongly suggests that several crucial RT-IN interactions are conserved between SNV and MLV. These results indicate that protein interactions in the reverse transcription and pre-integration complexes for MLV and SNV are conserved, and that the individual proteins can interact with other distantly related retroviral proteins.

The C-terminal region of IN encompassing residues 345-365 of MLV has nearly 100% sequence identity with residues 329-349 of SNV IN (Figure 1); this could comprise

a region of IN that contains the crucial contacts that allow for the interchangeability of SNV and MLV IN. This is of significance because it highlights a possible mechanism for evolution of new species of retroviruses, and perhaps viruses with increased pathogenicity or immune evasion for the host. Future studies to explore IN and RT complementation between SNV and MLV could investigate the role of these conserved amino acids in reverse transcription and replication. Establishing cell lines from single cell clones expressing the MLV RT mutant may improve complementation; this was necessary for other studies investigating MLV RH and RT complementation (16). Complementation between MLV and SNV Pol proteins cannot be further evaluated without improving the sensitivity of the assay. Various strategies were employed to attempt improvement of this system, including transient transfection assays and concentrating virus pseudotyped with vesicular stomatitis virus G (VSV-g) envelope prior to complementation titer measurement (32).

Specifically, the work presented in AIM 2 has furthered understanding of interactions between Pol molecules occurs, which may provide insight into mechanisms of retrovirus evolution. It is thought that SNV may have arisen from a recombination event between MPMV and GaLV . This recombination could not have taken place if a cell infected with both viruses had not expressed viral proteins that were able to interact with non-homologous viral RNA. In essence, the system in the second aim required the interaction of SNV and MLV viral proteins for effective replication. The fact that replication did occur with these two species forced cooperation underscores the genetic flexibility of these two simple retroviruses.

In the last aim, mechanisms of MLV assembly were investigated, specifically the contribution of a newly described region at the 3' end of CA. This region, termed the 'Charged assembly helix', was necessary for assembly of mature MLV viral particles. Furthermore, this region, consisting of highly charged acidic and basic residues, was widely conserved among gammaretroviruses, except for SNV. Interestingly, similar motifs are not easily observed in other retrovirus subgroups. The minimum length of CAH sequence necessary for efficient assembly was not determined, as the shortest deletion that did not work was a 13-amino acid deletion, while the longest deletion with wild-type phenotype was a 22 amino acid deletion. These results implied a highly ordered structural component to the charged assembly helix, and modeling studies were undertaken to elucidate its properties. A helical structure was postulated, and further experimental testing supported this proposal. Based on this hypothesis, structural modeling revealed a conserved cluster of neutral amino acids oriented in a similar way on all working charged assembly helix mutants. Follow-up studies of charged assembly helix function will further test the hypothetical alpha helical structure by analyzing the CD spectra of the charged assembly helix to determine the relative amount of alpha-helical structure contributing to the charged assembly helix.

Another possible function of the charged assembly helix that needs to be evaluated is its involvement in the recruitment of envelope proteins.

Such disparity exists between the defective charged assembly helix mutants and the functional charged assembly helix mutants that future studies should attempt to more finely define the mechanisms and role(s) of the charged assembly helix in MLV maturation. Specifically, the replication incompetent charged assembly helix mutants

were severely deficient in budding, virus production, and capsid maturation, as well as particle size regulation. This finding suggests an important function of the charged assembly helix early in the assembly process that all subsequent assembly steps depend upon. One possibility is that the highly ordered alpha helix formed by the charged assembly helix is essential for early Gag-gag interactions. It is tempting to speculate that the arrangement of the charged residues in the charged assembly helix is necessary for tight packing of Gag molecules together during assembly. The modeled charged assembly helix alpha helix contained alternating bands of positive and negative amino acids that could interact intermolecularly with charged assembly helix regions on other Gag proteins, pulling the Gag particles closely together. This theory is supported by the following findings. The replication-defective charged assembly helix mutants form virus particles that are, on average, much larger than wild type and lose their ability to tightly regulate virus particle size. This potentially implies a problem with condensation of the Gag proteins, which could easily be caused by charge-charge repulsions and steric inhibitions. Secondly, the charged assembly helix viruses that do get released are efficiently processed to their appropriate Gag protein components, which suggested that the charged assembly helix was also important in condensation of the virus core and not proteolytic processing, since even though processing occurs no mature particles were observed on EM. Thirdly, deletion of the 33 amino acids in the charged assembly helix seriously impaired virus production. This finding mapped a crucial assembly determinant to the charged assembly helix.

One weakness of the proposal that the charged assembly helix is involved in intermolecular Gag-Gag interactions is the fact that SNV, a related gammaretrovirus,

lacks a highly charged domain. Structural predictions *do* suggest a highly alpha-helical structure to the C-terminus of SNV CA (Figure 2); however, it is not comprised of highly charged residues and it is therefore unlikely to utilize charge-charge attractions to bring Gag molecules together. This suggests that either SNV has an entirely different mechanism for Gag:Gag interactions during assembly, or that it has the same mechanisms of Gag-gag interactions, but accomplishes assembly via mechanisms that do not require charged amino acids.

Structural studies are done with great difficulty, however, and care must be taken when proposing structural models in the absence of objective data. Algorithms for structure prediction are simply based on previously proven structures; these are not always predictive of actual new structural findings. In addition, although this sequence is highly conserved in gammaretroviruses, SNV lacks an charged assembly helix, which indicates that there must be ways at least one gammaretrovirus can fulfill the role of the charged assembly helix without an charged assembly helix-like sequence.

Perhaps the most interesting finding of these studies is the apparent correlation with the charged assembly helix domain and the roles of the p2 in HIV-1 and the spacer peptide (SP) in RSV. An alpha-helical structure has been proposed for the p2 of HIV-1, and deletion of this region results in mutants with structural similarities to our charged assembly helix mutants (1). In addition, deletion of the SP in RSV results in a similar phenotype for RSV; however, an alpha-helix structure for this peptide has not been established (19).

Regardless of what the structure of the charged assembly helix is eventually determined to be, this region has identified a new crucial determinant in MLV assembly,

and future studies of this region will likely provide great advances in our understanding of retrovirus assembly. Ultimately, it will be very important to elucidate the differences between SNV and MLV assembly and to determine what sequence(s) in SNV Gag polyproteins carry out the analogous function of the MLV charged assembly helix.

In summary, the work presented in this dissertation provided major insights into mechanisms of retrovirus reverse transcription and assembly. The first work established a proof-of-principle for a novel retrovirus gene therapy vector and proved that non-viral sequences can mediate minus-strand DNA transfer. The second aim established for the first time that MLV and SNV Pol proteins could interact and replicate as long as the Gag proteins are derived from SNV. Finally, the third aim identified a region in CA crucial for MLV assembly and proposed a working model for its function, which provides a framework upon which future studies can further dissect the role of the charged assembly helix in MLV biology.

In closing, the three works presented here have made important contributions to the current base of knowledge of retrovirus biology and have opened up new and interesting questions for further study.

FIGURE LEGENDS

Figure 1: Multiple sequence alignment of SNV and MLV Integrase. SNV Integrase is shown on the top line and MLV Integrase is on the bottom line. Asterisks denote areas of identity. The red letters denote an area of high homology between SNV and MLV IN near the C-terminal region of integrase.

Figure 2:Chou-Fasman structure prediction values for SNV CA. The sequence is given in the top line. The actual predictor number is given in the second line; any value >1.0 is strongly predictive of an alpha helix. The third line shows the predicted structure for each amino acid. Capital 'B' denotes strong beta sheet-forming, Capital 'H' denotes strong alpha-helix forming. Lower-case 'i' denotes an indifferent predictor value, and capital 'I' denotes slightly in favor of forming a helix. Lower-case 'h' predicts helix formation, lower-case 'b' denotes slight beta-sheet formation.

SNV Capsid Chou-Fasman Structure prediction

p l r e t g e r d m t g r p m r t y v p f t t s d l y n w k n q n p s
0.57 1.21 0.98 1.51 0.83 0.57 1.51 0.98 1.01 1.45 0.83 0.57 0.98 0.57 1.45 0.98 0.83 0.69 1.06 0.57 1.13 0.83 0.83 0.77 1.01 1.21 0.69 0.67 1.08 1.16 0.67 1.11 0.67 0.57 0.77
B H i H i B H i I H i B i B H i i b h B h i i i I H b b h h b h b B i

f s q a p d e v i s l l e s v f y t h q p t w d d c q q l l r t l f t
1.13 0.77 1.11 1.42 0.57 1.01 1.51 1.06 1.08 0.77 1.21 1.21 1.51 0.77 1.06 1.13 0.69 0.83 1.00 1.11 0.57 0.83 1.08 1.01 1.01 0.70 1.11 1.11 1.21 1.21 0.98 0.83 1.21 1.13 0.83
h i h H B I H h h i H H H i h h b i I h B i h I I i h h H H i i H h i

t e e r e r v r t e s r r e v r n d q g v q v t d e r e i e a q f p a
0.83 1.51 1.16 0.98 1.51 0.98 1.06 0.98 0.83 1.51 0.77 0.98 0.98 1.51 1.06 0.98 0.67 1.01 1.11 0.57 1.06 1.11 1.06 0.83 1.01 1.51 0.98 1.51 1.08 1.51 1.41 1.11 1.13 0.57 1.42
i H H i H i h i i H i i i H h i b I H B h H h i I H i H h H H H h B H

t r p d w d p n t g n d n l e r y r q i l l r g l r a a a r k p t n l s
0.83 0.98 0.57 1.01 1.08 0.77 0.57 1.08 1.01 0.57 0.98 1.42 0.57 1.51 0.67 1.11 1.42 1.42 1.08 1.06 1.08 1.11 1.13 1.06 0.77 1.11 0.77 1.42 0.57 1.01 1.08 0.98 1.16 1.16 1.08 1.11
i i B I h i B h I B i H B H b H H H h h h H h h i H i H B I h i h h h H

k i t e v r q g a d e s p t a y l e r l y q a y r t w s p i d p r a p
1.16 1.08 0.83 1.51 1.06 0.98 1.11 0.57 1.42 1.01 1.51 0.77 0.57 0.83 1.42 0.69 1.21 1.51 0.98 1.21 0.69 1.11 1.42 0.69 0.98 0.83 1.08 0.77 0.57 1.08 1.01 0.57 0.98 1.42 0.57
h h i H h i h B H I H i B i H b H H i H b h H b i i h i B h I B i H B

e n q a a i v i q f v s a p d i r k k i q k i d g f q g k s l s e l
1.51 0.67 1.11 1.42 1.42 1.08 1.06 1.08 1.11 1.13 1.06 0.77 1.42 0.57 1.01 1.08 0.98 1.16 1.16 1.08 1.11 1.16 1.08 1.01 0.57 1.13 1.11 0.57 1.16 0.77 1.21 0.77 1.51 1.21
H b h H H h h h h h h i H B I h i h h h h h h I B h h B h i H i H H

v a i a q k v f d q r e d p a k a t h e l t q k m a k v l
1.06 1.42 1.08 1.42 1.11 1.16 1.06 1.13 1.01 1.11 0.98 1.51 1.01 0.57 1.42 1.16 1.42 0.83 1.00 1.51 1.21 0.83 1.11 1.16 1.45 1.42 1.16 1.06 1.21
h H h H h h h h I h i H I B H h H i I H H i h h H H h h H

REFERENCES

1. **Accola, M. A., S. Hoglund, and H. G. Gottlinger.** 1998. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J Virol* **72**:2072-8.
2. **Adam, M. A., W. R. Osborne, and A. D. Miller.** 1995. R-region cDNA inserts in retroviral vectors are compatible with virus replication and high-level protein synthesis from the insert. *Hum Gene Ther* **6**:1169-76.
3. **Allain, B., J. B. Rasclé, H. de Rocquigny, B. Roques, and J. L. Darlix.** 1998. CIS elements and trans-acting factors required for minus strand DNA transfer during reverse transcription of the genomic RNA of murine leukemia virus. *J Mol Biol* **277**:225-35.
4. **Certo, J. L., B. F. Shook, P. D. Yin, J. T. Snider, and W. S. Hu.** 1998. Nonreciprocal pseudotyping: murine leukemia virus proteins cannot efficiently package spleen necrosis virus-based vector RNA. *J Virol* **72**:5408-13.

5. **Cheslock, S. R., J. A. Anderson, C. K. Hwang, V. K. Pathak, and W. S. Hu.** 2000. Utilization of nonviral sequences for minus-strand DNA transfer and gene reconstitution during retroviral replication. *J Virol* **74**:9571-9.
6. **Culver, K. W., W. R. Osborne, A. D. Miller, T. A. Fleisher, M. Berger, W. F. Anderson, and R. M. Blaese.** 1991. Correction of ADA deficiency in human T lymphocytes using retroviral-mediated gene transfer. *Transplant Proc* **23**:170-1.
7. **D'Andrea, A. D., J. F. Moreau, and M. O. Showers.** 1992. Molecular mimicry of erythropoietin by the spleen focus-forming virus gp55 glycoprotein: the first stage of Friend virus-induced erythroleukemia. *Biochim Biophys Acta* **1114**:31-41.
8. **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.
9. **De Cock, K. M., D. Mbori-Ngacha, and E. Marum.** 2002. Shadow on the continent: public health and HIV/AIDS in Africa in the 21st century. *Lancet* **360**:67-72.

10. **Derdeyn, C. A., J. M. Decker, J. N. Sfakianos, X. Wu, W. A. O'Brien, L. Ratner, J. C. Kappes, G. M. Shaw, and E. Hunter.** 2000. Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J Virol* **74**:8358-67.
11. **Feuer, G., and I. S. Chen.** 1992. Mechanisms of human T-cell leukemia virus-induced leukemogenesis. *Biochim Biophys Acta* **1114**:223-33.
12. **Garrett, L.** 1994. *The Coming Pague: Newly Emerging Diseases in a World Out of Balance.* Penguin Books, New York.
13. **Hazuda, D., P. Felock, J. Hastings, B. Pramanik, A. Wolfe, G. Goodarzi, A. Vora, K. Brackmann, and D. Grandgenett.** 1997. Equivalent inhibition of half-site and full-site retroviral strand transfer reactions by structurally diverse compounds. *J Virol* **71**:807-11.
14. **Hazuda, D. J., P. Felock, M. Witmer, A. Wolfe, K. Stillmock, J. A. Grobler, A. Espeseth, L. Gabryelski, W. Schleif, C. Blau, and M. D. Miller.** 2000. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**:646-50.

15. **Herman, S. A., and J. M. Coffin.** 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science* **236**:845-8.
16. **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.
17. **Johnson, P., and S. Benchimol.** 1992. Friend virus induced murine erythroleukaemia: the p53 locus. *Cancer Surv* **12**:137-51.
18. **Kirn, D., I. Niculescu-Duvaz, G. Hallden, and C. J. Springer.** 2002. The emerging fields of suicide gene therapy and virotherapy. *Trends Mol Med* **8**:S68-73.
19. **Krishna, N. K., S. Campbell, V. M. Vogt, and J. W. Wills.** 1998. Genetic determinants of Rous sarcoma virus particle size. *J Virol* **72**:564-77.
20. **Kulkosky, J., and R. J. Pomerantz.** 2002. Approaching eradication of highly active antiretroviral therapy-persistent human immunodeficiency virus type 1 reservoirs with immune activation therapy. *Clin Infect Dis* **35**:1520-6.

21. **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.
22. **Larder, B. A., and S. D. Kemp.** 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* **246**:1155-8.
23. **Miller, A. D., J. V. Garcia, N. von Suhr, C. M. Lynch, C. Wilson, and M. V. Eiden.** 1991. Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J Virol* **65**:2220-4.
24. **Mitsuya, H., R. Yarchoan, and S. Broder.** 1990. Molecular targets for AIDS therapy. *Science* **249**:1533-44.
25. **Palmarini, M., J. M. Sharp, M. de las Heras, and H. Fan.** 1999. Jaagsiekte sheep retrovirus is necessary and sufficient to induce a contagious lung cancer in sheep. *J Virol* **73**:6964-72.
26. **Richman, D. D.** 1992. HIV drug resistance. *AIDS Res Hum Retroviruses* **8**:1065-71.

27. **Svarovskaia, E. S., S. R. Cheslock, W. H. Zhang, W. S. Hu, and V. K. Pathak.** 2003. Retroviral mutation rates and reverse transcriptase fidelity. *Front Biosci* **8**:D117-34.
28. **Temin, H. M.** 1976. The DNA provirus hypothesis. *Science* **192**:1075-80.
29. **Whitcomb, J. M., B. A. Ortiz-Conde, and S. H. Hughes.** 1995. Replication of avian leukosis viruses with mutations at the primer binding site: use of alternative tRNAs as primers. *J Virol* **69**:6228-38.
30. **(WHO), J. U. N. P. o. H. A. U. a. W. H. O.** 2002, posting date. AIDS epidemic update: December 2002. [Online.]
31. **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.
32. **Yang, Y., E. F. Vanin, M. A. Whitt, M. Fornerod, R. Zwart, R. D. Schneiderman, G. Grosveld, and A. W. Nienhuis.** 1995. Inducible, high-level

production of infectious murine leukemia retroviral vector particles pseudotyped with vesicular stomatitis virus G envelope protein. *Hum Gene Ther* **6**:1203-13.

33. **Yin, P. D., V. K. Pathak, A. E. Rowan, R. J. Teufel, 2nd, and W. S. Hu.** 1997.

Utilization of nonhomologous minus-strand DNA transfer to generate recombinant retroviruses. *J Virol* **71**:2487-94.

APPENDIX

Cloning Strategies for Charged Assembly Helix Mutants. Construction of the CAH mutants was done with PCR mutagenesis of vector pWZH30, an MLV *gag-pol* expression vector. Two primers were designed to amplify the region for mutagenesis in each new mutant, and then two primers containing the mutated sequence were designed for use with these primers. The cloning strategy will be explained in detail for the cloning of pSR33; the table below includes the primer sequences for pSR33 and the other mutants constructed in this project.

To make a 33-amino acid deletion in pWZH30 to make pSR33, a two-step PCR reaction was needed. Using pWZH30 as a template, the first PCR reactions amplified bp 1885-2440 in pWZH30 and bp 2440-2951. The primers used for the former reaction were P7MCA (5' CGGTGCGGGGCGATGATGGGC 3') and delEWrev (5' CCAAAAGCTTGCTCATCTCTCTATGCGGGGTTTCTCGTTTATTAAAG 3') . These primers amplified the region 1885-2440. The latter reaction was done with primers RPM2951 (5' AACCTGAGCTCCTGATCCCTC 3') and delEWfor (5'GCATAGAGAGATGAGCAAGCTTTGGCCACTGTCGTTAGTGGA 3'). The products obtained from these reactions were purified and amplified together with primers P7MCA and RPM2951. The resulting fragment was 0.9 kb in length. This fragment was digested with ClaI and XhoI. The fragment was ligated into pWZH30 backbone that had been digested with ClaI and XhoI and treated with calf intestinal phosphatase. All resulting mutants were sequenced on both strands to verify that the correct sequence had been mutated. In the table below, the primers used to construct the remaining mutants as well as the mutations that resulted are detailed.

Mutant Name	CAH sequence	Forward Primer	Reverse Primer
Wild Type	peereerirreteekeerrrtedeqkekerdrrrhremskll		
PSR33	p-----hremskll	EWdelfor	EWdelrev
SR11	peereerirreteekeer-----rdrrrhremskll	EW11aaF	Raa11DWE
SR15	peereerirretee-----rdrrrhremskll	EWD15aaF	Raa51DWE
EW-13	peereerirrete--eer-----rdrrrhremskll	EW-17-1for	EW17-1-rev
EW-19-1	peereerirrete-----rrhremskll	EW19-1-for	EW19-1-rev
EW 19-2	pee----irretee-----rdrrrhremskll	EWdel19-2-for	EWdel19-2-rev
EW18	pe-----reteekeer-----rdrrrhremskll		
EW22U	PEEREERIRRET█-----HREMSKLL		
EW22D	PEEREERIRRETE-----█REMSKLL		
EW17-3res	PEEREERIRRETEE-----RE█RHREMSKLL	EWdel19-2-for	EWdel19-2-rev
EW20		EW21-2-for	EW21-2-rev

Table 1. Primers used to construct EW mutants

Primer Name	Sequence
P7MCA	5' cgg tgc ggg gcg atg atg ggc-3'
RPM2951	5' aac ctg agc tcc tga tcc ctc-3'
EWdelfor	5'-gca tag aga gat gag caa gct ttt ggc cac tgt cgt tag tgg act tta ata aac gag aaa ccc cgc ata gag aga tga gca agc ttt tgg cca ctg tcg tta gtg ga-3'
EWdelrev	5' cca aaa gct tgc tca tct ctc tat gcg ggg ttt ctc gtt tat taa ag- 3'
EWD15aaF	5'-aga gat cgt agg aga cat aga gag atg agc aag ctt tta gcc act gtc gtt agt gga cag-3'
Raa51DWE	5'-aag ctt gct cat ctc tct atg tct cct acg atc tct ttc ctc tgt ttc tct cct gat acg ttc c-3'
EW11aaF	5'-aga gat cgt agg aga cat aga gag atg agc aag ctt ttg gcc act gtc gtt agt gga c-3'
Raa11DWE	5'-aag ctt gct cat ctc tct atg tct cct acg atc tct gcg ttc ttc ttt ttc ctc tgt ttc tct cc-3'
EW-17-1for	5'-cga gaa acc cct gag gaa aga gag gaa cgt atc agg aga gaa aca gag gaa gaa cgc aga gat cgt agg-3'
EW17-1-rev	5'-acg atc tct gcg ttc ttc ctc tgt ttc tct cct gat acg ttc ctc tct ttc ctc agg gg-3'
EW19-1-for	5'-cga gaa acc cct gag gaa aga gag gaa cgt atc agg aga gaa aca gag agg aga cat aga gag atg agc aag cta ttg gcc-3'
EW19-1-rev	5-ggc caa tag ctt gct cat ctc tct atg tcc cct ctc tgt ttc tct cct gat acg ttc ctc tct ttc ctc ag-3'
EW21-2-for	5'-gaa aaa att ttt aat aaa cga gaa acc cgg gaa aga gaa aca gag gaa aaa gaa gaa cgc aga gat cgt agg aga cat aga gag atg agc-3'
EW21-2-rev	5'-gcg ttc ttc ttt ttc ctc tgt ttc tct ttc cgg ggt ttc tcg ttt att aaa aat ttt ttc-3'
EW17-3res-for	5' cct gag gaa aga gag gaa cgt atc agg aga gaa aca gag gaa aga gag aga cat aga gag atg agc aag cta ttg gcc
EW17-3res-rev	5' gct cat ctc tct atg tct ctc tct ttc ctc tgt ttc tct cct gat acg ttc c-3'
EW22-U-for	5' tc ttt aat aaa cga gaa acc cct gag gaa aga gag gaa cgt atc agg aga gaa aca agg aga cat aga gag atg agc aag cta ttg gcc 3'
EW22-U-rev	5' act aac gac agt ggc caa tag ctt gct cat ctc tct atg tct cct tgt ttc tct cct gat acg ttc ctc tct ttc c-3'
EW22-D-for	5' tc ttt aat aaa cga gaa acc cct gag gaa aga gag gaa cgt atc agg aga gaa aca gag agg aga aga gag atg agc aag cta ttg gcc 3'
EW22-D-rev	5' act aac gac agt ggc caa tag ctt gct cat ctc tct tct cct ctc tgt ttc tct cct gat acg ttc ctc tct ttc c-3'
EWdel19-2-for	5'-agg cag aaa ttt aaa ttt tta ata aac gag aaa ccc cgg aag aaa tca gga gag aaa cag agg aaa gag atc gta gg-3'
EWdel19-2-rev	5'-tcc tac gat ctc ttt cct ctg ttt ctc tcc tga ttt ctt cgg ggg ttt ctc gtt tat taa aaa ttt ttt ctg cc- 3'

Data from unpublished mutant 19-2.

Mutant 19-2 is not replication-competent. 293T cells were transfected with 19-2, SR2, and mlv-env expressing plasmids. Virus was harvested from the supernatants of the transfected cells and used to infect D17 cells. Five independent experiments with 19-2 were conducted. Four of the infections showed that 19-2 produced a titer of <1 cfu/mL (Table 1A and B). The fifth experiment was used for RNA packaging studies only, so titers were not measured with this set. A portion of the virus from each experiment was also used to measure the RT activity of the virus. For each experiment, the RT activity of mutant 19-2 was not significantly above background. The average RT activity for all five experiments was 0.6%. This indicates therefore that mutant 19-2 was not replication competent.

Table 1A.

Titers	number of exp.	titer, cfu/mL	RT assay, % wt
SR2	6	<1	0%
WZ30	5	(avg 4) 8.6×10^3	100%
SR33	5	<1	0.2%
SR11	3	(avg 2) 4×10^3	400%
SR15	4	(Avg 3) 2×10^4	88%
FeEW	3	(avg 2) 3×10^4	126%
17-2	4	(avg 3) <1	0
17-3	4	(avg 3) <1	0
19-1	4	(avg 3) 1.3×10^4	1140%
19-2	5	<1	0.6%
21-1	3	<1	1%
13-1	2	<1	1%
18-2	2	<1	1%

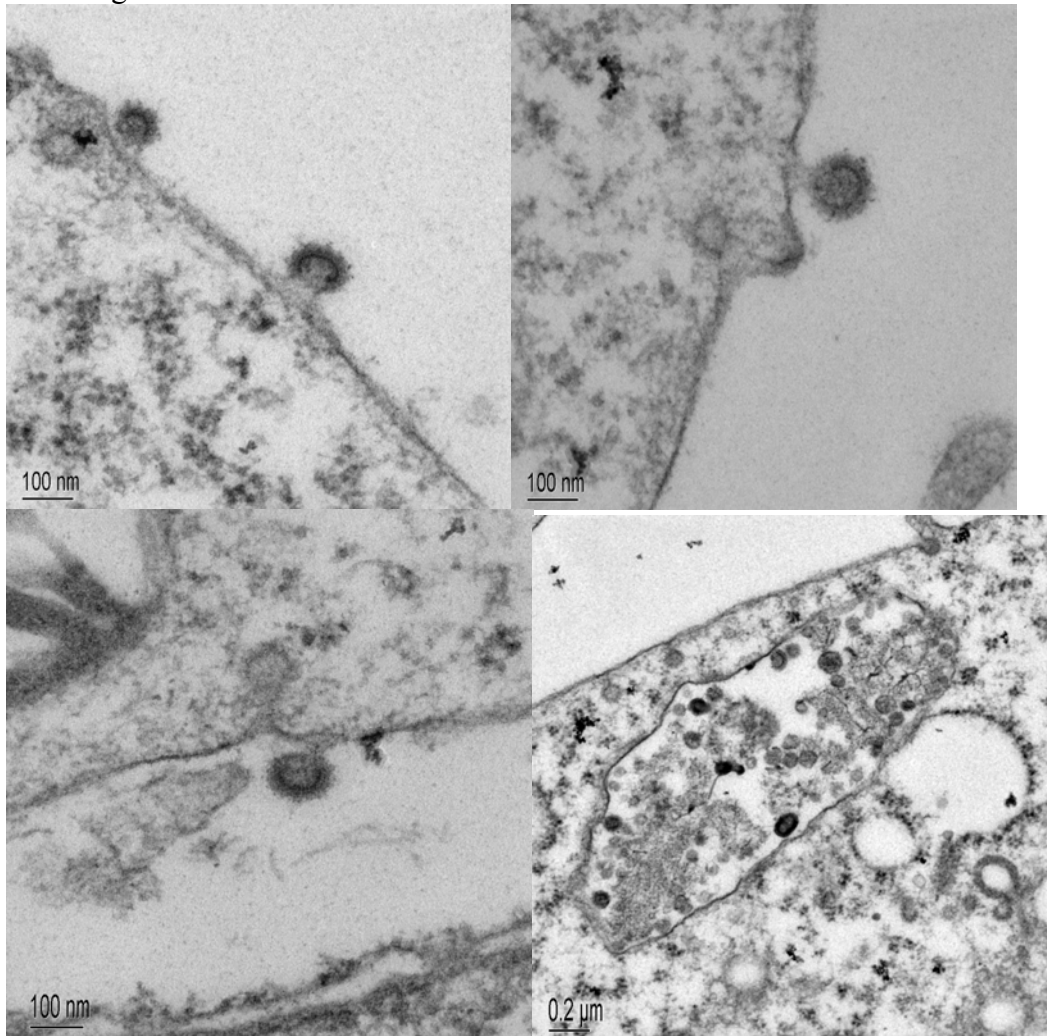
Table 1B

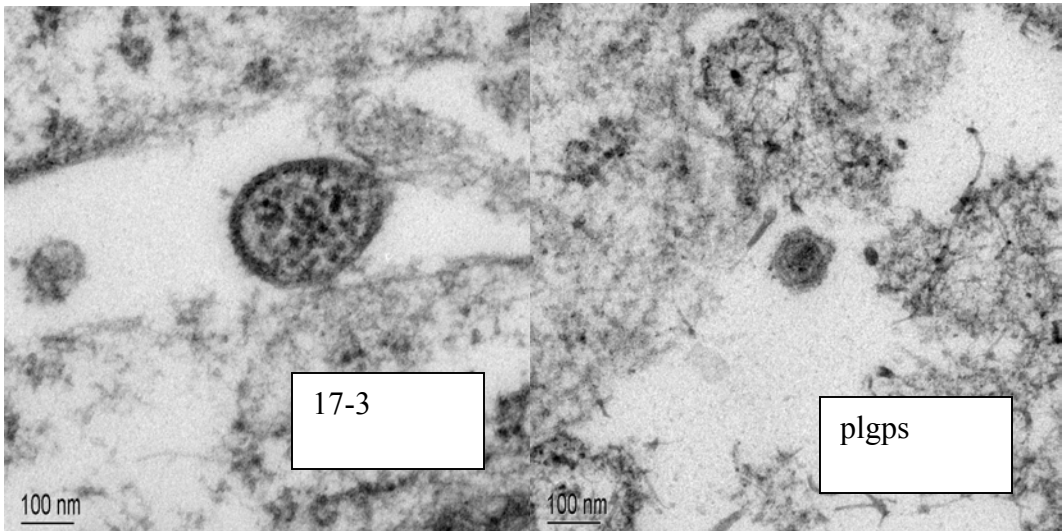
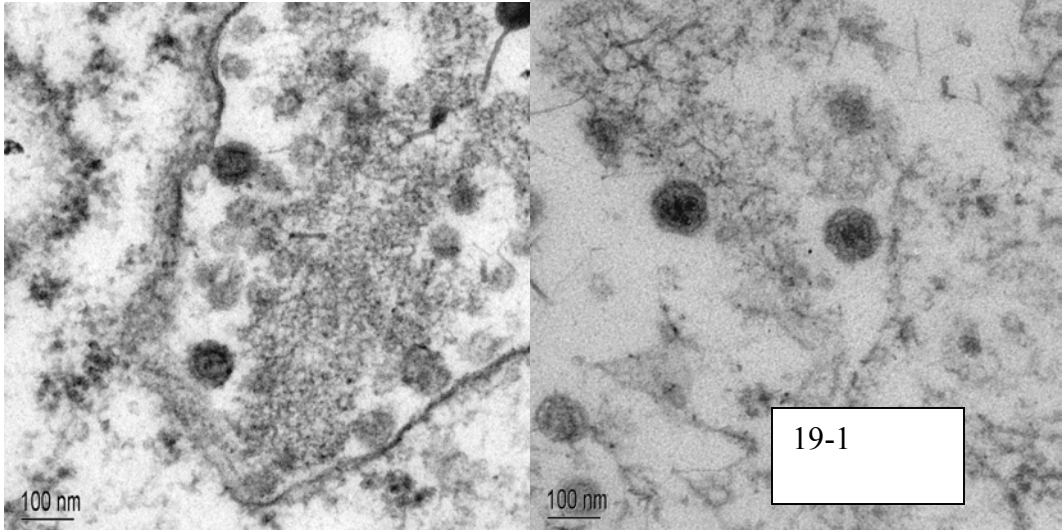
Exp #/date	Vector	RT Assay	% wt RT activity	Titer
2/11/02	19-2	5354	0	<1
2/5/02	19-2	11177	1%	
4-14-02	19-2	2414	0	<1

4-8-02	19-2	5022	2%	<1
3-24-02	19-2	6290	0	ND

The budding properties of mutant 19-2 are different from wild type and other EW mutants. **Following transfection of 293T cells, a portion of transfected cells were sent for electron microscopy analysis. Five pictures of cells producing vector 19-2 were obtained. Unlike the other replication-incompetent EW mutants, 19-2 maintains an ability to form wild-type sized particles. However, several particles were observed to be apparently budding *into* the cell, instead of to the extracellular space.**

EM Images of 19-2





Western on cell lysates for 19-2:

Westerns on Virus Lysates

