

**NOVEL RNA AND PROTEIN SEQUENCES INVOLVED IN
DIMERIZATION AND PACKAGING OF HIV-1 GENOMIC
RNA**

by

Rodney S. Russell

Microbiology & Immunology

McGill University, Montreal

May, 2004

A thesis submitted to McGill University in partial fulfillment of the requirements of the
degree of Doctor of Philosophy.

© Rodney S. Russell, 2004



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 0-494-06336-X
Our file *Notre référence*
ISBN: 0-494-06336-X

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

ABSTRACT

During HIV-1 assembly, the Gag structural protein specifically encapsidates two copies of viral genomic RNA in the form of a dimer. An RNA stem-loop structure (SL1) in the 5' untranslated region, known as the dimerization initiation site (DIS), is important for dimerization and packaging of HIV-1 genomic RNA; however, the mechanisms involved are not fully understood. The major goal of this PhD study was to further understand HIV-1 RNA dimerization, and to study the role of the Gag protein in the dimerization and packaging processes. Despite the known involvement of the DIS in RNA dimerization, DIS-mutated viruses still contain significant levels of dimerized RNA, and electron microscopy studies suggest that the RNA molecules are linked at the extreme 5' end. We show here that RNA sequences on both sides of the DIS are also required for HIV-1 genome dimerization, suggesting that multiple RNA elements are involved. We have also examined the contribution of specific amino acids within Gag to the dimerization and packaging processes. Previous work showed that partial deletion of the DIS impacted on viral replication capacity, but could largely be corrected by compensatory point mutations within Gag. To further elucidate the mechanism(s) of these compensatory mutations, we generated DIS mutants lacking the entire SL1, or only the SL1 loop sequences, and combined these deletions with various combinations of compensatory mutations. Analysis of virion-derived RNA showed that the relevant mutant viruses contained increased levels of spliced viral RNA compared to wild type, indicating that a defect in genome packaging specificity was present. However, this defect was corrected by our compensatory mutations, and a T12I substitution in p2 was shown to be solely responsible for this activity. These results suggest that the p2 spacer peptide plays a critical role in the specific packaging of viral genomic RNA. In summary, these findings provide new insight into the RNA-RNA and RNA-protein interactions involved in HIV-1 RNA dimerization and packaging.

RÉSUMÉ

Pendant l'assemblage du VIH-1, la protéine structurale Gag encapside spécifiquement deux copies de l'ARN génomique viral sous la forme d'un dimère. Une structure d'ARN tige-boucle (SL1), située dans la région non-transcrite en 5', nommée site d'initiation de dimérisation (DIS), est importante pour les processus de dimérisation et d'encapsidation. Toutefois, les mécanismes impliqués ne sont pas complètement élucidés. Le but majeur de ce projet fut de caractériser davantage la dimérisation de l'ARN du VIH-1, ainsi que d'étudier le rôle de la protéine Gag dans la dimérisation et l'encapsidation. Malgré l'implication connue du DIS dans le processus de dimérisation, les virus mutés dans cette région contiennent des niveaux significatifs d'ARN dimérisé, et des études de microscopie électronique suggèrent que les molécules d'ARN sont liées par leur extrémité en 5'. Nous démontrons que des séquences d'ARN flanquant de part et d'autre le DIS sont aussi requises pour le processus de dimérisation d'ARN, suggérant que de multiples éléments d'ARN peuvent être impliqués. Nous nous sommes aussi intéressés à la contribution des acides aminés de Gag dans la dimérisation et l'encapsidation. Des travaux précédents ont démontrés que la délétion partielle du DIS affectait la capacité de réplication virale mais était largement corrigée par des mutations ponctuelles compensatoires dans Gag. Pour élucider les mécanismes de ces mutations compensatoires, nous avons générés des mutants DIS manquant la région complète SL1, ou manquant la séquence bouclée du SL1, et nous avons combiné ces délétions à différentes combinaisons de mutations compensatoires. L'analyse de l'ARN dérivé de virus a montré que les virus mutants avaient contenu des niveaux supérieurs d'ARN viral épissé comparativement au type sauvage, indiquant un défaut dans la spécificité d'encapsidation du génome. Toutefois, ce défaut a été corrigé par les mutations compensatoires et cette activité a été attribué à la présence de la substitution T21I dans la région p2. En résumé, ces résultats donnent une nouvelle perspective des interactions ARN-ARN et protéines-ARN impliquées dans la dimérisation et l'encapsidation du génome du VIH-1.

PREFACE

This thesis was written in accordance with McGill University's "Guidelines for Thesis Preparation. The format of the thesis conforms to the "Manuscript-based thesis" option which states:

"Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis."

The contribution of co-authors to submitted or published articles, as well as journal of submission and information from published articles, appears on the title page of each chapter. Other manuscripts not included in this thesis, but to which a significant contribution was made by the candidate, are listed as follows:

Russell RS, Liang C, Wainberg MA. Recent studies in HIV-1 RNA dimerization. Manuscript in preparation for *Retrovirology*.

Roldan A, **Russell RS**, Marchand B, Götte M, Liang C, Wainberg MA. Early events in the protein-RNA interaction of HIV-1: An *in vitro* study. Manuscript in preparation for *J Biol Chem*.

Boulanger M-C, **Russell RS**, Liang C, Wainberg MA, Richard S. Arginine methylation of HIV-1 Tat. Manuscript in preparation for *J Virol*.

Roy BB*, **Russell RS***, Turner D, Liang C. Rescue of mutated nucleocapsid sequence by a second-site mutation within the SP1 region of human immunodeficiency virus type 1 Gag. Manuscript in preparation for *J Virol*. (*equal contribution)

Liang C, Hu J, **Russell RS**, Kameoka M, Wainberg MA. 2004. Spliced human immunodeficiency virus type 1 RNA is reverse transcribed into cDNA within cells. *AIDS Research and Human Retroviruses*. 20(2):203-11.

Guo X, Hu J, Whitney JB, **Russell RS**, Liang C. 2004. An important role for the CA-NC spacer region in the assembly of the bovine immunodeficiency virus Gag protein. *J Virol*. 78(2):551-60.

Liang C, Hu J, **Russell RS**, Wainberg MA. 2002. Translation of Pr55(gag) augments packaging of human immunodeficiency virus type 1 RNA in a *cis*-acting manner. *AIDS Res Hum Retroviruses*. 18(15):1117-26.

Liang C, Hu J, **Russell RS**, Roldan A, Kleiman L, Wainberg MA. 2002. Characterization of a putative alpha-helix across the capsid-SP1 boundary that is critical for the multimerization of human immunodeficiency virus type 1 gag. *J Virol*. 76(22):11729-37.

Kameoka M, Morgan M, Binette M, **Russell RS**, Rong L, Guo X, Mouland A, Kleiman L, Liang C, Wainberg MA. 2002. The Tat protein of human immunodeficiency virus type 1 (HIV-1) can promote placement of tRNA primer onto viral RNA and suppress later DNA polymerization in HIV-1 reverse transcription. *J Virol*. 76(8):3637-45.

Rong L, **Russell RS**, Hu J, Guan Y, Kleiman L, Liang C, Wainberg MA. 2001. Hydrophobic amino acids in the human immunodeficiency virus type 1 p2 and nucleocapsid proteins can contribute to the rescue of deleted viral RNA packaging signals. *J Virol*. 75(16):7230-43.

Kameoka M, Rong L, Götte M, Liang C, **Russell RS**, Wainberg MA. 2001. Role for human immunodeficiency virus type 1 Tat protein in suppression of viral reverse transcriptase activity during late stages of viral replication. *J Virol*. 75(6):2675-83.

Liang C, Rong L, **Russell RS**, Wainberg MA. 2000. Deletion mutagenesis downstream of the 5' long terminal repeat of human immunodeficiency virus type 1 is compensated for by point mutations in both the U5 region and *gag* gene. *J Virol*. 74(14):6251-61.

ACKNOWLEDGEMENTS

The first two people that I would like to acknowledge are my supervisor, Dr. Mark Wainberg, and my co-supervisor, Dr. Chen Liang. With them as mentors, I have truly had the best of both worlds during my PhD. Dr. Wainberg has provided me with all the resources I needed in order to carry out the work described in this thesis. His ambition and desire to push the boundaries of basic scientific research has inspired me to work hard, and his excellent reputation has opened doors for me that I never thought possible. Though Dr. Wainberg, I have seen what it takes to run a successful lab in a fast-moving field of research. But at the same time, there was Dr. Liang, who, when I started in the lab, still spent most of his time still doing bench work. Chen took me under his wing and taught me most of the scientific techniques that I know today. Working side by side, I think we made an excellent team, and I would like to think that my own hard work contributed, at least in part, to his establishing himself as one of the country's prominent young investigators. Through Chen, I learned that "you have to do good work, before you can do good science."

I would also like to acknowledge the various professors who sat on my committees over the course of my program. Drs. Nicholas Acheson and Lawrence Kleiman were especially helpful on my supervisory committee. I thank the Department of Microbiology & Immunology for giving me a chance to pursue a PhD at McGill, and a special thanks to Jennifer, for all her help and extended deadlines! I also have to thank Matthias Götte, who was never officially involved in my supervision, but from whom I learned a great deal.

I would like to sincerely thank the technicians that I've worked with over the last few years for their work and friendship that helped me get through this program, Cesar, Mervi, Maureen, Daniella, and Jing. Estrella and Monica, you guys always helped me, even with my most annoying requests! Thank you. Bonnie and Bluma, we have had many heated discussions, but in the end, I always knew I had both of your support when I needed it. During my time in the lab, I have met many people, from all over the world. Through them, I have learned about cultures and traditions that I never knew existed. And, I have made friends that I plan to keep forever. After almost five years, there are too many to name, but my true friends know who they are.

To the seven people who shared most of my Friday nights, we laughed, and we argued, but most of the time, we laughed hard. I doubt I'll find a group anywhere else in the world to replace you. Thank you for your support on so many levels, I will miss you all. I thank my family and friends back home. I've missed you as well, but when I needed you, you were always there. I especially have to thank my mother, who didn't want me to leave, but wouldn't let me stay, thank you mom.

And finally, Véro, you stood by me through a difficult, and the most uncertain time of my life. I can't describe in French or English how much your support really meant to me. I don't know what the future holds, but I thank you for all the love you gave me.

TABLE OF CONTENTS

ABSTRACT	ii
RÉSUMÉ	iii
PREFACE	iv
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES AND TABLES	xii
LIST OF ABBREVIATIONS AND DEFINITIONS	xv
1. CHAPTER 1 – INTRODUCTION	1
1.1 Global status of HIV/AIDS	2
1.2 The HIV-1 Virion	4
1.2.1 The Viral Genome and Proteins	4
1.2.2 The Structure of the Virion	10
1.3 The Viral Life Cycle	13
1.3.1 Attachment and Entry	13
1.3.2 Uncoating and Reverse Transcription	16
1.3.3 Nuclear Localization and Integration	16
1.3.4 Proviral Transcription and Protein Production	17
1.3.5 Assembly and Budding	19
1.3.6 Vif: A Mystery is Solved	25
1.4 Current and Future Therapies	27
1.5 HIV-1 RNA Structure and Function	29
1.5.1 Structural Elements in the HIV-1 5' UTR	29
1.5.2 HIV-1 RNA Packaging	33
1.5.3 The Discovery of Retroviral RNA Dimerization	36
1.5.4 Identification of the HIV-1 DIS	41
1.5.5 Viral Proteins Involved in Dimerization	47
1.5.6 Dimerization and Virus Replication	50
1.5.7 Remaining Questions	62

2. CHAPTER 2 – DEFICIENT DIMERIZATION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 RNA CAUSED BY MUTATIONS OF THE U5 RNA SEQUENCES	65
2.1 Preface to Chapter 2	66
2.2 Abstract	67
2.3 Introduction	68
2.4 Materials and Methods	69
2.4.1 Plasmid construction	70
2.4.2 Cell culture, transfection and infection	71
2.4.3 Analysis of viral proteins by Western blots	72
2.4.4 Analysis of viral RNA by Northern blots and RT-PCR	72
2.5 Results	74
2.5.1 Deletions of the GU-rich sequences in the HIV-1 U5 region cause deficient viral RNA dimerization and packaging	74
2.5.2 A second-site mutation can correct deficient RNA packaging, but not the diminished dimerization, caused by mutations of the GU-sequence [nt 99 to 108]	81
2.5.3 The palindrome residing on the top of the poly(A) hairpin does not contribute to RNA dimerization	84
2.5.4 The T12I (in p2) and V13G (in NC) point mutations restore the replication capacity of Δ GU3 but not of Δ GU2 to nearly wild-type levels	85
2.5.5 The T12I and V13G mutations compensate for defective RNA packaging caused by the Δ GU3 deletion	86
2.6 Discussion	89
3. CHAPTER 3 – SEQUENCES DOWNSTREAM OF THE 5' SPLICE DONOR SITE ARE REQUIRED FOR BOTH	

PACKAGING AND DIMERIZATION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 RNA	95
3.1 Preface to Chapter 3	96
3.2 Abstract	97
3.3 Introduction	98
3.4 Materials and Methods	99
3.4.1 Plasmid construction	99
3.4.2 Cell culture, transfection, and infection	101
3.4.3 Native Northern blotting	102
3.4.4 RNase protection assays	103
3.5 Results	104
3.5.1 SL3 is needed for both viral RNA packaging and dimerization	104
3.5.2 Mutations in GA-rich RNA sequences at nt 325 to 336 significantly reduce levels of both viral RNA dimerization and packaging	117
3.5.3 Additive effects of SL3 and GA-rich sequences on viral RNA packaging and dimerization	121
3.5.4 Mutation of the CU-repeats at nt 227 to 233 results in diminished levels of RNA dimerization and packaging	122
3.5.5 Mutations in two stretches of G-rich sequences neither affect viral replication nor viral RNA dimerization	125
3.6 Discussion	135
4. CHAPTER 4 – DELETION OF STEM-LOOP 3 IS COMPENSATED FOR BY SECOND-SITE MUTATIONS WITHIN THE GAG PROTEIN OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1	141
4.1 Preface to Chapter 4	142
4.2 Abstract	143
4.3 Introduction	144
4.4 Materials and Methods	145

4.4.1	Plasmid construction	145
4.4.2	Cell culture, transfection and infection	146
4.4.3	Analysis of viral proteins	147
4.4.4	Analysis of viral RNA by Northern blots	148
4.4.5	Identification of second-site mutations	149
4.5	Results	149
4.5.1	Long-term culture of the $\Delta(306-325)$ mutated viruses in MT-2 cells leads to the appearance of suppressor mutations in the <i>gag</i> coding region	149
4.5.2	The A11V mutation in p2 and the I12V mutation in NC increase levels of RNA packaging and dimerization in the $\Delta(306-325)$ mutated viruses	155
4.5.3	The A11V and I12V substitutions do not correct defective RNA dimerization caused by the $\Delta(306-325)$ deletion in the context of protease-negative viruses	163
4.6	Discussion	164
 5. CHAPTER 5 – EFFECTS OF A SINGLE AMINO ACID SUBSTITUTION WITHIN THE P2 REGION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 ON PACKAGING OF SPLICED VIRAL RNA		 167
5.1	Preface to Chapter 5	168
5.2	Abstract	169
5.3	Introduction	170
5.4	Materials and Methods	172
5.4.1	Plasmid construction	172
5.4.2	Cell culture, transfection, and infection	173
5.4.3	Native Northern blotting	174
5.4.4	RNase protection assays	175
5.5	Results	176
5.5.1	Compensatory mutations increase the replication capacity of the	

Δ DIS mutant	176
5.5.2 Compensatory mutations do not restore wild-type RNA dimerization to Δ DIS	183
5.5.3 The compensatory mutations restore wild-type RNA packaging to the Δ DIS and Δ Loop viruses while excluding spliced viral RNA from virions	187
5.5.4 The MP2 mutation alone corrects defects in packaging specificity seen in the Δ DIS and Δ Loop mutants	191
5.5.5 Mutation of T12 within p2 causes defective packaging of wild-type genomic RNA	191
5.6 Discussion	198
6. CHAPTER 6 - GENERAL DISCUSSION	203
6.1 RNA-RNA Interactions Involved in Dimerization and Packaging	204
6.2 RNA-Protein RNA Interactions Involved in Dimerization and Packaging	210
6.3 Novel Therapeutic Targets for HIV-1	220
6.3.1 The DIS as a therapeutic Target?	220
6.3.2 p2 as therapeutic target?	222
6.4 Final Summation	224
CONTRIBUTION TO ORIGINAL KNOWLEDGE	225
DEDICATION	229
REFERENCES	230

LIST OF FIGURES AND TABLES

Figure 1.1. The HIV-1 Genome	7
Table 1.1. Main Functions of HIV-1 Proteins and Peptides	8,9
Figure 1.2. The HIV-1 Virion	12
Figure 1.3. The Viral Life Cycle	15
Figure 1.4. HIV-1 5' RNA Structural Elements	32
Figure 1.5. The Kissing-Loop Model of HIV-1 RNA Dimerization	46
Figure 2.1. Various effects of two stretches of GU-rich sequences, located within the U5 region, on viral gene expression, viral replication, viral RNA packaging and dimerization	78, 80
Figure 2.2. Substitutions of the poly(A) stem or loop sequences and their effects on viral infectivity and viral RNA dimerization	83
Figure 2.3. Compensation of the Δ GU3 deletion by second-site mutations in Gag	88
Figure 3.1. Description of the mutations generated in the HIV-1 RNA sequences downstream of the 5' major splice donor (SD) site	106
Figure 3.2. Infectiousness of mutant viruses MD1 to MD3 and MS1 to MS6 in permissive cell lines	108
Figure 3.3. Effects of various mutations on viral RNA dimerization and packaging	112,114
Figure 3.4. Measurement of RNA dimer stability	116
Figure 3.5. Effects of RNA concentration on viral RNA dimerization	119
Figure 3.6. Mutation of the CU-repeat RNA sequence [nt 227 to 233] affects viral RNA dimerization and packaging	124
Figure 3.7. Analysis of the putative RNA-RNA interactions between the CU-repeats [nt 227 to 233] and the GA-rich sequences [nt 325 to 336]	127
Figure 3.8. Effects of the MG1, MG2, and MG12 mutations on viral replication	129

Figure 3.9 Effects of the MG1, MG2, and MG12 mutations on viral RNA dimerization and packaging	131
Table 3.1. Effects of various mutations on viral RNA packaging and dimerization as well as viral replication	132
Figure 3.10. Structural analysis of mutated and wild-type HIV-1 RNA sequences spanning nt 1 to 360 on the basis of the M-Fold program	134
Figure 4.1. Analysis of the $\Delta(306-325)$ deletion mutant	152
Figure 4.2. Compensation of the crippled replication of the $\Delta(306-325)$ mutated viruses by second-site mutations within the Gag protein	154
Figure 4.3. Effects of the A11V and I12V second-site mutations on deficient RNA packaging and dimerization of $\Delta(306-325)$	157
Figure 4.4. Thermostability of RNA dimers prepared from the $\Delta(306-325)$, $\Delta(306-325)$ -A11V-I12V and BH10 viruses	160
Figure 4.5. Effects of the $\Delta(306-325)$ deletion on RNA dimerization in the PR ^r virus	162
Figure 5.1. Illustration of the various RNA structural elements located within the HIV-1 5' region, including TAR, poly(A), U5-PBS, SL1 (DIS), SL2, SL3 (Ψ), and the <i>gag</i> gene	179
Table 5.1. Viruses generated in this study	180
Figure 5.2. Viral replication kinetics of the Δ DIS virus in the presence of various combinations of compensatory mutations	182
Figure 5.3. Effects of the Δ DIS and Δ Loop deletions on viral RNA dimerization	186
Figure 5.4. Effects of the Δ DIS and Δ Loop deletions on viral RNA packaging efficiency and specificity	189
Figure 5.5. Effects of the MP2 and MNC mutations on selective packaging of the full-length Δ DIS and Δ Loop RNA molecules	193
Figure 5.6. Effects of amino acid substitutions at position 12 in p2 on packaging specificity of wild-type RNA	196

Table 5.2. Effects of substitutions at T12 of p2 on overall versus specific packaging	197
Figure 6.1. Model of the MP2-induced alternate RNA-protein interaction	216

LIST OF ABBREVIATIONS AND DEFINITIONS

A (Ala)	alanine
A	adenine
ACS	autocomplementary sequence
Ag	antigen
AIDS	acquired immunodeficiency syndrome
AKR	an endogenous murine ecotropic virus
Ala (A)	alanine
ALV	avian leukosis virus
AMV	avian myeloblastosis virus
APOBEC3G	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
AZT	3'-azido-3'-deoxythymidine
BAF	barrier to autointegration
BIV	bovine immunodeficiency virus
BKD	an endogenous baboon virus
BMH	branched multiple hairpin
C (Cys)	cysteine
C	cytosine
CA (p24)	capsid
CA1	I91T revertant mutation CA
CBMCs	Cord blood mononuclear cells
CCR5	C-C chemokine receptor 5
cDNA	complementary DNA
CPE	cytopathic effects
cpm	counts per minute
CRM1	cellular export pathway
CXCR4	C-X-C chemokine receptor 4
Cys (C)	cysteine
D (Asp)	aspartic acid

dCTP	deoxycytidine triphosphate
DIS	dimerization initiation site
DLS	dimer linkage structure
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
E (Glu)	glutamic acid
E	encapsidation signal
EIAV	equine infectious anemia virus
ELISA	enzyme-linked immunosorption assay
EM	electron microscopy
<i>env</i>	HIV envelope gene
Env (gp160)	envelope glycoprotein
ESCRT	endosomal sorting complex required for transport
EtBr	ethidium bromide
F (Phe)	phenylalanine
FV	Friend virus
G (Gly)	glycine
G	guanine
G418	geneticin
Gag (Pr55)	group-specific antigen
Gag-Pol (Pr160)	Gag-Pol precursor polyprotein
<i>gag</i>	HIV Gag gene
Gly (G)	glycine
Gp41 (TM)	transmembrane glycoprotein; fusion peptide
Gp120 (SU)	surface glycoprotein
Gp160 (Env)	Envelope precursor glycoprotein
H (His)	histidine
HaSV	Harvey sarcoma virus
HIV-1	human immunodeficiency virus type 1
HIV-2	human immunodeficiency virus type 2
HIV-IIIB	an HIV-1 subtype B virus isolate

HMGI(Y)	high mobility group I(Y)
hr	hour
HXB2	an HIV-1 subtype B virus isolate
I (Ile)	isoleucine
ICAM-1	intercellular adhesion molecule-1
IL-2	interleukin 2
Ile (I)	isoleucine
IN	integrase
INI1	integrase interactor 1
K (Lys)	lysine
kb	kilobase
kcal	kilocalories
kD	kiloDalton
KiSV	Kirsten sarcoma virus
KLD	kissing-loop domain
L (Leu)	leucine
Lai	an HIV-1 subtype B virus isolate
LDI	long-distance interaction
LFA-1	leukocyte-function associated antigen
L-Glu	L-glutamate
LTR	long terminal repeat
M (Met)	methionine
M	molar; mol/L
MA (p17)	matrix
MA1	V35I revertant mutation in MA
MAbs	monoclonal antibodies
MAGI	multinuclear activation of a galactosidase indicator
Mal	an HIV-1 subtype A virus isolate
mCi	milliCurie
Met (M)	methionine
Mg ²⁺	magnesium ion

MgCl ₂	magnesium chloride
min	minute
ml	milliliter
MLV	Murine leukemia virus
mM	millimolar; mmol/L
MMTV	mouse mammary tumor virus
M-MuLV	Moloney murine leukemia virus
MNC	T24I revertant mutation in NC
MoMuLV	Moloney murine leukemia virus
MPMV	Mason-Pfizer monkey virus
MP2	T12I revertant mutation in p2
mRNA	messenger ribonucleic acid
MSV	Moloney Sarcoma Virus
MTV	mouse mammary tumor virus
MuLV	murine leukemia virus
MVB	multivesicular bodies
N (Asn)	asparagine
NaCl	sodium chloride
NaOAc	sodium acetate
NC (p7)	nucleocapsid
Nef	negative regulatory factor
NES	nuclear-export sequence
NFAT	nuclear factor of activated T cells
NF-κB	nuclear factor-kappaB
ng	nanogram
nm	nanometer
NMR	nuclear magnetic resonance
nt	nucleotide
NZB	an endogenous xenotropic virus
P	phosphorus
P (Pro)	proline

PAGE	polyacrylamide gels
PBMCs	peripheral blood mononuclear cells
PBS (buffer)	phosphate-buffered saline
PBS (RNA sequence)	primer-binding site
PCR	polymerase chain reaction
Phe (F)	phenylalanine
PIC	preintegration complex
<i>pol</i>	HIV Pol gene
poly(A)	polyadenylation signal
Pr55 (Gag)	Gag precursor polyprotein
Pr160 (Gag-Pol)	Gag-Pol precursor polyprotein
psiRNA	promoter-driven small interfering RNA
P-TEFb	transcription-elongation factor b
P(T/S)AP	Late domain motif
p1 (SP2)	spacer peptide 2
p2 (SP1)	spacer peptide 1
p6	peptide 6
p7 (NC)	nucleocapsid
p17 (MA)	matrix
p24 (CA)	capsid
Q (Gln)	glutamine
R	repeat region of the LTR
R (Arg)	arginine
RD-114	an endogenous feline retrovirus
RE	rabbit ears
REV	avian reticuloendotheliosis virus
Rev	regulator of gene expression
RNA	ribonucleic acid
RNAi	RNA interference
RNAPII	RNA polymerase II
RPA	RNase protection assay

rpm	rotations per minute
RRE	Rev response element
RSV	Rous sarcoma virus
RT	reverse transcriptase
RTC	reverse transcription complex
RT-PCR	reverse transcriptase-PCR
R5	CCR5-using
R5X4	CCR5- and CXCR4-using isolates; dual tropic
S (Ser)	serine
S	sulfur
S	Svedberg unit
SD	splice donor
SDS	sodium dodecyl sulfate
siRNA	small interfering RNA
SIV	simian immunodeficiency virus
SL	stem-loop
SP1 (p2)	spacer peptide 1
SP2 (p1)	spacer peptide 2
SU (gp120)	surface glycoprotein
SV40	simian virus 40
T (Thr)	threonine
T	thymine
TAR	transactivation response element
Tat	transcriptional transactivator
TCID ₅₀	tissue culture 50% infectious dose
Thr (T)	threonine
T _m	melting temperature
TM (gp41)	transmembrane glycoprotein
TN	tris sodium chloride
Tris-HCl	tris-hydrochloric acid buffer
T-20	fusion inhibitors

U	units of enzyme
U	uracil
UTP	uridine triphosphate
UTR	untranslated region
U3	unique 3' RNA sequences
U5	unique 5' RNA sequences
V (Val)	valine
v	volts
Val (V)	valine
Vif	viral infectivity factor
VLP	virus-like particle
Vpr	viral protein R
Vpu	viral protein U
W (Trp)	tryptophan
WoMV	woolly monkey sarcoma virus
X4	CXCR4-using
Y (Tyr)	tyrosine
β -gal	β -galactosidase
ΔG	free energy; kcal/mol
μ l	microliter
ψ	Encapsidation/packaging signal

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

In this chapter, I will provide a brief description of the current global status of HIV/AIDS, and highlight some of the significant advances that have been made in the field of HIV research since the discovery of the virus. I will then give a detailed description of the viral genome, structure, and life cycle, followed by a comprehensive review of the literature that forms the basis for the objectives and hypotheses that will be addressed later in this thesis.

1.1 Global Status of HIV/AIDS

In July of 1999, when I sat down to write my MSc thesis, it was estimated that there were nearly forty million people worldwide living with HIV, and almost twelve million people had already died¹⁷⁷. Today, in March of 2004, while I sit down to write my PhD thesis, there are still forty million people in the world living with HIV, but the number of deaths has risen to over twenty million. And the most upsetting statistic is that the majority of the people living with HIV can still be found in sub-Saharan Africa and Southeast Asia, where most of them don't have access to antiretroviral therapy, and many are children under the age of fifteen²⁶⁹.

Despite these discouraging statistics, the body of knowledge that has been gathered on the topics of HIV and AIDS in only two decades is quite astonishing. Although it has been debated as to whether AIDS represents the most devastating disease known to man, there is no argument against the fact that HIV represents the single most studied pathogen

in history. In only twenty years of known existence, there have been over 135,000 scientific articles published on HIV, compared to malaria at 36,000, and influenza at 35,000, both of which had 30 years of documented research before HIV was discovered. In just the five years that I've been doing my PhD, there have been 55,000 articles related to HIV registered in the PubMed database. I haven't had time to read them all. But despite this enormous body of literature, there is still no effective vaccine available, and although current antiretroviral therapy has significantly decreased AIDS morbidity and mortality, these drugs do not cure, and problems of drug resistance and establishment of latent pools of virus are inevitable in almost every infected individual. So what have we learned in twenty years?

After Luc Montagnier's group in France obtained experimental evidence for an association between a retrovirus and AIDS in 1983¹⁵, and Robert Gallo's group at the NIH proved that HIV was indeed the cause of AIDS²²⁴, a frenzy of research in the field of molecular virology led to the rapid characterization of the nine viral genes and the identification of all fifteen viral proteins¹⁰⁵. Since then, it has been determined that HIV-1 probably originated as a result of an accidental species jump from the *Pan troglodytes troglodytes* species of chimpanzees⁹³. Sensitive and specific antibody tests were developed in the same year that the virus was identified as the causative agent for AIDS²⁴¹, which allowed large-scale screening of blood and blood products, resulting in the prevention of potentially millions of new infections. The immunopathogenesis of HIV infection, although quite complex and still not fully understood, has been studied intensively²³¹. And in trying to understand the complex interplay between HIV and the human immune system, many basic immunological concepts have been rapidly

characterized, of which the chemokines and their receptors represents an excellent example⁷⁶. The first effective drug against HIV, i.e. AZT, was identified only five years after the discovery of the virus itself¹⁹², and today there are twenty different approved compounds that together target four different steps in the viral life cycle, and many more are in development⁸³. With respect to vaccines, there is still heated debate over what actually defines protective immunity against HIV infection, but significant progress has been made²⁰⁸, and numerous vaccine candidates are now in various stages of preclinical and clinical development⁸³.

Despite all of these significant advances, the HIV pandemic still rages throughout the world, and even if an effective vaccine is developed soon, which is highly unlikely, we will still face the challenge of treating the forty million that are already infected with HIV. Therefore, the need for new and better anti-HIV therapy is as strong as ever, and the development of such therapies will depend on continued basic virological research and the complete elucidation of the HIV-1 life cycle. In this regard, I will focus the remainder of this chapter on the description of what is known about the viral life cycle, and highlight some questions that still remain.

1.2 The HIV-1 Virion

1.2.1 The Viral Genome and Proteins

The HIV-1 proviral DNA consists of a typical retroviral genome with *gag*, *pol*, and *env* genes flanked by 5' and 3' long terminal repeats (LTRs) containing U3, R, and U5

regulatory elements (Fig. 1.1, top). Once the provirus is transcribed by the host cell's transcriptional machinery, a 9.2kb mRNA is produced that is capped and polyadenylated like normal cellular mRNA (Fig. 1.1, middle). Within the 5' untranslated region of this mRNA, which also acts as the viral genomic RNA, are a number of stem-loop (SL) RNA elements that are important for various aspects of the viral life cycle. These include the transactivation response element (TAR), the primer-binding site (PBS), the dimerization initiation site (DIS), the 5' major splice donor (SD) and the major packaging signal (SL3; ψ). The genome contains nine open reading frames (Fig. 1.1, bottom; viral proteins and functions listed in Table 1.1), the first of which codes for the group-specific antigen (Gag; Pr55), the major structural protein, which is cleaved by the viral protease into matrix (MA; p17), capsid (CA; p24), nucleocapsid (NC; p7), p6, and two spacer peptides, p2 (SP1) and p1 (SP2). The *pol* gene encodes the three essential viral enzymes, protease (PR), integrase (IN), and reverse transcriptase (RT), which also contains an RNase H domain. An infrequent -1 ribosomal frameshift at a slippery sequence near the 3' end of the *gag* coding region results in a readthrough event and allows the production of a Gag-Pol (Pr160) precursor protein. Splicing of the 9.2kb mRNA results in the production of a 160kD envelope glycoprotein (gp160), which is cleaved by cellular proteases into the surface (SU; gp120) and transmembrane (TM; gp41) glycoproteins. HIV also encodes six proteins that used to be known as accessory proteins, but have since been realized to play various important roles in the viral life cycle. These include the transcriptional transactivator (Tat), regulator of gene expression (Rev), negative regulatory factor (Nef), viral infectivity factor (Vif), viral protein R (Vpr), and viral protein U (Vpu) (Reviewed in Coffin *et al.*⁵³)

Figure 1.1. The HIV-1 Genome. Schematic representation of the HIV-1 proviral DNA and genomic RNA organization, as well as the open reading frames coding for the 15 viral proteins. The proviral DNA consists of a typical retroviral genome with *gag*, *pol*, and *env* genes flanked by 5' and 3' LTRs containing U3, R, and U5 sequences. The primary mRNA transcript, which is capped, and polyadenylated, serves as both mRNA and genomic RNA. MA, CA, and NC, along with 3 peptides, p2, p1, and p6, are cleaved from Gag, while PR, RT, and IN are generated from Pol. Env, Vpr, Vpu, and Vif are translated from partially spliced transcripts, while mRNAs coding for Tat, Rev, and Nef are generated by multiple splicing events. (Adapted from Coffin *et al.*⁵³)

Figure 1.1. The HIV-1 Genome.

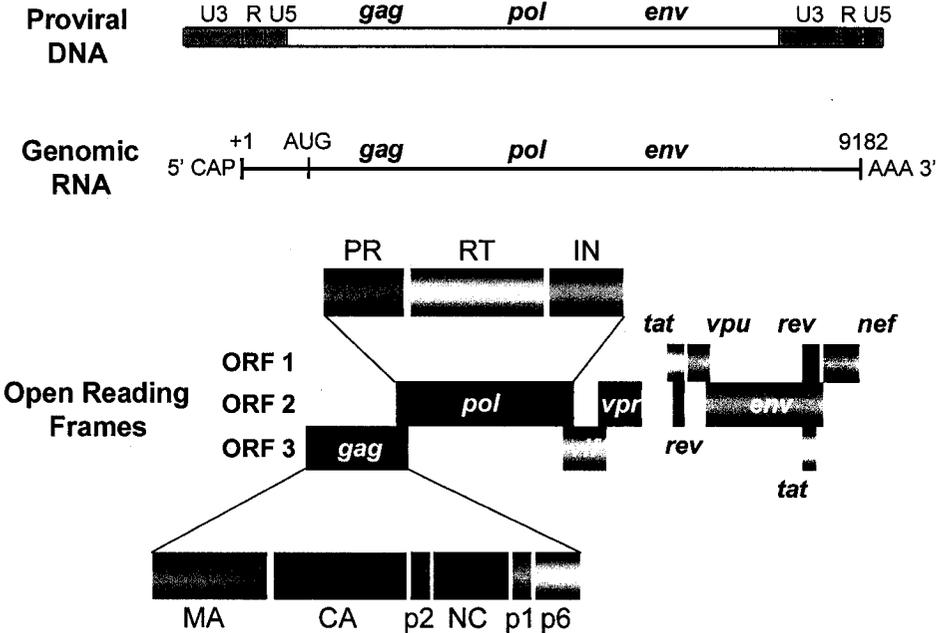


Table 1.1. Main Functions of HIV-1 Proteins and Peptides

Protein	Function
Gag (Pr55)	Precursor polyprotein for MA, CA, p2, NC, p1, and p6; main structural component of the immature virion; specifically binds to viral genomic RNA to facilitate encapsidation ⁵³ ; directs the virion assembly process; targets assembly to the plasma membrane; self-multimerizes to allow particle formation; directs the budding and release of the virus by recruiting vacuolar sorting complexes ⁴⁶
Matrix (MA; p17)	Main structural component of the mature virion; contains the M domain, consisting of a myristoylation signal and a basic amino acid cluster that help Gag target to the plasma membrane; proposed to form trimers to assist in virion assembly ⁴⁶ ; assists in nuclear localization of the PIC ¹⁰⁵
Capsid (CA; p24)	Multimerizes to form a cone-shaped core ⁵³ ; binds cyclophilin A ¹⁰⁵
p2 spacer peptide (SP1)	In the context of Gag, required for higher order multimerization ¹⁹⁹ and virus production ^{2,141} ; contains key amino acid residues that affect Gag multimerization and membrane targeting ¹⁶⁵ ; affects genome packaging specificity ¹³¹
Nucleocapsid (NC; p7)	Contains the I domain which provides the major contribution to Gag-Gag interaction; binds specifically to viral genomic RNA to direct packaging; binds non-specifically to viral genomic RNA within the core ⁴⁶ ; increases the efficiency of tRNA placement on the PBS; assists in strand transfer and RNA dimer maturation ²²⁶
p1 spacer peptide (SP2)	Proposed to cooperate with the L domain to facilitate budding ²⁵⁷
p6 peptide	Binds Vpr to incorporate it into the virion ¹⁰⁵ ; contains the L domain that binds Tsg101 to facilitate virus budding and release ⁹⁴ ; binds AIP1 to facilitate virus budding and release ²⁵⁸
Gag-Pol (Pr160)	Brings PR, RT, and IN into the virion as part of the precursor polyprotein; cleavage generates MA, CA, p2, NC, PR, RT, and IN ⁵³ ; required for packaging of tRNA ¹⁷³
Protease (PR; p10)	Cleaves viral precursor polyproteins (Gag and Gag-Pol) into mature final products ⁵³
Reverse transcriptase (RT; p66/51)	Reverse transcribes single-stranded viral genomic RNA into double-stranded proviral DNA; contains an RNase H domain which digests viral RNA as reverse transcription proceeds ⁵³

Table 1.1. (cont.) Main Functions of HIV-1 Proteins and Peptides

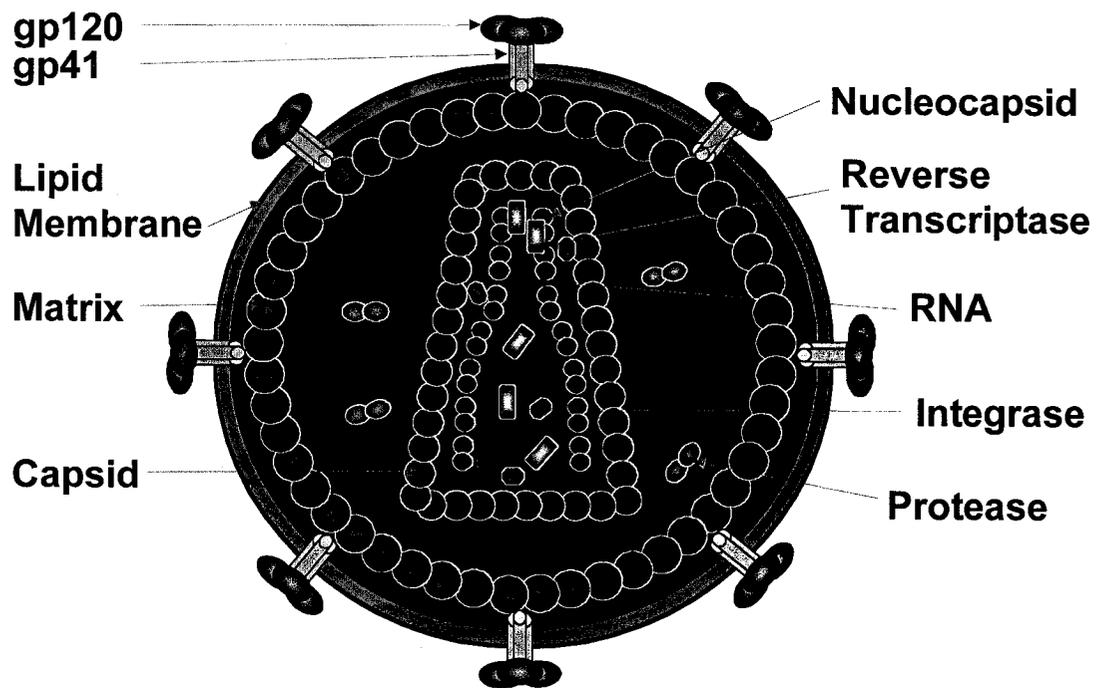
Protein	Function
Integrase (IN; p32)	Integrates proviral DNA into the host chromosome ⁵³ ; influences nuclear localization of the preintegration complex (PIC) ¹⁰⁵
Envelope (Env; gp160)	Cleaved by cellular proteases to produce SU and TM ⁵³ ; interacts with Gag during virion assembly and may affect assembly site selection ⁴⁶
Surface glycoprotein (SU; gp120)	Viral receptor; responsible for binding to CD4 and to coreceptors ¹⁰⁵
Transmembrane glycoprotein (TM; gp41)	Fusion peptide; mediates fusion of the viral envelope with the plasma membrane ¹⁰⁵
Tat (p14)	Binds TAR RNA structure to enhance RNAPII elongation of the viral DNA template; proposed to inhibit premature reverse transcription ¹⁶⁶
Rev (p19)	Binds RRE to inhibit viral RNA splicing; promotes nuclear export of incompletely spliced viral RNA ¹⁰⁵
Nef (p24)	Downregulates surface expression of CD4 and MHC I; inhibits apoptosis of infected cells; enhances virion infectivity; alters activation state of infected cells; increases the rate of disease progression in HIV-infected individuals ¹⁰⁵
Vpu	Promotes CD4 degradation; influences virion release ¹⁰⁵
Vpr (p15)	Induces cell cycle arrest in infected cells; influences nuclear localization of the PIC ¹⁰⁵
Vif (p23)	Binds viral genomic RNA in regard to an unknown function ⁷⁰ ; blocks the activity of an innate antiviral factor, CEM15 ²⁴⁵ (APOBEC3G)

1.2.2 The Structure of the Virion

As shown in Fig. 1.2, the HIV-1 virion contains two identical copies of positive-sense full-length viral genomic RNA, non-covalently linked at the 5' ends in the form of a dimer. The RNA is extensively coated by the NC protein, and closely associated with RT. This ribonucleoprotein complex, along with the IN enzyme, are contained in a cone-shaped core consisting of hexameric rings of CA protein. An outer shell, consisting of MA protein, reinforces the structure of the virus and maintains the integrity of the virion itself (Reviewed in Coffin *et al.*⁵³). The viral envelope is composed of a host cell-derived lipid bilayer that is acquired during the budding process, and therefore contains an array of cellular membrane proteins, such as leukocyte-function associated antigen (LFA-1) and intercellular adhesion molecule-1 (ICAM-1), both of which may actually play a role in HIV-1 infection¹¹⁷. The viral gp120 and gp41 glycoproteins are embedded within this lipid envelope and fit together in a knob-and-socket-like trimeric structure and serve as the viral receptors²⁴³. The viral proteins PR, Nef, Vpr,⁸⁹ and Vif¹³² are also incorporated into the virion, along with a variety of cellular factors whose functions will be discussed in detail below²¹².

Figure 1.2. The HIV-1 Virion. Schematic representation of the mature virion, showing the host cell-derived lipid membrane envelope and the protruding virus receptor, gp160. MA and CA proteins make up the virion and core structures, respectively. The genomic RNA, shown in red, is coated with NC. Essential viral enzymes, RT, IN, and PR are also shown. Colors correspond to the proteins shown in Fig. 1.1. (Adapted from Coffin *et al.*⁵³)

Figure 1.2. The HIV-1 Virion.



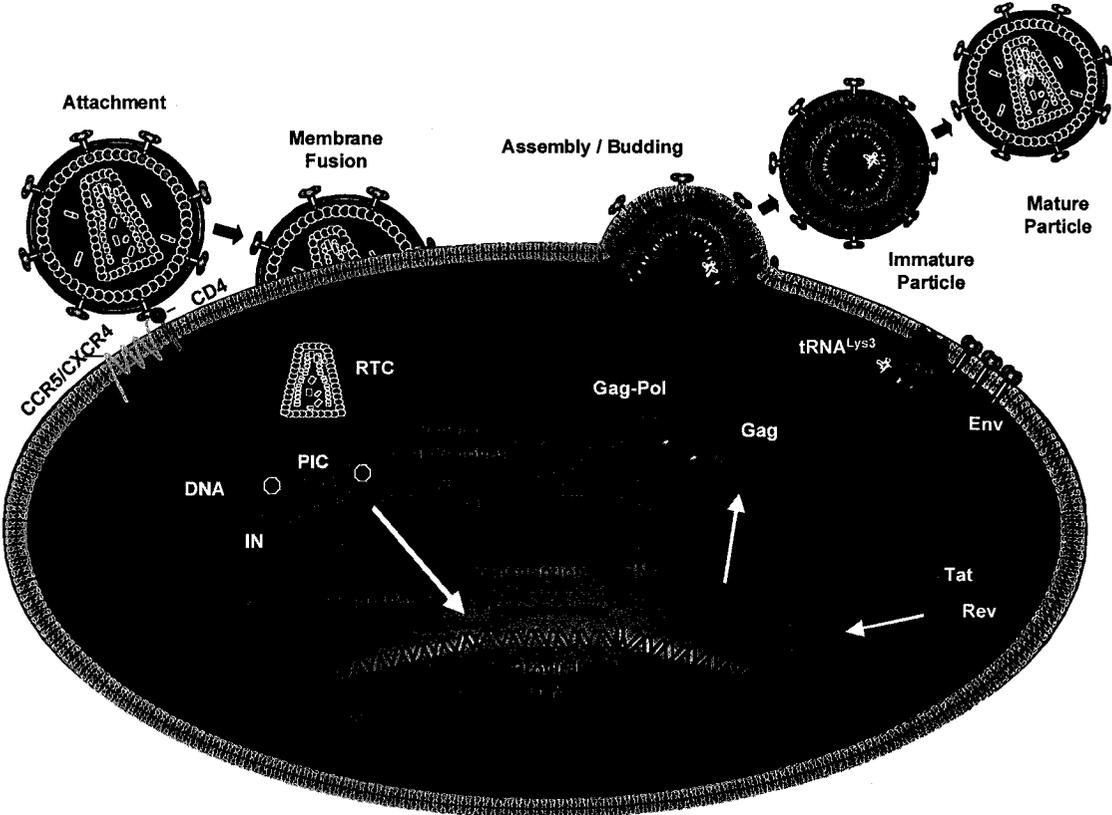
1.3 The Viral Life Cycle.

1.3.1 Attachment and Entry

Like all viruses, HIV-1 takes advantage of many cellular proteins and processes in order to make its way into the host cell and carry out its life cycle. The first such cellular factor to be identified was the CD4 receptor^{58,138}, to which HIV directly binds through a specific interaction with the viral receptor gp120¹⁹¹ (Fig. 1.3). Although the CD4 molecule was found to be the cellular receptor for HIV entry very soon after the virus was identified, it wasn't until ten years later that it was realized that the virus also needs to bind to a coreceptor in order to gain entry into the cell. These coreceptors, which turned out to be chemokine receptors normally involved in lymphocyte homing, bind to the viral envelope protein, gp120, after a conformational switch induced by the initial interaction with CD4^{6,44,69,73,75,84}. HIV can utilize a number of these chemokine receptors to gain entry into a variety of cell types, but it is clear that the two most commonly used are CCR5 and CXCR4⁷². Today, HIV isolates are classified as R5 (CCR5-using), X4 (CXCR4-using), or R5X4 for dual tropic isolates, based on coreceptor preference. One interesting point from tropism studies is that R5 viruses appear to mediate both mucosal and intravenous transmission of HIV, whereas X4 isolates are more commonly found at later stages of disease⁷⁶. Equally important was the finding that a small proportion of the world's population carries a naturally occurring 32 base-pair deletion in the gene encoding CCR5 that is associated with almost 100% resistance to HIV infection^{121,167}.

Figure 1.3. The Viral Life Cycle. The virus attaches to the cell via an interaction between gp120 and the host receptors CD4 and CCR5 or CXCR4. This interaction exposes the fusion peptide, gp41, which triggers the fusion of the viral and plasma membranes, allowing the virus to enter the cell. Once inside, the virus uncoats and reverse transcription begins in the reverse transcription complex (RTC). After the viral RNA has been converted to double-stranded DNA, the preintegration complex (PIC) translocates to the nucleus where the viral DNA is integrated into the host cell's genome. Transcription of the provirus generates full-length viral RNA molecules that are spliced in the nucleus and transported to the cytoplasm to be translated. Once all of the necessary viral components are produced, the viral proteins and genomic RNA assemble at the membrane and an immature virus particle buds away from the cell. After budding from the plasma membrane, the virion undergoes a PR-dependent maturation into an infectious virus particle.

Figure 1.3. The Viral Life Cycle.



The fact that these individuals display no apparent deleterious immunological phenotype suggests that interference with the expression of this protein should not have any harmful side effects, and therefore identified CCR5 as a potential target for antiretroviral and gene therapy^{13,280}.

The molecular interactions between gp120, CD4, and the coreceptor induce a second conformational change that exposes gp41, also known as the fusion peptide. A trimer of gp41 projects into the lipid bilayer of the target cell resulting in the fusion of the viral and cell membranes, facilitating the release of the viral core into the interior of the cell⁴².

1.3.2 Uncoating and Reverse Transcription

The events that take place immediately after the core enters the cell are not nearly as well characterized as other steps in the viral life cycle. But what is known is that the virus somehow disassembles and forms what is referred to as the reverse transcription complex (RTC). The RTC contains the two copies of viral genomic RNA, tRNA^{Lys3}, which primes the reverse transcription reaction, RT, IN, MA, NC, Vpr, and a number of host proteins¹²⁸. Once liberated from the plasma membrane, the RTC associates with actin microfilaments³⁷. At this point, reverse transcription, which has already been initiated within the virion, continues via a complex mechanism involving two strand jumps, and results in the conversion of the single-stranded RNA genome into a double-stranded viral cDNA⁵³.

1.3.3 Nuclear Localization and Integration

Once reverse transcription is complete, a structure known as the preintegration complex (PIC) forms, which is composed of double-stranded viral cDNA, RT, IN, MA, Vpr, and a cellular factor termed the high-mobility group DNA-binding protein [HMGI(Y)]¹⁹⁶. The mechanism by which the PIC transports to the nucleus is not entirely clear, but it has been suggested that it actually moves toward the nucleus on microtubule-associated dynein motors¹⁹⁰. Nuclear-localization signals found on IN, MA, and Vpr, along with a cDNA flap that is generated during reverse transcription, all help the PIC migrate to, and enter, the nucleus. Although, it has also been shown that Vpr can mediate direct binding of the PIC to components of the nuclear pore complex, which would allow it to enter the nucleus independently of the importin system¹⁰⁵. In the nucleus, IN, along with host factors HMGI(Y) and BAF (barrier to autointegration) whose specific contributions remain unknown, efficiently integrates the viral DNA into the host chromosome⁴³.

1.3.4 Proviral Transcription and Protein Production

Once integrated into the host chromosomal DNA, the HIV provirus acts like a normal cellular gene. The U3 region of the 5'LTR contains a number of transcriptional promoter elements, including an initiator, a TATA-box, and three Sp1 sites. These elements are able to position RNA polymerase II (RNAPII) and initiate transcription of the provirus; yet, elongation is very inefficient at this point, producing only incomplete transcripts. However, the U3 regulatory region also contains a number of upstream transcriptional

enhancer binding sites for NF- κ B, NFAT, and Ets family members (Reviewed in Jones and Peterlin¹²⁵). Once the cell is activated, NF- κ B and NFAT migrate to the nucleus, and increase the efficiency of initiation and elongation of proviral transcription enough to generate a low number of full-length HIV transcripts. These first RNA molecules are spliced by the host cell's splicing machinery, and the regulatory proteins Tat and Rev, as well as Nef, are produced (Reviewed in Greene and Peterlin¹⁰⁵). Once Tat is made in the cytoplasm, it shuttles back into the nucleus where, in association with cyclin T1, it binds the transactivation response element (TAR) within the R region of newly-transcribed viral mRNAs. After binding to TAR, Tat recruits Cdk9 as part of the transcription-elongation factor b (P-TEFb) complex, which in turn phosphorylates the C-terminal domain of RNAPII, thereby increasing the efficiency of transcription elongation by as much as 100-fold (Reviewed in Liang and Wainberg¹⁶⁶).

After Tat has boosted the production of full-length viral genomic RNA into high gear, many of the transcripts get rapidly spliced and transported to the cytoplasm, but some singly spliced and unspliced transcripts remain in the nucleus. In order for the viral life cycle to move into the next phase, i.e. virus production, these viral RNA transcripts must be transported to the cytoplasm, and this step depends mainly on the levels of Rev protein in the nucleus⁵⁷. As Rev is produced from the multiply spliced transcripts and shuttles back into the nucleus, it binds to an RNA stem-loop known as the Rev response element (RRE) located in the *env* gene¹⁷⁴. The binding of Rev to the RRE leads to a multimerization of Rev, and once a threshold level of Rev has multimerized on a given RRE, the RNA transcript is transported out of the nucleus and into the cytoplasm via the nuclear-export sequence (NES) on Rev and the CRM1 cellular export pathway⁵⁷. Then, in

the cytoplasm, the partially spliced viral RNA transcripts are translated to produce Env, Vif, Vpr, and Vpu. The full-length unspliced viral RNA transcripts are translated, producing Gag and Gag-Pol precursor proteins, and also serve as genomes to be packaged into newly forming virions (Reviewed in Coffin *et al.*⁵³).

1.3.5 Assembly and Budding

With all of the components required to form a new infectious particle now available in the cytoplasm, retroviruses can generally utilize one of two types of assembly pathways. Virus particles can be preformed in the cytoplasm and transported to the plasma membrane; this is classified as assembly types B and D, as seen in the case of Mouse mammary tumor virus (type B) and Mason-Pfizer monkey virus (type D). HIV, on the other hand, is an example of a retrovirus that uses the type C pathway of virion assembly. In this case, virion constituents accumulate at the plasma membrane at which the necessary interactions and multimerizations take place that allow the virus to assemble into a particle and bud away from the cell⁵³. However, recent studies in macrophages have shown that HIV-1 virions can assemble and bud into multivesicular bodies (MVBs) within the cell, so this dogma might eventually change (Reviewed in Amara and Littman⁷). But for now, most of what we know about HIV-1 assembly has been derived from studies involving transfection of cell lines such as COS-7, HeLa, and 293T fibroblast-like cells with plasmids that express the complete viral genome, or even Gag alone.

Expression of the retroviral Gag polyprotein by itself is sufficient to direct the formation of virus-like particles (VLPs), and therefore Gag is believed to be the major player in the HIV-1 assembly process. In order to assemble a virus particle and get it out of the cell, Gag must perform a number of important functions, such as membrane targeting and genomic RNA binding, and therefore must contain all of the necessary signals and domains required to carry out these activities (Reviewed in Cimarelli and Darlix⁴⁶). The first of these activities, plasma membrane targeting, is facilitated by an N-terminal 31 amino acid stretch, known as the M domain²⁹⁰, which is sufficient to target heterologous proteins to the plasma membrane²²⁷. A cotranslational modification of Gag removes the first Met and adds a myristate to the first Gly¹⁰¹. This M domain also contains a cluster of basic amino acids between residues 17 and 31. The myristate is thought to insert hydrophobically into the lipid bilayer, and the basic amino acids form electrostatic interactions with the charged headgroups of the phospholipids on the inner leaflet of the membrane. It has also been shown that Gag does not bind to just any membrane, but has a preference for specific microdomains, known as rafts, that are enriched in sphingomyelin, glycosphingolipids, and cholesterol (Reviewed in Cimarelli and Darlix⁴⁶).

Another important feature of the Gag polyprotein that is key to virus assembly is its ability to multimerize. This feature allows approximately 1500-2000 molecules of Gag to self-associate, and form the basic virion particle structure²⁷⁶. Although the details of the multimerization process have yet to be fully determined, this process probably involves multiple contacts between different domains of Gag (Reviewed in Cimarelli and Darlix⁴⁶). The NC domain, for example, is proposed to be the major contributor to the

Gag-Gag interaction, and has therefore been termed the I domain²⁰. It has been suggested that the binding of NC, in the context of full-length Gag, to the viral genomic RNA, might concentrate Gag proteins onto one or more RNA molecules, thereby facilitating Gag-Gag multimerization in a template-driven manner. In this way, the viral genomic RNA has been termed as a structural element, or scaffold, on which the virion can assemble²⁰⁶. The crystal structure of MA is said to be trimeric¹¹⁵, and trimers of Gag that are dependent on MA have been found in solution¹⁹⁸, suggesting a role for MA in the multimerization process. *In vitro* studies using purified CA or CA-NC fusion proteins have shown that these domains can form core-like structures, which is strong evidence for the involvement of the CA domain in Gag multimerization⁴⁶. However, studies aimed at identifying the minimal determinants for multimerization have shown that only the C-terminal domain of CA along with the p2 spacer peptide are required for proper assembly of VLPs^{3,34}. It is also important to keep in mind that for every twenty Gag proteins in the virion, there is also one Gag-Pol molecule present, which raises the possibility for additional PR-PR, IN-IN, and RT-RT interactions to contribute to the multimerization process²⁰⁶.

Once Gag has carried out its membrane binding and multimerization functions, the virus particle begins to take shape just beneath the plasma membrane and is ready to begin its budding and release from the cell. Despite the fact that this VLP quite closely resembles a nascent virion, it is not yet an infectious virus particle. In order to be infectious, the virus must incorporate a number of components that contribute to its infectivity, such as the Env and Vif proteins. The most important component still absent from the virion at this point is viral genomic RNA that must be encapsidated into all

infectious retroviral particles. This critical function of RNA packaging is also carried out by the Gag polyprotein via an RNA-protein interaction between the NC domain of Gag and the packaging signal contained within the 5' untranslated region (UTR) of the viral genomic RNA. This interaction involves specific binding between the Cys-His boxes (Zn fingers) in NC and a series of four RNA stem-loop structures (SL1-4 in Fig. 1.4) known as the ψ or E element (Reviewed in Berkowitz *et al.*²⁷, Cimarelli and Darlix⁴⁶, Rein *et al.*²²⁶). Within this packaging element, the loop of SL3 has been shown by nuclear magnetic resonance (NMR) to bind with high specificity to both Zn fingers of NC, and is therefore considered to be the major packaging signal. Conserved basic amino acids clustered around the Zn fingers also contribute non-specifically to this interaction by forming a 3_{10} helix that penetrates the major groove of the SL3 stem⁶⁷. As a result of this interaction, two copies of viral genomic RNA are specifically selected from the pool of more abundant spliced viral and cellular RNA, that are present within the cytoplasm, and packaged into the newly forming infectious virion. Interestingly, one report has shown that the p2 spacer peptide at the N-terminus of NC can contribute to the specific packaging of genomic RNA¹³¹.

The final step that takes place during assembly is budding. Mutations in the C-terminal region of Gag that block this step induce a phenotype in which the budding process appears to arrest just prior to viral detachment from the cell. Since this event takes place during what is considered to be a late stage of the viral life cycle, the domain responsible for this phenotype was termed the L, for late, domain, and was mapped to a P(T/S)AP motif in p6⁴⁶. Incidentally, this phenotype, which was first described in 1991¹⁰² had been a mystery for about 10 years; recently, it was shown that a cellular protein,

Tsg101, binds directly to the PTAP motif⁹⁴. The fact that this protein is known to be involved in the vacuolar protein sorting pathway suggests that HIV-1 might usurp this mechanism in the same manner that cellular membranes bud away from each other, in order to facilitate release of the virus from the plasma membrane during budding. This is indeed the case, and the last three years have seen an explosion in this area with the identification of a number of cellular factors required for assembly and budding of HIV and other viruses. Since this is an active area of research at this time, I will now give it special attention.

It is known that an inactive ubiquitin-binding domain at the N-terminus of Tsg101 binds directly to the PTAP motif in the HIV-1 p6 domain, and that in yeast, the homologue of this protein is a component of a 350kD complex required for the sorting of endosomal membrane cargo proteins into MVBs. This complex, known as the endosomal sorting complex required for transport, or ESCRT-I, is recruited to the membrane of endosomes, where it recruits two more large complexes, ESCRT-II and ESCRT-III. These complexes then function together to initiate MVB formation by invagination of membrane patches into the lumen, a process that topologically resembles virus budding. During the final stages of HIV-1 assembly and budding, the PTAP motif is thought to recruit Tsg101, which, in turn, recruits members of the ESCRT pathway, in order to perform the final pinching off that is required for the virus to leave the cell (Reviewed in Amara and Littman⁷). This model fits well with the L domain phenotype described over 10 years ago¹⁰², since mutation of the PTAP motif would eliminate the recruitment and binding of Tsg101, and subsequently, the ESCRT complexes. This would give rise to a phenotype in which the virus appears to be assembled normally but remains tethered to

the cell membrane. Interestingly, knock-down of Tsg101 by small interfering RNAs in cells producing HIV-1 duplicates the exact phenotype observed in L domain mutants, thus confirming the role of Tsg101 and the ESCRT pathway in HIV-1 assembly⁹⁴. Research devoted to this topic over the last three years has not only solved the longstanding mystery of the L domain phenotype, but, as a direct result of this work, the interaction network of the human ESCRT protein pathways are now better characterized^{185,258,270}. This represents another example of how HIV research has hastened the rate at which information in a different field has been attained.

During the assembly and budding process, the Gag and Gag-Pol polyprotein precursors are processed. The virus particle that begins to bud away from the cell is considered to represent an “immature” particle, which is characterized morphologically by the appearance of an electron-lucent center, and one or two concentric electron dense rings, based on electron microscopy (EM). The Gag and Gag-Pol polyproteins are unprocessed, and the two copies of full-length viral genomic RNA are noncovalently linked in the form of a loose unstable dimer. The maturation of the virus particle to the “mature” form is regulated by the activation of PR, and although the exact timing of this event remains unclear, it is generally believed that proteolysis takes place after assembly and perhaps during the budding process. Autoproteolysis of PR liberates it from the Gag-Pol precursor and allows the sequential cleavage of both Gag and Gag-Pol polyproteins, which results in the reorganization of the particle structure into the mature form. EM analysis of mature virus particles displays a condensed conical core of electron-dense material made up of cleaved CA protein. During the virus maturation process, the loose unstable RNA dimer is converted to a stable extended duplex that is typical of all

retroviruses. Once the cleavage of the virion-associated polyproteins and the conversion of the loose RNA dimer to a stable one have taken place, the new infectious virus is ready to initiate another round of the viral life cycle (Reviewed in Coffin *et al.*⁵³).

1.3.6 Vif: A Mystery is Solved

As mentioned earlier, the major function of most HIV-1 proteins was determined not long after the virus itself was identified. However, the true role of the Vif protein remained a mystery for about fifteen years. Recently, however, a paper describing the function of Vif in association with a cellular factor in HIV-1 replication was published²⁴⁵. Not since the discovery of the coreceptors has there been so much excitement in the field of HIV research, and because this work has had major impact, it also deserves special attention.

Vif was discovered in 1987^{87,259}, and it was quickly realized that HIV-1 mutant viruses lacking Vif (Δ Vif) displayed two infectivity phenotypes that depended on the cell line from which the virus was produced, hence its name, 'virion infectivity factor'. Basically, Δ Vif viruses generated by 'non-permissive' producer cells are unable to replicate when subsequently incubated with target cells. On the other hand, 'permissive' cell lines produce Δ Vif viruses that subsequently replicate well in target cells. This fascinating paradigm led to the hypothesis that a factor must be present in some cell lines but not others that could potentially inhibit HIV-1 replication; and, that the antiviral activity of this factor could be overcome by the Vif protein (Reviewed in Pomerantz²²¹). This factor was assumed to have a negative effect on HIV-1 replication because elegant

heterokaryon studies showed that the non-permissive phenotype was dominant over the permissive^{171,249}. Based on this assumption, a PCR-based complementary DNA subtraction strategy was used to study two genetically related cells lines, one non-permissive and one permissive, and identified the factor, which was called CEM15, after the cell line in which it was discovered²⁴⁵. Protein sequence analysis of CEM15 revealed it to possess significant amino acid similarity with apobec-1, the cytidine deaminase that is the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme. The significance of this homology led to the proposal that CEM15 might represent a form of innate antiviral immunity, against which HIV evolved to produce Vif²⁴⁵. This work also points to Vif as a novel target for anti-HIV drug development.

These findings have been confirmed by many reports, and CEM15 is now referred to as APOBEC3G (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G). In the absence of Vif, APOBEC3G is packaged into the assembling virion, and, at the reverse transcription stage of the next round of infection, APOBEC3G extensively edits dC nucleotides to dU on the DNA minus strand, thereby disrupting proper provirus formation and inducing G-A hypermutations in the subsequent progeny positive strand viral RNA^{110,175,181,285}. However, Vif blocks this innate antiviral effect by specifically binding to APOBEC3G during virus assembly to prevent its incorporation into the virus particle, and induces ubiquitination of APOBEC3G, thereby targeting it for degradation in the proteasome^{56,127,181,182,193,246,256,282}. More recently, it was shown that a single amino acid in the host APOBEC protein determines whether or not Vif proteins from various HIV and SIV strains can block the antiviral activity of APOBEC^{33,176,242,279}. This finding may help to explain why certain strains of SIV have been able to make the species jump

from primates to humans while others have not. A direct application of this finding is that it might eventually lead to the design of HIV mutants with broader host ranges; these would be valuable in the study of HIV pathogenesis and vaccine development.

1.4 Current and Future Therapies

As stated earlier, the enormous body of knowledge on the topic of HIV over the last two decades is impressive, and the work of the last two or three years with respect to the relationship between Gag and Tsg101, and the Vif/APOBEC story, has been exciting. In my opinion, both of these stories represent excellent examples of how answers to longstanding basic questions can lead to the unforeseen discovery of novel strategies that may ultimately broaden the arsenal of drugs available to fight this virus. Currently available anti-HIV therapies include numerous combinations of drugs that target RT and PR, and more effective drugs of these classes are continually being developed²²². Compounds designed to inhibit viral adsorption and entry are currently in development, and fusion inhibitors, such as T-20, have now been approved for use²⁶⁴, although their usage is presently limited by high costs. IN and NC inhibitors have been in development for some time, without clinical success to date⁶⁶. Future targets may involve the disruption of interactions between viral and host components. In this regard, many questions still remain regarding the HIV-1 life cycle and interactions with host factors. For example, the first event after viral entry into cells is uncoating. While this process is poorly understood, one can imagine that its timing must be tightly regulated by the virus, so that the RNA is protected from degradation by the cell until the proviral DNA is

formed. Once this process is better elucidated, drugs might be designed to either block the uncoating process or even cause it to happen earlier than it should. Both of these approaches would have the benefit of arresting the virus life cycle before the provirus is established, thereby blocking virus replication before drug resistance mutations can occur and decreasing the size of the latent reservoir of proviral DNA.

Multiple host proteins are believed to be required for the HIV replication process²¹², however, the exact role played by some of these proteins is not fully understood. For example, the integrase interactor 1 (INI1) is believed to be recruited by IN in order to take advantage of the chromatin remodeling machinery. Further characterization of the interaction between IN and INI1 might also uncover a region on the IN protein that could be a target for rationally designed inhibitors that would block the recruitment of chromatin remodeling components, thereby hindering the integration reaction¹⁰⁵. Assembly inhibitors are probably next on the horizon, and a few potential compounds have already been reported in the literature^{158,263,289}; these will be discussed in detail below. Small-molecule inhibitors aimed at disrupting the Gag/Tsg101 and the Vif/APOBEC interactions are already in development.

During assembly, the virus needs to encapsidate two copies of full-length viral genomic RNA. These two strands of genomic RNA have to be specifically selected from a pool of RNA that contains twice as much spliced viral RNA, and a hundred times as much cellular RNA, and they have to be dimerized at some point in time⁵³. Without proper encapsidation of this diploid dimeric genome, these assembled viruses would not be infectious. It is unlikely that viral RNA is passively recruited into the particle during the assembly process, since RNA can act as a template or scaffold for Gag assembly²⁰⁶.

Also viral RNA contributes to the stability of the particle itself, as virions containing less than adequate compliments of RNA genome are unstable and disassociate rapidly after leaving the cell^{272,273}. Thus, the role played by viral RNA, and the interactions that take place between viral RNA and other viral and/or cellular proteins during the assembly process needs to be better understood. The remainder of this chapter will provide a concise review of the processes of retroviral RNA dimerization and packaging, and how RNA sequences within the HIV-1 5' region are involved in these processes.

1.5 HIV-1 RNA Structure and Function

The HIV-1 genome has been referred to as the most extensively studied 9,700 bases of genetic sequence on the planet²²¹. By the same analogy, the 335 nucleotides upstream of the Gag start codon comprise one of the most frequently studied RNA segments. This region comprises the 5' UTR, and, as in all retroviruses, contains a number of sequence and structural elements that are crucial to various steps of the viral life cycle⁵³, though controversy exists in regard to the structure-function relationships that exist within the HIV-1 5' UTR. While many models have been proposed to represent overall RNA secondary structure within the region^{1,24,26,52,82,235,284}, there is disagreement as to which RNA sequences are absolutely required for the various functions attributed to it. We will now review what is known about some of these viral RNA elements and their suggested roles in the HIV-1 life cycle.

1.5.1 Structural Elements in the HIV-1 5' UTR

The first comprehensive model depicting all of the secondary structural elements present within the HIV-1 5' region was based on secondary structure analysis of conserved sequences with known function (Fig. 1.4) and shows the various RNA elements as a series of independent stem-loop structures²⁴. Although the actual RNA elements present in the virus particle and the cell undoubtedly exist as tertiary, and probably quaternary complexes, mutational analyses of most of these sequences suggest that these structures do exist at some point during the viral life cycle as independent stem-loops. The first is the TAR element, which is bound by Tat in order to transactivate transcription of the provirus, but has also been shown to affect reverse transcription and genome packaging. The poly(A) hairpin is also known to play a role in genome packaging, and is thought to repress the proximal polyadenylation signal. The U5-PBS complex contains the primer binding site (PBS) that is bound by tRNA^{Lys3} to prime reverse transcription. Downstream of the U5-PBS are a number of stem-loop structures (SL1-4) that bind the NC protein *in vitro*, and SL1, SL3, and SL4 are also known to be critical for packaging specificity *in vivo*. SL1 is known as the dimerization initiation site (DIS) since it is believed to facilitate the first contact between two strands of viral genomic RNA, thereby initiating dimer formation²⁴.

The main focus of the work in this PhD thesis has been the identification and characterization of novel viral RNA and protein sequences involved in the HIV-1 RNA dimerization and packaging processes. Therefore, I will now provide a brief review of what is known about HIV-1 RNA encapsidation, as genome encapsidation is a process that must be performed by all viruses. I will then provide a detailed compilation of the

Figure 1.4. HIV-1 5' RNA Structural Elements. Illustration of a working model of the HIV-1 5' UTR showing the various stem-loop structures important for virus replication. These are the TAR element, the poly(A) hairpin, the U5-PBS complex, and stem-loops 1-4 containing the DIS, the major splice donor, the major packaging signal, and the *gag* start codon, respectively. Nucleotides and numbering correspond to the HIV-1 HXB2 sequence. (Adapted from Clever *et al.*⁴⁸ and Berkhout and van Wamel²⁶)

discovery of retroviral RNA dimerization, which is essential for all retroviruses, followed by a review of what is known about HIV-1 RNA dimerization and a summary of the work that led to the studies undertaken in this PhD thesis.

1.5.2 HIV-1 RNA Packaging

Viruses are generally classified on the basis of the type of genome that they contain, i.e. RNA versus DNA, and the structure of that genome, be it linear, circular, single-stranded, double-stranded, etc. This classification puts retroviruses into a unique category, as the *Retroviridae* family of viruses all contain a single-stranded, but diploid, dimeric RNA genome. Why a class of viruses would evolve to have such a unique genomic structure is not really clear, but it is speculated that the availability of two copies of the genome would be advantageous during the complex reverse transcription process, that is key to the retroviral life cycle¹²⁰. Indeed, the dimeric nature of the genome is thought to be responsible for a high rate of recombination during infection. Regardless, the virus is faced with the challenge of specifically packaging two copies of full-length viral genomic RNA into the newly forming virion during the retroviral assembly process. Despite the fact that viral RNA represents <1% of the total RNA in an infected cell, >75% of the nucleic acid found in the virion is full-length viral genomic RNA⁵³, which indicates how specific this packaging process is. Even though there is twice as much spliced viral RNA in the cytoplasm as there is genomic, the viral RNA contained within wild-type virions is made up of 90-95% unspliced genomic. This highly specific packaging of viral genomic RNA strongly indicates that sequences absent from the

spliced viral RNAs are necessary for packaging. This, combined with the fact that Gag alone is able to form VLPs containing viral RNA, led to the hypothesis that the left half of the *gag* gene and its adjacent 5' UTR are responsible for the specific packaging of viral genomic RNA⁵³.

The first studies aimed at identifying the HIV-1 RNA packaging signal found that deletion of RNA sequences between the major splice donor (SD) and the *gag* coding region (i.e. SL3 and adjacent sequences) decreased the levels of genomic RNA packaged into virions^{5,47,157}. Since these sequences were downstream of the major 5' SD, and therefore would not be found in any spliced viral RNA species, it was plausible that this region could be responsible for the selective packaging of genomic RNAs. Analysis of the putative ψ locus from a variety of retroviruses showed that these sequences had the ability to direct selective encapsidation of heterologous RNAs to which they had been linked artificially^{4,11,17,79,140,179,261,274}. In HIV-1, such autonomous packaging signals were mapped to the regions extending 30-40 nt immediately upstream and downstream of the *gag* start codon¹¹³. However, subsequent studies showed that RNA sequences upstream of the 5' SD site also affected RNA packaging¹³⁴. It was also known that retroviral encapsidation required trans-acting amino acid sequences in the Gag protein^{5,74,100,197,209,253}, and several groups reported that HIV-1 Gag and NC exhibit specific binding affinity for the HIV-1 ψ site *in vitro*^{28,59,60,168,235}. These findings, combined with chemical and RNase accessibility mapping and computerized sequence analysis, led to the generation of a model for the HIV-1 ψ site that comprised four independent stem-loops⁴⁸ (SL1-4 in Fig. 1.4). These stem-loop structures were each shown to serve as specific, high-affinity binding sites for Gag *in vitro*, and were proposed

to contribute individually to overall packaging efficiency. Among these, SL1, SL3, and SL4 were later shown to be critical for packaging specificity *in vivo*^{187,188}. However, more recent work indicates that SL2 and SL3 display much higher affinities for NC than SL1 and SL4 *in vitro*^{8,67}. Based on these findings, the same group later proposed a model for the initial complex formed between the NC domains of assembling Gag molecules and the dimeric ψ region⁹. In this model, SL1 is shown to form an RNA duplex between the two strands, while SL4, instead of directly binding Gag, contributes additional RNA-RNA interactions that stabilize the tertiary structure of the ψ element. The conformation resulting from this folding pattern is thought to expose SL2 and SL3 for high-affinity binding to Gag.

Despite the clear results obtained from simplified *in vitro* studies such as those mentioned above, the SL1-4 region alone is not sufficient to target RNA into HIV-1 virions *in vivo*³⁰, and the minimal region required to confer autonomous packaging activity actually maps to a larger region covering the first 350-400 nt of the genome, including \approx 240 nt upstream of SL1^{114,130,189,219}. In agreement with these studies, mutations that alter the stability of the poly(A) hairpin stem region, or delete the upper part of the hairpin, severely inhibited HIV-1 replication⁶². And, these defects in replication were shown to correlate with reduced RNA packaging levels in virions, suggesting that the formation of the poly(A) hairpin is necessary for normal packaging of viral genomes. Subsequent research confirmed the importance of the poly(A) hairpin in the RNA packaging process⁵¹, and it was shown that similar disruption of base-pairing in the stem of the TAR element also caused profound defects in packaging^{51,114}. Finally, deletion analyses of RNA sequences between the poly(A) hairpin and SL1 suggest that

unspecified sequences within the U5-PBS region also contribute to HIV-1 RNA packaging^{51,189}.

In summary, all of the seven predicted stem-loop structures in the HIV-1 5' UTR have been shown to be important for genome encapsidation, and all of these RNA structural elements have also been assigned other functions in various steps of the viral life cycle, e.g. the role of SL1 in initiation of dimerization. The existence of such overlapping functions for these RNA structures raises the possibility that some of these functions, such as dimerization and packaging, might be linked. This question of a link between RNA dimerization and packaging will be discussed in greater detail below, and the concept of retroviral RNA dimerization and the RNA sequences involved in this process will be reviewed.

1.5.3 The Discovery of Retroviral RNA Dimerization

All modern virology textbooks state that a unique characteristic of the *Retroviridae* family of viruses is their possession of a dimeric RNA genome. It is interesting to review the key experiments, and their interpretation, that led to discovery of the retroviral RNA dimerization process.

One landmark study in 1967 demonstrated that viral RNA from RSV, AMV, MLV, and MTV, when subjected to zone sedimentation in sucrose gradients, displayed sedimentation constants between 64 and 74S²²⁸. It was also reported that the RNA had two components, a large $\approx 70S$ component, which was assumed to be the intact genome, and a smaller one of $\approx 4S$, which was thought to represent degraded viral RNA. It was

also suggested that the genomes of all four viruses were single-stranded, since only 2% of the RNA could be recovered after RNase digestion in single-strand-specific conditions. However, since these reported sedimentation constants and corresponding molecular weights were much larger than those of most other known viral RNAs, the structure of these RNA genomes became a matter of great interest. It now seems obvious that this seemingly large 70S component was actually a dimer of two copies of the smaller 35S genomic RNA, which would be comparable in size to other viral RNAs, and in 1968, it was shown that the fast-sedimenting (62S) RSV RNA was actually an aggregate of smaller (36S) RNAs⁷⁸. This was proven by experiments showing that the 62S RSV RNA species could be converted to a 36S species by heat treatment, suggesting a disaggregation of the 62S RNA into smaller RNAs. Further analysis showed that this conversion took place over a slow transition that depended on the incubation temperature and the duration of incubation. Based on these results, it was proposed that weak bonds, such as base-pairing, held the RSV RNA aggregate together.

With the clear demonstration that the 60-70S RNA aggregates were actually made up of 30-40S subunits, many questions arose about the nature of the linkage between the subunits, as well as the timing and function of such an RNA association. In one study aimed at addressing such questions, it was observed that RSV harvested from cells at intervals of 5 min contained heterogeneous RNA smaller than the 60-70S RNA of virus harvested at hourly intervals. This raised the possibility that the 30-40S subunits represented a precursor of the 60-70S RNA, and this was soon shown to be true, since RSV harvested from infected cells at intervals of 3 min contained 3-6 times more 30-40S RNA than 60-70S⁴⁰. Furthermore, in virus harvested after 3 min, but then incubated for

45 min at 40°C, most of the 30-40S RNA in the virion becomes converted to 60-70S RNA. More importantly, once the 30-40S RNA had been extracted from the 3 min virus, it could not be converted to the 60-70S form, which implied that some component of the virus was required for proper RNA conversion. Other research showed that the intersubunit linkages found in immature (3 min virus) B77 sarcoma virus RNA were unstable and rapidly dissociated upon extraction²⁵⁵. However, this linkage could be stabilized by extracting the RNA in the presence of high concentrations (0.5 M) of NaCl, as seen by an increase in the melting temperature (T_m) of the 60-70S RNA. The authors suggested that this increase in T_m results from an increase in the stability of base-paired regions and insightfully proposed possible mechanisms, one of which was that a rearrangement of the subunits in the virion might increase the length of base-paired regions. Years later, this model was shown to represent the mechanism by which the loose unstable RNA dimer present in the immature virion is rearranged into a more stable extended duplex by the activity of the NC protein.

Another major subject in the 1970s was the composition of these RNA aggregates. The first real understanding of this putative aggregate RNA structure came in 1975 in a paper in which RNA from the endogenous feline retrovirus, RD-114, was visualized by EM. This paper showed that the 52S RNA molecule existed as an extended single strand that contained a central Y- or T-shaped secondary structure¹⁴². It appeared that this 52S molecule actually consisted of two half-size molecules, joined together by the Y- or T-shaped structure, which was referred to as rabbit ears (RE). Melting of the RE using strongly denaturing conditions caused a dissociation of the 52S molecule into smaller subunits. In support of the studies mentioned above, the authors presumed a base-paired

structure between two identical subunits might give rise to a dimer, which was joined at the RE. In order to determine which regions of the RNA genome actually formed this RE, or dimer linkage structure, as it was renamed, a convenient EM label was developed that allowed the mapping of poly(A) sequences on large RNA molecules¹⁸. Another group had already constructed a label for poly(A) mapping by linearizing SV40 DNA circles with endonuclease, and then polymerizing short poly(dT) stretches on the 3' ends of the DNA using terminal deoxynucleotidyl transferase⁴¹. Unfortunately, when such a molecule was hybridized to a single-stranded RNA, it was difficult to distinguish between the RNA and the label. However, this system was improved in studies showing that terminal transferase could also extend poly(dT) tails from nicks in the circular SV40 DNA. This led to the creation of circular SV40 DNA molecules that contained an average of 2 poly(dT) tails per circle, with each tail being about 175 bases long. With such a molecule in hand, it was possible to localize the poly(A) on an RNA molecule by EM, based on the position of the SV40 DNA circle. Using this system, the 52S RNA from RD-114 virions was shown to form a circular structure with 2 free ends attached to a single SV40-poly(dT) circle at different dT tails, with the dimer linkage structure in the center of the RNA. This meant that the 52S RD-114 RNA had a poly(A) sequence at each of the two free ends, but more importantly, that nucleotides involved in the dimer linkage structure resided in the 5' region of the RNA genome. The same group also reported that two other type C RNA viruses, BKD, an endogenous baboon virus, and WoMV, a simian sarcoma virus isolated from a woolly monkey, also contained a 52S RNA dimeric genome, comprised of two monomer subunits joined at their non-poly(A) ends¹⁴⁴. Based

on these findings, the authors proposed that this dimer linkage structure might be a common feature of all type C virus RNAs.

The existence of a dimer linkage structure (DLS) was subsequently demonstrated in a number of type C viruses tested, including Friend virus⁷⁷ (FV), Moloney Sarcoma Virus¹⁷² (MSV), an endogenous murine ecotropic virus¹⁹ (AKR), an endogenous xenotropic virus¹⁹ (NZB), avian reticuloendotheliosis virus^{19,98} (REV), and Moloney murine leukemia virus²⁰⁷ (M-MuLV). In 1986, the hypothesis that conversion of the 35S RSV RNA to the more stable 70S form involved a rearrangement of the subunits within the virion, was shown to be true. A modified Northern Blot allowed the visualization of native RNA, which was then applied to the structural analysis of RNA viral genomes¹³³. Using this system, both the dimeric (70S) and monomeric (35S) forms of RSV RNA could be detected on a gel, and, it was shown that NC protein played a role as a cofactor in RNA dimer formation¹⁹⁴. Further analysis showed that RSV particles containing mutations in the Cys-His boxes of NC displayed abnormal ratios in regard to 70S and 35S RNAs, indicating that NC was indeed involved in the structural maturation of the 70S dimer¹⁹⁵. More importantly, these defects in dimer maturation correlated with decreased infectivity observed in these mutant viruses, suggesting that the formation and maturation of the 70S stable dimer was required for normal retroviral replication. By this time, HIV had been discovered and shown to be a retrovirus, and it was assumed that the dimerization process must be critical to the retroviral life cycle. Hence, inhibition of RNA dimerization was proposed as a possible therapy for HIV, and HIV-1 RNA dimerization became an intensely studied topic.

1.5.4 Identification of the HIV-1 DIS

Both *in vivo* and *in vitro* approaches have been used to study retroviral RNA dimerization. The *in vivo* approach is that whereby RNA is isolated from virions produced in tissue culture and then analyzed by native Northern Blotting¹³³. The other method involves synthesis of short segments of viral RNA *in vitro*, and then studying the ability of these segments to form dimers in an Eppendorf tube. Valuable information can be obtained by both methods. Unfortunately, both also have advantages and disadvantages, and it is often difficult to correlate results obtained with the different systems. The HIV-1 DLS was originally identified by *in vitro* analysis showing that an *in vitro* transcribed segment of HIV-1 RNA, representing nucleotides (nt) 28-1333, could form two major bands when incubated at 37°C for 15 min⁶⁰. The lower band had the expected size of ≈1300 nt, while the upper band corresponded to a dimer. In agreement with the *in vivo* RSV studies mentioned above, this *in vitro* dimer also disappeared upon heat denaturation at similar temperatures. These data provided strong evidence that the HIV-1 RNA genome indeed existed as a dimer linked at the 5' end. However, this group also attempted to map the minimal nucleotide sequence required for this dimerization activity, and found that RNA sequences from positions 316-346 and 311-415 were required and sufficient for dimerization of HIV-1 RNA. However, these segments were later shown not to be the most important dimerization sequences¹⁸⁴, which highlights the major disadvantage of the *in vitro* approach, since the outcome of the reactions are very sensitive to the conditions of the assay. This study also showed that the NC protein, when included in these dimerization assays, could bind to the HIV-1 RNA and increase the rate

of dimerization of the RNA fragments. Based on these results, a model was proposed in which NC binds to a specific site in the dimer domain and induces a conformational change in the RNA that facilitates the dimerization process⁶⁰.

In a similar report, the *in vitro* RNA dimerization process was shown to depend on the concentration of RNA, mono- and multivalent cations, temperature, and pH¹⁸³. Furthermore, high Mg²⁺ concentrations could yield HIV-1 RNA dimerization *in vitro* in the absence of NC, making subsequent *in vitro* studies easier to perform, since NC is difficult to purify. More importantly, it was shown that antisense HIV-1 RNA did not form dimers, from which it was inferred that sense RNA must adopt a specific structure required for proper dimerization. This group then reevaluated the dimerization activity of a number of HIV-1 RNA fragments, but this time with 0.1 mM MgCl₂ contained in the gel and the running buffer. Under these conditions, they found that an RNA fragment representing nt 1-311 of the HIV-1 RNA (Mal strain; subtype A) could not only form dimers, but that RNAs containing these first 311 nt could dimerize 10 times faster than the 311-415 RNA that had been previously shown to be sufficient for HIV-1 RNA dimerization¹⁸⁴. Based on these results, the authors concluded that sequences upstream of the splice donor site are involved in the dimerization process, and proposed that sequences in this region somehow hastened the reaction.

This surprising finding raised an important question about the link between dimerization and packaging of genomic RNA. The notion that sequences required for dimerization would reside downstream of the major splice donor fit nicely with the proposed model that only full length genomic RNA contained the RNA dimerization sequences, and that dimerization was a prerequisite for packaging. In this model, only

full-length genomic RNA could dimerize and then be subsequently recognized and packaged into the virion. The abundant spliced RNAs present in the cytoplasm would not contain the dimerization sequences and hence would not be packaged. However, these new findings implied that all viral RNAs, including the spliced species, contain sequences that are able to dimerize. The fact that spliced viral RNAs were not packaged at high levels implied that relevant interactions must also take place downstream of the splice donor.

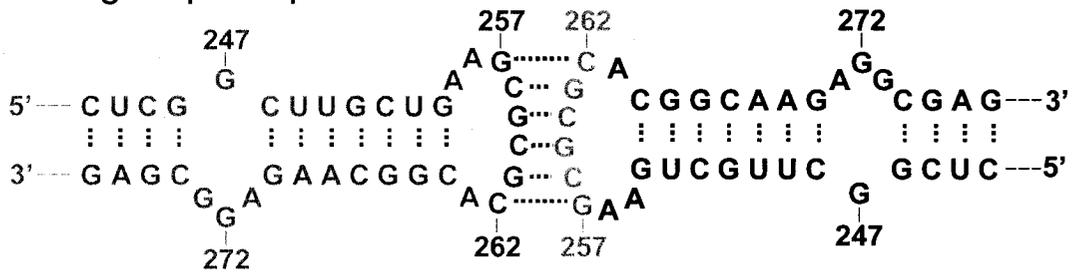
In a follow-up study, the same group set out to localize, at the nucleotide level, the sequences involved in HIV-1 RNA dimerization²⁵¹. Using chemical modification of RNA fragments representing the HIV-1 5' region, they showed that nucleotides whose modification was able to inhibit dimerization were centered around a palindromic sequence, 274-GUGCAC-279, between the PBS and the major splice donor. The authors noted that the RNA sequences on both sides of this palindrome could form a stem-loop structure with this palindrome in the hairpin loop. Deletion of this stem-loop motif (nt 265-287) completely abolished dimerization of the 1-615 HIV-1 RNA fragment *in vitro*. Based on NMR analysis of RNA fragments, a similar structure had been proposed for HIV-IIIB¹¹¹. The dimerization model, similar to the one shown in Fig. 1.5, whereby two monomers would recognize each other through a loop-loop interaction initially involving just the two palindromic sequences was proposed; this interaction would then induce subsequent annealing of the stems. The palindromic region was then termed the dimerization initiation site (DIS) and it was proposed that this structural element could be exploited for targeted antiviral therapy by antisense oligonucleotides²⁵¹.

Shortly thereafter, a publication appeared demonstrating that a 19 nt sequence upstream of the 5' major SD was part of the HIV-1 RNA dimerization domain¹⁴⁷. In an attempt to map the 5' and 3' boundaries of the dimerization domain, it was shown that RNA fragments lacking DIS stem nt 233-251 (Lai strain; subtype B) were fully monomeric. The results in this article complemented those described above, and it was proposed that the 248-270 nt region might be a core dimerization domain. The importance of this autocomplementary sequence (ACS) was confirmed by others who observed that *in vitro* dimerization of a 224-402 nt RNA fragment was completely blocked by an antisense oligonucleotide that targeted the ACS²⁰³. This led to a "loop-loop kissing complex"²¹⁵ or "kissing-loop model"¹⁴⁷ of HIV-1 RNA dimerization, in which the 6 nt palindromes on two monomeric RNA molecules interact through Watson-Crick base-pairing. This initial interaction was said to shift the equilibrium toward the formation of the dimers, allowing the stems to melt and anneal to their complementary sequences on the other RNA molecule, thus forming the stable extended duplex (Fig. 1.5). This model fits with the idea that immature virions contain a less stable dimer involving only base-pairing of the palindromes, but that the mature virions contain a more stable structure, the extended duplex. Subsequent phylogenetic analysis of over 50 HIV-1, HIV-2, and SIV nucleotide sequences showed an absolute conservation of a predictable structure similar to the DIS, with the hallmark of the HIV-1 DIS motif being a 6 nt palindrome consisting of either a GCGCGC or a GUGCAC sequence^{25,151}. The kissing-loop model has also been proposed for a number of other retroviruses including ASLV^{88,220}, MLV^{68,96,97,268}, SIV¹²⁶, and HIV-2^{71,126}, although, in the latter case, a few

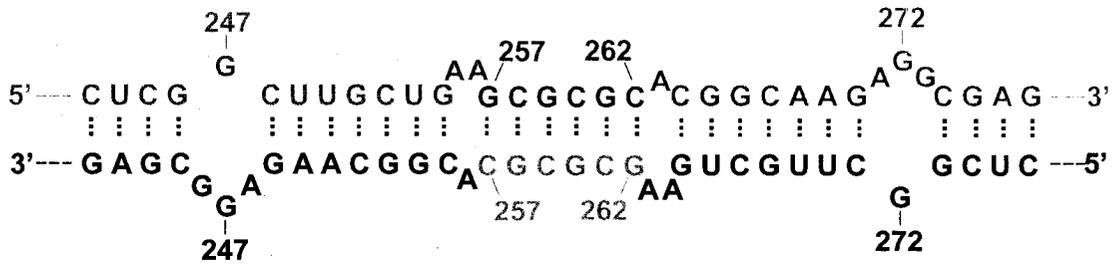
Figure 1.5. The Kissing-Loop Model of HIV-1 RNA Dimerization. HIV-1 RNA dimerization is initiated by a Watson-Crick base-pairing interaction between two palindromes in the loops of SL1 on two monomeric genomic RNAs. This interaction forms the loose unstable kissing-loop complex. Coincident with virus particle maturation, this unstable dimer is rearranged to form a more stable extended duplex that involves a mechanism whereby the base-pairs in the stems melt and then re-anneal to their complementary sequences on the opposite strand. Nucleotides and numbering correspond to the HIV-1 HXB2 sequence. (Adapted from Skripkin *et al.*²⁵¹ and Laughrea and Jetté¹⁴⁷)

Figure 1.5. The Kissing-Loop Model of HIV-1 RNA Dimerization.

Kissing-loop complex



Extended duplex



different RNA sequences have been proposed to act as the DIS, including the TAR element²², the PBS¹²⁶, and SL1²⁵. One report even showed that HIV-2 RNA could use either SL1 or the PBS to initiate dimerization of RNA fragments *in vitro*, depending on the different experimental conditions employed¹⁴⁵. Based on this finding, the authors concluded that HIV-2 is able to utilize alternate dimerization initiation sequences. However, this might be another example of how the experimental conditions used in *in vitro* dimerization experiments can affect outcome and raises questions about the relevance of approaches that rely exclusively on cell-free assays.

1.5.5 Viral Proteins Involved in Dimerization

Despite ample *in vitro* evidence supporting the above model of dimer maturation, it was not known where or when the dimer was actually formed *in vivo*. However, a report in 1993 used a native Northern Blotting¹³³ approach to analyze the conformation of dimeric RNA present in Moloney murine leukemia virus particles⁹⁰. They found that RNA extracted from MoMuLV virions every 5 min and electrophoretically analyzed, contained a single species of RNA that migrated more slowly than RNA isolated from 24 hr virus preparations. The authors interpreted this difference in mobility to indicate that dimeric RNA must indeed undergo some change after the virus is released from the cell. They also showed that the dimeric structure of RNA from rapid-harvest virus was less stable than in mature virus, as it dissociated into monomers at a slightly lower temperature, 51.3°C compared to 53.8°C. These authors then turned their attention to the question of the mechanism of this RNA maturation event, whose timing coincided with

the well-known cleavage of the Gag polyprotein, which is catalyzed by the viral PR enzyme and also takes place after virus release. Thus, they investigated the conformation of dimeric RNA in the context of a PR-deficient (PR-) mutant and found that RNA from two PR- mutants migrated more slowly, and displayed lower melting temperatures, than wild-type RNA. Based on these results, it was concluded that PR function is required for RNA maturation in MuLV. In the same paper, similar experiments with a related virus, KiSV, also suggested that the RNA maturation event required an intact, unsubstituted Cys array. On the basis of these results, they proposed a maturation pathway for MuLV in which Gag polyprotein molecules assemble into a nascent virion containing an immature dimer. The particle would then be released from the cell, and once Gag is cleaved by PR, NC would act on the immature dimer, converting it to the mature form.

With a clear link established between PR activity and dimer maturation in MuLV, this group then tested whether the same link existed in HIV. Previous work had shown that RNA from PR- HIV-1 mutant viruses also migrated more slowly, and displayed lower thermal stability, than wild-type RNA, indicating that RNA dimers in HIV-1 also undergo a PR-dependent maturation event⁹¹. The same PR-dependent model of dimer maturation could then be applied to the situation of HIV-1, but it remained to be determined which Gag cleavage product was actually involved in this mechanism.

The retroviral NC protein, a small, highly basic protein, is known to possess a nucleic-acid-chaperone activity that has been demonstrated in a number of critical events in the retroviral life cycle, including reverse transcription and genome packaging. Since NC proteins were known to associate with viral RNA within the mature virion, and because it was known to facilitate the formation of the most thermodynamically stable

structures in nucleic acids, NC became the obvious choice for the Gag cleavage product responsible for RNA dimer maturation (Reviewed in Rein *et al.*²²⁶). This hypothesis was directly confirmed by *in vitro* analysis of the effects of HIV-1 NC on Harvey sarcoma virus-derived dimeric RNAs⁸⁶. Fragments representing nt 34-378 of HaSV RNA were allowed to form dimers at 37°C and were then treated with NC. When the thermostabilities of these dimers were compared to those of untreated dimers, the NC-treated dimers were more stable, suggesting that NC could convert the less thermostable dimers to a more stable conformation. The authors then attempted to elucidate the mechanism responsible for this activity of NC by testing the ability of various NC mutants to induce this conversion. They found that if eight basic amino acids flanking the N-terminal zinc finger were changed to Ala, or if an Ala was substituted for a Phe at residue 16 in the first zinc finger, the mutant NC proteins were unable to stabilize the dimers. However, mutation of all six Cys residues within the two zinc fingers had no effect on the ability of NC to carry out this function. The authors proposed that NC-mediated RNA maturation involved electrostatic interactions between basic amino acids and phosphate groups on the RNA, as well as a stacking interaction between a zinc finger aromatic residue and bases of the RNA. These putative interactions were hypothesized to transiently break intramolecular bonds in the less stable dimer, thereby facilitating the conversion to the more stable form containing more intermolecular base-pairs. Similar results were obtained by others showing that HIV-1 NC could activate dimerization of a 77-402 nt fragment of HIV-1 Lai RNA²⁰⁴. This group also demonstrated the conversion of an unstable dimer, corresponding to the kissing complex, to a stable one, after incubation with NC. Taken together, the thermostability conversions described in these

two papers seem to resemble the RNA maturation reported *in vivo*, and strongly suggest that NC is responsible for the dimer maturation depicted in Fig. 1.5. Subsequent *in vivo* analysis of a panel of HIV-1 viruses containing mutations in NC showed that Cys-Ser substitution of amino acid residues within the second zinc finger decreased genomic RNA dimerization to the same extent as disruption of the DIS¹⁵³. This finding confirmed the involvement of NC in the dimerization process and suggests that the kissing-loop model also applies to the *in vivo* situation.

1.5.6 Dimerization and Virus Replication

With the viral RNA sequences involved in RNA dimerization now well characterized *in vitro*, and the existence of an RNA maturation event conclusively demonstrated *in vivo*, the next important question to be addressed was whether the same RNA and protein components identified *in vitro* were also involved in the dimerization process in the context of replicating virus. This question has been addressed by a number of groups, using a broad array of HIV-1 strains, mutants, assays, and quantitation methods. Hence, the results obtained, and the interpretation thereof, were often different.

The first report to address this issue showed that viruses containing the same mutations that disrupt dimer formation *in vitro* also lacked a stable dimeric RNA genome *in vivo*¹⁰⁸. More specifically, this group created an HIV-1 mutant virus in which the first three nucleotides of the GCGCGC palindrome were changed to AAA, thereby disrupting the complementarity in the loop sequence that was proposed to initiate the dimerization process. When they purified virus from infected cultures of C8166 cells and analyzed the

viral RNA by non-denaturing Northern Blotting, they found that the mutant viruses contained less dimeric RNA than did the wild-type viruses. This indicated that the same sequences required for dimerization of HIV-1 RNA *in vitro* strongly influenced genome dimerization *in vivo*. One surprising result, however, was that RNA dimers extracted from the mutant viruses were as thermostable as those of wild-type HIV, which is contrary to what one would expect if the kissing-loop model also applied to the *in vivo* situation. Although the mutant viruses grew more slowly than wild-type, they were still infectious in tissue culture. The authors concluded that the sequences required for formation of the kissing-loop dimer *in vitro* are not absolutely essential for virus replication, but are important for the *in vivo* dimerization process.

A comprehensive phylogenetic analysis of the untranslated leader regions of the RNA genomes of several HIV and SIV virus isolates has been performed²⁵. Interestingly, a structure similar to the DIS hairpin (SL1) proposed for HIV-1 RNA could be generated in similar regions for most virus isolates. And, despite considerable sequence heterogeneity in both the stem and loop domains, a 6 nt palindrome within the loop was conserved among the different viruses, consistent with the idea that HIV-1 RNA dimerization involves Watson-Crick base-pairing. In an effort to study the role of the DIS palindrome in HIV-1 replication, two mutant viruses were constructed. The wild-type palindrome, containing the GCGCGC sequence, was truncated to GC or enlarged to GCGCGCGCGC in mutants GC1 and GC5, respectively. When these mutants were cultured along with the wild-type virus in SupT1 T-cells, delayed replication kinetics were observed for both. However, in agreement with the above article, neither DIS mutant was fully replication impaired. Non-denaturing Northern analysis of the viral RNA showed that only one RNA

species was present in the wild-type and mutant virions, suggesting that both mutants were able to form dimeric RNA genomes, and RNA from all three viruses displayed identical melting temperatures. Based on these results it was concluded that the DIS motif is not essential for formation of an RNA dimer. However, when the levels of viral RNA contained within the virus particles were measured, both mutants apparently packaged lower amounts compared to wild-type, and these packaging defects were consistent with replication defects. It was suggested that the replication defect seen in their DIS mutants argues for an important role of this sequence in virus replication, but that the presence of a fully dimeric RNA genome shows that the DIS palindrome is not absolutely required for virus replication or RNA dimerization. They also proposed that one possible explanation for the DIS being less critical in virus infection studies could be that additional RNA elements might be involved in RNA dimerization, e.g. the upstream poly(A) hairpin which also contains a palindrome, or sequences downstream of the splice donor. In light of the packaging defects observed with their mutants, they proposed that their DIS loop mutants experienced a partial dimerization defect that led to inefficient RNA packaging, but that the fraction of RNA that ended up in the particle was dimeric, suggesting an intrinsic link between dimerization and packaging. Incidentally, this putative link is still strongly debated in the current literature.

The two papers mentioned above focused on the role of the DIS palindrome in the *in vivo* RNA dimerization process. Another addressed the function of the complete DIS stem-loop structure in the virus replication process through construction of a number of DIS mutant viruses, two of which lacked either the upper stem-loop sequences (nt 248-270), or the whole stem-loop structure (nt 243-277)²¹⁴. Unlike the viruses containing

minor DIS mutations above, these large deletion mutants grew poorly when cultured in SupT1 cells, displaying 1000-fold lower infectiousness than wild-type. Analysis of the RNA content in virions isolated from COS-7 cell transfections failed to show any significant differences in the RNA packaging levels of these mutants. However, virions produced during T cell infections contained substantially less viral RNA than wild-type, indicating that the complete DIS stem-loop structure contributes to the HIV-1 genomic RNA packaging signal. To further investigate the nature of the defects caused by these DIS mutations, this group also tested their effects on reverse transcription by analyzing viral DNA synthesized during SupT1 cell infections. Although the first-strand transfer activity was not significantly altered, when they amplified DNA using primers that detected products generated after the second-strand transfer had occurred, they observed 10-fold and 3-fold decreases for the large and small deletion mutants, respectively. The authors concluded that the complete DIS structure is a *cis* element acting on both genomic RNA packaging and synthesis of proviral DNA. The fact that these larger DIS deletions had such a pronounced impact on virus replication was interpreted to suggest melting of the stem and formation of an extended intermolecular interaction, i.e. the stable extended dimer, can occur *in vivo*. With respect to reverse transcription, they proposed that the ability of the DIS to bring the viral nucleic acids in close proximity would favor strand transfer.

Based on the above-cited *in vivo* studies, it is likely that the same RNA components shown to be important for *in vitro* dimerization are also involved in the dimerization process in the context of a replicating virus¹⁰⁸, suggesting that the kissing-loop model also applies to *in vivo* RNA dimerization. All three studies reported defects in virus

replication when DIS sequences were mutated^{25,108,214}, and it appears that larger mutations, e.g. deletion of the complete DIS stem-loop²¹⁴, had greater impact on virus replication. Despite the critical role of the DIS palindrome in *in vitro* studies, mutations within the palindrome caused only minor defects in virus replication^{25,108}, indicating that the DIS stem-loop structure might be more biologically important to the virus than the DIS palindrome itself. We can also conclude from these studies that DIS sequences affect both viral RNA packaging^{25,214} and reverse transcription²¹⁴, therefore highlighting the importance of the dimerized genome to other aspects of the viral life cycle. Finally, one paper reported that mutations in the DIS had no major effect on levels of dimerized RNA contained in viruses²⁵, and both studies that tested the thermostability of RNA dimers extracted from mutant viruses, found that these dimers were as stable as those extracted from wild-type virus^{25,108}. Thus, RNA sequences outside the DIS must be involved in the formation and/or maintenance of the RNA dimer.

Our laboratory had already been studying the effects of 5' UTR mutations on various aspects of virus replication, and Michael Laughrea's lab in our institute was well established in the *in vitro* aspects of HIV-1 RNA dimerization. In a collaborative effort, similar *in vivo* studies were initiated and a paper was published describing two HIV-1 kissing-loop domain (KLD; i.e. DIS) mutants, an ACS-, which contained a mismatch at the top of the stem-loop, and had 3 mutations in the upper loop that disrupted the ACS, and a deletion mutant (Δ 248-261), missing the right side of the upper stem, and the ACS¹⁵¹. The infectivity of these mutant viruses, based on TCID₅₀, was decreased by at least 99%. To investigate the cause of this low infectivity, genomic RNA encapsidation and dimerization were also analyzed. It was observed that genomic RNA encapsidation

levels were decreased by 75 and 50% of wild-type for the ACS- and $\Delta 248-261$ mutants, respectively. On the basis of these findings, it was concluded that mutations in the kissing-loop hairpin (DIS) have moderate effects on genomic RNA encapsidation. When the effects of these mutations on RNA dimerization were studied by non-denaturing Northern analysis, the proportion of dimeric genomic RNA was reduced from >80% for the wild-type to 40-50% for the mutants. In agreement with some of the results from each of the studies mentioned above^{25,108,214}, it was concluded that the kissing-loop hairpin plays a role in genomic RNA dimerization and encapsidation, and has dramatic effects on viral infectivity. It was also concluded that encapsidation and dimerization are to some extent coupled, and that the kissing-loop hairpin could be profitably targeted by antisense constructs within an AIDS therapeutic program.

By this time, the question of a link between dimerization and packaging had received much interest, since the location of the ψ locus was shown to coincide with the most stable point of interstrand binding of genomic RNA in many retroviruses, including SL3 and SL1 in HIV-1. A working model of the HIV-1 ψ locus had been proposed to comprise stem-loops 1-4⁴⁸, to which bacterially expressed Gag or NC had been shown to bind specifically^{27,29,48,235}, with SL1, SL3, and SL4 each providing independent high-affinity binding sites⁴⁸. Given the potential for overlapping dimerization and encapsidation signals within the ψ locus, and the required presence of Gag and/or NC for these activities *in vivo*, HIV-1 mutants were designed in an attempt to create viruses that would display selective defects in dimerization or encapsidation⁵⁰. By analyzing a panel of mutants containing substitutions or deletions in SL1 and SL3, it was found that deletion of either SL1 alone, or SL3 plus adjacent flanking sequences, reduced genome packaging

to 19 or 12%, respectively. Deleting both elements together caused a further reduction to 5%. One SL3 mutant, in which the left side of the stem was mutated to prevent base-pairing, reduced packaging to 42% of wild-type, but packaging efficiency was fully restored when the right side of the stem was also mutated to restore the base-pairing. Since mutations in the loop region of SL3 also had little effect on packaging specificity, it was concluded that the structure, and not the actual sequences, of SL3 contributes to genome packaging. Disruption of the base-pairing in the upper stem of SL1 also markedly inhibited packaging, but unlike SL3, compensatory mutations predicted to restore the base-pairing did not significantly restore packaging efficiency. Substitution and deletion of the 3-base bulge in SL1 also severely impaired genomic RNA content, indicating that stem and bulge sequences in SL1 were required for normal packaging efficiency. Incidentally, it was later shown that “stem-loop B”, which is comprised of the lower stem of SL1 and this “bulge”, plays a crucial role in genome dimerization and reverse transcription¹⁵². It was also shown that mutation of the palindromic sequence known to be important for initiating dimerization had only modest effects on encapsidation, implying that the loop palindrome was not critical for genomic RNA packaging⁵⁰. Another interesting result of this work, was that mutant viruses that packaged reduced amounts of viral genomic RNA tended to incorporate proportionally increased amounts of spliced viral RNA. The authors interpreted this to signify that spliced RNA may be encapsidated by the mutants in quantities sufficient to fill the space normally occupied by genomic RNA molecules. If this is the case, it implies that genomic RNA can be packaged as a monomer, and would be taken as evidence against dimerization being a prerequisite for packaging.

Infectivity assays performed on these viruses showed that mutants defective in genome encapsidation exhibited corresponding defects in infectivity. With respect to dimerization, the effects of the SL1 mutants were tested, and in agreement with the above reports, they observed that complete deletion of SL1, or even disruption of the base-pairing in the upper stem, resulted in the appearance of elevated amounts of monomer-sized RNA species on native Northern Blots, again confirming the importance of this region for the *in vivo* HIV-1 RNA dimerization process. Yet, these mutant genomes could still be packaged, which also suggested that HIV-1 RNAs need not be dimers in order to be packaged. Moreover, one mutant containing only a single substitution in the palindrome showed defects in infectivity that could not be explained by deficiencies in dimerization or packaging, suggesting that mutants lacking a functional DIS were defective in some later step in their life cycle which must also depend on dimerization.

To summarize, the two previous papers confirmed the importance of the DIS (KLD¹⁵¹, SL1⁵⁰) with respect to viral infectiousness, as reported above^{25,108,214}, and both papers also observed significant defects in RNA dimerization and packaging in the context of DIS-mutated viruses, further confirming the role of the DIS in these activities. RNA packaging analysis of both SL1 and SL3 mutant viruses showed an inverse relationship between the levels of spliced versus genomic RNA packaged into virus particles⁵⁰. This relationship suggests that the virus somehow prefers to package a certain critical quantity of viral RNA; when the specific packaging of genomic RNA is compromised, the virus may compensate by nonspecifically incorporating other RNA species.

By 1997, the dimeric RNA feature of the retroviral genome was considered to be important for the regulation of a number of critical steps in the viral life cycle. It was suggested that genomic RNA dimerization could be a negative regulator of *gag* and *gag-pol* expression¹⁶. Although heavily debated, the proximity between dimerization and packaging signals in numerous retroviral genomes led many groups to postulate that dimerization was linked to genome encapsidation^{5,32,47,60,61,113,129,157,178,180,183,225,281}. And, the presence of a dimeric genome was found to be important for recombination events during reverse transcription, thereby contributing to genetic diversity^{54,119,218,260,265}. Given the demonstrated importance of the DIS and various other 5' RNA sequences in these important aspects of the viral life cycle, our lab became interested in the structure and function of the HIV-1 5' UTR for two main reasons. First, due to the multiple functions of these RNA sequences, and the strong phenotypes observed when these sequences were mutated, our lab designed HIV-1 mutant viruses containing large deletions in these 5' RNA sequences that could potentially be used to test the feasibility of live attenuated vaccine candidates for HIV-1. Second, on a more basic level, the elucidation of the role of these 5' RNA sequences in various RNA-RNA and RNA-protein interactions would not only contribute to the basic understanding of the HIV-1 life cycle, but could potentially identify novel targets for anti-HIV therapy.

One of the first such studies initiated by Chen Liang formed the basis for much of the work described in this thesis. In a study aimed at further investigating the role of SL1 in HIV-1 replication, two DIS mutants were constructed, BH10-LD3, lacking nt 238-253 of SL1, and BH10-LD4, lacking nt 261-274¹⁶¹. Analysis of the BH10-LD3 mutant viruses generated from COS-7 cell transfections showed that this mutant was defective in both

packaging of viral genomic RNA and replication capacity. However, long-term culture in MT-2 cells for 18 passages resulted in the evolution of revertant viruses with near wild-type infectivity. When proviral DNA from these infected cells was sequenced, the original deleted sequences were still missing from the viral genome, but sequencing of the *gag* gene revealed two point mutations. The first was a C-T nt substitution resulting in a Thr-Ile mutation in the 12th position of the 14 amino acid p2 spacer peptide between CA and NC (termed MP2 for mutation within p2). The second was a C-T nt substitution that caused a Thr-Ile mutation at position 24 of the NC coding sequence within the first zinc finger (termed MNC for mutation in NC). To confirm the relevance of these compensatory point mutations, they were combined, alone and together, with the original BH10-LD3 deletion to generate mutant viruses LD3-MP2, LD3-MNC, and LD3-MP2-MNC. When these viruses were cultured in MT-2 cells it was observed that neither point mutation alone could restore wild-type replication capacity to the BH10-LD3 mutant. However, both MP2 and MNC together were able to restore replication capacity to almost wild-type levels. Next, to determine the mechanism of these compensatory point mutations, the levels of viral RNA packaged into wild-type and mutated viruses were analyzed. The results showed that BH10-LD3 packaged 4 to 5-fold lower levels of viral RNA. However, RT-PCR and slot blot analysis of viral RNA from the BH10-LD3-MP2-MNC revertant viruses showed that RNA packaging levels were restored to wild-type levels. The single mutant BH10-LD3-MNC also showed significant restoration of packaging, whereas viruses containing only the MP2 compensatory point mutation were still significantly impaired in this regard.

To test whether the effects of these mutations were specific for the BH10-LD3 deletion mutant, the same compensatory mutations were then combined with the BH10-LD4 deletion. MT-2 cell infections showed that the BH10-LD4 deletion mutant had a 6 day delay in peak viral replication compared to wild-type. The BH10-LD4-MP2-MNC mutant, on the other hand, showed only a 2 day delay in peak growth, indicating that these compensatory mutations could compensate for defects caused by two different DIS mutations. Overall, these results confirmed the involvement of SL1 in the efficient packaging of viral RNA. But more importantly, this strategy of forced evolution, identified a compensatory amino acid modification in the NC domain of Gag in response to deletion of RNA sequences in the 5' UTR. This finding represents *in vivo* evidence that a specific RNA-Protein interaction between SL1 and the first zinc finger of NC contributes to viral RNA packaging.

The contribution of the MP2 compensatory mutation to the overall reversion mechanism remained unclear at this point, but a subsequent paper showed that both the BH10-LD3 and BH10-LD4 mutants displayed abnormal ratios of CA to CA-p2 cleavage products¹⁶². These data suggested that both mutations resulted in a delay in Gag processing, and raised the possibility that non-coding viral RNA sequences may play a role in this process. Further evidence for this possibility was apparent when virus particles from BH10-LD5 (BH10-LD3 and BH10-LD4 deletions combined) were examined by EM. In the case of wild-type BH10, approximately 70% of the virus particles observed contained a condensed conical core, indicating that proper Gag cleavage and particle maturation had occurred. However, when virus particles from the BH10-LD5 mutant were subjected to the same analysis, only about 25% of the virions

appeared to be mature, consistent with delayed processing of the CA-p2 Gag cleavage product, as seen by radiolabeling and immunoprecipitation assays. Interestingly, when the MP2 and MNC compensatory mutations were combined with the LD5 deletion, both the Gag processing and particle maturation defects were corrected. Taken together, these data suggest that the MP2 and MNC compensatory mutations act through a concerted mechanism to restore replication capacity in which MNC increases packaging, while MP2 exerts its effects at the Gag processing level.

Although the BH10-LD3 mutant could replicate to near wild-type levels in the presence MP2 and MNC, these reverted viruses still showed a 2-3 day lag in growth compared to wild-type. In a follow-up study, the revertant virus isolated from the original 18th passage was continued in culture for an additional 11 passages, at which time it reacquired full replication capacity¹⁶³. When the 5'RNA and *gag* regions from 29th passage revertant virus were sequenced, the original BH10-LD3 deletion as well as the MP2 and MNC mutations still existed within the genome. However, the virus had acquired two additional mutations, a G-A nt substitution resulting in a Val-Ile mutation at position 35 in MA (termed MA1), and a T-C nt substitution causing an Ile-The mutation at position 91 in CA (termed CA1). Site-directed mutagenesis studies with various combinations of these 4 point mutations showed that each of these mutations contributed to the increase viability of the BH10-LD3 revertants, with MNC being the most important in this regard.

Similar extended culture was performed with the BH10-LD4 mutant virus, and, after 13 passages, a virus with wild-type replication properties emerged. As with BH10-LD3, sequencing of the 5'RNA and *gag* regions revealed that the original deleted nt were still

missing, but the revertant viruses had generated three point mutations in *gag*. These were the same MP2 and MNC described above, and an additional Val-Leu substitution at position 35 in MA (termed MA2). All three of these point mutations were required to restore infectiousness of BH10-LD4 to wild-type levels. One interesting, but at that time peculiar, observation was that MP2 alone was sufficient to confer some viability to BH10-LD4. The relevance of this observation will be more apparent below.

Unfortunately, the exact mechanisms employed by these compensatory point mutations to restore replication to the deletion mutants were still unclear. Since the original deletions were located within the DIS, it seemed plausible that these revertant mutations might somehow correct putative RNA dimerization defects caused by these deletions. However, this was not the case, and, it was later shown in collaboration with Michael Laughrea, that the BH10-LD3 mutant virus indeed showed decreased levels of dimeric genomic RNA compared to wild-type, but MP2 and MNC failed to increase the percentage of dimerized RNA contained within the virions²⁴⁸. Thus, these compensatory point mutations apparently act to restore viral replication to the deletion mutants by some mechanism that allows an overall increase in the amount of genomic RNA packaged into the virions, without any correction of dimerization defects. If such a mechanism were truly at work, it might mean that dimerization is not required for packaging, and these data would argue against the idea of a link between dimerization and packaging. However, it cannot be ruled out that the monomers detected on the gels were originally packaged as weak dimers that dissociated during extraction.

1.5.7 Remaining Questions

By the time I began my PhD project, a vast number of *in vitro* and *in vivo* studies had been carried out on the topic of RNA dimerization in HIV-1, but many important questions remained. Unfortunately, most of them still do. For example, it is still not known when Gag first interacts with genomic RNA during the encapsidation process, nor is it known whether the RNA is already dimerized during this initial interaction. As for Gag, it may also be in a dimeric or multimeric state when it first binds to the RNA. Dimerization is believed to take place during assembly and budding, but the exact timing of the formation of the loose dimer, and the subsequent rearrangement, is not known. In addition, the involvement of cellular factors in HIV-1 assembly and potentially, RNA dimerization, is of interest. It is also believed that the RNA may act as a scaffold for virus assembly²⁰⁶ and can affect virus particle morphogenesis^{38,39,92,206} and structural stability^{272,273}, but the mechanisms involved in these activities are currently unclear.

The first question addressed in my PhD project was whether RNA sequences outside the DIS could play a role in the HIV-1 RNA dimerization process. Results from *in vivo* dimerization studies showed that HIV-1 mutant viruses lacking large portions of the DIS, still displayed significant levels of dimerized RNA, and this mutant RNA dimer was as thermostable as that of wild-type. In contradiction to other retroviruses, EM analysis also showed that the free 5' ends of HIV-1 RNA could not be located, but instead, resembled a loop or ring-like structure. Assuming that the 3' contact point was generated by an interaction between DIS sequences on the two strands of RNA, then the other contact point must be somewhere upstream of the DIS. Ch. 2 of this thesis describes how we tried

to identify such upstream dimerization sequences and found that GU-rich regions in the poly(A) and U5-PBS stems were needed for efficient RNA dimerization.

Once we had established that RNA sequences outside the known DIS could play a role in HIV-1 RNA dimerization, and since these Poly(A) and U5-PBS regions, as well as the DIS, were all known to be important for genome packaging, we wondered whether other packaging signals might also contribute to the dimerization process. In Chs. 3 and 4 of this thesis, SL3, as the major packaging signal, and adjacent GA-rich sequences, were tested for their ability to affect dimerization as well as packaging. These chapters demonstrate that the structure but not the actual sequence of SL3 is important for HIV-1 RNA dimerization.

When I joined the lab, I was curious about the underlying mechanism(s) of the identified compensatory point mutations and performed work on this topic. We made two other DIS deletions, and combined these with various combinations of the compensatory point mutations that had been identified earlier by Dr. Liang. Even a virus completely lacking the DIS stem-loop, that never showed any sign of replication in tissue culture, was able to replicate to significant levels in the context of the compensatory point mutations, but the mechanism of these mutations still eluded us. Ch. 5 describes how we were eventually able to gain some insight into how one of these mutations works to restore replication.

CHAPTER 2

DEFICIENT DIMERIZATION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 RNA CAUSED BY MUTATIONS OF THE U5 RNA SEQUENCES

This chapter was adapted from an article that appeared in *Virology*, 2002, 303(1):152-63. The authors of this paper were **R.S. Russell**, J. Hu, M. Laughrea, M.A. Wainberg, and C. Liang. Over 50% of the data presented in this chapter were from experiments performed by myself under the supervision of Drs. Liang and Wainberg. J. Hu and Dr. Liang constructed some of the plasmids and performed all of the virus replication analyses. Dr. Laughrea critiqued the manuscript and offered suggestions for revision.

2.1 Preface to Chapter 2

Despite the fact that the DIS had been shown to play the major role in RNA dimerization initiation, and its biological significance had been demonstrated *in vivo*, results from many of these studies raised the possibility that RNA sequences outside the DIS might be involved in the HIV-1 RNA dimerization process. For example, the various DIS mutations decreased *in vivo* RNA dimerization to levels ranging from 0 to 40% of wild-type, and none of the mutations reported had any effect on the thermostability of the RNA dimer^{25,50,108,151,239,248}. One explanation for these observations is that other sequences must be involved in the formation and/or maintenance of the dimer structure. Other evidence comes from an EM study that was performed on HIV-1 genomic RNA. In this study, HIV-1 RNA, unlike that of most other retroviruses, lacked the characteristic Y- or T-like structures that were assumed to be the free 5' ends of the RNA¹¹⁸. Instead, HIV-1 RNA displayed two contact points that appeared to form a loop in the 5' region. Assuming that one of these contact points was generated by the DIS interaction, the other was then assumed to be formed by RNA sequences upstream of the DIS, possibly a palindromic sequence contained within the crown of the poly(A) hairpin, as suggested by others²⁵.

Based on these observations, we attempted to identify these putative RNA dimerization sequences. The first region we tested was a GU-rich region in the poly(A) and U5-PBS stems. We chose this region because it is known that G and U can form weak base-pairs, so we hypothesized that these GU-rich regions might bind to each other on the two strands of RNA to contribute to the formation and/or maintenance of the RNA dimer. The second region that we tested was the palindromic sequence contained within the loop of the poly(A) hairpin. Because this structure resembled the known DIS, we thought it might contribute to the RNA dimerization process through a kissing-loop mechanism similar to the one described for the DIS.

2.2 Abstract

The human immunodeficiency virus type 1 (HIV-1) virion contains two copies of genomic RNA that are non-covalently attached along a region at their 5' ends, in which two contact sites have been observed by electron microscopy. One of these sites is believed to be the stem-loop 1 (SL1) sequence which serves as the dimerization initiation site (DIS), and the other site, closer to the 5' end of the viral RNA, may involve the R or U5 RNA sequences. In this study, we present biochemical evidence showing that alteration of the U5 RNA sequence in the context of full-length viral RNA leads to diminished dimerization of virion RNA. In particular, two stretches of GU-rich sequences, that are located at nucleotides (nt) 99 to 108 and nt 112 to 123 within U5, were either deleted or substituted with exogenous sequences. The mutated viruses thus generated all exhibited deficient RNA dimerization. This dimerization deficit was not corrected by second-site mutations that preserved local RNA structures, such as the poly(A) hairpin, and was overcome to only a limited extent by compensatory mutations within Gag that were identified after long-term culture of the relevant mutant viruses in permissive cell lines and were able to restore viral infectiousness and RNA packaging to wild-type levels. Therefore, these GU-sequences do not regulate RNA dimerization by formation of local secondary structures nor by maintenance of efficient viral RNA packaging; instead, they may mediate direct RNA-RNA interactions in the dimer structure. In contrast, mutation of a palindrome 5'-AAGCUU-3', that resides within R and crowns the poly(A) hairpin, did not affect either RNA dimerization or packaging.

2.3 Introduction

All retroviruses possess diploid RNA genomes that are non-covalently dimerized in a parallel orientation near their 5' ends^{18,19,144,207}. This diploid status is believed to increase the rate of genetic recombination¹²⁰, and contribute to the overall genetic diversity^{55,99,271}. The dimer linkage site (DLS) of HIV-1 was originally mapped to a region downstream of the 5' major splice donor (SD) site that partially overlaps the *gag* coding region⁶⁰. Existence of purine-rich sequences in this DLS suggests formation of "purine quartet" structures during RNA dimerization^{16,183,262}. Subsequently, it was discovered that HIV-1 RNA sequences located upstream of the 5' SD site can form dimers in cell-free assays without the assistance of the originally defined DLS element¹⁸⁴. In this RNA fragment, a stem-loop structure, termed SL1, that is located between the primer binding site (PBS) and the 5' SD site, was shown to be indispensable for the dimerization process^{147,213,251}. The palindromic nature of the SL1 loop sequence (5'-GCGCGC-3' in HXB2) led to a "kissing-loop" model to explain the mechanisms involved in the initiation of dimerization; SL1 was thereafter termed the dimerization initiation site (DIS)²¹⁵. This hypothesis has been supported by results from both *in vitro* and *in vivo* studies^{49,50,108,148-152,215,216,248}.

However, the understanding of HIV-1 RNA dimerization has been complicated by the finding that significant amounts of dimerized viral genomic RNA can be found in mutant viruses that contain severe disruptions of the DIS structure; this implies the involvement of other viral RNA sequences in the dimerization process^{25,239}. Also, a recent electron microscopy (EM) study of HIV-1 genomic RNA revealed two linkage sites at the 5' ends

of dimerized RNA molecules¹¹⁸. One of these two linkage sites can be attributed to the DIS sequence, whereas the other one, closer to the 5' end of the viral RNA, may involve the R or U5 RNA sequences. Conceivably, even though disruptions of the DIS structure eliminate one linkage site, the other one may persist and thus still stabilize RNA dimers to a certain extent.

Sequence analysis reveals two types of RNA sequences within R and U5 that may participate in dimerization. One of these consists of two segments of GU-rich sequences that are located at nucleotides (nt) 99 to 108 and nt 112 to 123 in U5 (Fig. 2.1A). Since G and U can form base pairs, it follows that these GU-rich RNA sequences may bind to their counterparts in a second RNA molecule and therefore, contribute to the stabilization of RNA dimers. The other sequence is a 5'-AAGCUU-3' palindrome that is located at nt 77 to 82 (Fig. 2.2A). Since this palindrome is positioned at the top of the poly(A) hairpin (Fig. 2.2A), a feature that resembles the DIS structure, we speculate that it may also play a role in RNA dimerization via a "kissing-loop" mechanism similar to that of the DIS.

To assess the roles of the aforementioned RNA sequences in RNA dimerization, we mutated these RNA sequences in the context of a full-length viral genome and measured their effects on the conformation of virion-associated genomic RNA by performing native Northern blots. Our results revealed that the GU-rich sequences within U5, but not the 5'-AAGCUU-3' palindrome in the R region, were needed for efficient RNA dimerization.

2.4 Materials and Methods

2.4.1 Plasmid construction

The BH10 infectious HIV-1 cDNA clone was used as starting material for mutagenesis studies. The Δ GU2 and Δ GU3 deletions lack viral sequences at nt 99 to 108 and nt 112 to 123, respectively (Fig. 2.1A). The nucleotide positions refer to the first nt of the 5' R region. These two deletions were further recombined to generate construct Δ GU2-3. Δ GU2 was engineered by PCR using primer pair p Δ GU2 (5'-GCCTTGAGTGCTTCAACCCGTCTGTTGTGTGACTCTGG-3' [nt 83 to 130])/pA (5'-CCATCGATCTAATTCTCCC-3' [383 to 365]). The PCR product was used as a primer in a second round of PCR together with primer pHpa-S (5'-CTGCAGTTAACTGGAAGGGCTAATTCCTCCC-3' [nt -454 to -433]). The final product was digested with restriction enzymes *Hpa* I and *Bss*H II and inserted into BH10 to generate Δ GU2. Δ GU3 and Δ GU2-3 were generated with a similar strategy through use of primers p Δ GU3 (5'-GCTTCAAGTAGTGTGTGCCCACTCTGGTAACTAGAGATC-3' [nt 92 to 142]) and p Δ GU2-3 (5'-GCCTTGAGTGCTTCAACCCACTCTGGTAACTAGAGATC-3' [nt 83 to 142]), respectively.

Constructs AS4 and AS5 contain substituted sequences at nt 97 to 104 and nt 58 to 65, respectively (Fig. 2.2A). Δ (A) lacks the RNA sequence 5'-AAUAAA-3' at nt 73 to 78 (Fig. 2.2A). These constructs were engineered by PCR using primers pAS4 (5'-CTTGAGTGCTTCTCCATCATTGTGCCCGTCTGT-3' [nt 85 to 117]), pAS5 (5'-CAAGCTTTATTGAGGCTTCCATCATGGTTCCTAGTTAGCC-3' [nt 83 to 42]), and p(Δ A) (5'-CCCCTGCTTAAGCCTCGCTTGCCCTGAGTGCTTC-3' [nt 56 to 96]), respectively.

The T12I point mutation was generated as previously described¹⁶¹ (Fig. 2.3A). The V13G substitution in the NC protein (Fig. 2.3A) was constructed by PCR using primer pair pV13G (5'-CCAAAGAAAGATTGGTAAGTGTTCATTG-3' [nt 1490 to 1519])/pAPA-A (5'-CCTAGGGGCCCTGCAATTTCTG-3' [nt 1562 to 1541]). This PCR product was used as a primer in a second round of PCR together with primer pSph-S (5'-AGTGCATCCAGTGCATGCAGGGCC-3' [nt 977 to 1000]). The final PCR product was digested with *Apa* I and *Sph* I and inserted into BH10, ΔGU2 or ΔGU3. All primers were purchased from Invitrogen (Burlington, Ontario, Canada).

2.4.2 Cell culture, transfection and infection

COS-7 and MT-2 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) and RPMI 1640 medium, respectively, supplemented with 10% fetal calf serum. Transfection of COS-7 cells was performed with Lipofectamine (Invitrogen). Progeny viruses were harvested 48 hrs after transfection and quantities of viruses were determined by measuring levels of p24 (CA) antigen by enzyme-linked immunosorption assays (ELISA) (Vironostika HIV-1 Antigen Microelisa System, Organon Teknika Corporation, Durham, NC).

Amounts of virus equivalent to 3 ng of p24 were used to infect 3×10^5 MT-2 cells. Cells were washed twice at 2 hrs after infection and cultured in complete RPMI 1640 medium. Reverse transcriptase (RT) activity in culture fluids was measured at various times to monitor viral growth.

2.4.3 Analysis of viral proteins by Western blots

The culture fluids from transfected COS-7 cells were first clarified at 3,000 rpm in a Beckman GS-6R bench top centrifuge at 4°C for 30 min. Progeny viruses were then pelleted through a 20% sucrose cushion using a SW41 rotor in a L8-M ultracentrifuge at 40,000 rpm for 1 hr at 4°C. The virus pellet was suspended in NP-40 lysis buffer²⁴⁰. The transfected COS-7 cells were washed twice with cold phosphate-buffered saline and scraped from the plates. The cells were then lysed in NP-40 lysis buffer. Both virus and cell lysates were fractionated on sodium dodecyl sulfate (SDS)-12% polyacrylamide gels (PAGE) and proteins were transferred onto a PVDF membrane (Roche Inc., Laval, Quebec, Canada). Viral proteins were detected using IgG monoclonal antibodies (MAbs) against HIV-1 p24 (CA) antigen (ID Labs Inc., London, Ontario, Canada).

2.4.4 Analysis of viral RNA by Northern blots and RT-PCR

To prepare sufficient amounts of viral RNA from virus particles for native Northern blots, approximately 60 ml of culture supernatants from transfected COS-7 cells were collected and spun in 70 ml tubes at 35,000 rpm for 1 hr at 4°C using a Ti45 rotor in a Beckman L8-M ultracentrifuge. Virus pellets were suspended in 10 ml of DMEM and spun through a 20% sucrose cushion in TN buffer at 40,000 rpm for 1 hr at 4°C using a SW41 rotor. Virus pellets were resuspended in 300 µl of TN buffer; a 2 µl portion was removed for p24 determination, and remaining virions were disrupted using a 4X lysis buffer (final concentrations of 50 mM Tris-Cl [pH7.4], 10 mM EDTA, 1% SDS, 100 mM

NaCl) as previously described⁹¹. Samples were digested with proteinase K (100 µg/ml) in the presence of yeast tRNA (100 µg/ml) (Invitrogen) for 20 min at 37°C, then extracted twice with an equal volume of phenol/chloroform/isoamylalcohol (25:24:1) (Invitrogen), and once with chloroform. The RNA was precipitated in 2.5 volumes of 95% ethanol and a 0.1 volume of 3 M sodium acetate [pH5.2]. RNA pellets were washed with 70% ethanol and dissolved in TE buffer.

Native Northern blots were performed as follows. An amount of viral RNA equivalent to 300 ng of p24 was separated on 0.9% agarose gels at 100v in 1X TBE buffer for 4 hrs at 4°C. RNA was transferred to a nylon membrane (Amersham Pharmacia Biotech., Montreal, Quebec, Canada) and baked at 80°C for 2 hrs. The membranes were prehybridized for 3 hrs at 42°C in a hybridization buffer containing 6X SSPE [pH7.4], 50% formamide, 0.5% SDS, 5X Denhardt's, and 80 µg/ml herring sperm DNA (Invitrogen). A 2 kb HIV-1 DNA fragment at nt 1 to 2000 was nick-labeled with ³²P-α-dCTP (ICN, Irvine, CA) and used as a probe. After prehybridization, the membranes were hybridized with 1.5x10⁶ cpm (400 ng DNA) of probe for 18 hrs. Membranes were washed for 10 min at room temperature, then 65°C, with each of 2X-, 1X-, and 0.2X SSPE, 0.1% SDS and exposed to X-ray films.

To determine the thermostability of virion-derived RNA, RNA samples were first treated at different temperatures (i.e. 40°C, 45°C, 50°C, and 55°C) for 10 min in the presence of 100 mM NaCl before being fractionated on 0.9% agarose gels.

An amount of progeny virus equivalent to 10 ng of p24 was used to prepare viral RNA using the QIAamp Viral RNA Mini Kit (QIAGEN, Mississauga, Ontario, Canada). Viral RNA from viruses containing 0.2 ng of p24 was first treated with 20 units of

RNase-free DNase I (Invitrogen) to remove any possible DNA contamination and was then subjected to reverse transcription (RT)-PCR (Titan One-Tube RT-PCR system, Roche Inc.) using primer pair pGAG1/pST-A¹⁶¹. The RT-PCR products were fractionated on 5% native polyacrylamide gels and visualized by exposure to X-ray films.

Cellular RNA was studied by denaturing Northern blots as follows. The transfected COS-7 cells were washed twice with cold phosphate-buffered saline and lysed with NP-40 lysis buffer. The cell lysates were digested with proteinase K (100 µg/ml) for 20 min at 37°C, and RNA was extracted as described above. A 1% denaturing agarose gel containing 1X MOPS buffer and 18% formaldehyde was prepared as described²⁴⁰. RNA samples were denatured by heating in the presence of formamide and formaldehyde, then loaded onto the gels, and run at 100 volts for 4 hrs. After electrophoresis, the gel was rinsed in distilled water and RNA was transferred onto nylon membranes. Prehybridization and hybridization were performed as described above. Similar amounts of RNA were also run on a 1% agarose gel and stained with ethidium bromide (EtBr) in order to visualize 28S and 18S ribosomal RNAs.

2.5 Results

2.5.1 Deletions of the GU-rich sequences in the HIV-1 U5 region cause deficient viral RNA dimerization and packaging

Two stretches of GU-rich sequences are located at nt 99 to 108 and nt 112 to 123 within the U5 region (Fig. 2.1A). We assessed the roles of these RNA sequences in RNA

dimerization by deletion of nt 99 to 108 in construct Δ GU2 and deletion of nt 112 to 123 in construct Δ GU3. To study the potential synergistic effects of these two RNA stretches, both sequences were deleted to produce the Δ GU2-3 construct (Fig. 2.1A). When these mutant viral DNA and wild-type BH10 were transfected into COS-7 cells, the mutant viruses thus generated showed only minor effects in the processing of Gag proteins compared to wild-type, as demonstrated by the results of Western blots performed either with cell lysates or with virus particles (Fig. 2.1B). Next, we used denaturing Northern blots to examine the viral RNA species expressed upon transfection of these mutant DNA constructs into COS-7 cells. Three major viral RNA bands were observed for all of the DNA constructs tested, these represented full-length (9.2k nt) and spliced (4k and 2k nt) viral RNAs (Fig. 2.1C). Of the three mutant constructs, Δ GU2 and Δ GU2-3 exhibited moderate reductions in overall quantities of viral RNA (Fig. 2.1C).

Infectiousness of mutant viruses was then examined by infection of MT-2 cells with an amount of virus containing 3 ng of p24 antigen. In comparison to wild-type virus, the Δ GU3 deletion delayed viral replication by approximately 8 days. High levels of viral replication were not seen with the Δ GU2 mutant virus until the end of the fourth week in culture, whereas infectious virus was never recovered with the Δ GU2-3 mutation (Fig. 2.1D). Therefore, the Δ GU2 deletion caused a more severe deficit in viral replication than did Δ GU3, and both deletions together virtually eliminated viral infectiousness.

To investigate the effects of these deletion mutations on viral RNA dimerization, viral RNA was extracted from each of the Δ GU2, Δ GU3, and Δ GU2-3 mutant viruses under mild conditions and fractionated on native agarose gels, followed by Northern blot analysis. The results shown in Fig. 2.1E demonstrate that the BH10 wild-type virus

contained mainly the dimeric form of viral RNA. In contrast, each of the mutated viruses displayed predominantly the monomeric form of viral RNA. The relative percentages of dimer versus monomer for each mutant are summarized in the bar graph on the right. It was also observed that the Δ GU2 and Δ GU2-3 mutated viruses exhibited weak RNA signals due to low overall levels of viral RNA packaging. To clearly visualize the viral RNA bands associated with these two mutant viruses, additional analysis was performed with increased amounts of viral RNA and these results confirmed that Δ GU2 and Δ GU2-3 mutant viruses contained mainly monomeric RNA (Fig. 2.1E, bottom panel). To verify dimer versus monomer features of the observed RNA bands, RNA samples prepared from either Δ GU3 or BH10 viruses were first treated at various temperatures before being separated on native agarose gels. In the case of BH10, RNA dimers dissociated at 50°C and migrated as monomers on the gels (Fig. 2.1F). Similarly, the weak dimer bands associated with Δ GU3 disappeared at high temperatures and migrated as monomers (Fig. 2.1F). Taken together, we conclude that these GU-sequences are needed for efficient viral RNA dimerization.

It is also noticeable that these three deletion mutations exhibited lower levels of viral RNA (including both the dimeric and monomeric forms) than those of the wild-type BH10, with Δ GU2 and Δ GU2-3 showing the most adverse effects (Fig. 2.1E, the top panel). Since both the mutant and wild-type RNA were prepared from virus particles that contained 300 ng of p24, it follows that these GU-rich sequences, especially those located at nt 99 to 108, are apparently involved in RNA packaging. This conclusion was further supported by results of RT-PCR (Fig. 2.1G). A fragment of *gag* RNA was amplified to measure the quantities of full-length viral RNA prepared from either mutant or wild-type

Figure 2.1. Various effects of two stretches of GU-rich sequences, located within the U5 region, on viral gene expression, viral replication, viral RNA packaging and dimerization. (A) Illustration of the GU-sequences and their locations within U5. The sequences at nt 99 to 108 and nt 112 to 123 were deleted in constructs Δ GU2 and Δ GU3, respectively. Both nucleotide stretches were eliminated in Δ GU2-3. (B) Viral protein analysis by Western blots using antibodies against p24 (CA). Western blots were performed either with cell lysates from transfected COS-7 cells or with lysed virus particles harvested from culture fluids by ultracentrifugation. (C) Viral RNA analysis by denaturing Northern blots. Transfected COS-7 cells were first lysed in NP-40 lysis buffer. RNA was extracted from cell lysates using phenol/chloroform/isoamylalcohol and then subjected to denaturing Northern blots. To rule out possible saturation of the hybridization signals, two dilutions for each RNA sample (1:1 (lanes 1 to 5) and 1:3 (lanes 6 to 10)) were loaded on the gels. RNA was prepared from mock transfected COS-7 cells and also used in the hybridization experiments to serve as negative controls. Three major RNA bands were seen that represent full-length (9.2k nt) and spliced (4k and 2k nt) viral RNA species. Ribosomal RNAs (28S and 18S) were visualized by staining the agarose gels with EtBr, and the signals serve as internal controls for RNA integrity and to ensure that the same amounts of total cellular RNA had been loaded for each construct. (D) Infectiousness of the mutant and wild-type viruses in MT-2 cells. Progeny viruses generated by transfected COS-7 cells were used to infect MT-2 cells. Viral growth was monitored by measuring RT activities in the culture fluids at various time points.

Figure 2.1. Various effects of two stretches of GU-rich sequences, located within the U5 region, on viral gene expression, viral replication, viral RNA packaging and dimerization.

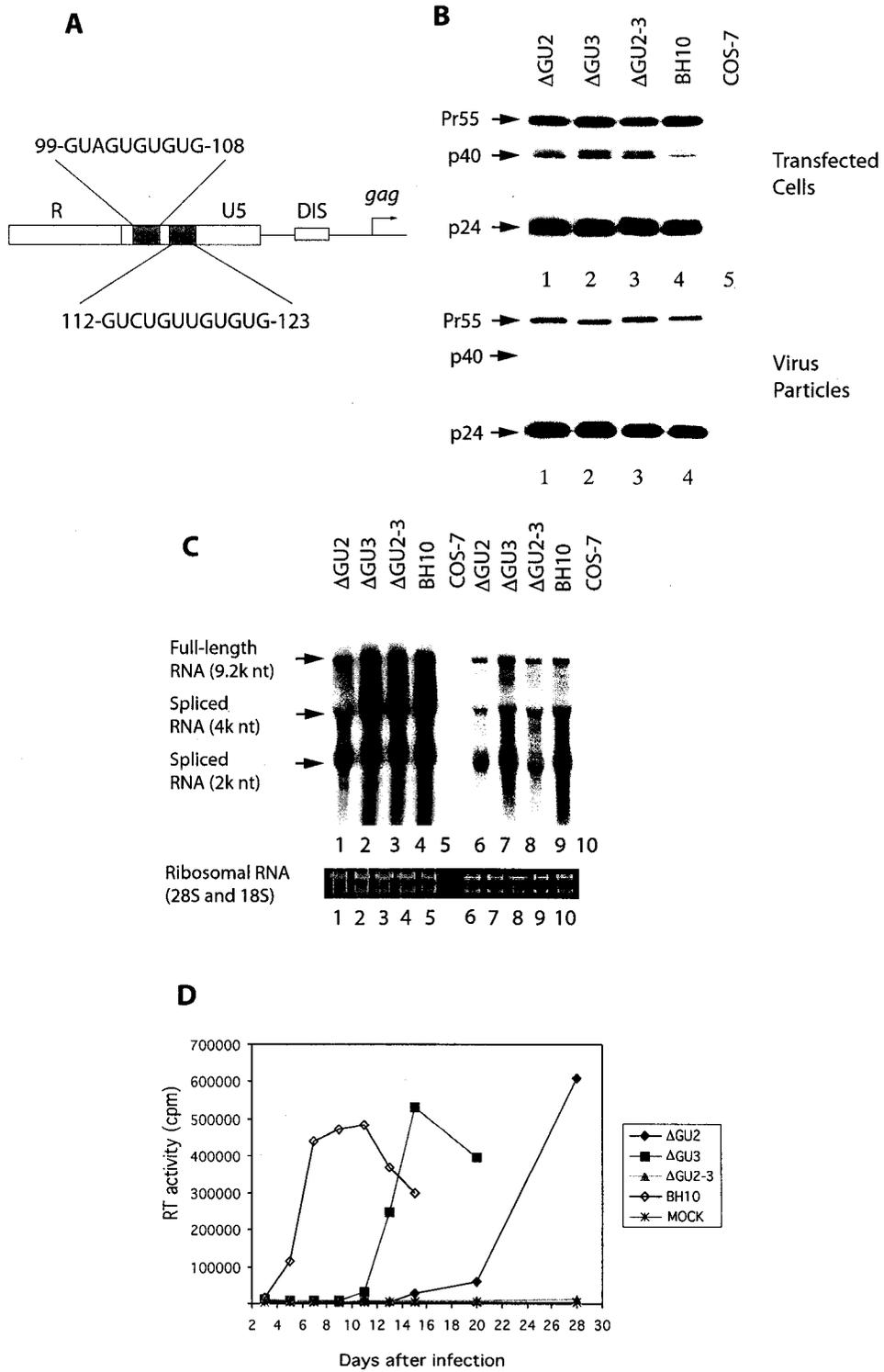
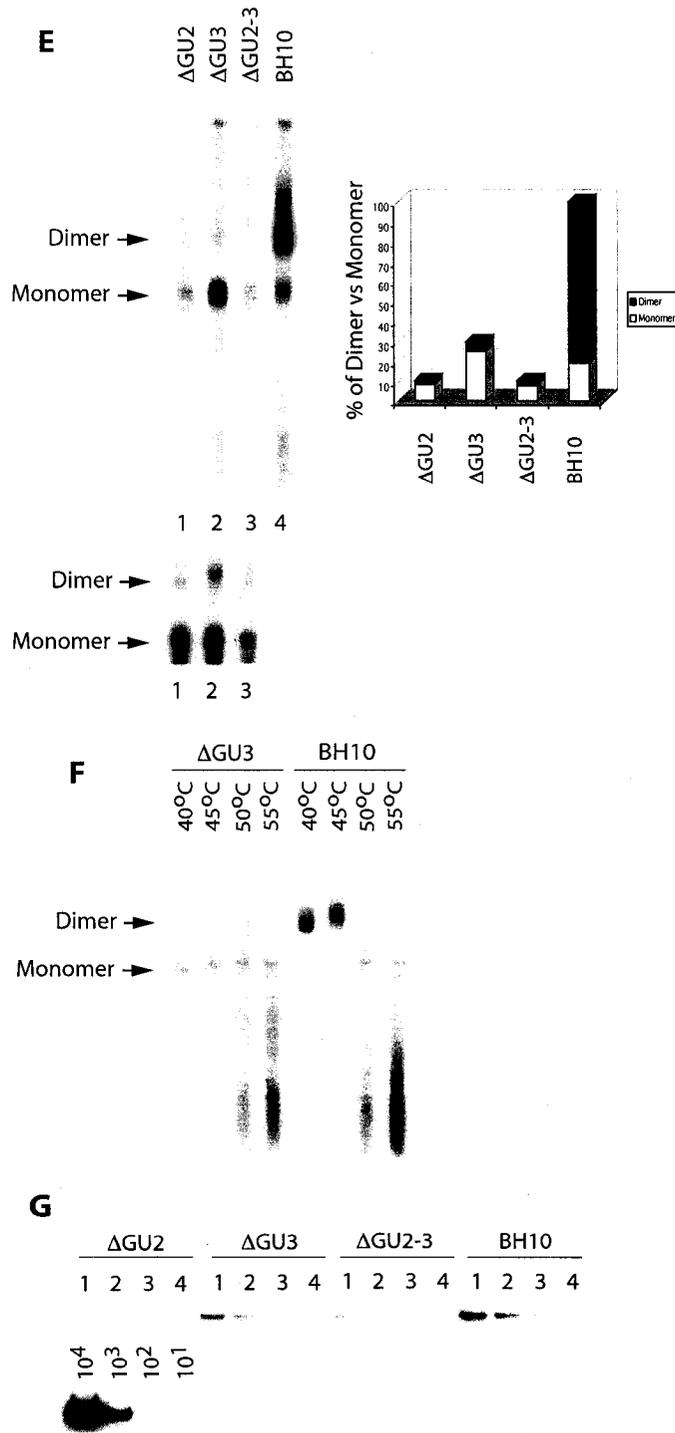


Figure 2.1. (cont.) Various effects of two stretches of GU-rich sequences, located within the U5 region, on viral gene expression, viral replication, viral RNA packaging and dimerization. (E) Analysis of viral RNA within virus particles by native Northern blots. Viral RNA for each construct was prepared from virus particles containing 300 ng of p24 (CA) under mild conditions (in the presence of 100mM NaCl) and then fractionated on native agarose gels before being subjected to Northern blotting. Positions of RNA dimers and monomers are labeled on the left side of the gels. The proportion of each RNA form in each construct was measured using the NIH Image program and plotted. The data shown are calculated from one representative experiment. RNA levels in wild-type BH10 were arbitrarily set at 100. In order to clearly visualize the two RNA forms in the mutant viruses, higher amounts of RNA samples were loaded on the gels and the results are shown in the bottom panel. (F) Thermostability of RNA dimers. Virion RNA was first treated at various temperatures (40°C, 45°C, 50°C, and 55°C) for 10 min in the presence of 100 mM NaCl before being separated on agarose gels. (G) Measurement of full-length viral RNA by semi-quantitative RT-PCR. Three dilutions of virion RNA for each construct were used in RT-PCR (1:1 (lane 1), 1:3 (lane 2) and 1:9 (lane 3)). RNase A-treated RNA samples were amplified by RT-PCR to serve as negative controls. Proviral DNA representing 10^1 , 10^2 , 10^3 and 10^4 copies of HIV were subjected to RT-PCR to demonstrate the linear range of the reactions.

Figure 2.1. (cont.) Various effects of two stretches of GU-rich sequences, located within the U5 region, on viral gene expression, viral replication, viral RNA packaging and dimerization.



virus particles. Consistent with the observations made with the native Northern blots (Fig. 2.1E), Δ GU2 and Δ GU2-3 caused significant reductions in RNA packaging, whereas Δ GU3 alone showed a modest defect in this regard (Fig. 2.1G).

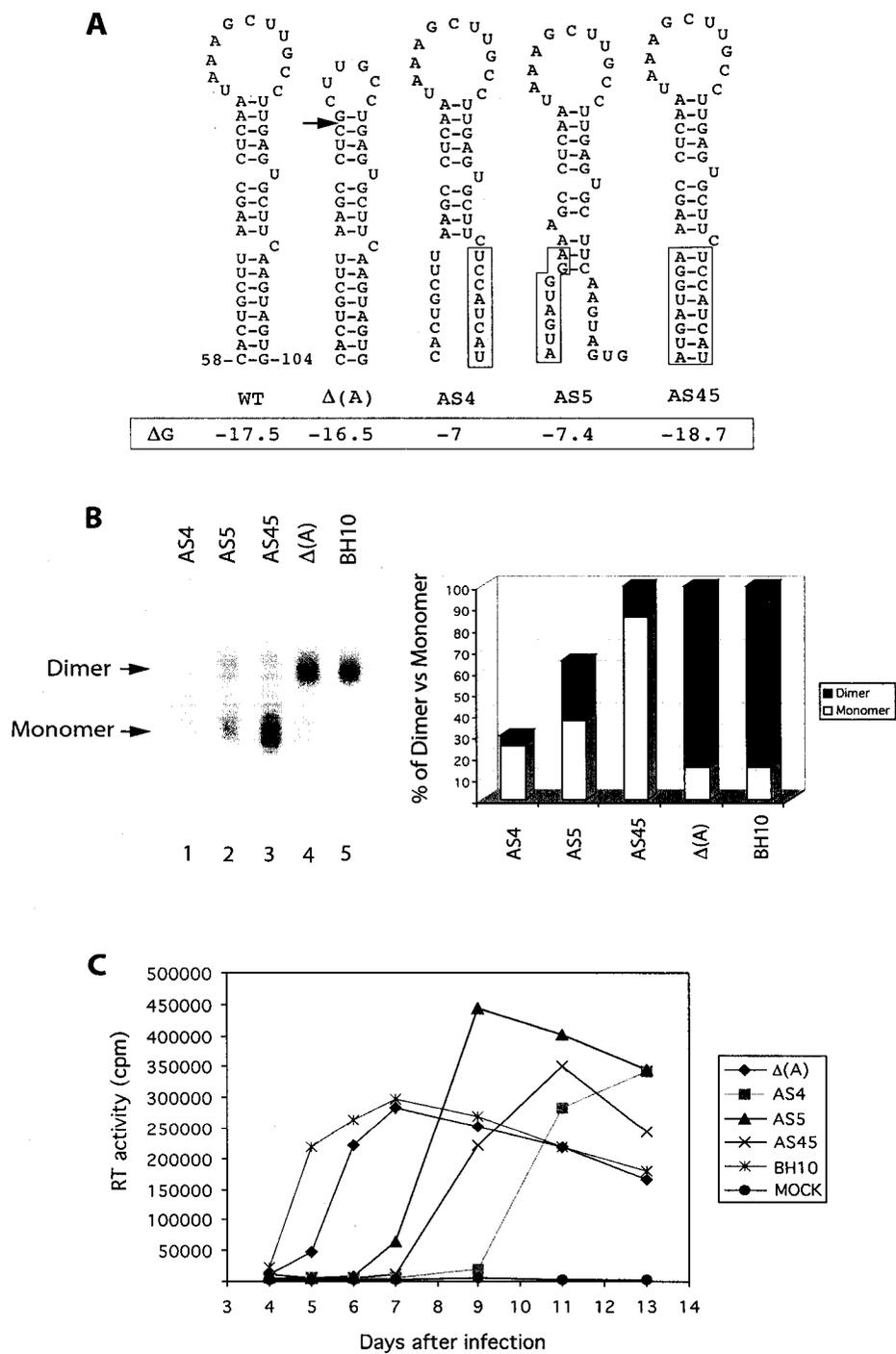
2.5.2 A second-site mutation can correct deficient RNA packaging, but not the diminished dimerization, caused by mutations of the GU-sequence [nt 99 to 108]

Next, we selectively mutated part of the GU-sequence [nt 99 to 108] and generated construct AS4 in which the nt sequences 5'-AAGUAGUG-3' [nt 97 to 104] were changed to 5'-UCCAUCAU-3' (Fig. 2.2A). This AS4 mutant virus exhibited severe defects in both RNA dimerization and packaging (Fig. 2.2B), concurrent with significantly attenuated infectivity in permissive cells (Fig. 2.2C).

Since these GU-sequences [nt 99 to 108] constitute part of the poly(A) hairpin (Fig. 2.2A), an RNA structure that has been shown to participate in RNA packaging^{51,62,189} it is possible that the AS4 mutation exerted its effects on RNA dimerization and packaging by disrupting this RNA structure. Indeed, the results of RNA secondary structure modeling based on the M-Fold program^{186,291}, revealed a negative impact of AS4 on the poly(A) structure (Fig. 2.2A). To further evaluate the sequence versus structural effects of the AS4 mutation, a second-site mutation AS5 was designed and further recombined with AS4 in order to restore the poly(A) stem (Fig. 2.2A). The resultant structures were predicted on the basis of the M-Fold program^{186,291}, and are shown in Fig. 2.2A. The AS5 mutation alone caused moderate defects in viral RNA packaging and dimerization

Figure 2.2. Substitutions of the poly(A) stem or loop sequences and their effects on viral infectivity and viral RNA dimerization. (A) Illustration of the $\Delta(A)$, AS4, AS5, and AS45 mutations. The $\Delta(A)$ mutation lacks the RNA sequence 5'-AAUAAA-3' at nt 73 to 78. The bottom stem structure of the poly(A) hairpin was destabilized by mutations AS4 and AS5 and restored in construct AS45. The modified sequences in mutations AS4, AS5, and AS45 are highlighted. The given structures and the relevant ΔG values (kcal/mol) were predicted by the M-Fold program^{186,291}. (B) Effects of these mutations on viral RNA dimerization and packaging as determined by native Northern blots. Viral RNA was prepared from virus particles containing 300 ng of p24 antigen. Intensities of RNA signals were measured using the NIH Image program and data from one representative experiment are shown. RNA levels from wild-type BH10 were arbitrarily set at 100. (C) Replication kinetics of the AS4, AS5, AS45, or $\Delta(A)$ mutant viruses. Virus growth in MT-2 cells was monitored by measuring RT activity in culture fluids at various times.

Figure 2.2. Substitutions of the poly(A) stem or loop sequences and their effects on viral infectivity and viral RNA dimerization.



(Fig. 2.2B). Recombination of AS4 and AS5 in AS45 led to wild-type levels of RNA packaging in contrast to severely deficient dimerization. Consistent with the elevated RNA packaging, AS45 mutant viruses showed increased infectiousness compared with AS4 (Fig. 2.2C). Therefore, deficient RNA dimerization caused by mutations of the GU-sequence [nt 99 to 108] persisted in the face of wild-type RNA packaging restored by a second-site mutation that preserves the poly(A) RNA structure. These results suggest that the GU-sequences studied are directly involved in RNA-RNA interactions within the dimer structure.

2.5.3 The palindrome residing on the top of the poly(A) hairpin does not contribute to RNA dimerization

Next, we deleted part of the loop sequence at nt 73 to 78 in construct $\Delta(A)$ to determine whether the palindrome 5'-AAGCUU-3' [nt 77 to 82], that crowns the 5' poly(A) hairpin, is able to regulate dimerization via a "kissing-loop" mechanism as does the DIS motif (Fig. 2.2A). It is important to note here that no mutations were made in the 3' poly(A) region, thus the polyadenylation of viral RNA transcripts would not be affected in any of these mutants since the 5' poly(A) is silenced^{23,62,65,103,136,137}. The results of Fig. 2.2B show that the $\Delta(A)$ mutant virus contained wild-type levels of dimeric RNA. Consistently, the $\Delta(A)$ deletion had no apparent effect on viral replication (Fig. 2.2C), as would be expected since this mutation would quickly revert to the wild-type sequence during reverse transcription due to its position in R. Based on these results, we

believe that the loop sequence of the poly(A) hairpin is not involved in viral RNA dimerization.

2.5.4 The T12I (in p2) and V13G (in NC) point mutations restore the replication capacity of Δ GU3 but not of Δ GU2 to nearly wild-type levels

The Δ GU3 mutated virus was able to produce high levels of viral RT activity and cause syncytium formation after 15 days, indicating that rapid viral replication had occurred (Fig. 2.1D). The Δ GU2 mutated virus yielded infectious virus after 28 days (Fig. 2.1D). In contrast, Δ GU2-3 failed to replicate over 60 days after a single infection event (data not shown). Viruses derived from the Δ GU2 and Δ GU3 infections were further used to infect fresh MT-2 cells over multiple infection cycles until wild-type replication kinetics had been established. Proviral DNA was then extracted from the infected MT-2 cells and assessed by PCR, cloning and sequencing. We found that the Δ GU3 deletion had been sustained during this time, while the viral sequences that had been deleted in Δ GU2 had been restored.

In order to identify any second-site mutations that might have conferred wild-type replication kinetics to Δ GU3, the 5' non-coding leader sequences and the *gag*-coding region of Δ GU3 proviral DNA were sequenced. Two point mutations were identified, i.e. a T12I substitution in p2 and a V13G substitution in NC (Fig. 2.3A). These two substitutions were then recombined either separately or together with the Δ GU3 deletion to test whether they were sufficient to rescue the defect caused by this mutation. The results in Fig. 2.3B show that V13G only slightly increased the infectivity of Δ GU3,

while T12I markedly stimulated viral replication. When both of these mutations were present (Δ GU3-T12I-V13G), replication kinetics close to wild-type levels were observed (Fig. 2.3B). Therefore, both T12I and V13G play key roles in the rescue of the Δ GU3 mutant virus.

Since the sequences that were deleted in Δ GU2 and Δ GU3 are similar, we wished to know whether the T12I and V13G substitutions, identified in Δ GU3, could also rescue Δ GU2. Toward this end, these two point mutations were inserted into Δ GU2 and the progeny viruses that were recovered after transfection with these recombinant constructs were used to infect fresh MT-2 cells. The results in Fig. 2.3C show that no viral infectiousness resulted from this step. Based on these results, we conclude that T12I and V13G were unable to rescue Δ GU2 and that the location of the deleted sequence within Δ GU2 as opposed to Δ GU3 may be of critical importance in this regard. In addition, the results of infectivity studies showed that neither V13G nor T12I alone, or in combination, compromised the replication capacity of the wild-type BH10 (Fig. 2.3C).

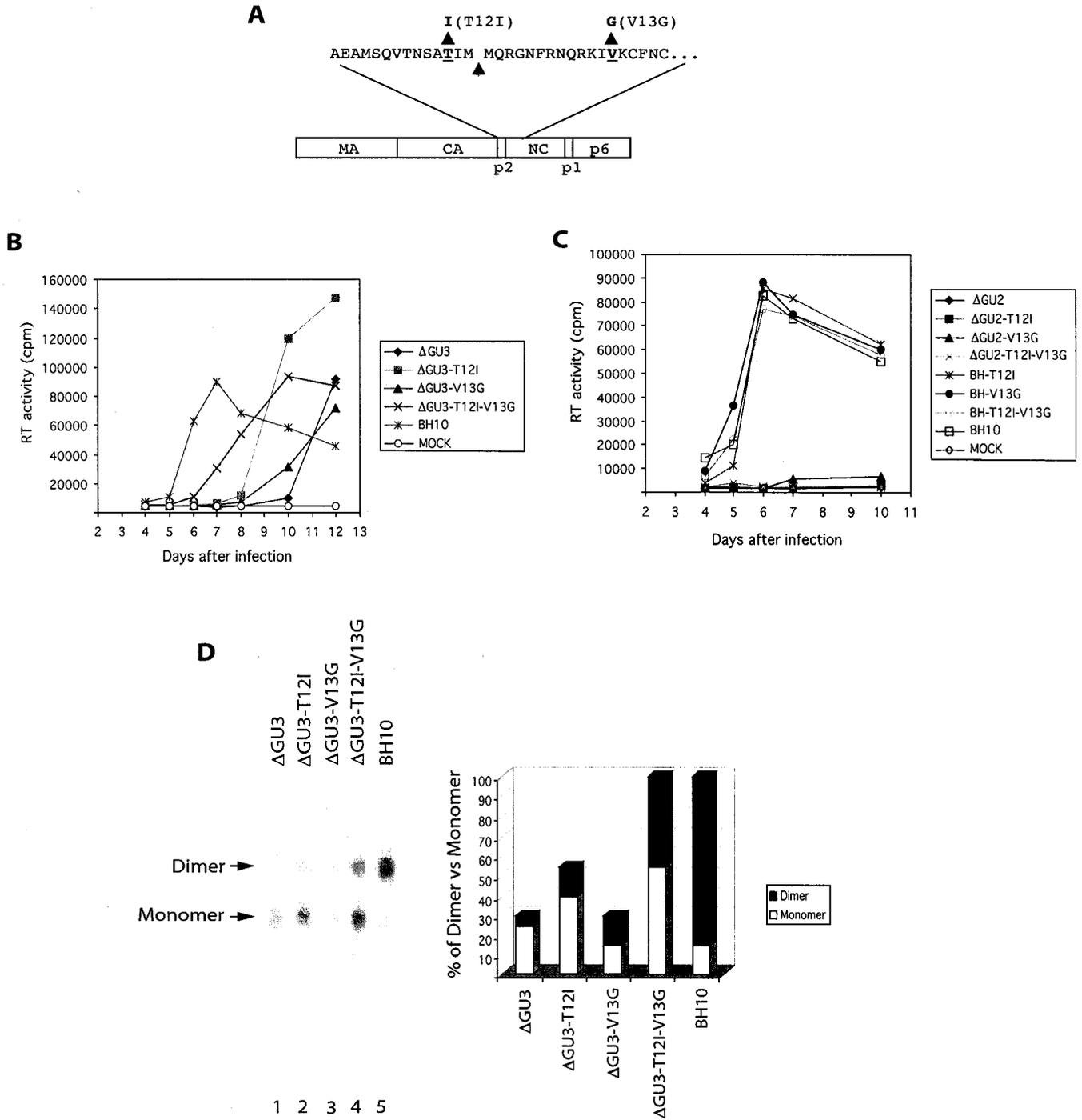
2.5.5 The T12I and V13G mutations compensate for defective RNA packaging caused by the Δ GU3 deletion

Next, we measured the levels of dimeric RNA in the recombinant viruses Δ GU3-T12I, Δ GU3-V13G, and Δ GU3-T12I-V13G by native Northern blots. In comparison with the Δ GU3 construct, Δ GU3-T12I displayed slightly higher levels of dimeric viral RNA (Fig. 2.3D). In the case of Δ GU3-V13G and Δ GU3-T12I-V13G, marked increases in the amounts of dimers were observed, however, these were still lower than that of wild-type

Figure 2.3. Compensation of the Δ GU3 deletion by second-site mutations in Gag.

(A) Illustration of the second-site mutations found in p2 (T12I) and NC (V13G). The mutated amino acids are underlined. The protease cleavage site between p2 and NC is indicated by an arrow. (B) Rescue of the Δ GU3 mutated viruses by the T12I and V13G mutations. T12I and V13G were recombined with the Δ GU3 deletion to generate constructs Δ GU3-T12I, Δ GU3-V13G, and Δ GU3-T12I-V13G. MT-2 cells were infected with relevant recombinant viruses and RT activity in culture fluids was measured at various times to monitor virus growth. (C) Effects of the T12I and V13G mutations on infectiousness of Δ GU2 and wild-type viruses. (D) Effectiveness of the T12I and V13G compensatory mutations in correcting the defective RNA dimerization and packaging associated with the Δ GU3 deletion. Viral RNA was extracted from virus particles containing 300 ng of p24 antigen and then analyzed by native Northern blots. RNA signals were measured using the NIH Image program and relative levels of viral RNA associated with each construct were plotted. RNA levels in wild-type BH10 were arbitrarily set at 100. Data shown are from one representative experiment.

Figure 2.3. Compensation of the Δ GU3 deletion by second-site mutations in Gag.



virus (Fig. 2.3D). It was also found that the Δ GU3-T12I-V13G recombinant virus contained wild-type levels of total viral RNA (including both dimeric and monomeric forms) (Fig. 2.3D). Therefore, the compensatory mutations T12I and V13G are able to improve viral RNA dimerization in Δ GU3 to higher levels, which, together with the restored RNA packaging, may account for the increased infectivity of the Δ GU3 revertants.

2.6 Discussion

EM observation of two tightly associated retroviral RNA molecules in diverse types of retroviruses represents visual evidence for the existence of viral RNA dimers^{18,19,77,143,144,172,207,254}. The region of RNA involved in dimerization commonly displays an X-like structure in EM, indicating a single contact site present between the two RNA molecules. In contrast, when the conformation of HIV-1 genomic RNA was studied by EM, the 5' regions of the two linked viral RNA molecules were found to display a ring-like structure¹¹⁸, suggesting two contact points between the RNA strands. One of these is thought to be facilitated by the DIS, while the other may involve the R and U5 sequences. Our studies now provide *in vivo* biochemical evidence supporting the participation of U5 RNA sequences in HIV-1 RNA dimerization.

In this study, we have focused on the potential dimerization roles of two stretches of GU-rich sequences located within the U5 region. Since G and U bases can form hydrogen bonds and thus contribute to interactions between RNA strands, we hypothesize that GU-sequences in U5 may assist in the stabilization of RNA dimer structures. Indeed, either

deletions (i.e. Δ GU2, Δ GU3, and Δ GU2-3) or substitutions (i.e. AS4) of these GU-sequences impaired RNA dimerization.

In order to understand the mechanisms by which these GU-sequences affect RNA dimerization, two issues need to be clarified. First, since these GU-sequences, especially the one located at nt 99 to 108, are believed to participate in complex RNA folding, it is speculated that mutation of these GU-sequences may have disrupted the relevant RNA structures and, as a consequence, indirectly compromised RNA dimerization. If RNA structures, rather than RNA sequences, play major roles in RNA packaging and dimerization, then preservation of these structures with exogenous sequences should ensure efficient RNA packaging and dimerization. However, when the disrupted poly(A) structure caused by the AS4 mutation was restored by a second-site mutation, i.e. AS5, the impairment in dimerization persisted, even though RNA packaging was restored to wild-type levels (Fig. 2.2B). Hence, it appears that the GU-sequences [nt 99 to 108], that were changed in the AS4 mutation, exert their effects on dimerization in a sequence-specific manner. We are now in the process of establishing an *in vitro* dimerization system to further analyze the contribution of these GU-sequences to the process of HIV-1 RNA dimerization. Another interesting possibility is that these GU sequences may actually be NC binding sites, as previously shown for other GU-rich sequences²¹. If so, our mutations might indirectly affect dimerization through some mechanism of reduced NC-mediated activation of dimerization, or dimer maturation. This possibility is currently under investigation, although it is doubtful since the dimers contained within the mutant viruses are as thermostable as that of wild-type (Fig. 2.1F).

The second issue concerns the potential mutual influences between RNA packaging and dimerization. Diminished RNA packaging was repeatedly observed together with impaired dimerization in various mutant viruses that harbored mutated GU-sequences. It is thus reasonable to postulate that the mutated GU-sequences actually caused deficient RNA packaging, which subsequently led to impaired dimerization, possibly due to insufficient amounts of viral RNA within virus particles. If this is the case, then the GU-sequences contribute to dimerization merely in an indirect manner. However, opposing evidence argues against this hypothesis. The first clue comes from the AS45 mutation, in which a second-site mutation, i.e. AS5, restored RNA packaging in the context of the AS4 mutation to wild-type levels, but failed to correct deficient RNA dimerization caused by AS4. These findings are similar to those reported in a recent article in which the authors describe HIV-1 mutants with defects in dimerization but display RNA encapsidation ability that was comparable to wild-type particles²³⁷. Second, compensatory mutations T12I (in p2) and V13G (in NC) increased the Δ GU3 RNA packaging to wild-type levels, but had a limited effect on the diminished levels of RNA dimers associated with mutant viruses. Although the mechanism by which these compensatory mutations contribute to the increase in packaging is unclear at this time, we may gain some insight if we consider the interactions between HIV-1 NC and SL3. Solution structure analysis of NC complexed with SL3 RNA predicts that the V13 residue of NC is actually in contact with the viral RNA⁶⁷. The authors report that nt G⁹ of the SL3 loop interacts with the FI knuckle of NC by binding to a hydrophobic cleft formed by the side chains of V13, F16, I24, and A25. Based on these observations, we propose that the substitution of the bulky methyl groups, associated with the Val residue,

for the hydrogen associated with the Gly residue of our V13G mutation, might reduce the extent of steric interaction in the cleft and allow the RNA to bind more tightly, thus contributing to the increased packaging seen in these mutants. It is also interesting to note that other groups have reported similar results that indicate a role for this region of NC in the encapsidation process. In one study, mutation of position 10 and 11 in NC, adjacent to the V13 mentioned above, decreased genomic RNA packaging levels, indicating the importance of this region for genome recognition and encapsidation⁴⁵. The location of our T12I compensatory mutation in the p2 domain of Gag, and its putative effect on genome packaging is also in agreement with previous findings suggesting a possible role for p2 in RNA encapsidation^{131,229}. In another study, RNA packaging defects, but not dimerization deficits, caused by elimination of a portion of the DIS sequences were corrected by compensatory mutations within Gag²⁴⁸. This latter observation indicates that the role played by DIS RNA in the dimerization process cannot be performed by modified Gag proteins. Conceivably, the failure of the T12I and V13G compensatory mutations to correct deficient RNA dimerization associated with the Δ GU3 deletion suggests a possible direct participation of this GU-rich fragment in RNA dimerization.

The ability of the Δ GU2 mutated viruses to reestablish wild-type sequences may be due to the fact that the Δ GU2 deletion is present only in the U5 region at the 5' end of proviral DNA. Although the 3' U5 region is normally removed at the termination stage of transcription, transcriptional error can occasionally result in retention of wild-type U5 sequences at the 3' end of viral RNA. During reverse transcription, the first strand transfer is usually directed by the binding of full-length minus (-) strand strong-stop cDNA to 3' R sequences²⁵⁰. However, premature reverse transcription that occurs before

nt 110 in U5 (Fig. 2.1) may result in a premature cDNA product being able to mediate the first strand transfer if a U5 is still available at the 3' end of viral RNA. Reverse transcription could then continue from the transferred cDNA to copy the wild-type 3' U5 sequences. Previous studies that employed natural tRNA^{Lys.3} as primer and synthetic viral RNA templates identified a strong pause site around nt 111¹⁶⁰. Such pausing may also occur during *in vivo* reverse transcription and help ΔGU2 recapture lost sequences via this mechanism.

These two GU-rich RNA stretches within U5 each contain a 5'-UGUGUG-3' element with a palindromic feature (Fig. 2.1A). It is postulated that the 5'-UGUGUG-3' sequences can pair with their counterparts on a second RNA strand to mediate dimerization. Although the G:U pairing may be less stable than that of G:C, the existence of two stretches of 5'-UGUGUG-3' within U5 may overcome any diminution in pairing stability. And, furthermore, viral protein factors such as the NC protein may further stabilize these interactions. GU-sequences may act either to trigger dimerization together with the DIS or to stabilize dimers that are already formed. Detailed genetic analysis will be required to sort out these possibilities.

We also confirmed that a 5'-AAGCUU-3' palindrome, located within the loop of the 5'-poly(A) hairpin, did not contribute to RNA dimerization¹⁵³. This failure may be due to certain structural features of this palindrome. Compared to the 5'-GCGCGC-3' palindrome within the DIS, 5'-AAGCUU-3' contains only one G and one C. Since G and C form a stronger base pair than that formed by A and U, it follows that interactions between RNA strands mediated by the 5'-AAGCUU-3' would be weaker than those mediated by the 5'-GCGCGC-3'. Therefore, the former palindrome may not be able to

initiate RNA interactions that are strong enough to trigger RNA dimerization. Moreover, structural studies reveal that the “A” residues within the DIS loop significantly contribute to the stability of RNA dimers, in addition to the role played by the 5'-GCGCGC-3' palindrome^{81,201,202}. Hence, specific structural features distinct from the palindrome sequence are needed for an RNA motif to serve as a dimerization initiation site.

In conclusion, we propose that the GU-sequences studied above are involved in the process of HIV-1 RNA dimerization, although this region does not necessarily constitute a new 'dimerization initiation site'. Rather, these sequences may function to stabilize the RNA dimer after the known DIS has started the process. This notion is compatible with previous reports suggesting that the dimerized RNA are bound together by direct RNA-RNA linkages along the entire genome²¹¹. In other dimer linkages, such as those seen in Harvey Sarcoma virus^{85,169}, and rat retrotransposon VL30^{266,267} the involvement of multiple regions in the dimerization process has been reported. It has also been reported that the Moloney murine leukemia virus contains dual DIS elements, both of which participate in the initiation of dimerization.^{170,210} The authors further hypothesize that the idea of multiple adjacent regions being involved in dimer linkages may prove to be a common theme among retroviruses. In line with this notion, we propose that the dimerization of the HIV-1 RNA genome *in vivo* may also involve multiple RNA elements.

CHAPTER 3

SEQUENCES DOWNSTREAM OF THE 5' SPLICE DONOR SITE ARE REQUIRED FOR BOTH PACKAGING AND DIMERIZATION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 RNA

This chapter was adapted from an article that appeared in *Journal of Virology*, 2003, 77(1):84-96. The authors of this paper were **R.S. Russell**, J. Hu, V. Bériault, A.J. Mouland, M. Laughrea, L. Kleiman, M.A. Wainberg, and C. Liang. Over 80% of the data presented in this chapter were from experiments performed by myself under the supervision of Drs. Liang and Wainberg. V. Bériault, from Dr. Mouland's lab, performed the RNase protection assays presented herein. J. Hu and Dr. Liang constructed some of the plasmids and performed all of the virus replication analyses. Dr. Kleiman critiqued the manuscript and offered valuable suggestions for revision. Dr. Laughrea critiqued an earlier version of the manuscript.

3.1 Preface to Chapter 3

The data presented in Ch. 2 confirmed our hypothesis that sequences other than the known DIS can play a role in the HIV-1 RNA dimerization process. So, at this point, the DIS and regions in the Poly(A) and U5-PBS stems had all been shown to contribute to RNA dimerization, but all of these regions are also known to represent important packaging signals^{51,62,189} (and Ch. 1). Since numerous reports had already proposed a link between dimerization and packaging^{25,91,151,214}, we wondered whether other known packaging signals, such as SL3²⁷, might also be involved in the dimerization process. In support of this hypothesis, one study had shown that an antisense oligonucleotide targeting SL3 and a downstream GA-rich region could inhibit dimerization of HIV-1 RNA fragments *in vitro*²⁸⁶. And, an earlier paper had shown that dimerized HIV-1 RNA fragments representing nt 311-415, downstream of the DIS and the major splice donor, were more heat stable than dimers formed between fragments representing nt 1-311, which contained the DIS¹⁸⁴. On the basis of these reports, we set out to determine whether sequences downstream of the 5' splice donor are required for dimerization, in addition to their known roles in packaging.

3.2 Abstract

Two copies of human immunodeficiency virus type 1 (HIV-1) RNA are incorporated into each virus particle and are further converted to a stable dimer as the virus particle matures. Several RNA segments that flank the 5' splice donor (SD) site [at nucleotide (nt) 289] have been shown to act as packaging signals. Among these, an RNA stem-loop 1 (SL1) [nt 243 to 277] can trigger RNA dimerization through a "kissing-loop" mechanism and thus is termed the dimerization initiation site (DIS). However, it is unknown whether other packaging signals are also needed for dimerization. To pursue this subject, we have mutated stem-loop 3 (SL3) [nt 312 to 325], a GA-rich region [nt 325 to 336], and two G-rich repeats [nt 363 to 367 and nt 405 to 409] in proviral DNA and assessed their effects on RNA dimerization by performing native Northern blots. Our results show that the structure, but not specific RNA sequence, of SL3 is needed not only for efficient viral RNA packaging but also for dimerization. Mutations of the GA-rich sequence severely diminished viral RNA dimerization as well as packaging; combination of mutations in both SL3 and the GA-rich region led to further decreases, implicating independent roles for each of these two RNA motifs. Compensation studies further demonstrated that the RNA packaging and dimerization activity of the GA-rich sequence may not depend on a putative interaction between this region and a CU-repeat sequence at nt 227 to 233. In contrast, substitutions in the two G-rich sequences did not cause any diminution of viral RNA packaging or dimerization. We conclude that both the SL3 motif and GA-rich RNA sequences, located downstream of the 5' SD site, are required for efficient RNA packaging and dimerization.

3.3 Introduction

Human immunodeficiency virus type 1 (HIV-1) contains a diploid RNA genome that is noncovalently associated in dimer form at its 5' ends in a parallel orientation. These attached viral RNA regions were first described as the dimer linkage structure (DLS) in monkey sarcoma virus¹⁴⁴ and this term has been subsequently used to describe dimerization in other retroviruses. Two models have been proposed to illustrate molecular interactions that constitute the HIV-1 DLS. The first involves a tetra-stranded RNA structure, termed a G-tetrad, that is formed by G-rich RNA sequences¹⁵⁴. This structure has been implicated in maintaining the integrity of chromosome telomeres in which stretches of G-rich nucleotide sequences are present¹⁰⁷. G-rich RNA regions were also identified at the 5' end of HIV-1 RNA downstream of the major splice donor (SD) site. It was therefore hypothesized that formation of G-tetrad structures may contribute to the maintenance of RNA dimers; this notion has been supported by studies performed with synthetic viral RNA fragments, yet has not been extensively tested in the context of the full-length viral RNA genome^{12,108,154,262}.

The second model involves a "kissing-loop" mechanism, and is derived from the observation that the stem-loop 1 (SL1) RNA segment, located upstream of the 5' SD site, was able to spontaneously form dimers under appropriate buffer conditions^{147,184}. This reaction is believed to be initiated by the SL1 loop palindrome sequence (e.g. 5'-GCGCGC-3' in HXB2D) via the formation of regular "Watson-Crick" base-pairs; SL1 was thus termed the dimerization initiation site (DIS)^{205,215,251}. Subsequent studies demonstrated that both the palindrome and the stem are essential for the dimerization

activity of SL1^{49,50,149,205}. However, substantial amounts of dimerized RNA were detected in mutant viruses containing altered loop sequences or a disrupted stem structure of SL1^{25,239}; thus, SL1 may constitute only part of the DLS structure.

The DLS of HIV-1 overlaps RNA packaging signals. Stem loop 3 (SL3) and its flanking RNA sequences, together with SL1, represent major packaging signals²⁷. In addition to the well-documented roles of SL1 in RNA dimerization, cell-free assays using synthetic RNA molecules showed that an antisense nucleotide oligomer that binds to SL3 and a downstream GA-rich RNA region was able to inhibit RNA dimerization²⁸⁶. In the present study, we have addressed this subject *in vivo* by mutating relevant RNA sequences in proviral DNA and analyzing virion-derived mutant RNA by native Northern blotting. Our data shows that both SL3 and the GA-rich RNA segments are required for both RNA packaging and dimerization.

3.4 Materials and Methods

3.4.1 Plasmid construction

An infectious HIV-1 cDNA clone BH10 was employed to generate the constructs described below and all mutations were introduced by PCR-based strategies using the *Pfu* enzyme (Stratagene, La Jolla, CA). MD1, MD2, and MD3 are deletion mutations that were engineered by PCR using primer pair pBssH-S (5'-CTGAAGCGCGCACGGCAAGAGG-3' [nt 252 to 273])/pD1 (5'-CCATCTCTCTCCTTCTAGCGCTAGTCAAATTTTTGGC-3' [nt 339 to 298]), pBssH-S/pD2 (5'-GCTCT

CGCACCCATCTCTTTCTAGCCTCCGC-3' [nt 349 to 315]), and pBssH-S/pD3 (5'-GCTCTCGCACCCATCTCTTTCTAGCGCTAGTCAAAATTTTTGGC-3' [nt 349 to 298]), respectively (Fig. 3.1A, B). The PCR products were then used as primers together with pSph-A (5'-GGCCCTGCATGCACTGGATGC-3' [nt 1000 to 980]) in a second round of PCR. Final PCR products were digested with the restriction enzymes *Bss*HII and *Sph*I and inserted into BH10. Constructs MS1 to MS6, MG1, MG2 and MG12 contain substitutions (Fig. 3.1A, B) and were generated with similar strategies using primers pS1 (5'-CCATCTCTCTCCTTCTAGCTAGCGCTAGTCAAAATTTTTGGC-3' [nt 339 to 298]), pS2 (5'-GCTCTCGCACCCATCTCTGATGTTCTAGCCTCCGC-3' [nt 349 to 315]), pS3 (5'-GCTCTCGCACCCATCTCTGATGTTCTAGCTAGCGCTAGTCAAAATTTTTGGC-3' [nt 349 to 298]), pS4 (5'-CTCTCTCCTTCTAGCCTCCTTGTCTCAAATTTTTGGCGTAC-3' [nt 335 to 294]), pS5 (5'-GCACCCATCTCTCTCCTTGACAACTCCTTGTCTCAAATTTTTGGCGTAC-3' [nt 349 to 294]), pS6 (5'-CGCTCTCGCACCCATATGTCTCCTTCTAGCCTC-3' [nt 350 to 318]), pG1 (5'-CCATCGATCTAATTCACCGCCGCTTAATACTGACGC-3' [nt 383 to 348]), pG2 (5'-TAATTTATATTTTTTCTTACCGCCTGGCCTTAACCG-3' [nt 428 to 393]), respectively (Fig. 3.1A, B). BH-GA contains a substitution of a 5'-CTCTCTC-3' [nt 227 to 233] by the sequence 5'-AGAG-3' (Fig. 3.6A) and was generated as described previously¹⁵⁹. Construct MS7 was designed to contain a compensatory mutation for the MS6 mutant, and was generated using primer MS6C-A (5'-CCTGCGTCGACATAGCTCCTCTGG-3' [nt 241 to 218]) in combination with primer U5-1S (5'-GAGATCCCTCAGACCCTTTTAG-3' [nt 137 to 158]), using the MS6 construct as a template. The PCR product was then used as a primer together with pSph-A in a second round of

PCR and the final PCR products were digested with *NarI* and *SphI* and inserted into BH10.

To generate an antisense riboprobe to be used in RNase protection assays (RPA), a 486 bp fragment was amplified from the BH10 proviral DNA sequence using primer pair RPA-S (5'-CagggcccGAGA GCTGCATCCGGAG-3' [nt -164 to -140]), which was modified to contain an *ApaI* restriction enzyme site (shown in lower case), and RPA-A (5'-CCTCCGgaattcAAAATTTTGGC G-3' [nt 321 to 297]), which was modified to contain an *EcoRI* restriction enzyme site (shown in lower case). The resulting PCR product was digested with *ApaI* and *EcoRI* and inserted into the pBluescript II KS⁺ cloning vector (Stratagene) that had been cut with the same enzymes to generate construct RPA1.

3.4.2 Cell culture, transfection, and infection

COS-7, MT-2, and Jurkat cells were grown in Dulbecco's modified Eagle's medium (DMEM) and RPMI 1640 medium, respectively, supplemented with 10% fetal calf serum. Transfection of COS-7 cells was performed using Lipofectamine (Invitrogen, Burlington, Ontario, Canada). Quantities of progeny viruses were determined by measuring levels of p24 antigen using enzyme-linked immunosorption assays (ELISA) (Vironostika HIV-1 Antigen Microelisa System; Organon Teknika Corporation, Durham, NC).

MT-2 or Jurkat cells (5×10^5) were incubated with aliquots of viruses equivalent to 5 ng of p24 in 2 ml of medium at 37°C for 2 hours. The cells were then washed twice and

maintained in 10 ml of medium. Culture fluids were collected at various times to determine levels of reverse transcriptase (RT) activity.

3.4.3 Native Northern blotting

Progeny viruses generated by transfected COS-7 cells were first clarified by centrifugation in a Beckman GR-6S centrifuge at 3,000 rpm for 30 min at 4°C and then pelleted through a 20% sucrose cushion by ultracentrifugation in a Beckman XL-80 ultracentrifuge using a SW41 rotor at 40,000 rpm for 1 hour at 4°C. Virus pellets were suspended in 300 µl of TN buffer; a 2 µl portion was removed for p24 determination and the remaining viruses were treated with virus lysis buffer (50 mM Tris-HCl [pH 7.4], 10 mM EDTA, 1% SDS, 100 mM NaCl, 50 µg of yeast tRNA/ml, 100 µg proteinase K/ml) for 20 min at 37°C. Samples were then extracted twice with phenol:chloroform:isoamylalcohol (25:24:1) and once with chloroform. Viral RNA was precipitated in 2.5 volumes of 95% ethanol. RNA pellets were washed with 70% ethanol and dissolved in TE buffer. An amount of viral RNA equivalent to 150 ng of HIV-1 p24 was fractionated on 0.9% native agarose gels in 1xTBE buffer at 100 volts for 4 hours at 4°C and analyzed by Northern blots²⁴⁰ using a [³²P]-α-dCTP (ICN, Irvine, CA) labeled 2 kb HIV-1 DNA fragment [nt 1 to 2000] as a probe. Bands were visualized by autoradiography, and quantified by digital image analysis using the NIH Image program. To determine the thermostability of viral RNA dimers, RNA samples were incubated at various temperatures (e.g. 25°C, 40°C, 45°C, 50°C and 55°C) for 10 min in a buffer containing 100 mM NaCl before being separated on native agarose gels.

3.4.4 RNase protection assays

Preparation of riboprobes and RPA experiments were performed based on previously described protocols^{48,50}. Briefly, radiolabelled probes were *in vitro* transcribed from *BspE1*-linearized RPA1 plasmid using T7 RNA polymerase in the presence of [³²P]- α -UTP (ICN, Irvine, CA). RNA was isolated from viruses as described above. Cytoplasmic RNA was isolated from transfected COS-7 cells by resuspending the cell pellets in 400 μ l of cell lysis buffer (10 mM Tris-HCl [pH 7.4], 140 mM NaCl, 1.5 mM MgCl₂, 0.5% Nonidet P-40, 1mM dithiothreitol, 100 U of RNase inhibitor (Invitrogen)), then incubated on ice for 5 min and centrifuged at 13,000 rpm for 10 min at 4°C. Supernatants were treated with 0.2% SDS and 125 μ g/ml of proteinase K at 37°C for 20 min, then subjected to phenol:chloroform extraction and ethanol precipitation as described above. Quantities of total cytoplasmic RNA were determined by spectrophotometry at 260 nm. Amounts of virion RNA equivalent to 25 ng of p24 capsid antigen, or 250 ng of cytoplasmic RNA, were treated with 10 U of DNase I (Invitrogen) for 30 min at 37°C to remove any plasmid contamination, and then subjected to phenol:chloroform extraction and ethanol precipitation, before analysis with the RPA II kit (Ambion Inc., Austin, TX). Briefly, the RNA was incubated at 42°C overnight with an excess of labeled riboprobe (10⁵ cpm), followed by digestion with RNases specific for single-strand RNA. Protected fragments were separated on 5% polyacrylamide-8M urea gels, visualized by autoradiography, and quantified by digital image analysis using the NIH Image program.

3.5 Results

3.5.1 SL3 is needed for both viral RNA packaging and dimerization

SL3 is known to be an important HIV-1 RNA packaging signal^{5,27,48,50,112,168,188}. To further assess the potential role of SL3 in viral RNA dimerization, we have generated a series of mutations. Loop sequence 5'-GGAG-3' [nt 317 to 320] was either deleted to create construct MD1 or replaced by a 5'-GCTA-3' sequence to generate construct MS1 (Fig. 3.1A, B); the latter sequence was designed so as not to disrupt the secondary structure of SL3 (Fig. 3.1B). Next, we destabilized the base-pairs within the stem by changing the sequence 5'-CTAGC-3' [nt 312 to 316] to 5'-GACAA-3' in construct MS4 (Fig. 3.1A, B). To restore the stem structure, the right portion of the stem sequence 5'-GCTAG-3' [nt 321 to 325] was changed to 5'-TTGTC-3' in construct MS5 (Fig. 3.1A, B).

These various DNA clones were transfected into COS-7 cells and the progeny viruses thus generated were used to infect MT-2 and Jurkat cells. Both of these cell lines are CD4⁺ and thus are permissive for HIV-1 infection. The results of Fig. 3.2A show that the MD1 and MS4 mutations caused delays in virus growth in MT-2 cells; in contrast, the MS1 and MS5 mutants grew almost as well as did wild-type virus. Differences in levels of infectivity between mutant and wild-type viruses were further verified in Jurkat cells (Fig. 3.2B). Since HIV-1 grew with slower kinetics in Jurkat cells than in MT-2 cells, as observed in the growth curves of wild-type BH10, replication differences between

Figure 3.1. Description of the mutations generated in the HIV-1 RNA sequences downstream of the 5' major splice donor (SD) site. (A) Mutations MG1, MG2, and MG12 are shown at the top; these altered the G-rich sequences without changing relevant amino acid sequences in MA. Mutations in SL3 and the GA-rich region are diagrammed at the bottom. Deleted nucleotides are indicated by a hyphen "-". Numbers of nucleotides refer to the first nt of the R region. A number of structural domains in the leader region are shown: these include TAR, poly(A), U5-PBS, SL1, SL2, and SL3. (B) Schematic representation of all mutants listed in (A). MS1 changed the loop sequence without disrupting the SL3 stem; MS4 replaced the left portion of the stem sequences and thereby destabilized stem base-pairing; MS5 restored the stem in SL3 by insertion of a compensatory mutation; MD1 represents a deletion of the SL3 loop; MD2 represents a deletion of the GA-rich sequence just adjacent to SL3; MD3 is a combination of the MD1 and MD2 deletions; MS2 contains a substitution of the GA-rich sequence just adjacent to SL3; MS3 is a combination of the MS1 and MS2 substitutions; MS6 contains substitutions of the two Gs at nt positions 332 and 334. RNA structures were predicted by the M-Fold program^{186,291}.

Figure 3.1. Description of the mutations generated in the HIV-1 RNA sequences downstream of the 5' major splice donor (SD) site.

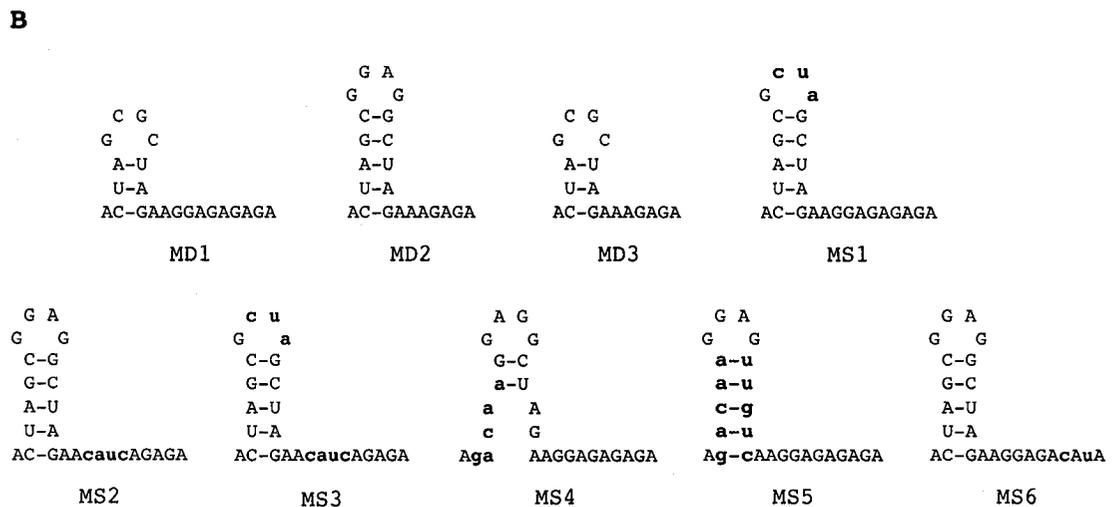
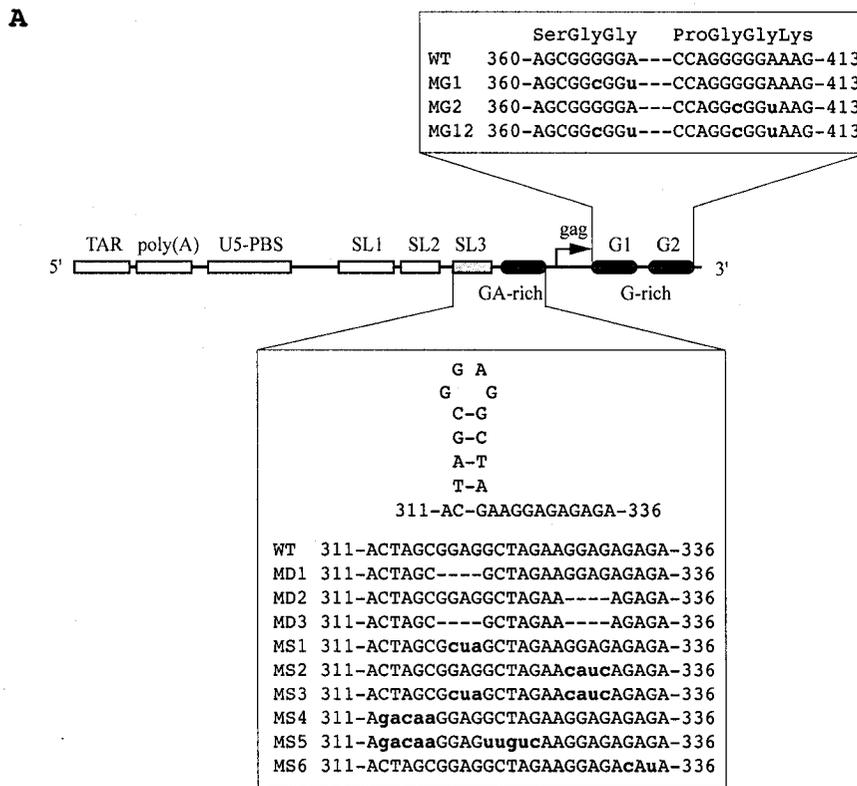
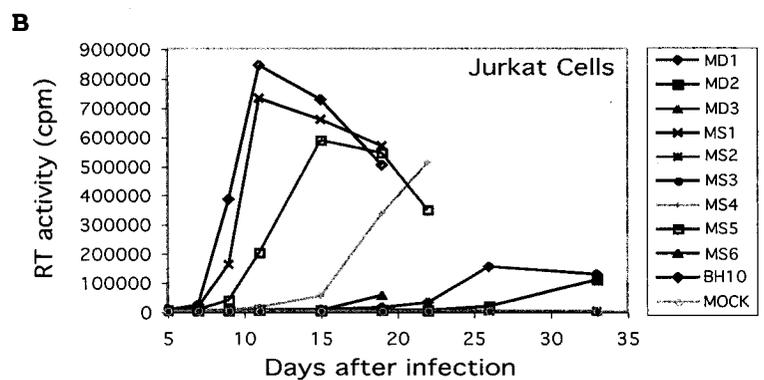
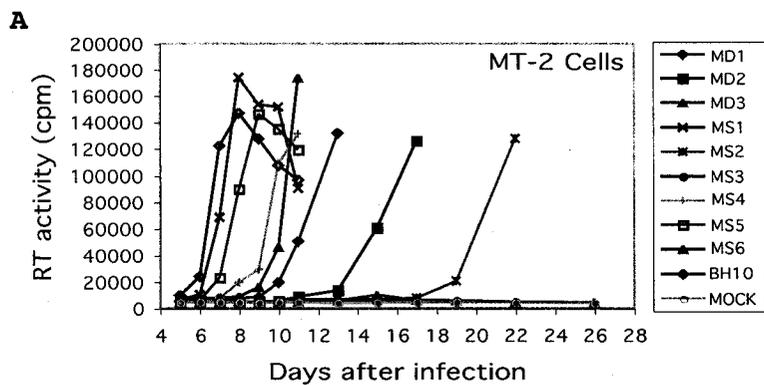


Figure 3.2. Infectiousness of mutant viruses MD1 to MD3 and MS1 to MS6 in permissive cell lines. MT-2 cells (A) and Jurkat cells (B) were infected by progeny viruses containing 5 ng of p24 antigen. Virus growth was monitored by measuring RT activity in culture fluids at various times. Mock infection represents exposure of cells to heat-inactivated wild-type viruses.

Figure 3.2. Infectiousness of mutant viruses MD1 to MD3 and MS1 to MS6 in permissive cell lines.



wild-type versus mutant viruses were more pronounced in the Jurkat system. The extent to which these four viruses were able to replicate is MS1>MS5>MS4>MD1.

Viral RNA dimerization and packaging studies were performed on virus particles derived from transfection of nonpermissive COS-7 cells. We used the latter nonpermissive cells so as not to allow viral mutants to undergo reversions that might regularly occur over multiple rounds of infection during culture in permissive cells. In order to determine the effects of these mutations on viral RNA dimerization and packaging, we analyzed virion-derived RNA on native Northern blots (Fig. 3.3A, B). In addition, viral RNA packaging was further investigated by RPA analysis (Fig. 3.3C-F). Since all of our mutations were situated in or just downstream of SL3, we designed an RPA probe that was complementary to sequences spanning the region from within U3 to just before SL3. This particular probe generates the same digestion pattern for wild-type HIV-1 RNA and each of our mutants, as depicted in Fig. 3.3C.

The results of these analyses demonstrate that the majority of wild-type viral RNA was present as a dimer (Fig. 3.3A, lane 10). In the case of construct MD1, overall levels of viral RNA were reduced compared to wild-type BH10 (Fig. 3.3A, lane 1 vs 11; and Fig. 3.3D, F) and more than 50% of the viral RNA migrated as monomers (Fig. 3.3A, lane 1; and Fig. 3.3B). However, substitution of the SL3 loop sequence in MS1 did not affect either RNA dimerization or packaging (Fig. 3.3A, D, lane 4; Fig. 3.3B, F), suggesting that the loop sequence itself is not required for either of these activities.

The negative effect of the MD1 mutation on RNA dimerization and packaging may be due to a destabilization of SL3 secondary structure. To address this issue, the base-pairing of the stem structure was disrupted by the MS4 mutation (Fig. 3.1B), and the

results show that this led to a decrease in overall levels of both viral RNA and dimeric RNA (Fig. 3.3A, D, lane 7; Fig. 3.3B, F). Restoration of the base-pairing in the stem structure by a compensatory mutation in MS5 (Fig. 3.1B) corrected these defects (Fig. 3.3A, D, lane 8; Fig. 3.3B, F). Therefore, SL3 not only acts as a packaging signal but also affects RNA dimerization, and SL3 secondary structure must play a major role in each of these activities.

The wild-type and mutant proviral constructs produced relatively similar amounts of viral RNA within the cytoplasm (Fig. 3.3E), with the exception of MS2 (lane 5), which may have been due to a difference in transfection efficiency. Analysis of the viral RNA expressed within the cells revealed that all mutants tested generated approximately 2-fold ratios of genomic to spliced viral RNA transcripts. As for viral RNA encapsidated into virions, wild-type HIV-1 packaged genomic RNA almost exclusively, while all mutants encapsidated detectable levels of spliced RNA. Therefore, SL3 and the downstream region play important roles in the specific packaging of full-length genomic RNA, with concomitant exclusion of the spliced. This observation is consistent with a previous report on this subject⁵⁰.

We then compared the thermostability of the mutant versus wild-type RNA dimers by incubating RNA samples at different temperatures before separation on native agarose gels. The results show that wild-type viral RNA dimers began to dissociate at 50°C and were completely converted to monomers at 55°C (Fig. 3.4). Similar results were obtained with RNA samples containing any of the MD1, MS1, MS4, or MS5 mutations (Fig. 3.4). Apparently, RNA dimers within these mutant virus particles displayed similar thermostability as did wild-type RNA dimers. Our mutant viruses can be ranked as

Figure 3.3. Effects of various mutations on viral RNA dimerization and packaging. (A) Viral RNA was prepared from mutant viruses MD1 to MD3 and MS1 to MS6 (lanes 1-9) and wild-type virus BH10 (lane 10), equivalent to 150 ng of p24 antigen, and fractionated on native agarose gels, followed by Northern blot analysis. Dimers and monomers are indicated on the left side of the gels. Results are shown from one representative gel. (B) Band intensities of dimer and monomer signals were measured using the NIH Image program, and relative levels for each construct were plotted. The results represent pooled data from three Northern blots using virion-derived RNA from three independent transfections of each mutant. (C) Schematic illustration of the RPA system used to quantify viral RNA, based on the strategy used by Clever and Parslow⁵⁰. Shown are the 5' LTR sequences including U3, R, U5, and stem loops 1-4. Below are the probe (569 nt) and the protected fragments that would be generated from the various viral RNA and DNA sequences; these include DNA (486 nt); full-length genomic RNA (310 nt; Fig. 3.3D, upper band), spliced RNA (288 nt; Fig. 3.3D, middle band), and 3' LTR sequence (243 nt; Fig. 3.3D, lower band) which serves as an internal control for total viral RNA.

Figure 3.3. Effects of various mutations on viral RNA dimerization and packaging.

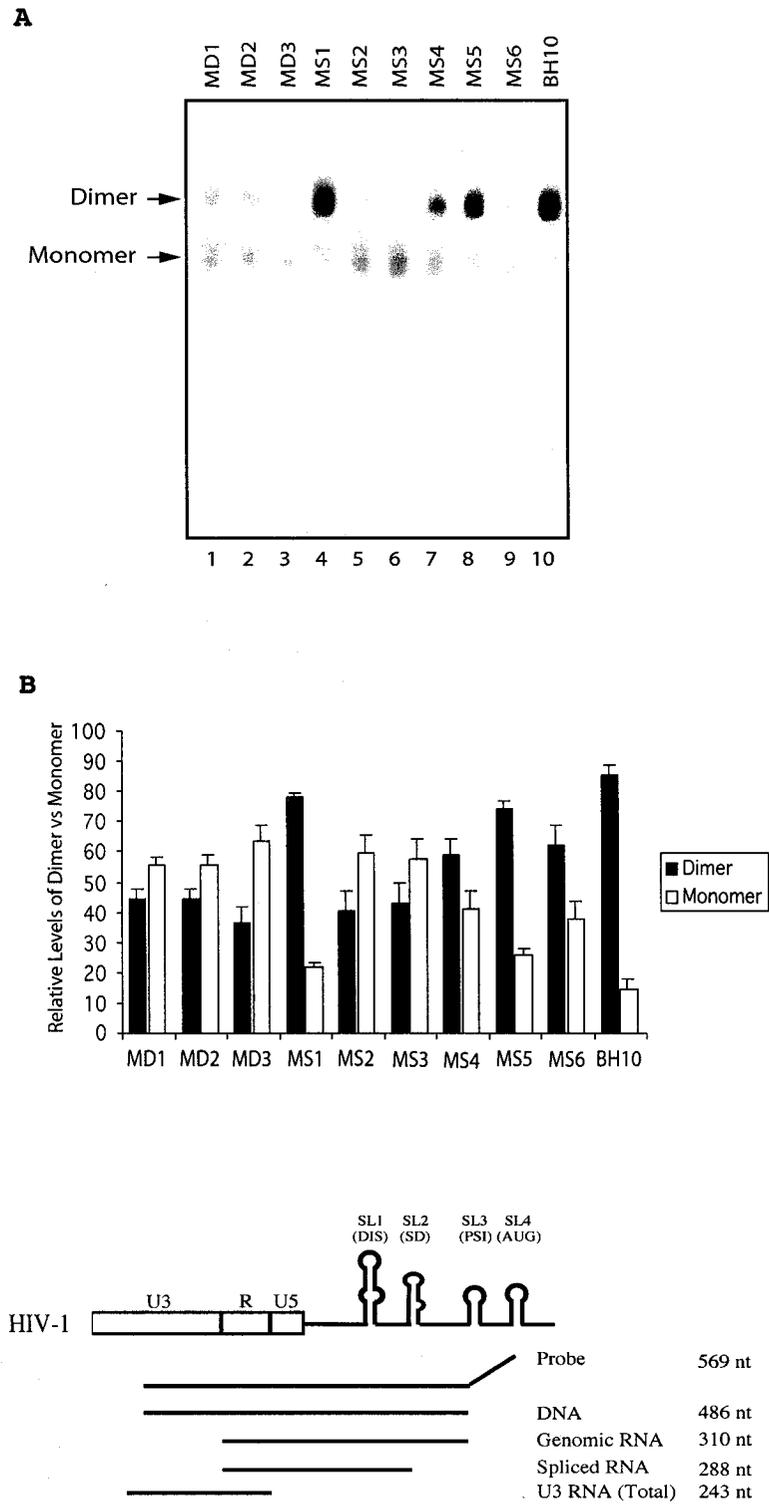
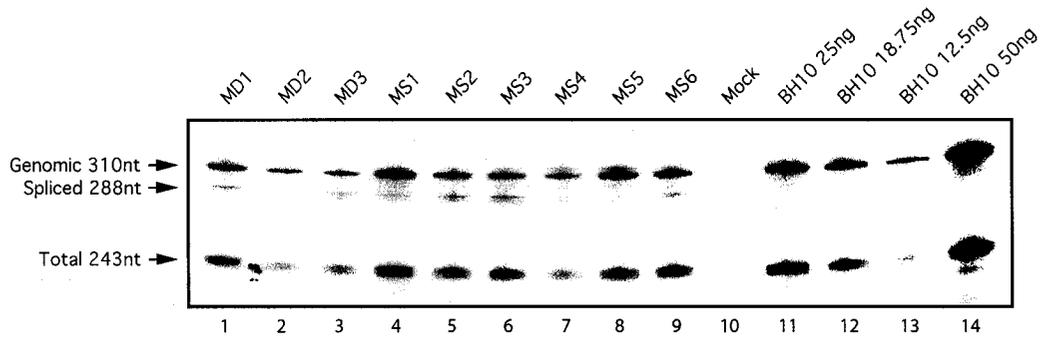


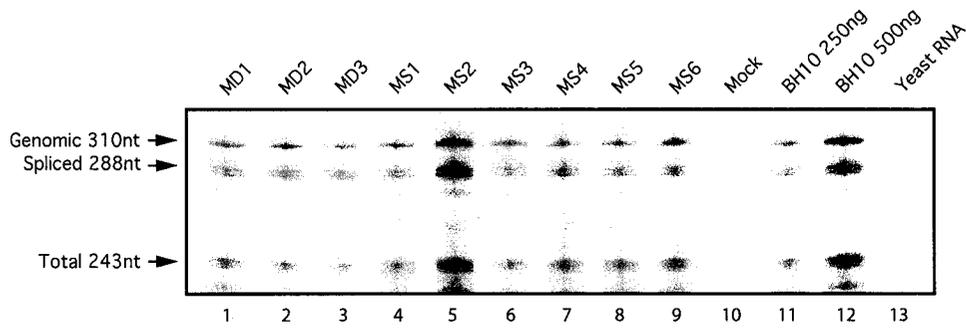
Figure 3.3. (cont.) Effects of various mutations on viral RNA dimerization and packaging. (D) RPA performed on virion-derived RNA from mutant MD1 to MD3 and MS1 to MS6 (lanes 1-9) and wild-type BH10 (lane 11) viruses. An amount of viral RNA equivalent to 25 ng p24 capsid Ag was annealed to 10^5 cpm of radiolabelled riboprobe, digested with RNases specific for single-strand RNA, and protected fragments were separated by 5% denaturing-PAGE. Transfection of the pSP72 cloning vector served as a mock experiment (lane 10). A dilution series of wild-type RNA was analyzed to show the linear range of the assay (25, 18.75, and 12.5 ng p24 in lanes 11-13, respectively). Wild-type RNA equivalent to 50 ng of p24 was also analyzed to demonstrate that the assay was not saturated at 25 ng p24 (lane 14). One representative gel is shown from two independent experiments. (E) RPA performed on cytoplasmic RNA from transfections of mutants MD1 to MD3 and MS1 to MS6 (lanes 1-9) and wild-type BH10 (lane 11) DNA constructs as described in (D) using 250 ng of RNA. A 2X sample (500 ng) of cytoplasmic RNA from transfection of wild type BH10 was analyzed to demonstrate that the assay was not saturated at 250 ng RNA (lane 12). A sample containing 10 μ g of yeast tRNA was used as a negative control (lane 13). One representative gel is shown from two independent experiments. (F) Packaging levels of mutant viral RNA expressed as a percentage of wild-type virus BH10 (arbitrarily set at 100%). The bar graph represents data pooled from three Northern blots and two RPA gels using RNA from five independent transfections of each mutant.

Figure 3.3. (cont.) Effects of various mutations on viral RNA dimerization and packaging.

D



E



F

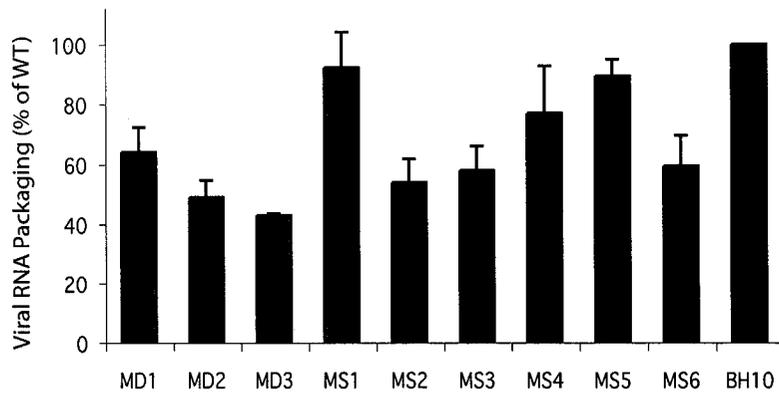
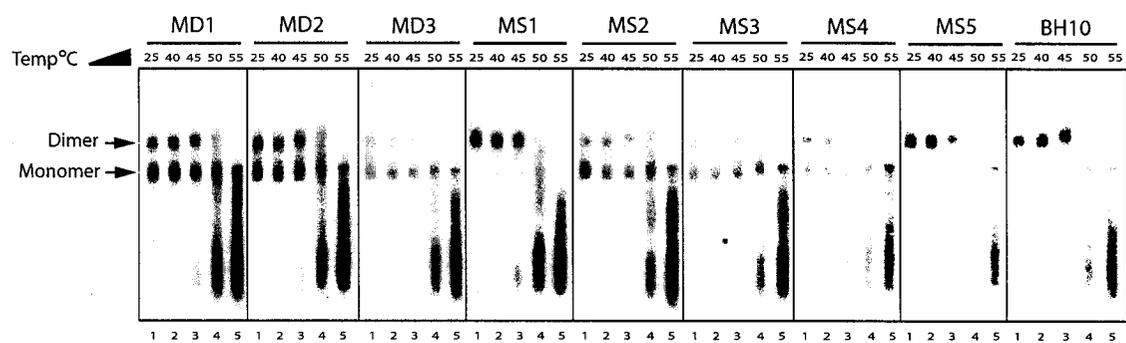


Figure 3.4. Measurement of RNA dimer stability. Virion-derived RNA was treated at different temperatures before electrophoresis on native agarose gels. Lanes 1-5 represent incubations performed at 25°C, 40°C, 45°C, 50°C and 55°C, respectively.

Figure 3.4. Measurement of RNA dimer stability.



MS1>MS5>MS4>MD1 on the basis of overall RNA levels, dimer levels within virions, and infectivity (summarized in Table 3.1).

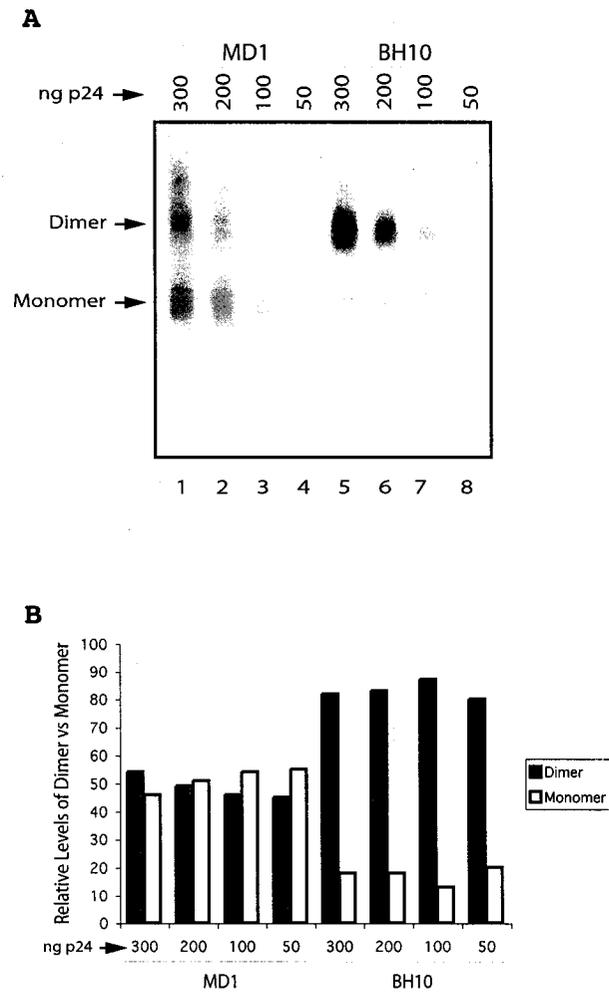
Since we were interested in studying the effects of our mutations on viral RNA dimerization as well as packaging, various amounts of viral RNA were loaded onto native Northern blots based on equivalent amounts of virus. However, some of our mutants displayed packaging defects, thus lower amounts of viral RNA from these mutants were assessed by Northern blot than for wild-type virus. In order to test the possibility that the dimerization defects seen in our mutants might indirectly result from a lower RNA concentration, we performed a Northern blot on RNA from one mutant which showed a packaging defect (i.e. MD1), along with wild-type BH10 RNA.

We loaded a series of diluted viral RNA samples equivalent to 300, 200, 100 and 50 ng of p24. The results show that dimer:monomer ratios for neither MD1 nor wild-type viral RNA were affected by decreasing the amount of RNA loaded in the well; approximately 85-90% of wild-type RNA was present as dimers regardless of the quantity of viral RNA employed (Fig. 3.5A, B). A comparison of lanes 2 and 7 (Fig. 3.5A) also shows that an increase in the amount of mutant RNA did not enhance RNA dimer levels. Thus, the dimerization defects seen on our gels are not an indirect result of RNA concentrations.

3.5.2 Mutations in GA-rich RNA sequences at nt 325 to 336 significantly reduce levels of both viral RNA dimerization and packaging

Figure 3.5. Effects of RNA concentration on viral RNA dimerization. (A) A series of RNA samples containing 300, 200, 100 and 50 ng of p24 equivalent of the MD1 mutant (lanes 1-4) and wild-type BH10 (lanes 5-8) virus RNA were analyzed by native Northern blotting. Dimers and monomers are indicated on the left side of the gels. (B) Band intensities of dimer and monomer signals were measured using the NIH Image program, and relative levels for each lane were plotted.

Figure 3.5. Effects of RNA concentration on viral RNA dimerization.



Previous findings have shown that a GA-rich RNA sequence downstream of the SL3 structure participates in viral RNA packaging^{5,47,168}. To assess the role of this RNA sequence in viral RNA dimerization, we either deleted the four nt, i.e. 5'-GGAG-3' [nt327 to 330], in construct MD2, or alternatively, changed them to 5'-CATC-3' in construct MS2 (Fig. 3.1A, B). The results of infection assays showed that both mutations markedly decreased viral replication capacity in each of MT-2 and Jurkat cells, with MS2 displaying the greater effect of the two (Fig. 3.2A, B); this is potentially attributable to the fact that MS2, unlike MD2, inserted C and T bases as well as removing G's and A's, thus diminishing overall GA content more effectively (Fig. 3.1A, B). Both MD2 and MS2 reduced viral infectivity to lower levels than did either the MD1 or MS4 mutations in SL3 (Fig. 3.2A, B).

We next assessed viral RNA packaging and dimerization in the MD2 and MS2 mutant viruses and found that both displayed significant reductions in levels of viral genomic RNA (Fig. 3.3D, lanes 2 and 5; and Fig. 3.3F). Over 50% of the viral RNA present in MD2 viruses existed as monomers as did 60% of MS2 viral RNA (Fig. 3.3A, lanes 2 and 5; and Fig. 3.3B). When RNA samples were heated at different temperatures before analysis on native agarose gels, mutant RNA dimers of MD2 and MS2 dissociated at approximately 50°C as did wild-type RNA (Fig. 3.4).

To further evaluate the role of the GA-rich sequence [nt 325 to 336] in RNA dimerization, two Gs at nt positions 332 and 334 were substituted in construct MS6 (Fig. 3.1A, B). MS6 resulted in diminished viral infectiousness in both MT-2 and Jurkat cells, (Fig. 3.2A, B) as well as significant reductions in RNA dimerization (Fig. 3.3A, lane 9; and Fig. 3.3B) and packaging (Fig. 3.3D, lane 9; and Fig. 3.3F) as shown by native

Northern blotting and RPA, respectively. The results of these experiments are summarized in Table 3.1. Thus, the GA-rich sequence at nt 325 to 336 is required for efficient viral RNA dimerization and packaging.

3.5.3 Additive effects of SL3 and GA-rich sequences on viral RNA packaging and dimerization

We next mutated both SL3 and the GA-rich sequences by either combination of the MD1 and MD2 deletions in construct MD3 or by recombination of the MS1 and MS2 mutations in construct MS3 (Fig. 3.1A, B). The MD3 deletion caused a further decrease in viral infectivity in comparison with either MD1 or MD2 (Fig. 3.2A, B); similar observations were made with the MS3 mutant virus (Fig. 3.2A, B). When virion-derived RNA samples were examined by native Northern blots, both the proportions of RNA dimers as well as overall levels of viral RNA in the MD3 mutant were shown to be lower than those observed in either MD1 or MD2 (Fig. 3.3A, D, lane 3 vs 1 and 2; and Fig. 3.3B, F). Thus, both SL3 and the GA-rich region must independently contribute to viral RNA dimerization and packaging. Combination of both the MS1 and MS2 substitutions in MS3 caused reductions in RNA dimerization and packaging similar to those observed in MS2 (Fig. 3.3A, D, lane 6 vs 5; and Fig. 3.3B, F), presumably because MS1 did not affect RNA dimerization or packaging (Fig. 3.3A, D, lane 4; and Fig. 3.3B, F) (see Table 3.1 for summary). Viral RNA dimers from MD3 and MS3 viruses dissociated at approximately 50°C, in similar manner as wild-type RNA dimers (Fig. 3.4).

3.5.4 Mutation of the CU-repeats at nt 227 to 233 results in diminished levels of RNA dimerization and packaging

Our data suggest that secondary structure, rather than specific SL3 RNA sequences, is required for dimerization and packaging. Computer modeling studies have also predicted that the GA-rich region binds to a segment of CU-repeats at nt 227 to 233 (Fig. 3.6A)¹¹¹, and that the SL1, SL2, and SL3 RNA motifs exist as independent domains (referred to as stems II, III, and IV in ref. 14). To test whether the defective RNA packaging and dimerization that was caused by mutations in the GA-rich region had resulted from destabilization of RNA-RNA interactions in this region, we replaced the CU-repeats at this site by the sequence 5'-AGAG-3' to generate construct BH-GA (Fig. 3.6A)¹⁵⁹. The results of native Northern blots showed that BH-GA contained reduced levels of viral RNA (Fig. 3.6D), as well as a diminished proportion of dimerized RNA (Fig. 3.6B, lane 1; and Fig. 3.6C), suggesting that RNA-RNA interactions between the CU-repeats [nt 227 to 233] and the GA-rich sequences [nt 325 to 336] might be important for both viral RNA packaging and dimerization.

To further understand the potential interactions between the CU-repeats [nt 227 to 233] and the GA-rich sequences [nt 325 to 336] and the importance of their interactions in viral RNA packaging and dimerization, we analyzed the local secondary structure using the M-fold algorithm^{186,291}. We found that both the BH-GA and the MS6 mutations disrupted predicted base-pairing between these sequences, but we were unable to design satisfactory compensatory mutations for the BH-GA mutant that would restore local base-pairing in this region, possibly because of the relatively large number of nucleotides

Figure 3.6. Mutation of the CU-repeat RNA sequence [nt 227 to 233] affects viral RNA dimerization and packaging. (A) Illustration of a structural domain formed by viral RNA sequences from nt 227 to 335. This domain contains the SL1, SL2, and SL3 RNA motifs and are isolated from other RNA structures by a stem formed by long range interactions between two stretches of RNA sequences at nt 227 to 231 and nt 332 to 336 (14). The BH-GA mutation replaced the RNA stretch at nt 227 to 233 with sequence 5'-AGAG-3', and presumably disrupted the highlighted stem. (B) Native Northern blots of BH-GA (lane 1) and BH10 (lane 2) RNA derived from virus particles equivalent to 150 ng of p24 antigen. The intensities of RNA signals were measured using the NIH Image program. Dimers and monomers are indicated on the left side of the gels. Results are shown of one representative gel. (C) Band intensities of dimer and monomer signals were measured using the NIH Image program, and relative levels for each construct were plotted. (D) RNA packaging levels were expressed as a percentage of the wild-type virus BH10 (arbitrarily set at 100%). Results shown in (C) and (D) represent pooled data from three Northern blots of three independent transfections of each mutant.

changed in BH-GA. However, we were able to restore putative stem base-pairing by mutating the two Cs at nt 223 and nt 225 to a G and A, respectively, in the context of the MS6 mutation (Fig. 3.7A, B). This newly generated mutation was termed MS7. Surprisingly, native Northern blotting showed that the viral RNA dimerization defect was not corrected despite the predicted restoration of base-pairing to a wild-type pattern (Fig. 3.7C, lane 2 vs 1; and Fig. 3.7D). Similarly, the packaging defect seen with the MS6 mutant was not overcome by the restoration of base-pairing in this region in MS7 (Fig. 3.7E). Data regarding these mutants are summarized in Table 3.1. It is likely that both the CU-repeats and the GA-rich sequences are important for viral RNA packaging and dimerization, but not necessarily through a direct base-pairing interaction.

3.5.5 Mutations in two stretches of G-rich sequences neither affect viral replication nor viral RNA dimerization

Previous studies suggested that two stretches of G-rich sequences at nt 363 to 367 and nt 405 to 409 contribute to viral RNA dimerization through formation of G-tetrad structures^{12,183,262}. To determine whether these two RNA segments are involved in genomic RNA dimerization, they were either individually mutated in constructs MG1 and MG2 or simultaneously changed in construct MG12 (Fig. 3.1A). The substituted nucleotides were designed such that relevant amino acids in MA were not altered. None of the three mutations affected viral growth in either MT-2 or Jurkat cells (Fig. 3.8A, B). In addition, wild-type levels of dimerized RNA were present in each of the MG1, MG2, and MG12 mutant viruses (Fig. 3.9A, B, C; and Table 3.1). Therefore, the G-rich features

Figure 3.7. Analysis of the putative RNA-RNA interactions between the CU-repeats [nt 227 to 233] and the GA-rich sequences [nt 325 to 336]. (A) Illustration of a structural domain formed by viral RNA sequences from nt 227 to nt 336. Nucleotides changed in MS6 and MS7 are underlined. (B) Secondary structure models representing the region shown in (A) for MS6, MS7, and wild-type BH10 based on the M-fold algorithm^{186,291}. (C) Native Northern blots of MS6, MS7, and BH10 (lane 1-3) RNA derived from virus particles equivalent to 150 ng of p24 antigen. The intensities of RNA signals were measured using the NIH Image program. Dimers and monomers are indicated on the left side of the gels. Results are shown from one representative gel. (D) Band intensities of dimer and monomer signals were measured using the NIH Image program, and relative levels for each construct were plotted. (E) RNA packaging levels were expressed as a percentage of wild-type virus BH10 (arbitrarily set at 100%). Results shown in (D) and (E) represent pooled data from three independent experiments for MS6 and BH10, and two for MS7.

Figure 3.7. Analysis of the putative RNA-RNA interactions between the CU-repeats [nt 227 to 233] and the GA-rich sequences [nt 325 to 336].

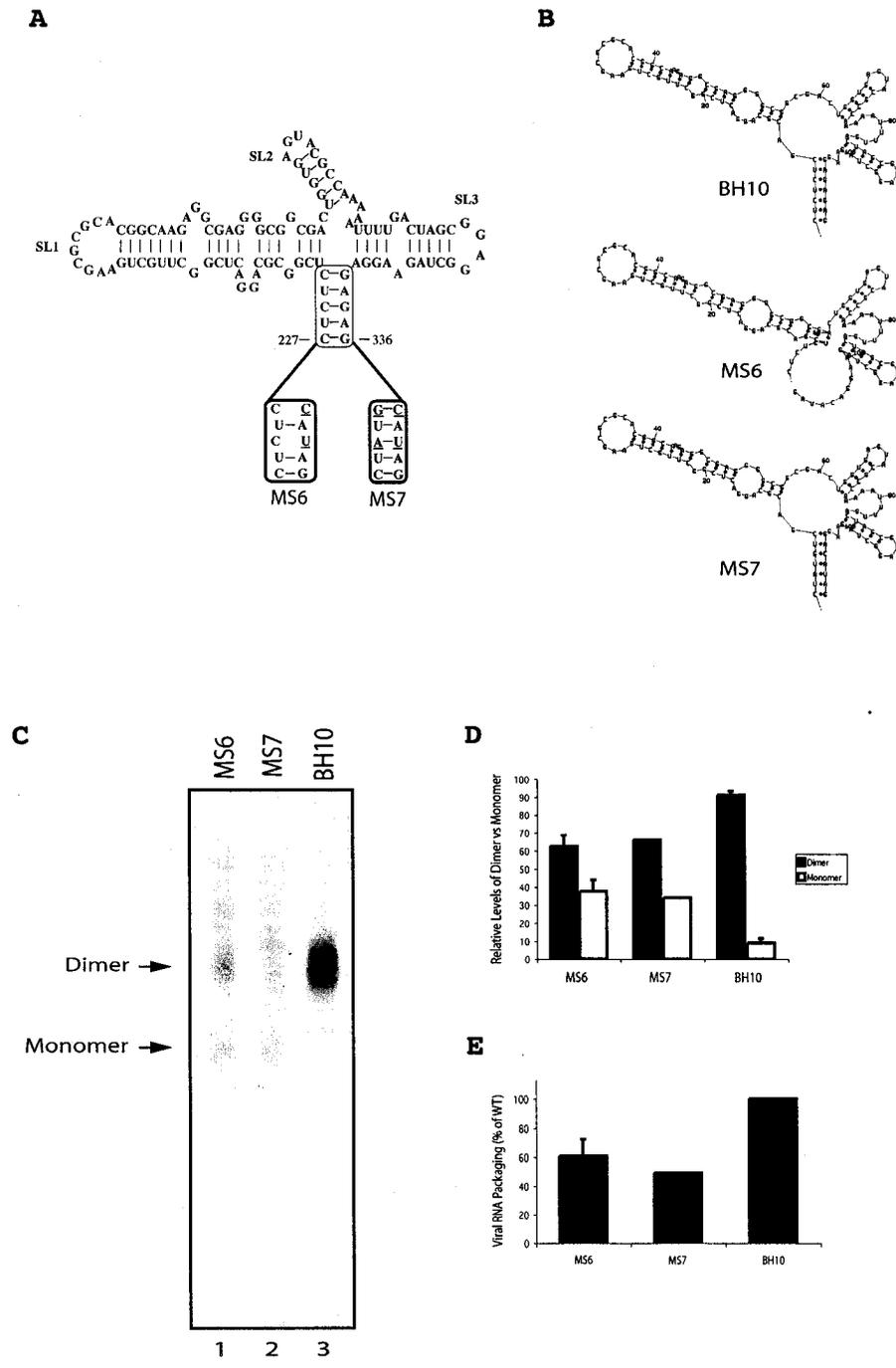


Figure 3.8. Effects of the MG1, MG2, and MG12 mutations on viral replication. Virus growth was monitored by infection of MT-2 cells (A) and Jurkat cells (B) followed by measurement of RT activity in culture fluids at various times.

Figure 3.8. Effects of the MG1, MG2, and MG12 mutations on viral replication.

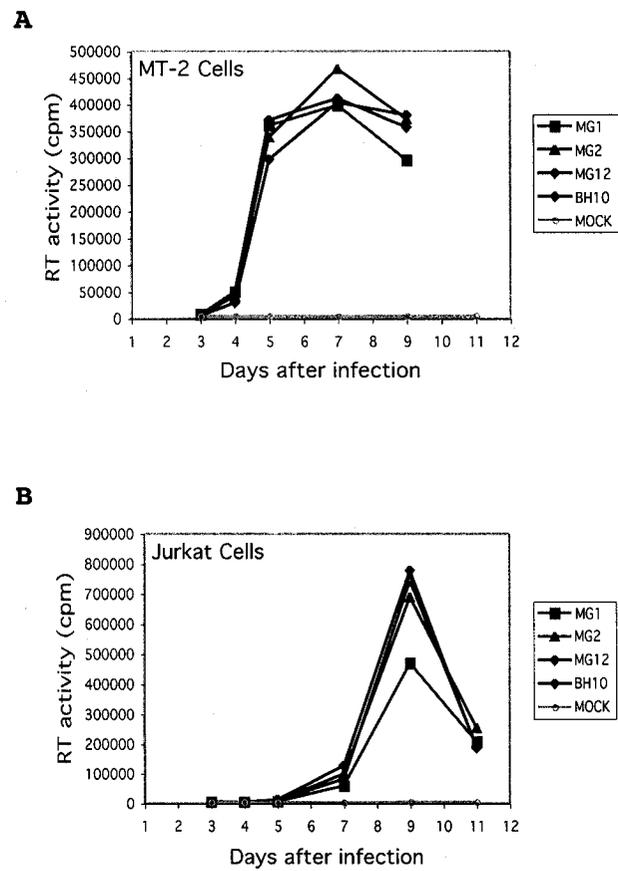


Figure 3.9. Effects of the MG1, MG2, and MG12 mutations on viral RNA dimerization and packaging. (A) Native Northern blots were performed on MG1, MG2, MG12 (lanes 1-3), and wild-type BH10 (lane 4) RNA derived from virus particles equivalent to 150 ng of p24 antigen. The intensities of RNA signals were measured using the NIH Image program. Dimers and monomers are indicated on the left side of the gels. Results are shown from one representative gel. (B) Band intensities of dimer and monomer signals were measured using the NIH Image program, and relative levels for each construct were plotted. (C) RNA packaging levels were expressed as a percentage of the wild-type virus BH10 (arbitrarily set at 100%). Results shown in (B) and (C) represent pooled data from three independent experiments.

Figure 3.9. Effects of the MG1, MG2, and MG12 mutations on viral RNA dimerization and packaging.

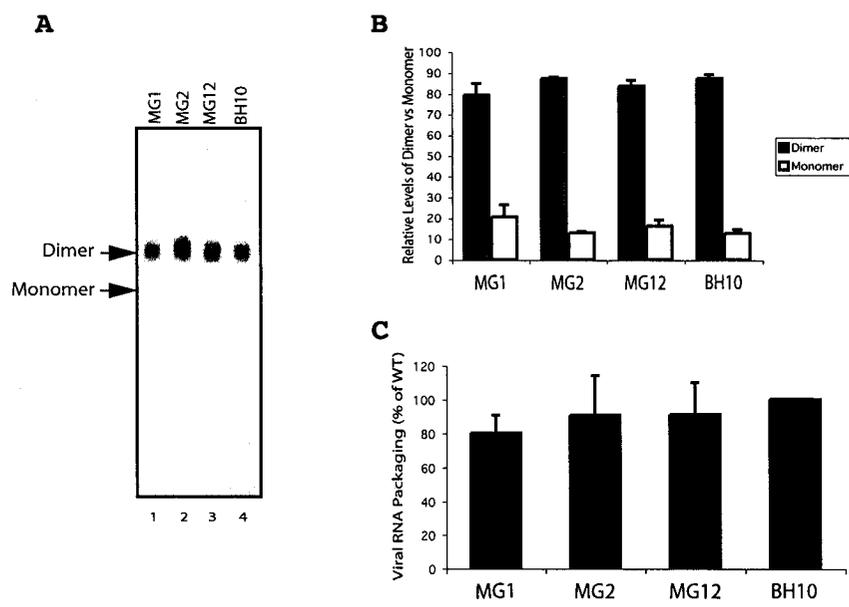


Table 3.1. Effects of various mutations on viral RNA packaging and dimerization as well as replication.

Virus	% Dimer ^A	% Packaging ^B	Replication ^C
MD1	45 ± 3	68 ± 11	+++
MD2	44 ± 3	49 ± 6	++
MD3	36 ± 5	51 ± 8	+
MS1	78 ± 1	87 ± 11	++++
MS2	41 ± 6	58 ± 11	++
MS3	43 ± 7	60 ± 8	+
MS4	59 ± 6	75 ± 10	+++
MS5	74 ± 2	79 ± 8	++++
MS6	62 ± 6	61 ± 11	+
MS7	66 ^D	49 ^D	ND ^E
BH-GA	47 ± 11	63 ± 8	ND ^E
MG1	79 ± 6	80 ± 10	++++
MG2	87 ± 1	91 ± 12	++++
MG12	84 ± 3	91 ± 9	++++
BH10	88 ± 3	100	++++

^A Percentages of dimerized viral RNA relative to total amount of viral RNA as determined by native Northern blot.

^B Levels of viral RNA packaged expressed as a percentage of wild type BH10.

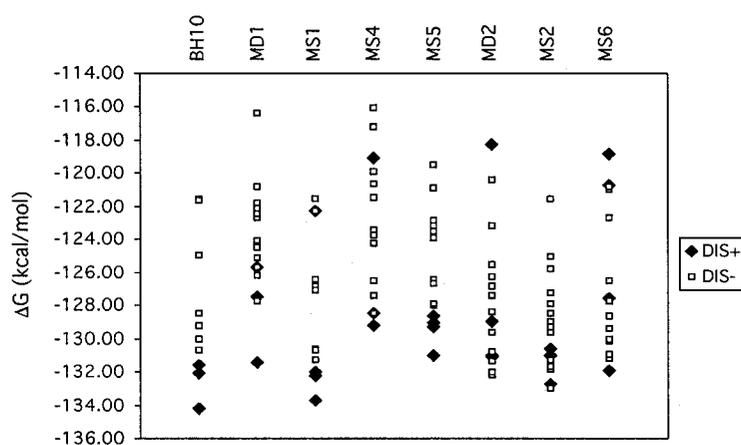
^C Replication capacity was scored based on infection studies performed in MT-2 and Jurkat cells.

^D All values represent means ± standard error for at least 3 independent experiments, except for MS7 which include data from 2 experiments.

^E ND denotes 'not determined'.

Figure 3.10. Structural analysis of mutated and wild-type HIV-1 RNA sequences spanning nt 1 to 360 on the basis of the M-Fold program^{186,291}. Each point in a series represents one possible structural prediction that the program generates. For sequences 300-400 nt in length, the program typically generates 12-15 possible structures and ranks them based on free energy (ΔG) calculations. Structures representing each mutation were divided into two groups based on the presence (\blacklozenge) or absence (\square) of the DIS (SL1) motif, and their ΔG values are plotted.

Figure 3.10. Structural analysis of mutated and wild-type HIV-1 RNA sequences spanning nt 1 to 360 on the basis of the M-Fold program.



of the RNA segments studied are important for neither viral replication, RNA dimerization, nor packaging.

3.6 Discussion

Although the RNA packaging activities of SL3 and GA-rich sequences [nt 325 to 336] have previously been reported^{5,27,47,50,112,168,188}, our results provide further *in vivo* evidence in this regard by measurements of viral RNA levels in the mutant viruses MD1 to MD3 and MS1 to MS6 (Fig. 3.3A-F). The data from our BH-GA, MS6, and MS7 mutant viruses (Figs. 3.6, 3.7) further demonstrate that this GA-rich RNA segment is important in the regulation of viral RNA packaging, but that its packaging activity may not depend on its interaction with a stretch of CU-repeats at nt 227 to 233. More importantly, our native Northern blot results suggest that both SL3 and GA-rich sequences are required for efficient RNA dimerization (Figs. 3.3A and B, 3.6B and C, 3.7C and D).

Of the four SL3 mutations studied, MD1 and MS4 caused a destabilization of SL3 structure and diminished levels of RNA dimerization; in contrast, MS1 and MS5 preserved SL3 structure and both maintained near wild-type levels of dimerization (Fig. 3.3A, B). These findings are consistent with other data showing that disruption of the SL1 stem led to aberrant RNA dimerization, and that restoration of the stem structure by compensatory mutations increased RNA dimerization to wild-type levels^{49,50}. Therefore, intact secondary structures of both SL1 and SL3 are needed for efficient viral RNA dimerization. Of these two RNA motifs, SL1 possesses a palindrome in the loop, the

alteration of which affects RNA dimerization^{49,215,251}. In contrast, replacement of the SL3 loop sequence did not interfere with viral RNA dimerization as long as the changes did not disrupt the SL3 secondary structure (Fig. 3.3A, B). SL1 apparently regulates viral RNA dimerization through its loop palindrome by a “kissing-loop” mechanism, while SL3 affects dimerization in a manner that is independent of the loop sequence.

SL3 shows high affinity for HIV-1 Gag and NC proteins⁴⁸. A nuclear magnetic resonance (NMR) structure of a complex formed by NC and SL3 has recently been resolved and has revealed specific interactions between these two molecules⁶⁷. We speculate that these specific RNA-protein interactions, thought to guide viral RNA packaging, may also be responsible for the role of SL3 in RNA dimerization.

We have shown that decreases in both RNA packaging and dimerization are the result of mutations in the SL3 and GA-rich regions (Figs. 3.3, 3.6, 3.7), but it is uncertain which of these processes is affected first. From a temporal perspective, studies with protease-negative HIV-1 suggested that viral RNA is packaged as immature dimers⁹¹, although this may not be true for other retroviruses such as avian leukosis virus (ALV)²¹¹. Interestingly, HIV-1 viral RNA may also be efficiently encapsidated in monomeric form based on the following observations: 1. less extensive mutations of the SL1 loop (i.e. a G-loop mutation) did not affect the viral RNA content but impaired RNA dimerization.⁵⁰; 2. compensatory mutations in Gag proteins (i.e. MP2 and MNC substitutions) restored viral replication and viral RNA packaging to near wild-type levels but did not correct defective viral RNA dimerization that was caused by deletion of the DIS²⁴⁸; 3. mutated viruses carrying a duplicated DIS contained wild-type levels of viral RNA, of which the majority were present as monomers^{236,237}; 4. alteration of Gag/Gag-Pol ratios in virions reduced

viral RNA dimer stability without decreasing levels of virion RNA content²⁴⁷. Therefore, wild-type RNA packaging can occur in the absence of efficient RNA dimerization.

In contrast, virus particles are able to encapsidate two copies of viral RNA despite decreased packaging efficiency. When basic amino acids that flank the zinc finger motif of NC were changed to neutral amino acids in Moloney murine leukemia virus (MoMuLV), RNA packaging was decreased; however, the viral RNA was still mainly dimeric¹²². Similar observations have been made with HIV-1¹⁵³. Therefore, RNA packaging and dimerization can be either individually or simultaneously affected by different mutations. Further studies are needed to determine which activity is affected first in the mutant viruses described herein.

In an attempt to further understand the role of the GA-rich sequences in the dimerization and packaging processes, we performed stem disruption/restoration mutagenesis experiments with the CU-repeats [nt 227 to 233] and GA-rich sequences [nt 325 to 336] to determine whether they are involved in a direct RNA-RNA interaction as depicted in Fig. 3.7A. Based on our dimerization data, both of these regions are important for viral RNA packaging and dimerization. However, despite the seemingly stable secondary structure predictions generated with the M-fold program, our MS7 compensatory mutation, which would theoretically restore base-pairing in this region, did not correct the packaging and dimerization defects seen with the MS6 mutant. These results suggest that the mechanism by which these sequences are involved in dimerization and packaging may not be through direct RNA-RNA interactions between these sequences.

Interactions between the CU-repeats [nt 227 to 233] and GA-rich sequences [nt 325 to 336] have been proposed on the basis of computer modeling and *in vitro* probing experiments, in which the RNA structure thus formed was referred to as stem I¹¹¹. However, the stability of stem I was questioned, as reactivity of G's in this region with kethoxal suggests that they are unpaired. On this basis, it was concluded that stem I may exist transiently or in equilibrium with an unpaired state. It is possible that the CU-repeats [nt 227 to 233] and GA-rich sequences [nt 325 to 336] may bind to other RNA regions as well, and thus regulate viral RNA dimerization and packaging. Alternatively, they may be involved in more than one type of RNA-RNA interaction. For example, recent *in vitro* data suggest that the HIV-1 leader RNA can form alternate structures that are proposed to regulate RNA dimer formation^{26,122}. Although these models have been generated using synthetic RNA transcripts, that for the most part do not include SL3, and await *in vivo* confirmation, it might be interesting to consider the involvement of SL3 and/or GA-rich sequences in these conformational switches with respect to dimerization and/or packaging.

The mutations we have generated in SL3 and the GA-rich region may also have disrupted DIS structure and indirectly affected RNA dimerization. Although structural studies of synthetic viral RNA fragments will provide insight on this topic, the putative effects of our mutations on DIS structure can also be assessed by the M-Fold program^{186,291}. We have divided predicted structures representing each mutation into two groups, based on the presence or absence of the DIS motif, and plotted ΔG values accordingly (Fig. 3.10). In the case of wild-type viral RNA, structures containing a DIS are preferable due to their low ΔG values. For MS1 and MS5, in which SL3 structure was

preserved, secondary structures containing a DIS were also strongly favored. In the cases of MD1 and MS4, structures containing a DIS tended to show ΔG values lower than those lacking a DIS. Therefore, mutations in SL3 may have affected RNA dimerization without a corresponding influence on DIS structure. In contrast, the MD2, MS2, and MS6 mutations in the GA-rich region failed to yield DIS-containing structures with favorable ΔG values (Fig. 3.10). This suggests that these mutations impacted adversely on formation of the DIS structure that, in turn, led to defective RNA dimerization and packaging. Since as few as two nt alterations in MS6 can alter DIS structure, we believe that more extensive GA-rich sequences at nt 325 to 336 may be essential in preserving the integrity of the HIV-1 RNA leader region and in presenting the DIS RNA motif for dimerization.

Purine-quartet structures were also proposed to play a role in RNA dimerization as shown by *in vitro* studies performed with synthetic viral RNA fragments^{12,183,262}. However, virion-derived RNA dimers are not stabilized by potassium, which in general, enhances purine-quartet structures⁹¹. Furthermore, interruption of the G-rich region at nt 817 to 821 in HIV-1 did not affect viral RNA dimerization¹⁰⁸. Finally, as shown here, mutations in two stretches of G-rich sequences did not impact on viral RNA dimerization or packaging, confirming that purine-quartets are not be involved in these processes.

In summary, our data provide evidence for the involvement of RNA packaging signals, that are located downstream of the 5' SD site, in HIV-1 RNA dimerization. These findings support the notion that multiple RNA elements may be needed for RNA dimerization in retroviruses, as was recently demonstrated in MLV^{170,210}. Since SL3

exhibits high affinity for NC protein, it is proposed that this RNA element may regulate HIV-1 RNA dimerization via interactions with NC.

CHAPTER 4

DELETION OF STEM-LOOP 3 IS COMPENSATED FOR BY SECOND-SITE MUTATIONS WITHIN THE GAG PROTEIN OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1

This chapter was adapted from an article that appeared in *Virology*, 2003, 314(1):221-8. The authors of this paper were L. Rong, **R.S. Russell**, J. Hu, M. Laughrea, M.A. Wainberg, and C. Liang. This project was originally initiated by L. Rong; however, I contributed all of the dimerization results presented herein, therefore Liwei and I are listed as co-first authors. The work was supervised by Drs. Liang and Wainberg. J. Hu and Dr. Liang constructed some of the plasmids and performed some of the virus replication analyses. Dr. Laughrea critiqued the manuscript and offered suggestions for revision.

4.1 Preface to Chapter 4

As mentioned in Ch. 1, previous work by Chen Liang showed that defects observed in HIV-1 DIS deletion mutants were compensated by compensatory point mutations in Gag^{161,163}. This work was then extended by a former PhD student in our lab, Liwei Rong, who made similar deletion mutants in SL3, the major packaging signal, and performed forced evolution studies in order to identify putative interactions between Gag and SL3 RNA sequences. In line with our other studies^{161,163} (and Ch. 2), she also found point mutations in Gag that restored replication capacity to SL3 deletion mutants. However, these point mutations were located in different positions than those seen in previous studies, implying that Gag may bind differently to SL3 than to SL1 and the GU-rich sequences reported in Ch. 2.

By that time, I had established RNA dimerization assays, and had results suggesting that SL3 sequences were not only required for RNA packaging, but also for RNA dimerization (Ch. 3). Therefore, we decided to study the effects of Liwei's SL3 mutations on HIV-1 RNA dimerization, and to test whether the second-site mutations she identified in culture were able to correct any dimerization defects that might be observed in these SL3 mutant viruses.

4.2 Abstract

Encapsidation of human immunodeficiency virus type 1 (HIV-1) RNA involves specific interactions between viral Gag proteins and viral RNA elements located at the 5' untranslated region (UTR). These RNA elements are termed packaging (Ψ) or encapsidation (E) signals and mainly comprise the stem-loop 1 (SL1) and SL3 RNA structures. We have previously shown that deletion of the SL1 sequences is compensated by second-site mutations within Gag. Similar studies are now extended to SL3 and the results demonstrate that deletion of this RNA structure is rescued by two point mutations, i.e. A11V in p2 and I12V in nucleocapsid (NC). These two compensatory mutations are different from those associated with the rescue of SL1 deletion, suggesting that SL1 and SL3 may bind to different residues of Gag during viral RNA packaging. Analysis of virion-derived RNA in native agarose gels shows that deletion of SL3 leads to decreases in both viral RNA packaging and dimerization. These defects are corrected by the compensatory mutations A11V and I12V. Yet, defects in viral RNA dimerization at an early stage that were caused by the SL3 deletion in the context of a viral protease-negative mutation cannot be overcome by these two suppressor mutations. Therefore, the positive effects of A11V and I12V on dimerization of the SL3-deleted RNA must have taken place at the maturation stage.

4.3 Introduction

Human immunodeficiency virus type 1 (HIV-1) selectively packages full-length viral RNA as its genome over a high background of cellular RNA in the cytoplasm. This process involves specific binding of HIV-1 Gag proteins to *cis*-acting viral RNA elements, termed Ψ or encapsidation (E) signals. Early mutagenesis studies have mapped the E signals to a region located between the 5' major splice donor (SD) site and the 5' portion of the *gag* gene; this region was subsequently shown to contain two RNA stem-loop (SL) structures, termed SL3 and SL4^{5,47,48,113,157,168}. Further studies revealed that RNA sequences located upstream of the 5' SD site also affected RNA packaging¹³⁴ these include the TAR structure, the poly(A) hairpin, the U5-PBS RNA complex, and SL1^{52,62,63,109,164,187-189}.

The nucleocapsid (NC) sequences of Gag protein function *in trans* to recognize the E signals and thereby direct RNA packaging^{5,29,31,74,223}. Of the multiple RNA structures within the 5' non-coding RNA leader region, SL1, SL3 and SL4 bind to NC with high affinity, as shown by *in vitro* binding assays^{28,29,48,235}. In contrast, the TAR, U5-PBS, and poly(A) elements display low affinity for NC and thus may indirectly participate in RNA packaging by regulating viral RNA metabolism or by helping to correctly present the SL1, SL3 and SL4 structures to NC⁴⁸.

Aside from its role in viral RNA packaging, SL1 has also been defined as the dimerization initiation site (DIS). This function of SL1 is attributable to the palindrome feature of its loop sequence which can trigger dimer formation through "Watson-Crick" basepairing^{49,147,149,184,201,203,204,213,215,251}. The importance of SL1 in RNA dimerization is

also supported by experiments using mutated viral RNA derived from virus particles^{25,50,108,151,248}. We have previously deleted the SL1 sequences in the context of full-length viral RNA genome and observed severe defects in both viral RNA packaging and dimerization²⁴⁸. Long-term culture of the mutated viruses in permissive cells led to isolation of compensatory mutations within Gag protein that restored viral infectivity to near wild-type levels¹⁶¹. Further analysis indicated that these compensatory mutations corrected the deficient viral RNA packaging but not RNA dimerization²⁴⁸. It is thus proposed that viral RNA can be packaged at high efficiency in the monomeric form.

SL3 represents another major RNA packaging signal²⁷. We have recently shown that disruption of the SL3 structure affected both RNA packaging and dimerization²³³. In order to further understand the mechanisms underlying the regulatory roles of SL1 and SL3 in HIV-1 RNA packaging and dimerization, we have conducted a compensation study to determine the second-site mutations that can rescue the deleted SL3 sequences. The results show that the suppressor mutations for the SL3 deletion are different from those for the SL1 deletion, indicating that SL1 and SL3 may bind to Gag proteins in different ways during encapsidation of viral RNA.

4.4 Materials and Methods

4.4.1 Plasmid construction

The BH10 infectious HIV-1 cDNA clone was employed as starting material. Mutagenesis studies were performed on the basis of polymerase chain reactions (PCR)

using *Pfu* enzyme with 3' to 5' proof-reading activity (Stratagene, La Jolla, CA). The primers used were purchased from Invitrogen (Burlington, Ontario, Canada). The $\Delta(306-325)$ deletion was engineered through use of primer pairs p Δ SL3(5'-CTGAAGCGCGCACGGCAAGAGGGCGAGGGGCGGCGACTGGTGAGTACGCCAA AAAAAGGAGAGAGATGGG-3' [nt 252 to 340])/pAPA-A(5'-CCTAGGGGCCCTGC AATTTCTG-3' [nt 1562 to 1541]). The PCR products were digested with restriction enzymes *BssH* II and *Apa* I (Invitrogen) and used to replace the same regions in BH10.

The A11V amino acid substitution in p2 was created using a PCR-Script Amp cloning kit (Stratagene) through use of primers p2-S2 (5'-GTAACAAATTCAGTTACCATAATGATGCAG-3' [nt 1443 to 1472])/p2-A2 (5'-GCA TCATTATGGTAACTGAATTTGTTACTTGG-3'[nt 1470 to 1439]). The I12V mutation in the NC sequence was engineered by PCR using primer pair pE (5'-GGAACCAAAGAAAGGTTGTTAAGTGTTC-3' [nt 1486 to 1514])/pNC-A (5'-TTAGCCTGTCTCTCAGTACAATC-3'[nt 1630 to 1608]). The PCR product was used as a primer in a second round of PCR together with primer pSph-S (5'-AGTGCATCCAGTGCATGCAGGGCC-3'[nt 977 to 1000]). The final PCR product was digested with *Sph* I and *Apa* I and inserted into $\Delta(306-325)$. The positions of primers are in relation to the first nt of the 5' R region. The mutations generated were confirmed by sequencing.

4.4.2 Cell culture, transfection and infection

COS-7 and MT-2 cells were grown in Dulbecco's modified Eagle's medium (DMEM) and RPMI 1640 medium, respectively, supplemented with 10% fetal calf serum (Invitrogen). Transfection of COS-7 cells was performed with Lipofectamine (Invitrogen). Progeny virus in culture fluids were collected 48 hours after transfection and quantified by measuring levels of HIV-1 CA (p24) antigen (Ag) by enzyme-linked immunosorption assays (ELISA) (Vironostika HIV-1 Antigen Microelisa System, Organon Teknika Corporation, Durham, NC).

Viruses containing 3 ng of p24 were used to infect 5×10^5 MT-2 cells in 2 ml RPMI 1640 medium. After 2 hours, cells were washed twice to remove unbound viruses and were grown in 10 ml complete medium. Growth of viruses was monitored by measuring reverse transcriptase (RT) activity of culture fluids at various times.

4.4.3 Analysis of viral proteins

COS-7 cells were transfected for 20 hours, followed by starvation in DMEM without Met and Cys at 37°C for 2 hours. Total cellular proteins were then radiolabeled by [³⁵S]-Met and [³⁵S]-Cys (ICN, Irvine, CA) at a concentration of 0.1 mCi/ml for 30 min at 37°C. The labeled cells were either lysed immediately after labeling or cultured for an additional hour in complete DMEM before being lysed in NP-40 lysis buffer. Cell lysates were incubated with monoclonal antibodies (MAbs) against HIV-1 p24 and the Ag:MAB complexes were absorbed with protein A-linked Sepharose 4B (Amersham Pharmacia Biotech., Baie d'Urfe, Quebec, Canada). After extensive washing, the immunoprecipitated viral proteins were fractionated on sodium dodecyl sulfate (SDS)-

12% polyacrylamide gels and visualized by exposure to X-ray films (Kodak, Rochester, NY). The [³⁵S] labeled virus particles in the culture fluids were collected through centrifugation and directly analyzed by running on sodium dodecyl sulfate (SDS)- 12% polyacrylamide gels.

4.4.4 Analysis of viral RNA by Northern blots

Progeny viruses generated by transfected COS-7 cells were first clarified by centrifugation in a Beckman GR-6S centrifuge at 3,000 rpm for 30 min at 4°C and then pelleted through a 20% sucrose cushion by ultracentrifugation in a Beckman XL-80 ultracentrifuge using an SW41 rotor at 40,000 rpm for 1 hour at 4°C. Virus pellets were suspended in 300 µl of TN buffer, a 2 µl portion was removed for p24 determination, and the remaining viruses were digested with 100 µg/ml protease K in the presence of 1% SDS, 10 mM EDTA, 100 mM NaCl, and 50 µg of yeast tRNA for 20 min at 37°C. Samples were then extracted twice with phenol:chlorophorm:isoamlyalcohol (25:24:1) and once with chlorophorm. Viral RNA was precipitated in 2.5 volumes of 95% ethanol. RNA pellets were washed with 70% ethanol and dissolved in TE buffer. An amount of viral RNA equivalent to 150 ng of HIV-1 p24 was electrophoresed on 0.9% native agarose gels in 1xTBE buffer at 100 volts for 4 hours at 4°C. The RNA was then transferred to a nylon membrane and analyzed by Northern blots using [³²P]-α-dCTP (ICN) labeled HIV-1 DNA probes. To determine the thermostability of viral RNA dimers, RNA samples were treated at various temperatures (e.g. 40°C, 45°C, 50°C and

55°C) for 10 min in a buffer containing 100 mM NaCl before being separated on native agarose gels.

In order to assess the expression of both full-length and spliced forms of viral RNA, total RNA was extracted from transfected COS-7 cells and separated on 1% agarose gels containing formaldehyde as a denaturant. RNA molecules were then transferred onto a nylon membrane and hybridized to HIV-1 probes as described above.

4.4.5 Identification of second-site mutations

MT-2 cells that had been infected with mutated viruses were cultured for prolonged periods until formation of syncytia and high levels of RT activity were observed. The infectious virus particles were further used to infect fresh MT-2 cells until wild-type replication kinetics were seen. Cellular DNA was then extracted from infected MT-2 cells and subjected to PCR and sequencing to identify newly emerged second-site mutations in proviral DNA¹⁶¹.

4.5 Results

4.5.1 Long-term culture of the $\Delta(306-325)$ mutated viruses in MT-2 cells leads to the appearance of suppressor mutations in the *gag* coding region

HIV-1 RNA sequences at nt 306 to 325 were deleted to remove the SL3 structure; the construct thus generated was termed $\Delta(306-325)$ (Fig. 4.1A). On the basis of the structure

prediction by the M-fold program^{186,291}, this deletion does not disrupt the flanking RNA secondary structures including SL1 and SL2 (Fig. 4.1B). Following transfection of COS-7 cells, the mutated DNA construct $\Delta(306-325)$ generated wild-type levels of full-length as well as spliced viral RNA as shown by the results of Northern blots (Fig. 4.1C). This indicates that the SL2 RNA structure, which contains the splice donor signal, is not affected by the $\Delta(306-325)$ deletion and thus allows normal splicing of viral RNA. In support of these results of viral RNA analysis, wild-type levels of both intracellular Gag proteins and extracellular virus particles were detected after transfection of this mutated DNA construct into COS-7 cells (Fig. 4.1D).

The infectivity of the virus particles that were generated either by the wild-type BH10 or the mutated $\Delta(306-325)$ DNA constructs was then studied through infection of MT-2 cells. The results show that growth of the $\Delta(306-325)$ mutant viruses was significantly delayed in comparison to the wild-type virus (Fig. 4.2A). The mutated viruses were further cultured in MT-2 cells over multiple passages until wild-type replication phenotypes were observed. Sequencing of proviral DNA at this stage indicated the retention of the original SL3 deletion and the emergence of novel second-site mutations in the *gag* coding region, i.e. an A11V amino acid substitution in p2 and an I12V amino acid substitution in NC (Fig. 4.2B). In addition, insertions of a number of As were seen between nt positions 305 and 306 in the non-coding RNA leader region. Site-directed mutagenesis experiments were then performed to insert these second-site mutations into the $\Delta(306-325)$ construct in order to determine whether these new mutations were sufficient and necessary to restore the infectivity of $\Delta(306-325)$ to wild-type levels. The results of infection assays showed that the A11V mutation in p2 markedly increased the

Figure 4.1. Analysis of the $\Delta(306-325)$ deletion mutant. (A) Schematic illustration of the $\Delta(306-325)$ deletion. Deleted sequences are indicated by dash lines. RNA secondary structures are shown for SL3. LTR: long terminal repeat; SD: splice donor. Numbering of nucleotides refers to the first nt in the R region. (B) Effects of the $\Delta(306-325)$ deletion on viral RNA secondary structures. Both the wild-type BH10 and the $\Delta(306-325)$ mutated HIV-1 RNA sequences, spanning nucleotide positions 243 to 361, were subjected to secondary structure analysis through use of the M-fold program^{186,291}. Locations of the SL1, SL2 and SL3 structures are indicated. (C) Expression of viral RNA within transfected COS-7 cells. Total RNA was prepared from COS-7 cells that had been transfected with either the BH10 or $\Delta(306-325)$ DNA constructs, followed by separation on agarose gels. Northern blotting was performed to detect viral RNA through use of HIV-1 specific probes. The full-length (9.2kb) as well as spliced (4kb and 2kb) forms of viral RNA are indicated on the right side of the gels. (D) Expression of viral proteins and production of virus particles. Transfected COS-7 cells were subjected to metabolic labeling through use of [³⁵S]-Methionine and [³⁵S]-Cysteine. The intracellular Gag protein and its derivatives were detected by immunoprecipitation using anti-HIV-1 p24 antibodies. [³⁵S]-labeled virus particles in the culture fluids were directly assessed on SDS-12% PAGE and visualized by exposure to X-ray films. Positions of the protein markers (kD) are indicated on the right side of the gels.

Figure 4.1. Analysis of the $\Delta(306-325)$ deletion mutant.

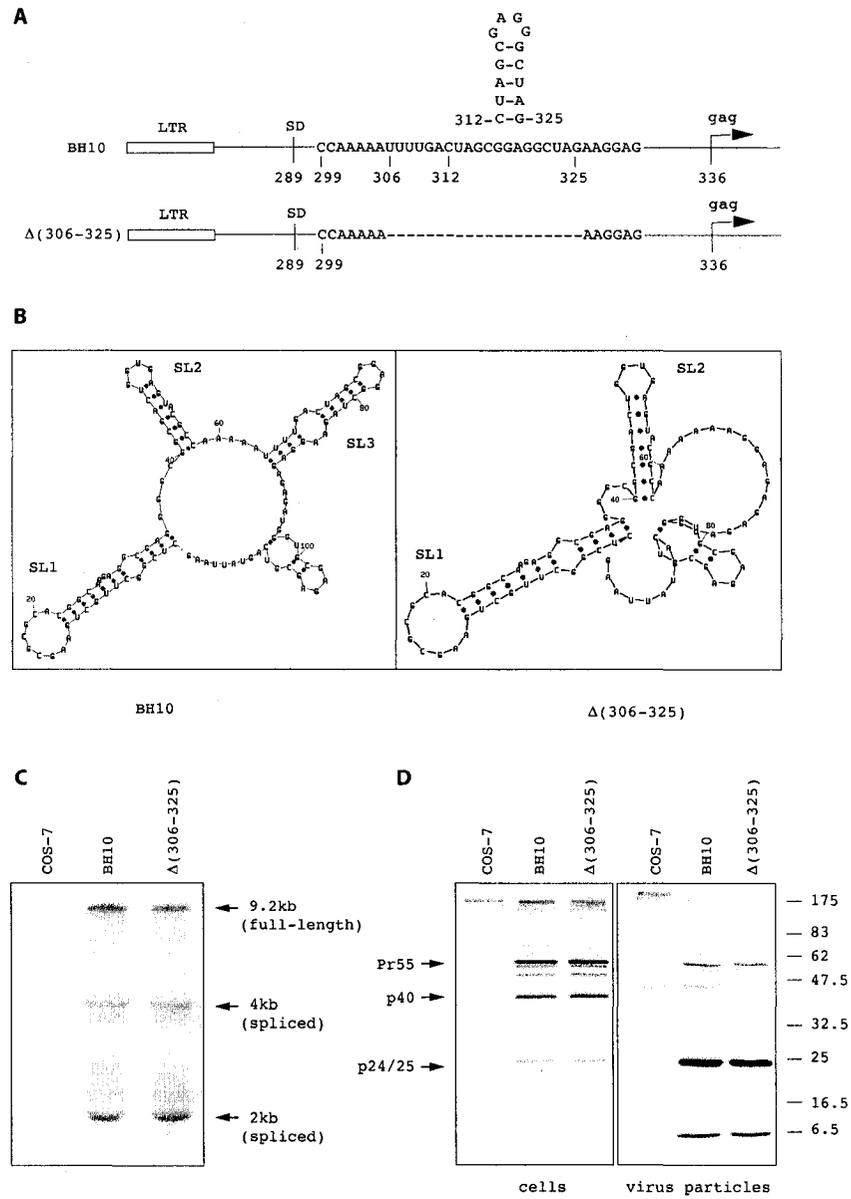


Figure 4.2. Compensation of the crippled replication of the $\Delta(306-325)$ mutated viruses by second-site mutations within the Gag protein. (A) Replication of mutated and wild-type viruses in MT-2 cells. Virus growth was monitored by measuring RT activity of culture fluids at various times. Mock infection represents exposure of MT-2 cells to heat-inactivated wild-type viruses. (B) Second-site mutations identified in the $\Delta(306-325)$ mutated viruses after long-term culture in MT-2 cells. The changed amino acid residues in the p2 and NC domains are underlined. The arrow indicates the protease cleavage site between p2 and NC.

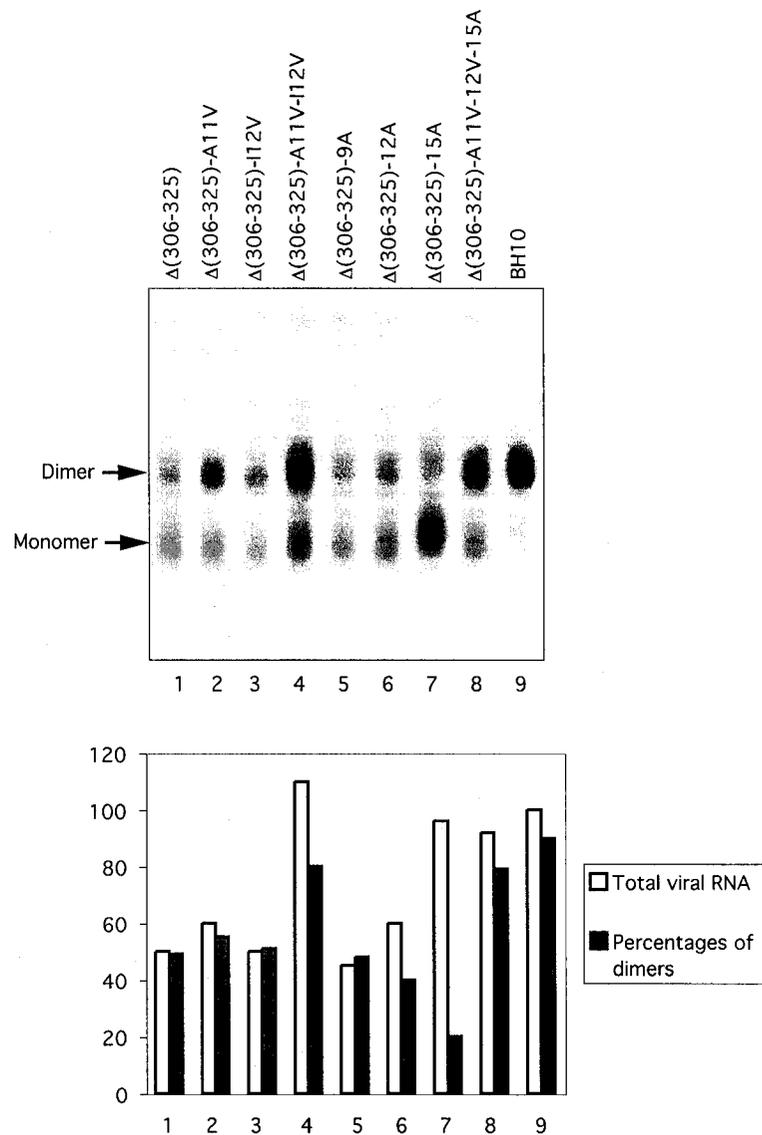
infectiousness of $\Delta(306-325)$; the I12V mutation alone barely augmented viral replication, but could synergistically increase viral growth together with A11V (Fig. 4.2A). The addition of A residues did not increase the infectivity of $\Delta(306-325)$, either alone or together with A11V and I12V (data are shown only in regard to the addition of 12 As) (Fig. 4.2A). Therefore, second-site mutations in both p2 and NC apparently played key roles in rescue of the defective replication of $\Delta(306-325)$.

4.5.2 The A11V mutation in p2 and the I12V mutation in NC increase levels of RNA packaging and dimerization in the $\Delta(306-325)$ mutated viruses

SL3 is known to direct specific packaging of HIV-1 genomic RNA. Thus, we hypothesize that the A11V and I12V mutations must have overcome deficits in packaging caused by the $\Delta(306-325)$ deletion. To test this hypothesis, viral RNA was prepared from virus particles through protease K digestion and phenol:chlorophorm extraction in the presence of 100 mM NaCl, fractionated on native agarose gels, and analyzed by Northern blotting. The results showed that the $\Delta(306-325)$ deletion led to severe reductions not only in the levels of total viral RNA but also in the relative levels of dimeric RNA (Fig. 4.3, lane 1). Following combination with the $\Delta(306-325)$ deletion, A11V gave rise to a measurable increase of viral RNA levels (Fig. 4.3, lane 2), while the I12V mutation alone was not effective in this regard (Fig. 4.3, lane 3). However, A11V and I12V together restored viral RNA packaging in $\Delta(306-325)$ to wild-type levels and also significantly increased dimer contents in the mutant virus particles (Fig. 4.3, lane 4). Therefore, A11V

Figure 4.3. Effects of the A11V and I12V second-site mutations on deficient RNA packaging and dimerization of $\Delta(306-325)$. Viral RNA was prepared from virus particles by phenol-chlorophorm extraction. An amount of viral RNA equivalent to 150 ng of p24 (CA) antigen was loaded onto native agarose gels for each construct. Following electrophoresis, viral RNA was transferred onto nylon membranes and further detected by HIV-1 DNA probes. Positions of dimers and monomers are indicated on the left side of the gels. Intensities of RNA signals were calculated using the NIH Image program and the results are summarized in the graph. The total levels of wild-type RNA in BH10 were arbitrarily set at 100. The relative proportion of dimers for each construct was also determined and the results are shown in the graph. The data shown are from one representative experiment.

Figure 4.3. Effects of the A11V and I12V second-site mutations on deficient RNA packaging and dimerization of $\Delta(306-325)$.



and I12V acted synergistically to rescue the deficient RNA packaging and dimerization in $\Delta(306-325)$.

We also examined the effects of additional inserted As on viral RNA dimerization and packaging in $\Delta(306-325)$. The results of Fig. 4.3 show that neither 9 nor 12 As, inserted at nt 305, led to any increase in the total levels of $\Delta(306-325)$ RNA (lanes 5 and 6). However, insertion of 15As restored wild-type RNA packaging; yet, the majority of viral RNA exhibited a monomeric form (lane 7). Again, the A11V and I12V mutations were jointly able to overcome most of the deficits in RNA packaging and dimerization caused by the $\Delta(306-325)$ deletion in the presence of the inserted 15 As (lane 8).

We next asked whether the corrected RNA dimers in the $\Delta(306-325)$ -A11V-I12V recombinant viruses possessed wild-type thermostability. To pursue this subject, viral RNA samples were treated at various temperatures in the presence of 100 mM NaCl before separation on native agarose gels. We first examined dimers retained in the $\Delta(306-325)$ mutated viruses. The results of Fig. 4.4 show that these RNA dimers dissociated over a range of temperatures similar to those seen with wild-type dimers. When the $\Delta(306-325)$ -A11V-I12V RNA dimers were analyzed, the same dissociation temperatures were again observed (Fig. 4.4). Therefore, the $\Delta(306-325)$ dimers that were formed in the presence of A11V and I12V mutations exhibit wild-type thermostability.

Figure 4.4. Thermostability of RNA dimers prepared from the $\Delta(306-325)$, $\Delta(306-325)$ -A11V-I12V and BH10 viruses. Viral RNA samples were treated at different temperatures (i.e. 40°C, 45°C, 50°C and 55°C) in a buffer containing 100 mM NaCl prior to separation on native agarose gels.

Figure 4.4. Thermostability of RNA dimers prepared from the $\Delta(306-325)$, $\Delta(306-325)$ -A11V-I12V and BH10 viruses.

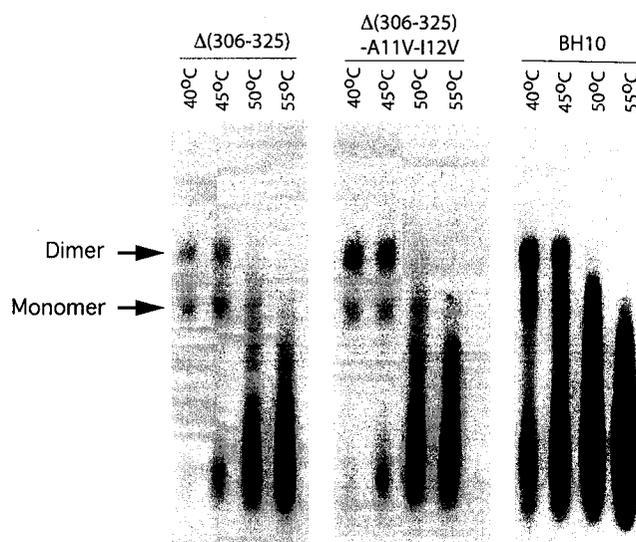
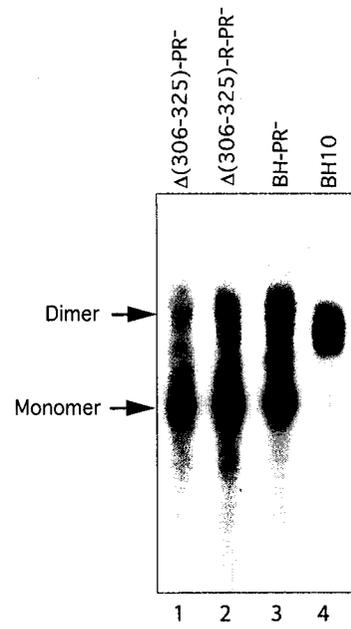


Figure 4.5. Effects of the $\Delta(306-325)$ deletion on RNA dimerization in the PR⁺ virus.
Construct $\Delta(306-325)$ -R-PR⁺ contains the $\Delta(306-325)$ deletion, the second-site mutations A11V and I12V, as well as PR mutation D25A.

Figure 4.5. Effects of the $\Delta(306-325)$ deletion on RNA dimerization in the PR- virus.



4.5.3 The A11V and I12V substitutions do not correct defective RNA dimerization caused by the $\Delta(306-325)$ deletion in the context of protease-negative viruses

HIV-1 RNA is initially packaged as immature dimers that are further converted into mature forms upon release of the NC protein by protease (PR) cleavage of the Gag protein⁹¹. The immature dimers are readily detected within the protease-negative virions and exhibit a slightly slower mobility during electrophoresis as well as lower stability in comparison to the mature dimers that are associated with the wild-type mature virus particles. We wished to determine whether the $\Delta(306-325)$ deletion had exerted adverse impacts on dimerization at the immature dimer stage. To pursue this subject, the $\Delta(306-325)$ deletion was combined with a point mutation D25A that eliminates the enzymatic activity of the HIV-1 PR²⁰⁰, and the construct thus generated was termed $\Delta(306-325)$ -PR⁻. The D25A substitution was also inserted into wild-type BH10 to generate BH-PR⁻. In the case of BH-PR⁻, a significant amount of viral RNA was observed as immature dimers that migrated at a slower rate than did the wild-type RNA dimers in BH10 (Fig. 4.5). This indicates that the immature dimers are in a relatively loose conformation compared to the mature dimers⁹¹. In contrast, RNA derived from $\Delta(306-325)$ -PR⁻ virus was present almost exclusively in monomeric form (Fig. 4.5). These findings suggest that the $\Delta(306-325)$ RNA is defective in the formation of immature dimers during virus assembly.

Next, we asked whether the compensatory mutations A11V and I12V were able to correct this early dimerization defect. Accordingly, $\Delta(306-325)$, A11V, and I12V were combined with the PR mutation D25A to yield a construct $\Delta(306-325)$ -R-PR⁻. The results showed that levels of RNA immature dimers in this latter virus were drastically lower

than those associated with BH-PR⁻ (Fig. 4.5). These data, together with those shown in Fig. 4.3, lead to the conclusion that rescue of deficient dimerization by A11V and I12V must have taken place during viral maturation following PR cleavage.

4.6 Discussion

In the present study we have identified two second-site mutations within the Gag protein that can rescue the deleted SL3 sequences. Interestingly, the two mutations, i.e. A11V in p2 and I12V in NC, are distinct from those previously characterized for the deleted SL1 sequences which include substitutions T12I in p2 and T24I in NC¹⁶¹. This difference suggests that even though both SL1 and SL3 participate in RNA packaging via binding to Gag⁴⁸, the detailed RNA-protein interactions may vary. In support of this notion, the A11V and I12V mutations could not rescue the SL1-deleted viruses in regard to its crippled viability as well as defective RNA packaging and dimerization (data not shown).

Compensation of the deleted SL1 and SL3 sequences by their relevant suppressor mutations involves restoration of wild-type RNA packaging. Yet, in contrast to mutations A11V and I12V that also corrected the defective dimerization caused by the deleted SL3 sequences, mutations T12I and T24I could not fix the defective dimers associated with the deleted SL1 sequences²⁴⁸. Thus, SL1 and SL3 may regulate viral RNA dimerization via distinct mechanisms. Results of cell-free assays show that SL1 is able to form dimers spontaneously without the assistance of any viral or cellular proteins by virtue of its palindrome loop sequence^{49,147,149,184,201,203,204,213,215,251}; in contrast, short synthetic viral

RNA fragments containing the SL3 sequence do not dimerize²⁶². Antisense DNA oligonucleotides, targeting either SL1 or SL3, were able to block dimerization of synthetic viral RNA fragments. However, the SL3-antisense molecules bound to RNA dimers while the SL1-antisense molecules did not, perhaps because SL1 already forms “Watson-Crick” base-pairs within dimers and SL3 may have been involved in other types of molecular interactions²⁸⁶.

Rescue of the deleted SL3 sequences by second-site mutations in p2 (A11V) and NC (I12V) involve increases in the levels of both RNA packaging and dimerization. It is thus possible that increased dimerization is a result of optimized RNA packaging. Although we cannot rule out this possibility at this stage, studies performed with PR-negative viruses indicate that the A11V and I12V mutations cannot rescue defective immature dimers caused by the $\Delta(306-325)$ deletion (Fig. 4.5). Together with the fact that these two substitutions can increase packaging of $\Delta(306-325)$ RNA to wild-type levels (Fig. 4.3), we conclude that to the least extent, the early defects in dimerization caused by $\Delta(306-325)$ prior to PR cleavage are irrelevant to the deficient packaging of $\Delta(306-325)$ RNA.

Since SL1 and SL3 are closely located within viral RNA genome, it is possible that deletion of SL3 may have interfered with the folding of the SL1 structure, and thereby indirectly affected dimerization. This seems unlikely, since the results of RNA secondary structure prediction by the M-Fold program show that deletion of SL3 still allows the correct folding of SL1 as well as the other structures within the non-coding RNA leader region (Fig. 4.1B). Therefore, although results of computer modeling need to be further verified by structural probing experiments, it is highly possible that the $\Delta(306-325)$

deletion may have disturbed RNA dimerization through mechanisms other than disruption of the SL1 structure.

We hypothesize that the involvement of SL3 in RNA dimerization may be related to its NC-binding activity^{28,29,48,235}. NC has been shown to promote dimerization of synthetic viral RNA fragments^{86,205}, and mutations within NC zinc finger motifs led to deficient viral RNA dimerization within virions¹⁵³. Therefore, the role of SL3 in dimerization may partially be due to its function in the specific recruitment of NC onto viral RNA.

Notably, compensation of mutated RNA packaging signals (e.g. SL1 or SL3) involves second-site mutations in the p2 sequence (e.g. T12I and A11V); this indicates a potential role of p2 in recognition of viral RNA during viral assembly. Indeed, presence of HIV-1 p2 sequence, together with NC, in an HIV-1/HIV-2 chimeric virus significantly augments specific packaging of HIV-1 RNA; this constitutes the first evidence that p2 may play important roles in HIV-1 RNA packaging¹³¹.

Together with our previous findings obtained with the deleted SL1 sequences¹⁶¹, this study further demonstrates that HIV-1 is capable of surviving partially removed RNA packaging signals by modifying the Gag sequences. Presumably, these modified Gag proteins must have possessed wild-type levels of binding affinity for the mutated viral RNA and thus restored wild-type RNA packaging.

CHAPTER 5

EFFECTS OF A SINGLE AMINO ACID SUBSTITUTION WITHIN THE P2 REGION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 ON PACKAGING OF SPLICED VIRAL RNA

This chapter was adapted from an article that appeared in *Journal of Virology*, 2003, 77(24):12986-95. The authors of this paper were **R.S. Russell**, A. Roldan, M. Detorio, J. Hu, M.A. Wainberg, and C. Liang. All data presented in this chapter, except for those related to CBMC infections (M. Detorio), were from experiments performed by myself under the supervision of Drs. Liang and Wainberg. J. Hu and Dr. Liang constructed some of the plasmids used. Dr. Roldan offered valuable suggestions for revision of the manuscript, and provided one of the plasmids used in this study.

5.1 Preface to Chapter 5

The work outlined in Ch. 5 was actually the first project that I started when I joined the lab. As mentioned in Ch. 1, I was interested in the mechanism of the compensatory point mutations that Dr. Liang had identified in the context of DIS deletion mutants^{161,163}. In an attempt to characterize the underlying mechanism of these compensatory mutations, we designed two other DIS mutations, and combined these with various combinations of the compensatory mutations previously identified. I then set up the RNA dimerization assay for the purpose of analyzing the effects of these point mutations on dimerization of viral genomic RNA. However, despite much work and frustration, the mechanism of these point mutations still eluded us. Therefore, we used the RNA dimerization assay to generate the data described in the previous three chapters.

However, I was still interested in the underlying mechanism employed by these compensatory point mutations. So I later returned to this project, and with the help of Véronique Bériault and Ariel Roldan, set up the RNase Protection Assay used in our lab. Using this assay, we planned to more accurately determine the effects of the compensatory mutations on the overall packaging of viral genomic RNA, as well as the packaging specificity for genomic versus spliced viral RNA, in our DIS mutants. The results of this analysis were quite surprising, and in my opinion, are the most interesting and significant findings in this thesis.

5.2 Abstract

Human immunodeficiency virus type 1 (HIV-1) encapsidates two copies of viral genomic RNA in the form of a dimer. The dimerization process initiates via a 6-nucleotide palindrome that constitutes the loop of a viral RNA stem-loop structure (i.e. stem loop 1, SL1, also termed the dimerization initiation site (DIS)) located within the 5' untranslated region of the viral genome. We have now shown that deletion of the entire DIS sequence virtually eliminated viral replication, but that this impairment was overcome by four second-site mutations located within the matrix (MA), capsid (CA), p2 and nucleocapsid (NC) regions of Gag. Interestingly, defective viral RNA dimerization caused by the Δ DIS deletion was not significantly corrected by these compensatory mutations which did, however, allow the mutated viruses to package wild-type levels of this DIS-deleted viral RNA while excluding spliced viral RNA from encapsidation. Further studies demonstrated that the compensatory mutation T12I located within p2, termed MP2, sufficed to prevent spliced viral RNA from being packaged into the Δ DIS virus. Consistently, the Δ DIS-MP2 virus displayed significantly higher levels of infectiousness than did Δ DIS virus. The importance of position T12 in p2 was further demonstrated by identification of four point mutations, T12D, T12E, T12G and T12P, that resulted in encapsidation of spliced viral RNA at significant levels. Taken together, our data demonstrate that selective packaging of viral genomic RNA is influenced by the MP2 mutation, and that this represents a major mechanism for rescue of viruses containing the Δ DIS deletion.

5.3 Introduction

The hallmark of all retroviruses, including the human immunodeficiency virus type 1 (HIV-1), is the need to generate a full-length double stranded proviral DNA molecule from positive-stranded RNA, following which integration of this DNA into the host cell genome takes place. Given that the length of the retroviral genome is >9kb and that most viral RNA molecules are nicked, one might expect that the successful generation of complete proviral DNA from such a genome would represent a difficult task. To overcome difficulties, retroviruses specifically package two copies of dimeric full-length viral genomic RNA that are non-covalently linked at their 5' ends⁵³. The availability of two copies of the RNA genome provides viral reverse transcriptase with an alternate template, should this enzyme encounter a break during transcription, and also contributes to overall viral genetic variability¹²⁰.

The mechanisms of retroviral RNA dimerization have been extensively studied, particularly in the case of HIV-1. The major determinant for HIV-1 RNA dimerization has been mapped to a stem-loop structure termed SL1, that is located within the 5' untranslated region (UTR) of the viral genome^{147,184,251}. A 6-nucleotide (nt) palindrome sequence within the loop region of SL1 initiates dimerization through a “kissing-loop” mechanism that involves the formation of base-pairs between the palindromes of two genomic RNA molecules^{49,149,205,215}. The loose dimer is converted to a more stable extended duplex with the help of the viral nucleocapsid (NC) protein^{86,150,205}. Accordingly, SL1 has been termed the dimerization initiation site (DIS)²⁵¹.

The features of SL1 that allow this RNA structure to function as the DIS have been further explored in a number of genetic and structural studies. HIV-1 RNA dimerization is affected not only by mutation of the palindrome loop sequence but also by alteration of the stem region^{49,152,248}. These latter changes may either affect the appropriate presentation of the palindrome within the loop, which is needed to initiate RNA dimerization, or prevent the transition of RNA dimers from the loose to the stable form. Detailed structures of the RNA dimer formed by SL1 have configured that both loop and stem sequences contribute to stabilization of the RNA duplex^{80,81,95,201,202,283}. As an example, the loop region contains three adenine residues that cannot form base pairs within the dimer, and as a consequence, might be expected to distort the RNA duplex. However, what actually happens is that these adenines initiate a distinctive pattern of interstrand stacking which helps to stabilize the dimer structure. These studies provide an explanation as to why RNA dimerization is initiated at SL1 and not at other viral RNA structures, such as the poly(A) hairpin, the loop region of which also contains a palindrome^{153,232}.

Consistent with these observations, the mutation of DIS sequences can result in severely diminished viral infectiousness^{25,50,112,134,151,152,214,248}. To further understand the role of the DIS, we have previously generated two DNA constructs, BH-LD3 and BH-LD4, that lacked portions of the stem sequence of the DIS structure^{161,163}. Replication of these two mutated viruses in permissive cells led to the outgrowth of revertant viruses that displayed wild-type infectiousness. Interestingly, these revertants retained the original DIS mutations, but acquired compensatory mutations within the Gag region. However, the mechanisms that underlie this compensatory activity have not been

characterized. In the present study, we show that the compensatory mutations involved were able to confer viability to viruses lacking the entire DIS sequences. Interestingly, restoration of viral replication was not accompanied by wild-type RNA dimerization but rather by the exclusion of spliced viral RNA from the infectious Δ DIS viruses that were produced, and MP2 played a major role in this process.

5.4 Materials and Methods

5.4.1 Plasmid construction

The infectious HIV-1 BH10 cDNA clone was employed to generate the constructs described below and all mutations were introduced by PCR-based strategies using the *Pfu* enzyme (Stratagene, La Jolla, CA). Δ DIS and Δ Loop are deletion mutations that were engineered by PCR using the sense primer pS (5'-AGA CCA GAT CTG AGC CTG GGAG-3' [nt 14-35, numbering from the 1st nt in R]) together with the antisense primers Δ DIS (5'-TAC TCA CCA GTC GCC GCC CTC CTG CGT CGA GAG AGC-3' [nt 293-227]) and Δ Loop (5'-CGC CGC CCC TCG CCT CCT GCC GCA GCA AGC CGA GTC CTG C-3' [nt 282-235]), respectively (Fig. 5.1). The PCR products thus generated were then used as primers together with primer pApa-A (5'-CCT AGG GGC CCT GCA ATT TCT G-3' [nt 1559-1538]) in a second round of PCR. Final PCR products were digested with the restriction enzymes *Nar*I and *Apa*I and inserted into a BH10 vector that had been cut with the same enzymes. To generate constructs containing the Δ DIS or Δ Loop deletions, or wild-type BH10, along with various combinations of compensatory

mutations, the *gag* gene was replaced with that of previously generated constructs containing combinations of various compensatory mutations (Table 5.1)¹⁶³. Accordingly, Δ DIS-MP2, Δ DIS-MNC, Δ Loop-MP2, and Δ Loop-MNC were generated by substituting the *gag* gene from the previously described BH10-MP2 and BH10-MNC into the Δ DIS and Δ Loop constructs²²⁹. The protease-negative (PR-)²⁰⁰, MD1²³³, and constructs containing substitutions of all 20 amino acids in position 12 of p2 have also been described²²⁹. All constructs generated were confirmed by sequencing.

To generate antisense riboprobes for use in RNase protection assays (RPA), DNA fragments of 486, 449, and 477 bp were amplified from BH10, Δ DIS, and Δ Loop proviral DNA, respectively, using primers RPA-S (5'-Cag ggc ccG AGA GCT GCA TCC GGA G-3' [nt -164 to -140]), which was modified to contain an *Apa*I restriction enzyme site (shown in lower case), and RPA-A (5'-CCT CCG gaa ttc AAA ATT TTT GGC G-3' [nt 321 to 297]), which was modified to contain an *Eco*RI restriction enzyme site (shown in lower case), as previously described²³³. The resulting PCR products were digested with *Apa*I and *Eco*RI and inserted into the pBluescript II KS⁺ cloning vector (Stratagene), that had been cut with the same enzymes to generate constructs RPA1, RPA-DIS, and RPA-Loop.

5.4.2 Cell culture, transfection, and infection

COS-7 cells were grown in Dulbecco's modified Eagle's medium (DMEM). HeLa-CD4-LTR- β -gal cells were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Michael Emerman¹³⁵, and were

cultured in DMEM containing 0.2 mg/ml G418 and 0.1 mg/ml hygromycin B. MT-2 and Jurkat cells were grown in RPMI 1640 medium containing 2 mM L-Glu. All cells were supplemented with 10% fetal calf serum. Cord blood mononuclear cells (CBMCs) were grown in RPMI 1640 supplemented with 10% fetal calf serum and 20 U/ml of IL-2. Transfection of COS-7 cells was performed using Lipofectamine (Invitrogen, Burlington, Ontario, Canada). Quantities of progeny viruses were quantified based on p24 antigen levels by enzyme-linked immunosorption assays (ELISA) (Vironostika HIV-1 Antigen Microelisa System; Organon Teknika Corporation, Durham, NC).

MT-2 or Jurkat cells (5×10^5) were incubated in 2 ml of medium at 37°C for 2 hours with aliquots of virus equivalent to 5 ng of p24. The cells were then washed twice and maintained in 10 ml of medium. CBMC infections were performed as described³⁶. Culture fluids were collected at various times to determine levels of reverse transcriptase (RT) activity. Viral infectivity was also measured using a multinuclear activation of a galactosidase indicator (MAGI) assay¹³⁵. HeLa-CD4-LTR- β -gal cells were plated in 24-well plates (4×10^4 cells/well) and cultured for one day before being infected in triplicate with an amount of virus equivalent to 10 ng of p24 (CA) in the presence of 20 μ g/ml DEAE Dextran. Forty-eight hours after infection, cells were fixed with 1% formaldehyde and 0.2% glutaraldehyde in PBS and stained with X-gal. Infectivity was scored based on the number of blue cells counted per well as described¹³⁵.

5.4.3 Native Northern blotting

Progeny viruses generated by transfected COS-7 cells were first clarified by centrifugation in a Beckman GR-6S centrifuge at 3,000 rpm for 30 min at 4°C and then pelleted through a 20% sucrose cushion by ultracentrifugation in a Beckman XL-80 ultracentrifuge using an SW41 rotor at 40,000 rpm for 1 hour at 4°C. Virus pellets were suspended in 300 µl of TN buffer; a 2 µl portion was removed for p24 determination and the remaining material was treated with virus lysis buffer (50 mM Tris-HCl [pH 7.4], 10 mM EDTA, 1% SDS, 100 mM NaCl, 50 µg of yeast tRNA/ml, 100 µg proteinase K/ml) for 20 min at 37°C. Samples were then extracted twice with phenol:chloroform:isoamylalcohol (25:24:1) and once with chloroform. Viral RNA was precipitated in 2.5 volumes of 95% ethanol and a 0.1 volume of 3M NaOAc [pH 5.2]. RNA pellets were washed with 70% ethanol and dissolved in TE buffer. An amount of viral RNA equivalent to 150 ng of HIV-1 p24 was fractionated on 0.9% native agarose gels in 1xTBE buffer at 100 volts for 4 hours at 4°C and analyzed by Northern blots²⁴⁰, using a [³²P]-α-dCTP (ICN, Irvine, CA) labeled 2 kb HIV-1 DNA fragment [nt 1 to 2000] as a probe. Bands were visualized by autoradiography and quantified by digital image analysis using the AlphaImager v.5.5 program. Briefly, object boxes were selected at the smallest size necessary to encase the largest band to be measured in a series. The same box was then used to measure all bands within that series. Based on manufacturer's recommendations, the 'Autobackground' setting which calculates an independent background for each object box derived from the average intensity of the 10 lowest pixels within each box.

5.4.4 RNase protection assays

Preparation of riboprobes and RPA experiments were performed as described²³³. Briefly, radiolabeled probes were transcribed in vitro from *BspE1*-linearized RPA1, RPA-DIS, and RPA-Loop plasmids using the T7-MEGAscript kit (Ambion Inc., Austin, TX) in the presence of [³²P]- α -UTP (ICN). RNA was isolated from viruses as described above. Amounts of virion RNA equivalent to 25 ng of p24 capsid (CA) antigen were treated with 10 U of DNase I (Invitrogen) for 30 min at 37°C to remove any plasmid contamination, and then subjected to phenol:chloroform extraction and ethanol precipitation, before analysis with the RPA II kit (Ambion Inc., Austin, TX). RNA was then incubated at 42°C overnight with an excess of labeled riboprobe (10⁵ cpm), followed by digestion with single-strand-specific RNases. Protected fragments were separated on 5% polyacrylamide-8M urea gels, visualized by autoradiography, and quantified by digital image analysis using the AlphaImager v.5.5 program.

5.5 Results

5.5.1 Compensatory mutations increase the replication capacity of the Δ DIS mutant

We have previously identified mutations within Gag protein that were able to rescue deletion of stem sequences within the DIS^{161,163}. We now wished to determine whether the complete absence of DIS sequences could still be compensated by second-site mutations. Accordingly, the DIS region spanning nt 243 to 277 was removed to generate

a construct termed Δ DIS (Fig. 5.1). Mutated DNA was transfected into COS-7 cells and the virus particles thus generated were used to infect MT-2 cells. The results of Fig. 5.2A show that RT activity was detectable in the wild-type BH10 virus culture by day 4, which coincided with the first appearance of cytopathic effects (CPE), and that this peaked at day 6. In contrast, the Δ DIS mutant culture was negative for RT activity and did not show CPE over three months. Similar results in regard to the Δ DIS mutant were obtained in Jurkat cells and human CBMC (Fig. 5.2B). Thus, Δ DIS viruses were unable to establish persistent infection in culture and did not have the opportunity to accumulate second-site mutations to improve infectiousness.

We further studied the production of p24 by Jurkat cells that had been infected by the Δ DIS virus and found low levels of capsid protein (approximately 35 pg/ml, as opposed to an average of 100,000 pg/ml for wild-type virus) during two weeks of culture. In the case of the Δ DIS, this subsequently became negative and remained so over two months (data not shown), suggesting that the virus was able to replicate at low levels, but too low to establish a persistent infection.

We next asked whether the compensatory mutations that had previously been identified with the BH-LD3 and BH-LD4 deletions were able to rescue the Δ DIS deletion. The results of Fig. 5.2A show that the combination of Δ DIS deletion with the compensatory mutations restored viral replication and that as few as two of these mutations (i.e. MNC-MP2 in DA or MNC-CA1 in DB) were sufficient in this regard. It should be noted that the DB mutant, which included the CA1 mutation instead of MP2, showed a lower rate of virus replication than did viruses containing other combinations of the compensatory mutations. This suggests that the MP2 mutation played a more

Figure 5.1. Illustration of the various RNA structural elements located within the HIV-1 5' region, including TAR, poly(A), U5-PBS, SL1 (DIS), SL2, SL3 (Ψ), and the *gag* gene. The secondary structure of SL1 is shown below with the 257-GCGCGC-262 loop palindrome highlighted in bold. Nucleotides that were deleted in the Δ DIS and Δ Loop mutants are indicated by arrows, with Δ Loop lacking the 9 nucleotides that comprise the loop of SL1, and Δ DIS missing all 35 nucleotides contained within SL1. The MA, CA, p2, NC, p1, and p6 domains of Gag are shown, with the positions of the MA1, CA1, MP2, and MNC compensatory point mutations indicated by asterisks.

Figure 5.1. Illustration of the various RNA structural elements located within the HIV-1 5' region, including TAR, poly(A), U5-PBS, SL1 (DIS), SL2, SL3 (Ψ), and the gag gene.

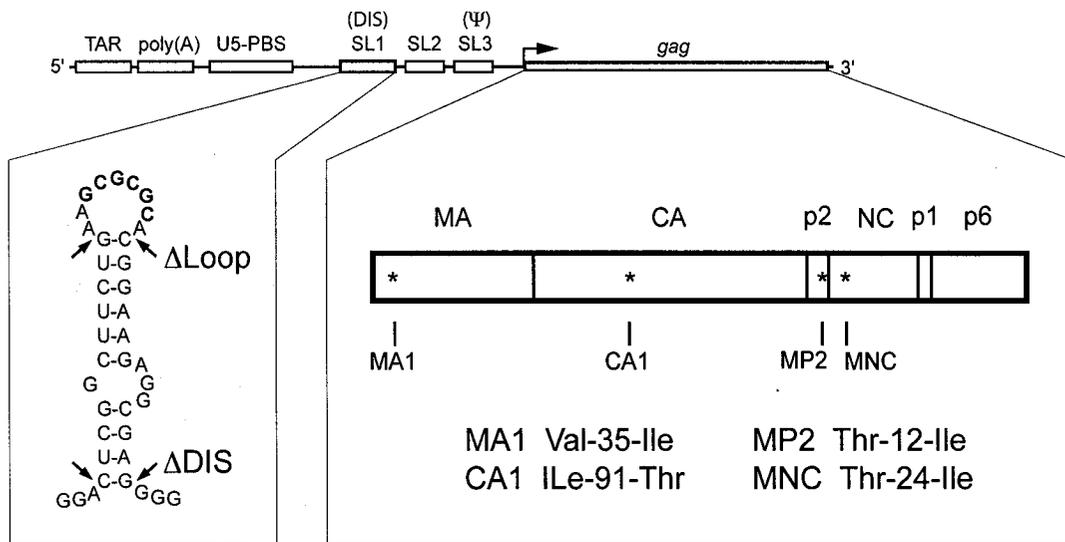
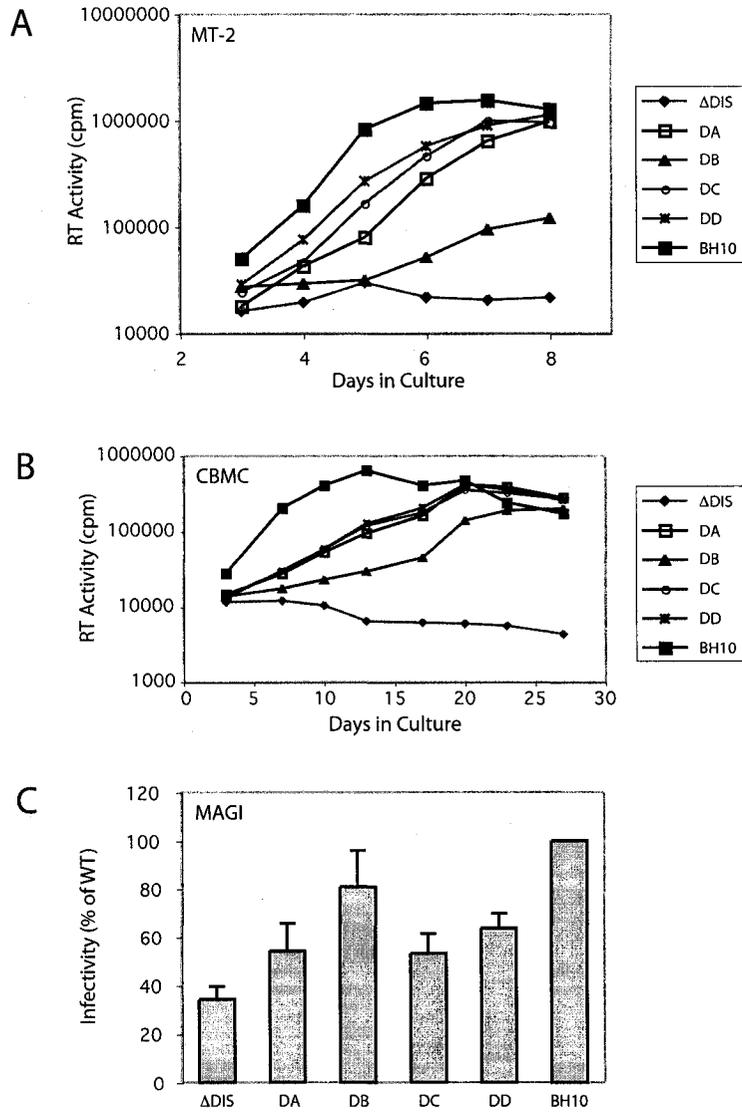


Table 5.1. Viruses generated in this study.

Virus	Deletion	Compensatory Mutations
Δ DIS	Δ DIS	None
DA	Δ DIS	MNC - MP2
DB	Δ DIS	MNC - CA1
DC	Δ DIS	MNC - MP2 - CA1
DD	Δ DIS	MNC - MP2 - CA1 - MA1
Δ Loop	Δ Loop	None
LA	Δ Loop	MNC - MP2
LB	Δ Loop	MNC - CA1
LC	Δ Loop	MNC - MP2 - CA1
LD	Δ Loop	MNC - MP2 - CA1 - MA1
BH10-D	None	MNC - MP2 - CA1 - MA1

Figure 5.2. Viral replication kinetics of the Δ DIS virus in the presence of various combinations of compensatory mutations. MT-2 cells (A) or CBMCs (B) were infected by an amount of progeny virus containing 5 or 200 ng of p24 antigen, respectively. Viral replication was monitored based on observations of syncytium formation and RT activity in culture supernatants at various times. For MAGI assays (C), viruses equivalent to 10 ng of p24 antigen were incubated with 4×10^4 HeLa-CD4-LTR- β -gal cells/well in triplicate for 48 hours. The number of blue cells in each well were counted (multinuclear syncytia were scored as one) and expressed as a percentage of wild-type to generate a bar graph representing relative infectivity. The number of β -gal-positive cells ranged from 50-200/well, and the mock infected wells contained a mean of 9 ± 3 . See Table 5.1 for the combinations of compensatory mutations represented by DA, DB, DC and DD.

Figure 5.2. Viral replication kinetics of the Δ DIS virus in the presence of various combinations of compensatory mutations.



important role in viral reversion than did CA1. When combinations of three (DC; MNC-MP2-CA1) or all four (DD; MNC-MP2-CA1-MA1) compensatory mutations were incorporated into the Δ DIS mutant, the resultant viruses grew to significantly higher levels than the Δ DIS alone, with rates of replication comparable to that of wild-type (Fig. 5.2A). The positive roles of these compensatory mutations were further confirmed by infection studies performed with CBMC (Fig. 5.2B) and Jurkat cells (data not shown).

We further quantified the levels of infectivity of our mutants in an one-round infection experiment termed the MAGI assay¹³⁵. The results of Fig. 5.2C show that the Δ DIS mutant exhibited an approximately 3-fold decrease in viral infectivity compared to the wild-type BH10. In agreement with the results of the spread infection studies shown in Figs. 5.2A and 5.2B, infectivity of the DIS deleted viruses was significantly increased in the presence of the compensatory mutations (Fig. 5.2C). It was also noted that the DB mutant, which lacks the MP2 mutation, displayed the greatest increase in infectivity in the MAGI assay among the four mutants DA, DB, DC and DD, as opposed to the lowest replication rate of DB in each of the MT-2, Jurkat and CBMC cells. Since the MAGI assay measures the transactivation levels of HIV-1 LTR by Tat protein, the observed discrepancy indicates that compensation of Δ DIS by the four suppressor mutations, particularly by MP2, occurs, to a large extent, at the late stages of viral replication.

5.5.2 Compensatory mutations do not restore wild-type RNA dimerization to Δ DIS

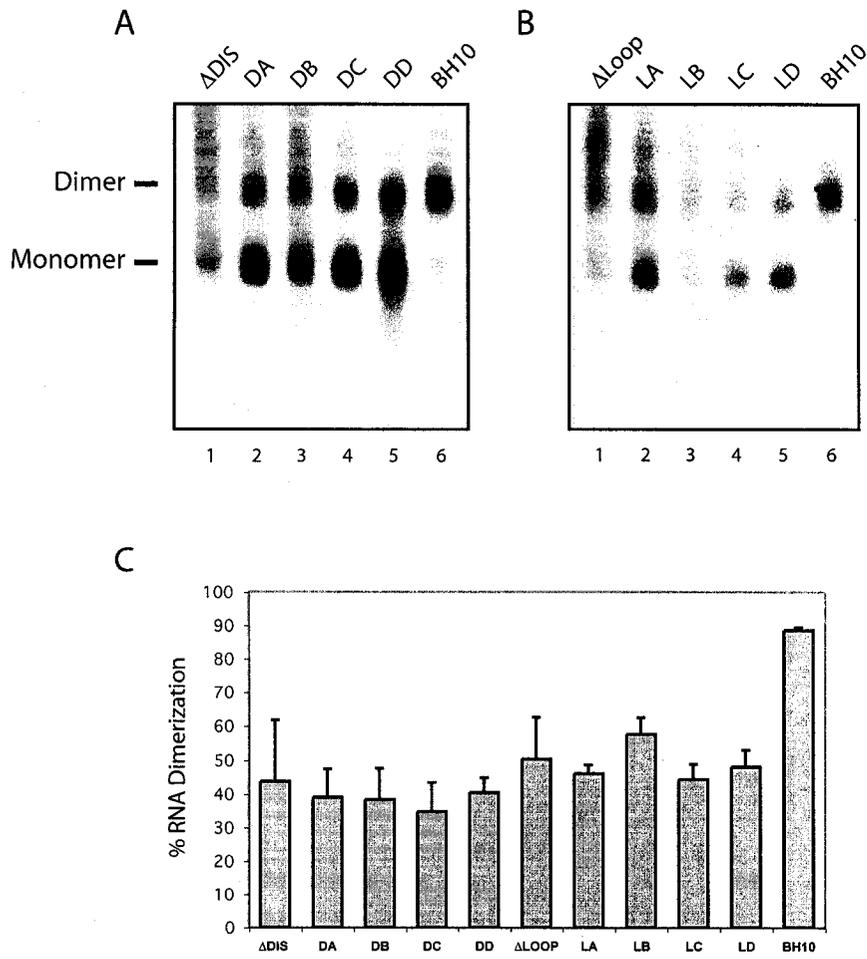
The DIS is the major signal for viral RNA dimerization^{147,184,251}. Not surprisingly, the Δ DIS mutant showed a decreased level of dimerized RNA compared to wild-type BH10

(Fig. 5.3A, lane 1 versus 11). Moreover, Δ DIS RNA formed complexes that migrated slower on gels than RNA dimers, a defect not seen with wild-type BH10 RNA. We next assessed whether the compensatory mutations could repair these RNA dimerization defects by native Northern blotting and found that introduction of the substitutions had only modest effects. Notably, approximately 55% of viral RNA was still in monomeric form compared with only 10% monomeric presence in the case of BH10 (Fig. 5.3A). However, large aggregates of viral RNA associated with Δ DIS were resolved by the presence of the compensatory mutations, particularly when at least three were present (Fig. 5.3A, lanes 4, 5). Therefore, the compensatory mutations did not lead to wild-type RNA dimerization but did have positive effects in regard to formation of RNA complexes.

An important determinant for RNA dimerization within the DIS is the palindrome loop sequence 257-GCGCGC-262 that triggers dimerization by formation of “Waston-Crick” base-pairs^{49,149,205,215}. Removal of the loop sequence in the construct Δ Loop (shown in Fig. 5.1) affected RNA dimerization, but to a lesser extent than the Δ DIS deletion (Fig. 5.3A, 3B). We next tested whether the compensatory mutations were able to correct the defect in RNA dimerization caused by Δ Loop. The results showed that significant levels of monomeric RNA were still associated with each of the viruses that contained the Δ Loop deletion in combination with various compensatory mutations (Fig. 5.3B). However, the latter viruses contained less of the high molecular weight RNA, that migrated slower than dimers on gels. Hence, wild-type RNA dimerization was not restored by compensatory mutations within Gag.

Figure 5.3. Effects of the Δ DIS and Δ Loop deletions on viral RNA dimerization. Viral RNA was prepared from mutant viruses Δ DIS, DA, DB, DC, DD (panel A, lanes 1-5), Δ Loop, LA, LB, LC, LD (panel B, lanes 1-5) and wild-type virus BH10 (panels A and B, lane 6), equivalent to 150 ng of p24 antigen, and fractionated on native agarose gels, followed by Northern blot analysis. Dimers and monomers are indicated on the left side of the gels. The gels shown are from one representative experiment. Band intensities of dimer and monomer signals were measured using the AlphaImager v.5.5 program, and relative dimer levels were plotted for each construct (C). The results represent pooled data from three Northern blots using virion-derived RNA from three independent transfection experiments. The interassay variation is reflected by the error bars; for instance, the DA, DC and DD mutants may display as low as 30% RNA dimers. See Table 5.1 for the combinations of compensatory mutations represented by DA, DB, DC and DD.

Figure 5.3. Effects of the Δ DIS and Δ Loop deletions on viral RNA dimerization.

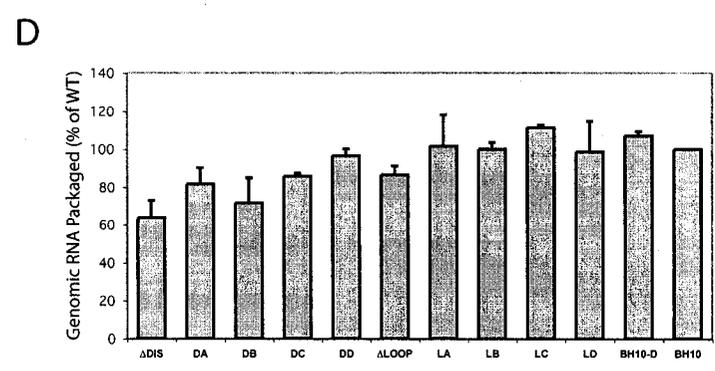
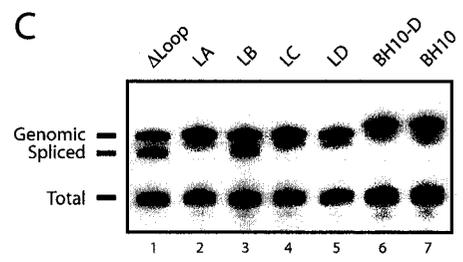
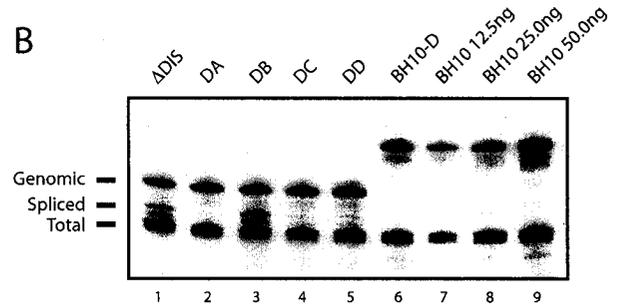
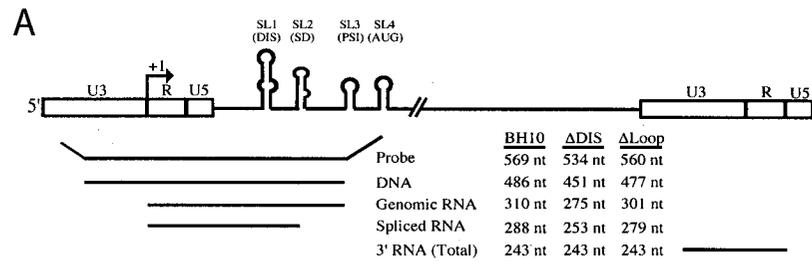


5.5.3 The compensatory mutations restore wild-type RNA packaging to the Δ DIS and Δ Loop viruses while excluding spliced viral RNA from virions

As stated, the compensatory mutations did not correct dimerization defects associated with Δ DIS, but they did lead to more intense dimer and monomer bands on gels (Fig. 5.3A). This suggests that these mutations may have helped to increase overall levels of viral RNA within the Δ DIS viruses. To validate this notion, we next assessed our viral RNA samples by RNase protection assays. The results showed that all mutants in transfected cells produced approximately equal levels of viral RNA, with similar ratios of genomic to spliced viral RNA as wild-type (data not shown). When RNA was extracted from the wild-type and mutant viruses, it was found by RPA that the Δ DIS mutant packaged approximately 35% less viral genomic RNA than did wild-type BH10 (Fig. 5.4B, lane 1 vs 8, and Fig. 5.4D). The presence of the compensatory mutations resulted in increased levels of genomic RNA packaged into the Δ DIS mutant (Fig. 5.4B, lanes 2-5, and Fig. 5.4D), and the DD virus (that contained all four compensatory mutations) showed the highest level in this regard (97%). Therefore, the compensatory mutations increased overall packaging efficiency of viral RNA, which may have accounted for the increased infectiousness of the Δ DIS virus. As a control, the BH10-D virus, which contained all four compensatory mutations in the context of wild-type 5' RNA sequences, was also analyzed by RPA and showed similar levels of viral RNA to wild-type virus (Fig. 5.4B, lane 6). Thus, the effects of the compensatory mutations on packaging efficiency of the Δ DIS mutant were specific to this deletion.

Figure 5.4. Effects of the Δ DIS and Δ Loop deletions on viral RNA packaging efficiency and specificity. (A) Illustration of the RPA system. On the top are the 5' and 3' LTR sequences and stem loops 1-4. Below are the respective probes used for BH10, Δ DIS and Δ Loop viral RNA molecules as well as the sizes of the protected fragments. These probes allow the detection of proviral DNA contamination of the virion-derived RNA. RNA species detected include genomic RNA (upper band), spliced RNA (middle band), and a total viral RNA band derived from binding of the probe to the 3' U3 and R sequences found on all viral RNA species (lower band). In some experiments, spliced and total RNA protected fragments appear as a doublet. In the case of the spliced RNA, this may result from heterogeneity in the initial sequences of the various spliced exons immediately following the major splice donor, allowing 1 or 2 extra nucleotides of homology with the probe. It is also suggested in the RPA II kit literature that A/U-rich sequences, often found in the 3' UTR of many transcripts, can be susceptible to RNase digestion due to local denaturation of the double-stranded RNA hybrid. Such doublets have been reported elsewhere^{31,112,287,288}. (B, C) Quantification of viral RNA by RPA. Viral RNA was prepared from mutant viruses Δ DIS, DA, DB, DC, DD, BH10-D (panel B, lanes 1-6), Δ Loop, LA, LB, LC, LD (panel C, lanes 1-5), and wild-type virus BH10 (panel B, lanes 7-9, panel C, lane 6). An amount of viral RNA equivalent to 25 ng p24 was annealed to 10^5 cpm of α -UTP-labeled riboprobe, digested with RNases specific for single-strand RNA, and protected fragments were separated by 5% denaturing-PAGE. A dilution series of wild-type RNA (12.5, 25.0, and 50.0 ng p24, Panel B, lanes 7-9, respectively) was analyzed to show the linear range of the assay. Two samples containing 10 μ g of yeast tRNA +/-RNase were included in all RPA experiments to demonstrate probe specificity, but are not shown due to the large difference in size between the probe and the relevant bands. One representative gel is shown from three independent experiments. (D) Relative packaging efficiency. Bands were measured using the AlphaImager v.5.5 program. Packaging levels of genomic RNA are expressed as a percentage of wild-type BH10 (arbitrarily set at 100%). The bar graph represents pooled data from three RPA gels using RNA from three independent transfections of each mutant. See Table 5.1 for combinations of mutations represented by DA, DB, DC and DD.

Figure 5.4. Effects of the Δ DIS and Δ Loop deletions on viral RNA packaging efficiency and specificity.



Consistent with previous studies showing that reduced HIV-1 genomic RNA packaging was constantly accompanied by increased incorporation of spliced viral RNA^{50,244}, a strong band corresponding to the spliced viral RNA was found to be associated with the Δ DIS virus, as distinct from the exclusive packaging of genomic RNA by wild-type BH10 (Fig. 5.4B). This suggests that Δ DIS caused defects in packaging efficiency and packaging specificity as well, both of which might have been related to its inability to replicate in culture. The presence of the compensatory mutations helped the Δ DIS virus to exclude spliced RNA molecules from being packaged (Fig. 5.4B), although they did not all function equally well in this regard. The DA (Δ DIS-MNC-MP2), DC (Δ DIS-MNC-MP2-CA1), and DD (Δ DIS-MNC-MP2-CA1-MA1) viruses recruited only trace amounts of spliced viral RNA (Fig. 5.4B, lanes 2, 4, and 5); in contrast, the DB (Δ DIS-MNC-CA1) virus exhibited high levels of spliced viral RNA concomitant with deficient genomic RNA packaging. More importantly, the DB virus replicated at a significantly slower rate than either DA, DC, or DD (Fig. 5.2). These results strongly suggest that correction of both packaging efficiency and specificity was important for compensation to occur.

The positive roles of these compensatory mutations in overcoming the deficits of packaging specificity were further confirmed by experiments performed in the context of the Δ Loop deletion (Fig. 5.4C). Similar to the DB virus, the LB virus packaged significant levels of spliced viral RNA. However, it was noteworthy that DB and LB are the only viruses that did not contain the MP2 mutation (Table 5.1); this suggests that MP2 is responsible for restoration of wild-type packaging to the Δ DIS and Δ Loop viruses.

5.5.4 The MP2 mutation alone corrects defects in packaging specificity seen in the Δ DIS and Δ Loop mutants

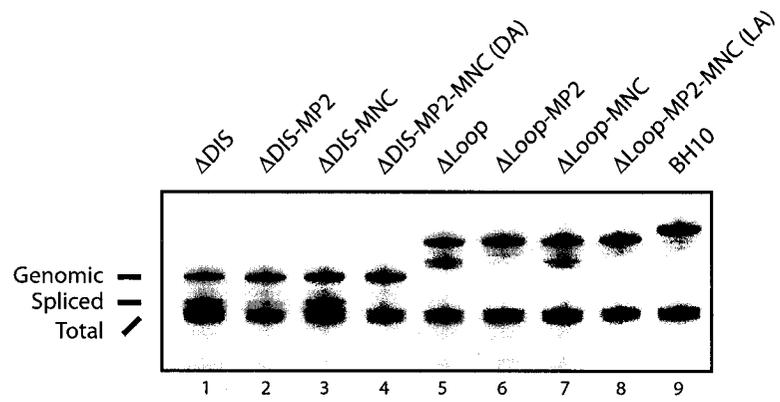
We next generated four DNA constructs that contained the Δ DIS or Δ Loop deletions, in combination with either the MP2 or MNC compensatory mutations. Analysis of virion-derived RNA from these mutants showed that Δ DIS-MP2 barely packaged any spliced viral RNA compared with Δ DIS and Δ DIS-MNC (Fig. 5.5). Similarly, MP2 alone sufficed to restore normal packaging specificity to the Δ Loop mutant (Fig. 5.5, lanes 5 vs 6). Consistently, the Δ DIS-MP2 virus resulted in CPE by day 14 after infection of MT-2 cells and showed peak levels of RT activity on day 17, compared with wild-type virus, that showed peak RT activity at day 7 (data not shown). These data demonstrate a role for MP2 in restoration of wild-type packaging specificity to Δ DIS as well as the importance of this mutation in augmenting the infectiousness of this virus.

5.5.5 Mutation of T12 within p2 causes defective packaging of wild-type genomic RNA

The positive effect of MP2 on selective packaging of full-length versus spliced Δ DIS RNA suggests a role for the T12 amino acid position in HIV-1 RNA packaging. To assess this possibility, we performed RNase protection assays on a series of viruses containing all 20 amino acid substitutions at position 12 of p2 in the context of wild-type

Figure 5.5. Effects of the MP2 and MNC mutations on selective packaging of the full-length Δ DIS and Δ Loop RNA molecules. RPA analysis was performed as described in Fig. 5.4 using virion-derived RNA from the Δ DIS (lane 1) and Δ Loop (lane 5) mutants in the presence of MP2 (lanes 2 and 6), MNC (lanes 3 and 7), or both compensatory mutations in combination (lanes 4 and 8), along with viral RNA from wild-type BH10 (lane 9). One representative gel is shown of two independent experiments.

Figure 5.5. Effects of the MP2 and MNC mutations on selective packaging of the full-length Δ DIS and Δ Loop RNA molecules.



RNA packaging signals²²⁹. The results show that wild-type BH10 virus packaged more than 95% full-length genomic RNA, despite the fact that spliced viral RNA species represented the majority of viral RNA that was present in the cytoplasm of BH10-transfected cells (Fig. 5.6, lane 23 versus 24). Most of the substitutions at position 12 of p2 had no effect on the selective packaging of full-length viral RNA (Fig. 5.6 and Table 5.2), but exceptions were the replacement of T by G, P, D, and E that caused significant increases in levels of the 288 nt band representing spliced viral RNA (Fig. 5.6, lanes 1, 12, 16, 17, and Table 5.2).

To validate that these bands represent packaged spliced RNA, as opposed to degradation products of the probe, the previously described MD1 mutant, which contains wild-type sequence except for deletion of the SL3 loop sequence 5'-GGAG-3' (nt 317-320), and which packages significant levels of spliced viral RNA²³³, was included as a control (Fig. 5.6, lane 21). As expected, the spliced RNA bands seen in the G, P, D, and E lanes correspond to 288 nt spliced viral RNA observed with the MD1 mutant. Since the mutated residue is located close to the protease (PR) cleavage site between p2 and NC, we also included a PR- control to rule out the possibility that any of the observed phenotypes might have been caused by improper Gag cleavage. The results of Fig. 5.6 (lane 20) show that PR- virus particles barely packaged any spliced viral RNA. This result also confirms that packaging specificity is indeed conferred by the Gag precursor prior to cleavage by viral protease. Taken together, these data demonstrate that T12 within p2 participates in the selection of full-length HIV-1 RNA for packaging.

Figure 5.6. Effects of amino acid substitutions at position 12 in p2 on packaging specificity of wild-type RNA. RPA was performed as described in Fig. 5.4 using virion-derived RNA from a panel of HIV-1 mutants containing each of the 20 amino acids at position 12 in p2²²⁹. Thr is shown in the wild-type BH10 (lane 23) and single letter codes are given for the other 19 amino acids (lanes 1-19). Lane 5 shows T12I previously identified as MP2. The PR- and Ψ-mutant MD1 (containing wild-type sequence except for deletion of the SL3 loop sequence 5'-GGAG-3' [nt 317-320]) were run as controls (lanes 20 and 21, respectively). A 12.5 ng p24 equivalent of wild-type BH10 RNA was run to show the range of the assay (lane 22). A 250 ng sample of cytoplasmic RNA from transfection of wild-type BH10 was also analyzed to show the abundance of spliced RNA in the transfected cells, and to confirm the identity of the spliced band reported in other lanes (lane 24). Band intensities were measured using the AlphaImager v.5.5 program, and the data are summarized in Table 5.2. One representative gel is shown of two independent experiments.

Figure 5.6. Effects of amino acid substitutions at position 12 in p2 on packaging specificity of wild-type RNA.

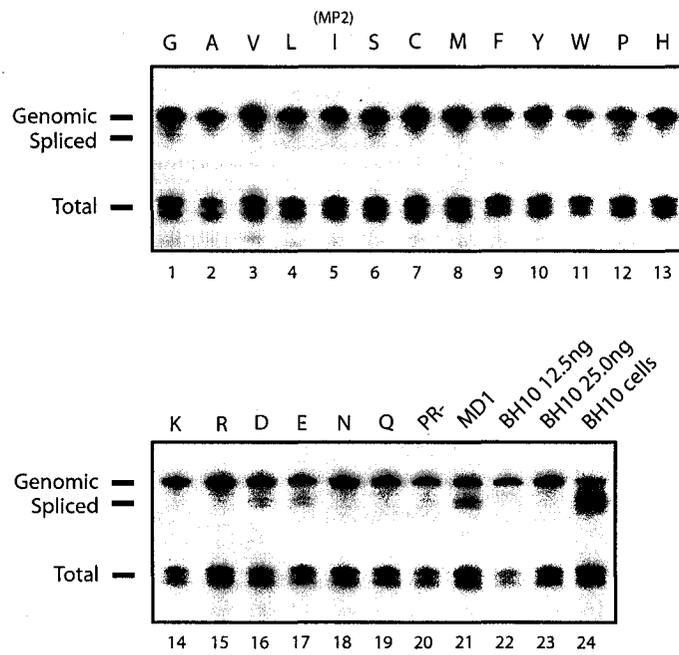


Table 5.2. Effects of substitutions at T12 of p2 on overall versus specific packaging.

Amino Acid	% Packaging ^{a,c} Specificity	% Genome ^{b,c} Packaging
G	43	87
A	75	77
V	65	87
L	83	81
I	96	82
S	65	87
C	76	90
M	75	86
F	103	70
Y	84	87
W	94	72
P	49	78
H	52	79
K	97	90
R	80	107
D	48	91
E	40	82
N	100	98
Q	92	93
T(BH10)	100	100

^a The ratio of genomic to spliced viral RNA incorporated into virions was expressed as a percentage of wild-type BH10 based on digital image analysis of RPA gels.

^b Viral genomic RNA packaged, as a percentage of wild-type BH10.

^c All values represent the average from two independent experiments. Interassay variation was less than 10%.

5.6 Discussion

The DIS represents the major dimerization signal for HIV-1 RNA^{147,184,251}. Partial deletion of DIS sequences, as found in the case of the BH-LD3 and BH-LD4 deletions^{161,163}, severely attenuated viral replication. However, the mutated viruses were able to establish persistent infection in permissive cells and eventually regained wild-type replication capacity due to the emergence of four compensatory mutations in Gag^{161,163}. In contrast, removal of the entire DIS effectively destroyed viral replication capacity. Interestingly, infectivity of the Δ DIS virus could be restored to near wild-type levels by the same four compensatory mutations that were originally associated with the rescue of the BH-LD3 and BH-LD4 deletions. Thus, HIV-1 can survive the loss of the DIS, reflecting the highly plastic nature of the virus and its genome.

It is of interest to understand how compensatory mutations within Gag, that are distal to the original deletions, could have rescued Δ DIS. In agreement with previous studies, the Δ DIS deletion severely compromised HIV-1 RNA dimerization, a defect that must have led to dramatic reductions in viral infectivity^{25,50,112,134,150-152,214,248}. It is reasonable to assume that the four compensatory mutations, which restored wild-type infectiousness to the Δ DIS virus, should also have corrected this dimerization defect, but this was not the case. Conceivably, the four compensatory mutations could still have promoted the association of the Δ DIS RNA, and that binding between the mutated RNA molecules may have been too weak to resist extraction and electrophoresis procedures. However, even if this is the case, we can still conclude that tightly-associated RNA dimers, as seen within

wild-type viruses, are not a strict prerequisite for efficient viral replication, since the Δ DIS viruses could replicate in the presence of the compensatory mutations.

Although the compensatory mutations did not help Δ DIS RNA to dimerize in wild-type fashion, they did exert positive effects on the folding and association of Δ DIS RNA molecules. In this context, Fig. 5.3 shows that significant levels of Δ DIS viral RNA migrated on gels at a rate slower than that of the dimer complexes. This migration defect was overcome by the compensatory mutations (Fig. 5.3). It is likely that lack of the Δ DIS element not only prevented RNA dimerization but may also have led to abnormal RNA folding. In support of this view, large viral RNA complexes have also been detected for a variety of mutations within the DIS element^{50,248}. Although the compensatory mutations were unable to rescue the dimerization function of the deleted DIS motif, they did help to reorganize Δ DIS RNA molecules within virus particles to promote discrete dimer or monomer forms.

Aside from its role in viral RNA dimerization, the DIS acts in concert with other viral RNA sequences, such as SL3, to regulate the specific encapsidation of viral RNA^{25,50,52,104,112,151,187,214}, for a review, see reference²⁷. Our data show that deleting the DIS interfered with viral RNA packaging, as shown by increased levels of spliced viral RNA associated with the Δ DIS and Δ Loop viruses (Fig. 5.4). Interestingly, the MP2 mutation was able to help the Δ DIS virus to exclude spliced viral RNA from being packaged. Since the selective encapsidation of two copies of full-length viral RNA is normally achieved by specific interactions between NC residues and RNA packaging signals located within the 5' UTR of viral RNA²⁷, we were surprised to find that the excessive encapsidation of spliced viral RNA into the Δ DIS virus was repaired by

changing a single amino acid at position T12 within p2 (i.e. MP2) rather than by mutations at NC residues, such as the MNC mutation. This demonstrates the pivotal role of the p2 region in HIV-1 RNA packaging, and this conclusion is further supported by the identification of four mutations, T12D, T12E, T12G and T12P, that led to packaging of spliced viral RNA in the context of wild-type BH10 virus (Fig. 5.6).

Consistent with its role in viral RNA packaging, the MP2 mutation was also able to increase the viability of the Δ DIS virus to significant levels. This suggests that the correction of non-specific viral RNA packaging may represent a major mechanism for compensation. The importance of MP2 in rescue of the Δ DIS deletion is also supported by its role in rescue of other mutated viruses that are deleted within the 5' UTR of HIV-1, e.g. U5²³², the region immediately downstream of the primer binding site (PBS)¹⁶⁴, as well as a GA-rich sequence adjacent to SL3 (unpublished data). More importantly, excessive packaging of spliced viral RNA caused by deletion of this GA-rich region could be corrected by the MP2 mutation (data not shown). Thus, MP2 is capable of repairing the defects in viral RNA packaging that are caused by mutation of RNA packaging signals within the 5' UTR. This may involve binding of modified Gag protein containing MP2 to viral RNA elements distinct from 5' packaging signals. This might then re-establish selectivity for mutated viral RNA.

The role of p2 in RNA packaging is also suggested by one study demonstrating that the presence of HIV-1 p2 within HIV-1/HIV-2 chimeric Gag viruses significantly enhanced the packaging of HIV-1 versus HIV-2 RNA¹³¹. Since viral RNA is recruited into virus particles prior to the processing of Gag by PR, this indicates that p2 may regulate viral RNA packaging in concert with the downstream NC domain through either

direct interaction with viral RNA or indirectly by helping NC to adopt a correct conformation.

Since maximal rescue of the Δ DIS deletion was seen when all four compensatory mutations were present, the correction of viral RNA packaging by the MP2 mutation may not represent the sole mechanism for compensation. Consistent with this belief, DIS sequences may have been involved in the regulation of HIV-1 reverse transcription^{25,214,248}, viral protein translation³⁵ and other activities in either a direct or indirect manner. The DIS can also participate in the overall folding of the HIV-1 5' UTR that can then assume distinctive conformations and perform more than one function, depending on the stage of viral replication^{1,123,217}. Conceivably, the DIS may be necessary for multiple steps of the virus life cycle. This may explain the fact that the Δ DIS mutant was able to infect a significant proportion of cells in the MAGI assay, but failed to establish a productive infection during continuous culture. It is also possible that compensatory mutations may have improved the function of viral components other than the DIS to stimulate viral replication. For example, a replication defect caused by insertion of an AUG translation initiation codon into HIV-1 5' UTR was overcome by second-site mutations within Env protein which presumably improved the Env function⁶⁴.

Interestingly, a recent article by Hill *et al.*¹¹⁶ reported that the HIV-1 DIS stem-loop was dispensable for viral replication in PBMCs, which is contrary to the non-infectiousness of our Δ DIS mutated viruses in CBMCs (Fig. 5.2B). This apparent discrepancy could be attributable to the fact our Δ DIS deletion lacked the complete DIS stem-loop, including the palindrome, whereas the DIS mutants described in the Hill *et al.* paper contained either the wild-type or an arbitrary palindrome sequence in place of SL1.

In summary, we have demonstrated that HIV-1 is able to replicate efficiently in the absence of the DIS through modification of Gag protein sequences. The modified Gag did not restore wild-type RNA dimerization but was able to augment the selective packaging of full-length versus spliced viral RNA molecules into virus particles.

CHAPTER 6

GENERAL DISCUSSION

The discussion of the articles presented in this thesis and their relevance to the current status of the HIV-1 RNA dimerization and packaging fields will be combined with the background literature in Ch. 1 to compile a review, which will be written by myself and submitted to the journal *Retrovirology*. The authors will be **R.S. Russell, C. Liang, and M.A. Wainberg.**

6.0 GENERAL DISCUSSION

The significance of the results in this thesis has been discussed in detail within each chapter. I will now describe how the data fit into the overall picture of HIV-1 RNA dimerization and packaging. There are two main themes that run throughout this thesis, RNA-RNA and RNA-protein interactions. Although these two types of interactions represent different concepts, the work described in this thesis clearly suggests that we have to consider both when studying either dimerization or packaging. The greater goal of our research is to identify targets for drug therapy, so I will also discuss how current research by ourselves and others might lead to the development of novel anti-HIV drugs.

6.1 RNA-RNA Interactions Involved in Dimerization and Packaging

By the time the projects described in this thesis were initiated, most of what is now known about HIV-1 RNA dimerization had already been discovered. The nucleotide sequences responsible for the initiation of dimerization, i.e. the DIS, had been identified^{147,251}, the involvement of PR and NC proteins in this process had been proven^{91,153,204}, and the biological significance of the DIS and genome dimerization had been demonstrated^{25,50,108,112,134,151,152,161,163,214}. However, the observation that DIS mutants still contained some dimerized genomic RNA, combined with the fact that RNA dimers extracted from DIS mutants were as thermostable as those from wild-type viruses^{25,50,108,151,239,248}, raised the possibility that RNA sequences outside the known DIS might also be required for proper HIV-1 RNA dimerization. Because EM analysis of

HIV-1 genomic RNA pointed to a second dimer linkage at the extreme 5' region of the genome¹¹⁸, we looked for sequences upstream of the DIS that could contribute to RNA-RNA interactions in the dimer structure. In Ch. 2, we showed that mutation of GU-rich sequences in the poly(A) and U5-PBS stems caused deficient RNA dimerization and packaging, and that these defects correlated with diminished virus replication capacity, thereby stressing the importance of the identified sequences and their biological significance to HIV-1 replication. Our data suggest that these GU-rich sequences might be directly involved in RNA-RNA interactions between the two copies of viral RNA. This work represents the first *in vivo* evidence that specific sequences outside the known DIS can play a role in the HIV-1 RNA dimerization process; similar findings were reported for other retroviruses^{85,169,170,210,266,267}. Therefore, we proposed that the involvement of multiple RNA regions in retroviral dimer linkages may be a common theme among retroviruses.

The results in Ch. 2 also established that RNA sequences upstream of the DIS could influence HIV-1 RNA dimerization. Since these new regions, and the DIS, are also important for genome packaging^{25,50,51,62,112,151,187,189,214}, we asked whether other known packaging signals might also play a role in the dimerization process. In Ch. 3, we mutated RNA sequences in both the major packaging signal, SL3, and adjacent GA-rich sequences, and found that the structure, but not the actual RNA sequence, of SL3 is needed for both efficient RNA packaging, and genome dimerization. Likewise, mutating GA-rich sequences downstream of SL3 severely diminished RNA dimerization and packaging, although it is not clear how these sequences are involved. In Ch. 4, we showed that complete deletion of SL3 caused defects in viral RNA dimerization and

packaging as well as viral replication capacity. However, as previously seen with the DIS¹⁶¹⁻¹⁶³ and GU-rich (Ch. 2) deletion mutants, these defects were corrected by second-site mutations in Gag that were acquired during long-term culture. At this point, it is still unclear why second-site mutations identified in the context of other HIV-1 mutants were not able to correct dimerization defects, while those identified together with the SL3 deletion mutant were able to do so. We assume that the role of SL3 in the dimerization process is dependent on NC, and that these second-site mutations restored the NC-mediated effects of the SL3 region on dimerization.

Regardless of the mechanisms whereby the poly(A), U5-PBS, SL3, and GA-rich sequences each contribute to the dimerization process, all of these RNA elements can affect overall RNA dimerization. In HIV-2, elements both upstream and downstream of the major splice donor have been shown to influence RNA dimerization *in vitro*¹⁴⁶. Based on the identification of novel RNA sequences involved in dimerization, we now propose a model to explain RNA dimerization in HIV-1. Sequence analysis of the HIV-1 5' UTR, and even the RNA sequences within the *gag* gene, suggests the possible formation of a number of other stem-loop structures²¹⁷. So we propose that the kissing-loop model applies *in vivo*, in which the initial unstable kissing complex is formed via the DIS, and is then bound by NC to melt the DIS stems and re-anneal the strands to form an extended duplex. However, once the RNA molecules are brought into close proximity, and are extensively bound by NC, the stems of other stem-loop structures, such as the poly(A) hairpin, or SL3, might also be melted by NC. These sequences could then form further extended base-paired regions on both sides of the DIS. This model may resemble a double zipper, which starts at the DIS, and is then pulled in opposite directions to close

the zipper from the middle out. Such a model would explain the earlier EM observations that the dimer linkage structures of some viruses appeared to cover an area of approximately 400-600 nucleotides in length¹⁴⁴. Unfortunately, this model would be difficult to test *in vivo*, since deleting all of these putative dimerization sequences would undoubtedly interfere with other functions attributed to the 5' UTR, and the virus would probably would not package enough RNA for gel analysis.

The alternative is that different RNA sequences might instead be involved in higher order intrastrand structures that favor the dimerization of the two RNA molecules. Such a model has been proposed¹²³, and is supported by numerous *in vitro* dimerization studies conducted on HIV-1, HIV-2, and SIV RNA^{1,71,122,123}. The model proposes that the HIV-1 5' UTR can form two alternating conformations, termed the long-distance interaction (LDI) structure and the branched multiple hairpin (BMH) structure. The LDI conformation is believed to exist when the RNA is in a monomer form, and is thought to form a long extended base-paired structure with almost all of the proposed stem-loop sequences buried. This structure is thought to be favored during certain steps of the life cycle, and to protect viral RNA from degradation due to extensive base-pairing. In this model, NC is thought to bind the LDI structure to induce a switch to the BMH structure, in which the DIS and ψ would then be exposed and be able to mediate dimerization and packaging. The existence of such a 'riboswitch' is an attractive hypothesis, especially since similar mechanisms have recently been proposed to account for previously unexplained results in the field of gene regulation¹³⁹, however, there is currently no *in vivo* evidence to support such a model in the case of retroviruses. Others have recently attempted to explain our *in vivo* data (Ch. 3) by applying their riboswitch model to our

mutant sequences and found that our results correlate well with the outcomes that they would expect (B. Berkhout, personal communication). It will be interesting to determine whether this model can explain other HIV-1 mutants, and shed light on mechanisms whereby certain sequences can affect the dimerization process.

One other possible explanation exists that might explain our observation that numerous RNA sequences affect the dimerization process. As mentioned above, most of the sequences we have shown to be involved in dimerization are also known to affect viral RNA packaging. And, *in vitro* dimerization depends on the concentration of viral RNA¹⁸³. If this concentration-dependence also exists *in vivo*, then any RNA or amino acid mutation that affects RNA packaging could theoretically decrease the RNA concentration at the site of assembly, thereby indirectly impacting on dimerization. However, this explanation also implies that genomic RNA can be packaged in the form of a monomer, and raises the question of the link between dimerization and packaging, which will be discussed in detail below.

As stated, both *in vitro* and *in vivo* approaches have been used to study retroviral RNA dimerization. However, contradicting results may be obtained depending on the method employed. In cell-free experiments, less than 5% of the viral genome is placed into a tube and heated in the absence of the proteins, membranes, and ion concentrations that might exist in a cell or virus in which dimerization normally takes place. These small fragments of RNA are then run on a gel and scored on the basis of whether or not they formed dimers; the outcome of such experiments are so dependent on experimental conditions that the interpretation of results is difficult. As mentioned, the true location of the DIS went unresolved for some time, because *in vitro* assays were performed in the

absence of magnesium. However, *in vivo* methods can also produce misleading results, although they involve the full-length viral genome and an RNA dimer that was formed under physiological conditions. If a mutation is made in viral RNA, and no effect results on either dimerization or packaging, it can be concluded that the mutated sequences are not involved in the dimerization or packaging processes. However, mutated sequences that do compromise RNA dimerization or packaging may act through direct mechanisms, or may even affect the overall structure and folding of the RNA itself. In this case, mutations may indirectly cause defects in dimerization or packaging, even though the mutated sequences may not be involved in interactions relevant to these processes.

Why have I gone through this long-winded technical comparison of these techniques – because, I think we may soon have a better way to study the structure and function of HIV-1 RNA. One group is now developing a novel *in vivo* chemical probing protocol that takes advantage of the chemical modification reagents that are commonly used for *in vitro* analysis of RNA structure. However, these experiments are performed on virus-producing cells. Hence, structural analysis of viral RNA, that is normally carried out *in vitro* on short fragments of artificially transcribed RNA, can now be performed on *in vivo*-generated authentic HIV-1 genomic RNA. This method also allows comparisons of cellular and virion-derived HIV-1 RNA. (J.-C. Paillart and R. Marquet, personal communication). Such a method represents a compromise between classic *in vitro* and *in vivo* approaches and may lead to agreement on the true structure of the HIV-1 leader and on which RNA substructures are involved in dimerization. Furthermore, this method might also have application in regard to *in vivo* foot-printing that might allow the study of RNA-protein interactions in the context of virus-producing cells.

6.2 RNA-Protein Interactions Involved in Dimerization and Packaging

In line with the goal of identifying novel RNA and protein sequences involved in HIV-1 RNA dimerization and packaging, we also studied the DIS and its putative interactions with Gag. The DIS has been shown to affect viral replication^{25,50,112,134,151,152,214,248}, reverse transcription^{10,14,214,248}, RNA dimerization^{50,150-152,238,248}, and packaging levels^{25,50,112,151,187,214,238}, as well as the specificity of viral RNA packaging⁵⁰. However, previous work in our lab has shown that defects in these functions caused by large deletions in the DIS can be corrected by a series of compensatory point mutations identified in MA, CA, p2, and NC¹⁶¹⁻¹⁶³. These findings imply that the RNA sequences comprising the DIS interact in some way with these domains of Gag, and that when these RNA sequences are mutated, the virus will acquire, over long-term culture, adaptive mutations that potentially restore putative RNA-protein interactions that might have been disrupted. However, the mechanisms whereby these compensatory point mutations correct for these defective virus functions have remained elusive. To better understand these mutations, we created two other DIS deletions and combined them with various combinations of the previously identified compensatory point mutations. In Ch. 5, we showed that these mutant viruses, Δ Loop (lacking the loop region of SL1) and Δ DIS (lacking the complete SL1) displayed defects in replication, RNA dimerization, and packaging, but that these defects were largely corrected by the compensatory point mutations in Gag. Even a virus completely lacking the DIS, e.g. Δ DIS, which never showed any signs of viral growth in tissue culture, could replicate to significant levels in

the context of the compensatory mutations. Although this Δ DIS virus was severely defective in genome dimerization, the compensatory mutations were able to restore virus replication without correcting the dimerization defects!

Since Gag can bind specifically to the DIS region, it is assumed that mutations within viral RNA might result in disruptions of the RNA-Gag interaction. Potentially, compensatory point mutations within Gag might then either restore specific interactions between the mutated viral RNA and Gag, or alternatively, induce Gag to adopt a different conformation that might bind better to the mutated RNA. Using an RNase protection assay that allows discrimination between spliced versus genomic viral RNA that is packaged into virions, we found that our DIS mutant viruses packaged increased levels of spliced viral RNA compared to wild-type. Such an increase in packaging of spliced RNA indicated a defect in packaging specificity that had previously been shown to correlate with decreased levels of packaged genomic RNA and also resulted in decreased virus replication capacity⁵⁰. More importantly, we observed that all of the combinations of compensatory point mutations that we tested corrected this packaging specificity defect, except for one that did not include the T12I mutation in p2, i.e. MP2. We subsequently showed that this point mutation alone was capable of correcting defective packaging specificity for both the Δ DIS and Δ Loop deletions, and, also that MP2 by itself could restore replication capacity to the Δ DIS mutant virus. This was a surprising finding because NC is believed to be the domain responsible for the overall packaging activities of Gag, and only one other report had ever shown that the p2 spacer peptide could have an effect on packaging specificity¹³¹. Our data indicated that a single amino acid residue within the p2 spacer peptide could play a role in the exclusion of spliced viral RNA from

the virus particle, and we subsequently showed that mutation of this position in the context of wild-type 5' RNA sequences also caused an increase in the levels of spliced viral RNA that were encapsidated into wild-type virus. We proposed that either the amino acid at position 12 in p2 binds directly to viral genomic RNA to facilitate specific encapsidation or that it somehow contributes to the overall structure of Gag in a manner that allows Gag to bind with high affinity to genomic RNA, while simultaneously excluding spliced RNA from being packaged.

Since publication of that work (Ch. 5), I have tested the effects of the MP2 point mutation on the packaging specificity of the BH10-LD3 and BH10-LD4 DIS deletion mutants in which MP2 was originally identified, and found that these DIS mutants also package higher than normal levels of spliced viral RNA, and that MP2 corrected these defects. Thus, MP2 somehow restores the ability of Gag to specifically package DIS-deleted full-length genomic RNA. This is consistent with the notion that MP2 is an adaptive mutation that can compensate for the disruption of RNA-protein interactions that were lost when the RNA sequences were deleted in LD3 and LD4. However, MP2 also corrects packaging specificity for the Δ DIS and Δ Loop, two viruses that did not generate MP2 on their own. Hence, the broad activity of MP2 might be centered around SL1, regardless of the nature of the DIS mutation.

The p2 spacer peptide is a hotspot for the appearance of compensatory point mutations, and MP2 has itself been identified during forced evolution studies on a number of our 5' RNA-mutated viruses. The broad activity of MP2 in the context of the DIS mutants prompted us to test whether these other mutant viruses also had defects in packaging specificity, and whether MP2 could correct for these as well. The Δ GU3

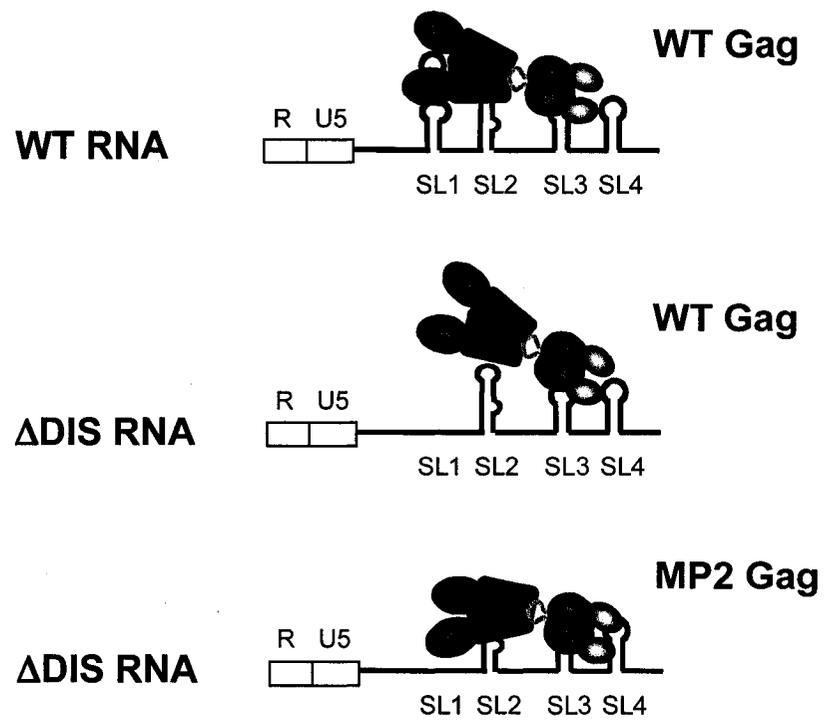
mutant, lacking GU-rich sequences in the lower stem of the U5-PBS complex (Ch. 2), and the MD2 mutant, lacking GA-rich sequences adjacent to SL3 (Ch. 3), also generated the MP2 compensatory point mutation during long-term culture. When RNase protection assays were performed on RNA from these mutant viruses, significantly higher levels of packaged spliced viral RNA than wild-type were obtained, and MP2 was able to repair this deficit in both instances. These results suggest that the effect of MP2 is, in fact, not centered around the DIS, but is even broader and relates to any 5' UTR mutant that we test. Even more surprising was the fact that the M3 mutant, lacking SL3 and upstream A-rich sequences, generated an A11V compensatory mutation within p2 over long-term culture (Ch. 4) that had almost no effect on packaging specificity, whereas MP2 restored the packaging specificity of M3 to wild-type levels (B.B. Roy, R.S. Russell, *et al.*, manuscript in preparation). These results strongly suggest that the p2 spacer peptide is critical for the specific packaging of HIV-1 genomic RNA, and these effects on packaging specificity correlate with virus replication capacity. Spacer peptides similar to p2 are conserved among other lentiviruses such as HIV-2, SIV, BIV, and EIAV; therefore, it will be interesting to test whether this effect on packaging specificity is a common feature of all lentiviral spacer peptides.

MP2 has broad activity in the correction of packaging specificity, but the mechanism(s) involved remain obscure. As mentioned above and in Ch. 1, the actual RNA structure is still debated, and we don't know whether viral RNA exists as a monomer or dimer when it is first recognized by Gag. Although Gag is responsible for the specific packaging of viral genomic RNA, its actual structure is also poorly understood, nor is it known what multimeric state Gag might be in when it first

encounters viral RNA. Even though a model may be premature, we propose one for purpose of discussion which shows that the NC domain of Gag can bind specifically to the major packaging signal, SL3, in an orientation that allows other domains of Gag to interact with regions of genomic RNA, such as SL1 or the poly(A) hairpin; these interactions might stabilize the overall genomic RNA-Gag interaction (Fig. 6.1, top). Deletion of the DIS (or other sequences) might disrupt these stabilizing effects, resulting in a reduced affinity for genomic RNA (Fig. 6.1, middle). This decreased affinity for genomic RNA could explain the extraneous packaging of spliced viral RNA, and probably cellular RNA as well, but we have not yet tested this possibility. In the context of the MP2 mutation, Gag might then adopt an alternate conformation that restores a high affinity interaction with mutated genomic RNA (Fig. 6.1, bottom). Such an 'induced fit' model, in which RNA-protein interactions involve conformational changes in both the RNA and protein, has recently gained interest, and is thought to apply for other HIV-1 RNA-protein interactions such as Tat-TAR and Rev-RRE (Reviewed in Williamson²⁷⁷). Interaction with the capsid protein can alter RNA structure and affects *in vitro* assembly pathways¹²⁴ in another RNA virus. Such 'induced fit' interactions might be required for proper Gag multimerization and RNA packaging during HIV assembly as well. These questions are now being addressed at the *in vitro* level by a postdoc in Dr. Wainberg's lab, who has generated results that provide insight into such putative induced fit interactions between Gag and viral RNA (A. Roldan, R.S. Russell, *et al.*, manuscript in preparation). Such *in vitro* analyses might also answer the question of whether a Gag protein containing MP2 really does have higher affinity for genomic versus spliced viral RNA (A. Roldan, unpublished).

Figure 6.1. Model of the MP2-induced alternate RNA-protein interaction. The NC domain of Gag binds specifically to the major packaging signal, SL3, in an orientation that allows other domains of Gag to interact with other regions of genomic RNA, such as SL1 or the poly(A) hairpin, to stabilize this RNA-protein interaction (top). Deletion of the DIS (or other sequences) could disrupt these stabilizing effects, resulting in a reduced affinity for genomic RNA (middle). In the context of MP2, Gag might itself adopt an alternate conformation that restores a high affinity interaction with mutated genomic RNA (bottom).

Figure 6.1. Model of the MP2-induced alternate RNA-protein interaction.



Recent data from our group have raised other questions regarding the mechanism(s) whereby MP2 can correct packaging specificity in the context of deletion mutations in the 5' UTR of viral RNA. Since we have proposed that the effect of MP2 was through an alternate induced fit interaction between the MP2 Gag and mutated 5' RNA sequences, then mutations outside of the 5' UTR that also affect packaging specificity, e.g. in NC, should not be overcome by MP2. To test this hypothesis, Dr. Liang's lab has generated a panel of HIV-1 mutants containing amino acid substitutions or deletions in the zinc fingers and basic amino acid residues that are known to be important for the RNA packaging activity of Gag. As expected, these mutants packaged lower levels of genomic RNA and higher levels of spliced viral RNA compared to wild-type, indicating a packaging specificity defect. Surprisingly however, MP2 was able to correct the aberrant packaging of spliced viral RNA in the context of these NC amino acid mutations (B.B. Roy, R.S. Russell, *et al.*, manuscript in preparation). This observation argues against our assumption that MP2 acts to induce an alternate interaction between the MP2 Gag and the mutated 5' RNA sequences, and suggests that the binding of MP2 Gag might instead induce a general conformation in the RNA that creates a higher affinity interaction between the protein and the RNA, regardless of RNA sequence (R. Marquet, personal communication). The paradox is that if MP2-Gag has a higher affinity for viral genomic RNA, then why is it not contained in the wild-type HIV-1 consensus sequence. We cannot confidently explain these results; one possibility is that it might be detrimental for the virus if Gag binds too tightly to the viral RNA, and the RNA-protein interaction may

need to be flexible in order to allow conformational rearrangements necessary for other aspects of the life cycle.

Another interesting question relates to the need to package spliced viral RNA species in the first place. In wild-type viruses, 90-95% of the viral RNA in the virion is full-length genomic. Aberrant packaging of spliced RNA in mutated viruses may be due to a requirement of the virus to fill a certain amount of space that is normally occupied by full-length genomic RNA. The presence of RNA or Gag mutations that decrease levels of genomic RNA in virions may cause the virus to incorporate spliced viral RNA to fill this space. However, the MP2 mutation causes spliced RNA to be excluded from the virion, without affecting levels of genomic RNA. Thus, overall RNA packaging and specific genome packaging are separate activities, possibly controlled by different domains within Gag.

The debate about the link between dimerization and packaging is relevant here since most 5' UTR mutant viruses that show decreased RNA dimerization also package lower levels of genomic RNA (along with higher levels of spliced RNA) compared to wild-type. Possible explanations are: (i) Viral RNA is normally packaged in the form of a dimer, but dimerization deficient mutants result in weak or less recognizable dimers being packaged into fewer viruses; hence, lower levels of genomic RNA are packaged per ng of p24-normalized virus. Spliced RNA is then incorporated non-specifically by assembling Gag molecules that have not established a high affinity interaction with a stable dimer in the cytoplasm. (ii) Mutated dimerization sequences weaken interactions between the two copies of full-length genomic RNA and allow the formation of weak heterodimers that consist of genomic and spliced RNA to be packaged, explaining both

the decreased levels of packaged genomic RNA and increased packaging of spliced RNA. Indeed, dimerization gels that analyzed mutants that package high levels of spliced viral RNA revealed only the high molecular weight bands representing dimers and monomers; hence, spliced RNA must migrate in a complex with one of both of these RNA species. (iii) The third possibility is that viral RNA is always packaged as a monomer and that it dimerizes during the assembly or budding process. This would explain our observations that DIS-deficient viruses package as much RNA as wild-type in the presence of the compensatory point mutations, but that this RNA is predominantly monomeric on a gel. Spliced RNA would then be packaged when the specific packaging of monomeric genomic RNA became compromised. Of course, these monomers might also have been packaged as weak dimers that dissociated during virus purification and RNA extraction.

The link between dimerization and packaging has been debated for some years^{25,50,90,91,151,236-238,248}, and, in my opinion, explanations (i) and (ii) which argue that dimerization must be a prerequisite for packaging make the most sense. Genomic RNA can be packaged as monomers^{50,236,237,247,248}, or at least as very weak dimers, but mutant viruses that exhibit dimerization defects do not grow as well as wild-type viruses. The fact that our Δ DIS-MP2 virus can replicate in tissue culture, despite being severely compromised in genome dimerization, is evidence that efficient dimerization is not required for packaging or replication. Other sequences that affect dimerization, such as those reported in Chs. 2-4, may form a weak dimer that allows the RNA to be recognized and packaged. The contribution of the DIS might then be to significantly increase the efficiency of the dimerization process, resulting in more efficient packaging and

replication. Comparison of cellular versus virion-derived RNA, using the *in vivo* chemical probing protocols mentioned above, might finally resolve this issue, since viral RNA may already be dimerized in the cytoplasm (J.-C. Paillart, personal communication).

6.3 Novel Therapeutic Targets for HIV-1

Research aimed at basic RNA-RNA and RNA-protein interactions not only contributes to the basic understanding of the retroviral life cycle, but might also potentially lead to the identification of novel targets for anti-HIV therapy. I now ask whether any of my observations may impact on HIV disease.

6.3.1 The DIS as a Therapeutic Target?

The unique aspect of the dimerized genome, along with the fact that this feature is so well conserved within the *Retroviridae* family, has prompted retrovirologists to consider this process as a target in HIV therapeutics. Indeed, the DIS was thought to be a potential therapeutic target, and antisense molecules were directed at this region of viral RNA^{251,252}, although there has been no practical outcome of this work. Other approaches directly target the HIV-1 kissing-loop complex which resembles the eubacterial 16S ribosomal aminoacyl-tRNA site, i.e. the target of aminoglycoside antibiotics such as paramycin and neomycin⁸², both of which specifically bind to the kissing-loop complex. Antibiotic modifications with high affinity and specificity for the DIS may be designed in

the future, although, the efficacy of such approaches might be limited, since HIV can replicate in the face of mutations that decrease genomic dimerization by more than 50% (Ch. 5).

RNA interference (RNAi) is a novel mechanism that regulates gene expression using small interfering RNAs to direct the targeted degradation of RNA in a sequence-specific manner. A quick PubMed search would lead one to think that RNAi, or some gene therapy-based application thereof, will cure all known diseases. Although RNAi is a powerful research tool, I am not convinced that its therapeutic potential will not fizzle. Several reports show that specific degradation of HIV-1 RNA is possible in an infected cell (Reviewed in Lee and Rossi¹⁵⁶), and inhibition of p24 levels by as much as 4 logs has been achieved by transfection of psiRNA directed against HIV-1 *tat* and *rev*¹⁵⁵. DNA vectors are currently being engineered that will allow long-term production of siRNAs for use against chronic diseases, such as HIV. I believe, in spite of what I stated above, that the DIS might be a good candidate for sequence-specific targeting of HIV by RNAi therapy. Why? The DIS is highly conserved among naturally occurring virus isolates, and, due to its position upstream of the major splice donor, it is contained in all HIV-1 RNA transcripts, both spliced and unspliced. Effective DIS-directed degradation of HIV RNA should confer the same viral phenotype as observed with our Δ DIS mutant, which never showed any signs of virus replication in either permissive T cell lines or CBMCs. One concern is how accessible certain viral RNA sequences might be to siRNAs. For example, complex secondary structures might cause some sequences to be buried and therefore inaccessible to siRNA. This might not be a concern with DIS-directed RNAi, since the DIS contains a 6 nt palindromic sequence that is believed to initiate the

dimerization process by binding to the same sequence on another molecule of genomic RNA. If two 6 nt stretches of RNA can find each other on two 9200 nt strands of highly structured RNA, they should also be accessible to siRNAs.

All of these DIS-directed strategies rely on specifically targeting the viral RNA itself, which might not be practical, due to our inadequate knowledge of the overall structure of the HIV-1 5' region. The fact that RNA sequences such as SL1 and SL3 are known to form relevant RNA-protein interactions raises the possibility that the protein component of these interactions might also provide potential targets for anti-HIV therapy. Such approaches are currently being explored in collaborative research aimed at designing inhibitors of the TAR-Tat RNA-protein interaction²⁷⁸.

6.3.2 p2 as a Therapeutic Target?

As discussed in Ch. 1, assembly inhibitors are probably next on the horizon for therapeutic approaches to combat HIV, and three recent studies have demonstrated proof of concept for such a class of molecules. The first compound, CAP-1, binds directly to the HIV-1 CA protein, and was shown to inhibit virus infectivity in a dose-dependent manner²⁶³. Virus particles generated in the presence of this inhibitor displayed heterogeneous size and abnormal core morphology, suggesting a novel antiviral mechanism. The second compound, PA-457, is termed a maturation inhibitor, since it acts by disrupting a late step in Gag processing, specifically, the conversion of the CA precursor (p25) to mature CA (p24)¹⁵⁸. Virions from PA-457-treated cultures were noninfectious, and, as with CAP-1, also exhibited aberrant particle morphology.

Although the exact mechanism of action of this compound is unknown, it is believed to block the CA-p2 cleavage site, since a mutation conferring resistance to the drug was mapped to the first amino acid in p2, at the p25/p24 cleavage site. More recently, a third compound, reported to be a small-molecule inhibitor of virion maturation²⁸⁹, showed a similar inhibitory profile as did PA-457.

p2 is required for higher-order multimerization of Gag¹⁹⁹, and deletion of p2, or even of its amino terminal residues, virtually abolishes virus production^{2,141}. Proper removal of p2 from the C-terminus of CA must take place during virus morphogenesis, otherwise the spherical CA shell fails to form a mature conical core^{106,275}. We have identified key amino acid residues within p2 that affect Gag complex formation¹⁶⁵ and specific packaging of viral genomic RNA (Ch. 5), both considered essential events in the assembly process. Theoretically, small-molecule or rationally designed inhibitors that specifically block the activity of these key interactions in which p2 is involved might result in the phenotypes mentioned above, and might represent potential assembly inhibitors. Based on the activities of the compounds described above, and the strong phenotypes that we, and others, have reported in viruses containing mutations in p2, I believe that the development of an inhibitor targeting this assembly domain of Gag is a realistic possibility. Such an inhibitor would have an advantage over the three drugs described above, since none of them showed more than a 50% decrease in virus production, and, therefore, might still allow for newly released virions to carry out early steps of the virus life cycle, e.g. reverse transcription, the stage at which drug-resistant mutations commonly arise. Theoretically, drug resistance could quickly develop to inhibitors such as PA-457, in which case a single mutation in p2 was shown to result in

wild-type replication in the presence of drug. In contrast, a tight-binding inhibitor of p2 might completely block virus production and the virus might not have the opportunity to undergo reverse transcription, thereby decreasing the risk of drug resistance.

6.3 Final Summation

I believe that the research presented in this PhD thesis has answered important basic questions regarding RNA-RNA and RNA-protein interactions that occur during the HIV-1 RNA dimerization and packaging processes. Relevant questions have also been raised, the answers to which will hopefully further our knowledge and understanding of these processes, and, more importantly, help us to design novel therapies that might potentially block critical steps in the virus life cycle. It is my hope that my identification of novel RNA and protein sequences involved in two important aspects of the retroviral life cycle, and the interactions in which these sequences are involved, will lead to novel drugs that will make a difference for HIV-infected individuals.

CONTRIBUTION TO ORIGINAL KNOWLEDGE

The work presented below is a summary of the major findings contained within Chs. 2-5 of this thesis, and has all been published in refereed scientific journals^{230,232-234}.

Chapter 2. Prior to the initiation of the projects detailed in this thesis, the dimerization initiation site (DIS) was shown to be the key region of the HIV-1 5' RNA involved in the dimerization process. However, three pieces of evidence strongly suggested that the DIS did not act alone to facilitate this process. First, EM analysis of HIV-1 RNA showed the 5' region to contain two stable contact points between the two RNA molecules, suggesting that the DIS was not the only region of RNA that was dimerized. Second, all DIS mutant viruses studied still contained a significant amount of dimerized genomic RNA within the virion, meaning that the genome was able to dimerize in the absence of the DIS sequences. And third, RNA dimers isolated from these DIS-mutated viruses were as thermostable as those isolated from wild-type viruses. These observations implied that RNA-RNA interactions outside the DIS must contribute to the interaction between the two RNA molecules. However, at the time that I began my PhD, no RNA sequences other than the known DIS had been shown to affect HIV-1 RNA dimerization *in vivo*. So first, I set up and established this *in vivo* dimerization assay in our lab, and then used this assay to study RNA sequences involved in the dimerization of HIV-1 RNA. The major contribution of the work in Ch. 2 was that we demonstrated for the first time that specific RNA sequences upstream of the DIS could affect the levels of dimerized RNA contained within the virus. Specifically, we showed that the GU-rich sequences within the poly(A) and U5-PBS stems were required for normal genome dimerization. The fact that these identified sequences were located upstream of the DIS explains the previous EM data, and these sequences may form the extreme 5' interaction that was observed in the EM pictures. We also showed that this dimerization deficit was not corrected by RNA mutations that preserved local RNA structures, such as the poly(A) hairpin, and was overcome to only a limited extent by compensatory mutations within Gag that were identified after long-term culture. These results imply that the GU-rich sequences

contribute to RNA dimerization by mediating direct RNA-RNA interactions in the dimer structure. These data represent the first *in vivo* identification of specific RNA sequences outside the DIS that are able to contribute to the dimerization process²³².

Chapter 3. Now that we had demonstrated that sequences other than the DIS could indeed play a role in the HIV-1 RNA dimerization process, we wondered if other regions of the 5' UTR could also affect genome dimerization. It turns out that the DIS, and the Poly(A) and U5-PBS RNA sequences that we showed to affect dimerization (Ch. 2), had all been previously shown to represent important packaging signals as well. Since numerous reports had already proposed a link between dimerization and packaging, we wondered if other known packaging signals, such as SL3, might also be involved in the dimerization process. One study had also shown that an antisense oligonucleotide targeting SL3 and a downstream GA-rich region could inhibit dimerization of HIV-1 RNA fragments *in vitro*. In Ch. 3 we showed that the structure, but not specific RNA sequence, of SL3 is needed not only for efficient viral RNA packaging but also for dimerization. We also showed that mutation of the GA-rich sequences downstream of SL3 also severely diminished viral RNA dimerization and packaging. These data represent the first *in vivo* evidence that RNA sequences downstream of the DIS can play a role in RNA dimerization. These findings, combined with the data in Ch. 2, raise the possibility that dimerization is initiated by the DIS, but that the RNA dimer also involves RNA-RNA interactions on both sides of the DIS. If these putative interactions are still intact in the context of a DIS- RNA, then these interactions could maintain a low level of thermostable dimer that is observed in DIS-mutated viruses. Earlier *in vitro* dimerization studies proposed that the formation of G-quartets might also contribute to the dimerization of retroviral RNA. However, this hypothesis was later refuted, based on the fact that potassium ions, which are known to stabilize G-quartet, had no effect on HIV-1 RNA dimerization *in vivo*. To confirm this work, we specifically mutated a series of G nucleotides within the *gag* gene and tested whether or not these mutations had any effect on RNA dimerization. We indeed found that the absence of the G nucleotides had no effect on the levels of dimerized RNA contained in virus particles, confirming that G-

quartet formation in *gag* gene sequences does not contribute to the formation of the RNA dimer²³³.

Chapter 4. A former student in our lab had performed forced evolution studies with a different series of SL3-mutated viruses and had identified compensatory point mutations that could restore replication and RNA packaging activity to these mutants. Given that we now had evidence for the involvement of SL3 in the dimerization process, we asked whether these other SL3 mutations caused dimerization defects that contributed to the reduced replication capacity of these mutant viruses. Since compensatory mutations that restored replication to these viruses had been identified, we could test whether these compensatory mutations could correct RNA dimerization in a *trans* fashion. In agreement with the results in Ch. 3, we found that complete deletion of SL3 led to decreases in both viral RNA packaging and dimerization. However, unlike the compensatory mutations identified in the context of DIS-mutated viruses, dimerization defects caused by SL3 mutations were corrected by the compensatory mutations in p2 and NC. These results imply that the role of SL3 in the dimerization process is more dependent on Gag, and that these second-site mutations might act by restoring the NC-mediated effects of the SL3 region on dimerization²³⁰.

Chapter 5. In Ch. 5, we focused our attention away from the RNA-RNA interactions involved in RNA dimerization and packaging, and tried to understand the putative RNA-protein interactions affecting these processes that were evident from our frequent identification of compensatory point mutations in our 5' UTR mutant viruses. Based on our previous studies showing that DIS deletion mutants could be rescued by compensatory mutations in MA, CA, p2, and NC, we created two additional DIS deletion mutations and combined them with various combinations of the compensatory mutations. The first important result to come out of this work was that a virus lacking the complete DIS was noninfectious in tissue culture, confirming the biological relevance of the DIS and the dimeric feature of the HIV-1 genome. However, we were surprised to see that this severely compromised virus was able to replicate to almost wild-type levels in the context of all combinations of compensatory mutations tested. Even in the presence of

only the T12I compensatory mutation identified in p2, our Δ DIS mutant was able to establish a persistent infection in a human T-cell line. This is arguably the most significant finding of my PhD research, since it represents an extreme example of how resilient HIV is. The Δ DIS virus contains a 35 nt deletion in the 5' UTR that we showed to decrease RNA dimerization, packaging, and packaging specificity, and this virus was, to all intents and purposes, dead. However, a single nt substitution, almost a kilobase away from the original deletion, resulting in a single amino acid change, restored replication capacity to this dead virus. The implication of this work extends to the idea of live-attenuated vaccine approaches for HIV, since it shows that even a highly attenuated virus could have the potential to revert to an infectious virus if it is able to replicate well enough to generate as few as one or two nt substitutions. The other important result to come out of the work described in Ch. 5 is that we are now one step closer to fully understanding the mechanism(s) employed by these compensatory mutations to facilitate restoration of replication to our mutant viruses. With the help of Véronique Bériault, a student from Dr. A. Mouland's lab, and Ariel Roldan, I set up and established an RNase protection assay that allowed us to discriminate between genomic versus spliced viral RNA packaged into HIV-1 virions. Using this assay we found that DIS-deleted viruses aberrantly packaged higher levels of spliced viral RNA than did wild-type virus. These data confirmed that the SL1, i.e. the DIS, is required for specific packaging of genomic versus spliced viral RNA. However, the surprising result was that the MP2 compensatory point mutation was able to independently correct this packaging specificity defect. The p2 spacer peptide in HIV-1 Gag had already been shown to affect packaging specificity, but we now showed for the first time that specific amino acid residues within p2 could have major effects on the exclusion of spliced viral RNA from being packaged into the virion. And, we went on to show that this effect was not restricted to the 5' RNA mutants we had been studying, but that different amino acids at the same position in p2 could affect the packaging specificity of the wild-type virus as well. These data provide strong evidence that NC is not solely responsible for the packaging function of Gag, and demonstrates that the p2 spacer peptide might play an active role in packaging of HIV-1 RNA²³⁴.

DEDICATION

This thesis is dedicated to two people who now only exist in our memories. One, an uneducated man, who lived a hard life, in a hard time, growing up during The Depression. He died of cancer at 76 years of age, and probably never really understood what HIV was. The other, an educated woman, who also lived a hard life, in a hard time, growing up during the early years of the AIDS epidemic. She died of AIDS at 30 years of age, and came to understand too well what HIV was. The man was my grandfather, the woman was my friend, they both died while I was doing this PhD. They lived very different lives, and they inspired me in very different ways.

For Pop and Becky

REFERENCES

- 1 **Abbink, T. E., and B. Berkhout.** 2003. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J Biol Chem* **278**:11601-11.
- 2 **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.
- 3 **Accola, M. A., B. Strack, and H. G. Gottlinger.** 2000. Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J Virol* **74**:5395-402.
- 4 **Adam, M. A., and A. D. Miller.** 1988. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* **62**:3802-6.
- 5 **Aldovini, A., and R. A. Young.** 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* **64**:1920-6.
- 6 **Alkhatib, G., C. Combadiere, C. C. Broder, Y. Feng, P. E. Kennedy, P. M. Murphy, and E. A. Berger.** 1996. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**:1955-8.
- 7 **Amara, A., and D. R. Littman.** 2003. After Hrs with HIV. *J Cell Biol* **162**:371-5.
- 8 **Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, and M. F. Summers.** 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* **301**:491-511.
- 9 **Amarasinghe, G. K., J. Zhou, M. Miskimon, K. J. Chancellor, J. A. McDonald, A. G. Matthews, R. R. Miller, M. D. Rouse, and M. F. Summers.** 2001. Stem-loop SL4 of the HIV-1 psi RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J Mol Biol* **314**:961-70.
- 10 **Andersen, E. S., R. E. Jeeninga, C. K. Damgaard, B. Berkhout, and J. Kjems.** 2003. Dimerization and template switching in the 5' untranslated region between various subtypes of human immunodeficiency virus type 1. *J Virol* **77**:3020-30.
- 11 **Armentano, D., S. F. Yu, P. W. Kantoff, T. von Ruden, W. F. Anderson, and E. Gilboa.** 1987. Effect of internal viral sequences on the utility of retroviral vectors. *J Virol* **61**:1647-50.
- 12 **Awang, G., and D. Sen.** 1993. Mode of dimerization of HIV-1 genomic RNA. *Biochemistry* **32**:11453-7.
- 13 **Bai, X., J. D. Chen, A. G. Yang, F. Torti, and S. Y. Chen.** 1998. Genetic co-inactivation of macrophage- and T-tropic HIV-1 chemokine coreceptors CCR-5 and CXCR-4 by intrakines. *Gene Ther* **5**:984-94.

- 14 **Balakrishnan, M., P. J. Fay, and R. A. Bambara.** 2001. The kissing hairpin sequence promotes recombination within the HIV-1 5' leader region. *J Biol Chem* **276**:36482-92.
- 15 **Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dautet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**:868-71.
- 16 **Baudin, F., R. Marquet, C. Isel, J. L. Darlix, B. Ehresmann, and C. Ehresmann.** 1993. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol* **229**:382-97.
- 17 **Bender, M. A., T. D. Palmer, R. E. Gelinas, and A. D. Miller.** 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the gag region. *J Virol* **61**:1639-46.
- 18 **Bender, W., and N. Davidson.** 1976. Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* **7**:595-607.
- 19 **Bender, W., Y. H. Chien, S. Chattopadhyay, P. K. Vogt, M. B. Gardner, and N. Davidson.** 1978. High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J Virol* **25**:888-96.
- 20 **Bennett, R. P., T. D. Nelle, and J. W. Wills.** 1993. Functional chimeras of the Rous sarcoma virus and human immunodeficiency virus gag proteins. *J Virol* **67**:6487-98.
- 21 **Berglund, J. A., B. Charpentier, and M. Rosbash.** 1997. A high affinity binding site for the HIV-1 nucleocapsid protein. *Nucleic Acids Res* **25**:1042-9.
- 22 **Berkhout, B., B. B. Essink, and I. Schoneveld.** 1993. In vitro dimerization of HIV-2 leader RNA in the absence of PuGGAPuA motifs. *Faseb J* **7**:181-7.
- 23 **Berkhout, B., B. Klaver, and A. T. Das.** 1995. A conserved hairpin structure predicted for the poly(A) signal of human and simian immunodeficiency viruses. *Virology* **207**:276-81.
- 24 **Berkhout, B.** 1996. Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol* **54**:1-34.
- 25 **Berkhout, B., and J. L. van Wamel.** 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol* **70**:6723-32.
- 26 **Berkhout, B., and J. L. van Wamel.** 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *Rna* **6**:282-95.
- 27 **Berkowitz, R., J. Fisher, and S. P. Goff.** 1996. RNA packaging. *Curr Top Microbiol Immunol* **214**:177-218.
- 28 **Berkowitz, R. D., J. Luban, and S. P. Goff.** 1993. Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* **67**:7190-200.
- 29 **Berkowitz, R. D., and S. P. Goff.** 1994. Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. *Virology* **202**:233-46.

- 30 **Berkowitz, R. D., M. L. Hammarskjold, C. Helga-Maria, D. Rekosh, and S. P. Goff.** 1995. 5' regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. *Virology* **212**:718-23.
- 31 **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.
- 32 **Bieth, E., C. Gabus, and J. L. Darlix.** 1990. A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis in vitro. *Nucleic Acids Res* **18**:119-27.
- 33 **Bogerd, H. P., B. P. Doehle, H. L. Wiegand, and B. R. Cullen.** 2004. A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc Natl Acad Sci U S A* **101**:3770-4.
- 34 **Borsetti, A., A. Ohagen, and H. G. Gottlinger.** 1998. The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. *J Virol* **72**:9313-7.
- 35 **Brasey, A., M. Lopez-Lastra, T. Ohlmann, N. Beerens, B. Berkhout, J. L. Darlix, and N. Sonenberg.** 2003. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J Virol* **77**:3939-49.
- 36 **Brenner, B. G., J. P. Routy, M. Petrella, D. Moisi, M. Oliveira, M. Detorio, B. Spira, V. Essabag, B. Conway, R. Lalonde, R. P. Sekaly, and M. A. Wainberg.** 2002. Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *J Virol* **76**:1753-61.
- 37 **Bukrinskaya, A., B. Brichacek, A. Mann, and M. Stevenson.** 1998. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. *J Exp Med* **188**:2113-25.
- 38 **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**:6487-97.
- 39 **Campbell, S., and A. Rein.** 1999. In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. *J Virol* **73**:2270-9.
- 40 **Canaani, E., K. V. Helm, and P. Duesberg.** 1973. Evidence for 30-40S RNA as precursor of the 60-70S RNA of Rous sarcoma virus. *Proc Natl Acad Sci U S A* **70**:401-5.
- 41 **Carbon, J., T. E. Shenk, and P. Berg.** 1975. Construction in vitro of mutants of simian virus 40: insertion of a poly(dA-dT) segment at the Hemophilus parainfluenza II restriction endonuclease cleavage site. *J Mol Biol* **98**:1-15.
- 42 **Chan, D. C., and P. S. Kim.** 1998. HIV entry and its inhibition. *Cell* **93**:681-4.
- 43 **Chen, H., and A. Engelman.** 1998. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci U S A* **95**:15270-4.
- 44 **Choe, H., M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P. D. Ponath, L. Wu, C. R. Mackay, G. LaRosa, W. Newman, N. Gerard, C. Gerard, and J. Sodroski.** 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**:1135-48.

- 45 **Cimarelli, A., S. Sandin, S. Hoglund, and J. Luban.** 2000. Rescue of multiple viral functions by a second-site suppressor of a human immunodeficiency virus type 1 nucleocapsid mutation. *J Virol* **74**:4273-83.
- 46 **Cimarelli, A., and J. L. Darlix.** 2002. Assembling the human immunodeficiency virus type 1. *Cell Mol Life Sci* **59**:1166-84.
- 47 **Clavel, F., and J. M. Orenstein.** 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* **64**:5230-4.
- 48 **Clever, J., C. Sasseti, and T. G. Parslow.** 1995. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* **69**:2101-9.
- 49 **Clever, J. L., M. L. Wong, and T. G. Parslow.** 1996. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* **70**:5902-8.
- 50 **Clever, J. L., and T. G. Parslow.** 1997. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* **71**:3407-14.
- 51 **Clever, J. L., D. A. Eckstein, and T. G. Parslow.** 1999. Genetic dissociation of the encapsidation and reverse transcription functions in the 5' R region of human immunodeficiency virus type 1. *J Virol* **73**:101-9.
- 52 **Clever, J. L., D. Mirandar, Jr., and T. G. Parslow.** 2002. RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome. *J Virol* **76**:12381-7.
- 53 **Coffin, J., Hughes, SH, Varmus, HE.** 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 54 **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol* **42**:1-26.
- 55 **Coffin, J. M.** 1996. HIV viral dynamics. *Aids* **10 Suppl 3**:S75-84.
- 56 **Conticello, S. G., R. S. Harris, and M. S. Neuberger.** 2003. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* **13**:2009-13.
- 57 **Cullen, B. R.** 1998. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249**:203-10.
- 58 **Dalgleish, A. G., P. C. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves, and R. A. Weiss.** 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**:763-7.
- 59 **Dannull, J., A. Surovoy, G. Jung, and K. Moelling.** 1994. Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues. *Embo J* **13**:1525-33.
- 60 **Darlix, J. L., C. Gabus, M. T. Nugeyre, F. Clavel, and F. Barre-Sinoussi.** 1990. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol* **216**:689-99.
- 61 **Darlix, J. L., C. Gabus, and B. Allain.** 1992. Analytical study of avian reticuloendotheliosis virus dimeric RNA generated in vivo and in vitro. *J Virol* **66**:7245-52.

- 62 **Das, A. T., B. Klaver, B. I. Klasens, J. L. van Wamel, and B. Berkhout.** 1997. A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *J Virol* **71**:2346-56.
- 63 **Das, A. T., B. Klaver, and B. Berkhout.** 1998. The 5' and 3' TAR elements of human immunodeficiency virus exert effects at several points in the virus life cycle. *J Virol* **72**:9217-23.
- 64 **Das, A. T., A. P. van Dam, B. Klaver, and B. Berkhout.** 1998. Improved envelope function selected by long-term cultivation of a translation-impaired HIV-1 mutant. *Virology* **244**:552-62.
- 65 **Das, A. T., B. Klaver, and B. Berkhout.** 1999. A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *J Virol* **73**:81-91.
- 66 **De Clercq, E.** 2002. New developments in anti-HIV chemotherapy. *Biochim Biophys Acta* **1587**:258-75.
- 67 **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**:384-8.
- 68 **De Tapia, M., V. Metzler, M. Mougel, B. Ehresmann, and C. Ehresmann.** 1998. Dimerization of MoMuLV genomic RNA: redefinition of the role of the palindromic stem-loop H1 (278-303) and new roles for stem-loops H2 (310-352) and H3 (355-374). *Biochemistry* **37**:6077-85.
- 69 **Deng, H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, and N. R. Landau.** 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**:661-6.
- 70 **Dettenhofer, M., S. Cen, B. A. Carlson, L. Kleiman, and X. F. Yu.** 2000. Association of human immunodeficiency virus type 1 Vif with RNA and its role in reverse transcription. *J Virol* **74**:8938-45.
- 71 **Dirac, A. M., H. Huthoff, J. Kjems, and B. Berkhout.** 2001. The dimer initiation site hairpin mediates dimerization of the human immunodeficiency virus, type 2 RNA genome. *J Biol Chem* **276**:32345-52.
- 72 **Doms, R. W., and D. Trono.** 2000. The plasma membrane as a combat zone in the HIV battlefield. *Genes Dev* **14**:2677-88.
- 73 **Doranz, B. J., J. Rucker, Y. Yi, R. J. Smyth, M. Samson, S. C. Peiper, M. Parmentier, R. G. Collman, and R. W. Doms.** 1996. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* **85**:1149-58.
- 74 **Dorfman, T., J. Luban, S. P. Goff, W. A. Haseltine, and H. G. Gottlinger.** 1993. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **67**:6159-69.
- 75 **Dragic, T., V. Litwin, G. P. Allaway, S. R. Martin, Y. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore, and W. A. Paxton.** 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**:667-73.

- 76 **D'Souza, M. P., and V. A. Harden.** 1996. Chemokines and HIV-1 second receptors. Confluence of two fields generates optimism in AIDS research. *Nat Med* **2**:1293-300.
- 77 **Dube, S., H. J. Kung, W. Bender, N. Davidson, and W. Ostertag.** 1976. Size, subunit composition, and secondary structure of the Friend virus genome. *J Virol* **20**:264-72.
- 78 **Duesberg, P. H.** 1968. Physical properties of Rous Sarcoma Virus RNA. *Proc Natl Acad Sci U S A* **60**:1511-8.
- 79 **Embretson, J. E., and H. M. Temin.** 1987. Lack of competition results in efficient packaging of heterologous murine retroviral RNAs and reticuloendotheliosis virus encapsidation-minus RNAs by the reticuloendotheliosis virus helper cell line. *J Virol* **61**:2675-83.
- 80 **Ennifar, E., M. Yusupov, P. Walter, R. Marquet, B. Ehresmann, C. Ehresmann, and P. Dumas.** 1999. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure Fold Des* **7**:1439-49.
- 81 **Ennifar, E., P. Walter, B. Ehresmann, C. Ehresmann, and P. Dumas.** 2001. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol* **8**:1064-8.
- 82 **Ennifar, E., J. C. Paillart, R. Marquet, B. Ehresmann, C. Ehresmann, P. Dumas, and P. Walter.** 2003. HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. *J Biol Chem* **278**:2723-30.
- 83 **Fauci, A. S.** 2003. HIV and AIDS: 20 years of science. *Nat Med* **9**:839-43.
- 84 **Feng, Y., C. C. Broder, P. E. Kennedy, and E. A. Berger.** 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**:872-7.
- 85 **Feng, Y. X., W. Fu, A. J. Winter, J. G. Levin, and A. Rein.** 1995. Multiple regions of Harvey sarcoma virus RNA can dimerize in vitro. *J Virol* **69**:2486-90.
- 86 **Feng, Y. X., T. D. Copeland, L. E. Henderson, R. J. Gorelick, W. J. Bosche, J. G. Levin, and A. Rein.** 1996. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *Proc Natl Acad Sci U S A* **93**:7577-81.
- 87 **Fisher, A. G., B. Ensoli, L. Ivanoff, M. Chamberlain, S. Petteway, L. Ratner, R. C. Gallo, and F. Wong-Staal.** 1987. The sor gene of HIV-1 is required for efficient virus transmission in vitro. *Science* **237**:888-93.
- 88 **Fosse, P., N. Motte, A. Roumier, C. Gabus, D. Muriaux, J. L. Darlix, and J. Paoletti.** 1996. A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. *Biochemistry* **35**:16601-9.
- 89 **Freed, E. O.** 2001. HIV-1 replication. *Somat Cell Mol Genet* **26**:13-33.
- 90 **Fu, W., and A. Rein.** 1993. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* **67**:5443-9.
- 91 **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.

- 92 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.
- 93 Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* **397**:436-41.
- 94 Garrus, J. E., U. K. von Schwedler, O. W. Pornillos, S. G. Morham, K. H. Zavitz, H. E. Wang, D. A. Wettstein, K. M. Stray, M. Cote, R. L. Rich, D. G. Myszka, and W. I. Sundquist. 2001. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**:55-65.
- 95 Girard, F., F. Barbault, C. Gouyette, T. Huynh-Dinh, J. Paoletti, and G. Lancelot. 1999. Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex. *J Biomol Struct Dyn* **16**:1145-57.
- 96 Girard, P. M., B. Bonnet-Mathoniere, D. Muriaux, and J. Paoletti. 1995. A short autocomplementary sequence in the 5' leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry* **34**:9785-94.
- 97 Girard, P. M., H. de Rocquigny, B. P. Roques, and J. Paoletti. 1996. A model of PSI dimerization: destabilization of the C278-G303 stem-loop by the nucleocapsid protein (NCp10) of MoMuLV. *Biochemistry* **35**:8705-14.
- 98 Gonda, M. A., N. R. Rice, and R. V. Gilden. 1980. Avian reticuloendotheliosis virus: characterization of the high-molecular-weight viral RNA in transforming and helper virus populations. *J Virol* **34**:743-51.
- 99 Goodenow, M., T. Huet, W. Saurin, S. Kwok, J. Sninsky, and S. Wain-Hobson. 1989. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* **2**:344-52.
- 100 Gorelick, R. J., S. M. Nigida, Jr., J. W. Bess, Jr., L. O. Arthur, L. E. Henderson, and A. Rein. 1990. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J Virol* **64**:3207-11.
- 101 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.
- 102 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**:3195-9.
- 103 Graveley, B. R., E. S. Fleming, and G. M. Gilmartin. 1996. Restoration of both structure and function to a defective poly(A) site by in vitro selection. *J Biol Chem* **271**:33654-63.
- 104 Greatorex, J., J. Gallego, G. Varani, and A. Lever. 2002. Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J Mol Biol* **322**:543-57.
- 105 Greene, W. C., and B. M. Peterlin. 2002. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat Med* **8**:673-80.
- 106 Gross, I., H. Hohenberg, T. Wilk, K. Wieggers, M. Grattinger, B. Muller, S. Fuller, and H. G. Krausslich. 2000. A conformational switch controlling HIV-1 morphogenesis. *Embo J* **19**:103-13.

- 107 **Guschlbauer, W., J. F. Chantot, and D. Thiele.** 1990. Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *J Biomol Struct Dyn* **8**:491-511.
- 108 **Haddrick, M., A. L. Lear, A. J. Cann, and S. Heaphy.** 1996. Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1. *J Mol Biol* **259**:58-68.
- 109 **Harrich, D., C. W. Hooker, and E. Parry.** 2000. The human immunodeficiency virus type 1 TAR RNA upper stem-loop plays distinct roles in reverse transcription and RNA packaging. *J Virol* **74**:5639-46.
- 110 **Harris, R. S., K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, and M. H. Malim.** 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**:803-9.
- 111 **Harrison, G. P., and A. M. Lever.** 1992. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol* **66**:4144-53.
- 112 **Harrison, G. P., G. Miele, E. Hunter, and A. M. Lever.** 1998. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J Virol* **72**:5886-96.
- 113 **Hayashi, T., T. Shioda, Y. Iwakura, and H. Shibuta.** 1992. RNA packaging signal of human immunodeficiency virus type 1. *Virology* **188**:590-9.
- 114 **Helga-Maria, C., M. L. Hammarskjold, and D. Rekosh.** 1999. An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA. *J Virol* **73**:4127-35.
- 115 **Hill, C. P., D. Worthylake, D. P. Bancroft, A. M. Christensen, and W. I. Sundquist.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc Natl Acad Sci U S A* **93**:3099-104.
- 116 **Hill, M. K., M. Shehu-Xhilaga, S. M. Campbell, P. Poubourios, S. M. Crowe, and J. Mak.** 2003. The dimer initiation sequence stem-loop of human immunodeficiency virus type 1 is dispensable for viral replication in peripheral blood mononuclear cells. *J Virol* **77**:8329-35.
- 117 **Hioe, C. E., L. Bastiani, J. E. Hildreth, and S. Zolla-Pazner.** 1998. Role of cellular adhesion molecules in HIV type 1 infection and their impact on virus neutralization. *AIDS Res Hum Retroviruses* **14 Suppl 3**:S247-54.
- 118 **Hoglund, S., A. Ohagen, J. Goncalves, A. T. Panganiban, and D. Gabuzda.** 1997. Ultrastructure of HIV-1 genomic RNA. *Virology* **233**:271-9.
- 119 **Hu, W. S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. *Science* **250**:1227-33.
- 120 **Hu, W. S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**:1556-60.
- 121 **Huang, Y., W. A. Paxton, S. M. Wolinsky, A. U. Neumann, L. Zhang, T. He, S. Kang, D. Ceradini, Z. Jin, K. Yazdanbakhsh, K. Kunstman, D. Erickson, E. Dragon, N. R. Landau, J. Phair, D. D. Ho, and R. A. Koup.** 1996. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* **2**:1240-3.

- 122 **Huthoff, H., and B. Berkhout.** 2001. Two alternating structures of the HIV-1 leader RNA. *Rna* **7**:143-57.
- 123 **Huthoff, H., and B. Berkhout.** 2002. Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry* **41**:10439-45.
- 124 **Johnson, J. M., D. A. Willits, M. J. Young, and A. Zlotnick.** 2004. Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus. *J Mol Biol* **335**:455-64.
- 125 **Jones, K. A., and B. M. Peterlin.** 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem* **63**:717-43.
- 126 **Jossinet, F., J. S. Lodmell, C. Ehresmann, B. Ehresmann, and R. Marquet.** 2001. Identification of the in vitro HIV-2/SIV RNA dimerization site reveals striking differences with HIV-1. *J Biol Chem* **276**:5598-604.
- 127 **Kao, S., M. A. Khan, E. Miyagi, R. Plishka, A. Buckler-White, and K. Strebel.** 2003. The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. *J Virol* **77**:11398-407.
- 128 **Karageorgos, L., P. Li, and C. Burrell.** 1993. Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS Res Hum Retroviruses* **9**:817-23.
- 129 **Katoh, I., T. Yasunaga, and Y. Yoshinaka.** 1993. Bovine leukemia virus RNA sequences involved in dimerization and specific gag protein binding: close relation to the packaging sites of avian, murine, and human retroviruses. *J Virol* **67**:1830-9.
- 130 **Kaye, J. F., J. H. Richardson, and A. M. Lever.** 1995. cis-acting sequences involved in human immunodeficiency virus type 1 RNA packaging. *J Virol* **69**:6588-92.
- 131 **Kaye, J. F., and A. M. Lever.** 1998. Nonreciprocal packaging of human immunodeficiency virus type 1 and type 2 RNA: a possible role for the p2 domain of Gag in RNA encapsidation. *J Virol* **72**:5877-85.
- 132 **Khan, M. A., C. Aberham, S. Kao, H. Akari, R. Gorelick, S. Bour, and K. Strebel.** 2001. Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. *J Virol* **75**:7252-65.
- 133 **Khandjian, E. W., and C. Meric.** 1986. A procedure for Northern blot analysis of native RNA. *Anal Biochem* **159**:227-32.
- 134 **Kim, H. J., K. Lee, and J. J. O'Rear.** 1994. A short sequence upstream of the 5' major splice site is important for encapsidation of HIV-1 genomic RNA. *Virology* **198**:336-40.
- 135 **Kimpton, J., and M. Emerman.** 1992. Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene. *J Virol* **66**:2232-9.
- 136 **Klasens, B. I., A. T. Das, and B. Berkhout.** 1998. Inhibition of polyadenylation by stable RNA secondary structure. *Nucleic Acids Res* **26**:1870-6.
- 137 **Klasens, B. I., M. Thiesen, A. Virtanen, and B. Berkhout.** 1999. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. *Nucleic Acids Res* **27**:446-54.

- 138 **Klatzmann, D., E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J. C. Gluckman, and L. Montagnier.** 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**:767-8.
- 139 **Knight, J.** 2003. Gene regulation: switched on to RNA. *Nature* **425**:232-3.
- 140 **Knight, J. B., Z. H. Si, and C. M. Stoltzfus.** 1994. A base-paired structure in the avian sarcoma virus 5' leader is required for efficient encapsidation of RNA. *J Virol* **68**:4493-502.
- 141 **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.
- 142 **Kung, H. J., J. M. Bailey, N. Davidson, M. O. Nicolson, and R. M. McAllister.** 1975. Structure, subunit composition, and molecular weight of RD-114 RNA. *J Virol* **16**:397-411.
- 143 **Kung, H. J., J. M. Bailey, N. Davidson, P. K. Vogt, M. O. Nicolson, and R. M. McAllister.** 1975. Electron microscope studies of tumor virus RNA. *Cold Spring Harb Symp Quant Biol* **39 Pt 2**:827-34.
- 144 **Kung, H. J., S. Hu, W. Bender, J. M. Bailey, N. Davidson, M. O. Nicolson, and R. M. McAllister.** 1976. RD-114, baboon, and woolly monkey viral RNA's compared in size and structure. *Cell* **7**:609-20.
- 145 **Lanchy, J. M., and J. S. Lodmell.** 2002. Alternate usage of two dimerization initiation sites in HIV-2 viral RNA in vitro. *J Mol Biol* **319**:637-48.
- 146 **Lanchy, J. M., C. A. Rentz, J. D. Ivanovitch, and J. S. Lodmell.** 2003. Elements located upstream and downstream of the major splice donor site influence the ability of HIV-2 leader RNA to dimerize in vitro. *Biochemistry* **42**:2634-42.
- 147 **Laughrea, M., and L. Jette.** 1994. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* **33**:13464-74.
- 148 **Laughrea, M., and L. Jette.** 1996. HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain. *Biochemistry* **35**:9366-74.
- 149 **Laughrea, M., and L. Jette.** 1996. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation. *Biochemistry* **35**:1589-98.
- 150 **Laughrea, M., and L. Jette.** 1997. HIV-1 genome dimerization: kissing-loop hairpin dictates whether nucleotides downstream of the 5' splice junction contribute to loose and tight dimerization of human immunodeficiency virus RNA. *Biochemistry* **36**:9501-8.
- 151 **Laughrea, M., L. Jette, J. Mak, L. Kleiman, C. Liang, and M. A. Wainberg.** 1997. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J Virol* **71**:3397-406.

- 152 **Laughrea, M., N. Shen, L. Jette, and M. A. Wainberg.** 1999. Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry* **38**:226-34.
- 153 **Laughrea, M., N. Shen, L. Jette, J. L. Darlix, L. Kleiman, and M. A. Wainberg.** 2001. Role of distal zinc finger of nucleocapsid protein in genomic RNA dimerization of human immunodeficiency virus type 1; no role for the palindrome crowning the R-U5 hairpin. *Virology* **281**:109-16.
- 154 **Lee, J. S., D. H. Evans, and A. R. Morgan.** 1980. Polypurine DNAs and RNAs form secondary structures which may be tetra-stranded. *Nucleic Acids Res* **8**:4305-20.
- 155 **Lee, N. S., T. Dohjima, G. Bauer, H. Li, M. J. Li, A. Ehsani, P. Salvaterra, and J. Rossi.** 2002. Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol* **20**:500-5.
- 156 **Lee, N. S., and J. J. Rossi.** 2004. Control of HIV-1 replication by RNA interference. *Virus Res* **102**:53-8.
- 157 **Lever, A., H. Gottlinger, W. Haseltine, and J. Sodroski.** 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* **63**:4085-7.
- 158 **Li, F., R. Goila-Gaur, K. Salzwedel, N. R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D. E. Martin, J. M. Orenstein, G. P. Allaway, E. O. Freed, and C. T. Wild.** 2003. PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc Natl Acad Sci U S A* **100**:13555-60.
- 159 **Liang, C., X. Li, Y. Quan, M. Laughrea, L. Kleiman, J. Hiscott, and M. A. Wainberg.** 1997. Sequence elements downstream of the human immunodeficiency virus type 1 long terminal repeat are required for efficient viral gene transcription. *J Mol Biol* **272**:167-77.
- 160 **Liang, C., X. Li, L. Rong, P. Inouye, Y. Quan, L. Kleiman, and M. A. Wainberg.** 1997. The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *J Virol* **71**:5750-7.
- 161 **Liang, C., L. Rong, M. Laughrea, L. Kleiman, and M. A. Wainberg.** 1998. Compensatory point mutations in the human immunodeficiency virus type 1 Gag region that are distal from deletion mutations in the dimerization initiation site can restore viral replication. *J Virol* **72**:6629-36.
- 162 **Liang, C., L. Rong, E. Cherry, L. Kleiman, M. Laughrea, and M. A. Wainberg.** 1999. Deletion mutagenesis within the dimerization initiation site of human immunodeficiency virus type 1 results in delayed processing of the p2 peptide from precursor proteins. *J Virol* **73**:6147-51.
- 163 **Liang, C., L. Rong, Y. Quan, M. Laughrea, L. Kleiman, and M. A. Wainberg.** 1999. Mutations within four distinct gag proteins are required to restore replication of human immunodeficiency virus type 1 after deletion mutagenesis within the dimerization initiation site. *J Virol* **73**:7014-20.
- 164 **Liang, C., L. Rong, R. S. Russell, and M. A. Wainberg.** 2000. Deletion mutagenesis downstream of the 5' long terminal repeat of human

- immunodeficiency virus type 1 is compensated for by point mutations in both the U5 region and gag gene. *J Virol* **74**:6251-61.
- 165 **Liang, C., J. Hu, R. S. Russell, A. Roldan, L. Kleiman, and M. A. Wainberg.** 2002. Characterization of a putative alpha-helix across the capsid-SP1 boundary that is critical for the multimerization of human immunodeficiency virus type 1 gag. *J Virol* **76**:11729-37.
- 166 **Liang, C., and M. A. Wainberg.** 2002. The role of Tat in HIV-1 replication: an activator and/or a suppressor? *AIDS Rev* **4**:41-9.
- 167 **Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, and N. R. Landau.** 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**:367-77.
- 168 **Luban, J., and S. P. Goff.** 1994. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J Virol* **68**:3784-93.
- 169 **Ly, H., D. P. Nierlich, J. C. Olsen, and A. H. Kaplan.** 1999. Moloney murine sarcoma virus genomic RNAs dimerize via a two-step process: a concentration-dependent kissing-loop interaction is driven by initial contact between consecutive guanines. *J Virol* **73**:7255-61.
- 170 **Ly, H., and T. G. Parslow.** 2002. Bipartite signal for genomic RNA dimerization in Moloney murine leukemia virus. *J Virol* **76**:3135-44.
- 171 **Madani, N., and D. Kabat.** 1998. An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J Virol* **72**:10251-5.
- 172 **Maisel, J., W. Bender, S. Hu, P. H. Duesberg, and N. Davidson.** 1978. Structure of 50 to 70S RNA from Moloney sarcoma viruses. *J Virol* **25**:384-94.
- 173 **Mak, J., M. Jiang, M. A. Wainberg, M. L. Hammarskjold, D. Rekosh, and L. Kleiman.** 1994. Role of Pr160gag-pol in mediating the selective incorporation of tRNA(Lys) into human immunodeficiency virus type 1 particles. *J Virol* **68**:2065-72.
- 174 **Malim, M. H., L. S. Tiley, D. F. McCarn, J. R. Rusche, J. Hauber, and B. R. Cullen.** 1990. HIV-1 structural gene expression requires binding of the Rev transactivator to its RNA target sequence. *Cell* **60**:675-83.
- 175 **Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono.** 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**:99-103.
- 176 **Mangeat, B., P. Turelli, S. Liao, and D. Trono.** 2004. A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. *J Biol Chem.*
- 177 **Mann, J. M., and D. J. Tarantola.** 1998. HIV 1998: the global picture. *Sci Am* **279**:82-3.
- 178 **Mann, R., R. C. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* **33**:153-9.
- 179 **Mann, R., and D. Baltimore.** 1985. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs. *J Virol* **54**:401-7.

- 180 **Mansky, L. M., A. E. Krueger, and H. M. Temin.** 1995. The bovine leukemia virus encapsidation signal is discontinuous and extends into the 5' end of the gag gene. *J Virol* **69**:3282-9.
- 181 **Mariani, R., D. Chen, B. Schrofelbauer, F. Navarro, R. Konig, B. Bollman, C. Munk, H. Nymark-McMahon, and N. R. Landau.** 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**:21-31.
- 182 **Marin, M., K. M. Rose, S. L. Kozak, and D. Kabat.** 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* **9**:1398-403.
- 183 **Marquet, R., F. Baudin, C. Gabus, J. L. Darlix, M. Mougel, C. Ehresmann, and B. Ehresmann.** 1991. Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res* **19**:2349-57.
- 184 **Marquet, R., J. C. Paillart, E. Skripkin, C. Ehresmann, and B. Ehresmann.** 1994. Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res* **22**:145-51.
- 185 **Martin-Serrano, J., T. Zang, and P. D. Bieniasz.** 2003. Role of ESCRT-I in retroviral budding. *J Virol* **77**:4794-804.
- 186 **Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner.** 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**:911-40.
- 187 **McBride, M. S., and A. T. Panganiban.** 1996. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* **70**:2963-73.
- 188 **McBride, M. S., and A. T. Panganiban.** 1997. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J Virol* **71**:2050-8.
- 189 **McBride, M. S., M. D. Schwartz, and A. T. Panganiban.** 1997. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J Virol* **71**:4544-54.
- 190 **McDonald, D., M. A. Vodicka, G. Lucero, T. M. Svitkina, G. G. Borisy, M. Emerman, and T. J. Hope.** 2002. Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol* **159**:441-52.
- 191 **McDougal, J. S., M. S. Kennedy, J. M. Sligh, S. P. Cort, A. Mawle, and J. K. Nicholson.** 1986. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**:382-5.
- 192 **McLeod, G. X., and S. M. Hammer.** 1992. Zidovudine: five years later. *Ann Intern Med* **117**:487-501.
- 193 **Mehle, A., B. Strack, P. Ancuta, C. Zhang, M. McPike, and D. Gabuzda.** 2004. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J Biol Chem* **279**:7792-8.
- 194 **Meric, C., and P. F. Spahr.** 1986. Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging. *J Virol* **60**:450-9.

- 195 **Meric, C., E. Guilloud, and P. F. Spahr.** 1988. Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): deletions of Cys-His boxes. *J Virol* 62:3328-33.
- 196 **Miller, M. D., C. M. Farnet, and F. D. Bushman.** 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 71:5382-90.
- 197 **Morellet, N., N. Jullian, H. De Rocquigny, B. Maigret, J. L. Darlix, and B. P. Roques.** 1992. Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *Embo J* 11:3059-65.
- 198 **Morikawa, Y., W. H. Zhang, D. J. Hockley, M. V. Nermut, and I. M. Jones.** 1998. Detection of a trimeric human immunodeficiency virus type 1 Gag intermediate is dependent on sequences in the matrix protein, p17. *J Virol* 72:7659-63.
- 199 **Morikawa, Y., D. J. Hockley, M. V. Nermut, and I. M. Jones.** 2000. Roles of matrix, p2, and N-terminal myristoylation in human immunodeficiency virus type 1 Gag assembly. *J Virol* 74:16-23.
- 200 **Morin, N., E. Cherry, X. Li, and M. A. Wainberg.** 1998. Cotransfection of mutated forms of human immunodeficiency virus type 1 Gag-Pol with wild-type constructs can interfere with processing and viral replication. *J Hum Virol* 1:240-7.
- 201 **Mujeeb, A., J. L. Clever, T. M. Billeci, T. L. James, and T. G. Parslow.** 1998. Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol* 5:432-6.
- 202 **Mujeeb, A., T. G. Parslow, A. Zarrinpar, C. Das, and T. L. James.** 1999. NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett* 458:387-92.
- 203 **Muriaux, D., P. M. Girard, B. Bonnet-Mathoniere, and J. Paoletti.** 1995. Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* 270:8209-16.
- 204 **Muriaux, D., H. De Rocquigny, B. P. Roques, and J. Paoletti.** 1996. NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* 271:33686-92.
- 205 **Muriaux, D., P. Fosse, and J. Paoletti.** 1996. A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry* 35:5075-82.
- 206 **Muriaux, D., J. Mirro, D. Harvin, and A. Rein.** 2001. RNA is a structural element in retrovirus particles. *Proc Natl Acad Sci U S A* 98:5246-51.
- 207 **Murti, K. G., M. Bondurant, and A. Tereba.** 1981. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J Virol* 37:411-19.
- 208 **Nabel, G. J.** 2001. Challenges and opportunities for development of an AIDS vaccine. *Nature* 410:1002-7.
- 209 **Omichinski, J. G., G. M. Clore, K. Sakaguchi, E. Appella, and A. M. Gronenborn.** 1991. Structural characterization of a 39-residue synthetic peptide

- containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett* **292**:25-30.
- 210 **Oroudjev, E. M., P. C. Kang, and L. A. Kohlstaedt.** 1999. An additional dimer linkage structure in Moloney murine leukemia virus RNA. *J Mol Biol* **291**:603-13.
- 211 **Ortiz-Conde, B. A., and S. H. Hughes.** 1999. Studies of the genomic RNA of leukosis viruses: implications for RNA dimerization. *J Virol* **73**:7165-74.
- 212 **Ott, D. E.** 2002. Potential roles of cellular proteins in HIV-1. *Rev Med Virol* **12**:359-74.
- 213 **Paillart, J. C., R. Marquet, E. Skripkin, B. Ehresmann, and C. Ehresmann.** 1994. Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* **269**:27486-93.
- 214 **Paillart, J. C., L. Berthoux, M. Ottmann, J. L. Darlix, R. Marquet, B. Ehresmann, and C. Ehresmann.** 1996. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol* **70**:8348-54.
- 215 **Paillart, J. C., E. Skripkin, B. Ehresmann, C. Ehresmann, and R. Marquet.** 1996. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* **93**:5572-7.
- 216 **Paillart, J. C., E. Westhof, C. Ehresmann, B. Ehresmann, and R. Marquet.** 1997. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J Mol Biol* **270**:36-49.
- 217 **Paillart, J. C., E. Skripkin, B. Ehresmann, C. Ehresmann, and R. Marquet.** 2002. In vitro evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA. *J Biol Chem* **277**:5995-6004.
- 218 **Panganiban, A. T., and D. Fiore.** 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* **241**:1064-9.
- 219 **Parolin, C., T. Dorfman, G. Palu, H. Gottlinger, and J. Sodroski.** 1994. Analysis in human immunodeficiency virus type 1 vectors of cis-acting sequences that affect gene transfer into human lymphocytes. *J Virol* **68**:3888-95.
- 220 **Polge, E., J. L. Darlix, J. Paoletti, and P. Fosse.** 2000. Characterization of loose and tight dimer forms of avian leukosis virus RNA. *J Mol Biol* **300**:41-56.
- 221 **Pomerantz, R. J.** 2002. HIV: a tough viral nut to crack. *Nature* **418**:594-5.
- 222 **Pomerantz, R. J., and D. L. Horn.** 2003. Twenty years of therapy for HIV-1 infection. *Nat Med* **9**:867-73.
- 223 **Poon, D. T., J. Wu, and A. Aldovini.** 1996. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J Virol* **70**:6607-16.
- 224 **Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo.** 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**:497-500.
- 225 **Prats, A. C., C. Roy, P. A. Wang, M. Erard, V. Housset, C. Gabus, C. Paoletti, and J. L. Darlix.** 1990. cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA. *J Virol* **64**:774-83.

- 226 **Rein, A., L. E. Henderson, and J. G. Levin.** 1998. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **23**:297-301.
- 227 **Resh, M. D.** 1999. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochim Biophys Acta* **1451**:1-16.
- 228 **Robinson, W. S., H. L. Robinson, and P. H. Duesberg.** 1967. Tumor virus RNA's. *Proc Natl Acad Sci U S A* **58**:825-34.
- 229 **Rong, L., R. S. Russell, J. Hu, Y. Guan, L. Kleiman, C. Liang, and M. A. Wainberg.** 2001. Hydrophobic amino acids in the human immunodeficiency virus type 1 p2 and nucleocapsid proteins can contribute to the rescue of deleted viral RNA packaging signals. *J Virol* **75**:7230-43.
- 230 **Rong, L., R. S. Russell, J. Hu, M. Laughrea, M. A. Wainberg, and C. Liang.** 2003. Deletion of stem-loop 3 is compensated by second-site mutations within the Gag protein of human immunodeficiency virus type 1. *Virology* **314**:221-8.
- 231 **Rowland-Jones, S. L.** 2003. Timeline: AIDS pathogenesis: what have two decades of HIV research taught us? *Nat Rev Immunol* **3**:343-8.
- 232 **Russell, R. S., J. Hu, M. Laughrea, M. A. Wainberg, and C. Liang.** 2002. Deficient dimerization of human immunodeficiency virus type 1 RNA caused by mutations of the u5 RNA sequences. *Virology* **303**:152-63.
- 233 **Russell, R. S., J. Hu, V. Beriault, A. J. Mouland, M. Laughrea, L. Kleiman, M. A. Wainberg, and C. Liang.** 2003. Sequences downstream of the 5' splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J Virol* **77**:84-96.
- 234 **Russell, R. S., A. Roldan, M. Detorio, J. Hu, M. A. Wainberg, and C. Liang.** 2003. Effects of a single amino acid substitution within the p2 region of human immunodeficiency virus type 1 on packaging of spliced viral RNA. *J Virol* **77**:12986-95.
- 235 **Sakaguchi, K., N. Zambrano, E. T. Baldwin, B. A. Shapiro, J. W. Erickson, J. G. Omichinski, G. M. Clore, A. M. Gronenborn, and E. Appella.** 1993. Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc Natl Acad Sci U S A* **90**:5219-23.
- 236 **Sakuragi, J., T. Shioda, and A. T. Panganiban.** 2001. Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions. *J Virol* **75**:2557-65.
- 237 **Sakuragi, J., A. Iwamoto, and T. Shioda.** 2002. Dissociation of genome dimerization from packaging functions and virion maturation of human immunodeficiency virus type 1. *J Virol* **76**:959-67.
- 238 **Sakuragi, J., S. Ueda, A. Iwamoto, and T. Shioda.** 2003. Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging. *J Virol* **77**:4060-9.
- 239 **Sakuragi, J. I., and A. T. Panganiban.** 1997. Human immunodeficiency virus type 1 RNA outside the primary encapsidation and dimer linkage region affects RNA dimer stability in vivo. *J Virol* **71**:3250-4.

- 240 **Sambrook, J., Fritsch, EF, Maniatis, T.** 1989. *Molecular Cloning: A Laboratory*
Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 241 **Sarngadharan, M. G., M. Popovic, L. Bruch, J. Schupbach, and R. C. Gallo.**
1984. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in
the serum of patients with AIDS. *Science* **224**:506-8.
- 242 **Schrofelbauer, B., D. Chen, and N. R. Landau.** 2004. A single amino acid of
APOBEC3G controls its species-specific interaction with virion infectivity factor
(Vif). *Proc Natl Acad Sci U S A* **101**:3927-32.
- 243 **Schulz, T. F., B. A. Jameson, L. Lopalco, A. G. Siccardi, R. A. Weiss, and J.**
P. Moore. 1992. Conserved structural features in the interaction between
retroviral surface and transmembrane glycoproteins? *AIDS Res Hum Retroviruses*
8:1571-80.
- 244 **Schwartz, M. D., D. Fiore, and A. T. Panganiban.** 1997. Distinct functions and
requirements for the Cys-His boxes of the human immunodeficiency virus type 1
nucleocapsid protein during RNA encapsidation and replication. *J Virol* **71**:9295-
305.
- 245 **Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim.** 2002. Isolation of a
human gene that inhibits HIV-1 infection and is suppressed by the viral Vif
protein. *Nature* **418**:646-50.
- 246 **Sheehy, A. M., N. C. Gaddis, and M. H. Malim.** 2003. The antiretroviral
enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif.
Nat Med **9**:1404-7.
- 247 **Shehu-Xhilaga, M., S. M. Crowe, and J. Mak.** 2001. Maintenance of the
Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA
dimerization and viral infectivity. *J Virol* **75**:1834-41.
- 248 **Shen, N., L. Jette, C. Liang, M. A. Wainberg, and M. Laughrea.** 2000. Impact
of human immunodeficiency virus type 1 RNA dimerization on viral infectivity
and of stem-loop B on RNA dimerization and reverse transcription and
dissociation of dimerization from packaging. *J Virol* **74**:5729-35.
- 249 **Simon, J. H., N. C. Gaddis, R. A. Fouchier, and M. H. Malim.** 1998. Evidence
for a newly discovered cellular anti-HIV-1 phenotype. *Nat Med* **4**:1397-400.
- 250 **Skalka, A., Goff, SP.** 1993. *Reverse Transcriptase.* Cold Spring Harbor
Laboratory Press, Cold Spring Harbor, NY.
- 251 **Skripkin, E., J. C. Paillart, R. Marquet, B. Ehresmann, and C. Ehresmann.**
1994. Identification of the primary site of the human immunodeficiency virus type
1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A* **91**:4945-9.
- 252 **Skripkin, E., J. C. Paillart, R. Marquet, M. Blumenfeld, B. Ehresmann, and**
C. Ehresmann. 1996. Mechanisms of inhibition of in vitro dimerization of HIV
type I RNA by sense and antisense oligonucleotides. *J Biol Chem* **271**:28812-7.
- 253 **South, T. L., and M. F. Summers.** 1993. Zinc- and sequence-dependent binding
to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein:
NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci*
2:3-19.
- 254 **Stokrova, J., J. Korb, and J. Riman.** 1982. Electron microscopic studies on the
structure of 60-70S RNA of avian myeloblastosis virus. *Acta Virol* **26**:417-26.

- 255 **Stoltzfus, C. M., and P. N. Snyder.** 1975. Structure of B77 sarcoma virus RNA: stabilization of RNA after packaging. *J Virol* **16**:1161-70.
- 256 **Stopak, K., C. de Noronha, W. Yonemoto, and W. C. Greene.** 2003. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* **12**:591-601.
- 257 **Strack, B., A. Calistri, and H. G. Gottlinger.** 2002. Late assembly domain function can exhibit context dependence and involves ubiquitin residues implicated in endocytosis. *J Virol* **76**:5472-9.
- 258 **Strack, B., A. Calistri, S. Craig, E. Popova, and H. G. Gottlinger.** 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **114**:689-99.
- 259 **Strebel, K., D. Daugherty, K. Clouse, D. Cohen, T. Folks, and M. A. Martin.** 1987. The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**:728-30.
- 260 **Stuhlmann, H., and P. Berg.** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. *J Virol* **66**:2378-88.
- 261 **Sullenger, B. A., and T. R. Cech.** 1993. Tethering ribozymes to a retroviral packaging signal for destruction of viral RNA. *Science* **262**:1566-9.
- 262 **Sundquist, W. I., and S. Heaphy.** 1993. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc Natl Acad Sci U S A* **90**:3393-7.
- 263 **Tang, C., E. Loeliger, I. Kinde, S. Kyere, K. Mayo, E. Barklis, Y. Sun, M. Huang, and M. F. Summers.** 2003. Antiviral inhibition of the HIV-1 capsid protein. *J Mol Biol* **327**:1013-20.
- 264 **Tashima, K. T., and C. C. Carpenter.** 2003. Fusion inhibition--a major but costly step forward in the treatment of HIV-1. *N Engl J Med* **348**:2249-50.
- 265 **Temin, H. M.** 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci U S A* **90**:6900-3.
- 266 **Torrent, C., T. Bordet, and J. L. Darlix.** 1994. Analytical study of rat retrotransposon VL30 RNA dimerization in vitro and packaging in murine leukemia virus. *J Mol Biol* **240**:434-44.
- 267 **Torrent, C., C. Gabus, and J. L. Darlix.** 1994. A small and efficient dimerization/packaging signal of rat VL30 RNA and its use in murine leukemia virus-VL30-derived vectors for gene transfer. *J Virol* **68**:661-7.
- 268 **Tounekti, N., M. Mougel, C. Roy, R. Marquet, J. L. Darlix, J. Paoletti, B. Ehresmann, and C. Ehresmann.** 1992. Effect of dimerization on the conformation of the encapsidation Psi domain of Moloney murine leukemia virus RNA. *J Mol Biol* **223**:205-20.
- 269 **UNAIDS** 2004, posting date. AIDS epidemic update. UNAIDS. [Online.]
- 270 **von Schwedler, U. K., M. Stuchell, B. Muller, D. M. Ward, H. Y. Chung, E. Morita, H. E. Wang, T. Davis, G. P. He, D. M. Cimbara, A. Scott, H. G. Krausslich, J. Kaplan, S. G. Morham, and W. I. Sundquist.** 2003. The protein network of HIV budding. *Cell* **114**:701-13.
- 271 **Wain-Hobson, S.** 1992. Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo. *Curr Top Microbiol Immunol* **176**:181-93.

- 272 Wang, S. W., and A. Aldovini. 2002. RNA incorporation is critical for retroviral particle integrity after cell membrane assembly of Gag complexes. *J Virol* **76**:11853-65.
- 273 Wang, S. W., K. Noonan, and A. Aldovini. 2004. Nucleocapsid-RNA interactions are essential to structural stability but not to assembly of retroviruses. *J Virol* **78**:716-23.
- 274 Watanabe, S., and H. M. Temin. 1982. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5' long terminal repeat and the start of the gag gene. *Proc Natl Acad Sci U S A* **79**:5986-90.
- 275 Wieggers, K., G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, and H. G. Krausslich. 1998. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* **72**:2846-54.
- 276 Wilk, T., I. Gross, B. E. Gowen, T. Rutten, F. de Haas, R. Welker, H. G. Krausslich, P. Boulanger, and S. D. Fuller. 2001. Organization of immature human immunodeficiency virus type 1. *J Virol* **75**:759-71.
- 277 Williamson, J. R. 2000. Induced fit in RNA-protein recognition. *Nat Struct Biol* **7**:834-7.
- 278 Xie, B., V. Calabro, M. A. Wainberg, and A. D. Frankel. 2004. Selection of TAR RNA-binding chameleon peptides by using a retroviral replication system. *J Virol* **78**:1456-63.
- 279 Xu, H., E. S. Svarovskaia, R. Barr, Y. Zhang, M. A. Khan, K. Strebel, and V. K. Pathak. 2004. A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. *Proc Natl Acad Sci U S A*.
- 280 Yang, A. G., X. Bai, X. F. Huang, C. Yao, and S. Chen. 1997. Phenotypic knockout of HIV type 1 chemokine coreceptor CCR-5 by intrakines as potential therapeutic approach for HIV-1 infection. *Proc Natl Acad Sci U S A* **94**:11567-72.
- 281 Yang, S., and H. M. Temin. 1994. A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA. *Embo J* **13**:713-26.
- 282 Yu, X., Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, and X. F. Yu. 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**:1056-60.
- 283 Yuan, Y., D. J. Kerwood, A. C. Paoletti, M. F. Shubsda, and P. N. Borer. 2003. Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry* **42**:5259-69.
- 284 Zeffman, A., S. Hassard, G. Varani, and A. Lever. 2000. The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *J Mol Biol* **297**:877-93.
- 285 Zhang, H., B. Yang, R. J. Pomerantz, C. Zhang, S. C. Arunachalam, and L. Gao. 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**:94-8.
- 286 Zhang, S., H. Hara, H. Kaji, and A. Kaji. 1997. Inhibition of HIV type 1 RNA dimerization by antisense DNA corresponding to the 17-nucleotide sequence

- downstream from the splice donor site of HIV type 1 RNA. *AIDS Res Hum Retroviruses* **13**:865-73.
- 287 **Zhang, Y., and E. Barklis.** 1995. Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J Virol* **69**:5716-22.
- 288 **Zhang, Y., and E. Barklis.** 1997. Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation. *J Virol* **71**:6765-76.
- 289 **Zhou, J., X. Yuan, D. Dismuke, B. M. Forshey, C. Lundquist, K. H. Lee, C. Aiken, and C. H. Chen.** 2004. Small-molecule inhibition of human immunodeficiency virus type 1 replication by specific targeting of the final step of virion maturation. *J Virol* **78**:922-9.
- 290 **Zhou, W., L. J. Parent, J. W. Wills, and M. D. Resh.** 1994. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J Virol* **68**:2556-69.
- 291 **Zuker, M., Mathews, DH, Turner, DH.** 1999. *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology.* Kluwer Academic, Dordrecht/Norwell, MA.