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

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

Role of the Leader Sequence of Human Immunodeficiency Virus Type 1 in Viral Replication, Genome Dimerization, Encapsidation, and Proviral DNA Synthesis

By Ni Shen

A thesis submitted to the faculty of Graduate studies and Research, McGill University, in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Department of Medicine, Division of Experimental Medicine, McGill University, Montreal, Canada

August 2002

Your file Votre référence
ISBN: 0-612-88578-X
Our file Notre référence
ISBN: 0-612-88578-X

To my parents, for their unconditional love and cultivation

To my parents-in-law, for their constant support

To my husband, for his love and support

To my big sisters Wendy R. Zeng and Marcia W. Tan, for their long-lasting friendship and encouragement

To my little daughter, for the joy she brings to our life

Abstract

Human immunodeficiency virus type 1 (HIV-1) genome consists of two identical RNAs that appear noncovalently linked near their 5' ends. The 5' untranslated region is called leader region. The 3' part of leader, i.e. nucleotides U200 to G335 in HIV-1 genomic RNA, between the primer binding site and the gag gene, can fold into 3 stem-loops: the kissing-loop domain (KLD) or stem-loop 1 (SL1), the 5' splicing junction hairpin (SD) or SL2 and SL3. The KLD, from nucleotide (nt) 243 to 277, forms a stem-loop (kissing loop hairpin) seated on top of a small stem bulge (stem B and loop B). The kissing-loop hairpin, or dimerization initiation site (DIS) hairpin consists of stem C and loop C. Loop C contains the autocomplementary sequence (ACS) GCGCGC262 or GUGCAC262, also called DIS.

In the kissing-loop model of HIV-1 genome dimerization, HIV-1 RNA dimerization is initiated by base pairing between the ACS of one RNA monomer and that of an adjacent monomer.

To understand the role of the ACS in HIV-1 replication and HIV-1 genomic RNA dimerization, we replaced the central CGCG261 (or tetramer) of the HIV-1 Lai ACS by other tetramers. Genomic RNAs containing the UUAA tetramer (non-HIV-1 tetramer) were half dimeric, but UUAA genome packaging was unaffected. This was the first evidence that genomic RNA dimerization and packaging can be dissociated (Chapter 2). Destroying stem-loop C reduced genomic RNA dimerization by ~ 50%, proviral DNA synthesis by ~ 85%, and reduced viral infectivity by ~ 3 logarithmic units. Destroying stem-loop B had similar effects on genome dimerization, reverse transcription, and viral infectivity. We also observed that mutations in stem-loop B and in the DIS hairpin were "non additive" (Chapter 2).

The existence of stem-loop C is supported by phylogenetic evidence, while that of stem-loop B is not, namely, its sequence is completely conserved. We investigated the role of stem B and loop B nucleotides in viral replication, and genomic RNA dimerization. The putative CUCG246/CGAG277 duplex was replaced by 9 alternative complementary sequences, 4 likely to base pair in long (~500 nts) RNAs, as assessed by the algorithm mfold. Among the 4 sequences, 3 preserved genome dimerization, 1 did

not significantly inhibit it, and 2 preserved viral replication. We also asked if 9 deletions or nucleotide substitutions within nucleotide 200 to 242 and/or 282 to 335 could influence genome dimerization. Δ200-226 and Δ236-242 genomic RNAs dimerized relatively poorly despite having neutral or positive influences on stem B, loop B and klh folding (Chapter 3).

Mutations within the Matrix, Capsid, p2 and nucleocapsid genes suppress several functional defects caused by KLD destruction. We tested the effect of these suppressor mutations on genome dimerization and infectiousness of viruses bearing moderate to crippling KLD mutations. Our conclusion is that these suppressor mutations can restore genomic RNA dimerization when DIS is weakened, but not when DIS is denatured or the KLD is destroyed (Chapter 4).

Résumé

Le génome du Virus de l'immunodéficience humaine 1 (VIH-1) est formé de deux ARNs identiques qui semblent liés de façon non covalente près de leur extrémité 5'. La region 5' non-traduite est appelée "leader". La portion 3' du "leader", i.e., les nucléotides U200 à G335 dans l'ARN génomique de VIH-1, situées entre le site de fixation du "primer" et le gène gag, peut former 3 tige-boucles: le "kissing-loop" domain (KLD) ou tige-boucle 1 (SL1), l'épingle d'épissage 5' (SD) ou SL2 et SL3. Le KLD, i.e., les nucleotides (nts) 243 à 277, forme une tige-boucle (kissing-loop hairpin) reposant sur une petite tige-bosse (tige B et boucle B). Le kissing-loop, ou dimérisation initiation site (DIS) hairpin, est formé d'une tige C et d'une boucle C. La boucle C contient la sequence autocomplémentaire (ACS) GCGCGC262 ou GUGCAC262, aussi appelée DIS.

Dans le modèle kissing-loop de la dimérisation du génome de VIH-1, la dimérisation de l'ARN VIH-1 est initiée par pairage de bases entre l'ACS d'un ARN monomérique et celui d'un monomère adjacent.

Pour comprendre le rôle de l'ACS dans la réplication de VIH-1 et dans la dimérisation de son ARN génomique, nous avons remplacé le CGCG261 central (ou tétramère) de l'ACS de VIH-1Lai par d'autes tétramères. Les ARN génomiques contenant le tétramère UUAA n'étaient que 50% dimériques, mais leur encapsidation est demeurée inchangée. Ceci représente la première indication que la dimérisation de l'ARN genome et son encapsidation peuvent être dissociées (Chapitre 2). La destruction de la tige-boucle C abaisse de 50% la dimérisation de l'ARN génomique, de 85% la synthèse d'AND proviral, et abaisse l'infectivité de 3 unités logarithmiques. La destruction de la tige-boucle B a eu des effets similaires sur la dimérisation du génome, la transcription reverse, et l'infectivité virale. Nous avons aussi observé que les mutations dans la tige-boucle B et dans le DIS hairpin étaient non-additives (Chapitre 2).

L'existence de la tige-boucle C est supportée par la phylogénétique, tandis que celle de la tige-boucle B ne l'est pas, puisque sa séquence est complètement

conservée. Nous avons investigué le rôle de la tige B ainsi que des nucléotides de la boucle B dans la réplication virale et la dimérisation de l'ARN génomique. Le supposé duplex CUCG246/CGAG277 a été remplacé par 9 séquences complémentaires alternatives, 4 susceptibles de former un duplex dans les ARNs longs (500 nts), selon l'algorithme mfold. Parmi ces 4 séquences, 3 ont préservé la dimérisation génomique, une ne l'a pas significativement inhibée, et 2 ont préservé la replication virale. Nous nous sommes aussi demandés si 9 délétions ou substitutions de nucléotides parmi les positions 200-242 et/ou 282-335 pouvaient influencer la dimérisation génomique. Les ARNs génomiques portant les délétions Δ200-226 ou Δ236-242 dimérisaient faiblement malgré que ces mêmes mutations aient un effet neutre ou positif sur la formation de la tige B, de la boucle B et du kissing-loop hairpin (Chapitre 3).

Des mutations dans les gènes de la matrice, de la capside, de p2 et de la nucléocapside ont supprimé plusieurs défauts fonctionnels causés par la destruction du KLD. Nous avons vérifié l'effet des ces mutations suppressives sur la dimérisation génomique et l'infectivité de virus portant des mutation modérés ou incapacitantes dans le KLD. Nous concluons que ces mutations suppressives peuvent restorer la dimérisation de l'ARN génomique quand le DIS est affaibli, mais non quand le DIS est dénature ou que le KLD est détruit (Chapitre 4).

Acknowledgements

(Names are listed according to the alphabetic order)

I would like to pay special thanks to my supervisor Dr. Michael Laughrea, who has been a tremendously great supervisor in the past five years. I am very grateful for all his supervision, his support, his always being there whenever we need his help. It's him who has made my dream of having a higher degree come true. I would also like to thank him for critical reading of my thesis and translating the abstract.

I would like to express my sincere thanks to my committee supervisors: Drs. Nick Acheson, Lorraine Chalifour, Emmanuel Faust, and Lawrence Kleiman for their supervision and all their precious advice and encouragement for all these years.

I would like to thank our director Dr. Gerald B Price for giving me the precious chance of being a Ph.D. student in our department.

I would like to thank Ms. Dominique Besso for all her care and support.

I would like to thank Dr. Mark A. Wainberg for providing the Biocontainment laboratory, the necessary facilities to make my *in vivo* research work possible.

I would also like to thank Dr. Ralph Germinario and Dr. Susan Colby-Germinario for their concern and encouragement.

I would like to express my sincere thanks to my big sister Wendy R. Zeng and her husband Junco Wang for all their help and support.

I would like to pay special thanks to Louis Jetté. Louis Jetté, my colleague and dear friend, has been a great help in my first years starting my Ph.D. studies and has made all the years working and studying interesting and the environment so friendly. I thank him for all the technique supports, as well.

I would like to thank Dr. Chen Liang for providing some of the plasmids described in chapter 2) and 4). I would like to thank Drs. Chen Liang, Yudong Quan, and Liwei Rong for their enlightening discussions and suggestions. I also thank Dr. Yudong Quan for his valuable suggestions on my thesis.

I would like to thank Drs. Shan Cen and Xin Wei for their precious help and advice during my Ph.D. study.

I would like to thank the "Bourret group": Shan Cen, Chen Liang, Yudong Quan, Liwei Rong, Xin Wei, Zhiqiang Yuan, and Yaoping Zhang for their friendship, help, concern, and for making the lonely life much easier and enjoyable.

I thank my friend Dr. Zhiheng He for his help during my first year of Ph.D. study.

I would like to thank my colleagues, and many of them are my friends as well: Dion Codrington, Ceasar Collazos, Mervi Acuzar Detorio, Karidia Diallo, Juliana Gabor, Yongjun Guan, Rabih Halwani, Hassan Javanbakht, Mayla Hsu, Yue Huang, Ahmad Khorchid, Yuhong Lu, Aretia Malapetsa, Meijuan Niu, Maureen Oliveira, Rodney Russell, Yingnian Shen, Bonnie Spira, Yuzhen Tao, Jing Wang, Sylvia Wise, Zhiyuan Xu, James Whitney, etc., for making the working environment so friendly and cooperative.

I would like to thank Rhona Rosenzweig, and Sandra Fraiberg for preparing the manuscript. I would like to thank Rita Notte for her help and support.

I would like to thank Gabriella Di Pancrazio for her help and support.

I would like to thank Robert Auerbach, Ida Caputo, Miriam Rosenzweig, etc., for administrative work.

I would like to thank all those that have shaved me to be the present me.

Thank you all, my teachers and friends, mentioned or not-mentioned, you will always be in my heart, wherever I go.

P. S.: I would like to thank Drs. Chantal Autexier and Jean-Pierre Perreault for critical reading of the thesis and their precious suggestions.

I would like to thank my oral defense committee supervisors: Drs. Chantal Autexier, Lea Brakier-Gingras, Anne Gatignol, Ralph Germinario and Matthias Gotte for their precious suggestions on my thesis.

Preface

The dissertation is written in the form of manuscript-based thesis, according to the "Guidelines for thesis preparation" from "the Faculty of Graduate Studies and Research (FGSR) of McGill University ". The manuscript-based thesis must conform to the following guidelines:

"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 of one or more published papers. The texts must be bound together as an integral part of the thesis."

"The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory."

" The thesis must conform to all other requirements of the "Guidelines for thesis preparation" in addition to the manuscripts."

"The thesis must include the following: a table of contents; an abstract in English and French; an introduction which clearly states the rational and objectives of the research; a comprehensive review of the literature (in addition to that covered in the introduction to each paper); a final conclusion and summary."

"...where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis."

"The candidate must have made a substantial contribution to co-authored papers included in the thesis. The candidate is required to make an explicit statement of the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled "Contribution of Authors" as a preface to the thesis."

I have included in the thesis two published papers, which are in chapter 2 and 3, and one manuscript sent for publication, which is in chapter 4. Chapter 1 is a general review of literature. Chapter 2, 3, and 4 contain their own abstract, introduction, methods, results, discussion, acknowledge-ments, and references. Before Chapters 2, 3, and 4, I have included a preface according to the Guidelines. A general discussion and references are included in Chapter 5. The importance and originality of the research are described in the last chapter, which is Chapter 6.

The papers and manuscript presented in the thesis are the following:

Chapter 2) **Shen, N.**, L. Jetté, C. Liang, M. A. Wainberg, and M. Laughrea. 2000. Impact of human immunodeficiency virus type 1 RNA dimerization and reverse transcription and dissociation of dimerization from packaging. J. Virol. 74:5729-5735.

Chapter 3) **Shen, N.**, L. Jetté, M. A. Wainberg, and M. Laughrea. 2001. Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization. J. Virol. 75:10543-10549.

Chapter 4) **Shen, N.**, L. Jetté, C. Liang, M. A. Wainberg, and M. Laughrea. 2002. Intergenic suppressor mutations in the human immunodeficiency virus Gag region restore genomic RNA dimerization and viral replication in viruses bearing a G/C-poor DIS, but not in viruses bearing a denatured DIS. Sent for publication.

Contribution of co-authors: The candidate contributed all the *in vivo* data in chapters 2) to 4), and was responsible for all the research described in chapters 2) to 4). Louis Jetté provided technical assistance in some of the plasmids preparations and Dr. Chen Liang provided some of the plasmids in 2) and 4). Louis Jetté contributed the *in vitro* data in 3). Dr. Mark A. Wainberg provided the biocontainment facilities

throughout this study. All the research was done under the supervision of Dr. Michael Laughrea.

Other relevant research that the candidate has done but not in the thesis:

4) Laughrea, M., **N. Shen**, L. Jetté, 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.

5) Laughrea, M., **N. Shen**, L. Jetté, 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: 226-234.

Table of contents

List of figures and tables

List of abbreviations

ACS: autocomplementary sequence, nucleotide 258 to 262 in HIV-1Lai genomic RNA, a synonym of DIS (dimerization initiation site); also called palindrome

Asp-Thr-Gly : Aspartic acid-Threonine-Glycine

BMH: branched multiple hairpin

CA: capsid protein

CCHC: Cysteine-Cysteine-Histidine-Cysteine

CD4 : cluster determinant

cDNA: complementary deoxyribonucleic acid

CRM1 : chromosomal region maintenance

CypA : cyclophilin A

Cys-His : Cysteine-Histidine

D: dimer

DC-SIGN : dendritic cells specific ICAM-3 grabbing nonintegrin

D-D-35-E: Aspartic acid-Aspartic acid-35-Glutamic acid

DIS: dimer initiation site

DLS: dimer linkage structure

DNA: deoxyribonucleic acid

Env: envelope

ER: endoplasmic reticulum

Gag: group antigen (viral structural protein precursor)

Gag-pol: group antigen-polymerase fusion protein (viral enzymatic protein precursor)

HIV-1: human immunodeficiency virus type 1

HIV-2: human immunodeficiency virus type 2

hnRNP: heterogeneous nuclear ribonucleoproteins

hRip: human Rev interaction protein

HTLV-BLV: Human T-cell leukemia virus-bovine leukemia virus

ICAM-3: intercellular adhesion molecule-3

ICAP: intracytoplasmic A-type particle

IN: integrase

Kb: kilobases

KD: kilodalton

KLD: kissing loop domain; also called SL1

KLH: kissing-loop hairpin; also called DIS hairpin, or DIS stem-loop

LDI: long distance interaction

Leu: Leucine

LFA-1: leukocyte function-associated antigen-1

Loop B: nucleotide 247 and 271 to 273 in HIV-1Lai genomic RNA

Loop C: nucleotide 255 to 263 in HIV-1Lai genomic RNA

LTR: long terminal repeat

M: monomer

MA: matrix protein

MHC: major histocompatibility complex

MHR: major homology region

MMTV: mouse mammary tumor virus

M-PMV: Mason-Pfizer monkey virus

NC: nucleocapsid protein; called NC, or NC domain, when part of the p55gag polyprotein, and NCp7 after proteolytic maturation

Nef: negative factor

NES: nuclear export signal

NF-ATc: nuclear factor of activated T cells

NLS: nuclear localization signal

NMR: nuclear magnetic resonance spectroscopy

Nt: nucleotide

PBS: primer binding site

PPT: polypurine tract

PR: protease

P-T/S-A-P: Proline-Threonine/Serine-Alanine-Proline

PTAP: Proline-Threonine-Alanine-Proline

R: repeat sequence at the 5' of the genome RNA

Rab: Rev/Rex activation domain-binding protein

RNA: ribonucleic acid

RRE: Rev responsive element

RT: reverse transcriptase

SIVcpz: Simian immunodeficiency virus, Chimpanzee

SIVsm: Simian immunodeficiency virus, Sooty Mangabey monkey

SL1-4: stem loop 1-4

snRNP: small nuclear ribonucleoproteins

sssDNA: strand strong stop DNA

Stem B: nucleotide 243 to 246 and 274 to 277 in HIV-1Lai genomic RNA

Stem C: nucleotide 248 to 254 and 264 to 270 in HIV-1Lai genomic RNA

SU: surface protein

TAR: trans-activating response element

TCID50: 50% tissue culture infectious dosage

TM: transmembrane

Tsg101: human tumor susceptibility gene 101

U3: unique sequence at the 3' of the genome RNA

U5: unique sequence at the 5' of the genome RNA

Vif: viral infectious factor

Vpr: viral protein R

Vps4: vacuolar protein sorting 4

Vpu: viral protein U

Chapter 1

Literature review

1.1. Introduction to Retroviruses

Retroviruses, to which HIV-1 (human immunodeficiency virus type 1) belongs, consist of a diverse family of enveloped RNA viruses and belong to the retroviridae family. They have been divided into two categories-simple and complex according to the organization of their genomes (30, 125). All retroviruses contain three major coding regions, which encode for gag, pol and env polyproteins. Those that contain only three coding domains are termed simple retroviruses, while those that also encode for additional proteins are named complex retroviruses. Gag directs the synthesis of at least three proteins: matrix, capsid and nucleoprotein. Pol encodes for at least two proteins: reverse transcriptase and integrase. Env encodes for the surface and transmembrane proteins.

Based on pathogenicity, the retroviruses used to be divided into three subfamilies: the oncoviruses, the lentiviruses and the spumaviruses (130). According to evolutionary relatedness, retroviruses are subdivided into seven groups or genuses: alpharetrovirus, betaretrovirus, gammaretrovirus, deltaretrovirus, epsilonretrovirus, lentivirus, and spumavirus (72). The first five groups are oncoviruses. HIV, the causative agent of acquired immunodeficiency syndrome (AIDS), belongs to the lentiviruses group (from the Latin *lentus* for 'slow'). Spumaviruses cause no known disease and relatively less is known about them (167).

All oncoviruses excluding the human T-cell leukemia virus-bovine leukemia virus (HTLV-BLV) are simple retroviruses. HTLV-BLV, lentiviruses and spumaviruses are complex retroviruses (167).

Retroviruses have a density of approximately 1.16 g/ml, and a diameter of 80-120 nm in thin-section electron microscopy (168). Several distinctive morphologies of retroviral particles have been identified by thin-section techniques and retroviruses can be classified into four types according to virion structure: type A, type B, type C and type D, which have been associated with different pathways of viral assembly (Fig. 1). There are two major patterns of retroviral assembly that have been observed under electron microscopy (55).

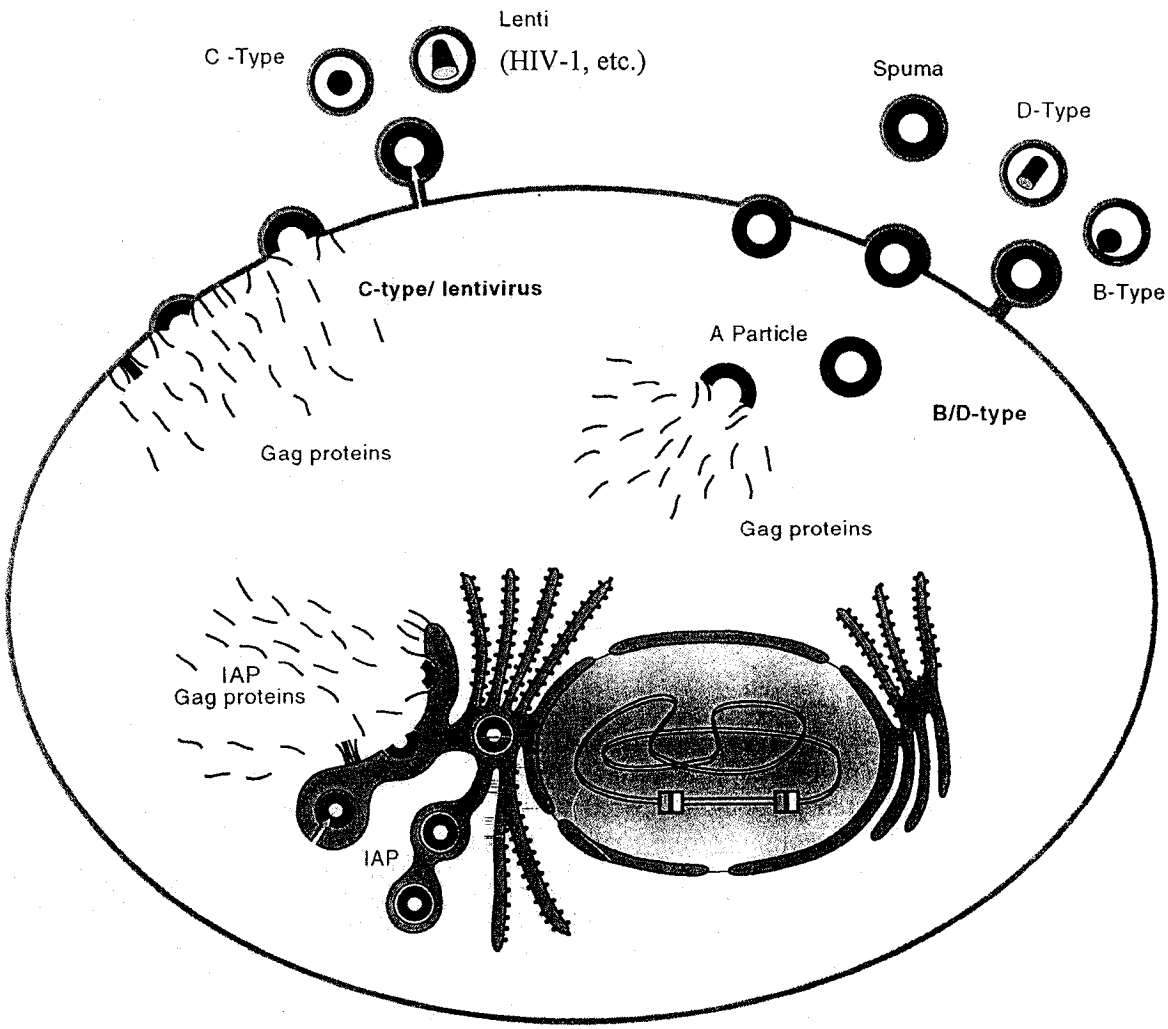


Fig. 1. Retroviral classification according to distinctive morphologies by thin-section electron microscopy, representing two different patterns for retroviral assembly (modified from R. Swanstrom, 1997).

Type A particles: in the first pattern of retroviral assembly, the Gag-containing polyproteins assemble in the cytoplasm to form a stable structure, called type A particle or intracytoplasmic A-type particle (ICAP). These particles are later transported to the cellular membrane and envelope protein is acquired during budding (156). They appear to have an electron-lucent center and one or two concentric electron-dense rings, looking like a doughnut in thin sections (168). Immature retroviruses often resemble type A particle in morphology. Type B, and D viruses, and spumaviruses take this mode of assembly pathway (156).

Type B viruses have an eccentrically placed round core, the prototype being mouse mammary tumor virus (MMTV) (168).

Type D viruses show a bar-shaped core, the prototype being Mason-Pfizer monkey virus (M-PMV) (168).

Type C particles: in the second pattern of retroviral assembly, retroviruses do not appear to assemble their Gag and Gag-Pol proteins into ICAPs, at least according to the electron microscopy. Instead, they assemble their cores at the plasma membrane prior to envelopment and release. This type of assembly has been defined as type C, which is the pattern most often seen for retroviruses, including HIV-1 and the other lentiviruses. Type C particles contain a round or sometimes slightly angular core centered in the middle of the particle. Many mature, infectious viral particles having a condensed core of electron-dense material will resemble type C morphology (167)

Retroviruses are enveloped, positive-stranded RNA viruses that contain a dimeric genome. Their genomes have several interesting characteristics (reviewed in 130): 1) retroviruses are the only viruses that contain dimeric genome (8, 24); 2) the plus strand RNA genome is not used directly as mRNA immediately after infection; 3) retroviruses contain a specific cellular tRNA as a primer (106) and the RNA genome will be reverse transcribed by the viral reverse transcriptase into a complementary DNA (4, 163) and integrated into host genome as a provirus; 4) the genome is synthesized and processed by cellular machinery.

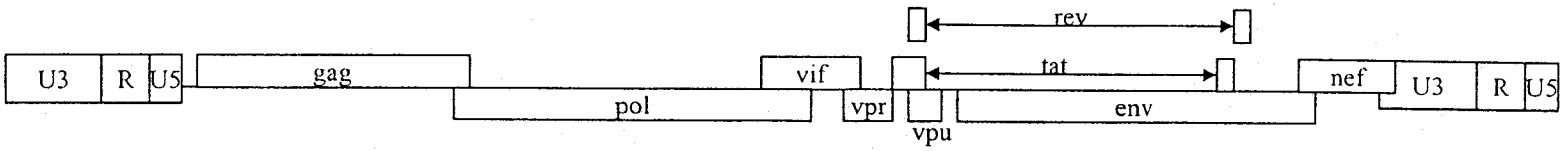
1. 2. General description of HIV-1

1. 2. 1. Overview

HIV-1 has been the subject of intense investigation for twenty years. It is one of the causative agents for acquired immunodeficiency syndrome (AIDS). AIDS is characterized by progressive depletion of CD4+ T cells, leading to immunosuppression with opportunistic infections, neurological abnormalities, various neoplasms and other clinical manifestations. AIDS first appeared on the public agenda in 1981 (59, 109, 149). By 1983, the etiologic agent of AIDS, a retrovirus named human immunodeficiency virus was isolated (6). By the mid-1980s, it was evident that there were two rather different types of HIV: HIV-1 and HIV-2. Most of the infections were caused by HIV-1 and a small number of people were infected with HIV-2, which seemed to be somewhat less infectious than HIV-1 and had a slower pace of disease development (17, 79) (reviewed in 70). Until 2001, it has been estimated that more than 23 million people in the world have died of HIV-1 infection and almost 37 million people are now infected (171).

Phylogenetic analysis indicates that HIV-2 probably originated from SIVsm (simian immunodeficiency virus, sooty mangabey monkey), and HIV-1 came from chimpanzee SIVcpz (70). To date, HIV-1 can be divided into three groups: M (main), O (outlier) and N (new). The 'M' group can currently be subdivided into 9 subtypes (A to K, except E and I), which are the majority of circulating strains, with variable endemicity. For example, subtype B viruses cause most HIV-1 infections in North America and Europe, and have been the first AIDS viruses recognized. Subtype C predominates India and Southern Africa while subtype E predominates Southeast Asia. A larger variety of subtypes have spread in sub-Saharan Africa, mostly A, C and D (reviewed in 70). The group M subtypes are about 25% to 30% different from each other and share a single common ancestor that has been dated at around 1930 (1915-1941) (88). There are no subtypes identified for group O and N viruses (70).

1. 2. 2. HIV-1 genome and proteins



HIV-1 genomic RNA

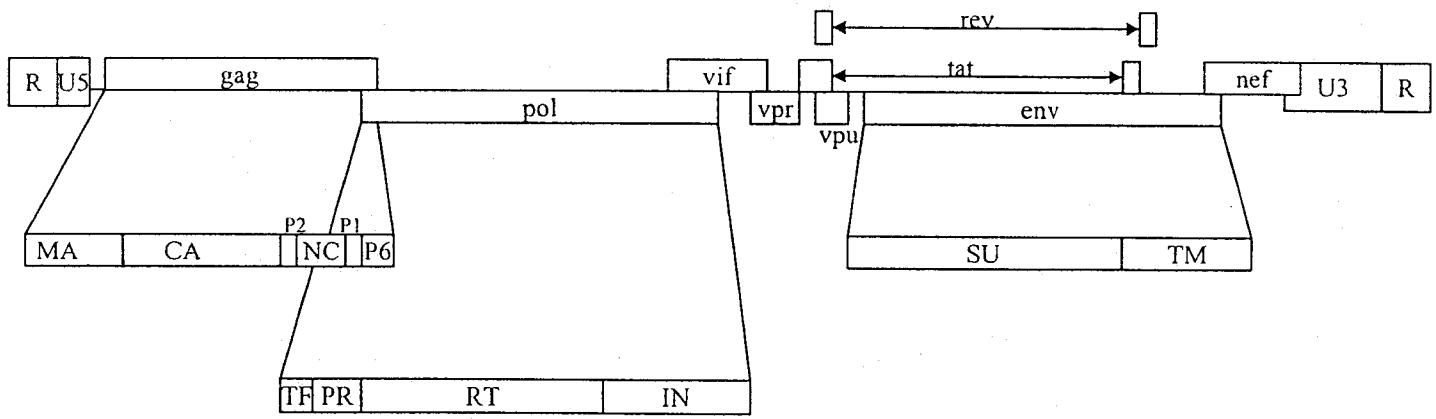


Fig. 2. Organization of HIV-1 viral sequence in proviral DNA and genomic RNA (adapted from ref. 168)

The HIV-1 genome consists of two identical RNAs that appear noncovalently associated near their 5' ends. The 9.2 kb RNA genome contains nine open reading frames, encoding altogether 15 proteins and two small spacer peptides (reviewed in 48, 49) (Fig. 2). Three of these open reading frames encode the Gag, Pol, and Env precursors, which are then further cleaved into individual proteins. The cleaved products from Gag and Env are structural proteins that form the viral core and outer membrane envelope. Pol encodes the viral enzymes and is also packaged in the virion particle (48). Gag proteins are initially produced as a polyprotein $pr55^{Gag}$, which is then processed into four proteins: MA (p17 matrix), CA (p24 capsid), NC (p7 nucleocapsid), and p6, as well as two small peptides: p2 and p1. The precursor protein of Env, gp160, is cleaved into two proteins: SU (surface or gp120) and TM (transmembrane or gp41) proteins. There are three Pol proteins: PR (protease), RT (reverse transcriptase) and IN (integrase). HIV-1 also encodes six additional accessory/regulatory proteins. Although they are not required for HIV replication in *in vitro* culture, they are essential for HIV pathogenesis. Vif, Vpr, and Nef are present in the viral particle. Tat and Rev, are important gene regulatory proteins. Finally, Vpu enhances virion assembly (48).

1. 2. 3. Model of HIV-1 virion

The outermost layer of a HIV-1 virion is the viral envelope, which is a lipid bilayer derived from the host plasma membrane. HIV-1 envelope proteins gp120 and gp41, noncovalently associated and oligomerized most likely into trimers, reside on the viral surface, protruding like a glycoprotein spike. The trimer of gp41 anchors the spike into the viral membrane and the trimer of gp120 can bind to the cellular CD4 receptor and coreceptors. Underneath the lipid membrane, is located the outer shell formed by MA proteins. Inside the MA layer is found the capsid formed by a shell of CA proteins. The capsid is also defined as the outer layer of the viral core, whose shape is conical in HIV (168). Inside the viral core, two identical RNA molecules are coated with NC proteins. A small number of RT and IN are present in the core and there might be interaction between RNA-NC complex and these two proteins. A tRNA

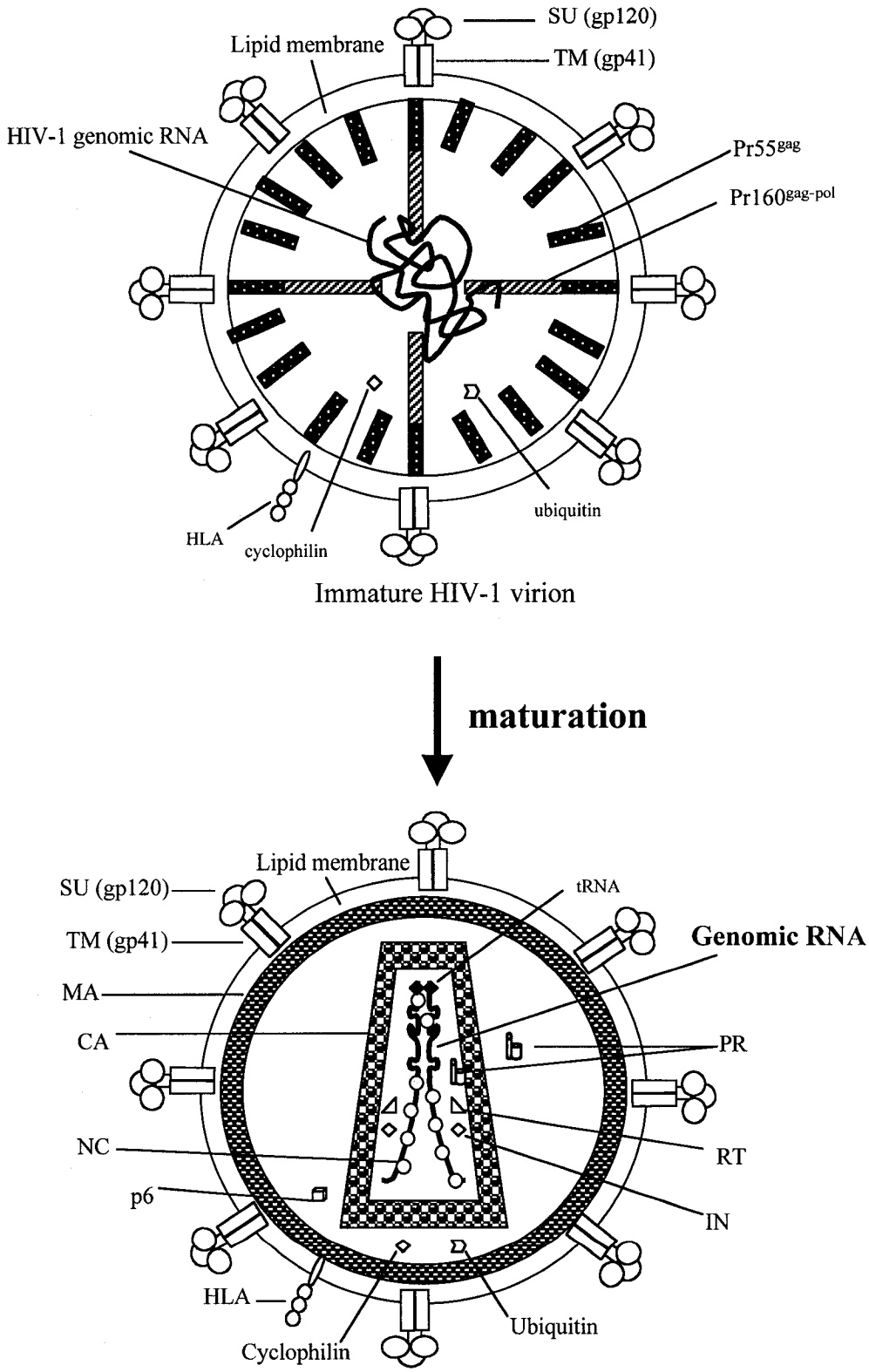


Fig. 3. Maturation of HIV-1 virion (modified from ref. 168)

primer is also packaged in the viral core. PR is present in the virion, perhaps both inside and outside the core. Some host proteins, like MHC (major histocompatibility complex) class I proteins, are incorporated in the viral membrane; Other cellular proteins, like cyclophilin A (47, 162) and ubiquitin (168), etc., have been found inside the HIV-1 virion (Fig. 3).

1. 2. 4. Gag gene products

MA (membrane-associated or matrix): it forms the N-terminal domain of the Gag polyprotein. MA plays several roles in viral replication, such as: binding and targeting of Pr55Gag and Pr160Gag-Pol into plasma membrane; incorporation of Env glycoproteins into newly formed virions; nuclear import of the preintegration complex and possibly export of Pr55Gag and genomic RNA (127). A N-terminal myristate group and basic residues located within the first 50 amino acids are associated with membrane targeting. The crystal structure of HIV-1 MA shows a trimeric form (68). Trimerization of MA might be very important because mutations that disrupt trimerization abolish viral assembly and because the basic residues directing membrane localization are present on the membrane binding surface of the trimer. The model for MA membrane binding has been suggested as following: the three myristate groups of the trimer will insert into the inner cellular lipid bilayer and the basic residues of the membrane-binding surface will interact with the acidic phospholipids of the cellular membrane, resulting in membrane binding (reviewed in 48). Env incorporation was shown to involve the long cytoplasmic tail of gp41 being accommodated by the holes in the MA lattice formed upon MA trimerization and further stabilized by MA-Env interaction (68, 121). It has been suggested that there are two nuclear localization signals, one located at the amino-terminal region of MA (NLS-1) (nuclear localization signal) (23, 170), and the other one (NLS-2) located at the carboxyl-terminal region of MA (62).

CA: CA lies in the second domain of the Gag polyprotein. The study of Rous Sarcoma Virus suggests that approximately 1500 copies of CA are present in each virus (169). The HIV-1 CA has 231 residues and comprises two domains: N-terminal

domain (residues 1 to 145), and C-terminal domain (from residues 151-231) (159). CA plays an essential role in virion assembly and incorporation of cellular protein cyclophilin A (CypA) (48). The N-terminal domain plays an essential role in early steps of the virus life cycle. Mutations in this domain result in deformed core structures, poor initiation of reverse transcription and loss of infectivity, without affecting the amount of virion particles (159). The N domain has been shown to be involved in the incorporation of cellular protein cyclophilin A (CyA). CA has been implicated in Gag-Pol incorporation, though the exact region of CA being involved is not quite clear (48). The C-terminal domain mainly plays a role in assembly and is involved in CA dimerization and Gag oligomerization (48). Located in the central region of CA, from residue 153 to 172, is the highly conserved major homology region (MHR). Mutations in this region affected many steps of HIV replication cycle, such as assembly, maturation and infectivity (49). MHR might be involved in Gag-Gag oligomerization (138) and might also influence membrane-binding affinity (48). Mutations in MHR impaired membrane binding, and Gag oligomerization. It has been suggested that MHR affected viral assembly by mediating the Gag conformation optimal for protein-protein (i.e. Gag-Gag, etc.) and protein-membrane (i.e. Gag with cellular membrane) interaction during this process (138).

NC: The nucleocapsid protein (NCp7, in HIV-1, following proteolytic maturation) lies in the third domain of Gag polyprotein. The major role of the NC domain of the Gag polyprotein is to bind specifically to the packaging signal of genomic RNAs and deliver them to the assembling virion. NCp7 is a 55-residue basic protein, which can also bind nonspecifically to nucleic acids (166), resulting in coating of the genomic RNA. Via this property, NCp7 may protect genomic RNAs from nucleases, compact them within the core, and promote other nucleic acid-dependent steps during viral replication, such as $tRNA^{lys3}$ primer annealing, melting of RNA secondary structures, DNA strand transfer (77, 140), efficiency of reverse transcription (140), or integration (48). NCp7 also promotes RNA dimerization, converting an unstable dimer into mature, stable dimer (45).

The structure of NC contains two zinc finger domains (CCHC) (Cysteine-Cysteine-Histidine-Cysteine) (from residue 13-30; 35-51) separated by basic linker

segment (residue 31-34) and flanked by amino- and carboxyl terminus (residue 1-12; 52-55) (35). Zinc finger domains are crucial for HIV replication (81) and have been involved in genome recognition and encapsidation (35). Mutations of the flanking base residues also impaired viral RNA encapsidation (81). The nuclear resonance magnetic studies of NCp7 bound to SL3 RNA stem-loop (the primary encapsidation signal) showed the two Zinc fingers are involved in highly specific interactions with bases in the SL3 loop (35). The N-terminal residue 3-11, which contains four conserved basic residues (Lys3, Arg7, Arg10 and Lys11) (Lys=lysine, Arg=arginine), make nonspecific, electrostatic interactions with phosphodiester groups of the SL3 stem. Highly conserved residue arginine (Arg32) interacts specifically with one base of the SL3 loop, and Asn5 (asparagine) with two bases from SL3 stem. It has been proposed that the two Zinger fingers contribute to sequence-specific RNA binding and hypothesized that the basic N-terminal domain is implicated in non-specific RNA interaction, leading to the nucleic acid chaperone activity of NCp7 protein. The detailed mechanism of the latter needs to be further clarified (reviewed in 40). The structure of NCp7, and the interaction of $tRNA^{lys3}$ with HIV-1 NCp7 have been shown by NMR studies (115, 154, 164).

p6: the p6 domain is located at the C terminus of the pr55Gag. p6 consists of 51 residues and plays a critical role in the release of virions from the membrane of infected cells. The N terminus of p6 contains a highly conserved motif, P-T/S-A-P (Proline-Threonine/Serine-Alanine-Proline), called late (L) domain, which is especially important for the p6 virus release function (36). Protein p6 is newly recognized as the predominant phosphoprotein of HIV-1 particle. There are several forms of phosphorylated p6 and the phosphorylation occurs at multiple residues (120). It is assumed that p6 acts as molecular adapters recruiting cellular budding proteins to the budding site and that the conserved L domain serves as docking site for those proteins. A number of p6 has been shown to be ubiquitylated, although the functional importance of p6 ubiquitylation is not very clear (12). The C terminus of p6 contains a Leucine-rich motif, which has been implicated in the incorporation of Vpr into virion particle during assembly (87, 132).

p2, p1: p2 and p1 are two spacer peptides. p2 lies between CA and NC. p1 lies between NC and p6. p2 has been shown to be important for viral assembly and infectivity (81). The p2 domain plays a key role in Gag multimerization (>1,000 kDa) (kilodalton), an intermediate step in viral assembly, which may occur shortly before Gag targeting to the plasma membrane (116). The p2 domain may contribute to the specific recognition of RNA packaging signal and stimulate genomic RNA encapsidation (81). Hydrophobic amino acids in the p2 domain were shown to rescue a genomic RNA packaging defect (141) and a viral replication defect (98). It is possible that p2 also plays a role in genome RNA dimerization (147). Mutations in p2/NC cleavage site in Gag precursor impaired genomic RNA dimer stability and virion core formation (145). The role of p1 is not clear.

1. 2. 5. Pol gene products

PR (protease): After the newly formed virion particle is released from the cell membrane, it is noninfectious because it has not yet undergone the maturation process, in which the Gag and Gag-Pol precursors are processed by the viral encoded PR, and conformational changes occur to produce infectious mature viral particle. HIV-1 PR cleaves at several polyprotein sites, with variable cleavage efficiencies among different sites. For HIV-1 Gag, the processing occurs sequencially at primary (p2/NC), secondary (MA/CA, p1/p6) and tertiary sites (CA/p2) (145). PR is part of Pol. It is crucial in the formation of infectious viral particle that the PR is activated and gets released from the Gag-Pol polyprotein. PR has been identified as an aspartic proteinase, whose common feature is the presence of two aspartic acid residues. Since each PR contains only one Asp-Thr-Gly (aspartic acid-threonine-glycine) motif, it needs to dimerize to make an active form. Active PR is a dimer consisting of two identical 11-kDa monomers, each with 99 amino acid residues. Each monomer provides an Asp25 catalytic residue from their conserved Asp-Thr-Gly triad. The three dimensional structure of HIV-1 PR is as follows: The two Asp25 in the active site each lies in a loop at the dimer interface. The two loops are interlocked and form the hydrophobic core of the PR enzyme. In the middle exists a long cleft between these

two monomers, where substrate can bind. The catalytic residue Asp25 lies on the floor of the cleft. Finally, each monomer has a flap stretching over the substrate-binding cleft. The flexible flap will undergo dramatic change when PR binds to substrates, allowing entry of the substrate and leaving of the cleaved products (156).

PR has been one of the main targets for drug design to inhibit HIV replication, since proteolytic processing by PR is an essential step in the viral life cycle. PR inhibitors have been widely used in treatment of HIV infection, however, under the treatment of these drugs, resistant viral variants have been observed (48).

RT (reverse transcriptase): All retroviruses contain the enzyme RT: it converts genomic RNA into a proviral DNA intermediate. RT possesses both RNA-dependent and DNA-dependent DNA Polymerase activities, and an RNase H activity that removes the RNA strand of RNA-DNA duplex formed during proviral DNA synthesis. HIV-1 RT uses $tRNA^{lys3}$ as primer. RT is a heterodimer, consisting of subunit p66 and p51, each of which contains four subdomains called fingers, palm, thumb, and connection, with p66 containing the RNase H activity (160). RT is the most extensively studied target for anit-HIV-1 drug design. There are two kinds of RT inhibitors that have been developed and used in clinic: nucleoside analogs and non-nucleoside inhibitors (48).

IN (integrase): After reverse transcription, the proviral DNA will be integrated into host chromosome by integrase. IN has been shown to possess: specific and nonspecific endonuclease (3'-end processing) activity; DNA joining activity (also called strand transfer); disintegration activity and DNA binding activity (reviewed in 20).

IN has 288 amino acid residues and is active in a multimeric form. It can be divided into three domains: the N-terminal domain (residues 1-55), the catalytic domain (residues 50-212) and C-terminal domain (residues 220-270).

The N-terminal domain contains a conserved Zinc-finger-like HHCC motif, which is essential for integration and viral replication. HHCC domain, in the presence of Zinc ion, might be very important for IN multimerization, which might allow the IN to form a stable complex with the viral DNA ends (20).

The catalytic domain contains a triad of acidic residues, the conserved D-X_{39-58}-D-X_{35}-E motif (D_{64}-D_{116}-E_{152} in HIV-1 IN, D stands for aspartic acid and E stands for glutamic acid; the D and E are always separated by 35 amino acids) (48). The three acidic residues are essential part of IN active site, all of which are required for integration in vivo. Divalent metal ions bound in this domain have been suggested to participate in the catalytic reaction of IN. The D-D-35-E motif also determines the sequence specificity of their DNA substrate (i.e. CA/TG sequence is always positioned two base pairs from the proviral DNA ends). The C-terminal domain, sometimes called DNA-binding domain, is the least conserved region in the IN. This region has intrinsic DNA-binding activity and binds to non-branched DNA substrates. This domain is essential for 3'-end processing and integration activity (20).

1. 2. 6. Env gene products

SU (Surface protein): Env polyproteins are synthesized and glycosylated in the endoplasmic reticulum and later cleaved into SU (gp120) and TM (gp41) subunits. gp120 binds to the cellular receptor CD4 and coreceptors, and determines the host range of the virus. The primary sequence of gp120 contains five variable regions (V1-V5) and four constant regions (C1-C5) (73). According to X-ray crystal structure, gp120 has 3 structural domains, inner and outer domain, connected by a small bridging domain. gp120 and gp41 might contact each other through the inner domain (42). The C4 region, as well as C1, C2 and C3 domains have been suggested to be involved in CD4 binding (73). The variable regions do not seem to play a direct role in receptor binding. V3 region is essential in determining HIV macrophage tropism, probably by interacting directly with the chemokine coreceptor (73). However, there might be other tropism determinants exist, like C4 region of gp120, and the fusion peptide region of gp41, etc.. V1/V2 might be involved in tropism determination and coreceptor binding as well (26).

TM (transmembrane protein): TM or gp41 is a 345-residue transmembrane protein. Going from its N-terminus to its C-terminus, one finds: the fusion peptide, which is hydrophobic, glycine-rich and crucial for membrane fusion; two hydrophobic

heptad repeats (called N- and C-peptides regions respectively), separated by a loop region containing two cysteines; the transmembrane segment (tm); and the cytoplasmic region (cyto) (25). In the native state of env glycoprotein complex, much of the gp41, including its fusion peptide, is embedded in the cellular membrane (42). Upon binding of gp120 to CD4 and coreceptor, conformational changes occur, leading to translocation of the N-peptide region and exposure of the fusion peptide region of gp41, followed by its insertion into the plasma membrane, forming the 'prehairpin intermediate'. Further conformational changes in gp41 may be needed to create the 'hairpin form' in which the viral and cellular membranes are brought together close enough to allow the fusion to occur (114). Details of this process remain to be clarified.

1. 2. 7. Accessory/Regulatory proteins

Vif: Viral infectivity factor (Vif) is a basic, phosphorylated 23-kDa protein that is expressed from a singly spliced mRNA in infected cells. Vif is essential for production of infectious virions and the requirement for Vif in infectivity is cell type dependent. According to the ability to support the replication of Vif$^-$ mutants, three types of cells are divided: nonpermissive, semipermissive and permissive (48). Vif is believed to be able to overcome the inhibitory effect of some of host cell factors in nonpermissive cells, which constitute most of the HIV natural target cells.

Most of Vif has been found in the cytoplasm of the infected cells. The specific incorporation of Vif into the mature virion is very controversial. Some studies suggested that small amount of Vif, approximately 20 to 100 molecules per virion is packaged into the virion (48).

Vif has been suggested to be involved in the late step of viral life cycle, affecting viral assembly and maturation, such as: processing of Gag and Gag-Pol: Vif has been shown to be associated with Gag and Gag-Pol (5, 18), and have an inhibitory effect on the proteolytic activity of PR, leaving intact Gag and Gag-Pol to be packaged; Modulating core morphology; Encapsidation of genome RNA: it has recently been shown that Vif can associate with genomic RNA, and affects viral RNA

folding and genome encapsidation (175). Vif might affect the early step of viral life cycle, i.e., Vif might affect the steps between entry and uncoating due to its effect on virion morphology, thus affecting the proviral DNA synthesis (64).

The C-terminal region of Vif is suggested to interact with some host cellular factor (5), and cellular membrane (57, 58, 151), and Gag (5, 18), and has been implicated in Vif multimerization, which might be a necessary form of an active Vif (64).

Vpr: viral protein R (Vpr) is a 96-residue, 14kDa small protein. Vif is packaged in the virion with an amount comparable to that of Gag, probably by interacting with p6 domain within the Gag precursor (48, 21).

Vpr plays a critical role in viral replication *in vivo,* and has two characteristic functions: one is that Vpr facilitates nuclear import of HIV-1 preintegration complex, which function is critical for viral replication in non-dividing cells (i.e., macrophages); another is that Vpr can induce G2 cell cycle arrest in proliferating cells (i.e., T lymphocytes), in which cell cycle regulation is essential for HIV infection. Besides these two functions, Vpr is also shown to stimulate transcription and regulate cell apoptosis, etc. (21).

Nef (negative factor): Nef contains 206 amino acid residues and there are around 70 molecules packaged into each virion. Nef was originally thought to have a negative effect on HIV-1 replication, thus gained its name "negative factor" (1, 134). Later, it turned out that Nef is essential for HIV pathogenesis: it promotes virion production, causing CD+ T cell depletion, and resulting in development of immunodeficiency syndrome. Nef has been shown to enhance viral infectivity and viral replication, and to downregulate CD4 receptor and MHC class I molecules (48). Besides, Nef has been found to modulate T cell signaling (44). Downregulation of CD4 can prevent superinfection by more HIV virions and prevent retention of env proteins by CD4 in the endoplasmic reticulum, thus facilitating the incorporation of env into the assembling virion (48). It has been suggested that downregulation of CD4 might also help to prevent some inhibitory signals transmitted by CD4 toward HIV transcription. Downregulation of MHC molecules can enable the infected cells to escape the surveillance of cytotoxic T cells. Nef mediated T cell signaling has been

shown to support viral replication and promote viral survival by way of immune evasion (in conjunction with downregulation of MHC molecules) and antiapoptosis (44).

Tat: The transcriptional transactivator Tat is a small, 86 to 102 amino acid protein. Tat regulates HIV gene expression and is essential for HIV replication. It binds to RNA hairpin TAR (transactivating response element: a hairpin at the 5' end of genomic RNA), not like the typical transcription factors which bind to DNA site. In eukarytic transcription system, transcription needs the assembly of RNA polymerase II and basal transcription factors to mRNA and initiates the transcription (139). The rate of transcription is determined by the frequency of initiation and the efficiency of elongation. Without Tat, the RNA polymerase II showed very poor processive elongation, and the basal transcription level is very low. Tat enhances the processivity of transcription and may promote the initiation of transcription as well (139).

Many cellular factors are involved and required in this process. Tat has been suggested to form transcription complexes and recruit cellular factors to phosphorylate RNA polymerase II C-terminal domain (CTD), a process necessary to enable RNA polymerase from an initiating enzyme to be an elongation enzyme (48). Cyclin T1, the cyclin component of CDK9 (cyclin dependent kinase), is one of the cellular factors recruited to phosphorylate CTD, resulting in increased processivity of RNA polymerase II and increased transcription elongation (105). Tat has been shown to bind directly with components of TFIID, the basal transcription factor (139). The transcriptional coactivators P300 and CBP (cAMP-responsive protein) are also recruited to promote transcriptional initiation by linking the basal transcription machinery to transcriptional activators (105).

Rev: It is a 116-amino acid residue protein that is synthesized at early time of infection. It is required to export out of the nucleus unspliced and singly spliced viral RNAs that contain an RRE (Rev response element). RRE is a 234-351 nucleotide RNA sequence in the env-coding region, which contains several stem-loop structures. Rev contains a basic, arginine-rich region at its amino terminus that is important for Rev multimer to bind to RRE. NMR studies have shown that the specific interactions occur between Rev and RRE (139). Rev also contains a leucine-rich effector domain

that mediates Rev functions, possibly with the involvement of cellular proteins. An amino acid sequence "LQLPPLERLTL" (Leu=leucine, Q=gultamine, P=proline, E=glutamic acid, R=arginine, and T=threonine) in this domain has been identified as nuclear export signal (NES), which resembles the peptides in cellular proteins involved in nuclear export, suggesting that Rev-mediated nuclear export might be linked to cellular export system.

HIV RNAs exit the nucleus through the nuclear pore. In order to do that, the RNAs have to bind to Rev through NES, then to an export receptor. Human CRM1 (chromosome region maintenance) has been identified as nuclear export receptor for NES and renamed as Exportin 1. Some cellular proteins, such as hnRNP A1 (heterogeneous nuclear ribonucleoproteins) that also carries export signal might bridge HIV-1 RNA to Exportin 1. Exportin will bind to the nuclear pore through interaction with nucleoporins (proteins involved in nuclear pore formation). Rip/Rab (human Rev interacting protein or Rev/Rex activation domain-binding protein) are nucleoporin-like proteins identified, acting like a dock for exportins by forming a Rip/Exportin 1/Rev-NES complex (48).

Besides nuclear export function, Rev may also be involved in RRE-containing RNA processing, like: splicing, RNA stability, and mRNA translation; Rev has also been shown to counteract the inhibitory effects of CRS (*cis*-acting regulatory sequence) or INS (inhibitory sequence) of HIV-1 RNAs, which inhibit HIV RNA export and promote RNA degradation (139).

Vpu: viral protein U is an 81-residue phosphoprotein primarily located at the internal membrane of the infected cells. Vpu contains a hydrophobic N-terminal membrane spanning domain and a C-terminal cytoplasmic tail (48). Vpu has two distinct functions: one is to downregulate CD4 and another is to enhance the virion release. The first function probably involves the C-terminal domain and the second function may relate to the N-terminal domain. Vpu promotes ubiquitin-mediated degradation of CD4, which is complexed with newly synthesized env, by targeting to the proteasome, thus releasing env and allowing them to reach the plasma membrane to be released. Vpu was shown to be able to oligomerize and form ion channels in the

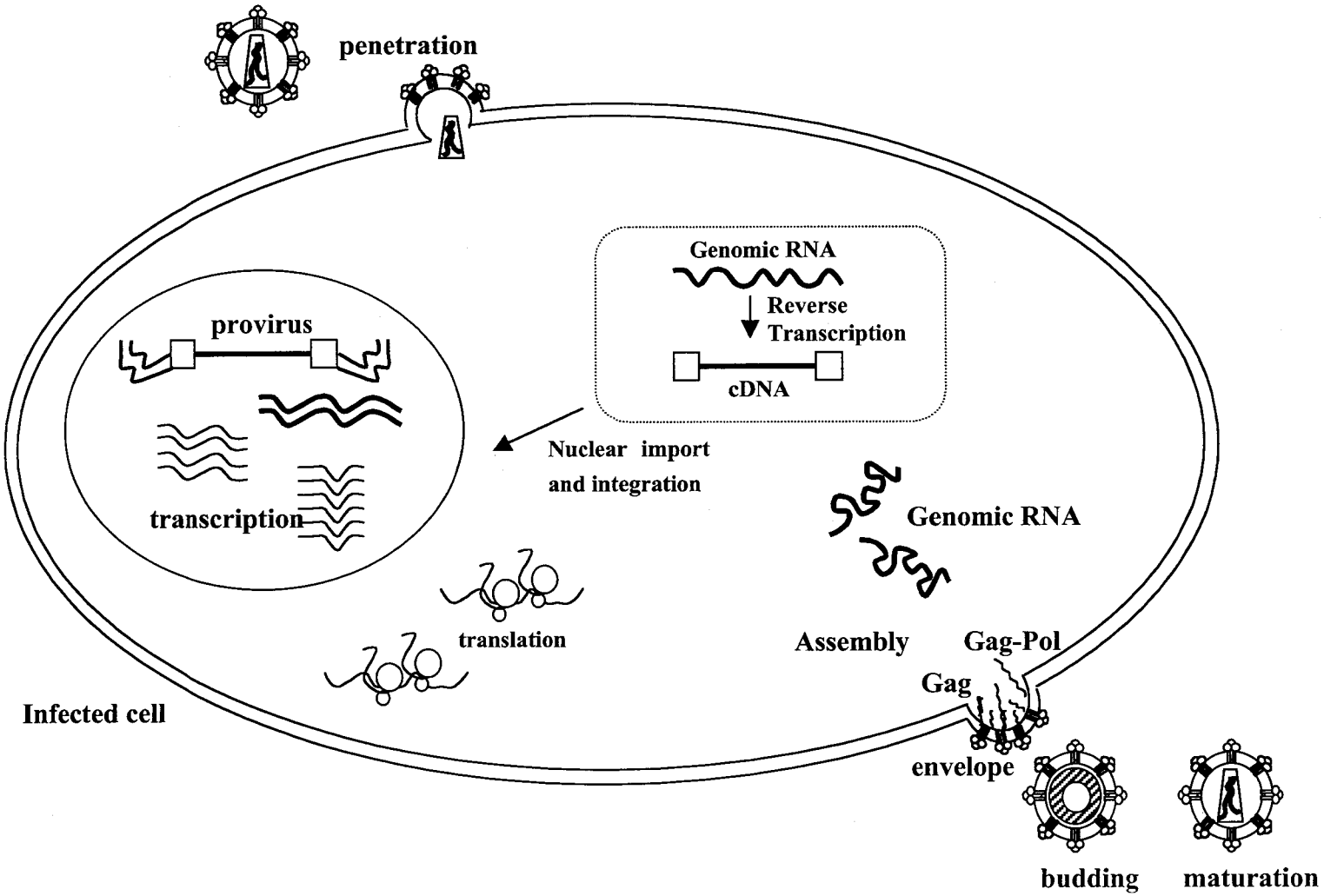


Fig. 4. HIV-1 replication cycle (modified from ref. 104)

plasma membrane, thus facilitating the virion release. Besides the above two functions, Vpu has also been found to downregulate MHC class I molecule (48).

1.2. 8. Replication cycle of HIV-1

The replication cycle of HIV-1 can be divided into several steps: 1) virus-cell attachment, 2) core entry, 3) uncoating, 4) nuclear import, 5) reverse transcription, 6) integration, 7) transcription and splicing, 8) outport, 9) translation, 10) viral core assembly, 11) Env release, 12) Env transport, 13) degradation of CD4, 14) budding and release, 15) viral maturation (Fig. 4)

Step 1): HIV-1 first binds to target cells using semi- or non-specific interactions between the viral envelope and cell surface glycans such as: a) heparan sulphate proteoglycans (113, 117); b) cell-surface lectins, e. g., DC-SIGN [dendritic cells-specific ICAM-3 (intercellular adhesion molecule-3) grabbing nonintegrin] (54); c) cell-associated adhesion factors (114) such as ICAM-1 and LFA-1 (leukocyte function-associated antigen-1) (99). The gp120 envelope glycoprotein then interacts with the CD4 receptor on the target cells, which *in vivo* are mainly T-cells, macrophages and probably dendritic cells (26). This interaction initiates a conformational change in gp120 that facilitate its binding to a chemokine coreceptor, CCR5 or CXCR4 *in vivo*. This is followed by further conformational changes in the gp120-gp41 complex, resulting in fusion of viral and cellular membrane.

Step 2): Once the virion has fused with the target cell, the viral core is released into the cellular cytoplasm.

Step 3): The poorly understood process of 'uncoating' occurs: The capsid core is disassembled; Viral genome RNA is released into the cytoplasm, along with its associated virion proteins. Cellular factor cyclophilin A (CypA), which is packaged within viral particles and interacts with the capsid, may participate in the uncoating process (19, 47, 162).

Step 4): Viral RNA is reverse transcribed into a double stranded complementary DNA, termed 'provirus' (114). This cDNA is synthesized by a reverse transcription complex that contains MA, RT, IN, Vpr, and RNA. Nucleocapsid protein

(NC) may also be involved in reverse transcription, apart from its role in genome encapsidation. The cellular transcription factor NF-ATc (nuclear factor of activated T cells) may participate in the process, acting as a cellular cofactor.

Step 5): On completion of reverse transcription, the complex, now called pre-integration complex, roughly the size of a ribosome, is transported into the host cell nucleus, where the viral cDNA is integrated into the host's DNA (63). This translocation is facilitated by viral proteins, such as MA (22), Vpr (174), and IN (51, 135, 152), which all contain nuclear localization signals (NLS). A 99-base-pair DNA central flap, generated during the synthesis of plus-strand viral cDNA, is a *cis*-acting element that also promotes the cDNA nuclear import efficiency (174).

Step 6): The viral integrase enzyme (IN), located within the pre-integration complex, then catalyzes the integration of the viral cDNA into the host chromosome. In short, IN cleaves and eliminates two terminal nucleotides at the 3' end of each strand of viral cDNA, producing recessed CA_{OH} ends. Second, the 3'-OH groups are used to attack phosphodiester bonds on opposite strands of the target chromosomal DNA, and the viral 3' ends are joined to the target DNA in a direct transesterification reaction; this process is called strand transfer (63). Some DNA synthesis takes place to fill gaps and trim mismatched viral 5' ends (160). Ligation completes the integration of the viral cDNA into the host chromosome. Evidence shows cellular machinery might be involved in this process (165).

Step 7): Viral cDNA is transcribed by the cellular machinery, using the promoter located in the 5' long terminal repeat (LTR), the rate of transcription being greatly enhanced by Tat protein. This generates the full-length viral RNA genome, as well as shorter viral RNAs resulting from single and double splicing of viral genomic RNA; all of these RNAs will be used as templates for viral protein production.

Step 8): The spliced and unspliced RNAs are exported from the nucleus to the cytoplasm in order to be translated or encapsidated. The unspliced transcript (9.2 kb in length) will be used to produce the Gag and Gag-Pol polyproteins. Singly spliced RNAs (4.3 kb in length) will be used to make Vif, Vpr Vpu and Env. Doubly spliced (1.8 kb) RNAs will be used to produce Tat, Rev and Nef (104, 139) (Fig. 5). The

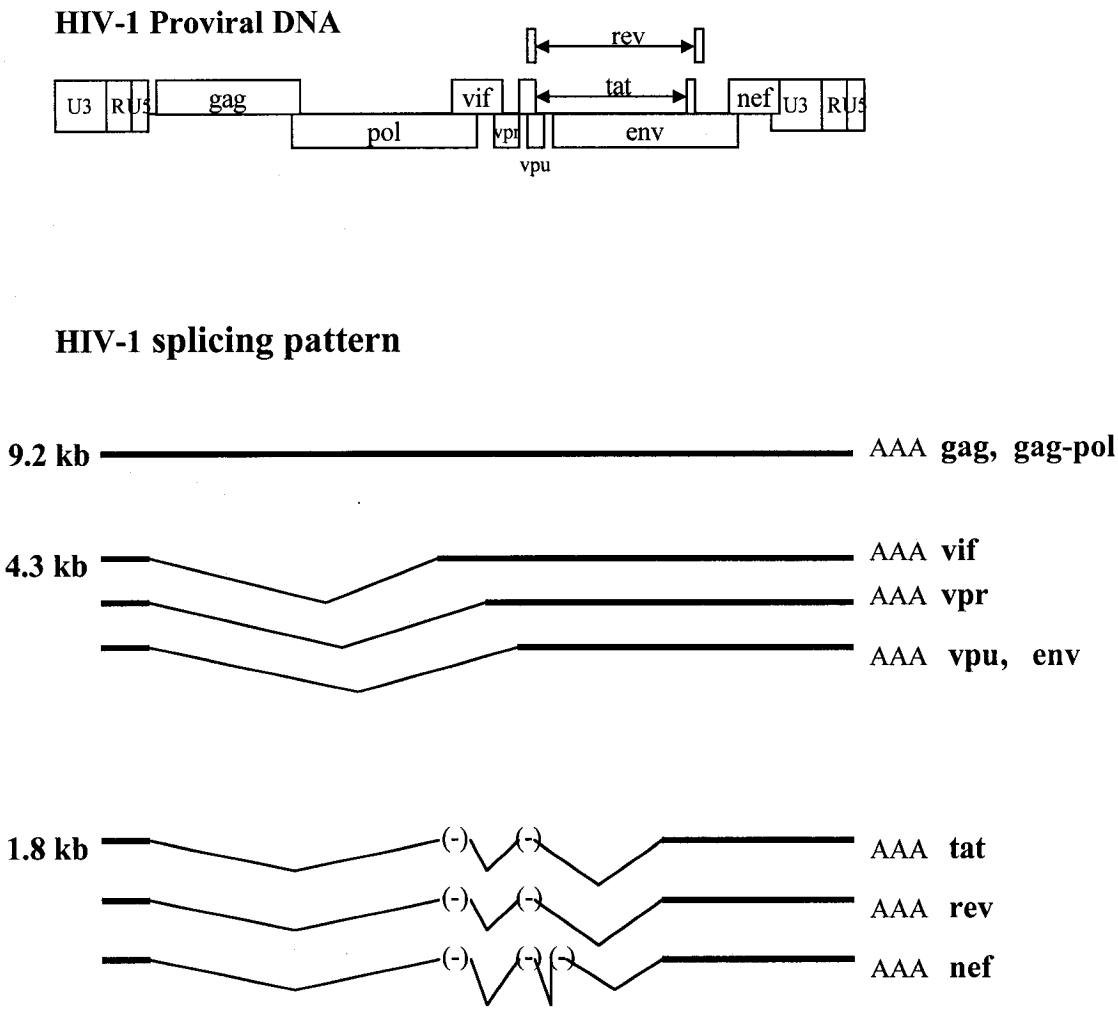


Fig. 5. Multiple splicing pattern for HIV-1 genomic RNA (adapted from ref. 39)

doubly spliced 1.8 kb RNAs form the early class of viral RNAs. They are the first HIV-1 nuclear RNAs to enter the cytoplasm and get translated. Once synthesized, Rev enters the nucleus to facilitate the export of unspliced and singly spliced viral RNAs to the cytoplasm; Tat also enters the nucleus to facilitate the elongation step during viral RNA synthesis (85). Unspliced and singly spliced RNAs will accumulate in the cytoplasm.

Step 9): Gag and Gag-Pol polyproteins are synthesized from unspliced genomic RNAs in the cytosolic compartment and later transported to the cytoplasmic face of the plasma membrane to mediate budding. Gag and Gag-Pol must maintain a certain ratio (~20:1 in HIV-1), since large numbers of Gag molecules are needed to serve as precursors to the structural proteins, whereas smaller amount of Gag-Pol polyproteins are needed to produce the catalytic enzymes (PR, RT, and IN). Gag-Pol synthesis is achieved by ribosomes slipping backward one nucleotide (-1 frameshift i.e., in the 5' direction) during genomic RNA translation (156). Env polyproteins are made from singly spliced genomic RNA on the endoplasmic reticulum (ER) membrane.

Step 10): Gag, Gag-Pol polyproteins, regulatory proteins: Vif, Vpr, Nef, genomic RNA and primer $tRNA^{lys3}$ begin to assemble into the viral core particle.

Step 11): After they are produced, Env polyproteins must be properly folded and oligomerized in ER, probably as trimers, before they can be transported to the cell surface. The fully folded and oligomerized Env proteins have the tendency to interact with their receptors CD4, which are also expressed in ER. This interaction has to be prevented in order for Env to be transported to the cell membrane. Vpu can serve this purpose by binding to the cytoplasmic domain of CD4 and inducing the degradation of CD4, thus allowing the release of Env.

Step 12): Env polyproteins are transported to the external surface of the cell membrane through secretory pathway. During this process, Env proteins will be cleaved into SU (surface subunit) and TM (transmembrane unit) by a cellular protease, possibly fusin, in the Golgi apparatus. Env polyproteins are also further modified along the whole secretion process (156).

Step 13): CD4 is downregulated by Nef, and, to a lesser extent Vpu, probably at several aspects: Some CD4 from the Golgi is redirected to lysosome; Some CD4 is prohibited from recycling to the cell surface and is redirected to the lysosome instead; Endocytosis of CD4 at the cell surface is increased due to Nef. Downregulation of CD4 has several benefits, such as: Releasing env from CD4 to be available to be assembled; Releasing a cellular factor $p56^{lck}$ (lymphoid cell protein-tyrosine kinase) from CD4, resulting in increased T cell activation and preventing hazardous superinfection (134).

Step 14): Genomic RNA is packaged by the Gag protein (39, 173), along with $tRNA^{lys3}$, and the viral particle, coated with SU and TM, is released from the cell surface (budding process). HIV-1 does not posses its own membrane fission machinery and thus may need the assistance from cellular machinery to bud. Two cellular proteins: Tsg101 (tumor susceptibility gene 101) and Vps4 (vacuolar protein sorting), belonging to the endosomal machinery, are essential for HIV-1 viral budding (52). Tsg101 contains UEV (ubiquitin enzyme variant) domain that can bind the PTAP motif of HIV-1 $p6^{Gag}$ late assembly (L) domain, which is required for retroviral budding. Depletion of Tsg101 or inhibition of Vps4 all blocked HIV-1 release in infected cells (52), as if Vps4 and Tsg101 acted in a shared pathway. A model for the participation of Vps pathway in HIV-1 budding was proposed: first, Tsg101 is recruited by $p6^{Gag}$ to the site of budding (ubiquitylation of $p6^{Gag}$ might enhance this interaction), then Tsg101 recruits Vps machinery, such as Vps4, which, alone or together with viral or other cellular proteins, form a viral core that is capable of budding (133).

Step 15): As the viral particle is released from the host cellular membrane coated with SU and TM, it is immature: completion of viral assembly requires Gag and Gag-Pol cleavage by PR, maybe with the involvement of Vif (48). The mature virion is now infectious.

1. 3. Reverse transcription

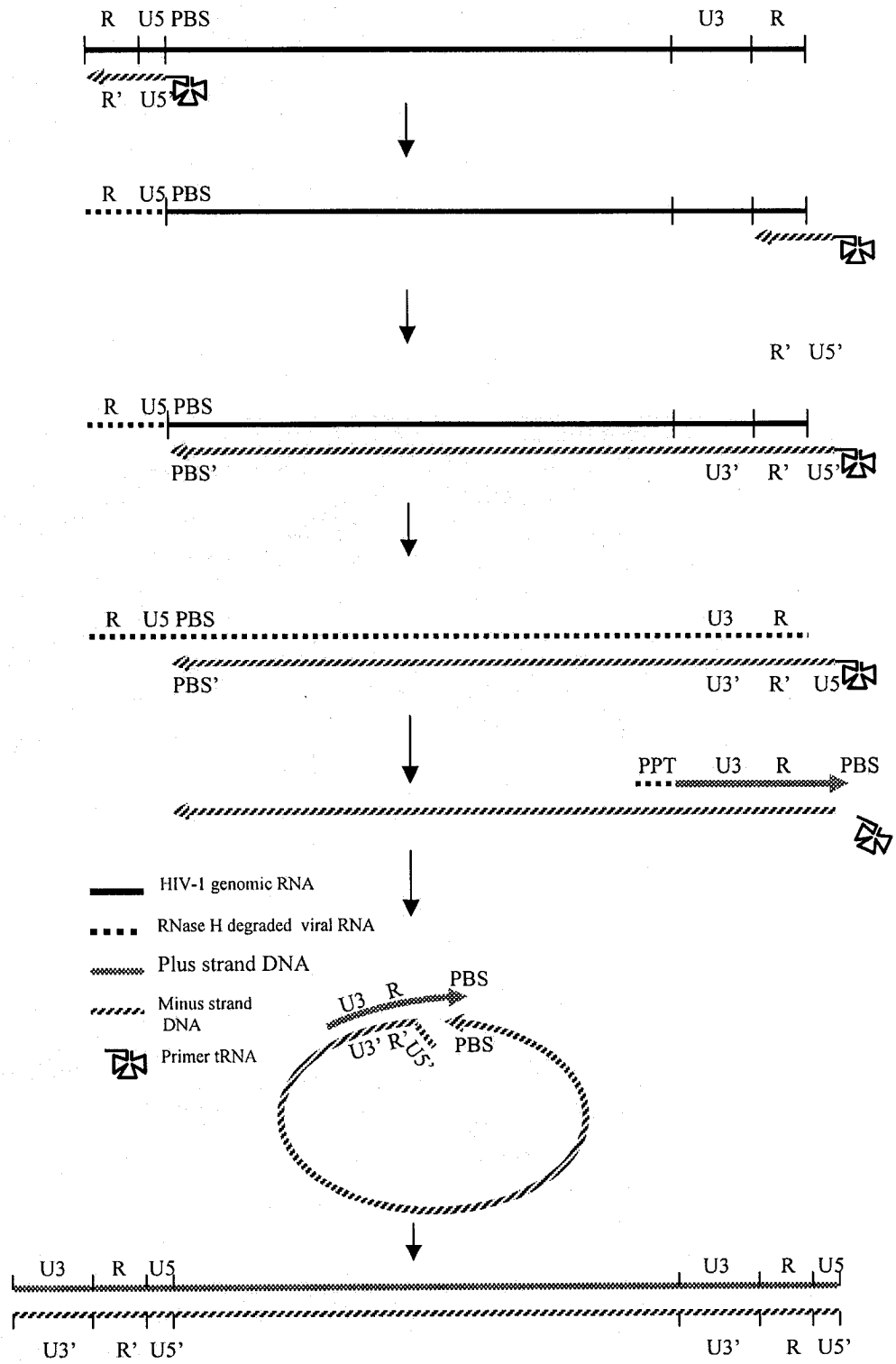


Fig. 6. Overview of the reverse transcription of HIV-1 genomic RNA (modified from ref. 160)

The process of HIV-1 reverse transcription can be divided into several steps (160) (Fig. 6):

1. The synthesis of minus-strand strong stop DNA (-sssDNA): The -sssDNA synthesis is initiated by the HIV-1 reverse transcriptase (RT), using a cellular $tRNA^{lys3}$ which is annealed at the primer binding site near the 5' end of genomic RNA, as a primer, and the viral genome as a template. It continues up to the 5' end of genomic RNA template.

2. The first strand transfer: The first strand transfer is facilitated by a R (repeat) sequence that appears at both 5' and 3' ends of the genomic RNA. The –sssDNA has already copied the R sequence at the 5' end of genome and therefore contains a 3' sequence that is complementary to both R sequences. Following degradation of the RNA strand of the RNA: DNA duplex, thanks to the RNase H activity of RT, the –sssDNA can base pair with the R sequences at the 3' end of genome RNA. This is called strand transfer, even though the "transfer" maybe from one end to the other of the same RNA strand.

3. Minus strand DNA synthesis: After the –sssDNA anneals to the R sequences at the 3' end of the genome, minus strand DNA synthesis resumes, still using RNA genome as template. Minus strand DNA synthesis is accompanied by the degradation of RNA template of the RNA: DNA duplex, though leaving a short polypurine tract (PPT) intact, which is relatively resistant to RNase H digestion.

4. Plus strand strong stop DNA synthesis (+sssDNA): The undegraded PPT, and some internal plus strand primers (174), also originated from genomic RNA, would be used as primers; the newly synthesized minus strand DNA would serve as a template. The synthesis of a DNA segment that is complementary to the minus strand DNA is initiated, and continues until part of the tRNA primer is copied. This short DNA copy is called +sssDNA. The RNA: DNA hybrid formed by part of $tRNA^{lys3}$ and part of +sssDNA would result in the removal of tRNA by RT's RNase H activity. Somehow the PPT primer gets removed as well.

5. Second strand transfer: The +sssDNA contains at its 3' end a PBS sequence complementary to an anti-PBS sequence located at the 3' end of the minus strand

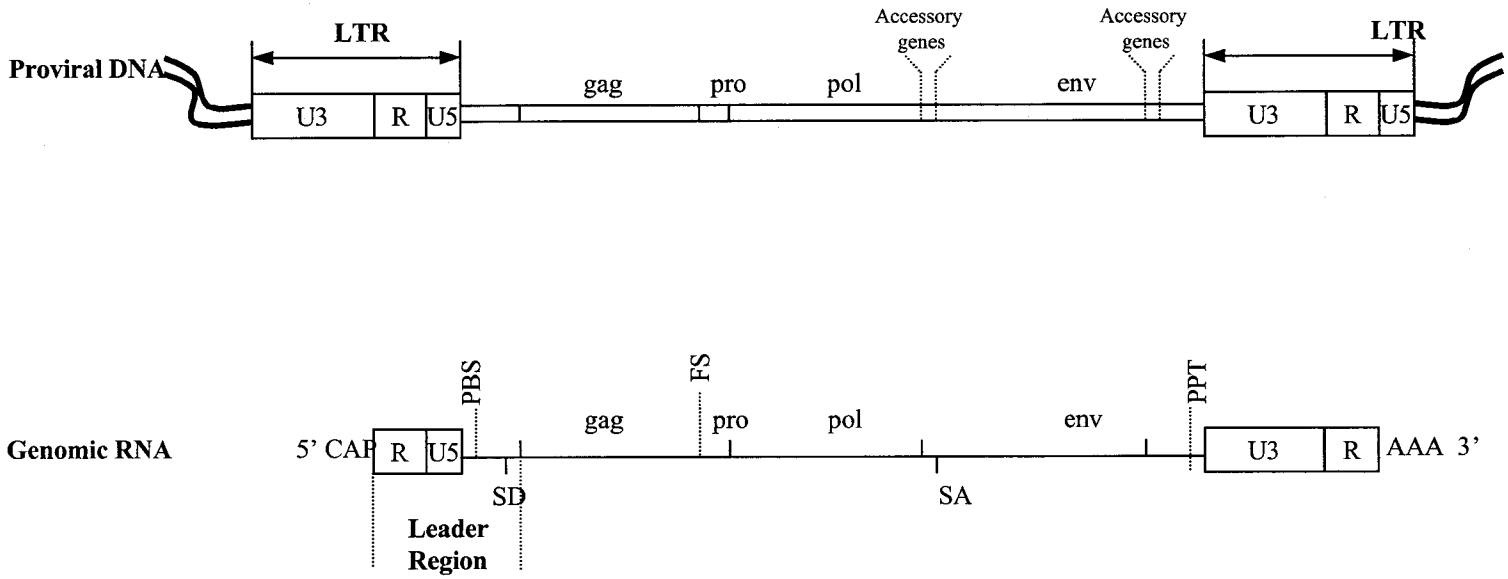


Fig. 7. HIV-1 LTR and leader region (adapted from ref. 168)

DNA. The second strand transfer represents base-pairing between the PBS sequence of the +sssDNA and the anti-PBS sequence of minus strand DNA.

6. Completion of proviral DNA synthesis: After the second strand transfer, elongation of minus strand and plus strand DNA would resume. When RT continues to synthesize the 3' end of minus strand, by copying the plus strand, displacement of the 5' end of minus strand from the 5' end of plus strand template must take place, presumably by RT and possibly along with NC. The complete double-stranded proviral DNA will be a blunt end linear DNA, which will be integrated into host genome by integrase (IN).

In addition to HIV-1 reverse transcriptase, $tRNA^{lys3}$ and genomic RNA, viral proteins, such as, Tat, Nef, Vif, Vpr, IN, MA, NCp7, and cellular proteins have been found to affect reverse transcription efficiency as well. The RNA stem structure, like TAR, the A-rich loop from adjacent to PBS has also been shown to affect reverse transcription (64).

1. 4. LTR

Proviral DNA is flanked by identical long terminal repeats (LTRs) at both ends (Fig. 7). These two LTRs are generated when reverse transcriptase jumps from one end of the template to the other. LTR contains U3, R, and U5 elements, with U3 representing the sequence unique to the 3' end of the RNA, R representing a sequence repeated at both ends of the RNA, and U5 representing the sequence unique to the 5' end of the RNA (168). HIV-1 genome contains ~9,200 nucleotides, with U3 having 450 nucleotides (nts), R 100 nts, and U5 80 nts.

HIV-1 5' LTR contains many *cis*-acting regulatory elements that regulate viral RNA transcription. U3 contains, from 5' to 3': modulator region, enhancer region and promoter region. The enhancer region contains two binding sites for nuclear factor-kB (NF-kB): they are at nt –105 to –96 and –91 to –82. The promoter region contains three binding sites (from nt –46 to –78) for Sp1 protein, and TATA box (nt -24 to –27). TATA box is essential for basal HIV-1 transcription. TATA box and Sp1 activity are also important for Tat transcription and support viral replication. At the boundary

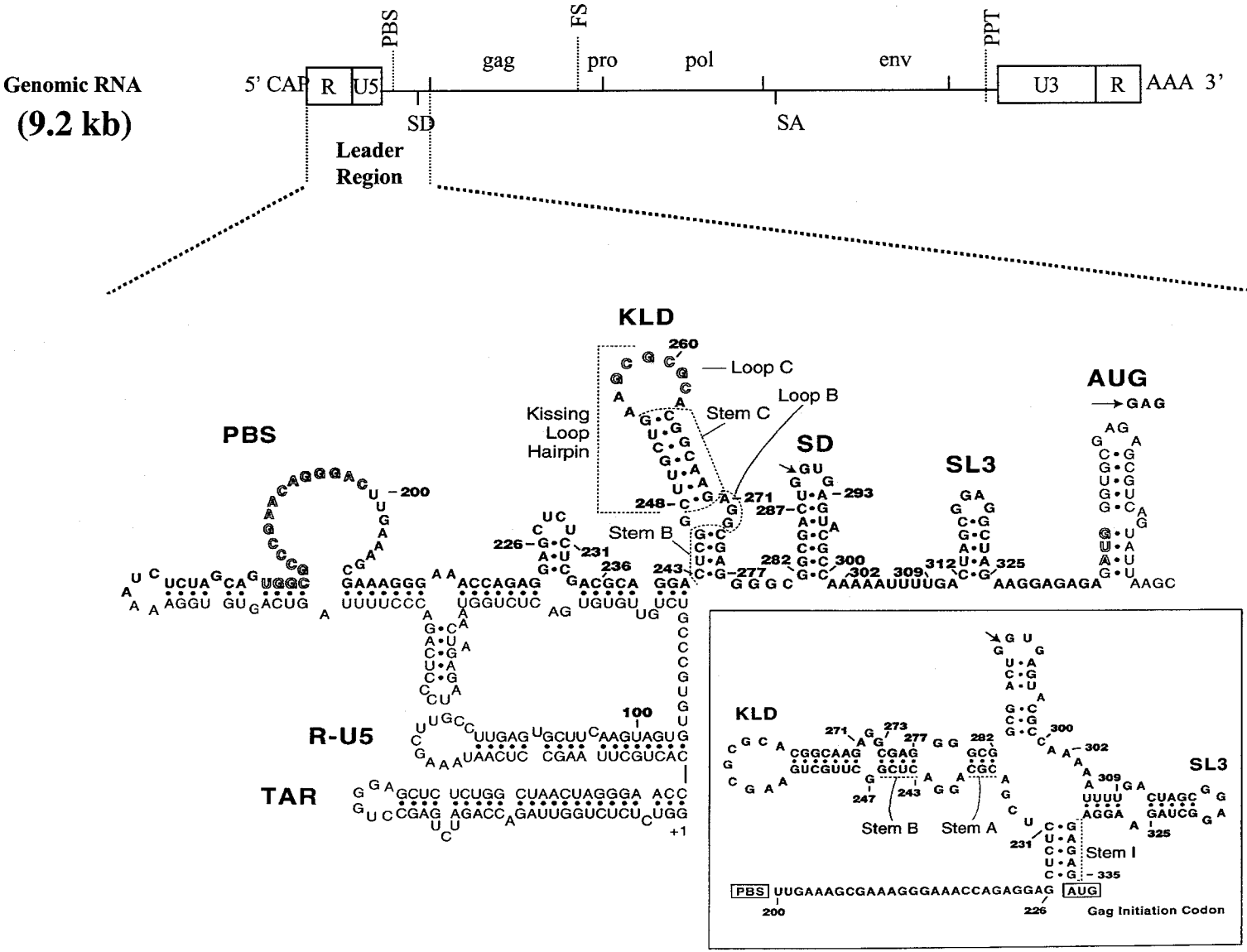


Fig. 8. Postulated stem-loop diagram of the leader region of HIV-1 genomic RNA (modified from ref. 148)

of U3 and R, there is transcription initiation site. Located immediately downstream is the *trans*-activating response element (TAR, from +1 to +80), which forms a stem-loop structure, and is a major control element for HIV-1 transactivation system (53).

The 3' LTR contain the *cis*-acting control elements critical for posttranscriptional processing of the 3' end of the HIV-1 RNA transcripts. There is functional poly (A) addition site in the R region of 3' LTR (139).

1. 5. Leader region

The 5' untranslated region of HIV-1 genomic RNA is called leader. The 5' part of the leader has the capacity to fold into several hairpin structures, namely, the TAR hairpin (7), which is the *trans*-activating region; the R-U5 hairpin (16, 33); and the PBS hairpin, which contains the primer binding site (Fig. 8). Besides its effect on virus replication mechanism (11), TAR has also been shown to affect RNA packaging and reverse transcription (12, 27, 28, 34, 65, 67). The 3' part of the leader region also has the potential to fold into several stem-loop structures: SL1 (KLD), SL2 (SD), SL3 and SL4. SL2 contains the major splicing donor site, SL3 contains an encapsidation signal, and SL4 contains the Gag start codon. SL1 plays several functional roles, such as encapsidation, dimerization, reverse transcription (131), recombination, etc. We will get into details in later chapters.

1. 6. Encapsidation

Encapsidation (packaging) is the incorporation of genomic RNA into the virus particle during viral assembly. It is highly specific since viral genomic RNA might only comprise 1% or less of the total mRNA in the infected cell, therefore viral genomic RNA has to be selectively packaged and spliced viral RNA and cellular mRNAs are largely excluded (100). Genomic RNA packaging requires the NC domain of the Gag polyprotein, as well as packaging signals located within the genomic RNA.

The primary packaging signal for HIV-1 is located between PBS and the gag gene (110, 111). It has the capacity to form four hairpin structures (or stem-loops)

(reviewed in 76): SL1, SL2, SL3, and SL4. SL1, SL3, and SL4 have been shown to contribute to genome RNA encapsidation, and base substitutions mutants in these hairpins showed relative encapsidation level around 39%, 31%, and 44% respectively, compared to that of wild type. SL3 has been suggested to be the major packaging signal motif. SL2, which spans the major splice donor site seemed to play a minor role in packaging (around 80%) (66, 103, 111). Deletion of SL1, SL3 and SL1+SL3 showed more profound effect on packaging (20%, 15%, 10% compared to that of wild type). The positions of the SL1 and SL3 stem-loops are not interchangeable, indicating the hairpin structure have to function within the right context. SL1, located upstream of the major splice donor site, probably together with SL3 and SL4, contributes to the selective pakaging of genomic versus subgenomic RNA (111). It has been suggested that SL1, SL3, and SL4 might form a higher-order structure to be recognized as packaging signal.

Although SL1, SL3, and SL4 are necessary, they are not sufficient for genome RNA packaging. Sequences upstream of SL1, such as TAR (67), r-U5, and U5-l (112) regions promote encapsidation. Deletions in these three regions showed relative encapsidation efficiency 26%, 11%, and 4%, respectively, versus control 93%. Sequences downstream of SL4 and within gag region and the RRE (Rev-responsive element) in env region (82) may promote encapsidation, probably contributing to the selective packaging of unspliced genomic versus spliced subgenomic RNA since these sequences are only present in the unspliced genomic RNA. As a whole, it has been indicated that the whole 5' end of genome RNA, from nucleotide +1 to 1017, are sufficient for HIV-1 encapsidation, with the presence of RRE to facilitate the RNA nuclear export (112).

Viral proteins are involved in this process. It has been suggested that the unprocessed Pr55Gag captures the genomic RNA (80, 144). NC domain of Gag has been shown to play important role in HIV genome encapsidation (49). NC contains two Zinc fingers and flanking basic residues as mentioned previously. Disrupting the Zinc binding activity of the first and and the second Zinc finger in the NC domain showed encapsidation efficiency 20% and 73%, respectively, as of wild type. Apparently the first Zinc finger has a more profound effect on encapsidation. The first

Zinc finger mutant also showed increased packaging of subgenomic RNA. Mutating the flanking basic residues generally reduced genome RNA encapsidation in vivo (40, 136, 137) and reduced the binding activity of NC to RNA in vitro (143). NMR studies of NCp7 protein bound with SL3 hairpin showed that the guanosines on SL3 loop can interact with NC (35). Although NC domain is necessary for RNA encapsidation, it does not contribute to selective packaging of genomic RNA, maybe suggested by its non-specific binding activity. It has been suggested that other regions in Gag might contribute to the selective packaging of genomic RNA. P2 region has been suggested to be one of the regions that contribute to optimal genome RNA encapsidation (81).

1.7. HIV-1 RNA dimerization

All retroviruses contain two identical single stranded positive RNAs that appear noncovalently associated near their 5' ends by a dimer linkage structure (DLS). Genomic RNA dimerization is of interest because it has been suggested to regulate several key steps in retroviral life cycle (reviewed in 130), such as: 1) encapsidation: the genome RNA encapsidation signal overlaps with DLS, and the two events often seem to be closely related. It is not completely clear yet as whether they can be dissociated. 2) genomic RNA translation: the highly structured DLS-encapsidation signal might inhibit the translation of HIV-1 genome and spare the RNAs for packaging. 3) recombination: Since long RNAs often have nicks, two molecules of RNA might provide the virus progeny with variability and viability (71); and the DLS would bring the two molecules closer to facilitate recombination. It was recently demonstrated that the kissing hairpin sequence within the DLS promotes recombination 4-fold more per unit RNA length than sequences from the pol region (3). 4) reverse transcription: the ordered structure of DLS might affect reverse transcription (131).

1. 7. 1. Evidence for the dimeric nature of the retroviral genome

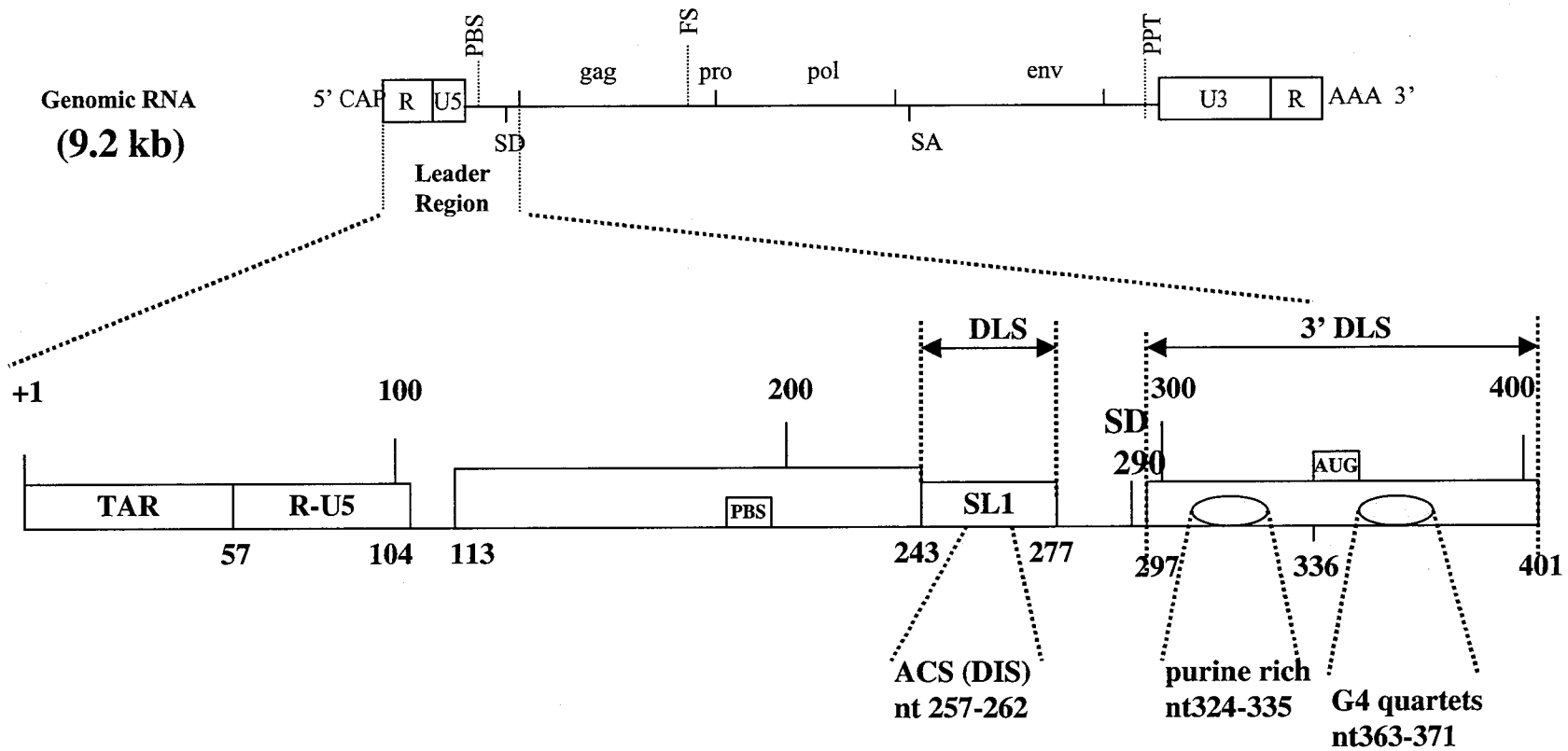


Fig. 9. HIV-1 5' end of genomic RNA, showing the consensus DLS and 3' DLS, numbers corresponding to that of the HIV-1 Lai genomic RNA (modified from ref. 130).

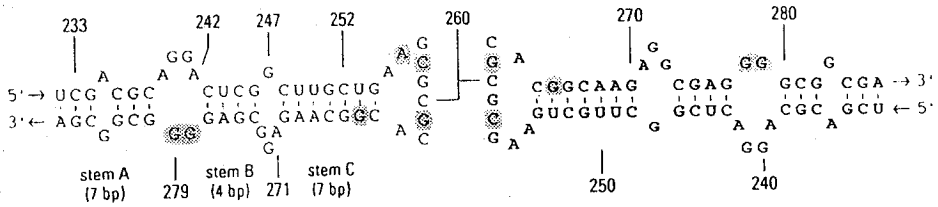
Evidence for retroviral genome being a dimer first came from sedimentation analysis (153) and electron microscopic studies (10, 126). The dimeric genome extracted from RSV (Rous Sarcoma Virus) under electron microscopy showed that the RNAs were associated near the 5' end of the genome, in a region called dimer linkage structure (DLS) (reviewed in 130). It has been suggested that the dimer linkage structure runs antiparallel while the general orientation of the two RNA molecules is parallel (9, 60, 89). It has been demonstrated that viral genomic RNA, including that of HIV-1, run on nondenaturing Northern gels, appeared as two RNA bands, corresponding to the size of a dimer and a monomer (reviewed in 60). The dimeric RNA can readily dissociate into monomeric RNA under mild conditions, suggesting the dimer linkage is noncovalent (9, 10, 89, 153).

HIV-1 is the retrovirus whose genome dimerization is the most studied, thus I will concentrate mainly on HIV-1 dimerization in the following description.

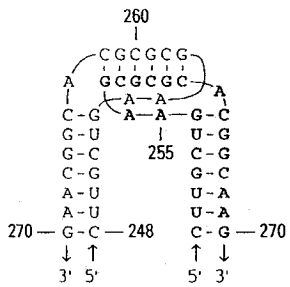
1.7.2. 3' DLS, RGGARA motif, G quartets

More detailed understanding of DLS started in the late 1980s when in vitro transcribed HIV-1 short RNAs representing the 5' end of the genome were shown to be able to dimerize (32). In 1990, it was identified that a 100 nt sequence located immediately downstream of the splice donor (SD) site (corresponding to nt297-401in HIV-1 Lai genome RNA), which was later called 3' dimer linkage structure (3' DLS), promoted in vitro dimerization in the presence of nucleocapsid precursor NCp15 (32). Soon after, it was suggested that in this 3' DLS region, a conserved purine-rich motif (RGGARA, R=purine) within nt 324 to 335 in HIV-1Lai RNA might be involved in *in vitro* HIV-1 RNA dimerization (107). However, *in vitro* RNA transcripts of the 5' HIV-2 leader sequence, lacking this motif were still able to form dimers (15).

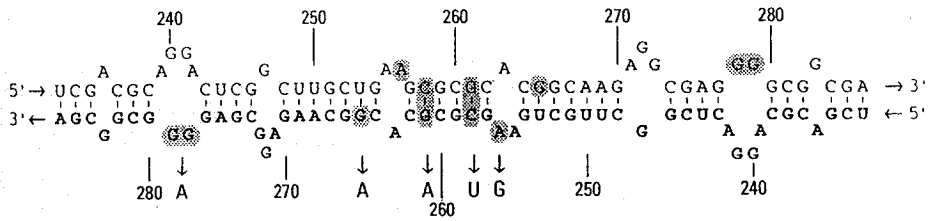
There are three purine tracts in the 3' DLS (corresponding to nt297-401 in HIV-1Lai RNA) identified by phylogenetic studies (reviewed in 130). In 1993, the second purine-rich motif (corresponding to nt363-371) in this region forming G quartets (G4) was also proposed to initiate *in vitro* dimerization (2, 155, 172). However, *in vitro* dimerization occurred in sequences that did not contain the G



(1) Monomer



(2) Loose dimer



(3) Tight dimer

Fig. 10. Loose and tight HIV-1 RNA dimers (M. Laughrea, 1996)

quartets (108, 124) or under ionic conditions that did not favor G4 formation (50), suggesting that these motifs are not essential for *in vitro* dimerization, although they might still have some effects on dimer stablization or maturation (60) (Fig. 9).

1.7. 3. DLS : SL1 (or KLD), DIS (or kissing-loop) hairpin

Further studies opposed the *in vitro* HIV-1 RNA dimerization model, which involves the formation of purine tetrads (50, 61, 90, 124, 150). In 1994, by using *in vitro* transcribed short HIV-1 RNA transcripts, it was identified that sequence upstream of the splice donor site is essential for in vitro HIV-1 RNA dimerization (90, 150). The sequence has the capacity to form SL1 [or kissing loop domain (KLD), nt243-277 in HIV-1Lai genomic RNA]. SL1 is in the form of a kissing-loop hairpin, which contains loop C and stem C, seated on top of a short, conserved, but putative structure: stem-loop B (95). Loop C contains an autocomplementary sequence (ACS), also called palindrome or DIS: GCGCGC in HIV-1Lai (subtype B), and GUGCAC in HIV-1Mal (subtype A).

It has been proposed that in vitro dimerization is initiated through the Watson-Crick interaction in a kissing-loop manner between the autocomplementary sequence of GCGCGC (90) or GUGCAC (150) from one RNA molecule with that of the other (29, 61, 90, 91, 124, 129, 150).

1.7.4. Loose (kissing-loop) dimer and tight (extended-duplex) dimer

In 1996, it was demonstrated that there are two kinds of HIV-1 RNA dimers in vitro. The spontaneously formed dimer under physiological temperature is called loose (or kissing-loop) dimer, which involves the base pairing between the ACS of one monomer with that of the other. Most of the intramolecular base pairing forming the monomer stem will remain intact. After heat treatment or after incubation with viral NCp7, the loose dimer can turn into tight dimer. The intramolecular base pairing of the stem will melt, and intermolecular base pairing forms between two monomers, resulting in the formation of extended duplex (45, 46, 91, 92, 122, 123) (Fig. 10).

1.7.5. Mutagenesis, NMR structure, chemical enzymatic probing, crystallography, and *in vitro* selection studies of dimerization

Studies by various methods, including site directed mutagenesis (91-95), chemical and enzymatic probing (7), molecular modeling (78, 128), NMR (nuclear magnetic resonance) studies (31, 56, 118, 119, 157), and X-ray crystallography (43) etc., have been used to construct a three dimensional model of the kissing-loop complex.

In 1998, the solution structure of the initial kissing-loop dimer formed by 23 nt RNA corresponding to truncated HIV-1Lai SL1, including the palindrome and two distal base pairs (13 nts out of 23 representing nt253-265 in HIV-1Lai genome RNA) was characterized by 2D NMR spectroscopy. The palindrome GCGCGC from one monomer will form Watson/Crick base pairs with that of the other and there are interstrand stackings among the adenine bases flanking the palindrome, which might cause some mechanical strains that can potentially melt the stems of both monomers during the future tight dimer formation process (118). The NMR study of kissing-loop dimer formed between two 19 nt RNAs representing the palindrome GUGCAC and part of the stem C from HIV-1Mal SL1, 17 nts out of 19 corresponding to nt251-267 in HIV-1Lai genome RNA, also gave similar results (31).

In 1999, NMR study of the above-mentioned 23 mer forming the mature dimer showed that the dimer is in the form of an extended duplex with two adenine bulges. The adenine bulge consists of three unpaired adenines flanking the palindrome (119). The crystal structure of mature dimer formed by 23 nt RNA corresponding to the whole kissing loop hairpin of HIV-1Mal, nt248-270 according to HIV-1Lai genome RNA, also showed that the dimer is an extended duplex with two bulges, each of which consists of a bulged-out adenine and a G-A mismatch (7). These studies help us understand the mechanism of the transition from a kissing-loop dimer to a more thermodynamic stable tight dimer. The energy resided in the mechanical distortion of the kissing-loop dimer might cause the melting of the stem in the monomer and refolding and formation of extended dimer (119).

The transition from a kissing-loop dimer to the extended duplex formed by the 23 nt RNAs representing nt 248 to 270 in HIV-1Lai SL1 was followed by UV and 1D NMR spectroscopy (163). It has been suggested that during this transition, the loop-loop interaction will remain intact, while the intramolecular base pairing of the stems within the monomer will rupture, and new intermolecular base pairing of the stems between the two monomers will form, resulting in the extended duplex (163).

In vitro selection study of dimerization competent sequences by randomizing the loop C sequences of HIV-1 subtype A and subtype B SL1 has been performed to understand the conservative nature of loop C (101). Further studies have identified the sequence and structure constraints for loop C, such as: the optimal number of nucleotides for loop C, the optimal size for the ACS, and the conformation of the purines flanking the ACS, etc. (102). It has been suggested that besides the importance of the ACS in dimerization, the flanking purines are also important for dimerization (102, 128).

1.7.6. *In vivo* study

In vitro data suggests that SL1 plays a key role in HIV dimerization. Questions were asked as whether SL1 plays a role in *in vivo* genome dimerization. The *in vivo* role of SL1 in HIV-1 genome dimerization is not as well studied as its *in vitro* role. There is also little known on the role of SL1 in proviral DNA synthesis. This dissertation is concerned with the role of SL1 in HIV-1 genomic RNA dimerization, encapsidation, and proviral DNA synthesis. In this thesis, we will provide a detailed *in vivo* analysis of the role of different parts of SL1 in HIV-1 genome dimerization, encapsidation, and proviral DNA synthesis. By showing that a point mutation in the palindrome of loop C would inhibit genome dimerization, we clearly show that loop C is of vital importance to HIV-1 genome dimerization. We also showed that stem C, stem B and loop B are as important for HIV-1 genome dimerization as the palindrome (146). Our results also show that SL1 affects HIV-1 viral replication, encapsidation and proviral DNA synthesis, indicating that SL1 is a multifunctional structure. While the existence of stem C and loop C is supported by phylogenetic evidence (13, 94), the

existence of stem B and loop B was putative. In this thesis, we present data clearly favoring the notion that stem B and loop B exist (148).

While mutating SL1 could abolish dimerization of *in vitro* transcribed RNAs (90, 124), *in vivo*, the destruction of SL1 resulted in a ~50% reduction in dimerization yield (146). It is possible that the tertiary structure of the whole HIV-1 genome influences SL1 function; or there could be some other DLS structures unidentified *in vivo*.

1.7.7. Proteins and proteolytic involvement

Viral and cellular proteins might be involved in HIV-1 genome dimerization. NCp7 has been demonstrated, paradoxically, to have strong nucleic acid annealing and nucleic acid unwinding ability (84, 166). Maybe due to these two activities, NCp7 has been shown to promote *in vitro* dimerization (45, 122), converting the unstable kissing-loop complex into a stable dimer.

A mutant with a defective MA nuclear export signal has been shown to package monomeric genomic RNA, suggesting that the MA region might also play some indirect role in HIV-1 genome dimerization (41).

The basic regions flanking the N-terminal zinc finger of HIV-1 NCp7 are required for dimer maturation of *in vitro* synthesized RNAs from HIV-1 (158) and Harvey sarcoma virus (HaSV) (45). The role of the zinc fingers of NCp7 in dimerization is very controversial. It was reported that residues within the NCp7 N-terminal zinc finger, not the zinc finger structure itself, were required for *in vitro* HaSV viral RNA dimer maturation (45). While the two zinc fingers appear dispensable for *in vitro* HIV-1 RNA dimer maturation (158), *in vivo*, the distal Zinc finger of NC domain plays a role in HIV-1 genomic RNA dimerization (96).

After long term culture, mutations within the matrix, capsid, p2, and nucleocapsid genes suppress several functional defects caused by SL1 destruction, but it is unknown if they can restore infectivity or genomic RNA dimerization to wild type level. We will demonstrate, in this thesis, that protein mutations can restore a genomic

RNA dimerization defect caused by a G/C-poor DIS, but not when DIS is denatured (147). Our results suggested protein involvement in genomic RNA dimerization.

Proteolytic processing of HIV-1 Gag and Gag-Pol polyproteins has been shown to be important for dimer formation since protease negative mutant contains lower percent of dimeric genomic RNA, which will be shown in this thesis (146) or contains labile dimers (50). It has been demonstrated that mutations in the primary cleavage site (p2/NC) result in the formation of unstable dimer (145), indicating the importance of cleavage of the p2/NC site, especially within Gag precursor, in HIV-1 genomic RNA dimer maturation. The ratio of Gag/Gag-Pol precursors has been shown to be critical for HIV-1 genome dimerization (145).

Recently, Sakuragi et al. showed that by duplicating an ~1000 nt-long RNA sequence of the 5' HIV-1 genome, including the DLS region, he could create mutants that contain increased amount of monomers, without affecting encapsidation, and virion formation (142), claiming dimerization and encapsidation could be dissociated.

1.7.8. Other regions might be involved

The DLS of full-length HIV-1 genomic RNA extracted from the HIV-1 virion has been visualized by electron microscopy (69), and the DLS has been shown to contain a 323± 44 nt loop. It was suggested by electron microscopy data and computer modeling that there are two contact points between the two RNA molecules, one is at the R-U5 region and the other is SL1 region. However, disrupting the palindrome of R-U5 hairpin did not show much impact on HIV-1 genome dimerization, inconsistent with the *in vitro* data (96).

In this thesis, we will show that regions other than SL1 might be involved in the dimerization process and that the leader region might form a compact structure (148). Running different length of *in vitro* transcribed RNAs encompassing the HIV-1 Lai leader region, in nondenaturing polyacrylamide gel, it has been shown that there were unusually fast migrating conformers, especially the RNAs containing nt 1-290 of genomic RNA. It has been suggested that the leader region RNA might form a compact structure and there might be long-range interaction between nt 1-105 and

nt245-290 (74). It has been proposed that the HIV-1 leader RNA exists in two different conformations: one is called LDI (long distance interaction), in which leader RNA forms a more stable compact conformer with long distance interaction between the loop and upper stem of poly A hairpin and with that of SL1 hairpin; another is called BMH (branched multiple hairpin), in which leader RNA locally forms individual hairpins. In the first conformation, the SL1 is trapped and dimerization signal is not exposed. In order to form dimer, the LDI conformation will have to undergo structural rearrangement into the BMH conformation so that dimerization signal can be exposed and dimerization can take place. NCp7, which has nucleic acid chaperone activity, can facilitate *in vitro* dimerization, probably by stablizing the BMH conformation (74).

Mutations in HIV-1 TAR hairpin have been demonstrated to disrupt the LDI conformation and favor the BMH formation, thus facilitate *in vitro* dimerization (75).

1.7.9. Dissociation of genomic RNA dimerization and encapsidation

Dimerization and encapsidation are likely to be closely related events, but whether dimerization is required for encapsidation is not yet resolved. However, accumulating data suggest that dimerization and encapsidation could be dissociated. In this thesis, we will show that dimerization is not a prerequisite for encapsidation (146). There are other studies supporting this notion (142, 145). It has been shown that in Rous sarcoma virus (RSV), the newly sheded virions encapsidated monomeric RSV genome, and dimerize over time. It has been shown that RSV protease defective mutant also packaged monomers (97). These suggest that dimerization is not a requirement for packaging, at least in RSV.

1.7.10. Conclusion

Dimerization is an important event for viral life cycle. It is related to encapsidation, reverse transcription, recombination, translation, etc.. It might be an important target for antiviral drugs.

1. 8. References

1. **Ahmad, N., and S. Venkatesan.** 1998. Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. Science **241:**1481-1485.
2. **Awang G, and D. Sen.** 1993. Mode of dimerization of HIV-1 genomic RNA. Biochemistry **32:**11453-11457.
3. **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:**36482-36492.
4. **Baltimore, D.** 1970. Viral RN-dependent DNA Polymerase in virions of RSV. Nature **226:**209-211.
5. **Bardy, M., B. Gay, S. Pebernard, N. Chazal, M. Courcoul, R. Vigne, E. Decroly, and P. Boulanger.** 2001. Interaction of human immuno-deficiency virus type 1 Vif with Gag and Gag-Pol precursors: co-encapsidation and interference with viral protease-mediated Gag processing. J. Gen. Virol. **82**:2719-2733.
6. **Barre'-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeytre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blue, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immuno-deficiency syndrome (AIDS). Science **220:**868-871.
7. **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:**382-397.
8. **Beemon, K, P. Duesberg, and P. Vogt.** 1974. Evidence for crossing-over between avian tumor viruses based on analysis of viral RNAs. Proc. Natl. Acad. Sci. USA **71:**4254-4258.
9. **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.
10. **Bender, W., Y. H. Chien, S. Chattopadkyay, 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:**888-896.

11. **Berkhout, B.** 1996. Structure and function of the human immunodeficiency virus leader RNA. Progress in Nucleic Acid Research and Molecular Biology **54:**1-34.
12. **Berkhout, B.** 2000. Multiple biological roles associated with the repeat region of the HIV-1 RNA genome. Adv. Pharmacol. **48:**29-73.
13. **Berkhout, B., and J. L. B. van Wamel.** 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. J. Virol. **70:**6723-6732.
14. **Berkhout, B., and J. L. B. van Wamel.** 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. RNA **6:**282-295.
15. **Berkhout, B., B. B. Oude Essink, and I. Schoneveld.** 1993. *In vitro* dimerisation of HIV-2 leader RNA in the absence of PuGGAPuA motifs. FASEB **7:**181-187.
16. **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:**276-281.
17. **Blattner, W. A., T. R. O'Brien, and N. E. Mueller.** 1997. Retroviruses-human immunodeficiency virus. In *Viral Infections of Humans: Epidemiology and Control* 4th Edition (ed. A. S. Evans & R. A. Kaslow), pp. 713-783. Plenum Medical Book Company, New York.
18. **Bouyac, M., M. Courcoul, G. Bertoia, Y. Baudat, D. Gabuzda, D. Blanc, N.chanzal, P. Boulanger, J. Sire, R. Vigne, and B. Spire.** 1997. Human immunodeficiency virus type 1 Vif protein binds to the Pr55Gag precursor. J. Virol. **71:**9358-9365.
19. **Braaten, D., C. Aberham, E. K. Franke, L. Yin, W. Phares, and J. Luban.** 1996. Cyclosporine A-resistant human immunodeficiency virus type 1 mutants demonstrate that gag encodes functional target of cyclophilin A. J. Virol. **70:**5170-5176.

20. **Brown, P. O.** 1997. Integration. p161-204. *In* J. M. Coffin, S. H. Hugbes, and H. E. Varmus (ed.), Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

21. **Bukrinsky, M., and A. Adzhubei.** 1999. Viral protein R of HIV-1. Rev. Med. Virol. **9:**39-49.

22. **Bukrinsky, M. I., N. Sharova, T. L. McDonald, T. Pushkarskaya, and W. G. Tarpley.** 1993. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc. Natl.Acad. Sci. USA **90:**6125-6129.

23. **Bukrinsky, M. I., S. Haggerty, M. P. Dempsey, N. Sharova, A. Adzhubei, L. Spitz, P. Lewis, D. Gldfarb, M. Emerman, and M. Stevenson.** 1993. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature **365:**666-669.

24. **Canaani E, K. V. D. Helm, and P. Duesberg.** 1973. Evidence for 30-40S RNA as precursor of the 60-70 S RNA of Rous sarcoma virus. Proc. Natl. Acad. Sci. USA **72:**401-405.

25. **Chan, D. C., and P. S. Kim**. 1998. HIV entry and its inhibition. Cell **93:**681-684.

26. **Clapham, P. R., and A. Mcknight.** 2001. HIV-1 receptors and cell tropism: British Medical Bulletin **58:**43-59.

27. **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:**3407-3414.

28. **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:**101-109.

29. **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:**5902-5908.

30. **Coffin, J. M.** 1992. Structure and classification of retroviruses. p.19-49. In J. A. Levy (ed.), The retroviridae, Plenum Press, NY.

31. **Dardel, F., R. Marquet, C. Ehresmann, B. Ehresmann, and S. Blanquet.** 1998. Solution studies of the dimerization initiation site of HIV-1 genomic RNA. Nucleic Acids Res. **26:**3567-3571.

32. **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 HIV-1. J Mol Biol **216:**689-699.

33. **Das A. T., B. Klaver, B. I. F. Klasens, J. L. B. 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:**2346-2356.

34. **Das, A. T., B. Klaver, and B. Berkhout.** 1998. The 5' and 3' TAR elements of the human immunodeficiency virus exert effects at several points in the virus life cycle. J. Virol. **72:**9217-9223.

35. **De Guzman, R. N., Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, and M. F. Summers.** 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science **279**:384-388.

36. **Demirov, D. G., J. M. Orenstein, and E. O. Freed,** 2002. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. J. Virol. **76**:105-117.

37. **Deora, A., P. Spearman, and L. Ratner.** 2000. The N-terminal Matrix domain of HIV-1 Gag is sufficient but not necessary for viral protein U-mediated enhancement of particle release through a membrane-targeting mechanism. Virology **269:**305-312.

38. **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:**8938-8945.

39. **Dickson, C. R. Eisenman, H. Fan, E. Hunter, and N. Teich.** 1984. Protein biosynthesis and assembly, p513-648. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA Tumor Viruses, Vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

40. **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 require only peptide domains outside the zinc fingers. Proc. Natl. Acad. Sci. USA **89:**6472-6476.

41. **Dupont, S., N. Sharova, C. Dehoratius, C. –M. A. Virbasius, X. Zhu, A. G. Bukrinskaya, M. Stevenson, and M. R. Green.** 1999. A novel nuclear export activity in HIV-1 matrix protein required for viral replication. Nature **402:**681-685.

42. **Eckert D. M., and P. S. Kim.** 2001: Mechanisms of viral membrane fusion and its inhibition. Annu. Rev. Biochem. **70:**777-810.

43. **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:**1439-1449.

44. **Fackler, O. T., and A. S. Baur.** 2002. Live and let die: Nef functions beyond HIV replication. Immunity **16:**493-497.

45. **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. Aca. Sci. USA **93:**7577-7581.

46. **Feng, Y. X., W. Fu, A. J. Winter, J. G. Levin, and A. Rein.** 1995. Mutiple regions of Harvey sarcoma virus RNA can dimerize *in vitro*. J. Virol. **69:**2486-2490.

47. **Franke, E. K., H. E. H. Yuan, and J. Luban.** 1994. Specific incorporation of cyclophilin A into HIV-1 virions. Nature **372:**359-362.

48. **Frankel, A. D., and J. A. T. Young.** 1998. HIV-1: Fifteen proteins and an RNA. Annu. Rev. Biochem. **67:**1-25.

49. **Freed, E. O.** 1998. HIV-1 Gag proteins: diverse function in the virus life cycle. Virology **251:**1-15.

50. **Fu, W., R. J. Gorelick, and A, Rein.** 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. J. Virol. **68:**5013-5018.

51. **Gallay, P., T. Hope, D. Chin, and D. Trono.** 1997. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. Proc. Natl Acad. Sci. USA **94:**9825-9830.

52. **Garrus, J. E., U. K. Von Schwedler, O. W. Pomillos, S. G. Morham, K. H. Zavitz, H. E. Wang, D. A. Wettstein, K. M. Stray, M. Cote, and 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

53. **Gaynor, R. B.** 1995. Regulation of HIV-1 gene expression by the transactivator protein Tat. Curr. Top. Microbiol. Immunol. **193:**51-77

54. **Geijtenbeek, T. B. H. R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, G. J. Adema, Y. van Kooyk, and C. G. Figdor.** 2000. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell **100:**575-585.

55. **Gelderblom, H.** 1990. Morphogenesis, maturation and fine structure of lentiviruses. p159-180. *In* Retroviral proteases: maturation and morphogenesis. L. H. Pearl (ed.), Stockton press, NY.

56. **Girard, F., F. Barbault, C. Gouyette, T. Huynh-Dinh, J. Paoletti, and G. Lancelot.** 1999. Dimer initiation sequence of HIV-1 Lai genomic RNA: NMR solution structure of the extended duplex. J. Biomol. Struct. Dyn. **16:**1145-1157.

57. **Goncalves, J., B. Shi, X. Yang, and D. Gabuzda**. 1995. Biological activity of human immunodeficiency virus type 1 Vif requires membrane targeting by C-terminal basic domains. J. Virol. **69:**7196-7204.

58. **Goncalves, J., P. Jallepalli, and D. H. Gabuzda.** 1994. Subcellular localization of the Vif protein of human immunodeficiency virus type 1. J. Virol. **68:**704-712.

59. **Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. –T. Fan, R. A. Wolf, and A. Saxon.** 1981. Pneumocystis carinii pneumonia and mucosal

candidiasis in previously healthy homosexual men. New Engl. J. Med. **305:**1425-1431.

60. **Greatorex, J., and A. Lever.** 1998. Retroviral RNA dimer linkage. J. Gen. Virol. **79:**2877-2882.

61. **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.

62. **Haffar, O. K., S. Popov, L. Dubrovsky, I. Agostini, H. Tang, T. Pushkarsky, S. G. Nadler, and M. Bukrinsky.** 2000. Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. J. Mol. Biol. **299:**350-368.

63. **Hansen, M. S., S. Carteau, C. Hoffmann, L. Li, and F. Bushman.** 1998. Retroviral cDNA integration: mechanism, applications and inhibition. Genet. Eng. **20:**41-61.

64. **Harrich, D., and B. Hooker.** 2002. Mechanistic aspects of HIV-1 reverse transcription initiation. Reviews in Medical Virology **12:**31-45.

65. **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:**5639-5646.

66. **Harrison, G. P., G. Miele, E. Hunter, and A. M. L. Lever.** 1998. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. J. Virol. **72:**5886-5896.

67. **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:**4127-4135.

68. **Hill, C. P., D. Worthylake, D. P. Bancroft, A. M. Christensen, and W. Sundquist.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1matrix protein: implications for membrane association and assembly. Proc. Natl. Acad. Sci. USA **93:**3099-3104.

69. **Hoglund, S., A. Ohagen, J. Goncalves, A. T. Panganiban and D. Gabuzda.** 1997. Ultrastructure of HIV-1 Genomic RNA. Virology **233:**271-279.

70. **Holmes, E. C.** 2001. On the origin and evolution of the human immunodeficiency virus (HIV). Biol. Rev. **76:**239-254

71. **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. USA **87:**1556-1560.

72. **Hull, R.** 2001. Classification reverse transcribing elements; a proposal and a challenge to the ICTV. Arch. Virol. **146:**2255-2261.

73. **Hunter, E.** 1997. Viral entry and receptors. *In* J. M. Coffin, S. H. Hugbes, and H. E. Varmus (ed.), Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

74. **Huthoff H., and B. Berkhout.** 2001: Two alternating structures of the HIV-1 leader RNA. RNA **7:**143-157.

75. **Huthoff, H., and B. Berkhout.** 2001. Mutations in the TAR hairpin affect the equilibrium between alternative conformations of the HIV-1 leader RNA. Nucleic Acids Res. **29:**2594-2600.

76. **Jewell, N. A., and L. M. Mansky.** 2000. In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. J. Gen. Virol. **81:**1889-1899.

77. **Johnson, P. E., R. B. Turner, Z. R. Wu, L. Hairston, J. Guo, J. G. Levin, and M. F. Summers.** 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. Biochemistry **39:**9084-9091.

78. **Jossinet, F., J. –C. Paillart, E. Westhof, T. Hermann, E. Skripkin, J. S. Lodmell, C. Ehresmann, B. Ehresmann, and R. Marguet.** 1999. Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. RNA **5:**1222-1234.

79. **Kanki, P. J., K. U. Travers, S. M'boup, C. C. Hsieh, R. G. Marlink, A. Gueye-Ndlaye, T. Siby, I. Thior, M. Hernandez-Avila, J. L. Sankale', I. Ndoye, and M. Essex.** 1994. Slower heterosexual spread of HIV-2 than HIV-1. Lancet **343:**943-946.

80. **Kaye J. F., and A. M. Lever.** 1996. Trans-acting proteins involved in RNA encapsidation and viral assembly in human immunodeficiency virus type 1. J. Virol. **70:**880-886.

81. **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.

82. **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:**6588-6592.

83. **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:**7252-7265.

84. **Khan, R., and D. P. Giedroc.** 1992. Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds rRNA. J. Biol. Chem. **267:**6689-6695.

85. **Kim, S. Y., R. Byrn, J. Groopman, and D. Baltimore.** 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: Evidence for differential gene expression. J. Virol. **63:**3708-3713.

86. **Kishi, M., Y. Nishino, M. Sumiya, K. Ohki, T. Kimura, T. Goto, M. Nakai, M. Kakinuma, and K. Ikuta.** 1992. Cells surviving infection by human immunodeficiency virus type 1: vif or vpu mutants produce non-infectious or markedly less cytopathic viruses. J. Gen. Virol. **73:**77-87.

87. **Kondo, E., F. Mama no, E. A. Cohen, and H. G. Gottlinger.** 1995. The p6 gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. J. Virol. **69:**2759-2764.

88. **Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya.** 2000. Timing the ancestor of the HIV-1 pandemic strains. Science **288:**1789-1796.

89. **Kung, H. J, S. Hu, W. Bender, J. M. Bailey, N. Davidson, M. O. Nicolson, and R. M. McAllister.** 1976. RD-114, baboon, and woolly monkey viral RNAs compared in size and structure. Cell **7:**609-620.

90. **Laughrea, M., and L. Jette.** 1994. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus type 1 genomic RNA. Biochemistry **33:**13464-13474.

91. **Laughrea, M., and L. Jette.** 1996a. 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:**1589-1598.

92. **Laughrea, M., and L. Jette.** 1996b. HIV-1 genomic dimerization: Formation kinetics and thermal stability of dimeric HIV-1 Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain. Biochemistry **35:**9366-9374.

93. **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:**9501-9508.

94. **Laughrea, M., L. Jette, J. Mak, L. Kleiman, C. Liang, and M. A. Wainberg.** 1997. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. J. Virol. **71:**3397-3406.

95. **Laughrea, M., N. Shen, L. Jetté, 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:**226-234.

96. **Laughrea, M., N. Shen, L. Jetté, J. -L. Darlix, L. Kleiman, and M. 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:**109-116.

97. **Lear, A. L., M. Haddrick, and S. Heaphy.** 1995. A study of the dimerization of the Rous sarcoma virus RNA in vitro and *in vivo*. Virol. **212:**47-57.

98. **Liang, C., L. Rong, M. Laughrea, L. Kleiman, 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.

99. **Liao, Z., J. W. Roos, and J. E. K. Hildreth.** 2000. Increased infectivity of HIV type 1 particles bound to cell surface and solid-phase ICAM-1 and VCAM-1 through acquired adhesion molecules LFA-1 and VLA-4. AIDS Res. Hum. Retroviruses **16:**355-366.

100. **Linial, M. L., and A. D. Miller.** 1990. Retroviral RNA packaging: sequence requirements and implications. Curr. Top. Microbiol. Immunol. **157:**125-152.

101. **Lodmell, J. S., C. Ehresmann, B. Ehresmann, and R. Marquet.** 2000. Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. RNA **6:**1267-1276.

102. **Lodmell, J. S., C. Ehresmann, B. Ehresmann, and R. Marquet.** 2001. Structure and dimerization of HIV-1 kissing loop aptamers. J. Mol. Biol. **311:**475-490.

103. **Luban, J. , and S. P. Goff.** 1994. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. J. Virol. **68:**3784-3793.

104. **Mak, J.** 1996. Mechanism of selective incorporation and genomic placement of primer tRNAlys3 into human immunodeficiency virus type 1. Thesis.

105. **Marcello, A., M. Zoppe, and M. Giacca.** 2001. Multiple modes of transcriptional regulation by the HIV-1 Tat transactivator. Iubmb Life. **51:**175-181.

106. **Marquet, R., C. Isel, C. Ehresmann, and B. Ehresmann.** 1995. tRNA as primer of reverse transcriptase. Biochimie **77:**113-124.

107. **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:**2349-2357.

108. **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:**145-151.

109. **Masur, H., M. A. Michelis, J. B. Greene, I. Onorato, R. A. Vandestouwe, R. S. Holzman, G. Wormser, L. Brettman, M. Lange, H. W. Murray, and S. Cunningham-Rundles.** 1981. An outbreak of community-acquired pneumocystis carinii pneumonia. New Engl. J. Med. **305:**1431-1438.

110. **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:**2963-2973.

111. **McBride, M. S., and A. T. Panganiban.** 1997. Position dependence of functional hairpins for HIV-1 RNA encapsidation in vivo. J. Virol. **71:**2050-2058.

112. **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:**4544-4554.

113. **Mondor, I., S. Ugolini, and Q. J. Sattentau.** 1998. Human immunodeficiency virus type 1 attachment to HeLa CD4 cells is CD4-independent and gp120 dependent and requires cell surface heparans. J. Virol. **72:**3623-3634.

114. **Moore J. P., and M. Stevenson.** 2000. New targets for inhibitors of HIV-1 replication: Nature Rev.: Mol. Cell Biol. **1:**40-49.

115. **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:**3059-3065.

116. **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.

117. **Moulard, M., et al.**. 2000. Selective interactions of polyanions with basic surfaces on human immunodeficiency virus type 1 gp120. J. Virol. **74:**1948-1960.

118. **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. Nature Struct. Biol. **5:**432-436.

119. **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:**387-392.

120. **Muller, B., T. Patschinsky, and H. –G. Krausslich.** 2002. The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. J. Virol. **76:**1015-1024.

121. **Murakami, T., and E. O. Freed.** 2000. Genetic evidence for an interaction between human immunodeficiency virus type 1 matrix and α-helix 2 of the gp41 cytoplasmic tail. J. Virol. **74:**3548-3554.

122. **Muriaux, D., H. de Roquigny, B. –R. Roques, and J. Paoletti.** 1996. NCp7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. J. Biol. Chem. **271:**33686-33692.

123. **Muriaux, D., P. Fosse, and J. Paoletti.** 1996. A kissing complex together with a stable dimer is involved in the HIV-1 Lai RNA dimerization process *in vitro*. Biochemistry **35:**5075-5082.

124. **Muriaux, D., P. M. Girard, B. Bonnet-Mathoniere, and J. Paoletti.** 1995. Dimerization of HIV-1 Lai RNA at low ionic strength. *An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide.* J. Biol. Chem. **270:**8209-8216.

125. **Murphy, F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (ed.).** 1994. Virus taxonomy: The classification and nomenclature of viruses, Retroviridae. Springer-Verlag, Vienna.

126. **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:**411-419.

127. **Ono, A., J. M. Orenstein, and E. O. Freed.** 2000. Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. J. Virol. **74:**2855-2866.

128. **Paillart, J. -C., E. Westhof, C. Ehesmann, 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.

129. **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:**27486-27493.

130. **Paillart, J. -C., R. Marquet, E. Skripkin, C. Ehresmann, and B. Ehresmann.** 1996. Dimerization of retroviral genomic RNAs: structural and functional implications. Biochimie **78:**639-653.

131. **Paillart, J.-C., L. Berthoux, M. Ottman, 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 synthesis. J. Virol. **70:**8348-8354.

132. **Paxton, W., R. I. Connor , and N. R. Landau.** 1993. Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of Gag and mutational analysis. J. Virol. **67:**7229-7237.

133. **Perez, O. D., and G. P. Nolan.** 2001. Resistance is futile: assimilation of cellular machinery by HIV-1: Immunity **15:**687-690

134. **Piguet, V., and D. Trono.** 1999. The Nef protein of primate lentiviruses. Rev. Med. Virol. **9:**111-120.

135. **Pluymers, W., P. Cherepanov, D. Schols, D. De Clercq, and Z. Debyser.** 1999. Nuclear localization of human immunodeficiency virus type 1 integrase expressed as a fusion protein with green fluorescent protein. Virology **258:**327-332.

136. **Poon, D. T. K., and 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:**6607-6616.

137. **Poon, D. T. K., G. Li, and A. Aldovini.** 1998. Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. J. Virol. **72:**1983-1993.

138. **Provitera, P., A. Goff, A. Harenberg, F. Bouamr, C. Carter, and S. Scarlata.** 2001. Role of the major homology region in assembly of HIV-1 Gag. Biochemistry **40:**5565-5572.

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

140. **Rein, A., L. E. Henderson, and J. G. Levin.** 1998. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. (review).Trends Biochem. Sci. **23:**297-301

141. **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.

142. **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:**959-967.

143. **Schmalzbauer, E., B. Strack, J. Dannull, S. Guehmann, and K. Moelling.** 1996. Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding *in vitro*. J. Virol. **70:**771-777.

144. **Schwartz, M. D., D. Fiore, and A. T. Panganiban.** 1997. District 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:**9295-9305.

145. **Shehu-Xhilaga, M., H. G. Kraeusslich, S. Pettit, R. Swanstrom, J. Y. Lee, J. A. Marshall, S. M. Crowe, and J. Mak.** 2001. Proteolytic processing of the

p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. J. Virol. **75:**9156-9164.

146. **Shen N., L. Jetté, 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:**5729-5735.

147. **Shen, N., L. Jetté, C. Liang, M. A. Wainberg, and M. Laughrea.** 2002. Intergenic Suppressor mutations in the human immunodeficiency virus Gag region restore genomic RNA dimerization and viral replication in viruses bearing a G/C-poor DIS, but not in viruses bearing a denatured DIS. Sent for publication.

148. **Shen, N., L. Jetté, M. A. Wainberg, and M. Laughrea.** 2001. Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization. J. Virol. **75:**10543-10549.

149. **Siegal, F. P., C. Lopez, G. S. Hammer, A. E. Brown, S. J. Kornfeld, J. Gold, J. Hassett, S. Z. Hirschman, C. Cunningham-Rundles, B. R. Adelsberg, D. M. Parham, M. Siegal, M. Cunningham-Rundles, and D. Armstrong.** 1981. Severe acquired immunodeficiency in Male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesions. New Engl. J. Med. **305:**1439-1444.

150. **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. USA **91:**4945-4949.

151. **Sova, P., W. Chao, and D. J. Volsky.** 1997. The redox state of cysteines in human immunodeficiency virus type 1 Vif in infected cells and in virions. Biochem. Biophys. Res. Comm. **240:**257-260.

152. **Stevenson, M.** 2000. HIV nuclear import: What's all the flap? Natrue Med. **6:**9-10.

153. **Stoltzfus, C. M. and P. N. Snyder.** 1975. Structure of B77 sarcoma virus RNA: stabilization of RNA after packaging. J. Virol. **16:**1161-1170.

154. **Summers, M. F., L. E. Henderson, M. R. Chance, J. W. Jr. Bess T. L. South, P. R. Blake, I. Sagi, G. Perez-Alvarado, R. C. 3rd. Sowder, and D. R. Hare.**

1992. Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein form HIV-1. Protein Sci. **1:**563-574.

155. **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. USA **90:**3393-3397.

156. **Swanstrom, R., and J. W. Wills.** 1997. Synthesis, assembly, and processing of viral proteins. p263-334. *In* J. M. Coffin, S. H. Hugbes, and H. E. Varmus (ed.), Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

157. **Takahashi, K., S. Baba, Y. Hayashi, Y. Koyanagi, N. Yamamoto, H. Takaku, and G. Kawai.** 2000. NMR analysis of intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. J. Biochem. **127:**681-686.

158. **Takahashi, K., S. Baba, Y. Koyanagi, N. Yamamoto, H. Takaku, and G. Kawai.** 2001. Two basic regions of NCp7 are sufficient for conformational conversion of HIV-1 dimerization initiation site from kissing-loop dimer to extended-duplex dimer. J. Biol. Chem. **276:**31274-31278.

159. **Tang, S., T. Murakami, A. B. Agresta, S. Campbell, E. O. Freed, and J. G. Levin.** 2001. Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. J. Virol. **75:**9357-9366.

160. **Telesnitsky, A., and S. P. Goff.** 1997. Reverse transcriptase and the generation of retroviral DNA. p121-160. *In* J. M. Coffin, S. H. Hugbes, and H. E. Varmus (ed.), Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

161. **Temin, H. M., and S. Mizutani.** 1970. RNA-dependent DNA Polymerase in virions of RSV. Nature **226:**211-213.

162. **Thali, M., A. Bukovsky, E. Kondo, B. Rosenwirth, C. T. Walsh, J. Sodroski, and H. G. Göttlinger.**1994. Functional association of cyclophilin A with HIV-1 virions. Nature **372:**363-365.

163. **Theilleux-Delanande, V., F. Girard, T. Huynh-Dinh, G. Lancelot, and J. Paoletti.** 2000, The HIV-1Lai RNA dimerization: Thremodynamic parameters

associated with the transition from the kissing complex to the extended dimer. Eur. J. Biochem. **267:**2711-2719.

164. **Tisne, C., B. P. Roques, and F. Dardel.** 2001. Heteronuclear NMR studies of the interaction of tRNAlys3 with HIV-1 nucleocapsid protein. J. Mol. Biol. **306:**443-454.

165. **Turelli, P., V. Doucas, E. Craig, B. Mangeat, N. Klages, R. Evans, G. Kalpana, and D. Trono.** 2001. Cytoplamic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. Mol. Cell **7:**1245-1254.

166. **Urbaneja, M. A., M. Wu, J. R. Casas-Finet, and R. L. Karpel.** 2002. HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. J. Mol. Biol. **318:**749-764.

167. **Vogt, P. K.** 1997. Historical introduction to the general properties of retroviruses. p 1-26. *In* J. M. Coffin, S. H. Hugbes, and H. E. Varmus (ed.), Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

168. **Vogt, V. M.** 1997. Retroviral virions and genomes. p 27-70. *In* J. M. Coffin, S. H. Hugbes, and H. E. Varmus (ed.), Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

169. **Vogt, V. M., and M. N. Simon**. 1999. Mass determination of Rous Sarcoma Virus virions by scanning transmission electron microscopy. J. Virol. **73:**7050-7055.

170. **Von Schwedler, U., R. S. Kornbluth, and D. Trono.** 1994. The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. Proc. Natl Acad. Sci. USA. **91:** 6992-6996.

171. **Weiss, R. A.** 2001. Gulliver's travels in HIVland . Nature **410:**963-967.

172. **Weiss, S., G. Hausl, M. Famulok, and B. Konig.** 1993. The multimerization state of retroviral RNA is modulated by ammonium ions and affects HIV-1 full-length cDNA synthesis *in vitro*. Nucleic Acids Res. **21:**4879-4885.

173. **Wills, J. W., and R. C. Craven.** 1991. Form, function, and use of retroviral Gag proteins. AIDS **5:**639-654.

174. **Zennou, V., C. Petit, D. Guetard, U. Nerhbass, L. Montagnier, and P. Charneau.** 2000. HIV-1 genome nuclear import is mediated by a central DNA flap. Cell **101:**173-185.

175. **Zhang, H., R. J. Pomerantz, G. Dornadula, and Y. Sun.** 2000. Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. J. Virol. **74:**8252-8261.

Preface (Chapter 2)

In this chapter, we studied the effect of mutations near the 3'end of the leader region, including kissing loop domain (KLD or SL1) on viral replication, genome dimerization, encapsidation and reverse transcription. KLD consists of a kissing loop hairpin (klh or DIS hairpin) and a putative stem B, loop B. klh is composed of stem C and loop C, which contains a palindrome, also called autocomplementary sequence (ACS). We have performed a detailed study of all the constitutive parts of KLD on the above mentioned aspects of viral life cycle. We studied the mutants of (1) deletion from downstream of PBS to the two adenines 5' to the palindrome (or ACS) (Δ200-256), (2) deletion of 3' of leader region (Δ241-256), (3) deletion of 5' of stem C (Δ248-256), (4) deletion of 5' of stem-loop B (Δ243-247), and (5) transversion mutations in loop C (change the palindrome from wild type GCGCGC262 to GUUAAC, GGGCGC, GGGCCC, GAGCUC. We also asked whether dimerization could be dissociated from packaging.

Chapter 2. Impact of Human Immunodeficiency Virus Type 1 RNA Dimerization on Viral Infectivtiy and of Stem-Loop B on RNA Dimerization and Reverse Transcription and Dissociation of Dimerization from Packaging

2. 1. ABSTRACT

The kissing-loop domain (KLD) encompasses a stem-loop, named kissing-loop or dimerization initiation site (DIS) hairpin (nucleotides [nt] 248 to 270 in the human immunodeficiency virus type 1 strains HIV-1_{Lai} and HIV-1_{Hxb2}), seated on top of a 12-nt stem-internal loop called stem-loop B (nt 243 to 247 and 271 to 277). Destroying stem-loop B reduced genome dimerization by ~50% and proviral DNA synthesis by ~85% and left unchanged the dissociation temperature of dimeric genomic RNA. The most affected step of reverse transcription was plus-strand DNA transfer, which was reduced by ~80%. Deleting nt 241 to 256 or 200 to 256 did not reduce genome dimerization significantly more than the destruction of stem-loop B or the DIS hairpin. We conclude that the KLD is nonmodular: mutations in stem-loop B and in the DIS hairpin have similar effects on genome dimerization, reverse transcription and encapsidation, and are also "nonadditive", i.e. a larger deletion spanning both of these structures has the same effects on genome dimerization and encapsidation as if stem-loop B strongly impacted DIS hairpin function and vice-versa. A C258G transversion in the palindrome of the kissing-loop reduced genome dimerization by ~50% and viral infectivity by ~1.4 logs. Two mutations, CGCG261→UUAA261 (creating a weaker palindrome) and a Δ241-256 suppressor mutation, were each able to reduce genome dimerization but leave genome packaging unaffected.

2. 2. Introduction

The kissing loop domain (KLD) encompasses a stem-loop, named kissing-loop hairpin (nucleotides [nt] 248 to 270 in human immunodeficiency virus type 1 strain HIV-1_{Lai} and HIV-1_{Hxb2} genomic RNA), seated on top of a short stem-internal loop called stem-loop B (nt 243 to 247 and 271 to 277) (18). The apical loop of the kissing-loop hairpin contains an almost invariant hexameric autocomplementary sequence (ACS) (see reference 17 and references therein), also called a palindrome. The palindrome is seen as the dimerization initiation site (DIS) of genomic RNA (13, 15, 31); thus the kissing-loop hairpin is also called the DIS hairpin. The level of genomic RNA dimerization within isolated HIV-1 viruses is influenced by the DIS hairpin (6, 9, 17) and p55Gag processing (8).

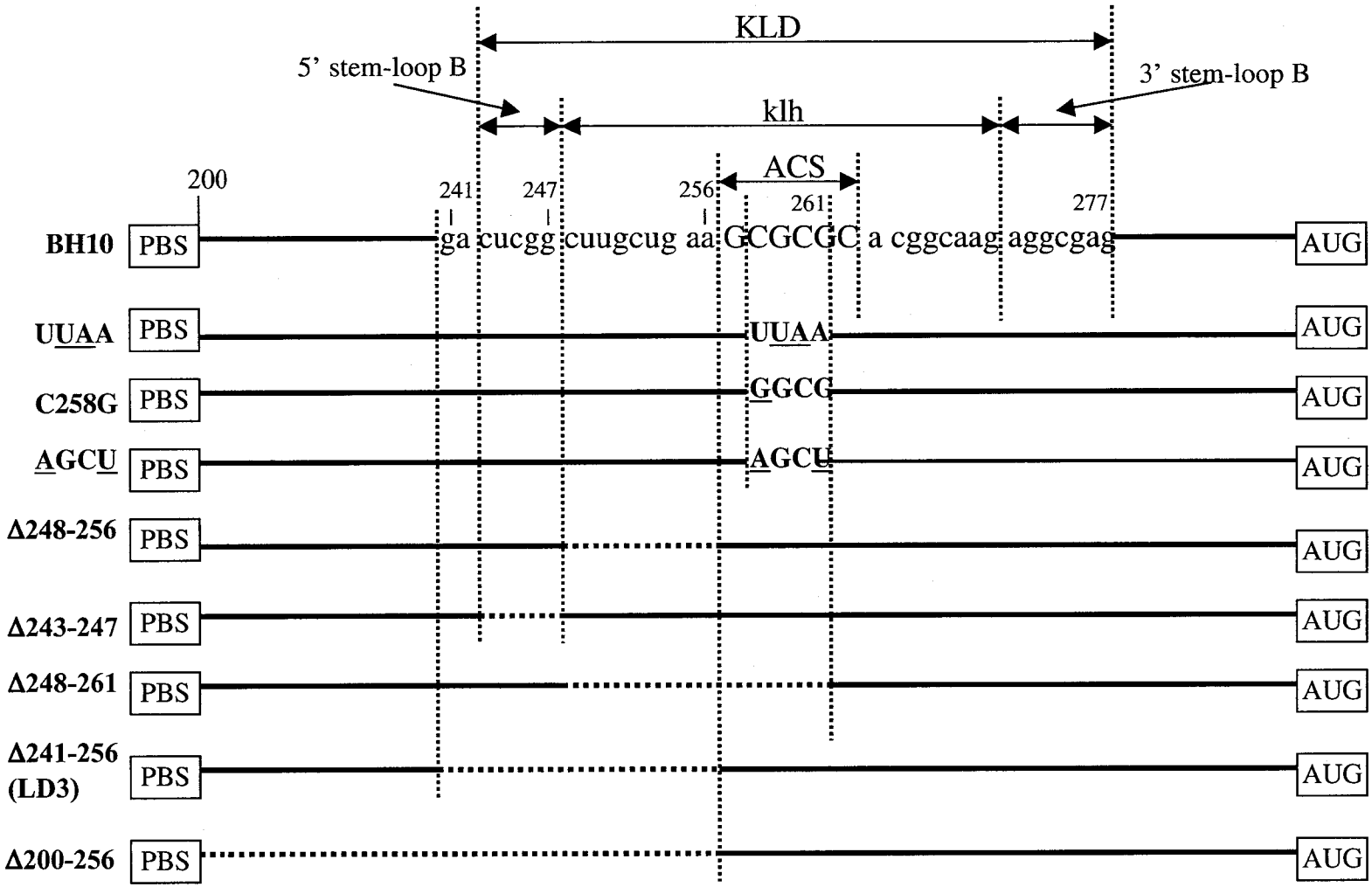
In the kissing-loop model of HIV-1 genome dimerization (13, 15, 31), stem-loop B has ill-defined roles (15, 17); one might be to properly orient the DIS hairpin within the covalently linked 9,000-nt-long tangle of secondary and tertiary structure (18). Experimentally, substantial deletions within stem-loop B or the DIS hairpin have identical impacts on viral infectivity and genomic RNA encapsidation (18). This raises the possibility that the KLD might be nonmodular, i.e., a highly integrated structure whereby stem-loop B and the DIS hairpin may have similar, if not identical, physiological impacts. To establish this, it is necessary to show that stem-loop B mutations inhibit genomic RNA dimerization and proviral DNA synthesis, two processes affected by the DIS hairpin (6, 9, 17, 25).

In this paper, we identify a crucial role of stem-loop B in genome dimerization and reverse transcription and compare its physiological impact to that of the DIS hairpin. We also relate genome dimerization to viral infectivity via studying a point mutation unlikely to directly impact more than genome dimerization. The transversion C258G transforms the GCGCGC ACS into the nonpalindrome GGGCGC reduces viral infectivity by ~1.4 logs [~2 logs less than larger mutations within the DIS hairpin (18)] and should not affect proviral DNA synthesis (25). Finally, we scrutinize the links between genome dimerization and genome packaging, via the study of two additional and nonoverlapping mutations. First, an ACS mutation (CGCG261→UUAA261) which

preserves the palindromic nature of the ACS strongly reduces viral infectivity and in vitro dimerization of RNA transcripts, while leaving genomic RNA packaging unaffected (18). Second, a double mutation in nucleocapsid protein NCp7 and the p2 peptide partially suppresses the effects of KLD destruction, i.e., reverts genome packaging to wild-type level and increases viral replication to a level of ~1.4 logs below that of wild-type (21, 23). We shall examine if the genomes of these two disparate and poorly infectious mutants, as well as the genome of the C258G transvertant, are poorly dimeric despite being adequately packaged.

2. 3. Experimental methods

Production of mutant viruses. COS-7 cells were transfected in parallel with equal amounts of wild-type plasmid pSVC21.BH10 and mutant plasmids such as pSVC21Δ243-247, Δ241-256, Δ200-256, LD3-MP2-MNC, GGCG, UUAA, GGCC, AGCU, Δ248-256, and Δ248-261. In pSVC21Δ243-247, CUCGG247 has been deleted; in pSVC21GGCG, CGCG261 has been replaced by GGCG mutatis mutandis for the other plasmids. Nucleotides differing from those in HIV-1_{Lai} and HIV-1_{Mal} (subtype ADI) are underlined. The sequence of the BH10 KLD is 243CUCGGCUUGCUGAAGCGCGCACGGCAAGAGGCGAG277; nucleotides forming stem B (13) and stem C (17) (the stem of the DIS hairpin) are underlined. Comparable amounts of viruses were produced at 48 h posttransfection. To investigate the effects of the mutations on genomic RNA dimerization, genomic RNA was extracted from the isolated viruses, electrophoresed on a nondenaturating agarose gel and detected by Northern blotting with a ^{35}S-labeled HIV-1 riboprobe (17).



LD3-MNC-MP2

Fig. 1'. List of mutants of KLD and leader region studied in this chapter


1 2 3 4
d –
m –

FIG. 1. Dimerization level of viral RNA isolated from BH10 (lanes 1 and 4), Δ248-256 (lane 2) and Δ243-247 (lane 3) viruses. HIV-1 genomic RNA was isolated and analyzed by nondenaturing Northern blot analysis as described previously (17). Viral RNAs were dissolved in 8 μl of buffer S (10 mM Tris [pH 7.5], 100 mM NaCl, 1 mM EDTA, 1% sodium dodecyl sulfate) and subjected to electrophoresis (70V, 5 h 20 min, 0.7% agarose in buffer TBE_2 [Tris-borate-EDTA] at 4°C). The samples were next Northern blotted, hybridized, and autoradiographed for 8 h. d, dimer; m, monomer. Each lane represents an independent transfection and contains viral RNA isolated from 3.7 to 9 tissue culture dishes. Without a CAp24 estimate, efficiency of RNA packaging cannot be derived because of variations in virus yield from one transfection to the next. Unless otherwise indicated, the samples of Fig. 2 to 4 were processed as in Fig. 1.

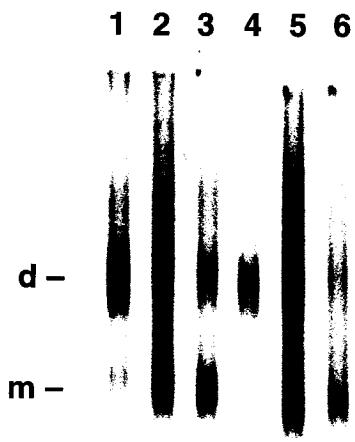

1 2 3 4 5 6
d –
m –

FIG. 2. Dimerization level of viral RNA isolated from BH10 (lane 1), Δ241-256 (lane 2) and LD3-MP2-MNC (lane 3) viruses, respectively, containing 8 x 10^{12}, 23 x 10^{12} and 8 x 10^{12} CAp24 copies. Electrophoresis: 5 h 15 min in 0.8% agarose. The autoradiographic exposure was for 2 h. Scanning lanes 1 to 3 and equivalent lanes from other Northern blots (not shown) indicate that Δ241-256 and LD3-MP2-MNC viruses, respectively, package genomic RNA 0.4 ± 0.08 and 0.81 ± 0.15 as well as BH10 viruses. These values are close to what was previously estimated with slot blot assays (21). Lanes 4 to 6, dimerization level of viral RNA isolated from 1.8 to 4.2 tissue culture dishes of BH10 (lane 4), Δ243-247 (lane 5), and Δ200-256 (lane 6) viruses. Electrophoresis was for 5 h 20 min in 0.7% agarose. The autoradiographic exposure was for 25 min.

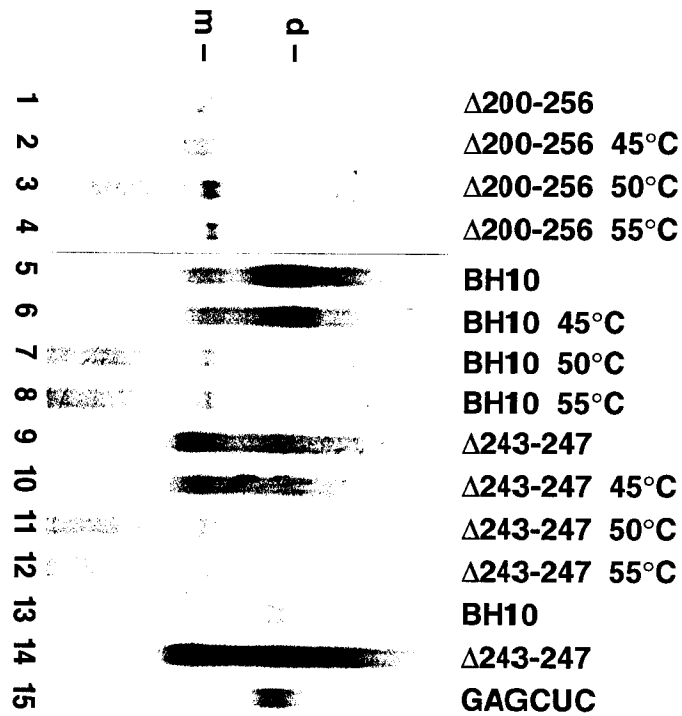

m –
d –
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Δ200-256
Δ200-256 45°C
Δ200-256 50°C
Δ200-256 55°C
BH10
BH10 45°C
BH10 50°C
BH10 55°C
Δ243-247
Δ243-247 45°C
Δ243-247 50°C
Δ243-247 55°C
BH10
Δ243-247
GAGCUC

FIG. 3. Dimerization level and thermal stability of viral RNA isolated from 2.5 to 4 tissue culture dishes of the Δ200-256, BH10, Δ243-247 viruses (lanes 1 to 12). Nonheated samples were left on ice for 10 min. The other samples were incubated in buffer S at the temperature indicated for 10 min. After incubation, all samples were loaded without delay and with the voltage on. Lanes 13 to 15 give the dimerization level of viral RNA isolated from two to eight tissue culture dishes of BH10, Δ243-247 and GAGCUC viruses (i.e., viruses whose GCGCGC262 palindrome has been replaced by GAGCUC). The autoradiographic exposure were for 6 h for lanes 1 to 4 and 3 h for lanes 5 to 15. Electrophoresis was for 4 h 30 min in 0.8% agarose.

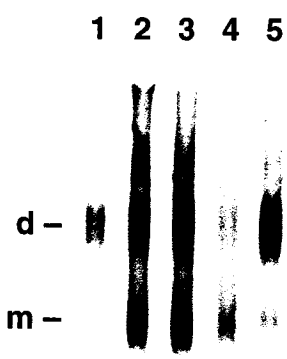

1 2 3 4 5
d –
m –

FIG. 4. Dimerization level of viral RNA isolated from BH10 (lane 1), GGCG (lane 2), UUAA (lane 3), and Δ248-256 (lane 4) viruses containing 1.3 x 10^{12}, 5 x 10^{12}, 5 x 10^{12}, and 4 x 10^{12} CAp24 copies. GGCG viruses have a C258G transversion. UUAA viruses have CGCG261 replaced by UUAA. Lane 5, dimerization level of viral RNA isolated from BH10 viruses containing 4 x 10^{12} CAp24 copies. Electrophoresis was for 4 h 35 min in 1% agarose. The autoradiographic exposure was for 1 h. Scanning lanes 1 to 3 and equivalent lanes form other Northern blots indicates that GGCG, UUAA and Δ248-256 viruses, respectively, package genomic RNA 0.78 ± 0.12, 1.1 ± 0.11 and 0.4 ± 0.06 as well as BH10 viruses. These numbers are consistent with previous estimates using dot blot assays (18).

TABLE 1. Mutant proviral DNA present in infected MT2 cells, as % of wild-type proviral DNA, 10 h after infection[a], and specific impact of stem-loop B and DIS hairpin destruction on early, intermediate and late events of reverse transcription.

Mutnat	% of wild-type DNA with primer pairs				
	pR-pMA'	pPBS-pMA' (~early events[b])	pR-pU5' (~intermediate events[c])	pPBS-pMA'/pR-pU5'	pR-pMA'/pPBS-pMA' (~late events[d])
Δ243-247	12± 4 (7± 2)	65±30 (40 ± 20)	80±20 (45±12)	80±45	18±11
Δ248-256 and Δ248-261[e]	13±6 (8±4)	60±15 (36±10)	75±25 (45±15)	80±35	22±11
control (BH10)	100[f] (100)	100[g] (100)	100[h] (100)	100	100

[a]PCR was as described in the text. Primers pR (5' AGA CCA GAT CTG AGC CTG GGA G35) and pMA' (CT GAC GCT CTC GCA CCC: [the antisense of nt 338 to 354]), respectively, hybridize to the minus strand of R DNA and to the plus strand of matrix DNA. Primers pPBS and pU5' (5' CCA CAC TGA CTA AAA GGG TC: [the antisense of nt 148 to 167]), respectively, hybridize to the minus strand of the PBS and to the plus-strand of U5 DNA. For pR-pMA', pPBS-MA', and pR-pU5', the amount of mutant proviral DNA per unit of ß-globin DNA was divided by the amount of wild-type proviral DNA per unit of ß-globin DNA, multiplied by 100, and corrected for deficient genomic RNA packaging in the mutant viruses (18). Values in parentheses: are the numbers obtained before correction for deficient genomic RNA packaging. No correction was made for the presumed two- to fourfold-higher proportion of subgenomic RNA present in the mutant viruses (6,24). This correction would have been negligible with primer pairs pR-pMA' and pPBS-pMA' (PCR signal lost by hypothetical transfer of minus-strand strong-stop DNA from the genomic to subgenomic strand would have been almost completely regained by transfer of minus-strand strong-stop DNA from the subgenomic to genomic strand) and small with the primer pair pR-pU5' (see footnote b).

[b]This reflects or exaggerates the specific effect of the mutations on minus-strand strong-stop DNA production (maximum extension of minus-strand DNA prior to minus-strand transfer). Why "exaggerates"? A worst case scenario follows. Assume that all wild-type proviral DNA synthesis was completed at 10 h (doubtful), while mutant proviral DNA synthesis was unaffected at the level of minus-strand strong-stop DNA synthesis, but 100% blocked at the level of minus-strand transfer (clearly not the case). A mutant PCR yield of 25% of wild-type level would be reported, because each mutant proviral DNA would have been single stranded (with one pR/pU5' binding site), while wild-type proviral DNA would have contained two LTRs (and thus would have had four pR/pU5' binding sites). If wild-type proviral DNA synthesis was completed at 10 h while mutant proviral DNA synthesis was 100% blocked at the level of plus-strand transfer (not quite the case), a mutant PCR yield of 50% of wild-type would be reported (because of one LTR per mutant DNA versus two in the wild-type). On the other hand, the higher relative proportion of subgenomic RNAs expected in Δ243-247, Δ248-256,and Δ248-261 viruses will increase this 50% yield to 55 to 59% (assuming 0.2 to 0.3 subgenomic RNAs per genomic RNA in mutant viruses versus 0.1 in wild-type viruses). Let's now describe

a reasonable scenario: if BH10 proviral DNA synthesis was 90% completed at 10 h while mutant proviral DNA synthesis was 80% blocked at the level of plus-strand transfer (last column) and mutant viruses contained 0.3 subgenomic RNA per genomic RNA versus 0.1 in wild-type viruses (see references 6 and 24 and text), a mutant PCR yield of 76% of BH10 would obtain (64% in the absence of correction for differential packaging of spliced viral RNAs). Interestingly, this matches the percentage reported in the column. This modeling and the attending uncertainties should not distract from the main experimental result: stem-loop B and DIS hairpin mutations have indistinguishable effects on reverse transcription. Also, the fact that the pR-pU5' column yields averages above 50% suggests that production of plus strand strong-stop DNA is not affected in the mutant viruses.

[c]This division reflects or possibly understimates by a factor ~2 the aggregate effect of the mutations on minus-strand transfer plus formation of 9.1-kb minus strand DNA (maximum extension of minus-strand DNA produced prior to plus-strand transfer). Indeed, since the percentage expressed in the pR-pU5' column may reflect more than the effect of mutations on minus-strand strong-stop cDNA synthesis (see above), it follows that dividing the pPBS-pMA' column by the pR-pU5' column may underestimate the effect of the mutations on minus-strand transfer plus formation of 9.1-kb minus-strand DNA.

[d]This division reflects the aggregate effect of the mutations on a mixture of late events including initiation of plus-strand DNA synthesis, plus-strand transfer, and displacement of the minus strand of the 5' LTR by the nascent extension of 9.1 kb minus-strand DNA.

[e]The effects of deleting nt 248 to 261 or 248 to 256 were insignificantly different. With each set of primers, Δ248-256 viruses were actually on average slightly less productive than Δ248-261 viruses, consistent with the idea that the ACS does not influence proviral DNA synthesis (25).

[f]On average, there were ~2,000 apparent DNA copies per 10^5 cells. At 2 h postinfection, there was <50 apparent DNA copies per 10^5 cells. One apparent DNA copy is the amount of proviral DNA giving the same PCR signal as one linearized SVC21.BH10 plasmid. The actual number depends on the number of PCR primer binding sites in the nascent proviral DNA relative to the number of PCR primer binding sites in linearized SVC21.BH10.

[g]On average, there were ~2,000 apparent DNA copies per 10^5 cells. At 2 h posttransfection, there were ~400 apparent DNA copies per 10^5 cells.

[h]On average, there were ~4,000 apparent DNA copies per 10^5 cells. At 2 h posttransfection, there were ~2,000 apparent DNA copies per 10^5 cells.

2. 4. Results and discussion

Stem-loop B mutation reduces genome dimerization as much as DIS hairpin destruction. Figure 1 compares Δ243-247 genomic RNA (lane 3) to Δ248-256 (lane 2) and BH10 (lanes 1 and 4) genomic RNA. There were two similarly labeled RNA species in the Δ243-247 and Δ248-256 RNA preparations, the lower band having the same mobility as monomeric genomic RNA. Scanning lanes 1 to 4, as well as many other gel lanes from independent transfections (not shown), reveals three novel pieces of informations. (i) Δ243-247 RNA was 48% ± 5% dimeric vs 85% ± 3% and 45% ± 5% dimeric for BH10 and Δ248-256 genomic RNAs. Destruction of stem-loop B had the same effect as destruction of stem C. (ii) The dimer band in the Δ243-247 and Δ248-256 samples contained 60% ± 15% more high-molecular-weight (trimer-like) complexes per unit of RNA loaded than the BH10 samples. (In BH10 RNA, the multimer shoulder amounted to 18% ± 4% of the dimer plus monomer peaks. Such shoulders are not unusual [6, 8, 9, 28].) (iii) The two mutant RNAs appeared electrophoretically diffuse as well because the trough separating the monomer and dimer peaks was not as RNA free as in the BH10 samples. To quantitate this impression, we measured, relative to the baseline of the scans, the height of the dimer peak (d), the height of the monomer peak (m) and the minimum height recorded in the trough (t). The t/(m + d) ratios were, respectively, 0.3 ± 0.02 and 0.13 ± 0.03 in the mutant and BH10 samples (averaged over four to nine Northern blots). As an additional comparison, we found that genomic RNA isolated from Δ248-261 viruses (20) was 40% ± 4% dimeric and had a t/(m + d) of 0.25 ± 0.03 (not shown; data from five Northern blots).

Stem-loop B and stem C mutations do not have additive effects on dimerization: each is as potent as deletion of nt 200 to 256. The genomic RNA of Δ241-256 viruses was 45% ± 4% dimeric (lane 2 of Fig. 2): mutations in stem-loop B and stem C do not have additive effects on genome dimerization. Furthermore, genomic RNA from Δ200-256 viruses was 42% ± 3% dimeric (lane 6 of Fig. 2). Thus, removal of all nucleotides separating the primer binding site (PBS) (nts 182 to 199) from the ACS had no more impact than destroying stem C or stem-loop B. The Δ241-256 and Δ200-256 RNAs contained 30% ± 20% more higher-molecular-weight complexes than the

BH10 samples, and their t/(m + d) was 0.28 ± 0.03. Hence the dimer and monomer bands were relatively diffuse.

Why does stem-loop B contribute as much to genome dimerization as the DIS hairpin, but in a nonadditive way? Stem-loop B may facilitate DIS hairpin folding (i.e., reduce the probability of alternative foldings) or orient the DIS hairpin "head" away from improper (interfering) contacts with downstream or upstream sequences, thereby increasing the probability of DIS hairpin dimerization. If genomic RNA was sufficiently shortened (sequences causing "alternative foldings" and "interfering contacts" getting deleted), the stem-loop B "neck" would no longer be needed. This head-and-neck model is supported by in vitro data showing that partial HIV-1 RNA transcripts lacking nt 243 to 247 can dimerize 10 times more poorly when they end hundreds of nts 3' of the KLD than when they end <20 nt 3' of the KLD (18). The model predicts that destruction of stem B should substantially inhibit genomic RNA dimerization, while destruction of the KLD should reduce genome dimerization no more than deleting of nt 248 to 256 or 248 to 261, which is what we found above. To reinforce this point, Fig. 3 shows that genomic RNAs from Δ200-256 (lanes 1 to 4) viruses appeared as thermostable as genomic RNA from BH10 (lanes 5 to 8) or Δ243-247 (lanes 9 to 12) viruses. BH10 RNA was monomeric at 50°C (lane 7) and 78% dimeric at 45°C (lane 6), consistent with a dissociation temperature (T_d) of ~47 to 48°C; Δ200-256 and Δ243-247 RNAs were monomeric or close to monomeric at 50°C (lanes 3 and 11) and, at 45°C, as dimeric as the unheated samples (lanes 2 and 10), consistent with a T_d of ~48 to 49°C.

C258G reduces genome dimerization as effectively as Δ200-256, without, however, generating diffuse dimer and monomer bands. We have shown that deletions of 5 to 56 nt, as long as they destroy stem-loop B or the DIS hairpin, reduce genome dimerization by half. The kissing-loop model predicts that these deletions act by obstructing ACS function. If true, transforming the ACS into a nonpalindrome via 1-nt substitution should have a comparable impact on dimerization. Accordingly, we replaced the BH10 ACS by the GGGCGC262 nonpalindrome to generate GGCG (or C258G) viruses. Lane 2 of Fig. 4 shows that genomic RNA from C258G viruses was 50% dimeric. Scanning this and several other C258G gel lanes (not shown) revealed that the mutant genomic RNA was 48% ± 4% dimeric, versus 44% dimeric for the average

deletion mutant and 42% dimeric for Δ200-256 RNA (see above; e.g., compare lane 2 of Fig. 4 to lane 6 of Fig. 2). The t/(m + d) ratio (0.16 ± 0.02) and the multimer content (17% ± 3%) of C258G RNA were as low as in BH10 RNA. C258G represents the smallest molecular change ever shown to affect HIV genomic RNA dimerization.

As a control, we engineered a C258G suppressor mutation, i.e., replaced the BH10 ACS by GGGCCC262, a theoretically valid palindrome nevertheless absent from all sequenced HIVs and simian immunodeficiency viruses (SIVs) (12, 18). The genome of the double transvertant viruses was 83% ± 3% dimeric, (i.e., wild-type like [data not shown]). The infectivity of the double transvertant was not significantly impaired (18), supporting the idea that wild-type infectivity requires wild-type levels of genome dimerization.

Good packaging stoichiometry but poor dimerization: three examples. Replacing the BH10 ACS by GUUAAC leaves genomic RNA packaging unaffected (18). Lane 3 of Fig. 4 shows that genomic RNA from UUAA viruses was half-dimeric and appeared as well-resolved dimer and monomer bands. Scans of this and other UUAA lanes (not shown) revealed that UUAA RNAs were 55% ± 4% dimeric, had a t/(m + d) ratio of 0.15 ± 0.02 and a multimer content of 21% ± 2%, (i.e., were electrophoretically indistinguishable from C258G RNAs). Northern blot analyses (intensity of the monomer and dimer bands per unit of CAp24 loaded) confirm that UUAA viruses packaged genomic RNA as well as BH10 (Fig. 4). UUAA genomic RNA was much less dimeric than genomic RNA from AGCU viruses (lane 15 of Fig. 3). AGCU is absent from the palindromes of allHIVs and SIVs so far sequenced (12, 16), while UUAA is found in the putative DIS hairpin of SIV_{mnd} (18).

Point mutations T24I in NCp7 and T12I in the p2 peptide suppress the deletion of nt 241 to 256in three ways: (i) Genome packaging is boosted from ~30% to ≥80% of wild-type (21 [see below]); ii) viral replication is boosted by at least 2 logs (21; [see below]). (iii) The maturation kinetics of the CAp24-p2 protein into CAp24 and p2 seems to be ≥70% restored (22). (The name of the Δ241-256 suppressor is LD3-MP2-MNC, with LD3 being a designation for Δ241-256.) However, lane 3 of Fig. 2 shows that genomic RNA from LD3-MP2-MNC viruses was 42% dimeric. The average over several Northern blots was 43% ± 6%, undistinguishable from the 45% ± 4% obtained with

Δ241-256 viruses. The NCp7 and p2 suppressor mutations failed to suppress the Δ241-256 defect in genome dimerization. On the other hand, the ratio t/(m + d) and the multimer content of LD3-MP2-MNC RNA were wild-type-like (lane 3 of Fig. 2; [not shown]). Northern blot analyses confirm that LD3-MP2-MNC viruses packaged genomic RNA almost as well as BH10 viruses (Fig. 2).

Finally, Northern blot analyses suggests that C258G viruses packaged genomic RNA 78% ± 12% as well as BH10 viruses (Fig. 4). This is consistent with a previous report showing that two palindrome point mutations (G261U and GCGCGC262 GGGGGG262) did not grossly reduce genomic RNA packaging in virions (6).

These three convergent observations do not necessarily mean that genome packaging has in all respects been dissociated from genome dimerization: a KLD-independent signal might exist which stimulates cytoplasmic formation of dimers prone to dissociate during genomic RNA extraction or electrophoresis. Conversely, the C258G, UUAA and LD3-MP2-MNC RNAs packaged in the mutant viruses may have been monomeric prior to protease-directed maturation of $p55^{Gag}$ (8). In addition, we cannot strictly exclude the idea that the majority (~60 to 65%) of the three mutant viruses contained one genomic RNA per 50% smaller protein shell: the diameter, surface (CAp24 content), and packaging efficiency (number of genomic RNAs per unit of CAp24) of these small viruses would, respectively, be ~79, ~63, and ~79% of those of BH10. However HIV-1 viruses severely deficient in packaging do not appear significantly smaller than the wild-type (1, 5). Finally, even if the mutant viruses could consistently package two genomic RNAs per protein shell, assembling a protein shell around two monomeric genomic RNAs (or a dimer lacking a functional KLD) might be kinetically slower than assembling it around one fully dimeric BH10 genome.

If losing KLD-directed genome dimerization does not inhibit packaging, why were Δ241-256 viruses encapsidation defective? Δ241-256 viruses may have lost two independent signals: a dimerization signal and a separate packaging signal. We imagine that the T24I mutation in the NCp7 of LD3-MP2-MNC viruses compensates for the latter loss by enhancing some form of genomic RNA binding beyond the capacity of BH10 NCp7 (described below).

Heterogenous migration of genomic RNAs from packaging-defective viruses: do they form complexes with subgenomic RNA? Δ243-247, Δ241-256, Δ248-256, and Δ248-261 viruses were packaging-defective (Fig. 2 and 4) (17, 18), and their RNA bands were diffuse (above). In contrast, C258G, UUAA and LD3-MP2-MNC viruses packaged genomic RNA ~90% as well as BH10 on average (Fig. 2 and 4) and presented sharp RNA bands. Since the amount of spliced viral RNA inside viruses mutated in the KLD is the same as that in wild-type viruses (6), if not larger (24), KLD mutants encapsidating half as well as BH10 should contain, per genomic RNA, two to four times more spliced viral RNA than BH10 viruses or KLD mutants encapsidating as well as BH10. If a large proportion of the RNAs moving as a shoulder behind the dimer band, and halfway between dimers and monomers, were dimers and monomers hybridized to subgenomic RNA(s), then a doubling or tripling of spliced viral RNAs per genomic RNA in Δ243-247, Δ241-256, Δ248-256 and Δ248-261 viruses might significantly increase the proportion of genomic RNAs moving as a shoulder behind the dimer band and halfway between dimers and monomers. This could explain why in Northern blots from these viruses, the relative height of the multimer shoulder was increased by ~45%, and the relative RNA content in the trough separating the monomer and the dimer band was increased ~ 2.2-fold.

Partial reduction in genome dimerization may cause a 25-fold reduction in viral infectivitiy. Using the 50% tissue culture infective dose method (17), we performed infectivity tests on, respectively, two and three independent preparations of LD3-MP2-MNC and Δ241-256 viruses: the logs of BH10 titer/mutant titer were, respectively, 1.35 ± 0.3 and 4.2 ± 0.5 (not shown) versus 1.45 ± 0.25 and 1.35 ± 0.25 for the C258G and UUAA viruses (18). The Δ241-256 and LD3-MP2-MNC titers confirm and extend similar titers independently obtained by Liang et al (23). LD3-MP2-MNC viruses are thus 11- to 50- fold less infectious than wild-type (20- to 25-fold on average), and they have a phenotype apparently indistinguishable from that of C258G and UUAA viruses, despite the nonoverlapping nature of the mutations: identical low infectivity, identically deficient genome dimerization, and little or no impact on genome packaging and other functions. It is tempting to suggest that the poor genome dimerization displayed by the C258G, UUAA and LD3-MP2-MNC strains causes their poor infectivity. If true,

reducing genomic RNA dimerization by 50% would reduce viral infectivity ~25-fold and at most 50-fold.

Stem-loop B mutation reduces proviral DNA production as much as DIS hairpin destruction. In the hope of discovering why destruction of stem-loop B inhibits dimerization no more than the C258G transversion but reduces viral infectivity by ~ 3.4 logs (18), we analyzed viral DNA present inside 2.5 x 10^6 MT-2 cells infected with equal amounts of DNase I-treated BH10 and mutant viruses, each containing 25 ng of CAp24 (17). Intracellular DNA harvested 10 h after infection (10,11,29,38) was subjected to 25 cycles of PCR (7) using primer pairs designed to amplify ß-globin DNA (27) and different regions of proviral DNA (20). Serial dilutions of pSVC21.BH10 linearized with SpeI served to quantitate intracellular proviral DNA and acted as a control for PCR linearity. The ß-globin gene acted as an internal control. The Δ248-256 and Δ248-261 data were pooled together as representing, for comparative purposes, DIS hairpin destruction (Table 1). We used primers pR and pU5' (Table 1) to quantitate DNA containing the RU5 sequence (i.e., minus-strand strong-stop DNA plus subsequent DNAs), primers pPBS and pMA' (Table 1 caption) to quantitate DNA containing the PBS-to-MA (matrix) DNA sequence (9.1-kb minus-strand DNA [Table 1] plus subsequent DNAs), and primers pR and pMA' to quantitate nearly complete proviral DNA.

The intracellular quantity of nearly complete proviral DNA was decreased ~15-fold when BH10 viruses were replaced by viruses mutated in stem-loop B or the DIS hairpin (pR-pMA' column of Table 1). After correction for poor genome packaging, the mutant reverse transcriptional machinery appeared approximately eightfold less productive than BH10, as if the stem-loop B and DIS hairpin poles of the KLD interfered strongly and equally with production of nearly complete proviral DNA. Is this inhibition due to decreased DNA synthesis or increased degradation? Assuming increased degradation leads to a difficulty. At 10 h post-infection, the RU5 and PBS-MA DNA would be minimally degraded (Table 1), while the more freshly synthesized R-MA DNA (merely RU5 and PBS-MA DNA linked via a 14-nt linker) would be extensively degraded prior to PCR: one would have to suppose that the 14-nt linker is about 50 times

more labile, per nucleotide residue, than flanking sequences. We conclude that our mutations are more likely to affect DNA synthesis than DNA degradation.

A late step of reverse transcription is preferentially inhibited. After correction for lower encapsidation, it was not clear that RU5 DNA production was decreased (pR-pU5' column of Table 1), and PBS-MA DNA production was not dramatically reduced (pPBS-pMA' column of Table 1), indicating that the destruction of stem-loop B preferentially impaired a step located between production of 9.1-kb minus-strand DNA and nearly complete proviral DNA (Table 1). Specifically, the steps leading from 9.1-kb minus-strand DNA to nearly complete proviral DNA were inhibited 70 to 90% (last column of Table 1), whereas all previous steps were inhibited less, if at all (pPBS-pMA' column of Table 1). Total cellular DNA harvested at 2 h postinfection gave PCR results consistent with the 10-h data. PCR at 10 and 2 h post-infection measures intraviral plus intracellular proviral DNA production (2, 26, 35, 39-41). Subtraction of the 2-h data from the 10-h data yields intracellular DNA production during the 8-h interval: we found that the ratios of mutant to BH10 production, after correction for RNA packaging, were very close to those reported in Table 1, namely ~ 0.6 with primer pairs pR-pU5' and pPBS-pMA' and ~ 0.15 with primer pair pR-pMA' (not shown).

Choking between synthesis of 9.1-kb minus-strand DNA and nearly complete proviral DNA could have three causes (34): (i) step A, poor formation of plus-strand strong-stop DNA (maximum extension of plus-strand DNA produced before plus-strand transfer). (This could be due to excessive degradation of the polypurine track, deficient initiation of plus-strand DNA synthesis, deficient DNA-directed elongation, poor displacement of the PBS RNA strand still hybridized to tRNAlys3 [3], or premature degradation of tRNAlys3.); (ii) step B, poor plus-strand transfer per se, i.e., poor annealing between plus-strand PBS DNA and minus-strand PBS DNA. (This could be due to step Bi poor accessibility of plus-strand PBS DNA because of slow degradation of $tRNA^{lys}_3$ (32), or step Bii, poor accessibility of minus-strand PBS DNA.); (iii) step C, poor displacement of the minus strand of the 5' long terminal repeat (LTR) (4).

Working hypothesis: KLD mutations result in a minus-strand PBS DNA folding which hinders annealing with plus-strand PBS DNA. KLD RNA cannot directly influence steps B and C, because it should by then be degraded. Even though

plus-strand strong-stop DNA is probably made some time before 9.1 kb minus-strand DNA (34,36), a direct effect of KLD RNA on step A seems unlikely; in addition, the relatively abundant accumulation of plus-strand strong-stop DNA in the cytoplasm of infected cells (34) suggests that formation of plus-strand strong-stop DNA is not rate limiting in plus-strand transfer. Minus-strand KLD DNA cannot influence steps A and Bi, because they are expected to precede its appearance (34). An effect solely on step C is ruled out, because blocking it would inhibit R-MA DNA production (relative to PBS-MA DNA production) by only 50% (because plus-strand DNA elongation would remain unhindered) versus the ~80% reported in Table 1; in addition, we can't easily conceptualize how minus-strand KLD DNA could affect 5' LTR strand displacement. An effect on step Bii is easier to imagine: mutant minus-strand KLD DNA, located only 43 nt upstream of minus-strand PBS DNA, might stimulate an aberrant minus-strand PBS DNA folding that hinders annealing with plus-strand PBS DNA. This model can be tested using a reconstituted reverse transcription system (38).

Nonobvious factors influencing late steps of reverse transcription. Vif protein (30, 37), NCp7 and the KLD are three nonobvious factors which can affect reverse transcription steps posterior to minus-strand strong-stop DNA production. NCp7 from HIV-1_{MN} stimulates annealing between plus-strand PBS DNA and minus-strand PBS DNA, at least in the context of proviral DNAs truncated before KLD DNA (38), and NCp7 from HIV-1_{NL43} may stimulate strand displacement posterior to plus-strand transfer (33). These NCp7 proteins, like those from LD3-MP2-MNC virus, have an Ile24 in their N-terminal zinc finger; only HIV-1_{Hxb2}, the source of BH10 viruses, and two subtype D viruses have a Thr24 in their NCp7 (12). Thus, NCp7 from LD3-MP2-MNC viruses is in this regard more conformist than wild-type NCp7 from BH10 and Δ241-256 viruses. Perhaps HIV-1_{Hxb2} is slightly underendowed in strand annealing and genomic RNA packaging: this underendowment may become rate-limiting in the context of a Δ241-256 deletion presumed to inhibit packaging in a dimerization-independent manner and to hinder annealing of plus-strand PBS DNA to minus-strand PBS DNA.

2. 5. Acknowledgements

This work was supported by grant MT-12312 from the Medical Research Council of Canada to M. Laughrea.

2. 6. REFERENCES

1. **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:**1920-1926.
2. **Arts, E. J., J. Mak, L. Kleiman, and M. A. Wainberg.** 1994. DNA found in human immunodeficiency virus type 1 may not be required for infectivity. J. Gen. Virol. **75:**1605-1613.
3. **Ben-Artzi, H., J. Shemesh, E. Zeelon, B. Amit, L. Kleiman, M. Gorecki, and A. Panet.** 1996. Molecular analysis of the second template switch during reverse transcription of the HIV RNA template. Biochemistry **35:**10549-10557.
4. **Boone, L.R., and A. M. Skalka.** 1993. Strand displacement synthesis by reverse transcriptase, p. 119-133. *In* A. M. Skalka and S. P. Goff (ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
5. **Clavel, F., and J. M. Orenstein.** 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. J. Virol. **64:**5230-5234.
6. **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:**3407-3414.
7. **Dieffenbach, C. W., and G. S. Dveksler.**(ed.). 1995. PCR Primer, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
8. **Fu, W., R. J. Gorelick, and A. Rein.** 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. J. Virol. **68:**5013-5018.
9. **Haddrick, M., A. L. Lear, A. J. Cann, and S. Heaphy.** 1996. Evidence that a kissing loop structure facilitates genomic RNA dimerization in HIV-1. J. Mol. Biol. **259:**58-68.
10. **Hahn, B. H., G. M. Shaw, S. K. Arya, M. Popovic, R. C. Gallo, and F. Wong-Staal.** 1984. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature **312:**166-169.

11. **Kim, S., R. Byrn, J. Groopman, and D. Baltimore.** 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J. Virol. **63:**3708-3713.

12. **Korber, B., C. Kuiken, B. Foley, B. Hahn, F. McCutchan, J. W. Mellors, and J. Sodroski,(ed).** 1998. Human retroviruses and AIDS. Los Alamos National Laboratory, Los Alamos, N.Mex.

13. **Laughrea, M., and L. Jetté.** 1994. A nineteen nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of the HIV-1 genome. Biochemisty **33:**13464-13475.

14. **Laughrea, M., and L. Jetté.** 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:**1589-1598.

15. **Laughrea, M., and L. Jetté.** 1996b. HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1*Lai* RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain. Biochemistry **35:**9366-9374.

16. **Laughrea, M., and L. Jetté.** 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:**9501-9508.

17. **Laughrea, M., L. Jetté, 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:**6003-6010.

18. **Laughrea, M., N. Shen, L. Jetté, 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:**226-234.

19. **Li, X., J. Mak, E. J. Arts, Z. Gu, L. Kleiman, M. A. Wainberg, and M. A. Parniak.** 1994. Effects of alterations of primer-binding site sequences on human immunodeficiency virus type 1 replication. J. Virol. **68:**6198-6206.

20. **Liang, C., X. Li, L. Rong, P. Inouye, U. 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:**5750-5757.

21. **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:**6629-6636.

22. **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:**6147-6151.

23. **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:**7014-7020.

24. **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:**2050-2058.

25. **Paillart, J.-C., L. Berthoux, M. Ottmann, J.-L. Darlix, R. Marquet, B. Ehresmann, and C. Ehresmann.** 1996. Dual role of the putative dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. J. Virol. **70:** 8348-8354.

26. **Quan, Y., L. Rong, C. Liang, and M. A. Wainberg.** 1999. Reverse transcriptase inhibitors can selectively block the synthesis of differently sized viral DNA transcripts in cells acutely infected with human immunodeficiency virus type 1. J. Virol. **73:**6700-6707.

27. **Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:**1350-1354.

28. **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:**3250-3254.

29. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual., 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

30. **Simon, J. H., and M. H. Malim.** 1996. The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes. J. Virol. **70:**5297-5305.

31. **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. USA **91:**4945-4949.

32. **Smith, C. M., J. S. Smith, and M. J. Roth.** 1999. RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription. J. Virol. **73:**6573-6581.

33. **Tanchou, V., D. Decimo, C. Péchoux, D. Lener, V. Rogemond, L. Berthoux, M. Ottmann, and J.-L. Darlix.** 1998. Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication. J. Virol. **72:**4442-4447.

34. **Telesnitsky, A., and S. P. Goff.** 1993. Strong-stop strand transfer during reverse transcription. p. 49-84. *In* A.M. Skalka, and S. P. Goff (ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

35. **Trono, D.** 1992. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. J. Virol. **66:**4893-4900.

36. **Varmus, H. E., S. Heasley, H.-J. Kung, H. Opperman, V. C. Smith, J. M. Bishop, and P. R. Shank.** 1978. Kinetics of synthesis, structure and purification

of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. J. Mol. Biol. **120:**55-82.

37. **von Schwedler, U., J. Song, C. Aiken, and D. Trono.** 1993. *vif* is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. J. Virol. **67:**4945-4955.

38. **Wu, T., J. Guo, J. Bess, L. E. Henderson, and J. G. Levin.** 1999. Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. J. Virol. **73:**4794-4805.

39. **Zhang, H., Y. Zhang, T. P. Spicer, L. Z. Abbott, M. Abbott, and B. J. Poiesz.** 1993. Reverse transcription takes place within extracellular HIV-1 virions: potential biological significance. AIDS Res. Hum. Retroviruses **9:**1287-1296.

40. **Zhang, H., O. Bagasra, M. Niikura, B. J. Poiesz, and R. J. Pomerantz.** 1994. Intravirion reverse transcripts in the peripheral blood plasma of human immunodeficiency virus type 1-infected individuals. J. Virol. **68:**7591-7597.

41. **Zhang, H., G. Dornadula, and R. J. Pomerantz.** 1996. Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenvironment: an important stage for viral infection of nondividing cells. J. Virol. **70:**2809-2824.

Preface (Chapter 3)

In the previous paper, we have showed that the klh (or DIS hairpin) played an important role in HIV-1 viral replication, genome dimerization, encapsidation, and reverse transcription. Stem-loop B played a similar role as that of DIS hairpin. However, contrary to klh (or DIS) hairpin, whose existence is supported by phylogenetic data, the existence of stem-loop B was putative. In this chapter, we provide evidence for the existence of stem-loop B.

Mutation in KLD resulted in 50% of reduction in genome dimerization. It raised the possibility that there are other regions in HIV-1 genome that might be involved in genome dimerization. In this chapter, we have also investigated the role of other part of 3' leader region in genome dimerization. Our data suggest that the whole 3' of leader region form a compact entity.

Chapter 3

Role of Stem B, Loop B, and Nucleotides next to the Primer Binding Site and the Kissing-Loop Domain in Human Immunodeficiency Vius Type 1 Replication and Genomic-RNA Dimerization

3. 1. ABSTRACT

Stem-loop B is a 12-nucleotide [nt]-long completely conserved sequence postulated to form a 4-bp stem and a 4-nt internal loop under the kissing-loop hairpin (klh) (nt 248 to 270) of human immunodefiency virus type 1 (HIV-1) genomic RNA. We investigated its role in viral replication, genomic RNA dimerization, and dimerization of partial HIV-1 RNA transcripts. The putative CUCG246-CGAG277 duplex was replaced by nine alternative complementary sequences, five likely to base-pair only in short RNAs and four likely to base-pair in long (~500-nt) RNAs, as assessed by the algorithm mfold. Among the five former sequences, none preserved genome dimerization and all reduced viral replication by 98 to 99.9%. Among the four latter sequences, three (MB6, -9 and -10) preserved genome dimerization, one (MB7) did not significantly inhibit it, and two (MB9 and -10) preserved viral replication. We conclude that duplex formation by stem B nucleotides is necessary for viral infectivity and complete genome dimerization. Deleting the 5' or 3' side of loop B or of stem B had little impact on dimerization of partial RNA transcripts and no impact on klh folding (and, for loop B mutations, on stem B folding), but each deletion inhibited genome dimerization almost as much as klh destruction. This suggests that loop B is required for complete genome dimerization and that loop B and stem B stimulate dimerization only in very long RNAs and/or in the presence of unidentified viral and cellular factors. Finally, we asked if nine deletions or nucleotide substitutions within nt 200 to 242 and/or nt 282 to 335 could influence genome dimerization. These mutations had intermediate inhibitory impacts consistent with their predicted influence on stem B, loop B and klh formation. Two exceptions were Δ200-226 and Δ236-242 genomic RNAs, which dimerized relatively poorly despite having neutral or positive influences on stem B, loop B, and klh folding.

3. 2. Introduction

A dimeric genome appears required for human immunodefinciency virus type 1 (HIV-1) replication (20). Stem-loop B (nucleotide [nt] 243 to 247 and 271 to 277 in HIV-1_{Lai} and HIV-1_{Hxb2} genomic RNA) is a completely conserved RNA sequence postulated to form a 4-bp stem and a 4-nt internal loop under the kissing-loop hairpin (or klh) (11, 13) (Fig. 1). Together with the klh, it forms the kissing-loop domain (KLD) (13), also called stem-loop 1 (SL1). The klh is crowned by a palindrome called the dimerization initiation site (DIS) (20, 21). The existence of the klh is supported by considerable phylogenetic evidence (11,13), and much information is available on its structure and function (e.g., see references 2, 4, 5, 7, 11-15, 17, 18, 20, 21). It is well established that the klh sequence stimulates genomic RNA dimerization (5, 7, 13, 20), genomic RNA encapsidation (2, 5, 13, 14, 17, 18, 20), and proviral DNA synthesis (18, 20). In contrast, little is known on the existence, structure and function of stem-loop B (4, 5, 14, 20, 22), despite the fact that it is more conserved than the klh (10, 13) and that deleting its 5' side inhibits viral replication as much as klh destruction (14). Since the existence of stem-loop B is neither proven nor universally accepted (Fig. 1 [see legend]), stem-loop diagrams appear where CUCG246 forms a loop (1), base-pairs to an upstream sequence (3, 6, 19), or base pairs to the tip of the AUG hairpin (8;) (Fig. 1). The first purpose of this paper is to test the stem-loop B hypothesis by investigating the role of stem B and loop B nucleotides in viral replication, genomic RNA dimerization, and in vitro dimerization of partial RNA transcripts. The second purpose is to ask if nine deletions or nucleotide substitutions near the KLD, namely, within nt 200 to 242 and/or nt 282 to 335 (which encompass residues stimulating reverse transcription, splicing, encapsidation and translation of viral RNA), could influence genome dimerization.

3. 3. Methods

Production and analysis of mutant viruses. COS-7 cells were transfected in parallel with wild-type plasmid pSVC21.BH10 and derived mutant plasmids, which are defined in Fig. 2. For PCR mutagenesis 5' of the klh, the PCR fragments synthesized extended from restriction site *NarI* to site *BssHII*; for mutagenesis 3' of the klh, they extended from *BssHII* to *SpeI*. Plasmid ΔP1 was altered by cassette mutagenesis (14) to construct C258GΔP1, in which C258 of the DIS was replaced by G and nt 301 to 319 (Fig. 1) were deleted. All cassette and PCR fragments introduced into pSVC21.BH10 and ΔP1 were completely sequenced to verify that the correct and no extraneous mutation had been introduced. Virus titer was determined by estimating the amount of capsid protein (CAp24) in purified viruses or in clarified COS-7 culture supernatant fluid (13, 14). Genomic RNA packaging was derived from densitometry of the Northern blots whenever the titer of the purified viruses was known. To assay infectivity, the number of 50% tissue culture infective doses contained in a mutant and a wild-type clarified COS-7 culture supernatant fluid was measured in two 96-well flat-bottom plates (15).

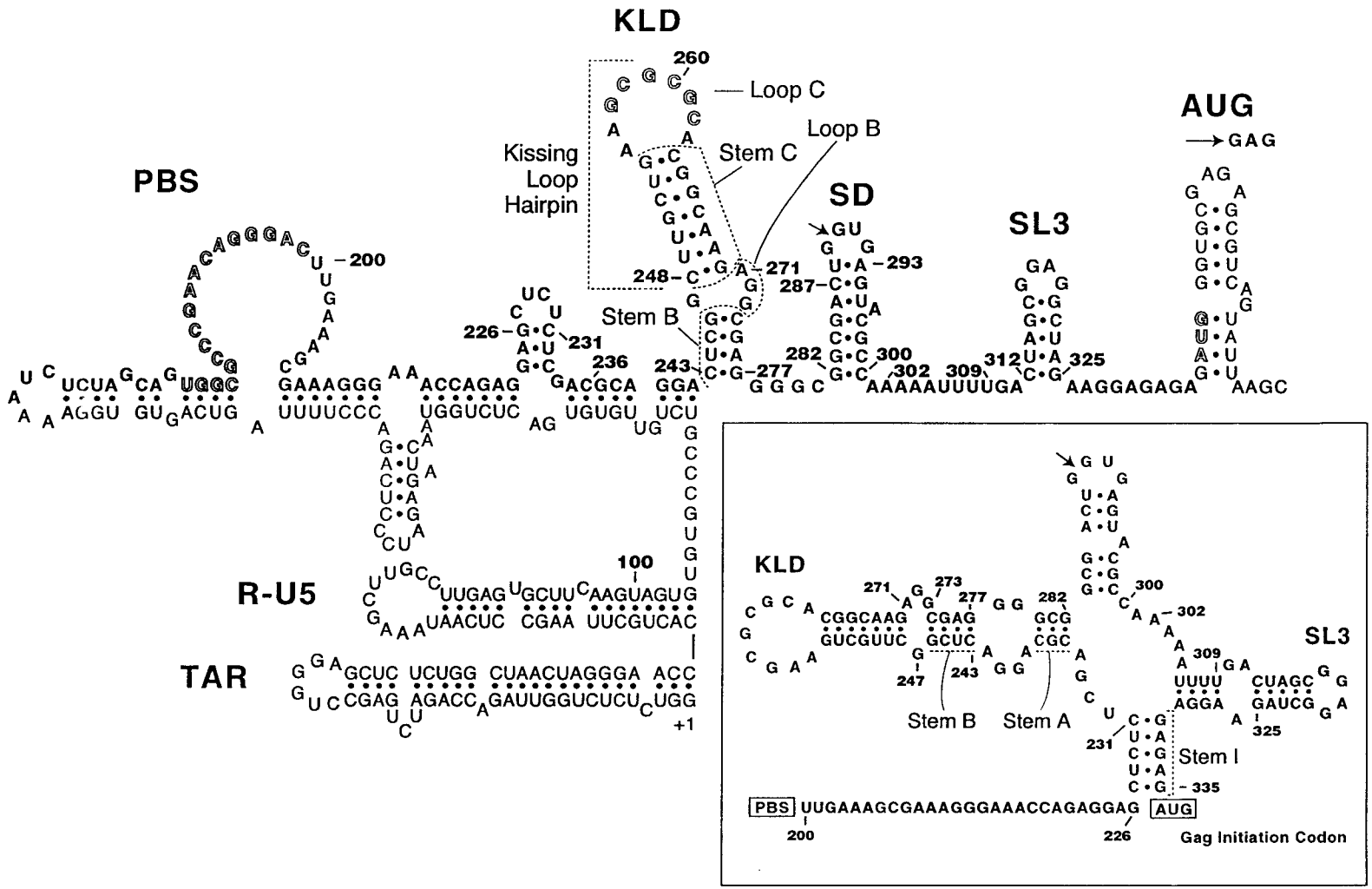

KLD
260
Loop C
Loop B
Kissing
Loop
Hairpin
Stem C
PBS
SD
SL3
AUG
GAG
Stem B
R-U5
TAR
100
+1
KLD
Stem B
Stem A
Stem I
SL3
PBS
UUGAAAGCGAAAGGGAAACCAGAGGAG
AUG
200
226
Gag Initiation Codon

Fig. 1. Postulated stem-loop diagram of the 5' untranslated region of HIV-1_{Hxb2} and HIV-1_{Lai} genomic RNA (3, 4, 6, 19). The cleavage site within the 5' major splice donor is marked by an arrow within the SD hairpin. The AUG initiation codon of the *gag* gene is highlighted and somewhat arbitrarily presented as hydrogen bonded in the interest of a compact figure. The PBS and the GCGCGC262 palindrome are also highlighted. Focusing on the L sequence, the model can be considered proven regarding SL3 and the klh (5, 17, 20). It is rather inconclusive regarding stem-loop B and the PBS hairpin; for example, there is little evidence for the formation of stem-loop B in transcripts starting near the physiological 5' end (1, 6, 8). A different model of the L sequence is shown in the insert (23). Note that stem-loop B and the klh can be theoretically formed from the sequence of any HIV-1 or simian immunodeficiency virus (strain cpz) (SIV_{cpz}) subtype (10, 13). In contrast, stem I is not found in subtype O and in SIV_{cpz}, stem-loop A is not found in subtypes D and O, the SD hairpin is not found in subtype N, and SL3 is not found in most SIV_{cpz} (10), indicating that these four separate folds are extensively but not as absolutely preserved as the stem-loop B and klh folds.

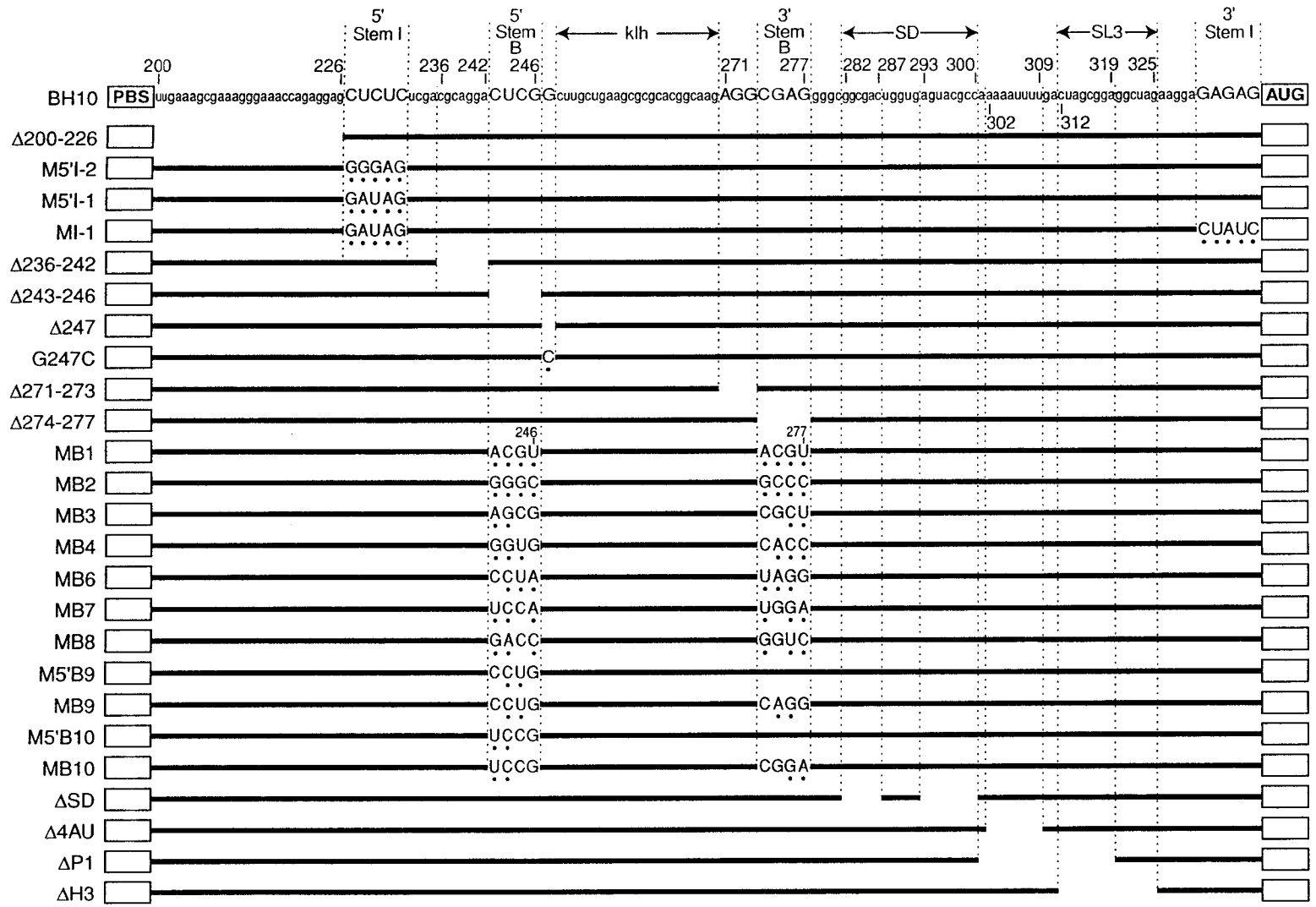
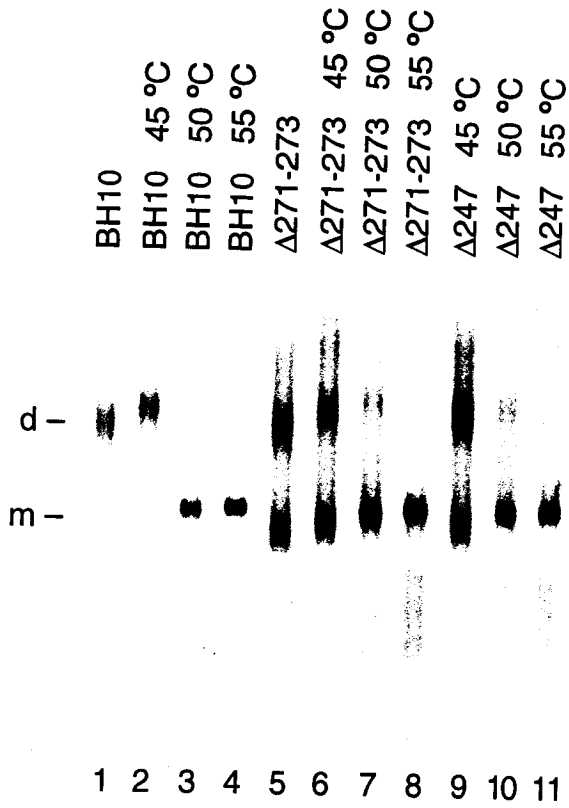

5' Stem I
5' Stem B
klh
3' Stem B
SD
SL3
3' Stem I
200
226
236 242
246
271
277
282 287 293 300
309
319 325
BH10
PBS
uugaaagcgaaagggaaaccagaggag
CUCUC
ucgacgcagga
CUCGG
cuugcugaagcgcgcacggcaag
AGGCGAG
gggcggcgacuggugaguacgccaaaaauuuugacuagcggaggcuagaaggaGAGAG
AUG
302
312
Δ200-226
M5'I-2
GGGAG
M5'I-1
GAUAG
MI-1
GAUAG
CUAUC
Δ236-242
Δ243-246
Δ247
G247C
C
Δ271-273
Δ274-277
246
277
MB1
ACGU
ACGU
MB2
GGGC
GCCC
MB3
AGCG
CGCU
MB4
GGUG
CACC
MB6
CCUA
UAGG
MB7
UCCA
UGGA
MB8
GACC
GGUC
M5'B9
CCUG
MB9
CCUG
CAGG
M5'B10
UCCG
MB10
UCCG
CGGA
ΔSD
Δ4AU
ΔP1
ΔH3

Fig. 2. Sequence of most mutations constructed for this study. PBS, $tRNA_{Lys3}$ (primer) binding site; AUG, initiation codon of the $Pr55^{gag}$ polyprotein. Numbering is in reference to the genomic RNA cap site. The secondary structure of stem I, stem B, the klh, the SD hairpin, and SL3 is shown in Fig. 1. ΔSD maintains a wild-type-consensus CUGGUGAR splice site sequence.

A


BH10
BH10 45 °C
BH10 50 °C
BH10 55 °C
Δ271-273
Δ271-273 45 °C
Δ271-273 50 °C
Δ271-273 55 °C
Δ247 45 °C
Δ247 50 °C
Δ247 55 °C
d –
m –
1 2 3 4 5 6 7 8 9 10 11

B

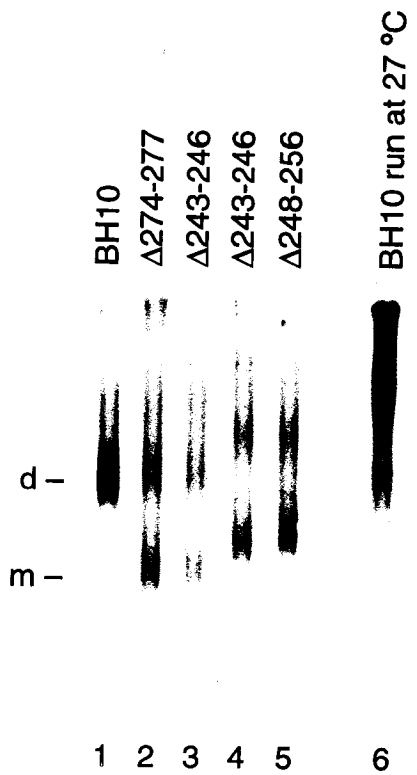
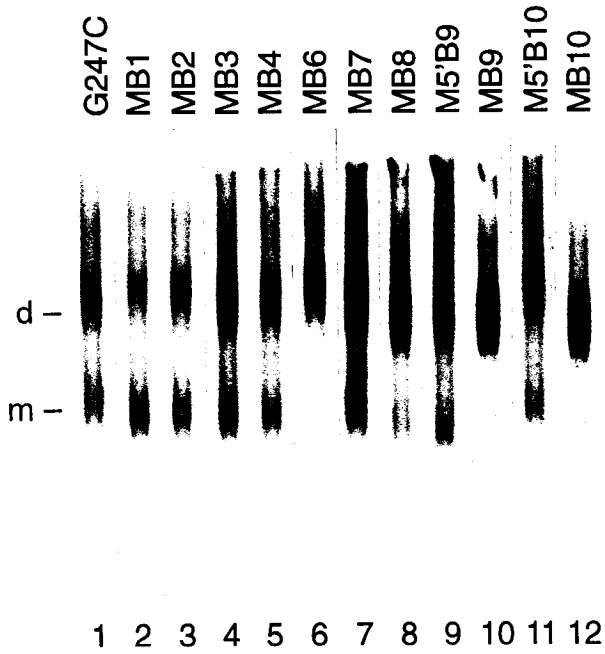

BH10
Δ274-277
Δ243-246
Δ243-246
Δ248-256
BH10 run at 27 °C
d –
m –
1 2 3 4 5 6

Fig. 3. (A) Dimerization level and thermal stability of the viral RNA isolated from BH10, Δ271-273, and Δ247 viruses, respectively, containing 1.5 x 10^{12}, 4.5 x 10^{12}, and 4.5 x 10^{12} CAp24. Samples were dissolved in 8 μl of buffer S (10 mM Tris [pH 7.5], 100 mM NaCl, 1 mM EDTA, and 1% sodium dodecyl sulfate) and incubated at the temperatures indicated for 10 min (samples in lanes 1 and 5 were left on ice for 10 min). After the incubations, all samples were loaded without delay and with the voltage on. After electrophoresis (70 V; 4 h; 1% agarose in TBE_2 [89 mM Tris, 89 mM Borate, and 2 mM EDTA] at 4 °C), the samples were Northern blotted, hybridized and autoradiographed for 4 h. (B) Lanes 1 to 5: dimerization level of viral RNA isolated from BH10, Δ274-277, Δ243-246, and Δ248-256 viruses respectively containing 3.6 x 10^{12}, 7.1 x 10^{12}, 7.1 x 10^{12}, 8.6 x 10^{12} , and 4 x 10^{12} CAp24; experimental conditions were unchanged except for 0.8 % agarose in lanes 1 to 3, and autoradiographic exposures of, respectively, 1h (lanes 1 to 3) and 2h 10 min (lanes 4 to 5). Lane 6: BH10 viral RNA electrophoresed (45 V, 3h 10 min, 1% agarose in TBE_2) in a room maintained at 27 °C, i.e., under conditions where only tight dimers of partial HIV-1 RNA transcripts remain dimeric. Each lane represents a different transfection. Densitometry of lanes 1, 5, and 9 of panel A, lanes 1 to 5 of panel B, and equivalent lanes from other Northern blots indicate that Δ271-273, Δ247, Δ274-277, Δ243-246, and Δ248-256 viruses package genomic RNA, respectively, 44% ± 4%, 65% ± 11%, 40% ± 6%, 25% ± 4%, and 40% ± 6% as well as BH10. Abbreviations: d, dimeric RNA; m, monomer.

A


G247C
MB1
MB2
MB3
MB4
MB6
MB7
MB8
M5'B9
MB9
M5'B10
MB10
d –
m –
1 2 3 4 5 6 7 8 9 10 11 12

B

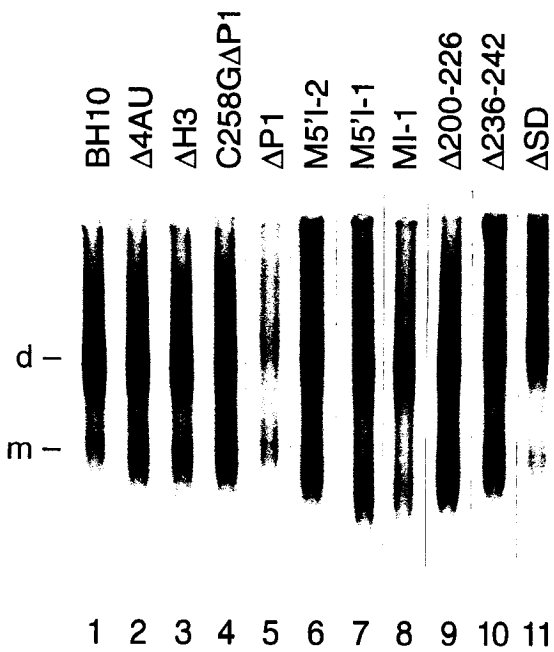
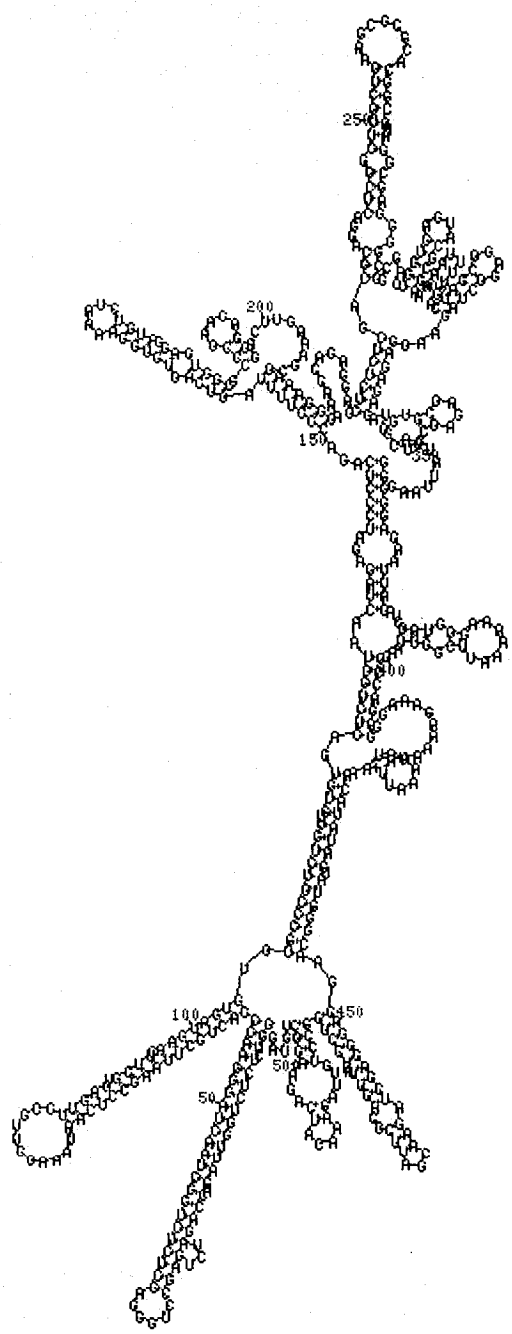

BH10
Δ4AU
ΔH3
C258GΔP1
ΔP1
M5'I-2
M5'I-1
MI-1
Δ200-226
Δ236-242
ΔSD
d –
m –
1 2 3 4 5 6 7 8 9 10 11

Figure 4. Dimerization level of viral RNA isolated from viruses purified from 10 petri dishes. Experimental conditions as in Fig. 3 except for 5 h electrophoresis. (A) G247C, MB1, MB2, MB3, MB4, MB6, MB7, MB8, M5'B9, MB9, M5'B10 and MB10 genomic RNA. (B) BH10, Δ4AU, ΔH3, C258GΔP1, ΔP1, M5'I-2, M5'I-1, MI-1, Δ200-226, Δ236-242, and ΔSD genomic RNAs.

plt22jpg by D. Stewart and M. Zuker



dG = -160.98 [initially -169.9] BH10 (nt1=508)

Fig. 5. An example of secondary structure of the first 508 nts of BH10 genomic RNA predicted by Mfold 3 (16, 24). The folding shows the formation of a KLD structure.

Table 1: Genomic RNA dimerization yield and infectious titers of 26 HIV-1 mutants[a]

Mutation location and sequence	% Dimerization C ± SD	Log BH10/mutant ratio ± SEM
Within KLD		
Δ247	54 ± 2.5	1.6 ± 0.45
Δ271-273	51.5 ± 3	2.0 ± 0.35
Δ243-246	48 ± 4.5	3.35 ± 0.20
Δ274-277	52.5 ± 2	3.5 ± 0.20
BH10	84 ± 3	0.0
G247C	62.5 ± 3	2.5 ± 0.15
MB1	38 ± 2.5	3.2 ± 0.15
MB2	55 ± 1.5	2.4 ± 0.1
MB3	61 ± 6	2.65 ± 0.25
MB4	64 ± 2.5	2.2 ± 0.25
MB6	85 ± 1	3.2 ± 0.15
MB7	73.5 ± 7.5	2.5 ± 0.2
MB8	70 ± 5	1.6 ± 0.1
MB9	85 ± 2	0.2 ± 0.25
MB10	86.5 ± 2	0.3 ± 0.2
M5'B9	69 ± 4	3.5 ± 0.2
M5'B10	63.5 ± 3	3.6 ± 0.1
5' or 3' of KLD		
Δ200-226	63 ± 3.5	3.5 ± 0.65
M5'I-1	64 ± 2	3.3 ± 0.65
MI-1	63 ± 5	3.1 ± 0.45
M5'I-2	67.5 ± 8.5	3.5 ± 0.3
Δ236-242	55 ± 3	>4.2
ΔSD	76 ± 2	>4.3
Δ4AU	73 ± 5.5	1.2 ± 0.15
ΔP1	56 ± 3.5	ND[b]
C258GΔP1	55.5 ± 2	ND
ΔH3	73 ± 2.5	1.5 ± 0.15

[a] In each infectivity test, the number of 50% tissue culture infective doses per ng of CAP24 was determined in octuplicate (14) by using a fourfold serial dilution against a control BH10 population. The number of independent infectivity tests per mutant was 2 to 4.
[b]ND, not done.

3. 4. Results and discussions

Impact on viral replication, genome dimerization, and genome packaging of deletions within postulated loop B or postulated stem B. COS-7 cells were transfected in parallel with equal amounts of plasmid pSVC21.BH10 and mutant plasmids pSVC21 Δ243-246, Δ247, Δ271-273, and Δ274-277 (Fig. 2). The infectious titers of the resulting viruses are presented in rows 1 to 4 of Table 1; loop B mutants were equally attenuated (deleting 1 nt of loop B was at least as damaging as inactivating the DIS [20]) and stem B mutants were also equally highly attenuated. This is most consistent with the existence of stem B.

Genomic RNA was extracted from the isolated viruses, electrophoresed on a nondenaturing agarose gel, and visualized by Northern blotting with a ^{35}S-labeled HIV-1 riboprobe specific for unspliced viral RNA (13,15). Densitometry of Fig. 3, as well as many other gel lanes from independent transfections, reveals that BH10, Δ247, Δ271-273, Δ243-246 and Δ274-277 genomic RNAs were, respectively, 84, 54, 51.5, 48, and 52.5% dimeric (Table 1), and that the dissociation temperatures of BH10, Δ271-273, and Δ247 genomic RNAs were similar (~49 to 50°C); deleting one side or the other of loop B, or deleting one side or the other of stem B thus inhibited genome dimerization almost as much as klh or KLD destruction (which results in 40 to 45% dimeric genomic RNA [13, 15, 20]; lane 5 of Fig. 3B]). This is most consistent with the existence of stem B at some point in the viral life cycle. Densitometry also indicates that Δ243-246 and Δ274-277 viruses encapsidated genomic RNA 2.5 to 4 times less efficiently than BH10 viruses (Fig. 3 [legend]). This is the first demonstration that mutating stem B inhibits genome encapsidation and the first demonstration that mutating loop B or stem B inhibits genome dimerization.

While klh mutations reduce the loose (4, 13, 14, 21) and tight (12,13) dimerization of partial HIV-1 RNAs by ≥90%, deletions Δ247, Δ271-273, Δ243-246, and Δ274-277 yielded, on average, ~ 0% inhibition of tight dimerization and an insignificant ~ 20% reduction of loose dimerization of partial transcripts ending at nt 295 or 508 (not shown). This suggests that loop B and stem B fully stimulate dimerization

only in RNAs longer than partial transcripts and/or in the presence of unidentified viral and cellular factors.

Base-pair-preserving mutations in postulated stem B preserve genome dimerization; mutations which change base pair number inhibit genome dimerization even when they preserve klh formation. Fig. 4A investigates the impact on genome dimerization of nucleotide substitutions in loop B (G247C; Fig. 2) or in stem B (MB1 to MB10, M5'B9, and M5'B10; Fig. 2). Nucleotide substitutions MB1 to MB10 preserve strand complementarity, and G247C makes possible a longer stem B and a smaller loop B. Densitometry of Fig. 4A and other gel lanes from at least two independent transfections reveals a range of dimerization yields; e.g., MB6, -9, and -10 did not affect genome dimerization, while MB1 was as inhibitory as klh or KLD destruction (Table 1).

We used the algorithm mfold (version 3.1) (16, 24) to fold the first 508 nt of BH10 and mutant genomic RNAs. All foldings minimizing ΔG or producing ΔGs within 5% of the predicted minimum—typically 16 per sequence—were accepted as equally probable (not shown). Mutating loop B or deleting the 5' or 3' side of stem B left klh folding intact: interference with klh formation is not a plausible explanation for the poor dimerization of Δ243-246, Δ247, G247C, Δ271-273, and Δ274-277 genomic RNAs (54% dimeric on average). MB6, MB7, MB9 and MB10 RNAs (82.5% dimeric on average) were about as likely as BH10 RNA to form a KLD; MB8 and M5'B9 RNAs (69.5% dimeric on average) had a small P (probability of forming a KLD) but a large P_{klh} (probability of forming a klh); MB2, MB3, MB4 and M5'B10 RNAs (61% dimeric on average) had a small P and a small but nonzero P_{klh}; and MB1 RNA (38% dimeric) had a P and a P_{klh} of 0. Thus, it appears that complete genome dimerization requires KLD formation: a klh unsupported by loop B and stem B typically loses its dimerization-potential, though it may in some mutants (two out of seven in our sample) keep half of it.

The G247C, Δ247, and Δ271-273 mutations left both klh and stem B folding intact (not shown; G247C stabilized stem B via addition of the C247-G273 base-pair). The poor genome dimerization associated with these mutations stress the importance of loop B in genome dimerization: loop B may orient the klh in a direction optimal for genome dimerization and/or interact with factors favoring dimerization. One such factor

could be the nucleocapsid (NC) protein, because it stimulates genome dimerization in isolated viruses (15) and a glutathione-S-transferase-NC fusion protein protects loop B from RNase T1 cleavage (6). Loop B may also favor dimerization by lowering the transition from loose to tight dimer (22). (Tight dimerization of HIV-1 RNAs ending at nt 295 or 508 was stimulated ~2-fold by deleting the 5' or 3' strand of stem B and inhibited ~3-fold by deleting the 3' side of loop B; Δ247 had neutral impacts [not shown]).

Two base pair-preserving stem B mutations (MB9 and MB10) preserve viral replication and genome dimerization. Mutants MB1 to MB8 were, per unit CAp24, 1.6 to 3.2 logarithmic units less active than BH10 (Table 1), indicating that mere strand complementarity did not preserve viral infectivity. But MB9 and MB10 replicated at near-wild-type levels, in contrast with M5'B9 and M5'B10 , which preserved neither genome dimerization nor viral replication (Table 1). However, MB6 and MB7 viruses were highly attenuated, suggesting that there may be no mechanistic overlap between how the stem B sequence controls genome dimerization and its other functions, which minimally include genome encapsidation (Fig. 3 [legend]) and reverse transcription (20). This notion is supported by our recent identification of klh mutations which inhibit genome dimerization without affecting genome encapsidation or reverse transcription (20), and of NC mutations which inhibit genome encapsidation without affecting genome dimerization (15). Note also that our 4 stem B or loop B deletions had identical effects on genome dimerization (Table 1) but variant effects on genome encapsidation (Fig. 3 [legend]).

Specifically inhibiting the dimerization function of the KLD reduces viral replication by 1.3 logarithmic units (20). Since mutations MB1 to MB8, M5'B9, and M5'B10 inhibited viral replication by ≥1.6 logarithmic units (2.7 on average), it follows that each one also inhibited functions other than genome dimerization, such as, we presume, proviral DNA synthesis and/or genome encapsidation.

Out of 255 alternative complementary stem B sequences tested by applying mfold to RNA 1 to 508, only 7 (e.g., MB6, -7 and -10) were as or more likely than BH10 to preserve KLD structure, and 210 (e.g., MB1, -2, -3, -4 and -8) had a *P* of 0 or close to 0 (not shown). This may be an important clue towards understanding why the stem-loop B sequence appears phylogenetically conserved. The simplest explanation for the results

with MB1 to MB10 is that the stem B helix exists at some point(s) in the viral cycle, that duplex formation by stem B nucleotides promotes genome dimerization, that few alternative complementary stem B sequences preserve stem B helix formation, and that still fewer preserve helix formation and viral replication.

Impact on genome dimerization and viral replication of mutations 5' and 3' of stem B. The primer binding site (PBS) and the *gag* gene are separated by 136 nt (the L sequence) which fold into stem B, loop B, the klh, and conceivably five other regions, such as nt 200 to 226, stem I, stem-loop A, the splice donor (SD) hairpin, and AAAAUUUU-SL3 (Fig. 1). We constructed nine mutant viruses bearing various deletions or nucleotide substitutions in each of these five other regions (first five and last four mutants in Fig. 2). (The mfold-derived probability of stem I formation [P_I]) was 0.31 in MI-1, 0.38 in BH10, and 0 in M5'I-1 and M5'I-2 viruses [not shown]). Genomic RNAs from these mutant viruses are shown in Fig. 4B. Their dimerization yields (from densitometry of Fig. 4B and many other gel lanes from independent transfections) are shown in Table 1. Δ236-242, ΔP1, Δ200-226, MI-1, M5'I-1 and M5'I-2 genomic RNAs had intermediate yields (55 to 56% for ΔP1 and Δ236-242; ~ 64% for Δ200-226 and stem I mutants), while ΔSD, ΔH3 and Δ4AU had rather large yields (73 to 76%). Dimerization yields were consistent with the mfold-predicted impact on KLD and klh folding, except that Δ200-226, Δ236-242 and ΔP1 genomic RNAs were less dimeric than predicted. The Δ200-226 and Δ236-242 results invite further studies on the role of the PBS hairpin (Fig. 1) in genome dimerization. ΔP1 destroys SL3 (Fig. 2), a stem-loop as important as the KLD for genome encapsidation (9, 17). But SL3 destruction may only have a weak or indirect impact on genome dimerization, for two reasons: ΔH3 genomic RNA specifically lacks SL3 but is closer to BH10 than to ΔP1 in dimerization yield (Table 1); ΔP1 plus DIS inactivation had no additive effects (C258GΔP1, C258G, and ΔP1 genomic RNAs were, respectively, 55.5, 48, and 56% dimeric [Table 1] [20]). Similarly, though disrupting nt 200 to 226, nt 227 to 231, nt 236 to 242, stem B, loop B or the klh inhibited genome dimerization by 25 to 50% in each instance, the inhibitions were not additive because Δ200-256 genomic RNA was indistinguishable from Δ248-256 genomic RNA in dimerization yield (20).

Mutants Δ200-226, MI-1, M5'I-1, M5'I-2, Δ236-242, and ΔSD were, per unit CAp24, at least 3 logarithmic units less active than BH10, while ΔH3 and Δ4AU viruses were ~ 1.35 logarithmic units less active than BH10 (Table 1). Thus, nt 200 to 226, nt 227 to 231, or nt 236 to 242 were as crucial for viral replication as stem B (Table 1) or the klh (13, 14) and at least twice as crucial as SL3 and the AU-rich sequence, whose impact was smaller than that of loop B and comparable to that of the central nucleotides of the DIS (14, 20). Since MI-1 viruses were as attenuated as M5'I-1 and M5'I-2 viruses, despite preserving the consensus sequence for protein synthesis initiation and despite a P_1 close to that of BH10, it follows that a stem I helix need not form during viral replication, or it is not easily replaced.

3. 5. Acknowledgements

This work was supported by a grant (MT-12312) from the Medical Research Council of Canada to Michael Laughrea.

3. 6. REFERENCES

1. **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:**382-397.
2. **Berkhout, B., and J. L. B. Van Wamel.** 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. J. Virol. **70:**6723-6732.
3. **Berkhout, B., and J. L. B. Van Wamel.** 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. RNA **6:**282-295.
4. **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:**5902-5908.
5. **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:**3407-3414.
6. **Damgaard, C. K., H. Dyhr-Mikkelsen, and J. Kjems.** 1998. Mapping the RNA binding sites for human immunodeficiency virus type 1 Gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. Nucleic Acids Res. **26:**3667-3676.
7. **Haddrick, M., A. L. Lear, A. J. Cann, and S. Heaphy.** 1996. Evidence that a Kissing Loop Structure facilitates genomic RNA dimerization in HIV-1. J. Mol. Biol. **259:**58-68.
8. **Hayashi, T., Y. Ueno, and T. Okamoto.** 1993. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. FEBS Lett. **327:**213-218.
9. **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 6198:**336-340.**
10. **Kuiken, C., B. Foley, B. Hahn, P. Marx, F. McCutchan, J. W. Mellors, J. Mullins, S. Wolinsky, and B. Korber (ed.).** 1999. Human retroviruses and AIDS. Los Alamos National Laboratory, Los Alamos, N.Mex.

11. **Laughrea, M., and L. Jetté.** 1994. A nineteen nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of the HIV-1 genome. Biochemistry **33:**13464-13475.

12. **Laughrea, M., and L. Jetté.** 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:** 1589-1598.

13. **Laughrea, M., L. Jetté, 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:**6003-6010.

14. **Laughrea, M., N. Shen, L. Jetté, 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:**226-234.

15. **Laughrea, M., N. Shen, L. Jetté, 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:**109-116.

16. **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:**911-940.

17. **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:**2050-2058.

18. **Paillart, J.-C., L. Berthoux, M. Ottmann, J.-L. Darlix, R. Marquet, B. Ehresmann, and C. Ehresmann**. 1996. Dual role of the putative dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. J. Virol. **70:**8348-8354.

19. **Rizvi, T. A., and A. T. Panganiban.** 1993. Simian immunodeficiency virus RNA is enficiently encapsidated by human immunodificiency virus type 1 particles. J. Virol. **67:**2681-2688.

20. **Shen, N., L. Jetté, 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:**5729-5735.

21. **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. USA **91:**4945-4949.

22. **Takahashi, K.I., S. Baba, P. Chattopadhyay, Y. Koyanagi, N. Yamamoto, H. Takaku, and G. Kawai.** 2000. Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. RNA **6:**96-102.

23. **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:**877-893.

24 **Zuker, M., D. H. Matthews, and D. H. Turner.** 1999. Algorithms and Thermodynamics for RNA secondary structure prediction: a practical guide. p.11-43 *In* J. Barciszewski, and B. F. C. Clark, (ed.), RNA biochemistry and bio/technology, NATO ASI Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Preface (Chapter 4)

Long term culture of KLD mutants results in mutations in the matrix, capsid, p2, and nucleocapsid genes that suppress several functional defects caused by KLD destruction. In this chapter, we raised the question as whether the four suppressor mutations would restore the functional defects caused by 1) changing the wild type palindrome to a weak one; 2) changing it into a non-palindrome; and 3) deletion of 5' of KLD. We found that the four suppressor mutations restore genomic RNA dimerization and viral replication in viruses bearing a weak palindrome, not in viruses bearing a non-palindrome.

Chapter 4

Intergenic Suppressor Mutations in the Human Immunodeficiency Virus Gag Region Restore Genomic RNA Dimerization and Viral Replication in Viruses Bearing a G/C-poor DIS, but Not in Viruses Bearing a Denatured DIS

4. 1. ABSTRACT

The kissing-loop domain (KLD) (nucleotides [nt] 243 to 277 of the human immunodeficiency virus type 1 [HIV-1] genomic RNA) is a stem-loop crowned by a GCGCGC262 (or GUGCAC) palindrome called dimerization initiation site (DIS). Nt substitutions in the DIS specifically impair HIV-1 genome dimerization while KLD destruction inhibits genome dimerization and unrelated functions. Mutations within the matrix, capsid, p2 and nucleocapsid genes suppress several functional defects caused by KLD destruction, but it is unknown if they can restore infectiousness or genomic RNA dimerization. Here we test the effect of these suppressor mutations on genome dimerization and infectiousness of viruses bearing moderate to crippling KLD mutations, namely: 1) mutation of the DIS into a weak GUUAAC palindrome; 2) mutation of the DIS into GGGCGC, a nonpalindrome; 3) deletion of the 5' strand of the KLD (each mutation reduces dimerization by 40 to 50% and replication rate by, respectively, ~1.3, ~1.3, and ~4 logs). In mutant 1, the suppressor tetrad restored genome dimerization and infectiousness to wild-type levels; in mutants 2 and 3, dimerization yield was not improved and infectiousness was increased to, respectively, ~0.4 and ~1.3 logs below wild-type, suggesting that ~70% of lost function was recovered. We conclude that the four suppressor mutations cannot suppress the replicative and genome dimerization defects caused by KLD destruction or even DIS denaturation but can suppress all functional defects caused by a G/C-poor DIS. This, together with other results, suggest that protein mutations, probably at the nucleocapsid level, can restore genomic RNA dimerization when DIS is weakened, but not when DIS is denatured or dimerization is inhibited by deleting nts immediately 3' of the primer binding site.

4. 2. Introduction

The HIV-1 genome is formed of two identical RNAs noncovalently linked *via* the kissing-loop domain (KLD), a complex stem-loop whose every part (stem B, loop B, stem C, and loop C [Fig. 1]) is essential for wild-type-like genome dimerization (1, 2). Dimerization of this diploid genome appears essential for HIV-1 function (1). However, since destroying the KLD reduces genome dimerization by 50% (1, 3, 4, 5), it is unclear whether the KLD is the dimerization initiation site of genomic RNA. The role of the KLD stem in genome dimerization includes, but is not limited to (2), endowing loop C with a shape optimal for initiation of KLD dimerization. The KLD dimerization initiation site (DIS) is a 6-nt palindrome (1, 2, 6, 7), located in the middle of loop C (Fig. 1). Its sequence is GCGCGC262 or GUGCAC262 in all HIV-1 and SIV_{cpz} clades. Studying two HIV-1 mutants, respectively, called UUAA and C258G, we have found that replacing CGCG261 by UUAA261 (replacing the DIS by a weak palindrome) impairs genome dimerization as much as destroying the KLD or replacing C258 by a G (replacing the DIS by a non-palindrome) 1 (Fig. 1). Destroying the KLD *via* deleting nt 241 to 256 (Δ241-256) or 264 to 277 (Δ264-277) reduces viral replication by about 4 logarithmic units (1, 8) (Fig. 1).

Suppressor mutations are second-site mutations which can restore a function, bring back a % of it, or promote an unrelated function that substitutes or partially substitutes for the missing one. Suppressor mutations MP2-MNC (corresponding at the protein level to Thr12Ile in the p2 polypeptide of p55gag polyprotein and Thr24Ile in the nucleocapsid protein (NC) of p55gag, and obtained after cultivating Δ241-256 viruses in MT-2 cells for 3 months) do not reverse the genome dimerization defect of Δ241-256 viruses (1), also called LD3 viruses, but correct their genome packaging defect and boost infectiousness to ~1.3 logs below BH10 (1, 8, 9). Interestingly, C258G and UUAA viruses, believed to be solely deficient in genomic RNA dimerization (1), are also ~1.3 logs less active than BH10. Cultivating Δ241-256 viruses for 4.5 months in MT-2 cells yielded suppressor mutations MA1-CA1-MP2-MNC, corresponding to Val35Ile, Ile91Thr, Thr12Ile and Thr24Ile in the matrix (MA), capsid (CA), p2 and NC proteins of p55gag (8)

Here we ask if poor genomic RNA dimerization can be reversed (or compensated for) by intergenic suppressor mutations in protein-coding genes. Notably, we ask if mutations MA1-CA1-MP2-MNC can reverse the poor genome dimerization and infectiousness of UUAA, C258G and Δ241-256 viruses or otherwise compensate for the dimerization defect. We also ask if mutations MP2-MNC can reverse the poor genome dimerization found in Δ200-226 viruses, i.e., viruses bearing the deletion of nt 200 to 226 (2), immediately 3' of the primer binding site (Long-term growth of Δ200-226 viruses in MT-2 cells yielded suppressor mutations, two of which, MP2 and MNC, were in the *gag* gene [10]).

4. 3. MATERIALS AND METHODS

COS-7 cells were transfected with equal amounts of wild-type plasmid pSVC21BH10 and mutant plasmids pSVC21C258G, UUAA, Δ241-256 (LD3), Δ264-277 (LD4), C258G-MA1-CA1-MP2-MNC, UUAA-MA1-CA1-MP2-MNC, LD3-MA1-CA1-MP2-MNC, LD3-MA1-MP2-MNC, LD3-CA1-MP2-MNC, LD4-MA1-MP2-MNC, LD4-MA2-MP2-MNC, Δ200-226, Δ200-226-MP2-MNC and BH10-MA1-CA1-MP2-MNC (1, 2, 5, 8, 13), Δ241-256, Δ264-277, and Δ200-226 mean that genomic RNA nt 241 to 256, 264 to 277 and 200 to 226 were respectively deleted. To construct pSVC21C258G-MA1-CA1-MP2-MNC and UUAA-MA1-CA1-MP2-MNC, pSVC21LD3-MA1-CA1-MP2-MNC was altered by PCR mutagenesis: the PCR fragments synthesized extended from restriction site *NarI* to site *SpeI* (2). Plasmid pSVC21Δ200-226-MP2-MNC was constructed by replacing the *BssHII* to *SalI* fragment of pSVC21Δ200-226 by the corresponding fragment of pSVC21LD3-MP2-MNC; pSVC21BH10-MA1-CA1-MP2-MNC was constructed by replacing the *NarI* to *BssHII* fragment of LD3-MA1-CA1-MP2-MNC by the corresponding fragment of BH10. To investigate genome dimerization, genomic RNA was extracted from the isolated viruses, electrophoresed on a nondenaturing agarose gel, and detected by Northern blotting with a ^{35}S-labelled HIV-1 probe (5). Viral infectiousness per ng of capsid protein was assayed in MT-2 cells, using the 50% tissue culture infective dose method ($TCID_{50}$) (4,5).

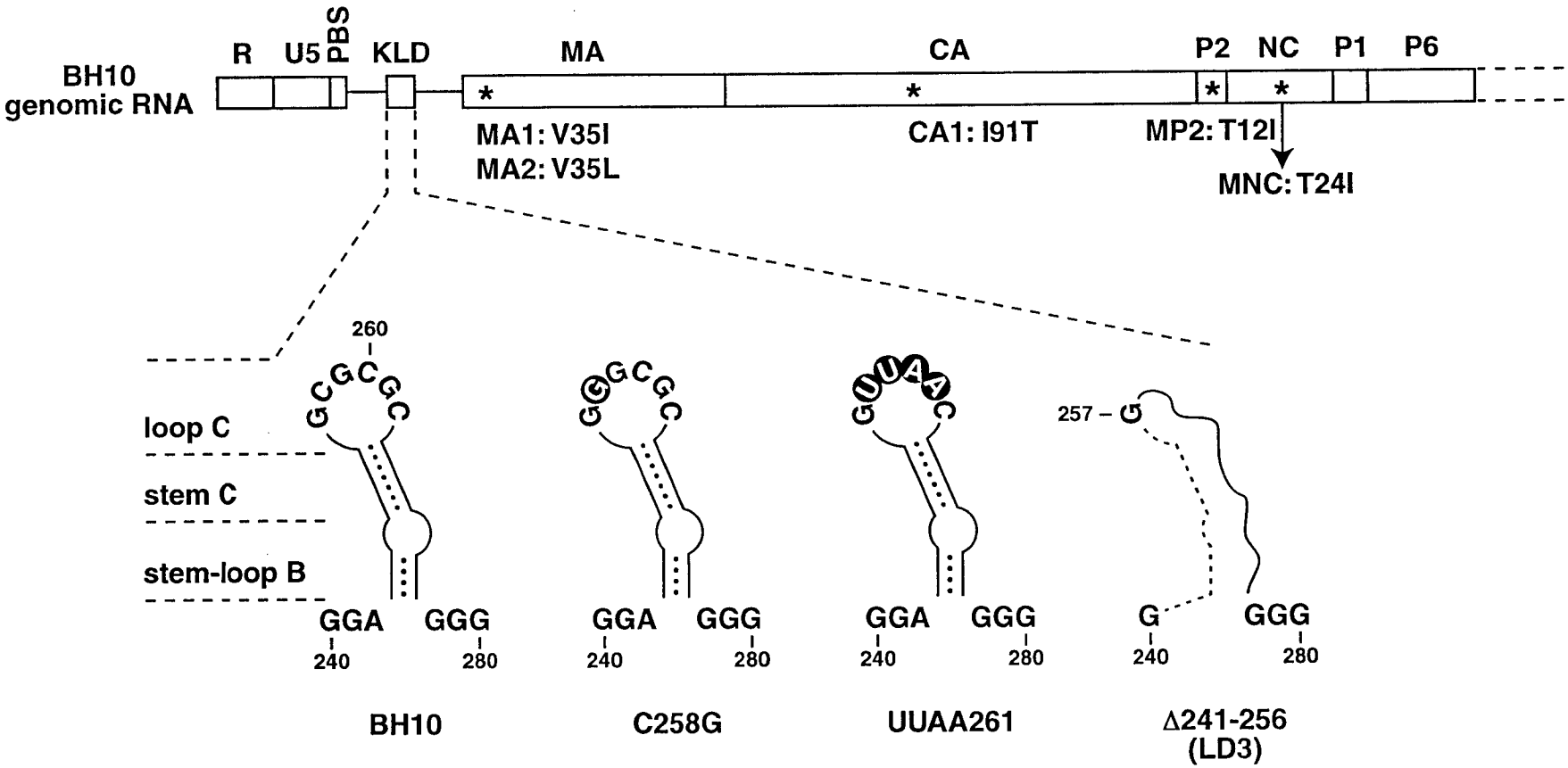

BH10
genomic RNA
R
U5
PBS
KLD
MA
CA
P2
NC
P1
P6
MA1: V35I
MA2: V35L
CA1: I91T
MP2: T12I
MNC: T24I
260
loop C
stem C
stem-loop B
GGA
GGG
240
280
257 – G
BH10
C258G
UUAA261
Δ241-256
(LD3)

FIG. 1. Schematic representation of the kissing-loop domain (KLD) and suppressor mutations studied in this paper. Not indicated is the LD4 (or Δ264-277) mutation, which destroys the 3' strand of the KLD stem. PBS, primer binding site.

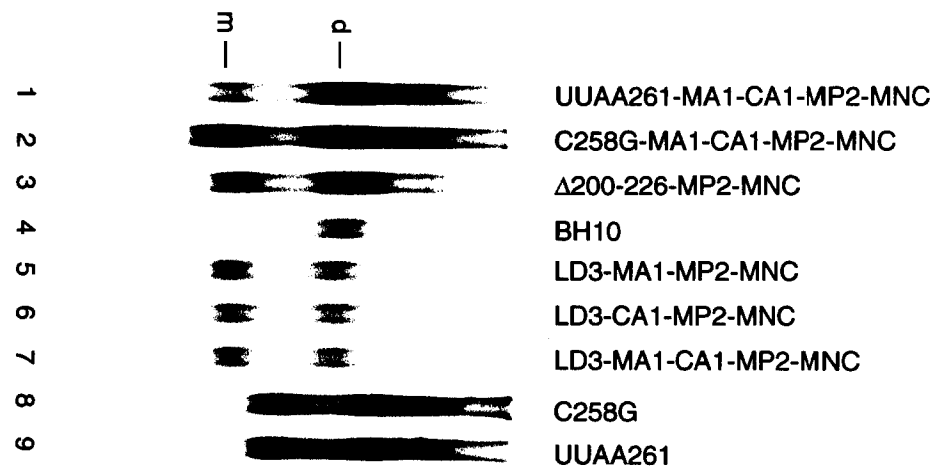

m
d
1
2
3
4
5
6
7
8
9
UUAA261-MA1-CA1-MP2-MNC
C258G-MA1-CA1-MP2-MNC
Δ200-226-MP2-MNC
BH10
LD3-MA1-MP2-MNC
LD3-CA1-MP2-MNC
LD3-MA1-CA1-MP2-MNC
C258G
UUAA261

FIG. 2. Dimerization level of viral RNA isolated from BH10 and mutant viruses purified from three (lanes 4 to 7), five (lanes 8 ,9), and ten (lanes 1 to 3) petri dishes. Viral RNAs were dissolved in 8 μl of buffer S (10 mM Tris (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1% sodium dodecylsulfate), and subjected to electrophoresis (70V, 4 to 5 h, 1% agarose in buffer TBE_2 [Tris-Borate-EDTA] at 4°C). The samples were next Northern blotted, hybridized, and autoradiographed. d, dimer; m, monomer.

Table 1. Infectious titer and dimerization level of HIV-1 BH10 vs. those of mutant viruses[a]

Mutant	log [BH10/mutant ratio][b]	% dimerization of genomic RNA
1) UUAA-MA1-CA1-MP2-MNC:	0.05 ± 0.1 (4)	78 ± 3
2) UUAA:	1.35 ± 0.25[c]	55 ± 4[c,d]
3) C258G-MA1-CA1-MP2-MNC:	0.4 ± 0.2 (6)	52 ± 7
4) C258G:	1.45 ± 0.25[c]	48 ± 4[c,e]
5) LD3-MA1-MP2-MNC:	1.35 ± 0.25 (2)	47.5 ± 3
6) LD3-CA1-MP2-MNC:	1.45 ± 0.35 (2)	47.5 ± 4
7) LD3-MA1-CA1-MP2-MNC:	1.3 ± 0.15 (4)	48.5 ± 2.5
8) LD3-MP2-MNC:	1.35 ± 0.3[c]	43 ± 6[c]
9) Δ241-256 (LD3):	4.2 ± 0.3[c]	45 ± 4[c]
10) LD4-MA1-MP2-MNC:	1.4 ± 0.25 (2)	-
11) LD4-MA2-MP2-MNC:	1.2 ± 0.25 (4)	-
12) LD4-MP2-MNC:	1.00[f]	-
13) Δ264-277 (LD4):	> 2.6[f]	-
14) Δ200-226-MP2-MNC:	~1.0 ± 0.3[g]	71 ± 4
15) Δ200-226:	3.5 ± 0.65[e]	63 ± 3.5[e]
16) BH10-MA1-CA1-MP2-MNC:	-	84.5 ± 2.5
17) BH10:	0	84 ± 2[c,d,e]

[a]All genome dimerization and infectivity results, except for three infectivity tests[f,g] were obtained concurrently. Thus footnotes c, d and e refer to the present study, some results of which were published a little earlier: because scientific publishing requires short thematic papers, we had to split a seemless concurrent whole into

installments, and installments are best written in succession rather than simultaneously.

[b]In each but three[f,g] of the infectivity tests, the number of $TCID_{50}$ per ng capsid protein (CAp24) was determined in octuplicate by using a 4-fold serial dilution against a control BH10 population. The number of independent infecitivty tests/mutant is indicated between parentheses; LD3-MA1-CA1-MP2-MNC titer experiments were independetly performed by N.S. and L.J. with undistinguishable results. -: not done. The standard errors are shown for the TCIDs and the standard deviations for the dimerization yields.

[c]Shen et al. (1)

[d]Laughrea et al. (5)

[e]Shen et al. (2)

[f]Liang et al. (8) The number of $TCID_{50}$ per ml supernatant was measured to monitor viral infection.

[g]Liang et al. (10) Reverse transcriptase activity (RT) per ng capsid protein was measured to monitor viral replication; we have translated this activity into a $TCID_{50}$ titer based on previous comparisons between RT-derived and $TCID_{50}$-derived infectious titers of LD3-MP2-MNC and LD4-MP2-MNC viruses (1, 8, 9). Based on RT-assays alone, Δ200-226-MP2-MNC was similar to LD4-MP2-MNC (8), less infectious than LD4-MA2-MP2-MNC (8) and more infectious than LD3-MP2-MNC(1, 8, 9).

4. 4. RESULTS

Lanes 1, 2, 4 and 7 to 9 of Fig. 2 show that mutations MA1-CA1-MP2-MNC restored genomic RNA dimerization in UUAA viruses but not in Δ241-256 and C258G viruses. This is quantitated in Table 1, which pools the results of scanning Fig. 1 and many other gel lanes representing two independent transfections per mutant: the suppressor tetrad boosted the genomic RNA dimerization yield of UUAA viruses from 55% ± 4% to 78% ± 3% (BH10 and BH10-MA1-CA1-MP2-MNC were ~84% dimeric), while C258G, C258G-MA1-CA1-MP2-MNC, Δ241-256 and LD3-MA1-CA1-MP2-MNC genomic RNAs were ~48% dimeric. Table 1 also shows that the suppressor tetrad restored infectivity in UUAA viruses and brought back ~70% of lost function in C258G and Δ241-256 viruses: the infectiousness of C258G and Δ241-256 viruses was boosted to respectively ~0.4 and ~1.3 logarithmic units below BH10. Since LD3-MA1-CA1-MNC-MP2 viruses were 95% less active than wild type, their infectivity was far from restored. This is contrary to past impressions based on measurements of infectivity per ml supernatant (8).

Moreover, lines 6 to 8 of Table 1 show that, contrary to previous impressions (8), MA1 and CA1 did not significantly improve suppressor activity. Confirming our observation that LD3-MP2-MNC and LD3-MA1-CA1-MP2-MNC viruses were phenotypically similar, i.e., defective in genome dimerization and 95% less infectious than BH10 (Table 1), we found that LD3-MA1-MP2-MNC and LD3-CA1-MP2-MNC viruses were identical to LD3-MP2-MNC in terms of poor infectiousness and poor genome dimerization (Table 1 and lanes 5-6 of Fig. 2), and that LD4-MA1-MP2-MNC and LD4-MA2-MP2-MNC viruses were as poorly infectious as LD3-MP2-MNC (lines 10-12 of Table 1), which is again contrary to previous impressions based on measurements of infectivity per ml supernatant (8) (Long-term growth of Δ264-267 viruses, also called LD4 viruses, in MT-2 cells yielded suppressor mutations MA2-MP2-MNC [8][Fig. 1]). Consistent with their poor infectiousness, it seems certain that LD4-MA1-MP2-MNC and LD4-MA2-MP2-MNC viruses are defective in genome dimerization.

Finally, lines 14 to 15 of Table 1 and lane 3 of Fig. 2 show that suppressor mutations MP2 and MNC had at best a modest effect on genome dimerization of Δ200-226 viruses, increasing dimerization yield by 13% ± 8.5% (for comparison, the suppressor tetrad increased the genome dimerization yield of UUAA viruses by 42% ± 5%).

4. 5. DISCUSSION

The simplest interpretation of our data is as follows. When genomic RNA kissing potential, i.e., the affinity of a KLD for another KLD, is weakened but not crippled (as in UUAA viruses), the suppressor tetrad can boost it sufficiently to restore genomic RNA dimerization; when kissing potential is crippled (DIS no longer palindromic [C258G] or KLD stem disrupted [Δ241-256]), the boost provided by the tetrad partially substitutes for the DIS or KLD-dependent functions (Table 1) but can't increase genome dimerization. If mutations in the central part of DIS affect only genome dimerization, then *gag* mutations can substitute for 70% of the function associated with KLD dimerization. We speculate that these suppressor mutations act predominantly at the protein level; this wouldn't be the first functional overlap between the RNA and protein worlds: ribozymes are RNAs with enzymatic functions, and protein synthesis elongation factor G-GDP is a protein which uses molecular mimicry to occupy a tRNA binding site (11).

Our data suggest that gag proteins cannot compensate for poor genome dimerization, when caused by KLD disruption (this is the simplest explanation for eight mutants of Table 1 converging at ~1.3 logarithmic units below BH10 in terms of infectivity), but that they can substantially compensate it, even reverse it, when the dimerization defect is caused by a modest sequence change in the DIS (see boost in C258G viral infectivity created by the suppressor tetrad). We conclude that mutations within four distinct gag proteins restore genome dimerization and viral replication in viruses bearing a G/C-poor DIS (e.g., UUAA viruses) but not in viruses bearing a denatured DIS (e.g., C258G, Δ241-256 or Δ264-277 viruses, where DIS has become nonpalindromic or has lost its dimerization-enhancing shape [12]).

Since the MA and CA suppressor mutations did not improve viral replication (lines 5-12 of Table 1), it is extremely likely that the MP2-MNC mutation is responsible for restoring genome dimerization and viral replication in UUAA viruses. Most HIV-1 NC proteins have an Ile24 in their N-terminal zinc finger: only HIV-1_{Hxb2}, the source of BH10 viruses, and two subtype D viruses have a Thr24 in their NC.[1] We had suggested that HIV-1_{Hxb2} might be slightly underendowed in strand annealing and genomic RNA

packaging, an underendowment which would become rate limiting in the context of a Δ241-256 deletion (1). We now add that this underendowment would also inhibit genome dimerization and viral replication when the DIS is G/C-poor. However MNC, even coupled with MP2, can't correct the poor genome dimerization associated with defects other than a G/C-poor DIS. For example, the dimerization defect caused by the deletion of nt 200 to 226, just 3' of the primer binding site, was not substantially corrected by the MNC-MP2 mutation (lines 14-15 of Table 1). Partial HIV-1 RNA transcripts bearing the GUUAAC DIS dimerize poorly (13), and incubating them with synthetic HIV-1_{Hxb2} NC does not stimulate dimerization (2). Might synthetic NC possessing an Ile at position 24 prove more stimulatory?

With regard to the inability of MA1, MA2 and CA1 to compensate for KLD destruction, note that MA1 reverses an Env incorporation defect created by point-mutations at Leu13 of MA (14) or a small deletion in the cytoplasmic tail of the transmembrane gp41 Env protein (15), it may be asking a bit much of the very subtle Val35Ile mutation (Fig. 1) that it should also suppress KLD mutations. Together with those of Freed and Martin (14) and Murakami and Freed (15), our results suggest that MA1 and the almost identical MA2 mutation may suppress Env and MA mutations more readily than KLD mutations.

What have we learned in this paper? (1) Four suppressor mutations in the *gag* gene can correct the dimerization defect created by a G/C-poor DIS, but cannot correct the dimerization defect created by a denatured DIS. (2) Since the MA1, MA2 and CA1 mutations do not significantly improve the suppressor activity of MP2-MNC, it appears that MP2-MNC reverses the poor genomic RNA dimerization associated with a G/C-poor DIS. (3) However MP2-MNC cannot reverse several kinds of dimerization defects: for example it cannot reverse the poor genomic RNA dimerization displayed by Δ200-226 viruses.

4. 6. ACKNOWLEDGEMENTS

This work was supported by a grant HOP-12312 from the Medical Research Council of Canada to M. Laughrea.

4. 7. REFERENCES

1. **Shen, N., L. Jetté , 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**:5729-5735.
2. **Shen, N., L. Jetté , M. A. Wainberg, B. P. Roques, and M. Laughrea.** 2001. Role of stem B, loop B and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization. J. Virol. **75**:10543-10549.
3. **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**:3407-3414.
4. **Laughrea, M., L. Jetté , 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. 71:6003-6010.
5. **Laughrea, M, N. Shen , L. Jetté, 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 280:109-116.
6. **Laughrea, M, and L. Jetté.** 1994. A nineteen nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of the HIV-1 genome. Biochemistry 33:13464-13475.
7. **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 USA 91:4945-4949.
8. **Liang, C., L. Rong, Y. Quan, M. Laughrea, L. Kleiman, 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. 73:7014-7020.

9. **Liang, C, L. Rong, M. Laughrea, L. Kleiman, 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. 72:6629-6636.

10. **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. 74:6251-6261.

11. **Nissen, P, M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B. F. C. Clark, and J. Nyborg:** 1995. Crystal structure of the ternary complex of Phe-$tRNA^{Phe}$, EF-Tu, and a GTP analog. Science 270:1464-1472.

12. **Laughrea, M, and L. Jetté:** 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 35:1589-1598.

13. **Laughrea, M, N. Shen, L. Jetté, 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 Biochemistry 38:226-234.

14. **Freed, E. O., and M. A. Martin.** 1996. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. J Virol. 70:341-351.

15. **Murakami, T., and E. O. Freed:** 2000. Genetic evidence for an interaction between human immunodeficiency virus type 1 matrix and alpha-helix 2 of the gp41 cytoplasmic tail. J Virol. 74:3548-3554.

Chapter 5. Discussion

5. 1. General discussion

In this thesis, we have studied the effect of the leader region (especially KLD or SL1 region) on HIV-1 viral replication, genome dimerization, encapsidation and reverse transcription. We have also studied the role of revertants in compensating for deficiencies in genome dimerization and viral replication. Our work clarifies the importance of the KLD in HIV-1 genome dimerization. Some unsolved questions are also being addressed below.

In vitro data have suggested that HIV-1 dimerization is initiated in a kissing-loop manner, i. e., by interstrand Watson-Crick base pairing between the autocomplementary sequence of KLD or SL1 from two adjacent RNAs (4, 6-10, 12-14, 19).

Prior to my Ph.D. work, there were controversial data concerning the *in vivo* role of KLD (or SL1) on genome dimerization. Some groups had seen a positive role of KLD on genome dimerization (3, 10), but some didn't (2, 15), possibly due to different experimental methods used.

The KLD consists of the kissing loop hairpin (klh, loop C plus stem C), plus putative stem B and loop B. In this thesis, we have shown that mutations in loop C, or in stem C of the KLD reduced genome dimerization by half (16), leaving the dissociation temperature unchanged. We also showed that by mutating the putative stem B and loop B, genome dimerization level was reduced to the same extent as by mutating the klh, indicating that stem B and loop B play an important role in genome dimerization (16). Our data also supported the notion that stem B and loop B exist (18). Up to then, the secondary structure of the stem-loop B sequence was very controversial (11).

In this thesis, we showed that a loop C mutant was defective in genomic RNA dimerization but wild-type-like in genomic RNA packaging, suggesting that genome dimerization is not a prerequisite for genome encapsidation.

The KLD plays important roles in genome dimerization, encapsidation, and reverse transcription. However, mutations in the KLD reduce viral infectivity much more than they reduce genome dimerization, packaging and reverse transcription. One

possibility might be that the method we used to measure infectivity exaggerated the infectivity defect (e. g., we did not specifically assess a single round infection). Another possibility is that KLD might affect other viral functions besides genome encapsidation, dimerization, and reverse transcription. It was shown that splicing was not affected by deleting the 5' of stem C plus two adenines 5' of the ACS (11). Kissing hairpin sequence was recently shown to promote recombination of the HIV-1 5' leader region (1).

The KLD mutations reduced genome dimerization by half, suggesting that there are other dimerization signals in HIV-1 genome. Efforts have been made to look for other dimerization sites. In the work presented in this thesis, we deleted the SL2 (or SD), SL3 (or H3) stem loop structure, and dimerization level was not changed in these mutants. We saw dimerization defect in mutating stem I, and in the mutant Δ200-226. Our data suggest that the 3' half of the leader region might form a compact entity. The possibility still remains that there are other possible regions involved *in vivo* genome dimerization.

Long term culture resulted in suppressor mutations that restored completely or partially the functional defects caused by KLD mutation. We have demonstrated that mutations within the matrix, capsid, P2 and nucleocapsid genes restored genome dimerization and viral replication in viruses bearing a G/C-poor DIS, but not in viruses bearing a denatured DIS (17). Our result suggests that proteins other than NC may be involved in HIV-1 genome dimerization. The matrix protein of RSV (Rous Sarcoma Virus) has been implicated in genome dimerization of RSV. The NCp7 protein of HIV-1 has been shown to be involved in dimer maturation (5). Mutations in matrix coding region has been shown to affect HIV-1 genome dimerization (20), although it is not clear whether the effect of the mutation is at the protein or RNA level.

More work is needed to find mutants that can reduce dimerization level to near zero percent. We need to identify these mutants before we can claim to have identified the DLS of HIV-1.

5. 2. References

1. **Balakrishnan, M., P. J. Fay, and R. A. Bambara.** 2001: The kissing hairpin sequence promotes recombination within the HIV-1 5' leader region. 276(39) Sept. 28, 36482-36492.
2. **Berkhout, B., and J. L. B. van Wamel.** 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. J. Virol. **70:**6723-6732.
3. **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:**3407-3414.
4. **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:**5902-5908
5. **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. Aca. Sci. USA **93:**7577-7581.
6. **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.
7. **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 type 1 genomic RNA. Biochemistry **33:**13464-13474.
8. **Laughrea, M., and L. Jette.** 1996a. 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:**1589-1598.
9. **Laughrea, M., and L. Jette.** 1996b. HIV-1 genomic dimerization: Formation kinetics and thermal stability of dimeric HIV-1 Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain. Biochemistry **35:**9366-9374.

10. **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. Biochem. **36:**9501-9508.

11. **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:**3397-3406.

12. **Muriaux, D., P. Fosse, and J. Paoletti .** 1996. A kissing complex together with a stable dimer is involved in the HIV-1 Lai RNA dimerization process in vitro. Biochemistry **35:**5075-5082.

13. **Muriaux, D., P. M. Girard, B. Bonnet-Mathoniere, and J. Paoletti.** 1995. Dimerization of HIV-1 Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. J. Biol. Chem. **270:**8209-8216.

14. **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:**27486-27493.

15. **Sakuragi, J. -I and A. T. Panganiban.** 1997. Human immuno-deficiency virus type 1 RNA outside the primary encapsidation and dimer linkage region affects RNA dimer stability. J. Virol. **71:**3250-3254.

16. **Shen N., L. Jetté, 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:**5729-5735.

17. **Shen, N., L. Jetté, C. Liang, M. A. Wainberg, and M. Laughrea.** 2002. Intergenic Suppressor mutations in the human immunodeficiency virus Gag region restore genomic RNA dimerization and viral replication in viruses bearing a G/C-poor DIS, but not in viruses bearing a denatured DIS. Sent for publication.

18. **Shen, N., L. Jetté, M. A. Wainberg, and M. Laughrea.** 2001. Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization. J. Virol. **75:**10543-10549.

19. **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. USA **91:**4945-4949.

20. **Dupont, S., N. Sharova, C. Dehoratius, C. –M. A. Virbasius, X. Zhu, A. G. Bukrinskaya, M. Stevenson, and M. R. Green.** 1999. A novel nuclear export activity in HIV-1 matrix protein required for viral replication. Nature **402:**681-685.

Chapter 6. Contribution to original knowledge

6. 1. Contribution to original knowledge

We are the first to provide a detailed analysis of the role of KLD in HIV-1 genome dimerization, viral replication, as well as to relate the genome dimerization defect with the extent of viral replication defect. We are the first to show evidence that genome dimerization and encapsidation can be dissociated. See Chapter 2 for details.

We are the first to perform a detailed analysis of the role of stem B and loop B sequence in genome dimerization, thereby lending support to the notion that stem B exists, at least at some point of the viral life cycle. We are the first to identify stem B mutants that preserved genome dimerization, viral infectivity, as well as stem formation. We are the first to study the role of SL2 (or SD), and SL3 (or H3) in HIV-1 genome dimerization. We are the first to study the nucleotides between the PBS and AUG start codon in genome dimerization. We found that Δ200-226 and stem I mutation affected dimerization. Taking all the results together, we suggested that the 3' leader region might act as integrity in genome dimerization. See Chapter 3 for details.

We are the first to study the effect of KLD revertants on genome dimerization. We found that four suppressor mutations in the Gag region restored the genomic RNA dimerization and viral replication in viruses bearing a G/C poor DIS, but not in viruses bearing a denatured DIS. Our results suggested that viral proteins could compensate for a dimerization-defective DLS. See Chapter 4 for details.

Chapter 6. Contribution to original knowledge

6. 1. Contribution to original knowledge

We are the first to provide a detailed analysis of the role of KLD in HIV-1 genome dimerization, viral replication, as well as to relate the genome dimerization defect with the extent of viral replication defect. We are the first to show evidence that genome dimerization and encapsidation can be dissociated. See Chapter 2 for details.

We are the first to perform a detailed analysis of the role of stem B and loop B sequence in genome dimerization, thereby lending support to the notion that stem B exists, at least at some point of the viral life cycle. We are the first to identify stem B mutants that preserved genome dimerization, viral infectivity, as well as stem formation. We are the first to study the role of SL2 (or SD), and SL3 (or H3) in HIV-1 genome dimerization. We are the first to study the nucleotides between the PBS and AUG start codon in genome dimerization. We found that Δ200-226 and stem I mutation affected dimerization. Taking all the results together, we suggested that the 3' leader region might act as integrity in genome dimerization. See Chapter 3 for details.

We are the first to study the effect of KLD revertants on genome dimerization. We found that four suppressor mutations in the Gag region restored the genomic RNA dimerization and viral replication in viruses bearing a G/C poor DIS, but not in viruses bearing a denatured DIS. Our results suggested that viral proteins could compensate for a dimerization-defective DLS. See Chapter 4 for details.