2008

STUDIES OF THE DIMERIZATION AND PACKAGING SIGNALS IN HIV-2 RNA

Tayyba Tabassum Baig
The University of Montana

STUDIES OF THE DIMERIZATION AND PACKAGING SIGNALS IN HIV-2 RNA

By

TAYYBA TABASSUM BAIG

B. Sc, Punjab University, Lahore, Pakistan, 1998
M. Sc, University of Agriculture, Faisalabad, Pakistan, 2000
M. Phil, University of Agriculture, Faisalabad, Pakistan, 2002

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Ph.D.
in Integrative Microbiology and Biochemistry

The University of Montana
Missoula, MT

Official Graduation Date *(*December *2008)*

Approved by:

Perry Brown, Associate Provost for Graduate Education, Graduate School

Dr. J. Stephen Lodmell, Research Advisor
Division of Biological Sciences

Dr. Michele McGuirl, Examination Chair
Division of Biological Sciences

Dr. Jean-Marc Lanchy, Research Co-Advisor
Division of Biological Sciences

Dr. Scott Samuels
Division of Biological Sciences

Dr. Mark Grimes
Division of Biological Sciences

Dr. J. B. Alexander Ross
Department of Chemistry and Biochemistry

ABSTRACT

Baig, Tayyba, Ph.D., Fall 2008 Integrative Microbiology and Biochemistry

Abstract Title: Studies of the dimerization and packaging signals in HIV-2 RNA

Chairperson: Dr. J. Stephen Lodmell

Co-Chairperson: Dr. Jean-Marc Lanchy

The investigation of sequences and structures in the 5' untranslated leader region (5'UTR) of HIV genomic RNA is essential for understanding viral replication because the 5'UTR regulates several essential functions by the alternate presentation and sequestration of signals through conformational changes. Our main focus in this study was to understand those sequences and structures that are involved in dimerization and packaging of HIV-2 RNA. Progressing from previous findings, we studied in detail a 10-nucleotide palindrome sequence (pal; 5'-GGAGUGCUCC-3') of the 5'UTR, located within the major packaging signal, upstream of the dimerization signal (SL1: stem loop-1). Pal has been shown to interact intramolecularly with SL1 to inhibit SL1-mediated HIV-2 leader RNA dimerization *in vitro*. We carried out three lines of experiments in an effort to understand the roles of pal in viral RNA dimerization and packaging.

The first study was achieved through randomization of pal and the subsequent selection from a population of random-pal RNAs those with enhanced or diminished dimerization properties *in vitro*. We showed that the 3'-pal motif (3'-pal; GCUCC-3') is involved in intramolecular interactions with a sequence downstream of SL1 that regulates SL1-mediated HIV-2 leader RNA dimerization. The second study was designed to investigate the role of 5'-pal motif (5'-GGAGU) in RNA packaging *in vivo*. Our findings indicated that the 5'-pal is essential for viral replication and genomic RNA packaging. Based on these findings, we proposed that 5'-pal is a binding element for the packaging proteins. Therefore, a third study was designed in which HIV-2 packaging proteins (Gag and NC) were expressed, purified, and assayed for binding with wild-type and mutant 5'-pal RNAs *in vitro*. These results suggested that the 5'-pal is a binding element for Gag protein. In summary, we have showed that pal is an important regulator of dimerization and packaging processes of HIV-2 RNA. We also demonstrated that pal is composed of two motifs with distinct functions. Overall, our study significantly contributes to the understanding of HIV-2 RNA dimerization and packaging, which may ultimately lead to the identification of novel antiretroviral targets.

ACKNOWLEDGEMENTS

My full appreciation and regards go to my research advisor, Dr. Stephen Lodmell, who is a cool-minded person and is always considered as a problem solver. Not only academic and lab matters, but also he was available to solve any personal matters. I found a very nice and unforgettable association with his mentorship. I appreciate his patience a lot that I was allowed to extract him from his busy schedule any time for discussions.

I extend my regards and honor to Dr. Jean-Marc Lanchy, who was always available to help and coordinate with me in all lab endeavors. He leads by example. He is an exemplary scientist. He taught me, how to take small steps one by one in designing and doing experiments.

I am grateful to my all committee members, who were very helpful to manage time for me from their busy schedules to arrange committee meetings. Dr. Michele McGuirl, being a female scientist was always a source of inspiration. I extend my heartiest thanks to her for helping and advising me in each and every aspect of academic and personal life. I am also thankful to her for allowing me to do a rotation project in her lab that provided me a ground for working with proteins. My extraordinary thanks go to Dr. Scott Samuels, Dr. Mark Grimes, and Dr. J. B. Alexander Ross, who were always available to help in any matter regarding academia.

I am extremely pleased to place on record my full appreciation and love to Dr. David Strobel, whom I consider as my spiritual father. During my stay at the UM, he played an exceptional role to fulfill my academic needs along with providing his best encouraging and loving behavior. I convey special thanks to Mary Kamensky and Effie Koehn for every support and facilitation.

I admire my friends, Maura Davenport, Debbie Weisser, Brittany Harris, Dalia Rokhsana, Ayesha Sharmin, Akriti Kothiwal, Chris Woolstenhulme, Leila Sears and Heather Cahoon, who provided me full friendly support and association during my stay in the United States.

In the end, my special thanks and love go to my uncle, Dr. Donald A Gatzke and my parents, M. Ashraf Baig and Kauser Tesneem, who made it possible for me to come abroad to fulfill my dreams. Their great association and every kind of support always cleared away all the difficulties of my life. I also acknowledge my sister, Sidra Baig, my brothers, Adnan Baig and Taimour Baig, who were always available for moral support and to provide all concerned facilities to fulfill my needs.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1. AIDS/HIV/Dimerization and packaging of HIV RNA

Since the discovery of AIDS, HIV infection has been diagnosed nearly all over the world, making AIDS one of the most devastating diseases in human history and has killed millions of people. In developed nations, AIDS is a controllable disease due to the accessibility of anti-HIV therapy, which inhibits viral replication by attacking key steps of the viral replication cycle thereby reducing viremia. The development of anti-HIV therapy became possible by obtaining a detailed understanding of essential steps of the HIV-1 replication cycle. However, the same essential steps of a second AIDS causing virus, HIV-2, have been less well studied and more research is warranted. Therefore, the impetus of the research presented in this dissertation was to understand in detail two crucial steps, RNA dimerization and packaging, of the HIV-2 replication cycle.

1.1.1. AIDS (acquired immunodeficiency syndrome)

In June 1981, the *Morbidity and Mortality Weekly* reported that an unusual form of *Pneumocystis carinii* pneumonia (PCP) causing immunosuppressive disease had been diagnosed among homosexual men living in Los Angeles, USA (Weekly, 1981). Further, the diagnosis of PCP with immunosuppressive disease in groups other than homosexuals at other places (Masur et al, 1981) announced the beginning of an epidemic that lead to a

major worldwide health dilemma (see review; Fauci, 2006). Because this disease is not genetically inherited and it suppresses the immune system, it was named AIDS (acquired immunodeficiency syndrome) (Marx, 1982).

The cause of AIDS was identified in 1983 in the lab of Dr. Luc Montagnier, who reported an association between the development of AIDS with the infection of a retrovirus (Barre-Sinoussi et al, 1983). The name of this AIDS retrovirus was given in 1986 as the HIV-1 (human immunodeficiency virus type 1) (Coffin et al, 1986). About that time, a second AIDS causing virus was isolated from West African patients and was called HIV-2 (human immunodeficiency virus type 2) (Clavel et al, 1986a). According to one estimate, over 50 million people have been infected with HIV. Of those, 33 million people are living and 25 million have died (see reviews; Fauci, 2006; Ho & Bieniasz, 2008).

1.1.2. HIV (human immunodeficiency virus)

1.1.2.1. HIV is a retrovirus

The Retroviridae comprise a family of RNA viruses whose genetic material is converted into double stranded proviral DNA in the host cell with the help of RT (reverse transcriptase) in a process called reverse transcription (Baltimore, 1970; Temin & Mizutani, 1970). Within the Retroviridae is a virus group called the lentiviruses (*lenti* is the Greek root for slow), which are characterized by producing slow progressing disease. Many of these viruses affect the cells of the immune system in animals and humans (see reviews; Shuljak, 2006; Sigurdsson, 1954). The detection of reverse

transcriptase activity in the cultured lymphocytes, obtained from AIDS patients revealed that the virus, which destroys the human immune system, is a retrovirus (Barre-Sinoussi et al, 1983). Due to the structural and sequential alignment with other lentiviruses, HIV was recognized as a lentivirus (Barre-Sinoussi et al, 1983; Gonda et al, 1985; Munn et al, 1985).

1.1.2.2. Origin and epidemiology of HIV

The origin of HIV became clearer with the finding of an evolutionary link between HIV and related retroviruses from African nonhuman primates called SIVs (simian immunodeficiency viruses) (see review; Hahn et al, 2000). HIV-1 seems to have been passed on to humans from SIVcpz (chimpanzee) (Gao et al, 1999). The HIV-1 strains present in infected human populations are classified into three groups. The first group is called the 'main group', group M, and this group is the major cause of infections in the world. The second group consists of highly divergent strains and is called the 'outlier group', group O, and the third group is called the 'non-M, non-O group', group N, and a small number of viruses belong to this group (se review; Sharp et al, 2001).

HIV-2 is closely related to the SIVs that exist in certain monkey populations. The HIV-2 seems to have been passed on to humans from SIVsmm (sooty mangabeys) (Chen et al, 1997; Gao et al, 1992). Phylogenetic analysis reveals that there are seven subtypes (A-G) of HIV-2, of which C, D, E, F, and G are rare. The subtype A is the major case of HIV-2 infections and is prevalent in Guinea-Bissau. Subtype B originated in West Africa and is occasionally found in Europe. Although HIV-2 has been mostly confined to West

Africa, the number of HIV-2 infections is increasing at a high rate in some other regions of the world including India and Europe (see review; Reeves & Doms, 2002).

1.1.2.3. Structure of HIV

The formation of HIV particles is started by the assembly of the Gag polyprotein into a shell beneath the membrane of the budding virions. After assembly and budding, the Gag polyprotein is processed into individual structural proteins, the MA (matrix), CA (capsid), p2, NC (nucleocapsid), p1 and p6 proteins by viral PR (protease). This processing converts the immature particles to mature infectious particles (Ganser-Pornillos et al, 2008; Henderson et al, 1992; Hoglund et al, 1992). The sizes of mature infectious particles range from 120-200 nm (average; ~145 nm) (Briggs et al, 2003) (Figure 1). Each particle contains an envelope with ~14 protruding spikes, corresponding to envelope glycoproteins (Zhu et al, 2006), the gp120 (glycoprotein 120) and gp41 (glycoprotein 41). These proteins are external and trans to the membrane respectively (Allan et al, 1985; Veronese et al, 1985). The viral envelope encircles the MA protein that lines the membrane and the core, which is present in the middle of the particle. The core is cone shaped and is composed of CA protein and NC-RNA complex (Briggs et al, 2003; Ganser et al, 1999; Gelderblom et al, 1987; Hoglund et al, 1992). The RNA is not a single molecule, two molecules exist within each particle (see reviews; Greatorex & Lever, 1998; Paillart et al, 2004b; Russell et al, 2004).

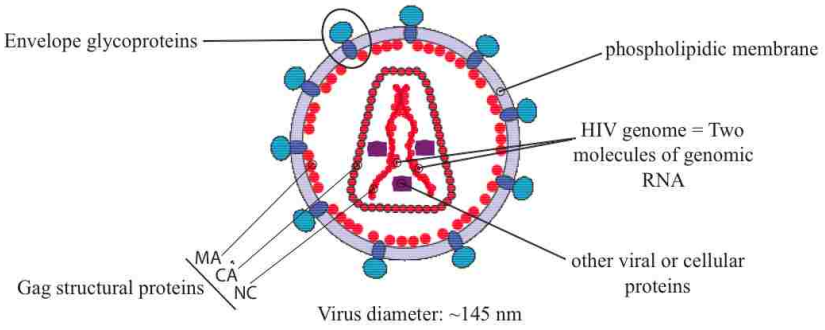


Figure 1. HIV mature virion particle.

1.1.2.4. Genome organization of HIV

The HIV RNA genome is capped and polyadenylated (Figure 2B). It is approximately 9.2 kb in length and encodes 15 different proteins (Frankel & Young, 1998; Wain-Hobson et al, 1985). Like other retroviruses, HIV genome contains three open reading frames: *gag, pol* and *env* (Ratner et al, 1985; Wain-Hobson et al, 1985) (Figure 2A).

The *gag* encodes the Gag (group specific antigen) precursor polyprotein, which is a structural protein and as described above, is cleaved by viral protease during maturation of virion particles yielding six proteins (MA, CA, p2, NC, p1 and p6) (Henderson et al, 1992). The functional roles of these proteins has been described in detail (see review; Frankel & Young, 1998), however, a few of those are illustrated here. The MA is involved in the stable association of Env glycoproteins to the viral particles (Dorfman et al, 1994). The NC binds with the full-length genomic RNA to package them into assembling virions (Aldovini & Young, 1990) and as described above, the CA makes the

shell of the central core (Ganser et al, 1999; Gelderblom et al, 1987). The p1 and p2 domains are required for the regulation of proteolytic cleavage of Gag for the production of mature infectious virions (Pettit et al, 1994; Wiegers et al, 1998). The p6 domain of Gag is also required for the production of infectious virions (Huang et al, 1995) as well as for the incorporation of an accessory protein, Vpr into viral particles during assembly process (Kondo & Gottlinger, 1996).

The *pol* encodes three enzymes, PR (protease), RT (reverse transcriptase) and IN (integrase), which are located at 5', middle and 3' sides of *pol* respectively (Muesing et al, 1985; Ratner et al, 1985). The PR performs autoprocessing and processing of polyproteins to yield the individual proteins (Debouck et al, 1987; Farmerie et al, 1987; Kramer et al, 1986). The RT is involved in retroviral specific reverse transcriptase (polymerase and RNase H) activity for conversion of RNA to DNA (di Marzo Veronese et al, 1986; Mizrahi et al, 1989) and the IN is involved in the integration process of HIV DNA (Bushman et al, 1990; Miller et al, 1997).

The *env* encodes the envelope precursor polyprotein that is cleaved into 2 subunits. One is external to the membrane called gp120 and the other is membrane associated called gp41 (Allan et al, 1985; Veronese et al, 1985; Willey et al, 1988). The gp120 mediates the binding of HIV virions to CD4 receptor (McDougal et al, 1986) and gp41 is responsible for the fusion between viral and cell membranes (Kowalski et al, 1987).

In addition to *gag, pol* and *env,* HIV contains an array of regulatory and accessory genes. These include two regulatory (*tat, rev*) and four accessory (*vif*, *vpr*, *vpu* and *nef*)

genes (see reviews; Cullen & Greene, 1990; Pavlakis & Felber, 1990) (Figure 2A). The *tat* encodes Tat (transcriptional activator) that is important for replication by mediating transactivation of transcription and regulation of gene expression. Tat performs these functions by binding to a structured region in RNA called TAR (transactive responsive) element (Dayton et al, 1986; Fisher et al, 1986; Garcia et al, 1989; Jakobovits et al, 1988). The *rev* encodes Rev (regulator of viral gene expression), which functions as a nuclear exporter to export viral RNA into the cytoplasm (Meyer & Malim, 1994).

The accessory (*vif*, *vpr*, *vpu* and *nef*) genes encode Vif (viral infectivity factor), Vpr (viral protein R), Vpu (viral protein U) and Nef (negative effector) proteins, which perform important functions during viral replication (see review; Frankel & Young, 1998). The Vif is involved in the import of pre-integration complex into the nucleus (see review; Frankel & Young, 1998) and is essential for the production of infectious virions by inhibiting the antiviral activity of a host factor, APOBEC3 (Mehle et al, 2004; Strebel et al, 1987). The Nef and Vpu are involved in the degradation of CD4 cells (Mangasarian & Trono, 1997; Tiganos et al, 1997; Willey et al, 1992).

The *gag*, *pol*, *env,* regulatory and accessory genes are flanked by R (repeat)-U5 (unique 5') regions at 5' UTR (untranslated leader region) and U3 (unique 3')-R regions at 3'UTR of genomic RNA respectively (Figure 2B). These regions are duplicated during reverse transcription and form U3-R-U5, termed the LTR (long terminal repeat) region at both 5' and 3' ends of proviral DNA (Starcich et al, 1985) (Figure 2A).

The HIV-1 5'UTR contains several conserved sequence and/or structure elements including TAR, poly A signal, PBS (primer binding site), SLI (stem loop-1; dimerization

signal), SD (splice donor site) and PSI (packaging signal) (Figure 3A). These elements are involved in the regulation of many essential steps of the viral replication cycle including transcription, polyadenylation, genomic RNA dimerization, reverse transcription, splicing, packaging and translation (Berkhout, 1996).

HIV-1 and HIV-2 viruses are less than 50% alike at the sequence level but contain a significant similarity at the genome organization level (Clavel et al, 1986b; Zagury et al, 1988). As in HIV-1, the HIV-2 genome also contains a 5' UTR, consisting of conserved structural elements (Figure 3B), important for several functions of the viral replication cycle (Berkhout, 1996).

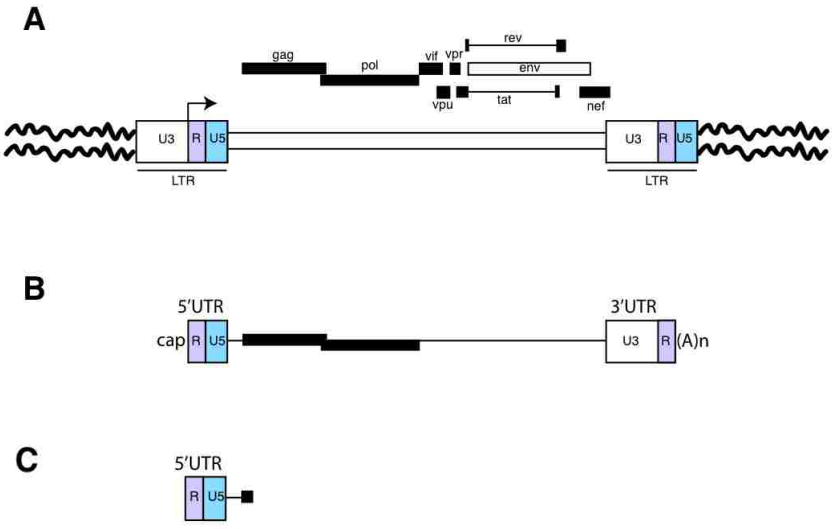


Figure 2. Organization of the HIV genome.

A. HIV proviral DNA. The arrow indicates the initiation site of transcription. (Top) Open reading frames for HIV structural, regulatory and accessory proteins. B. HIV RNA genome. C. The UTR (untranslated leader region) at the 5' end of RNA genome.

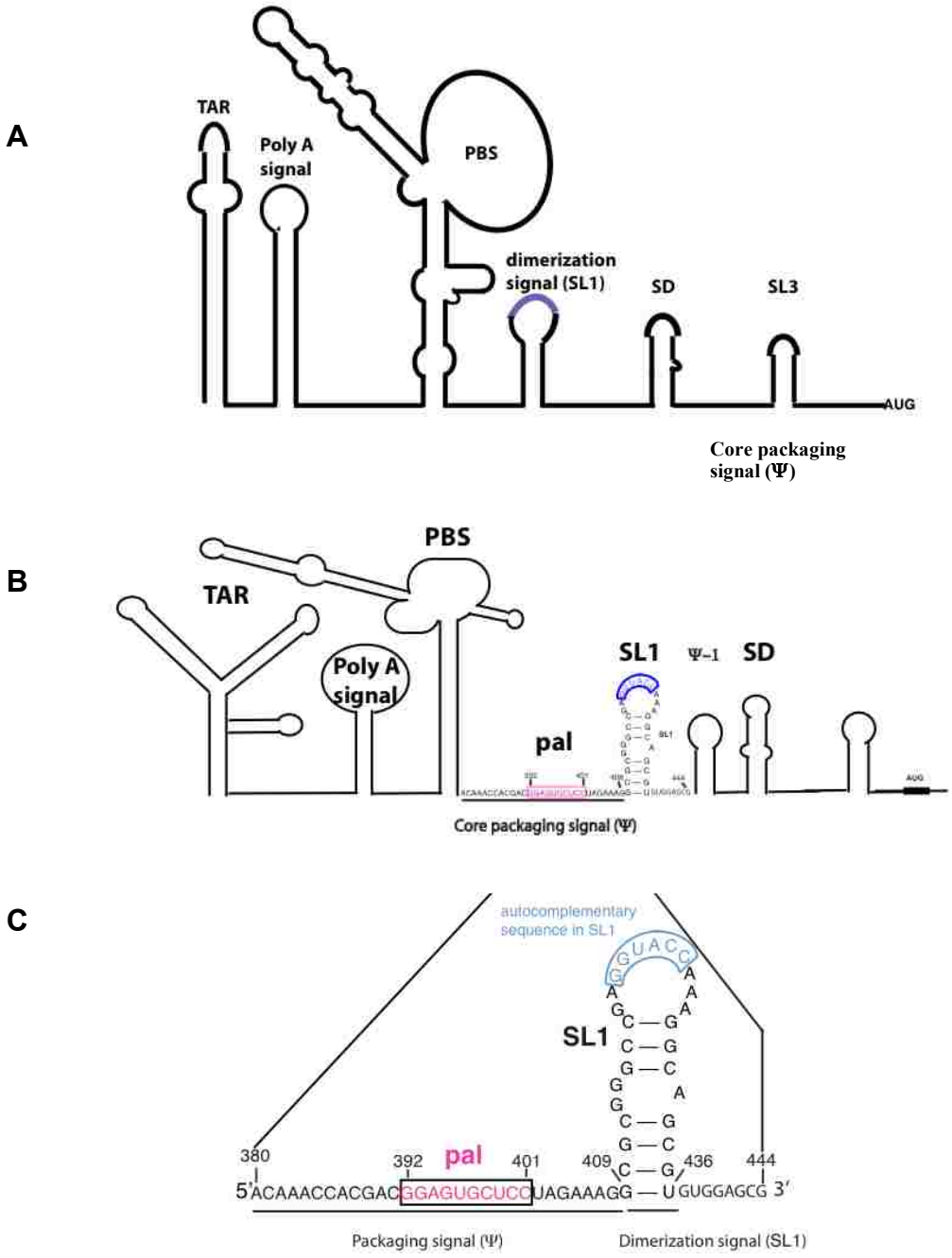


Figure 3. Model of the 5' UTR.

A. A model of the 5' UTR of HIV-1 RNA. B. A model of the 5' UTR of HIV-2 RNA. The structure motifs are represented by their putative functions on top and they include TAR (transactivation region), poly A (polyadenylation) signal, PBS (primer binding site), SL1 (stem-loop1; dimerization signal), SD (splice donor site) Ψ-core (core packaging signal), and AUG (gag initiation codon). C. The dimerization and the packaging signals of HIV-2 RNA, which are the focus of study in this dissertation are shown.

1.1.2.5. Replication cycle of HIV

HIV infects the CD4 lymphocytes of immune system and develops cytopathic effects resulting in the impairment of immunity (Bowen et al, 1985; Klatzmann et al, 1984) with subsequent vulnerability to opportunistic infections (Gottlieb et al, 1981). The infection starts with the binding of HIV virions to the cell surface. This binding is mediated by the interactions between envelope glycoprotein (gp120) of HIV and the cellular receptor CD4 along with coreceptors, CCR5 or CXCR4 (Choe et al, 1996; Deng et al, 1996; Doranz et al, 1996; Dragic et al, 1996; McDougal et al, 1986). Subsequently, viral and cellular membranes fuse (Kowalski et al, 1987) and the viral core is introduced into the cell followed by reverse transcription of the viral RNA genome to double stranded proviral DNA (see review; Harrich & Hooker, 2002). The proviral DNA is directed towards the nucleus with the aid of viral and host proteins and integrated into the host genomic DNA with the help of the integrase enzyme (Bukrinsky et al, 1993; Farnet & Bushman, 1997). Inside the nucleus, transcription factors bind the 5' LTR, which promotes the transcription of the proviral DNA (Garcia et al, 1989; Kao et al, 1987). The resulting viral RNAs are exported into the cytoplasm for the synthesis of viral proteins (Feinberg et al, 1986; Pollard & Malim, 1998). As described above, proteins are transported towards the cell membrane where they are assembled into viral particles, along with two molecules of full-length genomic RNA (Ganser-Pornillos et al, 2008). The assembled particles are released (budded) from the plasma membrane of the infected cell as immature particles and the viral protease becomes active to cleave the Gag and Gag-Pol polyproteins into individual proteins. The cleavage leads to the maturation of the

immature particles into infectious virions to infect other cells (Kaplan et al, 1994; Kohl et al, 1988; Peng et al, 1989) (Figure 4).

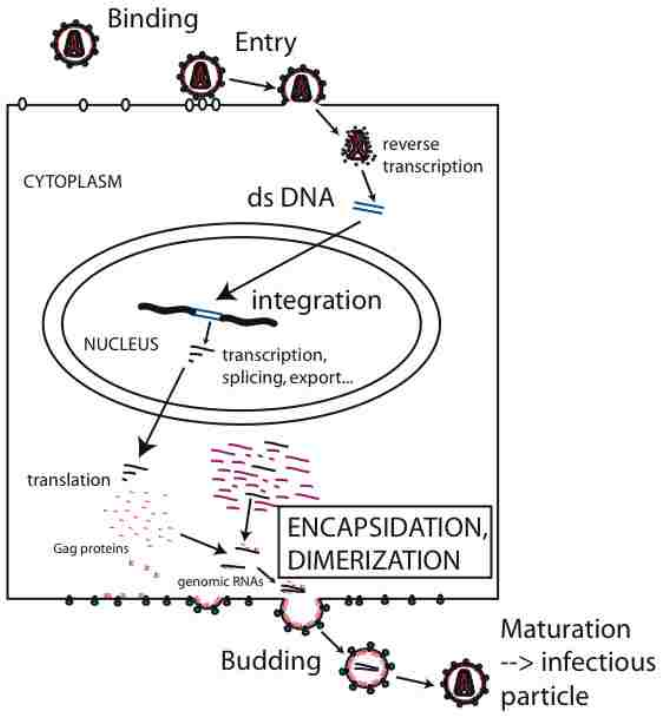


Figure 4. A schematic of the HIV replication cycle.

1.1.3. Genomic RNA dimerization of HIV

The genome inside a single HIV-1 particle is a dimer consisting of two identical copies of positive stranded RNA molecules joined through non-covalent interactions. (see reviews; Greatorex & Lever, 1998; Paillart et al, 2004b; Russell et al, 2004). The dimeric nature of HIV-1 genomic RNA was demonstrated by isolating the virion RNAs and analyzing them by non-denaturing gel electrophoresis. The analysis showed two RNA species: one corresponded to the single RNA molecule called monomer and the other to

two RNA molecules called dimer (Fu et al, 1994). Moreover, using EM (electron microscopy), the strongest RNA-RNA base-pairing interactions in the dimer of HIV-1 have been localized to the 5' end of the genome called the DLS (dimer linkage structure) (Hoglund et al, 1997).

1.1.3.1. Identification of contact sites in DLS

The DLS recognized by EM studies for HIV-1 genomic RNA as described above was originally identified in 1990, when Darlix and coworkers showed that *in vitro* transcribed HIV-1 RNAs corresponding to the 5' UTR of genomic RNA can dimerize *in vitro* even in the absence of proteins (Darlix et al, 1990). Subsequently, the first RNA-RNA contact sites in this DLS, involved in HIV-1 RNA dimerization were identified in HIV-1_{Mal} isolate by using short RNA constructs containing 5'UTR. This site is located between the PBS and the SD and was called the DIS or dimerization initiation site (Paillart et al, 1994; Skripkin et al, 1994). About the same time, a homologous DIS was found for another isolate, HIV-1_{Lai} (Laughrea & Jette, 1994; Muriaux et al, 1995). Because the DIS can fold into a hairpin structure, it was called stem-loop 1 (SL1; (McBride & Panganiban, 1996)) (Figure 3).

1.1.3.2. Model for SL1-mediated RNA dimerization

HIV-1 SL1 contains an autocomplementary sequence of 6-nts in its apical loop (5'-GCGCGC-3' or 5'-GUGCAC-3' depending on the isolate). Due to the presence of this autocomplementary sequence, it was suggested that dimerization between two RNA

molecules is initiated through SL1 loop-loop Watson-Crick base-pairing interaction. This model of dimerization initiation was called the kissing loop model (Laughrea & Jette, 1994; Paillart et al, 1994; Skripkin et al, 1994) (Figure 5A). The kissing complex may proceed further through an extended intermolecular interaction to form a stable extended duplex, at least when short RNA constructs are used (Greatorex & Lever, 1998; Laughrea & Jette, 1996; Muriaux et al, 1996b) (Figure 5A).

A similar autocomplementary sequence (5'-GGUACC-3') has been identified in the HIV-2 5' UTR (Dirac et al, 2001; Jossinet et al, 2001; Lanchy & Lodmell, 2002) and is located as a homologous SL1 structure (Berkhout, 1996). Using phylogenetic analysis of HIV-1 and HIV-2 strains, the HIV-2 DIS has been found to be a conserved structure (Berkhout & van Wamel, 1996) that may promote dimerization through similar kissing loop and extended duplex formation (Figure 5B).

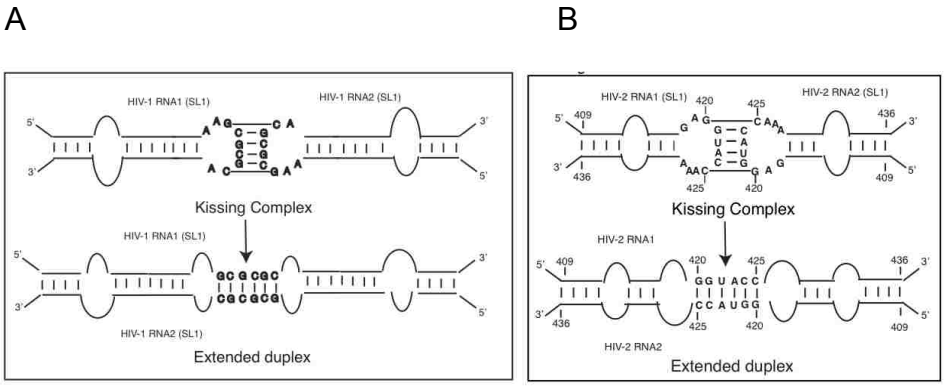


Figure 5. The SL1-mediated dimerization model.

A. The SL1-mediated dimerization model for HIV-1. B. The SL1-mediated dimerization model for HIV-2. The dimerization can be initiated by Watson-Crick base-pairing interactions between autocomplementary sequences of the loops between two HIV RNA molecules that can proceed through extended intermolecular interaction to make extended duplex.

1.1.3.3. Regulation of RNA dimerization

HIV-1 RNA dimerization has been found to be regulated *in vitro* by masking of SL1 through base-pairing interactions that silences its ability to initiate dimerization (Berkhout & van Wamel, 2000). SL1 is located in the 5' region of genomic RNA, which has the capability to adopt alternative conformations. Through structural rearrangements, the functional sequences within 5'UTR are either sequestered to silence a function or are accessible to perform a function (Abbink & Berkhout, 2003; Lanchy et al, 2003a). The *in vitro* studies have indicated that the 5'UTR of HIV-1 may exist in two mutually exclusive conformations called LDI (long distance interactions) or BMH (branched multiple hairpins). The SL1 is available to dimerize when the 5'UTR is in the BMH form because of the accessibility of autocomplementary sequence, and is unavailable to dimerize when the 5'UTR is in the LDI form because of its entrapment through intramolecular interactions with the other regions (Abbink & Berkhout, 2003).

The *in vitro* characterization of the dimerization properties of longer HIV-2 RNA constructs suggested that the HIV-2 5'UTR can also adopt different conformations and that sequences located downstream and upstream of SL1 influence SL1-mediated dimerization in these RNAs (Dirac et al, 2002a). Further characterization in our laboratory mapped these sequences by using the 5' and 3' truncated HIV-2 leader RNA constructs. These sequences were localized to nucleotides 189-196 and nucleotides 543-550 of the 5'UTR and were referred to as the C-box and G-box regions. In that study, it was proposed that intramolecular LDI between C-box and G-box prevents the use of SL1 as a dimer element by promoting the sequestration of SL1 in a stable SDI (short-distance

intramolecular interaction) (Lanchy et al, 2003b).

A sequence located just upstream of SL1 was identified through experiments utilizing mutagenesis and complementary DNA oligonucleotides as likely the sequestration element for SL1 (Lanchy et al, 2003a). This sequence is a palindromic element (pal; 5'-GGAGUGCUCC-3') located upstream of SL1 within the core packaging signal and was proposed to be a dimerization regulatory element.

1.1.4. Genomic RNA packaging of HIV

Inside the cytoplasm of an infected cell, there is a vast sea of cellular, spliced and unspliced RNAs. During assembly of virion particles in an infected cell, two molecules of full-length (unspliced) genomic RNAs are selected and transported into budding virion particles in a process called encapsidation or packaging (Figure 6). The primary determinants of genome packaging are RNA sequences or structures located towards the 5' end of genomic RNA called PSI; ψ (packaging signals) and the NC domain of the packaging protein, Gag (see review; D'Souza & Summers, 2005).

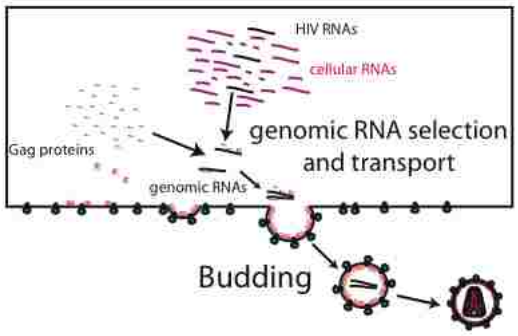


Figure 6. Schematic presentation of the packaging process.

Two molecules of full-length genomic RNA molecules are selected and transported into budding virion particles with the aid of Gag protein.

As described previously, the HIV-1 Gag polyprotein is a precursor protein in immature viral particles and is cleaved into individual proteins by the viral protease. NC is one of the proteolytic products of Gag (Henderson et al, 1992)and studies have shown that the NC domain, as a part of Gag polyprotein, is critical for genome recognition and packaging. Packaging is a selective process because the NC domain recognizes and captures mainly viral RNAs that have appropriate signals for packaging. It also discriminates between unspliced viral RNAs over spliced viral RNAs for this process (see review; D'Souza & Summers, 2005).

In the initial studies that aimed to find the packaging signals required for efficient HIV-1 genomic packaging, it was reported that those signals are located in the 5' UTR between the major SD site and the *gag* (Aldovini & Young, 1990; Lever et al, 1989; McBride & Panganiban, 1997). Following those studies, other elements of the 5'UTR, including nucleotides of U5, PBS, TAR and nucleotides within the *gag*, were also found to be important for HIV-1 genome packaging (Clever et al, 2002; D'Souza & Summers, 2005; Helga-Maria et al, 1999; Luban & Goff, 1994).

For HIV-2, the packaging sequences that mediate genomic RNA packaging into virion particles have also been localized to the 5'UTR. However, there is not universal agreement concerning the exact location of these sequences. In some studies, it has been shown that, like HIV-1, the nucleotides located between the SD site and *gag* are essential for HIV-2 genome packaging (Arya et al, 1998; Poeschla et al, 1998). Other studies indicated that another sequence is crucial for packaging and is located upstream of the SD

site. This sequence was called the Ψ-core (core packaging signal) because of its profound effects on HIV-2 packaging *in vivo* (Griffin et al, 2001; Kaye & Lever, 1999; McCann & Lever, 1997). In refinement of this work, a 10-nt palindromic sequence located within this signal was defined as a critical element for HIV-2 genomic RNA packaging *in vivo* (Lanchy & Lodmell, 2007).

In vitro binding studies between purified Gag and NC proteins of HIV-1 and *in vitro* synthesized RNA fragments corresponding to 5' end of HIV-1 genomic RNA have been useful in dissecting the packaging mechanism. These studies were helpful to identify sequences or structures at the 5' end of HIV-1 genomic RNA that act as binding sites for Gag and NC (Berkowitz et al, 1993; Sakaguchi et al, 1993). Based on these studies along with other additional studies, it was reported that the Ψ-site of HIV-1 RNA is composed of four stem-loops (SL1-SL4) (Clever et al, 1995).

For HIV-2, the interactions between packaging proteins and the RNA fragments corresponding to the 5' side of HIV-2 genomic RNA have also been investigated by *in vitro* binding studies. However, there is a limited number of studies and either small HIV-2 RNAs were used for binding with HIV-2 proteins or large HIV-2 RNAs were used with HIV-1 proteins. (Damgaard et al, 1998; Tsukahara et al, 1996). To more accurately identify the specific HIV-2 RNA-protein binding sites, it will be necessary to conduct binding studies on large HIV-2 RNA transcripts encompassing the entire 5' UTR with either HIV-2 Gag or NC proteins.

1.2. Investigation of the structural presentation of dimerization and packaging signals in 5'UTR

As described above, the 5' UTR is a conserved and essential region of HIV genomic RNA. It contains several sequence or structural signals that are important for various steps of viral replication, including RNA dimerization and packaging. The appropriate presentation of the RNA dimerization and packaging signals is important for temporal regulation of these processes. It has been proposed that the 5'UTR can adopt alternate conformations to expose or hide these signals. To investigate this notion, it is important to determine which specific RNA structures are essential for the regulation of the dimerization and packaging processes.

1.2.1. Combinatorial approaches for determining essential RNA structures

A straightforward way to determine whether a sequence or structure motif is essential is through site-directed mutagenesis. Loss of function through mutation of the motif is suggestive of its importance. However, many of the structures in the 5'UTR are poorly defined and thus the choice of nucleotides to mutate by site-directed mutagenesis

is difficult. A description appeared in 1990 for a method termed as SELEX (systematic evolution of ligands by exponential enrichment) (Tuerk and Gold, 1990). In this way, a huge number of possible mutations on a particular sequence of interest can be tested at the same time rather than testing mutations individually and sequences with functional structures can be selected from a pool of random sequences (Ellington & Szostak, 1990; Joyce, 1989; Tuerk & Gold, 1990). The SELEX method has been adapted for both *in vitro* and *in vivo* approaches.

1.2.1.1. *In vitro* SELEX

In vitro SELEX involves several rounds of selection for the purification of functional RNA molecules from a complex pool or population of RNA variants. One round of selection is comprised of three steps: 1. Synthesis of the randomized RNAs. 2. Selection 3. Amplification, as described below briefly. A general scheme of *in vitro* SELEX is shown in Figure 7.

1.2.1.1.1. Synthesis of the randomized RNAs

This step starts with the generation of a complex pool of DNAs that are randomized at a particular sequence. For example, if 9 nts (nucleotides) are randomized, the starting pool represents 262144 different possible molecules ($4^9 = 262144$, where 4 = A, C, T, G). The pool of randomized DNAs is transcribed to obtain the pool of

randomized RNAs. This pool of RNAs is called the initial pool and is an ensemble of functional and non-functional RNA molecules.

1.2.1.1.2. Selection

The initial pool of RNA molecules with random sequences is separated into pools of desired and undesired molecules by the use of various partitioning techniques. As a result of partitioning, the complexity of sequences becomes lower in the new pools as compared to the initial pool. Subsequently, these pools are subjected to further rounds of selection to achieve desired characteristics by enriching it with a great proportion of active molecules.

1.2.1.1.3. Amplification

To initiate every new round of selection, the selected RNA molecules from the previous step are used as templates for the synthesis of cDNA by reverse transcriptase, followed by amplification by PCR. With subsequent rounds of selection, functional sequences/structures dominate the pool and the non-functional sequences/structures are discarded. Ultimately, the activity of the pool is increased and the complexity is decreased.

The last step after the completion of the rounds of selection is to analyze the composition of the functional and non-functional pools. This is achieved by isolating

individual clones by cDNA cloning, followed by sequencing and or transcribing to individual RNAs for further characterization.

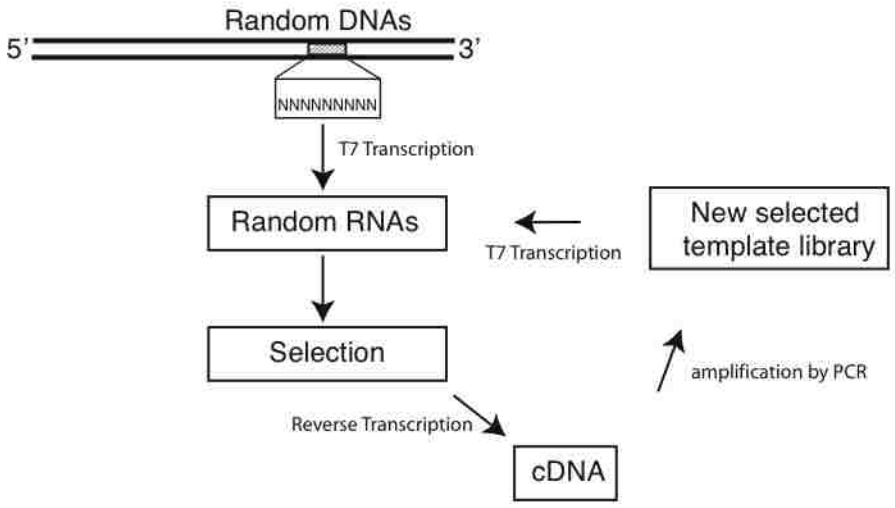


Figure 7. Schematic presentation of *in vitro* SELEX.

1.2.1.2. *In vivo* SELEX

In addition to employing the SELEX method *in vitro*, one can also use it inside the cells, termed *in vivo* SELEX. Berkhout and Klaver described the *in vivo* application of the SELEX to RNA viruses in 1993 for the first time (Berkhout & Klaver, 1993), after which other groups (Doria-Rose & Vogt, 1998; Morris et al, 2002) used it. This approach is an exciting tool for understanding the functional role of a sequence/structure in the life cycle of a virus describing the biological relevance of that sequence/structure. Using this approach, replication-competent viruses can be obtained from a pool of viable and non-viable proviral DNAs with random sequences at a particular region (Figure 8).

The primary step of *in vivo* SELEX involves the construction of a random pool of

full-length infectious DNA clones, which is technically more challenging than constructing short randomized DNA templates for *in vitro* SELEX. Subsequently, this pool is transfected and infected into eukaryotic cells and infection is allowed to persist for a certain time until only replication-competent viruses survive. After the completion of the entire selection for the dominant sequences, two steps are performed. First, selected viruses are processed for RNA extraction and cloning for sequence analysis. Second, these sequences are cloned into the full-length backbone of the original wt plasmid to test individually the impact of these sequences on viral replication. Both *in vitro* and *in vivo* SELEX methodologies were employed to obtain new structural information on the HIV-2 packaging and dimerization domains, presented in this dissertation.

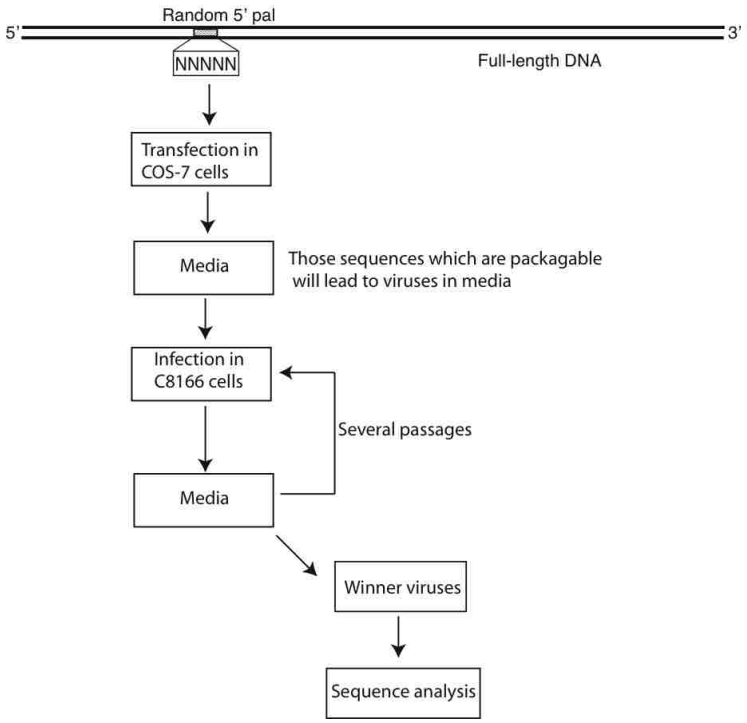


Figure 8. Schematic presentation of *in vivo* SELEX.

1.3. Specific aims

As mentioned before, the pal is a 10-nt palindrome sequence within packaging signal and is located upstream of the dimerization signal (Griffin et al, 2001; Lanchy et al, 2003a). In this dissertation research, the overall goal was to determine which structures and sequences are essential for the regulation of RNA dimerization and packaging, with attention focused on the pal region. This goal was divided into three specific aims, which are listed along with the overall findings below.

1.3.1. Specific Aim 1: Characterize the role of pal on dimerization *in vitro*

Since pal was shown to interfere with the SL1-mediated dimerization *in vitro* (Lanchy et al, 2003a), we hypothesized that pal acts as a regulator of SL1-mediated dimerization by base pairing with SL1. To study this hypothesis, we randomized pal and selected for those RNAs that were dimerization competent or dimerization-incompetent *in vitro.* This study constitutes specific aim 1 and is presented as a published manuscript (Baig et al, 2007) in chapter 2.

The first section of the research includes the experimental details about the randomization of pal motif in the context of HIV-2 leader RNA and selecting those RNAs with impaired and enhanced dimerization properties. The second section provides a detailed predicted secondary structure analysis of pal-SL1 region of these selected

RNAs. The third section is about the characterization of dimerization properties of a few of the individual RNAs. The results of this section provide evidence on the correlation of predicted secondary structures of pal-SL1 region with the dimerization properties of RNA. Overall, we demonstrate that an extended SL1 structure, which is formed as a result of base-pairing interactions between the 3' side of pal with a sequence downstream of SL1, is a regulator of dimerization.

1.3.2. Specific Aim 2: Characterize the role of pal *in vivo*.

Since pal is located within the packaging signal and has been shown to regulate dimerization *in vitro* through structural rearrangements, we hypothesized that pal is involved in regulating viral RNA packaging. The third chapter of this dissertation, which is an accepted manuscript (Baig et al, 2008) describes the research conducted to study this hypothesis. The research employs *in vivo* SELEX to study the role of 5' side of pal in packaging.

The first section describes the construction of a library of full-length HIV-2 proviral DNA genomes with random sequences in 5' side of pal. The second section indicates the selection for replication-competent (viable) viruses from the population of viruses that were randomized at the 5' side of pal. The third section shows that the 5' pal sequences of the viable viruses converged to the consensus sequence GGRGN. The last section provides evidence that this sequence at the 5' side of pal is critical for effective genomic RNA packaging and viral replication.

1.3.3. Specific Aim 3: Characterize the role of pal on RNA-protein interactions *in vitro*

The predicted secondary structure analysis of pal-SL1 region for selected and non-selected sequences from specific aim 2 suggested that extended SL1 is important for genomic RNA packaging, when the 5' side of pal is unpaired and exposed. Because of the similarities with known Gag and NC binding preferences, GGAG is likely to be an essential element for Gag and NC recognition and binding.

The fourth chapter in this dissertation includes experiments to study this hypothesis. The first section describes the expression and purification of Gag and NC proteins and synthesis of two radioactive RNAs corresponding to 5'UTR of HIV-2 genomic RNA. One of those was an unselected, non-consensus sequence obtained during the study of specific aim 2 and was packaging-defective. We hypothesized that the packaging defect may be attributable to poor Gag and/or NC binding to the non-consensus RNA sequence. The other RNA was the wt sequence and was also a selected individual in that study. The second section of this chapter shows the binding results between RNAs and the purified proteins. Our results indicate that Gag binds preferentially to wt RNA as compared to the packaging-defective RNA *in vitro,* which is consistent with *in vivo* results.

The fifth and last chapter of this dissertation describes the overall conclusions drawn from the research presented here. In addition, it provides some future directions for the continuation of studies described in the third specific aim.

Chapter 2: HIV-2 RNA dimerization is regulated by intramolecular interactions *in vitro*

Tayyba T. Baig, Jean-Marc Lanchy[1] & J. Stephen Lodmell

Division of Biological Sciences, The University of Montana, Missoula, MT 59812, USA

[1]Corresponding author: Jean-Marc Lanchy

This chapter is a modified version of the manuscript published in the RNA Journal in August 2007.

Baig, T.T., Lanchy, J.M. and Lodmell, J.S. (2007) HIV-2 RNA dimerization is regulated by intramolecular interactions in vitro. *RNA*, 13, 1341-1354.

This manuscript represents the experiments done by the author of this dissertation with the exception of the probing experiments of HIV-2 leader RNA, which were performed by Dr. Jean-Marc Lanchy.

2.1. Abstract

Genomic RNA dimerization is an essential process in the retroviral replication cycle. *In vitro*, HIV-2 RNA dimerization is mediated at least in part by direct intermolecular interaction at SL1 (stem-loop 1) within the 5' untranslated leader region (UTR). RNA dimerization is thought to be regulated via alternate presentation and sequestration of dimerization signals by intramolecular base pairings. One of the proposed regulatory elements is a palindrome sequence (pal) located upstream of SL1. To investigate the role of pal in the regulation of HIV-2 dimerization, we randomized this motif and selected *in vitro* for dimerization-competent and dimerization-impaired RNAs. Energy minimization folding analysis of these isolated sequences suggests the involvement of pal region in several short-distance intramolecular interactions with other upstream and downstream regions of the UTR. Moreover, the consensus predicted folding patterns suggest the altered presentation of SL1 depending on the interactions of pal with other regions of RNA. The data suggest that pal can act as a positive or negative regulator of SL1-mediated dimerization and that the modulation of base pairing arrangements that affect RNA dimerization could coordinate multiple signals located within the 5'UTR.

2.2. Introduction

Retroviruses package two copies of their positive sense single stranded RNA genome into a single viral particle. Analysis of the packaged genomes revealed that the two RNA molecules are linked by non-covalent bonds and form a dimeric RNA structure inside the viral particle (see reviews; Greatorex & Lever, 1998; Paillart et al, 2004b; Russell et al, 2004). The dimeric nature of the genomic RNA has been characterized by sedimentation analysis (Cheung et al, 1972) and non-denaturing gel electrophoresis (Fu & Rein, 1993). Moreover, electron microscopic studies indicated that the two RNA molecules are joined strongly with each other through a region close to their 5' ends, termed the DLS (dimer linkage structure; (Bender & Davidson, 1976; Hoglund et al, 1997)).

The mechanism of dimerization promoted by the DLS was further studied *in vitro* using RNA fragments encompassing the 5' end of the HIV-1 genomic RNA, which showed the spontaneous dimerization of these RNA fragments without any cellular or viral proteins (Darlix et al, 1990). The essential motif for dimerization of the HIV-1 genomic RNA was identified *in vitro* and called the DIS (dimerization initiation site; (Laughrea & Jette, 1994; Muriaux et al, 1995; Paillart et al, 1994; Skripkin et al, 1994)) or SL1 (stem-loop 1; (McBride & Panganiban, 1996)). *In vitro*, this motif mediates dimerization between two RNA molecules through a kissing loop interaction (Laughrea & Jette, 1994; Paillart et al, 1994; Skripkin et al, 1994) that may proceed further through an extended intermolecular interaction to form a stable extended duplex at least when short RNA constructs are used (Greatorex & Lever, 1998; Laughrea & Jette, 1996; Muriaux et al, 1996a).

The dimerization properties of the HIV-2 leader RNA are somewhat different from those of the HIV-1 leader. Multiple dimerization elements have been described within the HIV-2 leader RNA, including stem loop 1, which is homologous to the HIV-1 SL1 (Dirac et al, 2001; Lanchy & Lodmell, 2002), and nucleotides within the PBS (primer-binding site; (Jossinet et al, 2001; Lanchy & Lodmell, 2002)). Furthermore, a 10 nucleotide palindrome sequence called pal within the encapsidation signal of the HIV-2 leader RNA, has been proposed to be involved in dimerization or its regulation *in vitro* (Lanchy et al, 2003a).

The ability of genomic RNA to dimerize is a potentially important event for key steps of the viral replication cycle such as translation, encapsidation, and recombination (Abbink & Berkhout, 2003; Baudin et al, 1993; Chin et al, 2005; D'Souza & Summers, 2004; Darlix et al, 1990; Flynn et al, 2004; Fu et al, 1994; Fu & Rein, 1993; Hibbert et al, 2004; Hu & Temin, 1990; Jetzt et al, 2000; Mikkelsen et al, 2000; Rein, 1994; Sakuragi et al, 2003; Stuhlmann & Berg, 1992). Genomic RNA dimerization and encapsidation have been proposed to be regulated through different conformations of the leader RNA region that trigger or prevent these two processes. Two alternative conformations have been reported *in vitro* for HIV-1 leader RNA that show different dimerization properties (Berkhout & van Wamel, 2000; Huthoff & Berkhout, 2001). One of the proposed conformations exists as a rod like structure with a LDI (long-distance base pairing interaction) between the poly (A) signal stem loop and SL1 domains (Huthoff & Berkhout, 2001), therefore masking the SL1 and inhibiting dimerization. In the other proposed conformation, the leader region can refold into a BMH (branched structure with multiple hairpins) that exposes both the poly (A) and SL1 hairpins and favors

dimerization. Although disruption of the LDI causes phenotypic changes in HIV-1 replication (Abbink et al, 2005; Ooms et al, 2004), structural evidence for the two conformations has been elusive as chemical probing of the HIV-1 genomic RNA structure in infected cells and in viral particles did not reveal the existence of a predominant LDI conformation (Paillart et al, 2004a).

In contrast to HIV-1, the HIV-2 leader RNA fragments do not dimerize efficiently through SL1 *in vitro* (Dirac et al, 2002a; Lanchy et al, 2003a; Lanchy & Lodmell, 2002; Lanchy et al, 2003b). It has been proposed that similar to HIV-1, monomeric HIV-2 RNA can adopt two alternative conformations that sequester and release SL1, respectively (Dirac et al, 2002a; Lanchy et al, 2003a). The impaired dimerization of HIV-2 RNA is thought to be caused by an energetically favored folding that sequesters SL1 (Dirac et al, 2002a; Lanchy et al, 2003b). We suggested that the intramolecular sequestration of SL1 occurs through base-pairing with the upstream 10-nt pal sequence (Lanchy et al, 2003b).

Since the 10-nt pal sequence (392-401) is located upstream of the SL1 in the HIV-2 leader RNA (Figure 9) and can affect the SL1-mediated dimerization of HIV-2 leader RNA fragments *in vitro*, we have characterized in the present study the role of pal by randomizing this motif and selecting from a population of random-pal RNAs those with enhanced or diminished dimerization properties. We cloned and sequenced individual RNAs from the selected pools and used their sequences to analyze the functionally important secondary structures that might be involved in silencing or enhancing SL1's role as a dimerization element. In particular, the predicted folding models suggest that the non-native selected pal sequences (denoted pal* to distinguish from the wt pal sequence) were involved in different patterns of binding in the monomer and dimer RNA

populations. Moreover, the different pal* binding patterns are associated with different structural presentation of SL1 that correlated with the observed dimerization abilities of the selected RNAs. Our observations underscore the importance of pal* as a regulator of dimerization by modulating SL1 presentation. In addition, similarities in the dimerization behavior and predicted folding of the monomer pool RNAs with wt RNA suggest a general model for pal's involvement in the regulation of dimerization.

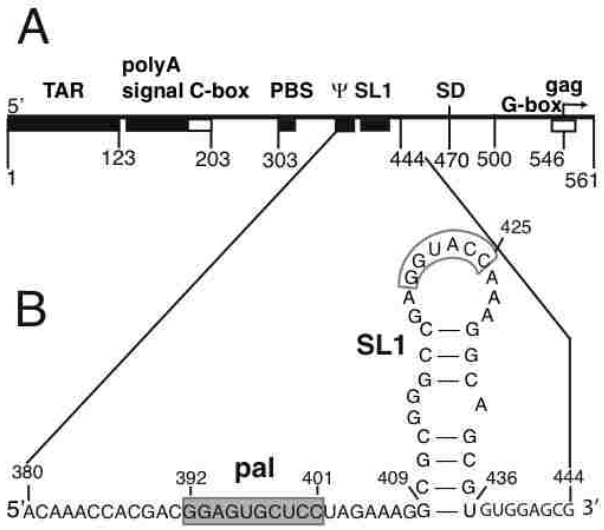


Figure 9. HIV-2 leader RNA and location of the pal and SL1 elements.

A. The 5' leader region of HIV-2 ROD genomic RNA is represented with boxes and numbers to indicate the landmark sequences with their names indicated above. TAR, poly A signal, C-box, PBS, ψ, SL1, SD, G-box and gag represent the trans-activation region, the poly (A) signal domain, C-rich sequence, the primer binding site, the encapsidation signal ψ, stem loop 1, the major splice donor site, G-rich sequence, and the 5' end of the Gag protein coding region, respectively. B. A simplified secondary structure of nucleotides 380-444. The gray box highlights the 10-nt palindrome sequence (pal) in the encapsidation signal and the 6-nt autocomplementary sequence essential for SL1-mediated dimerization in the apical loop of the SL1 is outlined.

2.3. Materials and Methods

2.3.1. Production of pool "0" DNA template

Several PCR reactions and ligation steps were used to build the pool 0 DNA library, which was degenerate at nine nucleotides in the pal (palindrome sequence). First, a sense primer containing a *Bam*HI site with a T7 RNA polymerase promoter (sBamT7R; Table 1) and an antisense primer containing the unique *Hpy*99I site (nts 388-392)(asHpy; Table 1) were used to amplify the first 392-nt long fragment of the HIV-2 leader region (ROD isolate, GenBank M15390). Second, a sense primer containing the *Hpy*99I site with the degenerate sequence in pal (sHpy-random; and an antisense primer containing an *Eco*R1 site (asEco561;

Table 1) were used to amplify the last 169-nt long fragment of the HIV-2 1-561 region (PfuUltra™ High Fidelity DNA polymerase, Stratagene). The 10-nt pal was degenerate at only nine nucleotides (393-401), because the first nucleotide (G392) of pal belongs to the *Hpy*99I recognition site used in our cloning strategy and thus needed to be kept intact to get cohesive *Hpy*99I ends of the fragments (Figure 10A). Third, the agarose gel-purified PCR fragments were digested with *Hpy*99I and ligated together using the Quick ligation kit (New England Biolabs). Another gel purification was used to purify the correct ligation product, *i.e*. the 392-nt fragment ligated to the 169-nt fragment. Fourth, the 561-nt long ligation product was re-amplified using sBamT7R and asEco561 primers, gel-purified and digested with *Bam*HI and *Eco*R1 to get the pool 0 DNA template.

sBamT7R	5’-TAG GAT CCT AAT ACG ACT CAC TAT AGG TCG CTC TGC GGA GAG-3’
asEco561	5’-AAG AAT TCA GTT TCT CGC GCC CAT CTC CC-3’
sHpy-random	5’-CAA CCA CGA CGN NNN NNN NNT AGA AAG GCG CGG GCC GAG G-3’
AsHpy	5’-AGG CAC TCC GTC GTG GTT TGT TCC TGC CGC CC-3’

Table 1. Oligonucleotides used in chapter 2.

The nucleotides forming the *Hpy*99I site are underlined (5'-CGACG-3').

2.3.2. Pool 0 RNA synthesis and purification

Pool 0 RNA was synthesized from the pool 0 DNA template with the AmpliScribe ™ T7 transcription kit (Epicentre). After transcription, the DNA template was digested with RNase-free DNase. The RNA transcripts were purified by phenol/chloroform extraction, ammonium acetate and ethanol precipitation followed by denaturing gel electrophoresis, excision, and extraction of the 1-561 RNA. A sample of the pool 0 DNA template was sequenced to verify degeneracy of the pal region (data not shown).

2.3.3. Selections for dimerization-competent and dimerization-impaired RNAs

The selection for dimerization-competent RNAs was carried out essentially as previously described (Lodmell et al, 2000). In this study, additional selections were made separately for dimerization-impaired RNAs. Briefly, an aliquot of the pool 0

degenerate RNA was denatured in 8 μl H_2O at 90°C for 2 min, then snap cooled on ice for 2 minutes. Dimerization was allowed to proceed for 30 min at 55°C after adding the dimer buffer (final concentration: 50 mM Tris-HCl, pH 7.5 at 37°C, 300 mM KCl and 5 mM $MgCl_2$). Following 30 min of incubation, samples were loaded with 2 μl of 6x glycerol loading dye (40 % glycerol, 44 mM Tris-borate, pH 8.3, 0.25 % bromophenol blue) onto a 0.8 % agarose gel. Electrophoresis was carried out for 90 min at room temperature (24˚C) in 44 mM Tris-borate, pH 8.3 and 1 mM EDTA (0.5x TBE). The ethidium bromide-stained gel was visualized on a Fuji FLA3000G Image Analyzer. The bands corresponding to the dimer and monomer RNA species were excised from the gel. To obtain the templates for further rounds of transcription and selection, a slice of each excised band was subjected to RT-PCR (reverse transcription and polymerase chain reaction) using AccuScript™ High Fidelity RT-PCR kit (Stratagene). The reverse transcription step was performed in a final volume of 10 μl to synthesize the cDNA from a 5 μl equivalent slice of each excised band with the asEco561 antisense primer (0.6 μM), dNTP mix (1 mM of each dNTP), 1x AccuScript™ RT buffer and 0.5 μl AccuScript™ High Fidelity RT enzyme according to the manufacturer's protocol. The reverse transcription reaction was incubated at 42°C for 30 minutes. The PCR step was performed in a final volume of 50 μl containing 4 μl of the cDNA template, 0.2 μM of sBamT7R and asEco561 primers (

Table 1), dNTP mix (0.2 mM of each dNTP), 1x PCR buffer and 2.5 units of PfuUltra™ High Fidelity DNA polymerase (Stratagene). Forty cycles of amplification with a 55°C annealing temperature were used. The resulting RT-PCR products were digested with *Bam*HI and *Eco*R1 prior to transcription or cloning. We carried out five and

six successive rounds of selections for dimerization-impaired and dimerization-competent RNAs, respectively. For sequence analysis, the digested RT-PCR products of the final rounds for both selections were ligated to the *Bam*HI and *Eco*RI sites of the pUC18 plasmid. The ligation reaction was transformed into competent DH5α *Escherichia coli* bacteria. Sixty individual DNA clones from each selection were obtained by DNA isolation from the bacterial colonies and sequenced.

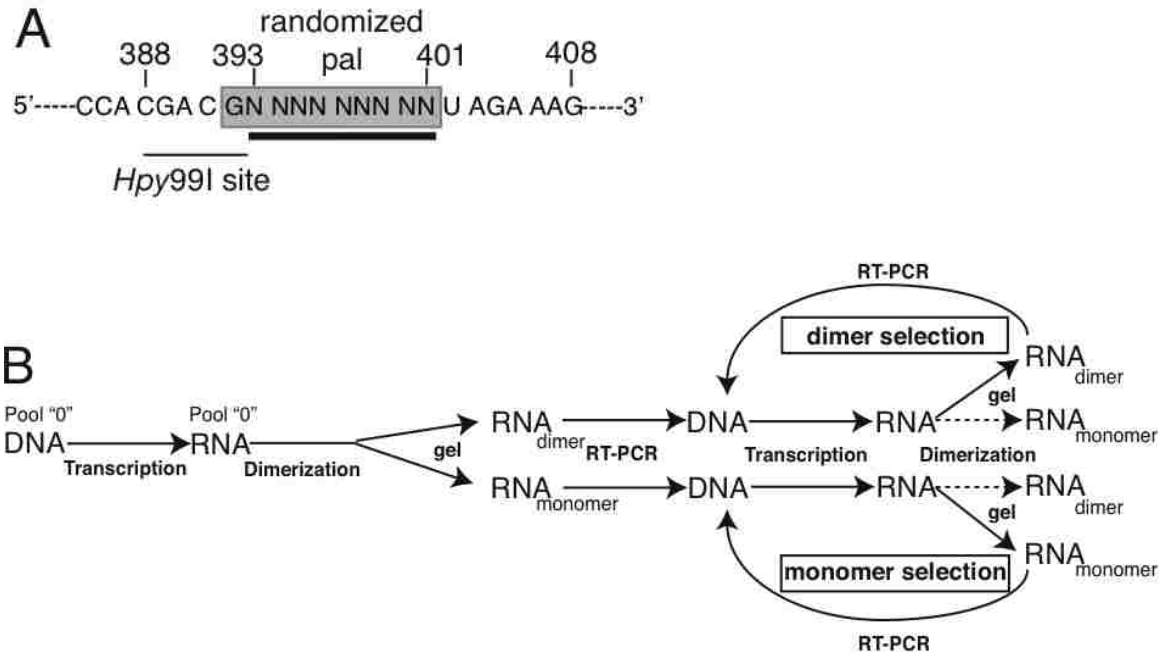


Figure 10. Randomized pal* sequence and *in vitro* selection methodology.

A. Detailed view of randomized pal region of the 561-nt pool 0 RNA used for *in vitro* selection. The pal element is a 10-nt palindrome sequence within the encapsidation signal ψ, located immediately upstream of SL1. Nine nucleotides (393-401) in the pal sequence were randomized in the template used for transcription of the pool 0 RNAs, while keeping the first nucleotide of pal (G392) constant so as not to disrupt the *Hpy*99I restriction site required for template construction (see Materials & Methods). B. Schematic of the SELEX approach used to select and amplify the dimerization-competent and dimerization-impaired RNAs from the dimer and monomer pools, respectively. The dimeric and monomeric RNAs were partitioned on non-denaturing agarose gel. The dimeric and monomeric RNAs were excised from the pool 0 gel and used as the templates for RT-PCR to generate the DNA template to transcribe RNAs for the subsequent dimer and monomer pools. This cycle was repeated several times separately for the dimer and monomer selections. The solid arrows represent the dimeric and monomeric RNAs used for each subsequent round of selection. The dotted arrows represent those RNAs that were not processed for selection.

2.3.4. Prediction of secondary structures

Mfold version 2.3 (Zuker, 2003) was used to predict the most stable secondary structure for 1-561 wild type HIV-2 ROD and the 120 individual clonal 1-561 RNA sequences isolated from the final dimer (7^{th}) and monomer (6^{th}) pools. The software used is found on the mfold server (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1-2.3.cgi) (Zuker, 2003). The folding temperature was set at 55˚C since the RNAs were incubated at 55°C in our dimerization protocol. The 120 most stable (*i.e.* optimal) secondary structures were visually analyzed and the organization of the pal* and SL1 elements and the base-pairing partners of the selected pal* nucleotides (nts 392-401) were recorded, allowing segregation of the 120 pal* sequences into distinct structural groups (see Figure 12 and Table 3).

2.3.5. Synthesis and purification of selected RNAs

Eight clones out of the 60 individual DNAs from each final dimer and monomer pools were chosen for characterization of dimerization properties. The plasmids were linearized with *Eco*RI prior to *in vitro* transcription. RNAs were synthesized with AmpliScribe ^{TM} T7 transcription kit (Epicentre). After synthesis, RNAs were purified by ammonium acetate precipitation and size exclusion chromatography (Bio-Gel P-4, Bio-Rad), followed by ethanol precipitation.

2.3.6. *In vitro* dimerization of selected RNAs

The dimerization efficiency of the 16 individual clonal RNAs was checked by non-denaturing gel electrophoresis as described for the *in vitro* selection of dimer and monomer pools.

2.3.7. RNA solution structure probing

In a standard experiment, 5 picomoles of 1-561 wild type HIV-2 genomic RNA was heated in water for 2 min at 90˚C, then quench cooled on ice. The folding was started by the addition of dimer buffer. After 0, 4 or 20 min incubation at 55˚C, dimethylsulfate (Aldrich) was added to the RNA and the mixture incubated 2 min at room temperature (27˚C). The final concentration of DMS was 0.5%. Chemical probing was stopped by the addition of 10 μg glycogen, EDTA (2 mM, final concentration), sodium acetate (300 mM, final concentration), followed by two and a half volumes of absolute ethanol. After a 30-min precipitation at –25˚C, the samples were pelleted by centrifugation at 15,000 rpm for 30 minutes, ethanol washed, vacuum dried, and resuspended in water. One fourth of the resuspended material was then reverse-transcribed with the avian myeloblastosis virus reverse transcriptase (Promega) and a 5' end-^{32}P-labeled radioactive DNA oligonucleotide primer (complementary to nts 470-494). After a 30-min primer extension at 42˚C, the samples were precipitated as described above and the dried pellet was resuspended in formamide loading and tracking dye and analyzed by denaturing PAGE. A DNA sequencing of the plasmid DNA used for the transcription was loaded together with the primer extension to identify the modified nucleotides.

2.4. Results

2.4.1. Construction of pool "0' RNA and selections for dimerization-competent and dimerization-impaired RNAs

The HIV-2 leader RNA fragment (561-nt long) was previously shown to dimerize inefficiently through SL1 *in vitro* (Lanchy & Lodmell, 2002; Lanchy et al, 2003b). The existence of an alternative conformation of the monomeric leader RNA was proposed to be responsible for this inability to dimerize efficiently, in which a short-range intramolecular interaction (pal-SL1) occludes the SL1 dimerization element (Lanchy et al, 2003a). The 10-nt pal sequence (392-401) is located upstream of the SL1 in the HIV-2 RNA (Figure 9). To further characterize the role of pal, we designed a 1-561 RNA construct harboring a randomized pal to determine how pal could act as a regulatory element for SL1. Nucleotides 393-401 of pal were randomized within the DNA transcription template used in this study. Hence, the pre-selection (pool 0) RNA species were 561-nt long and harbored a $_{392}$GNNNNNNNNN$_{401}$ sequence (where N has an equal probability of being A, C, G, or U) instead of the wt $_{392}$GGAGUGCUCC$_{401}$ sequence (Figure 9B, Figure 10A) Randomization of this region was verified by sequencing (data not shown).

An *in vitro* selection/amplification technique was used to enrich the randomized populations for RNAs that were dimerization-competent or-impaired. The approach used in the selections of both pools is explained in Materials and Methods and in (Figure 10B). Pool 0 RNA was fractionated into dimer and monomer species by non-denaturing gel

electrophoresis. RNA migrating as a dimer species was excised from the gel, subjected to RT-PCR, transcription, and the resulting RNA was used in a subsequent round of selection for dimerization-competent RNAs. Likewise, RNA migrating as a monomer species was similarly processed and the re-amplified RNA was used in a subsequent round of selection for dimerization-impaired RNAs (Figure 10B). After initial fractionation of pool 0 for dimer and monomer RNA species, further selections were made separately to minimize cross contamination. The percentage of RNA migrating in dimeric form increased with the consecutive rounds of dimerization-competent RNAs selection (Figure 11A). We stopped the dimerization-competent RNAs selection when the percentage of dimer reached a plateau and remained constant for several rounds of selection (Figure 11C). In parallel, the percentage of dimer decreased with the successive rounds of dimerization-impaired RNAs selection (Figure 11B). We discontinued the selection when the percentage of dimer remained the same for several rounds of selection (Figure 11C).

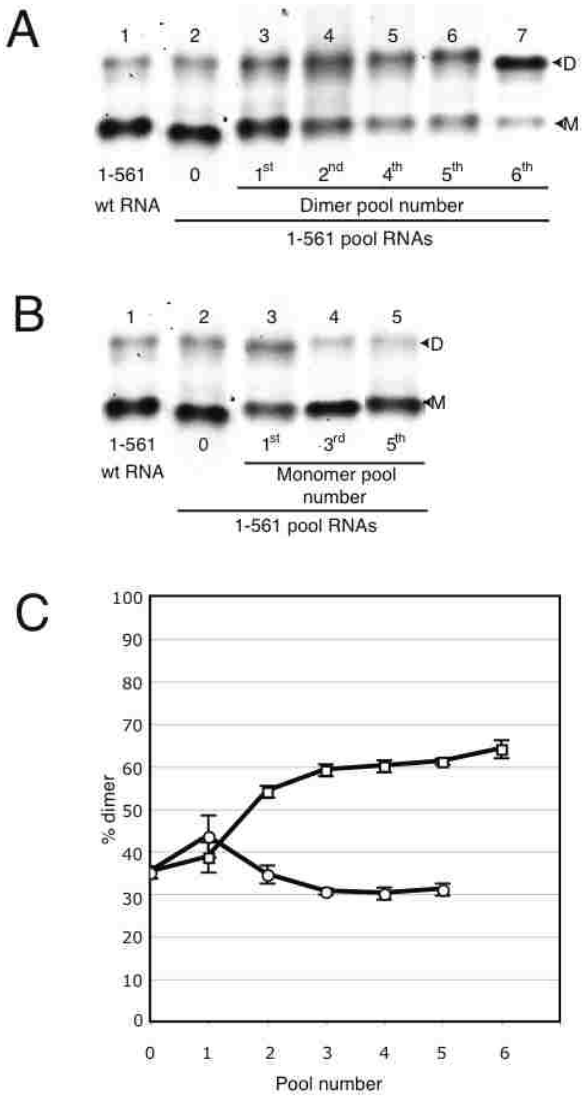


Figure 11. Evolution of dimerization characteristics after repeated rounds of selection for dimerization-competent and dimerization-impaired pool RNAs.

A. A non-denaturing dimerization gel is shown to represent the pool 0 (unselected RNA) and successive rounds of selection (1st, 2nd, 4th, 5th and 6th) of the dimer pools. The top band, migrating as dimer for pool 0 RNA, was used as the starting material for each subsequent round of selection for dimerization-competent RNAs, while the bottom band for pool 0 RNA was used during the separate selections for dimerization-impaired RNAs. The dimerization of the wt RNA is shown for comparison. B. A non-denaturing dimerization gel is shown to represent the pool 0 (unselected RNA) and successive rounds of selection (1st, 3rd and 5th) of the monomer pools. The wt RNA is also shown for comparison. C. The percent dimer yield from the dimer pools (open squares) and from the monomer pools (open circles) was quantified by phosphorimager scanning and analyzed with Fuji Image Gauge v3.3 software.

2.4.2. Interactions of *in vitro* selected pal sequences (pal*) within the 1-561-nt long RNA

The resulting cDNAs from the last dimerization-competent (6^{th} round) and dimerization-impaired (5^{th} round) selections were cloned into pUC18 for either sequencing or expression of the individual clonal RNAs. Sixty clones from each pool were sequenced and subjected to computer assisted lowest energy folding. We called these two pools monomer and dimer pools, since they originated from the last dimerization-impaired and -competent RNA selections, respectively. Mfold version 2.3 (Zuker, 2003) was used to predict the most stable secondary structure at 55°C for each of the 120 individual sequences. The folding temperature was set at 55˚C since the RNAs were incubated at 55°C in our selection protocol. The 120 optimal secondary structures were visually analyzed and the organization of the pal* and SL1 elements and the base-pairing partners of the selected pal* nucleotides was recorded and shown in (Figure 12) by model structures and dots adjacent to the leader RNA schematic, respectively.

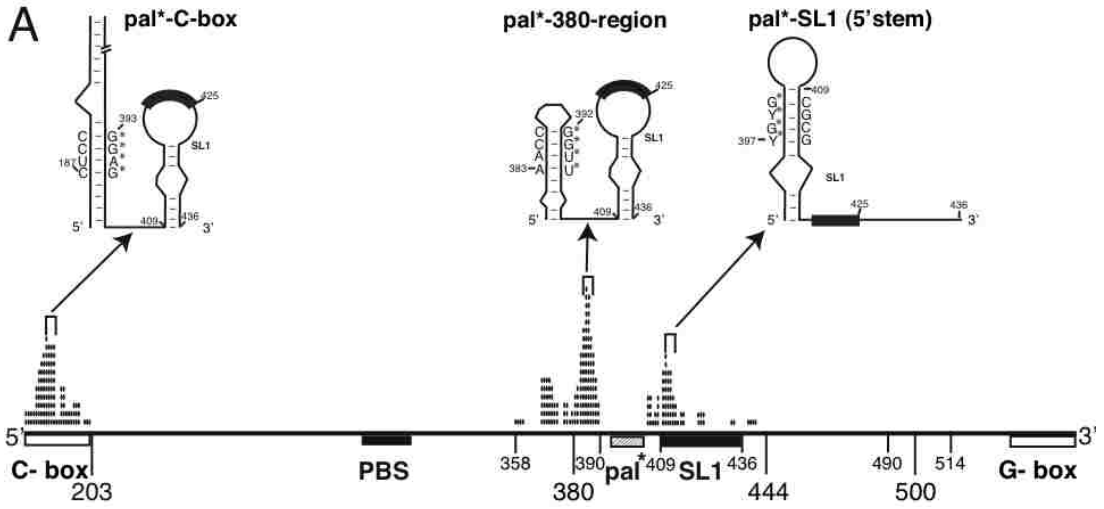
One striking pattern revealed by mfold analysis was the reproducible and relatively small number of intramolecular interactions that variant pal* made with upstream and downstream RNA sequences, as evidenced by the peaks of dots in (Figure 12). Distinct structural groups are formed by pal* interacting with either the C-box, 380, SL1, 444, or 500-regions. We included another structural group called free pal*, representing the variant pal* sequences where none of pal* nucleotides are base-paired in the optimal mfold-predicted structure (7 and 13% of dimer and monomer pool sequences, respectively; Table 3). The major partners of pal* among RNA species isolated from the dimer pool were C-box, 380-region and SL1 (Figure 12A). They represent 27, 45, and

20% of the dimer pool sequences respectively (Table 3). In the monomer pool, the major regions predicted to base-pair with pal* included the 380, SL1, 444 and 500-regions (Figure 12B). They represent 22, 20, 30, and 12% of the monomer pool sequences respectively (Table 3).

Another feature revealed by mfold analysis was that the structural environment of SL1 differs between dimer and monomer pools and correlates with the pal*-defined structural groups. In the dimer pool, SL1 is often presented as an individualized stem-loop structure (80% of the dimer pool sequences; Table 3) and (Figure 12A), while pal* interacts with sequences located upstream of itself, either the C-box element or the 380-region (27% and 45%, respectively; Table 3).

Although some pal* binding partners are located upstream of pal* (25%; Figure 12B; Table 3), monomer pool sequences are characterized by increased interaction of pal* with sequences located downstream of SL1, either with the 444-region or the 500-region (30% and 12%, respectively; Figure 12B; Table 3). The direct consequence is a local increase of SL1 stability, which may regulate HIV-2 RNA dimerization (see Discussion).

Interestingly, 20% of both monomer and dimer pools sequences optimal structures sequester SL1 by direct base pairing with part of SL1 (pal*-SL1; Table 3; Figure 12). The pal*-SL1 interactions could be further divided in two subgroups, depending on whether pal* interacts with the 5' stem of SL1 or its loop and 3' stem (Figure 12, compare the pal*-SL1 structures in panel A and B). Moreover, the pal*-(5' stem SL1) interactions are more represented among dimer pool sequences than among monomer pool sequences (13% and 7%, respectively; Table 3).


A
pal*-C-box
pal*-380-region
pal*-SL1 (5'stem)
SL1
5'
3'
C- box
PBS
pal*
G- box
203
358
380
390
409
436
444
490
500
514

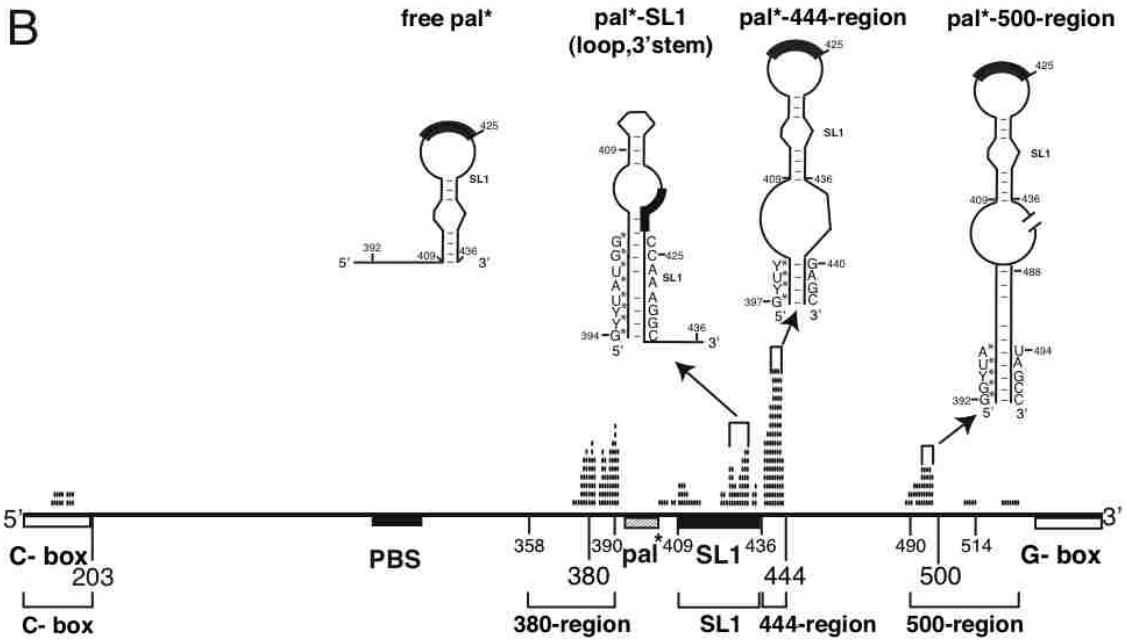

B
free pal*
pal*-SL1
(loop,3'stem)
pal*-444-region
pal*-500-region
SL1
5'
3'
C- box
PBS
pal*
G- box
203
358
380
390
409
436
444
490
500
514
C- box
380-region
SL1
444-region
500-region

Figure 12. Summary of the base-pairing interactions of *in vitro* selected pal* in the predicted optimal 1-561 RNA structures from the last dimer pool (panel A) and monomer pool (panel B) individual sequences.

A. The secondary structures of the selected RNAs from the dimer and monomer pools were predicted using the mfold program, v2.3 at 55˚C (Zuker, 2003). The base-pairing partners of pal* (nts 392-401) were taken from the most stable predicted structure for each of the 120 cloned sequences. Nucleotides predicted to interact with pal* in all of the most stable secondary structures are indicated by dots above the leader region schematic (*i.e*., one dot equals one base-pair with one pal* nucleotide). The high peaks of dots represent the major partners of pal*, and include the C-box (nts 186-199), the 380-region (nts 358-390), SL1 (nts 409-436), the 444-region (nts 437-444) and the 500-region (nts 490-514). The major pal* binding regions for both selections are indicated with brackets below panel B. Secondary structure of pal*, SL1, and pal*-interacting partners for the structural groups are indicated above the dots schematic in both panels. Only the pal* nucleotides that are consistently base-paired in each structural group are indicated. Y indicates a pyrimidine residue (C or U).

Dimer pool selection	
pal*- C-box	
193 T A C C C T C G T G T G T G	
D01/45/60 G T A G G A G A G A	
D17 G G A G A G A C A T	
D54 G G G A G A G A T A	
D23 G T G G A G A G A T	
D41 G G G A G C A T A A	
D51 G G G A G C A T T A	
D29 G G G A G C A A T G	
D30 G G G A G T A C A A	
D31/38 G G A G C A C A C A	
D56 G G G A G C T A A C	
D04 G G G A G A G A T A	
D39 G T G C A C T T A G	
D03 G G G G C G T T T A	
pal*- 380-region	
392 G C A G C A C C A A A C A A G G A	
D14 G T G G T T T A C A	
D02/34 G T G G T G C T A T	
D22 G T G G T C A T A C	
D10/21 G G T T T G C T T A	
D19 G G T T G G A T G C	
D44 G G T C A C G T A C	
D42 G C G T T G G T T A	
D06 G G A A G G T T A T	
D40 G A T G G T T C G T	
D58 G A G G T T C A G T	
D50 G C C T G G T T T T	
D18 G A T C A G G G T T	
D57 G C G A A T G T G T	
D12 G T C G T G G A A T	
D16 G T A T C G T A T G	
D05 G G A T G A C G T A	
D48 G A T G C G T G T A	
D59 G G A C T A T C A A	
D35 G G G A T G T T A A	
D46/47 G C C T A C A A C T	
D25 G C C C A C A T A G	
D09/11 G C C G T C T T C G	
D53 G T G C T C A A A G	
pal*- SL1 (5' stem)	
419 A G C C G G G C G C G G A A A G A T C	
D49 G C G A G G A T T T	
D13 G C G C A A C T T T	
D37 G C G T C C T T C G	
D24 G A T T T T T G C G	
D28 G G A C T G T G T A	
D07 G G T C C T T G C A	
D36/43 G C C T G G T T T T	
pal*- SL1 (loop, 3' stem)	
434 C G A C G G A A A C C	
D32/33 G G A A A T G T G G	
D55 G T G A G A G A A T	
D08 G C G A T C T T G A	
pal*- 444-region	
451 G A G G A G G G C G A G G T G T G C G	
D26 G G A C T C C T T T	
pal*- 500-region	
498 C C G A T G T C A A	
no sequence	
pal*-none (free pal*)	
no target	
D20 G A G A T T T A G C	
D27 G A T T T T A A A C	
D15 G A A C A A G A G A	
D52 G A T G A G T A A A	

Monomer pool selection	
pal*- C-box	
193 T A C C C T C G T G T G T G	
M18 G A G T T G G G C T	
M33 G A G T T T G G G A	
pal*- 380-region	
392 G C A G C A C C A A A A C A A	
M25 G G G T A C T T C T	
M28 G T C C G G T A T G	
M53 G T G G A T G T A C	
M20 G A T C C T G G A T	
M56 G A T C G T G A A T	
M32 G C A A A T C G T C	
M01 G G C A T C G A G T	
M51 G A G T A T G C A G	
M58 G G A A C G G T T G	
M23/43 G T T T T T A A T G	
M16 G T G T G A T C T A	
M12 G G G T G T T A A T	
pal*- SL1 (5' stem)	
419 A G C C G G G C G C G G A A A G	
M07 G G T C C A T G T C	
M26 G C G C A C A T T T	
M29 G A C T A T G C A T	
M50 G G C T T C T C T T	
pal*- SL1 (loop, 3' stem)	
434 C G A C G G A A A C C A	
M04/15 G A G T T T A T G G	
M54 G G T C A A T G G T	
M11 G G C C T T A A G G	
M13 G C G C T C G A T A	
M55 G C A G A C G G A T	
M08 G C G T G C A C G T	
M14 G G C C T T G G T A	
pal*- 444-region	
451 G A G G A G G G C G A G G T G T G C G A C	
M10 G C T T T G T T A G	
M40 G C T C T T T G A G	
M21 G T T T G A C A A T	
M24 G T T T C A T T T A	
M27 G C A T T G C T C T	
M05 G A C G A G C T T A	
M45 G T T C A C A T G C	
M52 G A C G T G T A G C	
M47 G T A C T A C C T T	
M41 G C A G A A G C A T	
M02 G A G C T A C A A A	
M38/42 G A C C T C C A C A	
M48 G A G T C C A T T C	
M37 G A A T C C A T T G	
M57 G G C C A T T T A T	
M34/39 G C A A G T A C G T	
pal*- 500-region	
498 C C G A T G T C A A	
M06 G G T T A C A A T T	
M44 G G T T A C C C A G	
M22/31 G G C T A A G T T C	
M36 G A G T C A T A G T	
M03 G T T T C G G T A G	
M59 G G A T A T C T G T	
pal*-none (free pal*)	
no target	
M30 G A C T G A A A C G	
M35 G A A T C G G C A A	
M19/46 G A A G C G A T A G	
M49 G A G C A T T A C A	
M17 G T A A C T T C C C	
M60 G G A A G T A A G T	
M09 G A T T A T A A T G	

Table 2. Structural groups of the individual pal* sequences from monomer and dimer pool selections.

M01-M60 and D01-D60 represent the selected pal* sequence of the 120 clones isolated from the final dimer (left panel) and monomer (right panel) pool selections written in a 5' to 3' orientation. The pal* sequence also includes the constant G392. The optimal (*i.e.* most stable) predicted secondary structures of the 120 individual clones were visually analyzed and the organization of the pal* and SL1 elements was recorded, allowing segregation of the 120 pal* sequences into distinct structural groups as described in Materials and Methods. Duplicated or triplicated sequences are represented by only one row (for example, D01/45/60). The sequence of the target region is indicated below the name of each structural group together with the number of the 3' nucleotide number, in a 3' to 5' orientation. The names of the individual clonal RNAs (M01-8 and D01-8) tested for dimerization are bolded. In order to highlight base-pairing consensus in each structural group, pal* nucleotides base-paired to target nucleotides in the predicted structures are aligned relative to the indicated target sequence. For example, a majority of sequences in the dimer pool's pal*-380-region group showed a 5'-GGTT motif base-paired to a target 3'-386-CCAA motif.

Structural groups		Pool selection	
		Dimer	Monomer
		(%)	(%)
pal*-C-box		27	3
pal*-380-region		45	22
pal*-SL1	SL1: 5' stem	13	7
	SL1: loop, 3' stem	7	13
pal*-444-region		1	30
pal*-500-region		0	12
pal*-none (free pal*)		7	13
Total		100	100

Table 3. Proportional representation of each structural group in the final dimer and monomer pools.

The percentage are obtained from the inventory of the structural organization of the pal* and SL1 elements in the mfold-predicted optimal structures of the individual monomer and dimer pools selected sequences (60 each; see Table 2 and Materials and Methods for details).

2.4.3. Characterization of the RNAs isolated from the last dimer and last monomer pools

Sixteen sequenced DNA plasmid clones, chosen from the last dimer and monomer pool selections were transcribed into pure clonal RNAs. These clonal RNAs were characterized for their dimerization properties by non-denaturing gel electrophoresis. The HIV-2 leader RNA was shown previously to be dimerization-impaired (Lanchy et al, 2003a). As expected, the dimerization yield of all eight clonal RNAs (M01-8) from the monomer pool selection was quite similar to the wt (Figure 13A, compare lanes 1 and 2-9) whereas the level of dimerization for six of the eight clonal RNAs (D01-6) from the dimer pool selection was higher than wt (Figure 13B, compare lanes 1 and 2-7). The dimerization efficiency of the other two clonal RNAs (D07 and D08) was similar to wt (Figure 13B, compare lanes 1 and 8-9). Although D07 and D08 RNAs were isolated from the last dimer pool selection, their impaired dimerization yields suggest that either they were selected as heterodimers with more dimerization-competent RNAs, such as D01-6 (Figure 13B), or that the stringency of the selection allows carryover of some monomer RNA species at each round. In fact, the re-partitioning of RNAs into dimer and monomer bands at each round, even though the RNA at each round is derived from template originating from solely dimer (or monomer) RNA in the previous round, suggests that the RNAs can adopt dimerization-competent or -impaired conformations.

Finally, analysis of the structural groups to which the 16 clonal RNAs belong, showed the same correlation between structural groups and dimerization phenotypes as the one we observed for the 120 sequenced individual clones (Table 2). Moreover, D07 and D08 belong to the pal*-SL1 group, a group represented equally in both monomer and

dimer pools (20%; Table 3), reinforcing the hypothesis that these were selected in the dimer pool as heterodimers or that they emerged through the selection despite their marginal dimerization capacity (see above).

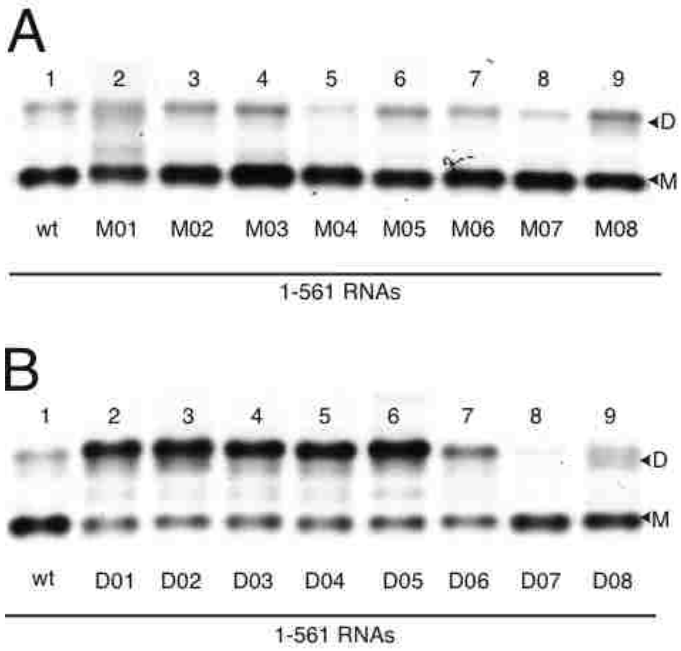


Figure 13. Characterization of selected RNAs from each of the last dimer and monomer pools.

Eight RNAs from each pool were selected to characterize their behavior in standard dimerization assays. A. Clonal RNAs from the final monomer pool are represented by M01-M08. Lanes 2-9 show the dimerization-impaired RNAs and lane 1 shows the wt RNA. B. Clonal RNAs from the final dimer pool are represented by D01-D08. Lanes 2-7 show the dimer-competent RNAs (D01-06). Lanes 8 and 9 show the dimerization-impaired D07 and D08 RNAs that persisted during the selection for dimerization-competent RNAs and lane 1 represents the wt RNA.

2.4.4. Wild type 1-561 HIV-2 RNA solution structure probing

To examine the conformation(s) of the wt HIV-2 leader RNA in solution, we performed chemical probing of the accessibility (reactivity) of the adenine residues in dimer buffer at 55˚C over time (0, 4, or 20 min; Figure 14A). Since the level of tight

dimers is low over the first 20 minutes of folding at 55˚C (Lanchy et al, 2003a), this experiment was designed to investigate the conformation of the monomeric RNA species together with the SL1-mediated kissing loop dimer species, which can also form under these conditions. Because solution structure probing yields data on the average accessibility of each nucleotide, the readout potentially represents the superimposed reactivities of several distinct conformations if more than one conformation exists in the reaction.

Indeed, most adenine residues of the pal and SL1 elements are rather accessible to dimethylsulfate (DMS) methylation (on the N1 atom of their Watson-Crick side; see (Ehresmann et al, 1987) for review)) (Figure 14). The most reactive adenine is A427 in SL1's loop and the most unreactive is A387 located upstream of pal (Figure 14). Furthermore, the A423 residue in SL1's palindromic sequence is weakly reactive, suggesting that a majority of RNA molecules may be involved in a "kissing" SL1 loop-loop interaction (loose dimers), since we know that only a small minority form tight dimers after 20 min of incubation (see Discussion and (Lanchy et al, 2003a)).

Finally, comparison of the change of reactivities over time (Figure 14A, lanes 2 to 4) indicated that the structure involving the pal and SL1 elements formed fast and did not change drastically during the first 20 min of folding, except that the A387 residue became unreactive after 4 min (Figure 14A). The chemical structure probing results are consistent with the extended SL1 model (Figure 14B).

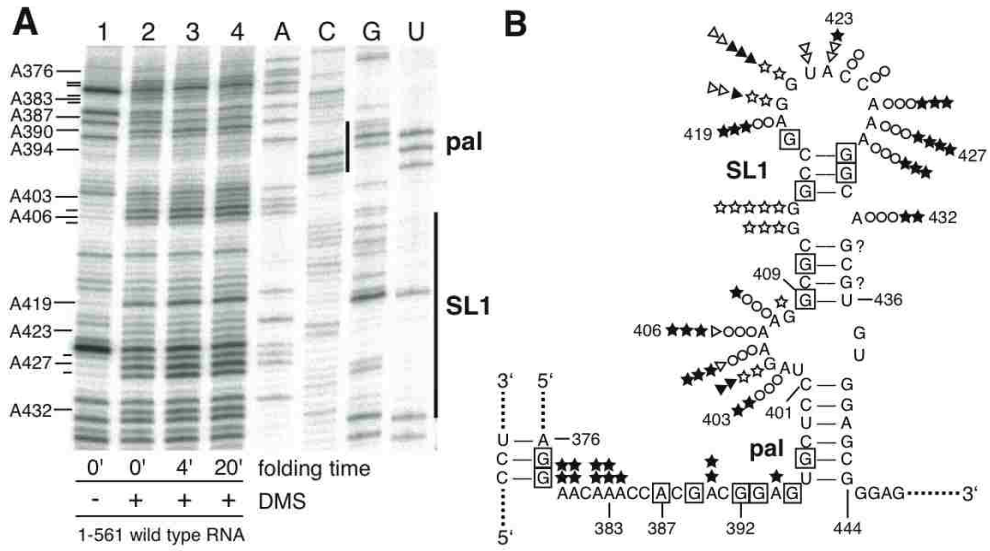


Figure 14. DMS probing of the pal and SL1 elements in the wild type HIV-2 leader RNA.

A. Wild-type 1-561 RNA was denatured, then incubated in the presence of dimer buffer at 55˚C for the times indicated. DMS was then added (0.5% final concentration) and incubation was continued at 27˚C for 2 min. Reactions were stopped and primer extensions were performed to visualize the sites and the intensities of adenine residues modifications as described in Materials and Methods (lanes 1 to 4). Lanes A, C, G, U represent the DNA sequencing of the plasmid DNA used to synthesize the RNA. The pal (nts 392-401) and SL1 (nts 409-436) regions are indicated by vertical lines along the DNA sequencing lanes. B. Summary of the principal nucleotide reactivities superimposed on a secondary structure model of the pal and SL1 elements for experiments done in this study (closed stars) or in previous publications: open stars for kethoxal probing of the guanine residues (Lanchy et al, 2004), open triangles for the single-strand-specific T1 nuclease (Damgaard et al, 1998), and open circles and closed triangles for DMS and T1 nuclease, respectively (Dirac et al, 2001). The boxed nucleotides indicate a lack of reactivity of the indicated adenines (this work) or guanines (Lanchy et al, 2004). The number of stars, triangles, or circles juxtaposed to a residue represent the relative levels of reactivity of this residue vis-à-vis the probing reagent (the stronger the reactivity, the greater the number of symbols).

2.5. Discussion

Among the compact genomes of the RNA viruses, the viral RNA encodes not only viral proteins, but also non-coding structural motifs that help to direct and regulate replicative functions. Because these regulatory entities need to be active only transiently, we and others have hypothesized that structural rearrangements of non-coding RNA lead to the alternate presentation and sequestration of signals in the HIV 5' untranslated leader region (Berkhout & van Wamel, 2000; Dirac et al, 2002a; Huthoff & Berkhout, 2001; Lanchy et al, 2003a; Lanchy et al, 2003b). We previously presented evidence that pal can interact with SL1 to inhibit HIV-2 RNA dimerization *in vitro*, but details of the interaction were unknown (Lanchy et al, 2003a). Because of the high degree of conservation of nucleotides in the region, comparative analysis of HIV-2 and SIV sequences was of little use. *In vitro* evolution of RNA molecules provides a convenient way to generate an artificial phylogeny of functionally similar molecules (Ellington & Szostak, 1990; Joyce, 1989; Tuerk & Gold, 1990).

By randomizing pal and selecting for molecules that were either dimerization-competent or -impaired, we could examine the ability of the pal region to regulate the SL1 dimerization signal without making *a priori* assumptions about structures involved. During the course of the study, we evolved pools of RNAs that had two distinct dimerization phenotypes. Mfold-mediated optimal secondary structure folding of 120 individual RNAs from these pools displayed patterns of intramolecular interactions with predictive value with respect to dimerization properties, at the level of both selection pools and individual clonal RNAs. In addition, similarities in the dimerization phenotype and predicted folding of the monomer pool RNAs with wt RNA suggest a general model

for pal's involvement in the regulation of SL1-mediated HIV-2 RNA dimerization *in vitro*.

Analysis of the lowest energy predicted folding structures for members of each evolved population provided general as well as specific insight into potential regulatory base-pairing interactions. The observation that most of the pal* interactions were local upstream and downstream base pairing was striking and in line with very recent observations of essential interactions of the pal in viruses in cell culture (see below). Although such predominance of local base pairing structures could not have been predicted, on theoretical grounds it is satisfying. To avoid large thermodynamic barriers against switching between conformations it is likely that regulatory conformational rearrangements are nearly isoenergetic and that switching between conformations should involve rearrangement of a limited number of base pairs; such criteria are most easily satisfied when base pairing changes are localized to a given domain. Furthermore, in the original population of randomized pal variants, the potential existed for pal* to hybridize anywhere along the leader region. The fact that discrete and predominantly local base pairing arrangements were found after selection suggests that either other arrangements were irrelevant to the selection or that the native folding of the 1-561 RNA limited the access of other regions of the leader RNA for hybridization with pal*, for instance the intrinsically stable TAR and PBS domains (encompassing nts 1-123 and 197-379, respectively (Berkhout & Schoneveld, 1993)).

Although most pal* base pairing involved short-range interactions with partners located less than 45 nts upstream and downstream (Figure 12), two exceptions were observed in both dimer and monomer pools. The first example is represented by 27% of

the dimer pool sequences (pal*-C-box; Table 3), where pal* was predicted to hybridize with the so-called C-box (nts 186-199; Figure 12A and Table 2), an element located between the poly A signal and PBS domains that we previously showed is capable of modulating dimerization through interactions with the G-box located at the 3' end of the leader RNA (Lanchy et al, 2003b). This result corroborated previous observations because pal* hybridizing to the C-box had the same effect on dimerization as mutagenesis of the C-box or hybridization of a small complementary oligonucleotide. We hypothesize that binding of pal* to the C-box increased the accessibility of SL1 for intermolecular interactions, thus explaining the enrichment of C-box-complementary pal* sequences (Table 2) in the dimer pool selection (27% versus 3% in monomer pool; Table 3).

The other example of a non-local base pairing interaction came from a structural group comprising 12% of the final monomer pool sequences (pal*-500-region; Table 3), where pal* was predicted to interact with the 500 region (nts 490-514; Figure 12B; Table 2). Although they are slightly different, we hypothesize that the structure centered around SL1 in pal*-500-region clones is functionally similar to the one in pal*-444-region clones with regard to their effect on SL1 presentation (Figure 12B). Indeed, the 445-485 region folds into two stable and conserved stem-loop structures (ψ1 and the major splice donor site stem-loop (Dirac et al, 2002b)) that, together with the pal*-500-region, form a constrained domain structurally extending SL1 in a way similar to the pal*-444-region fold. Such a constrained SL1 may impair the transition between monomers and tight dimers (see below), explaining the presence of the pal*-500-region selected RNAs in the monomer pool (12% versus none in dimer pool; Table 3).

The precise secondary structure(s) of the SL1 domain have not been definitively established for HIV-2 leader RNA, compared to the homologous region in HIV-1 (see below and (Shen et al, 2000; Shen et al, 2001)). Several secondary structure models have been proposed, some of which correspond to the structures predicted in this work. First, wt pal may interact with upstream nts 376-386 and form a stem-loop located between the PBS domain and SL1, similar to the pal*-380-region structure shown in (Figure 12A). This structure was proposed to form or to be part of a strong encapsidation signal for the HIV-2 genomic RNA (Griffin et al, 2001). Second, wt pal may interact with part of SL1, similar to the pal*-SL1 structures described in (Figure 12A). The altered presentation of SL1's palindromic sequence may lead to a decrease of its dimerization capacity (Lanchy et al, 2003b). However, our present study using pal*-randomized RNAs found that the pal interaction with the 5' side of SL1 was not frequently detected and, surprisingly, was slightly more represented in the final dimer pool (13% and 7% in dimer and monomer pools, respectively; Table 3). Third, the 3' end of wt pal may interact with a 5' GGAGC motif located downstream of SL1 ((Dirac et al, 2001; McCann & Lever, 1997) and Figure 14B) and form a structure similar to the one found in pal*-444-region species, which we called the extended SL1 (Lanchy & Lodmell, 2007).

Two interesting outcomes of this study are that the dimerization-impaired population is enriched with RNA species (30% of the monomer pool sequences; Table 3) showing a predicted, extended SL1 structure (*i.e.*, the pal*-444-region; Figure 12B) and that the wild type HIV-2 pal and SL1 elements may fold into a similar structure (Figure 14B). Wild type RNA, in line with this observation, forms tight dimers rather inefficiently, to an extent similar to the yields of monomer pool RNA species (Figure

11B) and (Figure 13A), yet SL1 in the extended SL1 structure seems clearly presented in as an individualized stem-loop with its palindromic sequence in the apical loop (Figure 14B). This apparent paradox between a presentation of SL1 apparently optimal for dimerization and an impaired dimerization phenotype may be explained if one considers both the mechanism of SL1-mediated tight dimerization and the relative stabilities of the SL1 structures. Although the mechanism of the transition between monomers and tight dimers is still a matter of discussion, studies using a 39-mer HIV-1 SL1 or larger leader RNA fragments revealed that this transition could be regulated by the stability of the stem of SL1 during the transition from kissing loops dimers to tight dimers *in vitro* (Baba et al, 2001; Bernacchi et al, 2005; Huthoff & Berkhout, 2002; Laughrea et al, 1999; Mujeeb et al, 2007; Muriaux et al, 1996b; Shubsda et al, 1999; Takahashi et al, 2000). Accordingly, the increased stability of extended (-11.3 kcal/mol) versus short (-7.5 kcal/mol) HIV-2 SL1 structures could have a similar negative effect on tight dimerization efficiency. Furthermore, the extended SL1 structure is even more stable in HIV-2 than the comparable 39-mer HIV-1 SL1 (-11.3 and -6.4 kcal/mol, respectively; mfold v2.3, 55˚C; data not shown), which could increase the energy barrier of the transition from kissing to tight dimers.

Although the extended SL1 structure characterized only one of the different RNA structural models revealed during the *in vitro* selections, recent results suggest that this conformation of the SL1 domain is functionally important *in vivo*. In particular, we showed that the formation of the extended SL1 structure is necessary for HIV-2 replication and genomic RNA encapsidation (Lanchy & Lodmell, 2007). Our experiments showed that one important role of the 3' end of pal is to base-pair with the 5'-

GGAGC motif located downstream of SL1 and that this interaction is necessary for an effective genomic RNA encapsidation to occur (Lanchy & Lodmell, 2007). Although a strong selective pressure during viral replication for an adenine at position 394 was observed upon mutation to a guanine residue (Lanchy & Lodmell, 2007), the precise role of the 5' end of pal is less understood. Nevertheless, it is notable that the extended SL1 structure is surrounded by two 5'-GGAG motifs, including one in pal (Figure 14B). It is possible that these purine motifs might be involved in the specific RNA-protein interactions taking place during genomic RNA selection and encapsidation. Likewise, deletion mutagenesis studies suggest that an extended SL1 element is involved in encapsidation and dimerization of the SIVmac239 genomic RNA (Strappe et al, 2003; Whitney & Wainberg, 2006).

In conclusion, our study shows that a few nucleotide changes in the pal element of HIV-2 leader RNA leads to altered dimerization phenotypes *in vitro*. The efficiency of SL1-mediated tight dimerization seems to correlate with the folding of the pal and SL1 elements, especially with the degree of structural constraint around the SL1 element. Extrapolating to wild type, our results suggest that the monomeric form of HIV-2 leader RNA adopts several local conformations that regulates the usage of SL1 as a dimerization element (Figure 15). Factors that contribute to the transition from dimerization-impaired to dimerization-competent RNA species possibly include viral proteins binding to this dynamic region, either Gag or NC or both (Figure 15) (Clever et al, 2000; D'Souza & Summers, 2005). Testing the protein-RNA interactions in this region represents our next endeavor (Figure 15).

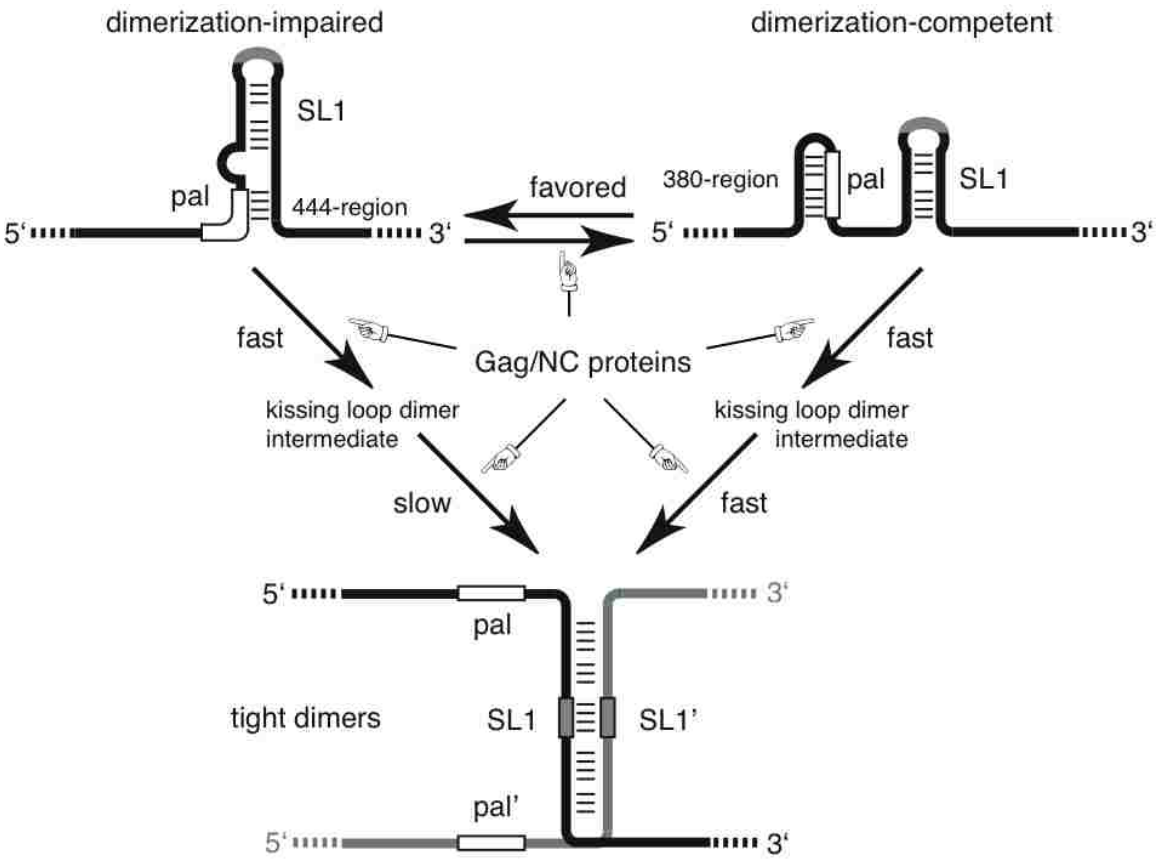


Figure 15. Model of pal's regulatory function in HIV-2 leader RNA SL1-mediated dimerization.

Different conformations of the pal and SL1 elements and different stages of the RNA tight dimerization process are represented. The top left structure corresponds to the extended SL1 structure where the 3' end of pal base-pairs to a sequence immediatly dowstream of SL1 (Figure 14B and (Dirac et al, 2001)). The top right structure shows individualized intra-ψ and SL1 small stem-loop structures (Figure 12A and (Griffin et al, 2001)). The dimerization-impaired phenotype for RNAs bearing extended SL1 may be due to a thermodynamically disfavored transition from kissing loop dimers to tight dimers, compared to dimerization-competent RNA species. However, one cannot rule out that a favored folding of extended SL1 may further amplify the dimerization-impaired phenotype of wild type HIV-2 leader RNA during *in vitro* dimerization assay. Finally, the virally encoded nucleocapsid protein NC and its Gag polyprotein precursor have RNA chaperone activity and may play a role in either the initial intramolecular folding or the different stages of SL1-mediated RNA dimerization.

2.6. Acknowledgements

This work is funded by the National Institutes of Health grant AI45388 to JSL. The plasmid pROD10 used as the original template to clone the wt 1-561 HIV-2 sequence (Lanchy & Lodmell, 2002) was provided by the EU Programme EVA/MRC Centralised Facility for AIDS Reagents, NIBSC, UK (Grant Number QLK2-CT-1999-00609 and GP828102). We would like to thank Dr. Michael Zuker for consultation using the several versions of mfold.

Chapter 3: Randomization and *in vivo* selection reveal a GGRG motif essential for packaging HIV-2 RNA

Tayyba T. Baig, Jean-Marc Lanchy & J. Stephen Lodmell[1]

Division of Biological Sciences, The University of Montana, Missoula, MT 59812, USA

[1]Corresponding author: J. Stephen Lodmell

This chapter is a modified version of the manuscript, accepted in the Journal of Virology in October 2008.

Baig, T.T., Lanchy, J.M. and Lodmell, J.S. (2008) Randomization and in vivo selection reveal a GGRG motif essential for packaging HIV-2 RNA. Epub ahead of print.

The experiments in this manuscript were done by the author of this dissertation.

3.1. Abstract

The packaging signal (ψ) of HIV-2 is present in the 5' non-coding region of RNA and contains a 10nt palindrome (pal; 5'-392-GGAGUGCUCC-3') located upstream of the dimerization signal stem loop 1 (SL1). Pal has been shown to be functionally important *in vitro* and *in vivo*. We previously showed that the 3' side of pal (GCUCC-3') is involved in base pairing interactions with a sequence downstream of SL1 to make an extended SL1, which is important for replication *in vivo* and the regulation of dimerization *in vitro*. However, the role of 5' side of pal (5'-GGAGU) was less clear. Here we characterized this role using an *in vivo* SELEX approach. We produced a population of HIV-2 DNA genomes with random sequences within the 5' side of pal and transfected these into COS-7 cells. Viruses from COS-7 cells were used to infect C8166 permissive cells. After several weeks of serial passage in C8166 cells, surviving viruses were sequenced. On the 5' side of pal there was a striking convergence toward a GGRG consensus sequence. Individual clones with consensus and non-consensus sequences were tested in infectivity and packaging assays. Analysis of individuals that diverged from the consensus sequence showed normal viral RNA and protein synthesis but had replication defects and impaired RNA packaging. These findings clearly indicate that the GGRG motif is essential for viral replication and genomic RNA packaging.

3.2. Introduction

The 5' leader RNA of retroviruses is involved in the regulation of many essential steps of the retroviral replication cycle. These steps include transcription, splicing, translation, genomic RNA dimerization, packaging, and reverse transcription (Abbink & Berkhout, 2008; Buck et al, 2001; Clever et al, 1999; Laughrea et al, 1997; McBride & Panganiban, 1996; McCann & Lever, 1997; Miele et al, 1996; Ohlmann et al, 2000; Patel et al, 2003). RNA packaging is a critical step during which two molecules of full length genomic RNA are encapsidated into the budding virion in the form of a dimer (see review; Rein, 1994). Furthermore, packaging is a highly selective and specific process in which unspliced genomic RNA is preferentially incorporated in viral particles over a vast excess of cellular mRNAs and spliced viral RNAs (Linial & Miller, 1990; Luban & Goff, 1994). This selectivity occurs through the interaction between trans-acting Gag polyprotein and cis-acting elements in the genomic RNA. The cis-elements are known as packaging signals (psi or Ψ) and have been mapped in the 5' leader region of unspliced genomic RNA in many retroviruses (Katz et al, 1986; Linial & Miller, 1990; Rein, 1994).

The packaging signal of HIV-1 has been extensively studied. Deletion analyses have found that the most important packaging signal (Ψ core) of HIV-1 genomic RNA lies between the major splice donor site (SD) and the gag initiation codon in the leader region (Aldovini & Young, 1990; Clavel & Orenstein, 1990; Lever et al, 1989). *In vitro* studies have suggested that the HIV-1 packaging signal is composed of at least four stem loops located both upstream and downstream of SD (Berkowitz & Goff, 1994; Clever et al, 1995). Although the HIV-1 Ψ core is located downstream of SD in the leader region, other packaging signals have been characterized either upstream of SD (Kim et al, 1994)

or even outside the 5' leader region (Berkowitz et al, 1995a). Thus, the full HIV-1 packaging signal is a complex, compound structure.

The HIV-2 packaging signal has been less extensively studied and conclusions about the exact location of the Ψ core have been more controversial (Garzino-Demo et al, 1995; Griffin et al, 2001; McCann & Lever, 1997; Poeschla et al, 1998). Several studies demonstrated the importance of sequences located downstream of SD (Garzino-Demo et al, 1995; Poeschla et al, 1998). However, other studies showed that deletion of a 28nt sequence located upstream of SD caused a severe packaging defect for HIV-2 genomic RNA packaging while mutation of sequences downstream of SD had mild effects on packaging (Griffin et al, 2001; McCann & Lever, 1997). The unusual localization of the HIV-2 Ψ core upstream of SD suggested that unlike HIV-1, the strongest HIV-2 packaging signal is present in both unspliced and spliced viral RNAs (McCann & Lever, 1997). In addition, a 10nt palindrome sequence called pal (5'-$_{392}$GGAGUGCUCC$_{401}$) was described within the 28nt Ψ core sequence, just upstream of the SL1 dimerization signal (Lanchy et al, 2003a) (Figure 16B). Mutations of the pal element revealed its involvement in RNA dimerization both *in vitro* and *in vivo* (Baig et al, 2007; L'Hernault et al, 2007; Lanchy et al, 2003a; Lanchy & Lodmell, 2007).

Since the pal motif is present in the HIV-2 packaging signal and can interfere with HIV-2 leader RNA dimerization *in vitro*, its role during viral replication was investigated. Deletion and substitution studies indicated that pal is required for efficient genome packaging and viral replication (Lanchy & Lodmell, 2007). Reversion analysis of pal mutant viruses in long term culture suggested that the 3' side of pal (GCUCC-3') interacts with a downstream sequence and forms an extension of SL1 called stem B that is

required for viral replication and genomic encapsidation (Lanchy & Lodmell, 2007). In parallel, an *in vitro* SELEX study found the 3' side of pal to be involved in the stem B structure that regulates SL1-mediated HIV-2 RNA dimerization *in vitro* (Baig et al, 2007). Finally, mutations in the pal/SL1 region showed pal to be the most important determinant for genomic RNA packaging and dimerization (L'Hernault et al, 2007).

These studies have clearly suggested a structural and functional role for the 3' side of pal *in vitro* and *in vivo*. To understand the role of the 5' side of pal, we used an *in vivo* randomization/selection methodology similar to one first employed by Berkhout and Klaver (Berkhout & Klaver, 1993) to study the role of the bulge and loop of the TAR element during viral replication in eukaryotic cells. Using this technique, we selected for replication-competent viruses from a population of viruses that were randomized in pal. We constructed a library of full-length HIV-2 proviral DNA genomes with random sequences in the 5' side of pal and selected viable viruses after transfection and subsequent infection of permissive cells. Our sequencing data showed that 5'pal nucleotides of the selected viable ("winner") viruses converged to the consensus sequence 5'-GGRGN. Analysis of individual clones that diverged from this sequence showed replication defects and impaired RNA packaging, while sequences of the type 5'-GGRGN restored these features. These data strongly support the hypothesis that pal is a critical packaging signal during the HIV-2 replication cycle.

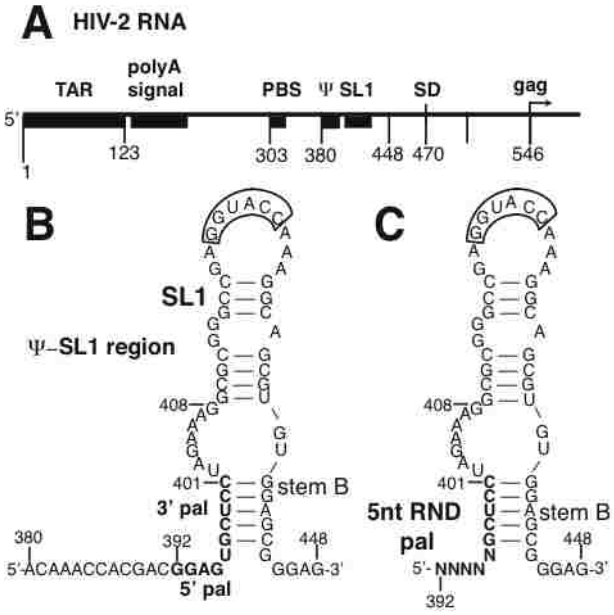


Figure 16. Schematic of the 5' untranslated leader region of HIV-2 genomic RNA.

A. The 5' leader region of HIV-2 RNA is represented with boxes and numbers to indicate the landmark sequences with their names indicated above. TAR, poly A signal, PBS, ψ, SL1, SD, and gag represent the trans-activation region, the poly (A) signal domain, the primer binding site, the packaging signal, the stem loop 1, the major splice donor site, and the 5' end of the Gag protein coding region, respectively. B. The secondary structure of Ψ-SL1 region (nucleotides 380-448) (Baig et al, 2007; Dirac et al, 2001; McCann & Lever, 1997). The pal element is a 10nt palindrome sequence (nts 392-401) within Ψ, located immediately upstream of SL1 and is represented by bold letters. Nucleotides 392-396 and 397-401 represent the 5'- and 3' sides of pal, respectively. The 3' pal sequence is involved in a base-pairing interaction with a sequence downstream of SL1 and creates stem B. The 5' side of pal is a purine rich motif upstream of 3' pal pyrimidine-rich sequence. The 6-nt autocomplementary sequence essential for SL1-mediated dimerization in the apical loop of SL1 is outlined. C. Five nts at the 5' side of pal (positions 392 to 396) were randomized in the insertion segment and inserted into the full-length proviral DNA as described in the text. The last five nts of pal (nts 397-401) were kept wild-type so as not to disrupt stem B.

3.3. Materials and Methods

3.3.1. Construction of plasmids for generation of randomized proviral DNA libraries

To rule out the chance of contamination by wild-type (wt) proviral DNA plasmid in the pal-randomized libraries, we used a parent plasmid called pSCR2, which harbors deleterious mutations in pal sequence as described previously (Lanchy & Lodmell, 2007) to construct our random library. pSCR2 was derived from modified pROD10 (a full-length plasmid containing infectious HIV-2 DNA sequence) and remained non-infectious even after several months in culture (Lanchy & Lodmell, 2007).

To create a vector for generating the randomized libraries, a derivative of pSCR2 plasmid called pSCR2AleIΔ(396-2030) was constructed, in which a fragment encompassing nucleotides 396-2030 of the viral genome including part of the non-coding region, and most of the Gag coding region up to the *Xho*I site was deleted and a nucleotide was substituted (A394T) to introduce a unique restriction site (*Ale*I) corresponding to nts 386-395. All nucleotide numbering in this study is referenced to the RNA sequence of HIV-2 (ROD isolate, GenBank M15390). The mutated vector sequence is as follows: 5'$_{380}$ACAAAC<u>CACGACGGtG</u>$_{395}$//$_{2030}$<u>CTCGAG</u>-3', where *Ale*I and *Xho*I sites are underlined and the small t indicates the changed nucleotide.

The pSCR2AleIΔ(396-2030) was constructed as follows. A fragment containing the long terminal repeat and the sequence up to nt 395 with the A394T substitution along

with 6-nt subsequent sequence of *Xho*I site was amplified by using sense primer (M13 Forward (-41) binding upstream of an unique *Aat*II site;

Table 4) and an antisense primer (asXhoI-AleI, containing the *Xho*I and *Ale*I sites;

Table 4). The amplified product was digested with *Aat*II/*Xho*I and ligated into the pSCR2 plasmid vector missing the original *Aat*II-*Xho*I fragment. The ligated product was transformed in *E. coli* DH5α cells, ampicillin resistant colonies were selected, followed by plasmid DNA purification.

3.3.2. Generation of pal-randomized proviral DNA library

To generate the proviral DNA library that is randomized at the 5' side of pal (5nt RND pal), we used an In-Fusion (Clontech) *in vitro* DNA recombination protocol, which obviates the need of a DNA ligase step and combines the insert and vector due to the homologous regions at both 5' and 3' ends of insert and digested vector. We constructed our vector and randomized insert with these homologous regions by the use of long primers as per the manufacturer's protocol.

To create the vector for use in the In-Fusion reaction, the pSCR2AleIΔ (396-2030) plasmid was digested with *Ale*I and *Xho*I and purified by agarose gel electrophoresis followed by gel extraction (5PRIME) prior to combining with the randomized insert.

To construct the randomized insert, a PCR product (nts 369-2050) was generated using a mutagenic sense primer (sPal-random-5;

Table 4) with degenerate nucleotides at positions 392-396 and an antisense primer (asROD2050;

Table 4) and using pSCR2 as a template. The PCR mixture contained 0.4 ng of pSCR2, 0.2 μM of the sense and antisense primers (

Table 4), dNTPs mix (0.2 mM), 1x Taq standard buffer, and 0.5 μl of Taq polymerase enzyme (New England Biolabs) in a final reaction volume of 50 μl. A 40 cycle PCR protocol with a 55°C annealing temperature was used. The PCR product (randomized insert) was purified by agarose gel electrophoresis.

The randomized insert was combined with the digested pSCR2AleIΔ(396-2030) vector at a 2:1 ratio and transferred to an In-Fusion reaction tube. The resulting recombined product was transformed into *E. coli* DH5α and DNA was extracted from bulk transformation reactions to obtain the randomized proviral DNA library. The degeneracy of pal region in the plasmids of this library was checked by sequencing (Table 5). Although there was a bias in oligonucleotide synthesis that favored the incorporation of purines somewhat over pyrimidines, each nucleotide was sufficiently represented to ensure adequate mutagenesis at each position in the reconstructed proviral DNA library.

3.3.3. Cell culture and transfections of 5nt RND pal library

COS-7 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, penicillin, and streptomycin (Invitrogen). COS-

7 cells were transfected with randomized library (5nt RND pal plasmids) using a Trans-IT-COS transfection kit (Mirus). Two days post-transfection, the cells and media were harvested and the presence of virus in the medium was monitored by p27-specific enzyme-linked immunosorbent assay (SIV p27 ELISA; Zeptometrix).

3.3.4. Cell culture and infections of RND viruses

C8166 cells (NIH AIDS reagent 404) were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% fetal calf serum, penicillin, streptomycin and glutamine (Invitrogen). The media of transfected cells containing 5nt-pal viruses was filtered through 0.22-µm-pore size filters. Viruses were collected from the media by centrifugation for 2 hours at 4°C and 8500 rpm (10,015 rcf). The resulting viral pellet was resuspended in 2.0 ml RPMI. Subsequently, 1.5 ml was used for infection in 10^5 C8166 cells. This 1.5 ml corresponded to 250ng of HIV-2 CAp27 for 5 nt RND pal viruses, as determined by ELISA. Aliquot of 0.5 ml was saved at -80°C for RNA extraction. After 12 h at 37°C, cells were washed and resuspended in fresh RPMI medium-fetal calf serum. Every 7-10 days, the viruses produced in the media were filtered through 0.22-µm syringe filters and were used to reinfect fresh uninfected 10^5 C8166 cells.

Virus propagation for 5nt RND pal library was monitored by quantifying the concentration of CAp27 protein (determined by ELISA; Zeptometrix) in the medium at different days post infection (d.p.i.) (data not shown). In addition, aliquots of medium

containing filtered viruses were saved at 4, 9, 15, and 41 d.p.i. for RNA isolation and analysis.

3.3.5. RNA isolation and analysis of non-selected and selected sequences from 5nt RND pal library

The total COS-7 intracellular vRNA and fractions of COS-7 and C8166 (19 and 41 d.p.i.) extracellular vRNAs from 5nt RND pal viruses were purified using a Stratagene Absolutely RNA microprep kit. To pellet the virus particles, saved aliquots were centrifuged for 2 hours at 4°C and 15,000 rpm (21,882 rcf). Purified RNAs were used for reverse transcription-PCR (AccuScript high-fidelity RT-PCR system; Stratagene) using sBamT7R and asEco561 primers (Table 4). An aliquot of each RT-PCR reaction was sequenced directly as a pool while the rest of the reaction was digested with *Bam*HI/*Eco*RI to clone individual sequences. The digested products were ligated with *Bam*HI and *Eco*RI sites of pUC18 vector and transformed in competent DH5α *E. coli*. Individual DNA plasmids were sequenced to examine the *in vivo* selection of the viable sequences and for further characterization.

3.3.6. Construction of individual plasmids with winner and loser sequences in the 5' pal region

To construct full-length plasmids with the selected and non-selected sequences in the 5' pal region, we employed the above-described In-Fusion strategy by using the same

vector but with different inserts that contained desired mutations in the 5' pal region. The inserts with non-selected "loser" (SM2 and TTB-16) and selected "winner" (TTB-78) sequences were constructed by using sense primers (sPal-SM2, sPal-TTB-16 and sPal-random-2, respectively (Table 4) and the antisense primer asROD2050 (Table 4) in separate PCR reactions. Since the wild-type (wt) sequence was one of the selected sequences (TTB-61 & 76, Table 7A & B), we used the modified pROD10 plasmid (Lanchy & Lodmell, 2007).

M13 Forward (-41)	CGC CAG GGT TTT CCC AGT CAC GAC
asXhoI-AleI	AAC TCG AGC ACC GTC GTG GTT TGT TCC TG
sPal-random-5	GGG CGG CAG GAA CAA ACC ACG ACN NNN NGC TCC TAG AAA GGC GCG GGC CGA GG
asROD2050	GGT GTC TCC CCC TGC TC
sBamT7R	TAG GAT CCT AAT ACG ACT CAC TAT AGG TCG CTC TGC GGA GAG
asEco561	AAG AAT TCA GTT TCT CGC GCC CAT CTC CC
sPal-SM2	GGG CGG CAG GAA CAA ACC ACG ACC CTC TGC TCC TAG AAA GGC GCG G
sPal-TTB-16	GGG CGG CAG GAA CAA ACC ACG ACA CCT GGC TCC TAG AAA GGC GCG G
sPal-random-2	GGG CGG CAG GAA CAA ACC ACG ACG GNG NGC TCC TAG AAA GGC GCG G

Table 4. Oligonucleotides used in chapter 3.
Sequences are shown in the 5' to 3' direction.

3.3.7. Transfections of non-selected and selected individual plasmids

Full-length plasmids with loser and winner sequences were transfected in COS-7 cells individually as described above.

3.3.8. Replication assay of non-selected and selected clones

200µl of medium from transfected COS-7 cells containing 0.6ng of HIV-2 p27 capsid protein, as determined by ELISA, was used to infect 10^5 C8166 cells. After 12 h, cells were washed twice and resuspended in fresh RPMI medium-fetal calf serum. Viral replication was followed for up to 16 days for the selected and non-selected individuals by quantifying the concentration of CAp27 protein (determined by ELISA; Zeptometrix) in the medium at d.p.i. (different days post infection).

3.3.9. Single round infectivity assay of non-selected and selected individuals

The infectivity of viruses in the medium of COS-7 cells was quantified using CMMT-CD4-LTR-βgal indicator cells. This cell line expresses the CD4 receptor and contains a β-galactosidase gene fused to the HIV long terminal repeat, thus detecting synthesis of the transactivating (Tat) protein after productive infection of the cells by the input virus (Chackerian et al, 1995). After two days of culture post-infection, the cells were stained for β-galactosidase activity, and the blue infected cells were counted. The numbers were normalized first to the amounts of input viruses, as determined by ELISA, and second to the level of infection by wt virus to give the infectivities of viruses relative to wt.

3.3.10. RNA isolation and RNase protection assay for non-selected and selected individuals

The COS-7 intracellular vRNAs were purified using a Stratagene Absolutely RNA miniprep kit. For extracellular vRNAs, a fraction (1ml) of the medium was centrifuged for 2 hours at 4°C and 15,000 rpm (21,882 rcf) to pellet the viral particles, followed by lysis and RNA purification. Purified RNAs were used separately for RNase protection assays (RPA III; Ambion), as described by the manufacturer. The RNase protection assay was carried out using a ^{32}P-labeled antisense RNA probe complementary to positions 401-562 of the HIV-2 RNA ROD isolate as described in (Lanchy & Lodmell, 2007). The antisense region was cloned into the Promega pGEM 7Zf(+) vector (Novagen) so that the T7 transcript had approximately 45 nucleotides of vector at its 5' end. The non-HIV-2 tail of the probe was used as a marker of the RNase digestion efficiency during the experiment.

3.4. Results

3.4.1. Generation and selection of 5nt RND pal library

We previously showed that a 10nt palindrome (pal) sequence (5'-GGAGUGCUCC-3'), located at the 3' end of Ψ was important for regulating SL1-mediated HIV-2 RNA dimerization *in vitro* (Lanchy et al, 2003a). Using SELEX methodology, we showed that this regulation is mediated by the formation of an extended SL1 structure created by the interaction of the 3' side of pal (GCUCC-3') with an immediate downstream sequence of SL1 (Baig et al, 2007). Moreover, formation of the extended SL1 structure is necessary during viral replication (Lanchy & Lodmell, 2007). Although both studies demonstrated the importance of the 3' side of pal, the role of the 5' side of pal (5'-GGAGU) was not well understood. Therefore, we have investigated the role of 5' side of pal in HIV-2 replication, by subjecting this region to randomization and subsequent *in vivo* selection.

The *in vivo* SELEX method starts with the generation of a pool of viral genomes containing many possible mutations at randomized sequences in a particular region (Berkhout & Klaver, 1993). We synthesized a pool of HIV-2 proviral DNA genomes with five randomized nucleotides of pal at positions 392-396 so that the DNA library harbored $_{392}$NNNNN$_{396}$ sequence (where N has an equal probability of being A, C, G, T; Figure 16C).

The recombinagenic method used for the generation of the 5nt RND pal library is detailed in the Materials and Methods. Briefly, a pSCR2AleIΔ(396-2030) vector and an insert (randomized at 5' pal) were assembled in an In-Fusion (Clontech) recombinagenic

reaction by virtue of the engineered homologous regions at their 5' and 3' ends. The recombined product was transformed in *E. coli* to produce a library of full-length HIV-2 proviral DNA plasmids with up to 1024 (*i.e.* 4^5) possible sequence variations at the 5' side of pal. To verify degeneracy of the library, a fraction of the bacteria transformed with the recombinagenic PCR products were counted and their plasmids sequenced. Sequencing of 38 individual clones (Table 5) revealed the presence of all four nucleotides at each position, although some purine bias in the synthesized oligonucleotides was noted. The diversity of the library allowed adequate interrogation of each position in the 392-396 pal region.

RND proviral DNA library					
	392	393	394	395	396
Clone					
WT	G	G	C	G	A
TTB83	G	G	C	G	A
TTB84	A	T	G	A	A
TTB85	C	G	A	G	A
TTB86	T	G	T	T	A
TTB87	C	A	T	G	T
TTB88	G	T	A	G	A
TTB89	T	A	C	A	C
TTB90	G	A	C	A	G
TTB91	C	C	A	G	T
TTB92	G	A	C	A	T
TTB93	G	A	T	G	A
TTB94	G	T	G	A	G
TTB95	G	G	A	A	C
TTB96	A	A	A	A	G
TTB97	G	G	G	G	A
TTB98	G	A	T	C	A
TTB99	G	A	G	A	A
TTB100	G	C	G	T	A
TTB101	G	G	T	A	G
TTB102	A	A	G	G	A
TTB103	G	A	A	A	A
TTB104	T	A	A	A	T
TTB105	A	C	G	T	C
TTB106	G	T	A	T	C
TTB107	G	G	A	G	G
TTB108	T	A	G	C	G
TTB109	G	A	T	C	G
TTB110	A	G	T	G	G
TTB111	A	A	A	A	G
TTB112	G	A	G	A	A
TTB113	G	G	A	T	C
TTB114	A	C	C	A	G
TTB115	C	G	C	G	T
TTB116	G	T	G	A	G
TTB117	A	A	G	A	T
TTB118	G	A	T	G	A
TTB119	G	C	A	G	A
TTB120	T	A	G	G	A

Table 5. Sequences of individual clones in the 5nt randomized proviral DNA library used for COS-7 cells transfection.

To initiate selection, the purified plasmid DNA library was first transfected into COS-7 cells to produce the library of 5nt RND pal viruses. To monitor the sequence diversity and to confirm that no individual sequence was dominating in the 5nt RND pal library at the transfection stage, we analyzed the leader sequences of intracellular and packaged genomic RNAs after RT-PCR. Sequence analysis of the RT-PCR product as a pool showed that COS-7 intracellular and packaged vRNAs contained random sequences at the 5' side of pal (data not shown). In addition, RT-PCR products were subcloned to obtain the sequence of individual clones. Sequence analysis of 25 and 21 individual clones from COS-7 intracellular and packaged (extracellular) vRNAs, respectively, confirmed that no single sequence was predominating in the library at this stage (Table 6A and B). However, a bias toward purines at nts 392-396 was already evident in the viral RNAs that were packaged and exported as viral particles into the media of the transfected COS-7 cells.

(Table 6A)

COS-7: intracellular vRNA nt number					
Clone	392	393	394	395	396
WT	G	G	A	G	U
TTB-1	A	G	C	U	U
TTB-2	A	A	U	U	U
TTB-3	U	C	U	A	A
TTB-4	A	G	G	A	A
TTB-5	G	G	G	A	G
TTB-6	C	A	G	G	G
TTB-7	G	G	C	C	G
TTB-8	A	U	G	C	U
TTB-9	A	A	G	U	C
TTB-10	A	A	A	A	A
TTB-11	A	G	G	A	A
TTB-12	A	G	A	A	C
TTB-13	U	A	A	C	G
TTB-14	G	G	C	A	G
TTB-15	U	A	G	G	U
TTB-16	A	C	C	U	G
TTB-17	U	U	A	G	U
TTB-18	G	U	G	G	C
TTB-19	G	U	U	U	U
TTB-20	A	A	G	G	C
TTB-21	G	G	C	A	G
TTB-22	A	C	G	G	G
TTB-23	A	C	U	G	G
TTB-24	U	A	C	A	G
TTB-25	G	G	A	A	G

(Table 6B)

COS-7: extracellular vRNA nt number					
Clone	392	393	394	395	396
WT	G	G	A	G	U
TTB-26	G	U	G	A	G
TTB-27	A	G	A	A	C
TTB-28	G	G	A	A	A
TTB-29	A	A	G	G	G
TTB-30	G	U	A	A	G
TTB-31	C	A	U	A	G
TTB-32	A	A	G	G	A
TTB-33	G	G	G	G	G
TTB-34	U	U	A	A	C
TTB-35	G	G	G	U	A
TTB-36	G	A	A	G	U
TTB-37	G	G	G	U	U
TTB-38	A	G	A	G	G
TTB-39	A	G	G	A	U
TTB-40	G	U	A	G	G
TTB-41	G	C	U	U	U
TTB-42	G	G	G	G	C
TTB-43	G	G	C	U	G
TTB-44	G	G	A	A	A
TTB-45	G	G	G	U	G
TTB-46	A	U	G	C	G

Table 6. Sequence of 5nt RND pal virus library inside and outside of COS-7 cells after transfection.

Viral RNAs extracted from COS-7 cells (intracellular) and media (extracellular) were subjected to RT-PCR, cloning, and sequencing as described in Materials and Methods. TTB-1 to 25 represent RNAs that were expressed intracellularly and may or may not be packageable. TTB-26 to 46 represent those RNAs that were exported as viral particles into the media of COS-7 cells and thus were at least marginally competent for packaging. The name of the individual clone (TTB-16) tested for infectivity and packaging is in bold text.

To study the selection of 5' pal sequences during viral replication and infection cycles, media from COS-7 transfected cells was used to infect permissive C8166 cells. Virus production was detected within a few days of infection in the media of C8166 cells by p27 capsid ELISA. Therefore, we harvested virus at 4 days post-infection, followed by vRNA purification, RT-PCR, and sequencing. Sequence analysis of the RT-PCR products showed selection at 4 d.p.i. for Gs at positions 392 and 393 (data not shown). Sequencing viral RNA from 9 and 15 d.p.i. showed that the selection was complete at position 392, and nearly complete for positions 393 & 395 (Table 7A). By 41 d.p.i., selection was obvious for the entire random region: viral RNAs had converged on the consensus sequence GGRGN (Table 7B), where N represents T, G or A at position 396 (no Cs were observed). The absence of C396 in the recovered viral RNAs may have been a result of negative selection, sampling artifact, or because of the relative under representation of C in the library. C396 viruses were present as packageable viruses in the post-transfection sampling (Table 6), so it appears that they are viable but are likely overrun during longer infection schemes (Table 6).

(Table 7A)

C8166: 9 d.p.i. nt number					
Clone	392	393	394	395	396
WT	G	G	A	G	U
TTB-47-48	G	G	A	A	A
TTB-49-50	G	G	A	G	G
TTB-51-52	G	G	G	G	G
TTB-53-54	G	G	A	G	A
TTB-55	G	G	G	U	A
TTB-56	G	G	G	C	G
TTB-57-58	G	G	G	C	A
TTB-59	G	G	U	A	G
TTB-60	G	G	U	A	A
TTB-61	G	G	A	G	U
TTB-62-65	G	A	G	G	G

(Table 7B

C8166: 41 d.p.i. nt number					
Clone	392	393	394	395	396
WT	G	G	A	G	U
TTB-66-71	G	G	A	G	A
TTB-72-75	G	G	A	G	G
TTB-76	G	G	A	G	U
TTB-77	G	G	A	A	G
TTB-78-80	G	G	G	G	U
TTB-81	G	G	G	U	A
TTB-82	G	G	G	U	G

Table 7. RNA sequences of those viruses that were selected after several rounds of infection in C8166 cells.

Viral RNAs extracted from viruses at 9 d.p.i. (Table A) and 41 d.p.i. (Table B) d.p.i. were subjected to RT-PCR and cloning as described in Materials and Methods. TTB-47 to 65 (9 d.p.i.) and TTB-66 to 82 (41 d.p.i.) clones were sequenced in the 5' pal region. Those sequences which appeared two or more times are shown in only one row (e.g., TTB-66-71). The name of the individual clones (TTB-76 (wt) and TTB-78) tested for infectivity and packaging are in bold text.

3.4.2. A purine rich sequence at the 5' Side of pal is important for viral replication

In wt HIV-2 the 5' side of pal contains a purine rich sequence (5'-GGAGU) and the *in vivo* 5nt RND pal selection demonstrated the importance of the 5'-GGRGN motif. We therefore compared viral replication properties of pyrimidine rich non-consensus and purine rich consensus clonal viruses. The pyrimidine rich clones used in this experiment were TTB-16, a non-surviving intracellular COS-7 sequence (5'-accug, where lower case

letters signify nucleotides different from the wt sequence; Table 6A) and SM2 (5'-cuccU; Figure 17), a replication-deficient 5' pal pyrimidine rich mutant previously described by Lever and co-workers (L'Hernault et al, 2007). The winner viruses were the purine rich clones TTB-76 (5'-GGAGU (wt sequence); Table 7B) and TTB-78 (GGgGU; Table 7B).

Duplicate samples of these four DNA constructs were transfected into COS-7 cells. Forty-eight hours after transfection, viruses were harvested and used to infect C8166 cells with an equivalent amount of p27 capsid, as measured by ELISA. Virus spread was followed by measuring p27 capsid in the C8166 media by ELISA. Infections with purine rich TTB-76 and TTB-78 viruses led to robust virus production, whereas those with pyrimidine rich SM2 and TTB-16 sequences showed low or no virus production for up to 16 d.p.i. (Figure 17). These results support our hypothesis that the purine rich 5'-GGRGN in the 5' side of pal is essential for viral replication.

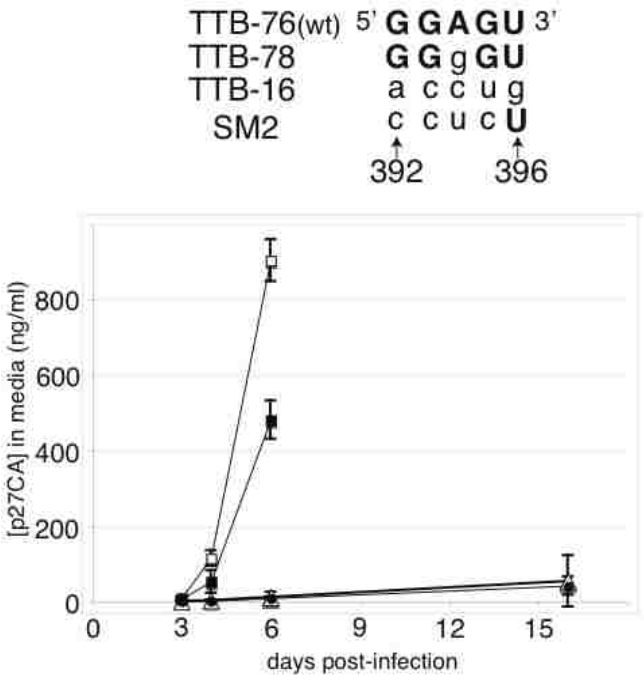


Figure 17. Mutant HIV-2 viruses harboring pyrimidines at the 5' side of pal exhibit replication defects.

(Top) Sequence at 5' side of pal (nts 392-396) in wt and mutant individuals used in this experiment. TTB-76 (wt) and TTB-78 contain the pal sequence derived from viral clones that were winners after extended serial passage (Table 7B). TTB-16 represents a sequence that was observed intracellularly after transfection but that did not survive during serial infections in C8166 cells (Table 6A). SM2 is a packaging-deficient mutant previously characterized by the Lever group that was used here for comparison with our own clones containing pyrimidines in the 5' side of pal (L'Hernault et al, 2007). The mutated residues are represented by lowercase letters. (Bottom) Replication kinetics of wt (closed squares), TTB-78 (open squares), TTB-16 (open circles) and SM2 (open triangles) viruses in C8166 cells. Equal amounts of viral particles (0.6ng of p27 capsid, as determined by ELISA), produced in the media by transfection of DNA plasmids in COS-7 cells were used to infect permissive C8166 cells. Virus production from these cells was assayed for the production of p27 capsid at intervals post infection. Results are based on infections of separate plates of permissive cells using virus harvested from two independent transfections for each clone. Error bars represent one standard deviation for a representative experiment.

3.4.3. A purine rich sequence (GGRGN) at the 5'side of pal is important for genomic RNA packaging

With the aim of further identifying which steps of HIV-2 viral replication in C8166 cells were affected by non-wt substitutions in the 5' side of pal, viruses produced from transfection of COS-7 cells with SM2, TTB-16, TTB-76 (wt), and TTB-78 proviral DNAs were assessed for infectivity/entry in CMMT-CD4-LTR-βgal indicator cells. The infectivity of these cells can be examined by the detection of a Tat-driven reporter gene whose expression is dependent on HIV entry into the cell (Chackerian et al, 1995). As observed in C8166 cells, SM2 and TTB-16 also showed a lower level of infectivity in CMMT-CD4-LTR-βgal indicator cells as compared to TTB-76 and TTB-78 (Figure 18A).

Since the packaging signals in retroviruses contain purine rich motifs (Berkowitz et al, 1996) and the 5' pal element is both purine rich and located within the HIV-2 packaging signal (Griffin et al, 2001; Lanchy & Lodmell, 2007), we investigated the role of 5' side of pal as an important RNA packaging element. We assessed and compared the packaging of RNAs containing consensus 5' pal sequence (TTB-76 (wt) and TTB-78) with packaging of RNAs containing the non-consensus 5' pal sequence (TTB-16 and SM2). We performed RNase protection assays (RPA) on TTB-16, SM2, TTB-76 (wt) and TTB-78 RNAs extracted from the cytoplasm of COS-7 transfected cells and from the COS-7-produced viruses. As seen in (Figure 18B), neither non-wt consensus nor the non-consensus sequences affected the level of genomic RNA in the cytoplasm of the transfected cells relative to wt. However, virions produced by these transfected cells

showed that the level of genomic RNA packaging was high in the consensus 5nt RND pal viruses (TTB-76 and TTB-78) but was reduced in the pyrimidine rich 5' pal viruses (TTB-16 and SM2) (10-15%) as compared wt (Figure 18C). Further, to ensure that mutations in 5' side of pal did not affect viral gene expression, we analyzed viral proteins from intra- and extracellular COS-7 fractions using western blot analysis. Our data show that mutation of 5' pal region does not affect viral gene expression (Figure 19). These data taken together suggest that the replication defect seen in viruses with a pyrimidine-rich 5' pal sequence is due to a packaging defect.

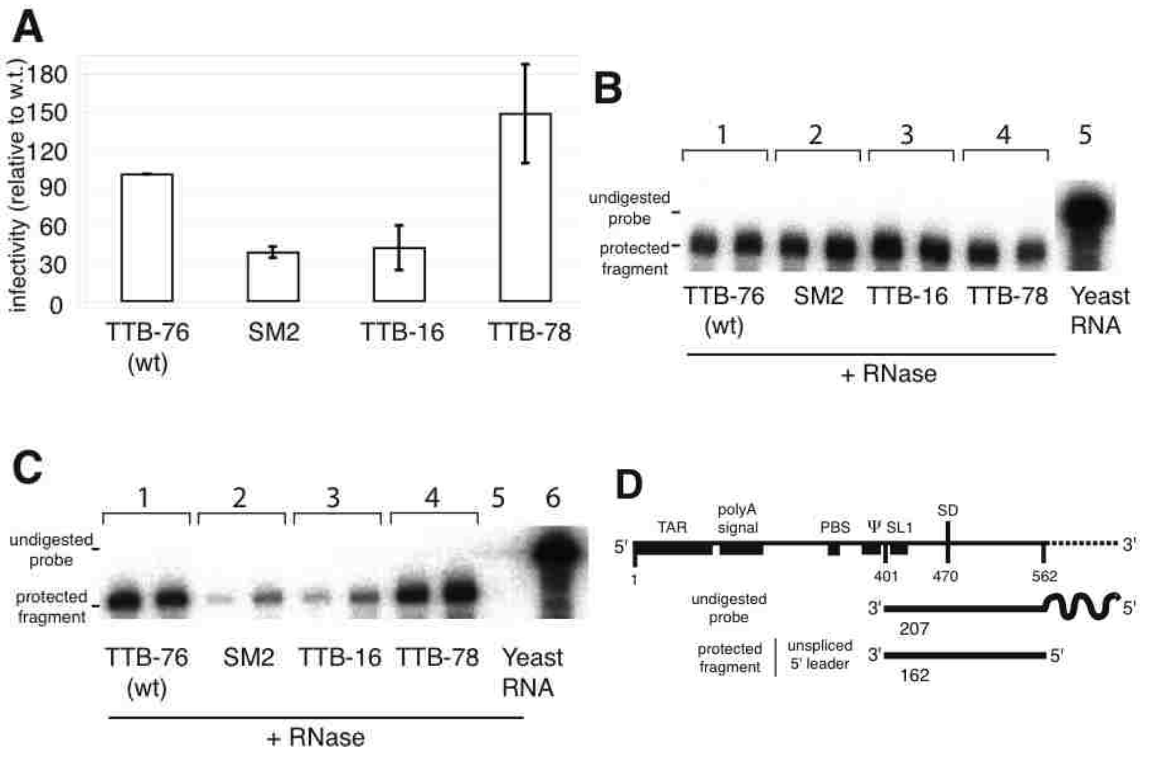


Figure 18. HIV-2 virus infectivity and genomic RNA packaging are decreased by the presence of pyrimidines at the 5' side of pal.

A. Single-round infectivity assay for wt and 5' pal mutant viruses isolated from transfected COS-7 cells using CMMT-CD4-LTR-βgal indicator cells (see Materials and Methods). Sequences of wt and mutant 5' pal for the viruses used in this experiment are shown at the top panel of (Figure 17). Virus generated by transfection of COS-7 cells was quantified by p27 ELISA and used to infect the indicator cells. Forty-eight hours post-infection, cells were stained to visualize cellular β-gal expression. Infectivity was normalized to the infectivity of wt virus, set at 100%. Infectivity was measured in two independent experiments and results are shown with error bars that represent the standard deviation. B. Quantification of full-length genomic RNA in the cytoplasmic fraction of COS-7 cells transfected with wt and 5' pal mutants using an RNase protection assay. Cells were harvested from two independent transfections of wt and 5' pal mutants. Three microgram of cytoplasmic RNA was probed with $2x10^6$ cpm of labeled riboprobe. Lanes 1, 2, 3 and 4 represent amounts of genomic RNAs from TTB-76 (wt) (5'-GGAGU), SM2 (5'-ccucU; (L'Hernault et al, 2007)), TTB-16 (5'-accug) and TTB-78 (5'-GGgGU) viruses, respectively. Lanes 5 represents yeast RNA used as a non-specific target RNA control, incubated without RNases. C. Quantification of full-length genomic RNA in wt and 5' pal mutant viruses using an RNase protection assay. Viruses were harvested from the media of transfected COS-7 cells from two independent transfections of wt and 5' pal mutants. First, the extracted RNAs form these viruses were normalized to the amount of viral particles (p27 capsid as determined by ELISA) and an equivalent amount of these RNAs were probed with $5x10^5$ cpm of labeled riboprobe. Lanes 1, 2, 3 and 4 represent amounts of genomic RNAs from COS-7 cells transfected with TTB-76 (wt) (5'-GGAGU) and 5' pal mutants (SM2 (5'-ccucU; (L'Hernault et al, 2007)), TTB-16 (5'-accug) and TTB-78 (5'-GGgGU)), respectively. Lanes 5 & 6 represent yeast RNA used as a non-specific target RNA control, incubated with or without RNases, respectively. D. Schematic representation of the radioactive probe used in this experiment and a predicted protected fragment corresponding to the unspliced leader region.

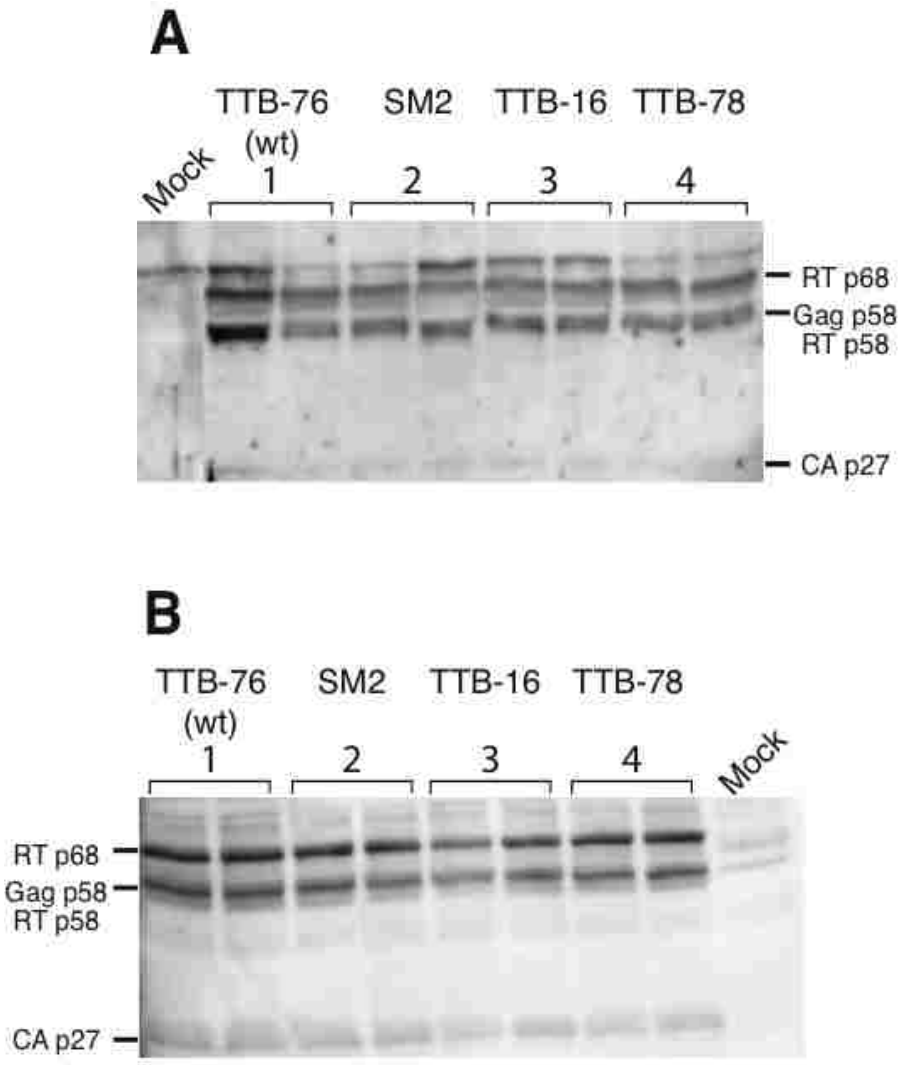


Figure 19. Mutations within the 5' side of pal do not affect viral gene expression.

A. Western blot analysis of COS-7 extracellular viral proteins. Protein samples from equivalent number of virions, as determined by p27 capsid ELISA were loaded in each lane. HIV-2 proteins were detected with human serum from an HIV-2 infected patient. Lanes 1, 2, 3 and 4 represent proteins from TTB-76 (wt) (5'-GGAGU), SM2 (5'-ccucU; (L'Hernault et al, 2007)), TTB-16 (5'-accug) and TTB-78 (5'-GGgGU) viruses, respectively. B. Western blot analysis of COS-7 intracellular viral proteins. Wild type and 5' pal mutants were tested for viral gene expression inside the COS-7 cells. Equal amount of cell lysate was loaded in each lane. HIV-2 proteins were detected with human serum from an HIV-2 infected patient (HIV-2 Reference sera, NIH AIDS Research & Reference Reagent Program 7G 1337 & 6F D150). Lanes 1, 2, 3 and 4 represent proteins from COS-7 cells transfected with TTB-76 (wt) (5'-GGAGU) and 5' pal mutants (SM2 (5'-ccucU; (L'Hernault et al, 2007)), TTB-16 (5'-accug) and TTB-78 (5'-GGgGU)), respectively. The positions of the detected proteins with their molecular size are shown.

3.5. Discussion

Retroviral packaging occurs by means of interactions between the nucleocapsid (NC) domain of the Gag polyprotein and the packaging signal (Ψ) of unspliced genomic RNA (see review; D'Souza & Summers, 2005). For HIV-2, Ψ is located upstream of the major splice donor site, SD, in the 5' untranslated leader region of genomic RNA (Figure 16). In this study, we investigated the role of pal, a 10nt palindrome sequence (5'-$_{392}$GGAGUGCUCC$_{401}$) situated within Ψ (Figure 16). The 3' side of pal (GCUCC-3') contributes to the formation of stem B, an extension of SL1 involved in the packaging of HIV-2 RNA, although the function of the 5' side of pal (5'-GGAGU) was less well defined (Lanchy & Lodmell, 2007). To ascertain the sequence constraints and role(s) of pal, we have used a randomization/*in vivo* selection methodology to investigate sequences in pal that are important for packaging and replication. *In vivo* selection approaches have been used previously to isolate replication-competent viruses from a complex population of viruses during replication in eukaryotic cells (Berkhout & Klaver, 1993; Doria-Rose & Vogt, 1998; Morris et al, 2002). This method allows for mutational analysis of an ensemble of variant viruses without making *a priori* assumptions that might influence the choice of mutations. In contrast to the methods used before, here we used a recombinagenic method that enables high efficiency incorporation of the randomized region into the proviral DNA vector. We were able to select from among a large number of mutant viral sequences only those that were viable.

After *in vivo* selection of the RND pal viral libraries, sequence analysis showed that nucleotide variation was not well tolerated at positions 392, 393 and 395 of the 5' pal region during viral replication, and G was strongly selected at each of these positions. For

position 394, the dominance of A among the viable winner viruses also reflects its importance for viral replication. This finding is supported by phylogenetic analysis (Figure 20B) and our previous phenotypic reversion study in which mutant pal viruses harboring an A394G substitution reverted back to A (Lanchy & Lodmell, 2007). Here, in the 5nt RND pal selection, a consensus sequence (5'-GGRG) emerged in all viruses surviving the competition after several passages. We interpret this to mean that the 5'-GGRG sequence is essential for optimal HIV-2 replication. In support of this, two non-selected loser viruses with non-consensus pyrimidine-rich sequences were deficient in infectivity and replication (Figure 17).

Compared to the strong selective pressure seen for positions 392-395, there were no obvious sequence constraints at position 396, except that we did not observe any C's at this position in the final pool. The lack of Cs at 396 may indicate that viruses with C396 are somewhat impaired and are successfully outcompeted in this assay, or may reflect a sampling artifact, since we have observed in previous studies that mutant viruses harboring A or C at position 396 replicated like wt during short time course infections (J.-M.L., unpublished). It should be noted that position 396 exhibits a lower degree of phylogenetic conservation than adjacent residues among HIV-2/SIV isolates (Leitner, 2005).

On the basis of previous (L'Hernault et al, 2007; Lanchy & Lodmell, 2007) and the present work, we propose that the 5' pal region comprises an essential component of the core packaging signal of HIV-2 RNA. First, as described above, winning viruses derived from the *in vivo* selection shared a consensus 5'-GGRG sequence (Table 7B). Second, this region is highly conserved among different strains of HIV-2 and SIV, where

37 of 37 published sequences show the 5'-GGAG sequence at this location (Figure 20B and (Leitner, 2005)). Third, packaging was strongly diminished in mutant viruses harboring sequences other than the 5'-GGRG sequence (Figure 18B and (L'Hernault et al, 2007)).

Several studies have demonstrated that the interaction between RNA and the packaging proteins Gag/NC is enhanced by the presence of unpaired G residues in the target RNA (De Guzman et al, 1998; Paoletti et al, 2002; South & Summers, 1993). In this study, secondary structure predictions of winner sequences suggest that, like wt, the 5'-GGRG sequence is unpaired (compare Figure 20A and C), while among the loser viruses the 5' side of pal was predicted to engage in stable base pairing with sequences downstream of SL1 (Figure 20D and E). Because of the similarities with known Gag/NC binding preferences (De Guzman et al, 1998; Fisher et al, 1998; L'Hernault et al, 2007; Lanchy & Lodmell, 2007; Paoletti et al, 2002), the 5'-GGRG consensus is likely to be an essential element for Gag/NC recognition and binding.

In addition to its probable importance for Gag/NC binding, the Ψ-SL1 region of HIV-2 RNA adopts a specific conformation (at least transiently) during replication that is essential for efficient packaging. This conformation involves base pairing between the 3' side of pal and sequence downstream of SL1 that results in the formation of stem B (Figure 16B; Figure 18C). Mutations disrupting stem B negatively affect RNA packaging (Baig et al, 2007; Griffin et al, 2001; Lanchy & Lodmell, 2007; McCann & Lever, 1997). Our present selection and RNA folding results raise the possibility that RNAs of TTB-16 and SM2 were packaging deficient due to the increased stability of extended SL1. ΔG values for TTB-16 and SM2 SL1 RNA structures were -9.9 and -11.7 kcal/mol,

respectively, while wt and winner RNA sequences folded with a predicted ΔG value of -8.3 kcal/mol. Hyperstabilization of stem B may lead to a less favorable interaction of NC with this region (Kerwood et al, 2001) or to the sequestration of normally unpaired G residues that may be critical NC binding determinants (De Guzman et al, 1998; Fisher et al, 1998; Paoletti et al, 2002).

Analysis of the predicted extended SL1 structures highlighted the presence of additional conserved 5'-GGAG sequences located downstream of stem B. In particular, packaging models presented by Summers and coworkers suggest that Gag/NC binds poorly to RNAs in which NC-binding elements are sequestered by base-pairing interactions with other regions (D'Souza & Summers, 2004; Dey et al, 2005). We suggest that the downstream 5'-GGAG sequence(s), could be a target for NC binding and cooperate with the 5' GGRG motif to mediate packaging of HIV-2 RNA. This might explain why TTB-16 and SM2, in which the second 5'-GGAG was sequestered in an extended stem B (Figure 20D and E), were packaging-impaired (Figure 18C).

On a technical note, we found that the use of the recombinagenic method provided a more efficient route to generate the randomized proviral DNA than traditional ligation protocols. *In vivo* selection has been used to characterize important viral RNA signals previously, but difficulty in generating the library of randomized proviral DNA limited the number of sequences that could be simultaneously examined (T.T.B & J.-M.L., unpublished and (Berkhout & Klaver, 1993; Doria-Rose & Vogt, 1998)). This enhanced synthesis of randomized proviral DNA allows a larger degenerate segment to be tested.

A growing body of evidence points to the importance of the pal element within the HIV-2 core packaging signal (Griffin et al, 2001; L'Hernault et al, 2007; Lanchy & Lodmell, 2007; McCann & Lever, 1997). The palindromic nature of this element has been enigmatic, but the current study demonstrates that pal is really composed of two complementary motifs, one that is required to be single stranded and a likely target for protein recognition, and the other that is crucial for an RNA-RNA interaction to form stem B. Because the major packaging signal (Ψ) in HIV-2 RNA is located upstream of SD, and is thus present in both spliced and unspliced RNA species (Figure 16), it is remarkable that only unspliced RNA is packaged. To gain further insights into the mechanism of HIV-2 packaging and RNA selectivity, it will be interesting to investigate the binding affinities of packaging proteins Gag/NC for spliced and unspliced RNA constructs. These experiments are underway in our laboratory.

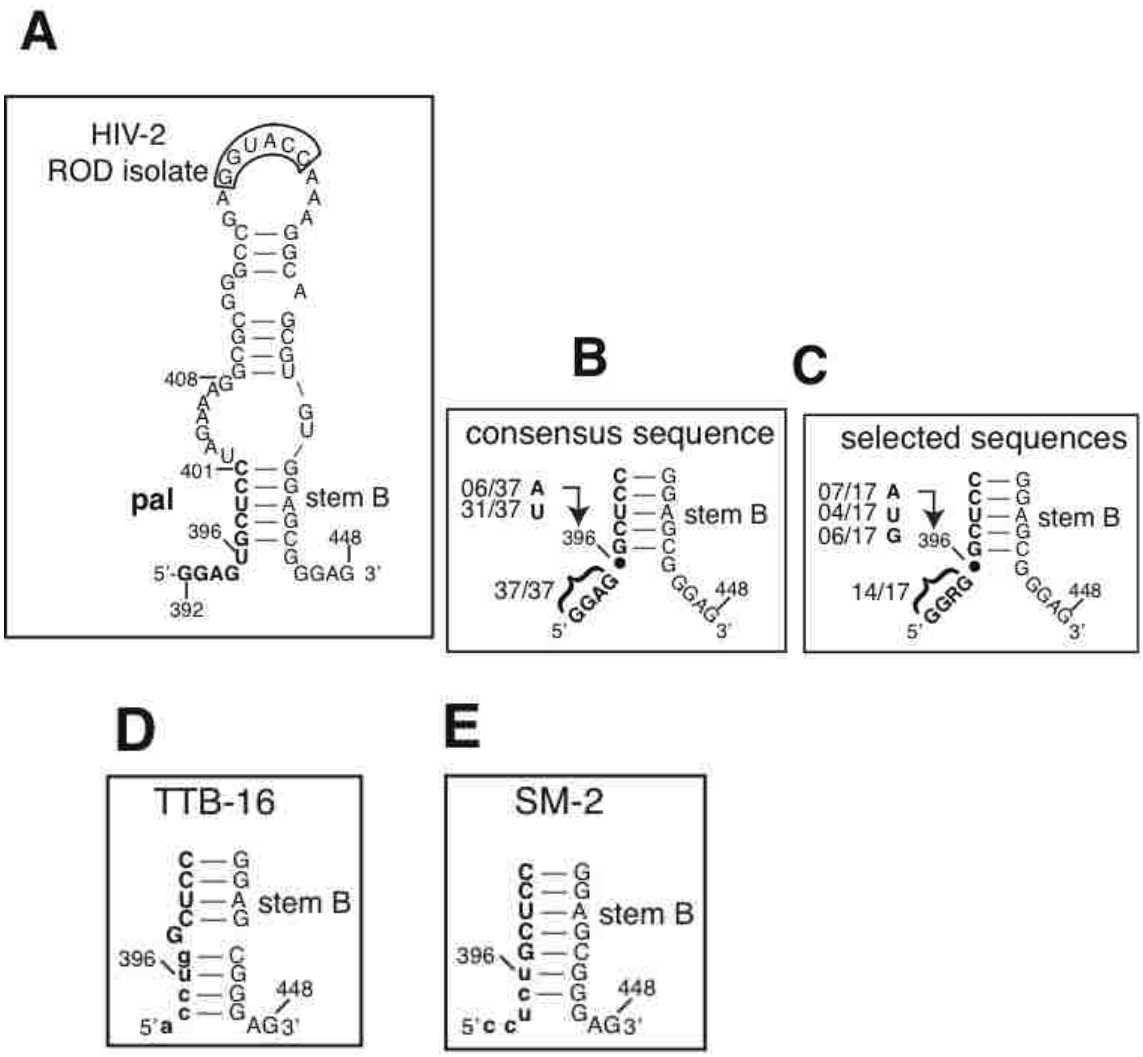


Figure 20. Predicted secondary structures of selected (winner) and non-selected (loser) RNAs.

A. Secondary structure model of pal-SL1 domain in HIV-2 ROD isolate (GenBank accession number M15390 (Baig et al, 2007; Lanchy & Lodmell, 2007)) as predicted by Mfold (Zuker, 2003). Stem B represents the interaction between 5'-GCUCC of pal and GGAGC-3' sequence just downstream of SL1. B. & C. Stem B structure in HIV-2/ SIV strains (Leitner, 2005) and in the selected (winner) sequences derived from serial passage of the 5nt RND library is indicated. The stem B structure was predicted using two-state hybridization model in the DINAMelt server (http://www.bioinfo.rpi.edu/applications/hybrid/twostate.php) using standard parameters (RNA at 37^0C, [Na^+]= 1M, [Mg^{2+}]= 0M and [strand]= 0.00001M (Markham & Zuker, 2005). The consensus purine rich motif and the proportion of bases at position 396 in HIV-2/SIV strains and the winner viruses is shown to the left. D & E. The GGAG sequence at 3' end of stem B in TTB-16 and SM2 is sequestered by base pairing with pyrimidines at the 5' end of pal and potentially makes stem B significantly more stable. The nts different from the wt sequence are shown with lowercase letters.

3.6. Acknowledgements

The authors gratefully acknowledge Christy Strong and Leila Sears for critical reading of the manuscript. This work was funded by the National Institutes of Health grant AI45388 to J.S.L. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, National Institutes of Health (C8166 cells, N° 404) and (HIV-2 Reference Sera, sample ID 7G 1337 and 6F D150). The plasmid pROD10 was provided by the EU programme EVA/MRC Centralised Facility for AIDS Reagents, NIBSC, UK (Grant Number QLK2-CT-1999-00609 and GP828102).

Chapter 4: Human immunodeficiency virus type 2 Gag protein binds to a GGAG motif in the packaging signal of HIV-2 RNA

This chapter consists of experiments done by the author of this dissertation.

4.1. Abstract

A GGAG motif in the packaging signal at the 5' end of HIV-2 genomic RNA has been found to be an essential element for packaging *in vivo*. Here we report studies aimed at characterizing this sequence as a binding element for Gag protein and for the NC domain of the Gag protein *in vitro*. The NC domain of Gag is known for its ability to bind to RNA packaging signals *in vivo*. Thus, the HIV-2 Gag and NC proteins were expressed and purified as C-terminal His-tag fusion proteins in order to use them in the binding studies with HIV-2 leader RNAs containing wt (GGAG) or non-wt (ACCU) sequences in the packaging signal. The gel shift assays revealed that NC protein binds identically with both RNAs, whereas the Gag protein binds specifically to HIV-2 RNA containing a GGAG sequence rather than RNA containing an ACCU sequence. These data suggest the binding of Gag to the GGAG sequence of packaging signal and playing an important role in mediating HIV-2 RNA packaging *in vivo*.

4.2. Introduction

During the HIV packaging process, the precise targeting of full-length genomic RNA into virion particles is facilitated by specific interactions between cis-acting sequences, or packaging signals, in full-length genomic RNA and trans-acting regions in the Gag polyprotein (see reviews; D'Souza & Summers, 2005; Lever, 2007; Linial & Miller, 1990; Russell et al, 2004). In HIV-1, the packaging signals have been mapped to motifs located downstream and upstream of the SD site in the 5' side of genomic RNA (Russell et al, 2004). In HIV-2, the packaging signal, comprised at least in large part of a 28-nt long sequence, is located upstream of the SD site (Griffin et al, 2001). In a study aimed at further characterizing the packaging signal, a 10-nt pal sequence located within this packaging signal was identified as critical for HIV-2 genomic RNA packaging by forming an essential RNA structure (Lanchy & Lodmell, 2007). Further, the 5' side of pal was shown to be essential for the HIV-2 packaging process presumably by forming a Gag binding site (Baig et al, 2008).

The trans-acting regions for HIV-1 packaging have been mapped to the NC domain of Gag polyprotein, specifically to the Cis-His boxes, also called CCHC zinc fingers (CCHC = C-X2-C-X4-H-X4-C, where X= variable amino acids) within the NC domain. The critical role undertaken by these motifs suggests that a specific interaction between Gag and viral RNA takes place during the packaging of genomic RNA into virion particles (Aldovini & Young, 1990; D'Souza & Summers, 2005; Dorfman et al, 1993).

Since the 5' side of pal (GGAG) has been found to be an essential element for

HIV-2 RNA packaging *in vivo* (Baig et al, 2008), we focused our attention on this region to study its interactions with Gag and NC. Here, we purified bacterially expressed HIV-2 Gag and NC proteins and examined their binding to the 5' side of pal by using wt and mutant RNA fragments corresponding to HIV-2 5'UTR. Using an EMSA (electrophoretic mobility shift assay), NC was found to bind to both wt and mutant HIV-2 RNAs. However, Gag protein displayed a distinct preference for interaction with wt over mutant RNAs. These data suggest that NC interacts with low specificity with RNAs, while Gag can discriminate binding targets.

4.3. Materials and Methods

4.3.1. Expression and purification of C-terminal His-tag NC fusion protein

The open reading frame of NC with the inserted sequence for C-terminal His-tag (histidine-tag) was obtained by PCR amplification using modified pROD10 (a full-length plasmid containing infectious HIV-2 DNA sequence) (Lanchy & Lodmell, 2007) as a template, a sense primer containing an *Nde*I site (sNCNdeI; 5'-TCCCATTCCATATGGCCCAGCAGAGAAAGGC-3'), and an anti-sense primer containing *Bam*HI site and 6-His-tag (asNCBamHIS; 5'-AAGGATCCTAGTGATGGTGATGGTGATGGGTCGTTGGGATATCGTAATCCTGAAAATACAGGTTTTATTCACCTGCCTGTCTATCTG-3'). The PCR product was gel purified and digested with *Bam*HI/*Nde*I enzymes before ligation into the pET3a expression vector (Novagen), which was also digested with the same enzymes. The ligated product was transformed into BL21-Gold (DE3) pLysS competent cells (Stratagene). After transformation, several clones were picked for plasmid DNA purification and sequence analysis to confirm the presence of the NC-His-tag fusion insert.

For protein expression, overnight cultures of the transformed BL21-Gold (DE3) pLysS cells were grown at 37°C in the presence of ampicillin and chloramphenicol. This was followed by 1:10 dilution next day with fresh LB in the presence of 0.2 mM $ZnCl_2$. Cultures were grown at 37°C to an optical density of 0.6-0.7 at 600 nm. Finally the expression of the NC protein was initiated by the addition of 1 mM IPTG (isopropyl beta-

D-thiogalactopyranoside), and continued for 5 hours.

After expression, the cells were pelleted by centrifugation at 2700 rcf for 20 minutes at 4°C, followed by lysis with Bugbuster protein extraction reagent (Novagen) in the presence of Lysonase^{TM} Bioprocessing reagent (Novagen) and protease inhibitor cocktail (Sigma Aldrich), shaking for one hour at room temperature. The lysed product was spun at 15300 rcf for 30 minutes at 4°C in order to collect the soluble protein in the supernatant. The supernatant was mixed with the Profinity^{TM} IMAC-Ni-charged resin (BioRad) for purification of protein through the affinity of His-tag with the resin. After shaking for one hour at 4°C, the mixture was centrifuged for three minutes at 1000 rcf to remove the supernatant. The resin was washed for five times with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 0.1 mM $ZnCl_2$, pH 8.0), followed by the elution of protein using elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, 0.1 mM $ZnCl_2$, pH 8.0). Subsequently the NC protein was concentrated using centrifugal concentrators (Amicon Ultra: Ultracell 5K) and stored in storage buffer (25 mM KCl, 25 mM K acetate, 0.1 mM beta-mercaptoethanol, 0.1 mM $ZnCl_2$, pH 7.5) at 4°C. The protein concentration was estimated by UV absorbance spectroscopy, using the extinction coefficient of 12 $mM^{-1}cm^{-1}$ and by SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis).

4.3.2. Expression and purification of C-terminal His-tag Gag fusion protein

The open reading frame of Gag with a C-terminal His-tag in the pET3a expression vector was obtained in two steps. In the first step, we introduced a His-tag in the mpROD10 plasmid at the C-terminus of Gag. This plasmid was called p333.3. Second, we used this plasmid as a template for the construction of Gag-expression vector. The p333.3 was built in the following steps: First, a PCR fragment was obtained from modified pROD10 (a full-length plasmid containing infectious HIV-2 DNA sequence) (Lanchy & Lodmell, 2007) using a sense primer (sXhoIHIS; 5'-TTCTCGAGCATCATCACCATCACCATCAGGGGGAGACACC-3') and an anti-sense primer (asROD2400; 5'-GGCTCTACTTTGGCGACTGG-3'). The amplified product was gel purified and digested with *Xho*I/*BSU*36I enzymes. This digested product was ligated into the mpROD10 vector, which was already digested with the same enzymes. After ligation, transformation, plasmid DNA purification, and the sequencing check, this plasmid (p333.3) was used as a template to produce a PCR product containing an open reading frame of Gag and a C-terminal His-tag. The PCR product was amplified using a sense primer (sGagNdeI; 5'-GATTGTGGCATATGGGCGCGAGAAACTCC-3') and an anti-sense primer (asBcl2114; 5'-TTTGAATGCTGTCGACTACTGGTCTTTTCCAAAGAGAG-3') on p333.3 template. The amplified product was gel purified and digested with *Bcl*I/*Nde*I enzymes to ligate with *Bam*HI/*Nde*I digested pet3a expression vector. The ligated product was transformed

in BL21-Gold (DE3) pLysS competent cells (Stratagene). From this, clones were selected for DNA purification and the sequences were verified to code for the Gag-His-tag fusion protein.

Gag protein expression and purification were performed in the same way as NC protein expression and purification described above. After purification, Gag protein was concentrated using centrifugal concentrators (Amicon Ultra: Ultrcell 10K) and stored in the storage buffer (20 mM Tris, 500 mM NaCl, 0.1 mM $ZnCl_2$, 10 mM beta-mercaptoethanol, pH 7.5) at 4°C. The concentration of protein was estimated by UV absorbance spectroscopy, using an extinction coefficient of 73 $mM^{-1}cm^{-1}$ (sequence analysis program) and by SDS-PAGE.

4.3.3. RNA synthesis

The plasmid DNAs obtained from a previous study of *in vivo* SELEX (Baig et al, 2008) were linearized with *Eco*RI before being transcribed into RNA. Radiolabelled RNAs (1-561 nucleotides) from these *Eco*RI linearized templates were synthesized with the transcription kit (Ambion) in 20ul reaction containing T7 enzyme, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 20 μCi of [^{32}P] UTP (800Ci/mmol; Perkin Elmer) for 30 minutes at 37°C, followed by digestion with RNase free DNase for 15 minutes at 37°C. After synthesis, RNA transcripts were purified by denaturing gel electrophoresis, were excised from the gel and were eluted in gel elution buffer (0.5 M $(NH_4)_2$ acetate, 0.2% SDS, 1 mM EDTA).

4.3.4. EMSA (electromobility shift assay)

Radiolabeled RNAs (50,000 cpm/µl) were denatured in H_2O at 90°C for two minutes then snap cooled on ice for two minutes. Increasing amounts of NC or Gag proteins (0 to 2µg for NC and 0 to 0.7µg for Gag) were added to a fixed amount of RNA. RNA-protein binding was allowed to proceed for 30 minutes at 37°C after adding the binding buffer (final concentrations: 50 mM Tris-HCl, pH 7.5 at 37°C, 40 mM KCl, 0.1 mM $MgCl_2$). Following incubation, samples were loaded with 2ul of 6X glycerol loading dye (40 % glycerol, 44 mM Tris-borate, pH 8.3, 0.25 % bromophenol blue) onto a 5% polyacrylamide gel. Electrophoresis was carried out for three hours at 87 V at 4°C in 44 mM Tris-borate, pH 8.3 and 1 mM EDTA (0.5X TBE). Gels were autoradiographed after drying.

4.4. Results and Discussion

To initiate *in vitro* binding studies, we recombinantly expressed two proteins of HIV-2, the Gag protein (p58) and, separately, the NC protein (p8), which is one of the domains of Gag protein in *E.coli*. Both proteins were expressed as fusion proteins with C-terminal His-tag. These were then purified by using Profinity^{TM} IMAC-Ni-charged resin (BioRad) in a batch purification method according to the manufacturer's protocol, described in Materials and Methods. Following purification, the proteins were analyzed by SDS-PAGE. Both proteins appeared on the gel corresponding to their expected size (Figure 21).

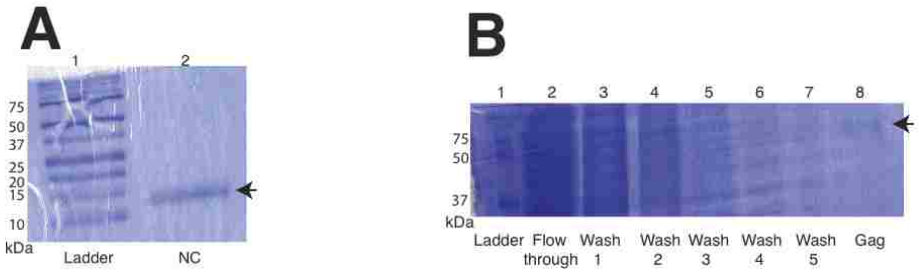


Figure 21. SDS-PAGE analysis of C-terminal His-tag NC fusion protein.

A. A bacterial pellet containing NC expressed protein was processed through lysis and then underwent batch purification as described in Materials and Methods. The purified protein was visualized by SDS-PAGE followed by coomassie brilliant blue staining. Lanes 1 and 2 correspond to the ladder (marker) and the purified NC protein respectively. B. SDS-PAGE analysis of C-terminal His-tag Gag fusion protein: The Gag fusion protein was also purified in a similar manner similar to NC protein. Lanes 1 and 8 correspond to the ladder (marker) and the purified Gag protein respectively. The lanes 2 to 7 correspond to several washing steps. The molecular size of the marker proteins are indicated on left. Note: Due to tilted gel, proteins seem to migrate at different positions other than their size.

The 5' side of pal (GGAG) was reported by our recent study, described in specific aim 2 as a critical element for HIV-2 genomic RNA packaging *in vivo*. We hypothesized that this sequence is a binding element for Gag protein during packaging *in vivo*. In order to study this hypothesis, here we used *in vitro* synthesized RNAs corresponding to 5'UTR of HIV-2 RNA. We synthesized two radioactive RNAs, TTB-16 and TTB-78, which are sequences found and characterized for their effects on packaging in our *in vivo* SELEX study (Baig et al, 2008). In that study, it was shown that the GGAG sequence at the 5' side of pal (as in TTB-78; wt) is important for efficient genomic RNA packaging, and that the ACCT in the same location (as in TTB-16) causes defective packaging of genomic RNA.

After purification of proteins and radiolabelled RNAs, the *in vitro* binding studies were performed using an EMSA. Initially, the NC protein alone was used to test its binding with wt RNA. Interestingly, we found two discrete bands for wt RNA on the gel in the absence of protein (Figure 22; lane 1). Since alternative conformations, as a result of long distance interactions, have been proposed for the HIV-2 leader RNA (Dirac et al, 2002a; Lanchy et al, 2003a; Lanchy et al, 2003b), it was surprising for us to see two distinct monomer bands for HIV-2 RNA on a native gel. We termed these lower and upper species as fast and slow migrating conformers, respectively. Berkhout and coworkers have shown that HIV-1 RNA also adopts two different monomeric conformations that migrate differently on gels (Huthoff & Berkhout, 2001). Secondary structural analysis of these species suggested that the fast migrating conformer can form a rod like structure due to LDI and that the slow migrating conformer was proposed to contain BMH structures (Huthoff & Berkhout, 2001).

It is known that HIV-1 NC has non-specific nucleic acid chaperone activity, which is accelerated by its binding to misfolded structures of nucleic acids and allowing them to rearrange (Cruceanu et al, 2006; Levin et al, 2005). Although HIV-1 NC protein's chaperone activity for converting the LDI conformer to a BMH conformer has been the subject of study (Huthoff & Berkhout, 2001), little is known about the chaperone activity of HIV-2 NC in this regard. We report in this study that HIV-2 NC also behaves as chaperone to convert one conformer of HIV-2 leader RNA into another. As seen in Figure 22, the fast migrating conformer was converted to the slow migrating conformer with increasing concentrations of NC. Interestingly, with the increasing concentrations of NC, the migration of the upper band was retarded, suggesting the formation of an RNA-protein complex (Figure 22).

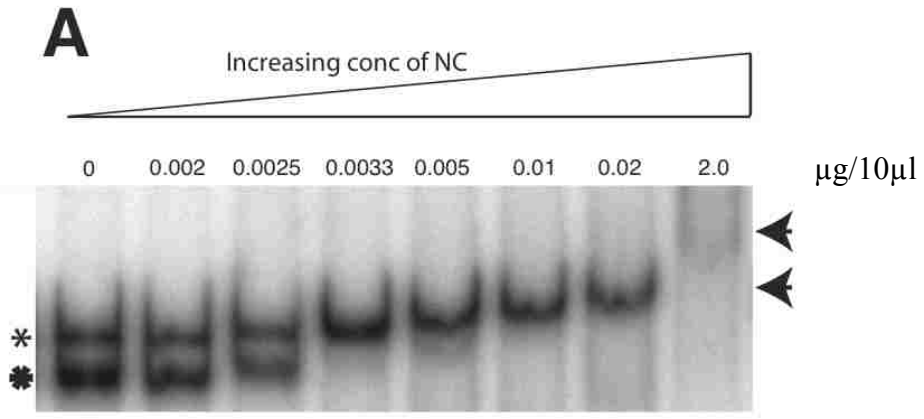


Figure 22. EMSA for NC- wt RNA interactions.

The radiolabelled wt HIV-2 leader RNA with GGAG sequence at the 5' side of pal at 50,000 cpm/µl/each lane was incubated with increasing concentrations of NC-fusion protein (0, 0.002, 0.0025, 0.0033, 0.005, 0.01, 0.02 and 2, each at µg/10ul reaction) for 30 minutes in the presence of binding buffer and loaded on native PAGE. The two conformers of monomeric RNA are represented by filled asterisk (fast conformer) and empty asterisk (slow conformer). Arrows indicate the retarded RNA-protein complexes.

In order to see whether non-GGAG pal sequences can also bind to NC *in vitro*, we used TTB-16, which harbors the packaging defective AACT sequence in pal.

Surprisingly, we could not see a distinctive mode of NC binding with this RNA as compared to the GGAG wt RNA (compare Figure 23 to Figure 22). Since it has been reported previously that HIV-1 NC protein can bind non-specifically with RNAs *in vitro* (Berkowitz et al, 1993; Clever et al, 1995), the similar pattern of binding observed here for wt and mutant RNAs suggests that the binding of HIV-2 NC for RNA is non-specific. In the future, this notion could be analyzed further with the use of competitor RNAs, used in binding studies of HIV-1 NC and those RNAs used in this study to reduce the background of non-specific interactions (Clever et al, 1995).

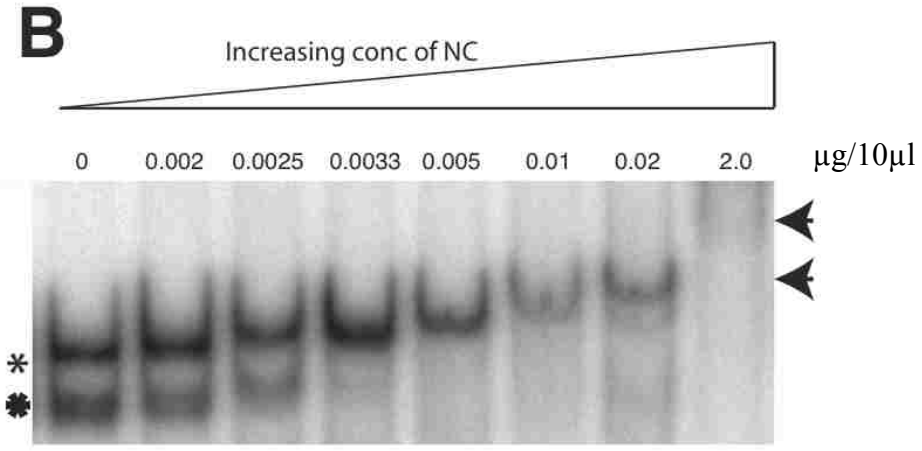


Figure 23. EMSA for NC-mutant RNA interactions.

The radiolabelled mutant RNA with ACCT sequence at the 5'side of pal at 50,000 cpm/µl/each lane was incubated with increasing concentrations of NC-fusion protein (0, 0.002, 0.0025, 0.0033, 0.005, 0.01, 0.02 and 2, each at µg/10ul reaction) for 30 minutes in the presence of binding buffer and loaded on native PAGE. The two conformers of monomeric RNA are represented by filled asterisk (fast conformer) and empty asterisk (slow conformer). Arrows indicate the retarded RNA-protein complexes.

It has been demonstrated that specific interactions between viral genomes and Gag take place via the NC domain within Gag in HIV-1 (Berkowitz et al, 1995b; Zhang & Barklis, 1995). Therefore, we tested to see whether a differential pattern of interaction could be seen between wt and mutant RNAs in the presence of Gag protein. As seen in Figure 24**,** both RNAs showed similar migration patterns with the appearance of two

bands in either absence or presence of low concentration of Gag. However, as the concentration of Gag was increased, the fast migrating band for wt disappeared as compared to TTB-16, which still had two distinct bands at this concentration of Gag (Figure 24; compare panel A with panel B). These results demonstrate that HIV-2 Gag protein also possesses the chaperone activity as seen for HIV-1 Gag (Cruceanu et al, 2006). Furthermore, it displays the specificity of interaction of Gag for wt RNA as compared to mutant RNA. These results correlate with our *in vivo* results, where GGAG was determined to be essential for HIV-2 genomic RNA packaging and ACCT was a determinant of packaging defects (Baig et al, 2008).

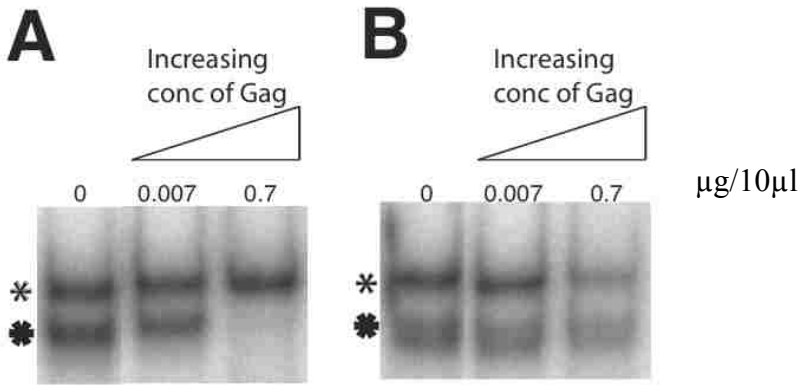


Figure 24. EMSA for Gag-RNA interactions.

The radiolabelled RNAs (wt RNA; TTB-78, Figure 24A and mutant RNA; TTB-16, Figure 24B) at 50,000 cpm/µl were incubated with increasing concentrations of Gag-fusion protein (0, 0.007 and 0.07, each at µg/10ul reaction) for 30 minutes in the presence of binding buffer and loaded on native PAGE. Two conformers of monomeric RNA are represented by filled asterisk (fast conformer) and empty asterisk (slow conformer).

Our results also corroborate the proposed folding patterns of the pal-SL1 region, which differ between wt and TTB-16 RNAs. In wt RNA, the predicted secondary structure for the pal-SL1 region showed an extended SL1 structure, where the 3' side of pal is bound with a sequence downstream of SL1 and the 5' side of pal is exposed and

present as an unpaired region. In this model, the GGAG sequence was proposed to be a binding target for the Gag binding (Baig et al, 2008) (Figure 25A). In contrast, 5' pal of packaging defective TTB-16 is entrapped by another sequence downstream of extended SL1, which was proposed to be an inappropriate presentation for Gag binding (Baig et al, 2008) (Figure 25B). We interpret the present *in vitro* Gag-RNA binding results to mean that the extended SL1 structure with an exposed 5' pal can bind with Gag to promote packaging, while the presence of a non-GGAG sequence at the 5' side of pal is either not recognized by Gag or is sequestered by aberrant base-pairing and can not be accessed by Gag.

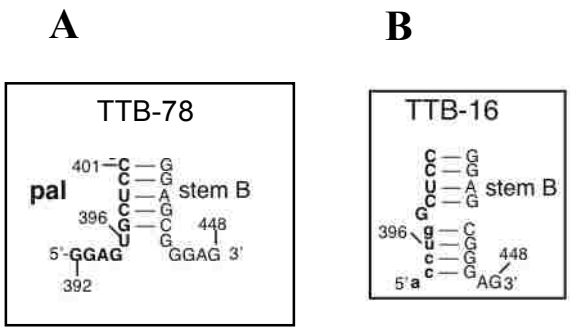


Figure 25. Predicted secondary structures of TTB-78 (wt) and TTB-16 RNAs.

A. Secondary structure model of pal-SL1 domain in HIV-2 ROD isolate (GenBank accession number M15390 (Baig et al, 2007; Lanchy & Lodmell, 2007)) and non-selected RNA from *in vivo* SELEX study as predicted by Mfold (Zuker, 2003). Stem B represents the interaction between 5'-GCUCC of pal and GGAGC-3' sequence just downstream of SL1. B. The GGAG sequence at 3' end of stem B in TTB-16 is sequestered by base pairing with pyrimidines at the 5' end of pal and potentially makes stem B significantly more stable. The nts different from the wt sequence are shown with lowercase letters.
The stem B structure was predicted using two-state hybridization model in the DINAMelt server (http://www.bioinfo.rpi.edu/applications/hybrid/twostate.php) using standard parameters (RNA at 37^0C, $[Na^+]$= 1M, $[Mg^{2+}]$= 0M and [strand]= 0.00001M (Markham & Zuker, 2005).

Chapter 5: Conclusions and Future Directions

5.1. Conclusions

The main intent of the research presented in this dissertation was to understand the regulation of the dimerization and packaging processes of HIV-2 RNA. The research emphasized a 10-nt palindrome sequence located in the leader region of HIV-2 RNA. Through randomization of this sequence within HIV-2 leader and full-length RNAs, followed by *in vitro* and *in vivo* selections, we demonstrated that pal is a critical element for the regulation of dimerization and packaging. Biochemical characterization supported the existence of essential RNA structures, and phenotypic characterization of mutant viruses showed the physiological significance of these structures and signals.

Both the *in vitro* and the *in vivo* selections identified an extended SL1 structure in which 3' pal is entrapped in a stem and 5' pal is exposed. This structure was discovered to be important for HIV-2 leader RNA dimerization *in vitro* and for genomic RNA packaging *in vivo*. The extended SL1 was also identified as a recognition element for Gag-RNA interactions *in vitro*.

5.2. Future Directions

Studies on the pal region of packaging signal provide evidence that it plays an important role in the packaging and viral replication of HIV-2. The findings presented in this dissertation provide additional support about the role of pal in packaging. However, the underlying mechanism of packaging is not resolved completely at the moment. Because packaging is an important step during viral replication, it may represent a new antiretroviral targets for anti-HIV therapy. Some suggestions and routes are presented below.

1. To further understand whether the extended SL1 structure, along with the exposed 5' pal, is the principal recognition element for Gag, a logical extension of the results presented here would be to search for Gag-binding RNA aptamers. RNA aptamers for HIV-1 Gag protein have been isolated using *in vitro* SELEX, in which short stretches of randomized nucleotides favored a single structural motif after several rounds of selection (Lochrie et al, 1997). Similar methods could be employed to select consensus secondary structures for the HIV-2 Gag.

After the generation and characterization of the aptamer, the sequence could be used to replace the pal-SL1 region of HIV-2 RNA in mutant viruses. Determination of the packaging phenotype would shed light on whether Gag recognition of RNA is sufficient to promote packaging.

2. The small RNA aptamers developed above could potentially also block viral replication (Joshi & Prasad, 2002). The Gag protein is critical for several steps of viral

replication including packaging and assembly (Kaye & Lever, 1999; Kim & Jeong, 2004; Luo et al, 1994). The results of our specific aim 3 also provide evidence that Gag can bind with HIV-2 RNA specifically, and can therefore facilitate the packaging of HIV-2 genomic RNA *in vivo*. This suggests that Gag protein could be a useful target for the inhibition of viral replication by the use of aptamers.

To explore this aspect, first an *in vitro* competition binding experiment would be carried out using Gag protein and aptamers versus HIV-2 leader RNA. If the aptamers successfully compete with wt leader RNA sequence for Gag binding *in vitro*, we shall test the inhibitory effect of these aptamers on genomic RNA packaging *in vivo* by using a method similar to that employed by Jeong's group (Kim & Jeong, 2004). If the high affinity aptamers can be efficiently expressed in HIV-2 infected cells, we would predict to see a packaging defect along with reduced viral replication. This study would be a good start towards the development of an inhibitor of Gag protein, which may ultimately lead to anti-HIV therapeutic strategies in the future.

Chapter 6: References

Abbink TE, Berkhout B (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J Biol Chem* **278**(13)**:** 11601-11611

Abbink TE, Berkhout B (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J Virol* **82**(6)**:** 3090-3098

Abbink TE, Ooms M, Haasnoot PC, Berkhout B (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* **44**(25)**:** 9058-9066

Aldovini A, Young RA (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* **64**(5)**:** 1920-1926

Allan JS, Coligan JE, Barin F, McLane MF, Sodroski JG, Rosen CA, Haseltine WA, Lee TH, Essex M (1985) Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science* **228**(4703)**:** 1091-1094

Arya SK, Zamani M, Kundra P (1998) Human immunodeficiency virus type 2 lentivirus vectors for gene transfer: expression and potential for helper virus-free packaging. *Hum Gene Ther* **9**(9)**:** 1371-1380

Baba S, Takahashi K, Nomura Y, Noguchi S, Koyanagi Y, Yamamoto N, Takaku H, Kawai G (2001) Conformational change of dimerization initiation site of HIV-1 genomic RNA by NCp7 or heat treatment. *Nucleic Acids Res Suppl*(1)**:** 155-156

Baig TT, Lanchy JM, Lodmell JS (2007) HIV-2 RNA dimerization is regulated by intramolecular interactions in vitro. *RNA* **13**(8)**:** 1341-1354

Baig TT, Lanchy JM, Lodmell JS (2008) Randomization and in vivo selection reveal a GGRG motif essential for packaging HIV-2 RNA. *J Virol (in press).*

Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**(5252)**:** 1209-1211

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**(4599)**:** 868-871

Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol* **229**(2)**:** 382-397

Bender W, Davidson N (1976) Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* **7**(4)**:** 595-607

Berkhout B (1996) Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol* **54:** 1-34

Berkhout B, Klaver B (1993) In vivo selection of randomly mutated retroviral genomes. *Nucleic Acids Res* **21**(22)**:** 5020-5024

Berkhout B, Schoneveld I (1993) Secondary structure of the HIV-2 leader RNA comprising the tRNA-primer binding site. *Nucleic Acids Res* **21**(5)**:** 1171-1178

Berkhout B, van Wamel JL (1996) Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol* **70**(10)**:** 6723-6732

Berkhout B, van Wamel JL (2000) The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA* **6**(2)**:** 282-295

Berkowitz R, Fisher J, Goff SP (1996) RNA packaging. *Curr Top Microbiol Immunol* **214:** 177-218

Berkowitz RD, Goff SP (1994) Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. *Virology* **202**(1)**:** 233-246

Berkowitz RD, Hammarskjold ML, Helga-Maria C, Rekosh D, Goff SP (1995a) 5' regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. *Virology* **212**(2)**:** 718-723

Berkowitz RD, Luban J, Goff SP (1993) Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* **67**(12)**:** 7190-7200

Berkowitz RD, Ohagen A, Hoglund S, Goff SP (1995b) Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric Gag polyproteins during RNA packaging in vivo. *J Virol* **69**(10)**:** 6445-6456

Bernacchi S, Ennifar E, Toth K, Walter P, Langowski J, Dumas P (2005) Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. *J Biol Chem* **280**(48)**:** 40112-40121

Bowen DL, Lane HC, Fauci AS (1985) Immunopathogenesis of the acquired immunodeficiency syndrome. *Ann Intern Med* **103**(5)**:** 704-709

Briggs JA, Wilk T, Welker R, Krausslich HG, Fuller SD (2003) Structural organization of authentic, mature HIV-1 virions and cores. *Embo J* **22**(7)**:** 1707-1715

Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF (2001) The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J Virol* **75**(1)**:** 181-191

Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M (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 U S A* **90**(13)**:** 6125-6129

Bushman FD, Fujiwara T, Craigie R (1990) Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249**(4976)**:** 1555-1558

Chackerian B, Haigwood NL, Overbaugh J (1995) Characterization of a CD4-expressing macaque cell line that can detect virus after a single replication cycle and can be infected by diverse simian immunodeficiency virus isolates. *Virology* **213**(2)**:** 386-394

Chen Z, Luckay A, Sodora DL, Telfer P, Reed P, Gettie A, Kanu JM, Sadek RF, Yee J, Ho DD, Zhang L, Marx PA (1997) Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J Virol* **71**(5)**:** 3953-3960

Cheung KS, Smith RE, Stone MP, Joklik WK (1972) Comparison of immature (rapid harvest) and mature Rous sarcoma virus particles. *Virology* **50**(3)**:** 851-864

Chin MP, Rhodes TD, Chen J, Fu W, Hu WS (2005) Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci U S A* **102**(25)**:** 9002-9007

Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, Ponath PD, Wu L, Mackay CR, LaRosa G, Newman W, Gerard N, Gerard C, Sodroski J (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**(7)**:** 1135-1148

Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C, et al. (1986a) Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**(4761)**:** 343-346

Clavel F, Guyader M, Guetard D, Salle M, Montagnier L, Alizon M (1986b) Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* **324**(6098)**:** 691-695

Clavel F, Orenstein JM (1990) A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* **64**(10)**:** 5230-5234

Clever J, Sassetti C, Parslow TG (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* **69**(4)**:** 2101-2109

Clever JL, Eckstein DA, Parslow TG (1999) Genetic dissociation of the encapsidation and reverse transcription functions in the 5' R region of human immunodeficiency virus type 1. *J Virol* **73**(1)**:** 101-109

Clever JL, Miranda D, Jr., Parslow TG (2002) RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome. *J Virol* **76**(23)**:** 12381-12387

Clever JL, Taplitz RA, Lochrie MA, Polisky B, Parslow TG (2000) A heterologous, high-affinity RNA ligand for human immunodeficiency virus Gag protein has RNA packaging activity. *J Virol* **74**(1)**:** 541-546

Coffin J, Haase A, Levy JA, Montagnier L, Oroszlan S, Teich N, Temin H, Toyoshima K, Varmus H, Vogt P, et al. (1986) What to call the AIDS virus? *Nature* **321**(6065)**:** 10

Cruceanu M, Urbaneja MA, Hixson CV, Johnson DG, Datta SA, Fivash MJ, Stephen AG, Fisher RJ, Gorelick RJ, Casas-Finet JR, Rein A, Rouzina I, Williams MC (2006) Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **34**(2)**:** 593-605

Cullen BR, Greene WC (1990) Functions of the auxiliary gene products of the human immunodeficiency virus type 1. *Virology* **178**(1)**:** 1-5

D'Souza V, Summers MF (2004) Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431**(7008)**:** 586-590

D'Souza V, Summers MF (2005) How retroviruses select their genomes. *Nat Rev Microbiol* **3**(8)**:** 643-655

Damgaard CK, Dyhr-Mikkelsen H, Kjems J (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**(16)**:** 3667-3676

Darlix JL, Gabus C, Nugeyre MT, Clavel F, Barre-Sinoussi F (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol* **216**(3)**:** 689-699

Dayton AI, Sodroski JG, Rosen CA, Goh WC, Haseltine WA (1986) The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* **44**(6)**:** 941-947

De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**(5349)**:** 384-388

Debouck C, Gorniak JG, Strickler JE, Meek TD, Metcalf BW, Rosenberg M (1987) Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. *Proc Natl Acad Sci U S A* **84**(24)**:** 8903-8906

Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**(6584)**:** 661-666

Dey A, York D, Smalls-Mantey A, Summers MF (2005) Composition and sequence-dependent binding of RNA to the nucleocapsid protein of Moloney murine leukemia virus. *Biochemistry* **44**(10)**:** 3735-3744

di Marzo Veronese F, Copeland TD, DeVico AL, Rahman R, Oroszlan S, Gallo RC, Sarngadharan MG (1986) Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science* **231**(4743)**:** 1289-1291

Dirac AM, Huthoff H, Kjems J, Berkhout B (2001) The dimer initiation site hairpin mediates dimerization of the human immunodeficiency virus, type 2 RNA genome. *J Biol Chem* **276**(34)**:** 32345-32352

Dirac AM, Huthoff H, Kjems J, Berkhout B (2002a) Regulated HIV-2 RNA dimerization by means of alternative RNA conformations. *Nucleic Acids Res* **30**(12)**:** 2647-2655

Dirac AM, Huthoff H, Kjems J, Berkhout B (2002b) Requirements for RNA heterodimerization of the human immunodeficiency virus type 1 (HIV-1) and HIV-2 genomes. *J Gen Virol* **83**(Pt 10)**:** 2533-2542.}

Doranz BJ, Rucker J, Yi Y, Smyth RJ, Samson M, Peiper SC, Parmentier M, Collman RG, Doms RW (1996) A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* **85**(7)**:** 1149-1158

Dorfman T, Luban J, Goff SP, Haseltine WA, Gottlinger HG (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **67**(10)**:** 6159-6169

Dorfman T, Mammano F, Haseltine WA, Gottlinger HG (1994) Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. *J Virol* **68**(3)**:** 1689-1696

Doria-Rose NA, Vogt VM (1998) In vivo selection of Rous sarcoma virus mutants with randomized sequences in the packaging signal. *J Virol* **72**(10)**:** 8073-8082

Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**(6584)**:** 667-673

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res* **15**(22)**:** 9109-9128

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**(6287)**:** 818-822

Farmerie WG, Loeb DD, Casavant NC, Hutchison CA, 3rd, Edgell MH, Swanstrom R (1987) Expression and processing of the AIDS virus reverse transcriptase in Escherichia coli. *Science* **236**(4799)**:** 305-308

Farnet CM, Bushman FD (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* **88**(4)**:** 483-492

Fauci AS (2006) Twenty-five years of HIV/AIDS. *Science* **313**(5786)**:** 409

Feinberg MB, Jarrett RF, Aldovini A, Gallo RC, Wong-Staal F (1986) HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* **46**(6)**:** 807-817

Fisher AG, Feinberg MB, Josephs SF, Harper ME, Marselle LM, Reyes G, Gonda MA, Aldovini A, Debouk C, Gallo RC, et al. (1986) The trans-activator gene of HTLV-III is essential for virus replication. *Nature* **320**(6060)**:** 367-371

Fisher RJ, Rein A, Fivash M, Urbaneja MA, Casas-Finet JR, Medaglia M, Henderson LE (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J Virol* **72**(3)**:** 1902-1909

Flynn JA, An W, King SR, Telesnitsky A (2004) Nonrandom dimerization of murine leukemia virus genomic RNAs. *J Virol* **78**(22)**:** 12129-12139

Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67:** 1-25

Fu W, Gorelick RJ, Rein A (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol* **68**(8)**:** 5013-5018

Fu W, Rein A (1993) Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* **67**(9)**:** 5443-5449

Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI (1999) Assembly and analysis of conical models for the HIV-1 core. *Science* **283**(5398)**:** 80-83

Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) The structural biology of HIV assembly. *Curr Opin Struct Biol* **18**(2)**:** 203-217

Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**(6718)**:** 436-441

Gao F, Yue L, White AT, Pappas PG, Barchue J, Hanson AP, Greene BM, Sharp PM, Shaw GM, Hahn BH (1992) Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* **358**(6386)**:** 495-499

Garcia JA, Harrich D, Soultanakis E, Wu F, Mitsuyasu R, Gaynor RB (1989) Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. *EMBO J* **8**(3)**:** 765-778

Garzino-Demo A, Gallo RC, Arya SK (1995) Human immunodeficiency virus type 2 (HIV-2): packaging signal and associated negative regulatory element. *Hum Gene Ther* **6**(2)**:** 177-184

Gelderblom HR, Hausmann EH, Ozel M, Pauli G, Koch MA (1987) Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* **156**(1)**:** 171-176

Gonda MA, Wong-Staal F, Gallo RC, Clements JE, Narayan O, Gilden RV (1985) Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. *Science* **227**(4683)**:** 173-177

Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A (1981) Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* **305**(24)**:** 1425-1431

Greatorex J, Lever A (1998) Retroviral RNA dimer linkage. *J Gen Virol* **79 (Pt 12):** 2877-2882

Griffin SD, Allen JF, Lever AM (2001) The major human immunodeficiency virus type 2 (HIV-2) packaging signal is present on all HIV-2 RNA species: cotranslational RNA encapsidation and limitation of Gag protein confer specificity. *J Virol* **75**(24)**:** 12058-12069

Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. *Science* **287**(5453)**:** 607-614

Harrich D, Hooker B (2002) Mechanistic aspects of HIV-1 reverse transcription initiation. *Rev Med Virol* **12**(1)**:** 31-45

Helga-Maria C, Hammarskjold ML, Rekosh D (1999) An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA. *J Virol* **73**(5)**:** 4127-4135

Henderson LE, Bowers MA, Sowder RC, 2nd, Serabyn SA, Johnson DG, Bess JW, Jr., Arthur LO, Bryant DK, Fenselau C (1992) Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J Virol* **66**(4)**:** 1856-1865

Hibbert CS, Mirro J, Rein A (2004) mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J Virol* **78**(20)**:** 10927-10938

Ho DD, Bieniasz PD (2008) HIV-1 at 25. *Cell* **133**(4)**:** 561-565

Hoglund S, Ofverstedt LG, Nilsson A, Lundquist P, Gelderblom H, Ozel M, Skoglund U (1992) Spatial visualization of the maturing HIV-1 core and its linkage to the envelope. *AIDS Res Hum Retroviruses* **8**(1)**:** 1-7

Hoglund S, Ohagen A, Goncalves J, Panganiban AT, Gabuzda D (1997) Ultrastructure of HIV-1 genomic RNA. *Virology* **233**(2)**:** 271-279

Hu WS, Temin HM (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**(4)**:** 1556-1560

Huang M, Orenstein JM, Martin MA, Freed EO (1995) p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol* **69**(11)**:** 6810-6818

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

Huthoff H, Berkhout B (2002) Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry* **41**(33)**:** 10439-10445

Jakobovits A, Smith DH, Jakobovits EB, Capon DJ (1988) A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV trans activator. *Mol Cell Biol* **8**(6)**:** 2555-2561

Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP (2000) High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol* **74**(3)**:** 1234-1240

Joshi P, Prasad VR (2002) Potent inhibition of human immunodeficiency virus type 1 replication by template analog reverse transcriptase inhibitors derived by SELEX (systematic evolution of ligands by exponential enrichment). *J Virol* **76**(13)**:** 6545-6557

Jossinet F, Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (2001) Identification of the in vitro HIV-2/SIV RNA dimerization site reveals striking differences with HIV-1. *J Biol Chem* **276**(8)**:** 5598-5604

Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. *Gene* **82**(1)**:** 83-87

Kao SY, Calman AF, Luciw PA, Peterlin BM (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**(6147)**:** 489-493

Kaplan AH, Manchester M, Swanstrom R (1994) The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *J Virol* **68**(10)**:** 6782-6786

Katz RA, Terry RW, Skalka AM (1986) A conserved cis-acting sequence in the 5' leader of avian sarcoma virus RNA is required for packaging. *J Virol* **59**(1)**:** 163-167

Kaye JF, Lever AM (1999) Human immunodeficiency virus types 1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *J Virol* **73**(4)**:** 3023-3031

Kerwood DJ, Cavaluzzi MJ, Borer PN (2001) Structure of SL4 RNA from the HIV-1 packaging signal. *Biochemistry* **40**(48)**:** 14518-14529

Kim HJ, Lee K, O'Rear JJ (1994) A short sequence upstream of the 5' major splice site is important for encapsidation of HIV-1 genomic RNA. *Virology* **198**(1)**:** 336-340

Kim MY, Jeong S (2004) Inhibition of the functions of the nucleocapsid protein of human immunodeficiency virus-1 by an RNA aptamer. *Biochem Biophys Res Commun* **320**(4)**:** 1181-1186

Klatzmann D, Barre-Sinoussi F, Nugeyre MT, Danquet C, Vilmer E, Griscelli C, Brun-Veziret F, Rouzioux C, Gluckman JC, Chermann JC, et al. (1984) Selective tropism of

lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science* **225**(4657)**:** 59-63

Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A* **85**(13)**:** 4686-4690

Kondo E, Gottlinger HG (1996) A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. *J Virol* **70**(1)**:** 159-164

Kowalski M, Potz J, Basiripour L, Dorfman T, Goh WC, Terwilliger E, Dayton A, Rosen C, Haseltine W, Sodroski J (1987) Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* **237**(4820)**:** 1351-1355

Kramer RA, Schaber MD, Skalka AM, Ganguly K, Wong-Staal F, Reddy EP (1986) HTLV-III gag protein is processed in yeast cells by the virus pol-protease. *Science* **231**(4745)**:** 1580-1584

L'Hernault A, Greatorex JS, Crowther RA, Lever AM (2007) Dimerisation of HIV-2 genomic RNA is linked to efficient RNA packaging, normal particle maturation and viral infectivity. *Retrovirology* **4:** 90

Lanchy JM, Ivanovitch JD, Lodmell JS (2003a) A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA* **9**(8)**:** 1007-1018

Lanchy JM, Lodmell JS (2002) Alternate usage of two dimerization initiation sites in HIV-2 viral RNA in vitro. *J Mol Biol* **319**(3)**:** 637-648

Lanchy JM, Lodmell JS (2007) An extended stem-loop 1 is necessary for human immunodeficiency virus type 2 replication and affects genomic RNA encapsidation. *J Virol* **81**(7)**:** 3285-3292

Lanchy JM, Rentz CA, Ivanovitch JD, Lodmell JS (2003b) Elements located upstream and downstream of the major splice donor site influence the ability of HIV-2 leader RNA to dimerize in vitro. *Biochemistry* **42**(9)**:** 2634-2642

Lanchy JM, Szafran QN, Lodmell JS (2004) Splicing affects presentation of RNA dimerization signals in HIV-2 in vitro. *Nucleic Acids Res* **32**(15)**:** 4585-4595

Laughrea M, Jette L (1994) A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* **33**(45)**:** 13464-13474

Laughrea M, Jette L (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**(5)**:** 1589-1598

Laughrea M, Jette L, Mak J, Kleiman L, Liang C, Wainberg MA (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**(5)**:** 3397-3406

Laughrea M, Shen N, Jette L, Wainberg MA (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**(1)**:** 226-234

Leitner T, B. Foley, B. Hahn, P. Marx, F. McCutchan, J.Mellors, S. Wolinsky, and B, Korber. (2005) *HIV Sequence Compedium 2005*, Los Alamos National LaboratorY: Los Alamos, NM.

Lever A, Gottlinger H, Haseltine W, Sodroski J (1989) Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* **63**(9)**:** 4085-4087

Lever AM (2007) HIV-1 RNA packaging. *Adv Pharmacol* **55:** 1-32

Levin JG, Guo J, Rouzina I, Musier-Forsyth K (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **80:** 217-286

Linial ML, Miller AD (1990) Retroviral RNA packaging: sequence requirements and implications. *Curr Top Microbiol Immunol* **157:** 125-152

Lochrie MA, Waugh S, Pratt DG, Jr., Clever J, Parslow TG, Polisky B (1997) In vitro selection of RNAs that bind to the human immunodeficiency virus type-1 gag polyprotein. *Nucleic Acids Res* **25**(14)**:** 2902-2910

Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (2000) Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. *RNA* **6**(9)**:** 1267-1276

Luban J, Goff SP (1994) Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J Virol* **68**(6)**:** 3784-3793

Luo L, Li Y, Dales S, Kang CY (1994) Mapping of functional domains for HIV-2 gag assembly into virus-like particles. *Virology* **205**(2)**:** 496-502

Mangasarian A, Trono D (1997) The multifaceted role of HIV Nef. *Res Virol* **148**(1)**:** 30-33

Markham NR, Zuker M (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33**(Web Server issue)**:** W577-581

Marx JL (1982) New disease baffles medical community. *Science* **217**(4560)**:** 618-621

Masur H, Michelis MA, Greene JB, Onorato I, Stouwe RA, Holzman RS, Wormser G, Brettman L, Lange M, Murray HW, Cunningham-Rundles S (1981) An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med* **305**(24)**:** 1431-1438

McBride MS, Panganiban AT (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* **70**(5)**:** 2963-2973

McBride MS, Panganiban AT (1997) Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J Virol* **71**(3)**:** 2050-2058

McCann EM, Lever AM (1997) Location of cis-acting signals important for RNA encapsidation in the leader sequence of human immunodeficiency virus type 2. *J Virol* **71**(5)**:** 4133-4137

McDougal JS, Kennedy MS, Sligh JM, Cort SP, Mawle A, Nicholson JK (1986) Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**(4736)**:** 382-385

Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D (2004) Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J Biol Chem* **279**(9)**:** 7792-7798

Meyer BE, Malim MH (1994) The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev* **8**(13)**:** 1538-1547

Miele G, Mouland A, Harrison GP, Cohen E, Lever AM (1996) The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J Virol* **70**(2)**:** 944-951

Mikkelsen JG, Lund AH, Duch M, Pedersen FS (2000) Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo. *J Virol* **74**(2)**:** 600-610

Miller MD, Farnet CM, Bushman FD (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* **71**(7)**:** 5382-5390

Mizrahi V, Lazarus GM, Miles LM, Meyers CA, Debouck C (1989) Recombinant HIV-1 reverse transcriptase: purification, primary structure, and polymerase/ribonuclease H activities. *Arch Biochem Biophys* **273**(2)**:** 347-358

Morris S, Johnson M, Stavnezer E, Leis J (2002) Replication of avian sarcoma virus in vivo requires an interaction between the viral RNA and the TpsiC loop of the tRNA(Trp) primer. *J Virol* **76**(15)**:** 7571-7577

Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA, Capon DJ (1985) Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* **313**(6002)**:** 450-458

Mujeeb A, Ulyanov NB, Georgantis S, Smirnov I, Chung J, Parslow TG, James TL (2007) Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acids Res* **35**(6)**:** 2026-2034

Munn RJ, Marx PA, Yamamoto JK, Gardner MB (1985) Ultrastructural comparison of the retroviruses associated with human and simian acquired immunodeficiency syndromes. *Lab Invest* **53**(2)**:** 194-199

Muriaux D, De Rocquigny H, Roques BP, Paoletti J (1996a) NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* **271**(52)**:** 33686-33692

Muriaux D, Fosse P, Paoletti J (1996b) A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry* **35**(15)**:** 5075-5082

Muriaux D, Girard PM, Bonnet-Mathoniere B, Paoletti J (1995) Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* **270**(14)**:** 8209-8216

Ohlmann T, Lopez-Lastra M, Darlix JL (2000) An internal ribosome entry segment promotes translation of the simian immunodeficiency virus genomic RNA. *J Biol Chem* **275**(16)**:** 11899-11906

Ooms M, Huthoff H, Russell R, Liang C, Berkhout B (2004) A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol* **78**(19)**:** 10814-10819

Paillart JC, Dettenhofer M, Yu XF, Ehresmann C, Ehresmann B, Marquet R (2004a) First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J Biol Chem* **279**(46)**:** 48397-48403

Paillart JC, Marquet R, Skripkin E, Ehresmann B, Ehresmann C (1994) Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* **269**(44)**:** 27486-27493

Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J (2004b) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* **2**(6)**:** 461-472

Paoletti AC, Shubsda MF, Hudson BS, Borer PN (2002) Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. *Biochemistry* **41**(51)**:** 15423-15428

Patel J, Wang SW, Izmailova E, Aldovini A (2003) The simian immunodeficiency virus 5' untranslated leader sequence plays a role in intracellular viral protein accumulation and in RNA packaging. *J Virol* **77**(11)**:** 6284-6292

Pavlakis GN, Felber BK (1990) Regulation of expression of human immunodeficiency virus. *New Biol* **2**(1)**:** 20-31

Peng C, Ho BK, Chang TW, Chang NT (1989) Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J Virol* **63**(6)**:** 2550-2556

Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* **68**(12)**:** 8017-8027

Poeschla E, Gilbert J, Li X, Huang S, Ho A, Wong-Staal F (1998) Identification of a human immunodeficiency virus type 2 (HIV-2) encapsidation determinant and transduction of nondividing human cells by HIV-2-based lentivirus vectors. *J Virol* **72**(8)**:** 6527-6536

Pollard VW, Malim MH (1998) The HIV-1 Rev protein. *Annu Rev Microbiol* **52:** 491-532

Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K, et al. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313**(6000)**:** 277-284

Reeves JD, Doms RW (2002) Human immunodeficiency virus type 2. *J Gen Virol* **83**(Pt 6)**:** 1253-1265

Rein A (1994) Retroviral RNA packaging: a review. *Arch Virol Suppl* **9:** 513-522

Russell RS, Liang C, Wainberg MA (2004) Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **1**(1)**:** 23

Sakaguchi K, Zambrano N, Baldwin ET, Shapiro BA, Erickson JW, Omichinski JG, Clore GM, Gronenborn AM, Appella E (1993) Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc Natl Acad Sci U S A* **90**(11)**:** 5219-5223

Sakuragi J, Ueda S, Iwamoto A, Shioda T (2003) Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging. *J Virol* **77**(7)**:** 4060-4069

Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH (2001) The origins of acquired immune deficiency syndrome viruses: where and when? *Philos Trans R Soc Lond B Biol Sci* **356**(1410)**:** 867-876

Shen N, Jette L, Liang C, Wainberg MA, Laughrea M (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**(12)**:** 5729-5735

Shen N, Jette L, Wainberg MA, Laughrea M (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**(21)**:** 10543-10549

Shubsda MF, McPike MP, Goodisman J, Dabrowiak JC (1999) Monomer-dimer equilibrium constants of RNA in the dimer initiation site of human immunodeficiency virus type 1. *Biochemistry* **38**(31)**:** 10147-10157

Shuljak BF (2006) Lentiviruses in ungulates. i. general features, history and prevalence. *Bulg J Vet Med* **9**(3)**:** 175-181

Sigurdsson B (1954) Maedi, a slow progressive pneumonia of sheep: An epizoological and pathological study. *Brit Vet J* **110:** 255-270

Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A* **91**(11)**:** 4945-4949

South TL, Summers MF (1993) Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci* **2**(1)**:** 3-19

Starcich B, Ratner L, Josephs SF, Okamoto T, Gallo RC, Wong-Staal F (1985) Characterization of long terminal repeat sequences of HTLV-III. *Science* **227**(4686)**:** 538-540

Strappe PM, Greatorex J, Thomas J, Biswas P, McCann E, Lever AM (2003) The packaging signal of simian immunodeficiency virus is upstream of the major splice donor

at a distance from the RNA cap site similar to that of human immunodeficiency virus types 1 and 2. *J Gen Virol* **84**(Pt 9)**:** 2423-2430

Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA (1987) The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**(6132)**:** 728-730

Stuhlmann H, Berg P (1992) Homologous recombination of copackaged retrovirus RNAs during reverse transcription. *J Virol* **66**(4)**:** 2378-2388

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

Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**(5252)**:** 1211-1213

Tiganos E, Yao XJ, Friborg J, Daniel N, Cohen EA (1997) Putative alpha-helical structures in the human immunodeficiency virus type 1 Vpu protein and CD4 are involved in binding and degradation of the CD4 molecule. *J Virol* **71**(6)**:** 4452-4460

Tsukahara T, Komatsu H, Kubo M, Obata F, Tozawa H (1996) Binding properties of human immunodeficiency virus type-2 (HIV-2) RNA corresponding to the packaging signal to its nucleocapsid protein. *Biochem Mol Biol Int* **40**(1)**:** 33-42

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**(4968)**:** 505-510

Veronese FD, DeVico AL, Copeland TD, Oroszlan S, Gallo RC, Sarngadharan MG (1985) Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science* **229**(4720)**:** 1402-1405

Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M (1985) Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**(1)**:** 9-17

Weekly M (1981) Pneumocystis Pneumonia-Los Angeles. **30**(21)**:** 1-3

Whitney JB, Wainberg MA (2006) Impaired RNA Incorporation and Dimerization in Live Attenuated Leader-Variants of SIVmac239. *Retrovirology* **3**(1)**:** 96

Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, Krausslich HG (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* **72**(4)**:** 2846-2854

Willey RL, Bonifacino JS, Potts BJ, Martin MA, Klausner RD (1988) Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proc Natl Acad Sci U S A* **85**(24)**:** 9580-9584

Willey RL, Maldarelli F, Martin MA, Strebel K (1992) Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J Virol* **66**(12)**:** 7193-7200

Zagury JF, Franchini G, Reitz M, Collalti E, Starcich B, Hall L, Fargnoli K, Jagodzinski L, Guo HG, Laure F, et al. (1988) Genetic variability between isolates of human immunodeficiency virus (HIV) type 2 is comparable to the variability among HIV type 1. *Proc Natl Acad Sci U S A* **85**(16)**:** 5941-5945

Zhang Y, Barklis E (1995) Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J Virol* **69**(9)**:** 5716-5722

Zhu P, Liu J, Bess J, Jr., Chertova E, Lifson JD, Grise H, Ofek GA, Taylor KA, Roux KH (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**(7095)**:** 847-852

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**(13)**:** 3406-3415