2011

Characterization of 5 UTR splicing and the CGI in HIV-2 RNA

Christy Strong
The University of Montana

CHARACTERIZATION OF 5'UTR SPLICING AND THE CGI IN HIV-2 RNA

By

CHRISTY LEE STRONG

B.S. Biology, Montana State University, Bozeman, MT, 2001
M.S. Biology, East Tennessee State University, Johnson City, TN, 2005

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

May 2011

Approved by:

Stephen Sprang, Associate Provost for Graduate Education
Graduate School

Dr. J. Stephen Lodmell,
Division of Biological Sciences

Dr. Jean-Marc Lanchy
Division of Biological Sciences

Dr. Scott Samuels,
Division of Biological Sciences

Dr. Scott Wetzel
Division of Biological Sciences

Dr. Kent Sugden
Department of Chemistry

ABSTRACT

Strong, Christy, Ph.D., Spring 2011 Integrative Microbiology and Biochemistry

Characterization of 5'UTR splicing and the CGI in HIV-2 RNA

Chairperson: Dr. J. Stephen Lodmell

Co-Chairperson: Dr. Jean-Marc Lanchy

HIV-2 tightly regulates several steps of its replication cycle via regulatory elements found within the 5' untranslated region of HIV-2 genomic RNA. Two elements of interest are the 5'UTR intron and the proposed long distance base pairing interaction between the C-box and G-box termed the CGI. This research focuses on the effects of 5'UTR splicing and the CGI in HIV-2 translation and replication. The central hypothesis is that both 5'UTR splicing and the CGI modulate HIV-2 translation and replication. This hypothesis was tested and supported by employing *in vitro* translation assays, SELEX, and cell culture studies. Results of these experiments demonstrate that splicing of the 5'UTR intron produces an isoform of *gag* mRNA that is specialized for high translational efficiency. Further, evidence is provided for a novel branched secondary structure within the CGI that is important for HIV-2 replication. In addition, we show that mutation of the C-box alone can enhance RNA encapsidation and mutation of the G-box can alter levels of Gag protein isoforms. These studies suggest coordinated regulation of RNA translation, dimerization, and encapsidation during HIV-2 replication. This research provides new insight into HIV translational mechanisms, in turn identifying potential antiviral targets that may be exploited for antiviral therapeutic strategies.

Acknowledgements

I would like to thank all of the people that guided and supported me during this endeavor. This work would not have been possible without you.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER ONE: INTRODUCTION

1.1 Acquired immunodeficiency syndrome

Acquired immune deficiency syndrome (AIDS) was first described in June 1981 (1,2). Four otherwise healthy homosexual men in Los Angeles, California presented with *Pneumocystitis carinii* pneumonia, a rare opportunistic infection. Over the next year, the number and type of opportunistic infections seen by the United States medical community increased significantly (3). Indeed, the affected pool population had now grown to include not only the male homosexual community but also to include intravenous drug users, hemophiliacs, as well as specific immigrant populations (4-7). The mortality rate for those afflicted was high, nearing 40% (3) and the etiologic agent was unknown.

The scientific community recognized routes of transmission to include sexual contact (genital fluids), blood transfusion and intravenous drug use (blood and blood products), as well as mother to child (*in utero* and via breast milk) (8). To diagnose AIDS, physicians looked for symptoms that included opportunistic infections, lymphopenia (low lymphocyte counts), altered T-cell ratios (T4+ helper/T8 suppressor), prolonged fever, and anergy to skin test antigens (failure to mount a response) (3,9). Alarmingly, a population of healthy homosexual males screened in New York City in 1982 displayed altered T-cell ratios, suggesting that physicians were just seeing "the tip of the iceberg" with regard to potential AIDS cases (3,10). This defined the need to determine and then be able to screen for the causative agent of AIDS.

1.2 Human immunodeficiency virus

Although the causative agent of AIDS was determined to be a retrovirus in 1983, the type of retrovirus was a source of controversy for several years (11-13). The Montagnier laboratory stated they had discovered a unique T-lymphotrophic retrovirus associated with AIDS, designated lymphadenopathy associated virus (LAV), and demonstrated LAV was cytopathic to CD4 cells (14,15). Robert Gallo's laboratory suggested that the virus associated with AIDS was in fact human T-cell

leukemia virus (HTLV), a member of the *Oncornavirus* genus and designated it HTLV-III (16-18). Eventually, it was determined that the LAV isolate and the HTLV-III isolate were identical (19), and that this unique retrovirus was not a member of oncogenic retroviruses comprising the *Oncornavirus* genus. The virus was then renamed Human Immunodeficiency Virus (HIV) in 1986 (20). It is notable that during this period, a reliable test for the retrovirus was developed (21). This allowed for screening of blood and blood products (22), as well as aided in the identification of high-risk asymptomatic HIV-infected individuals (23). In addition, screening with this test confirmed that a disease termed "slim" in Africa, was, in fact, AIDS (24). In addition, a second type of retrovirus (LAV-II) was identified in infected African individuals (25,26), which would later be designated HIV-2.

HIV-1 and HIV-2, retroviruses of the *Lentivirus* genus, are the causative agents of AIDS. Although, they display similar genomic organization (Figure 1), there are significant differences between these viruses. HIV-1 infection is found throughout the world, while HIV-2 infections are primarily concentrated in western African countries (27)

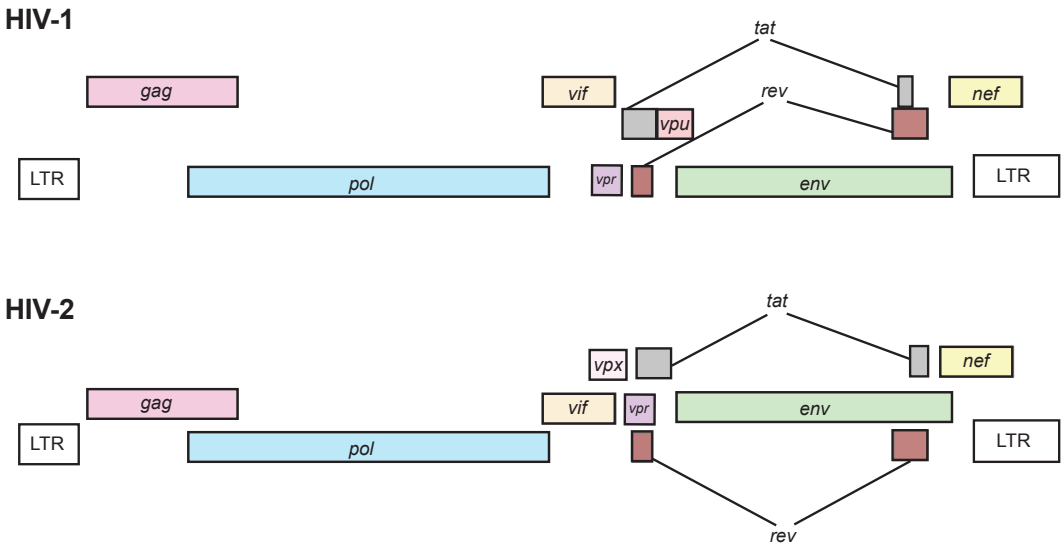

Figure 1. HIV-1 and HIV-2 genomes.
All nine encoded genes for each virus, HIV-1 and 2, are represented above. The LTR represents the long terminal repeat found at both ends of the genome. Adapted from (28).

HIV-2, an often less pathogenic virus, exhibits an overall slower progression to AIDS in infected individuals (29,30) in addition to a lower transmission rate compared to

HIV-1 (31-33). The lowered pathogenicity is associated with lower viral RNA levels in the plasma of HIV-2 infected patients. Some have attributed the lower viral RNA levels to lower HIV-2 viral DNA levels, however a consensus has not been reached on this subject (34-36). There is a significantly higher proportion of HIV-2 infected individuals termed long-term non-progressors than there is in the HIV-1 population (29,37). Long-term non-progressors or elite controllers carry lower viral loads and have lower mortality rates than other HIV-infected individuals (37,38). Indeed, the lower pathogenicity displayed by HIV-2 suggests a mechanism of control that warrants further investigation into regulation of HIV-2 replication (39)

1.3 HIV replication cycle

HIV replication begins with the binding of the virus to the cell (Figure 2). Following fusion and entry, the virus uncoats and begins the process of reverse-transcribing the vRNA genome into proviral DNA using the viral reverse transcriptase. The completed proviral DNA is transferred into the nucleus where it is integrated into a host chromosome. Once integrated, proviral DNA is transcribed by the host's RNA polymerase II into viral mRNAs.

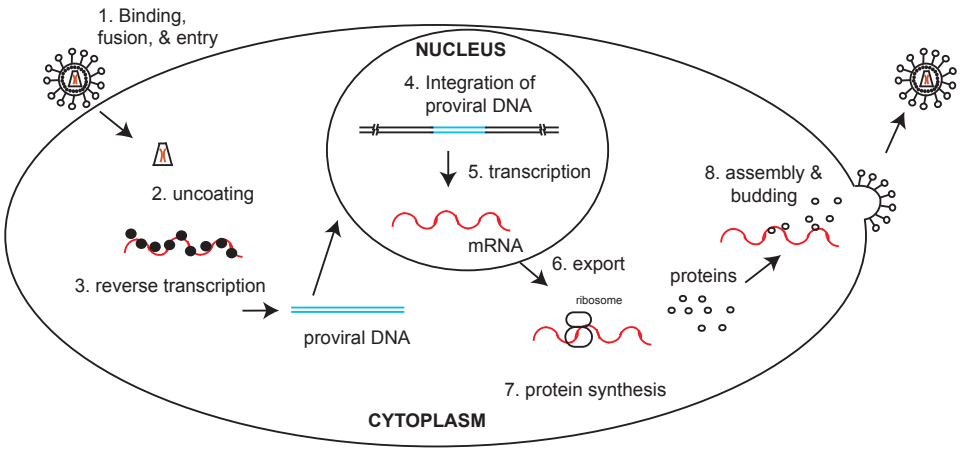

Figure 2. The HIV replication cycle.
A simplified representation of the HIV replication cycle in an infected cell. The cycle begins with binding (1) and progresses to assembly and budding (8). Both genomic RNA and mRNA are represented in red. Proviral DNA is represented in blue. Host chromosomal DNA is represented by the black lines in the nucleus.

Early in infection, multiply-spliced viral RNAs are produced and are exported out of the nucleus via host machinery. These viral RNAs include *rev* (export protein for partially spliced and unspliced viral RNAs), *nef* (a virulence factor with multiple functions), and *tat* (transactivator of viral transcription) mRNAs (40,41). Tat proteins act as transcription activators through recruitment of positive transcriptional elongation factor b (PTEFb) to the TAR (transactivation response region) element located at the 5' end of all HIV RNAs. This recruitment leads to C-terminal domain phosphorylation of the host RNA polymerase II, in turn increasing RNA synthesis. As the viral cycle progresses, other classes of viral RNAs are produced including partially spliced (*vif, vpx, gag, gag-pol*) and unspliced species (*gag, gag-pol*). These RNAs contain a Rev responsive element (RRE), which is recognized by the Rev protein (42). Rev proteins binds to the RRE of these viral RNAs allowing for export into the cytoplasm. Translation initiation of these viral RNAs occurs through multiple methods including ribosomal scanning, IRES, and frame-shifting, and results in production of various viral proteins including Gag and Gag-Pol (polyprotein precursors) (43). Once enough of these proteins has been synthesized, viral particle assembly at the plasma membrane can begin.

Meanwhile, in the rough ER, the envelope polyprotein precursor (gp140) is synthesized and post-translationally modified. As this precursor transverses the Golgi, it undergoes cleavage by a host protease resulting in two mature glycoproteins, surface unit gp125 and transmembrane gp36. These proteins then travel to the cell surface where they are incorporated into the viral particles.

Two copies of the RNA genome are encapsidated as a dimer into each viral particle (44). Dimerization of the genomic RNAs occurs through non-covalent intermolecular interactions. The strongest of these interactions involves a structural element located in the 5'UTR region, called stem loop 1 (SL1) or dimerization initiation site (45). Once the genomic RNA dimer is packaged, budding of the virus particle occurs, in tandem with activation of the viral protease. Processing of Gag and Gag-Pol by the viral protease (PR) within the virion is necessary for maturation and eventual infectivity of the virion. Cleavage of Pol gives protease (PR), reverse transcriptase (RT), and integrase (IN). Cleavage of the Gag polyprotein gives the

matrix (MA), capsid (CA), nucleocapsid (NC), and p6 proteins (40). To coordinate the steps of the replication cycle, HIV relies on interactions between viral and host proteins and on regulatory motifs contained within its RNA genome. The region with the highest density of regulatory motifs is the 5' untranslated region (5'UTR).

1.4 The 5' untranslated region of HIV

The highly conserved 5'UTR of HIV is packed with identified regulatory elements that modulate transcription and translation (TAR), packaging (pal), and genomic RNA dimerization (SL1) (46-48). These signals can be presented or masked at different stages of the replication cycle through conformational changes. Although HIV-1 and HIV-2 share overall organization of their 5'UTRs, there are significant differences between these retroviruses. The first major difference between HIV-1 and 2 is overall nucleotide length of their respective 5'UTRs (Figure 3). The 5'UTR of HIV-1 (LAI isolate) is 335 nt in length while the 5'UTR of HIV-2 (ROD isolate) is significantly longer at 545 nt. The second difference lies within the overall structure of TAR. HIV-1 has a 59 nt TAR structure that contains one major stem loop while HIV-2 has 123 nt TAR structure that contains 3 stem loops (49-51). The third difference is the presence of an intron in the 5'UTR of HIV-2 but not HIV-1 (52). This 5'UTR intron encompasses roughly half of TAR, the entire polyA signal domain, and the C-box region.

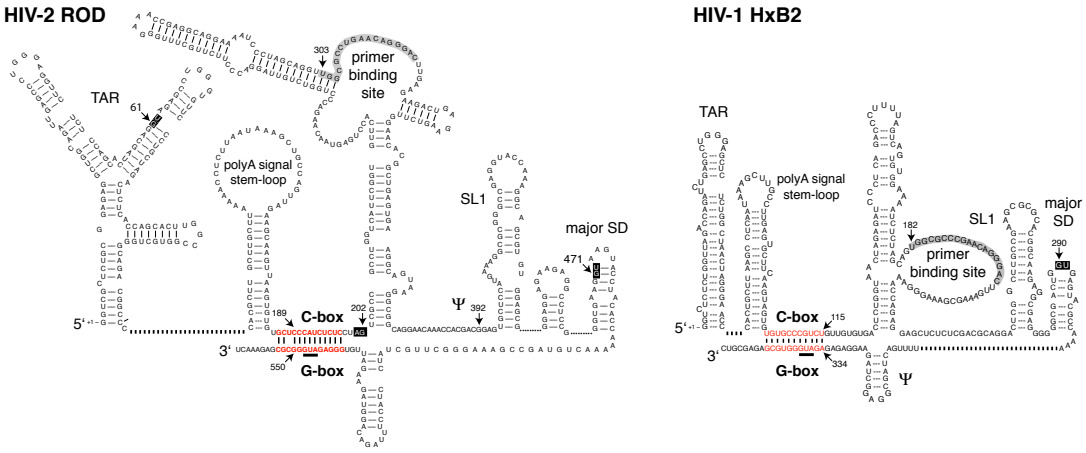

Figure 3. Schematic of Secondary Structure of HIV 5' Leaders.
The HIV-2 ROD leader region and the HIV-1 HxB2 leader region are represented on the left and right, respectively. TAR, polyA signal stem loop, C-box, primer binding site, Ψ, SL1, major SD and the G-box represent the transactivation region, the poly(A) signal domain, the core of the C-box, the tRNA primer binding site, the packaging signal, Stem loop 1, the major splice donor site, and the core of the G-box. The C-box-G-box interaction for both leader regions is highlighted in red. The HIV-2 5'UTR intron splice donor (nt 61-62) and splice acceptor sites (201-202) are highlighted in black, in addition to the major SD (471-472).

1.5 Dimerization through the 5'UTR in HIV

Each virion contains two genomic RNAs packaged in a dimeric form (44). The genomic RNA dimer is achieved through non-covalent intermolecular interactions initiated through the primary dimerization signal of each RNA. The primary dimerization signal, stem loop 1 (SL1), is located in the 5'UTR for both HIV-1 and HIV-2. Elements within the 5'UTR can regulate the presentation of SL1 (Figure 4). In HIV-1, a long-range base pairing interaction between two elements, the C-box and G-box, is found in a dimerization-competent structure (53). This dimerization-competent structure presents SL1 for dimerization (Figure 4A and 4C). In HIV-2, a similar long-range base pairing interaction results in a dimerization-incompetent structure (47). This dimerization-impairied structure masks the SL1 signal in a base pairing interaction with the packaging signal, ψ (Figure 4B and 4D).

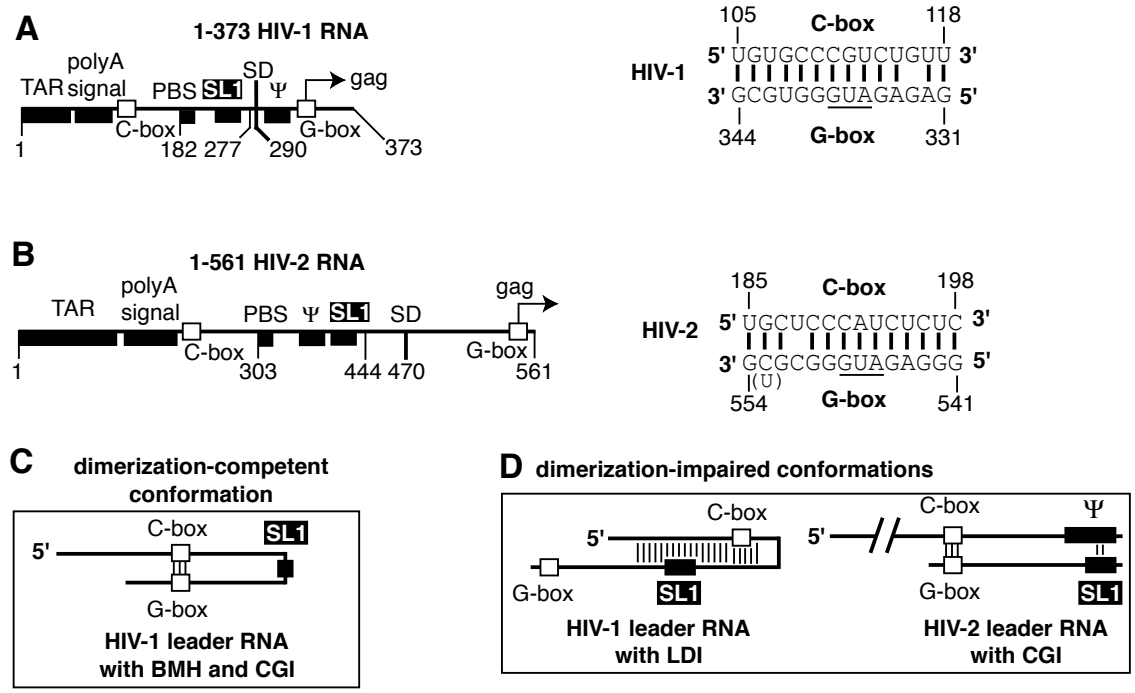

Figure 4. Dimerization-competent and incompetent RNAs.
A) A schematic of the HIV-1 leader region with a portion of the Gag coding region included. The HIV-1 CGI is represented on the top right panel. B) A schematic of the HIV-2 leader region with a portion of the Gag coding region included. The HIV-2 CGI is represented on the middle right panel. C) An illustration of the C-box, G-box, and SL1 in a HIV-1 dimerization competent conformation. D) An illustration of the C-box, G-box, and SL1 in HIV-1 and HIV-2 dimerization-impaired conformations. Adapted from (54).

1.6 The C-box and G-box of HIV

1.6.1 The C-box and the G-box

The C-box is a pyrimidine-rich regulatory element whose core encompasses nucleotides 189-196 of the 5'UTR (47,55). Studies have shown that the C-box affects both HIV-1 and HIV-2 RNA dimerization (association of two genomic RNAs) through a proposed long-range base pairing interaction (CGI) with a downstream element termed the G-box (47,53,55,56) (Figure 3). The G-box is a regulatory element that encompasses the Kozak sequence and start codon of Gag, in addition to the myristoylation signal coded in *gag*/*gag-pol*. The purine-rich G-box has many roles, one of which is the proposed CGI with the C-box.

1.6.2 The CGI in HIV

The 5' leader regions of both HIV-1 and -2 have been proposed to form two alternate conformations, the branched multiple hairpin form (BMH) and the long distance

interaction (LDI) form (57,58). Transitioning between the two conformations has been suggested to act as a riboswitch, regulating dimerization of the genomic HIV RNAs (53,57,59). The approximate location of the nucleotides involved in a long-range base pairing interaction within the LDI HIV-2 conformer (dimerization-incompetent form) was initially only inferred (58). However, a closer examination of the 5' leader using *in vitro* truncation and mutation studies revealed the identities of the two base pairing partners (47).

The C-box, whose core encompasses nucleotides 189-196, and the G-box, whose core encompasses nucleotides 543-550, were shown to inhibit dimerization through a long-range base pairing interaction that we now call CGI. Inhibition of HIV-2 RNA dimerization was suggested to occur through structural entrapment of the major dimerization site (stem loop one; SL1) by the CGI (54,60). Addition of nucleocapsid protein (NC) to dimerization reactions was shown to relieve this inhibition (47). Interestingly, formation of the CGI in HIV-1 may not be detrimental to dimerization (53,61,62). In fact, the HIV-1 CGI formation is associated with the dimerization-competent BMH conformation.

Although the long-range base pairing interaction (HIV-2 CGI or HIV-1 U5-AUG duplex) differs with respect to its regulation of HIV-1 and -2 dimerization, its possible role in modulation of *gag* translation was proposed to be similar (47,60,61). Because the CGI encompasses the *gag* translation initiation site, the formation of the CGI was postulated to influence Gag protein synthesis. An initial study in HIV-1 suggested that a leader RNA conformation that was supposed to occlude the start codon through base pairing did not affect translation, but the study did not directly probe the role of the G-box and C-box sequences when examining steady-state levels of Gag (53). At the same time, an examination of the role of the CGI in HIV-2 translation had not been undertaken, and the question of translational control through the CGI in HIV-1 or HIV-2 had not been unequivocally answered.

1.7 Translation

1.7.1 General

In order to understand the effects of 5'UTR splicing and the CGI in HIV-2 translation, an understanding of the basic mechanism of translation is needed. Eukaryotic translation of mRNA to protein is a multi-step process involving a variety of initiation and elongation factors that assist with progression of the cycle. The four mains steps of translation are initiation, elongation, termination, and ribosome recycling. To prepare for translation and generate free 40S ribosomes, recycling of inactive 80S ribosomes is necessary (Figure 5). To that end, select eIFs bind to the 40S subunit, causing dissociation from the 60S, yielding a small ribosomal subunit pool. Additional eIFs join the complex including eIF5 and the ternary complex consisting of eIF2-GTP bound to an initiating Met-tRNA$_i$ (63,64).

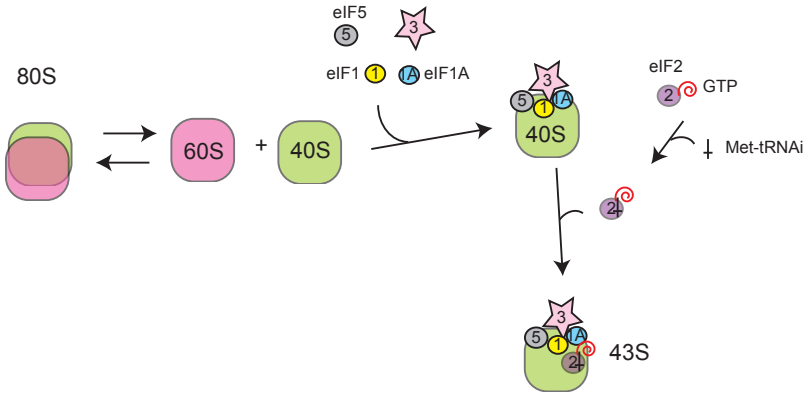

Figure 5. Preparing the ribosomal pool.
The 40S subunit pool is obtained from 80S inactive ribosome complexes. eIF1, 1A and 3 bind to the 40S resulting in its disassociation from the 60S subunit. eIF5 binds to the 40S as well as the ternary complex composed of eIF2-GTP bound to the Met-tRNA$_i$. Binding of all components results in the formation of the 43S ribosome. Adapted from (63).

Meanwhile, the mRNA to be translated undergoes activation (Figure 6) (64). This involves interaction of the mRNA with eIF4F. Once activated, the mRNA is now ready to bind to the 43S.

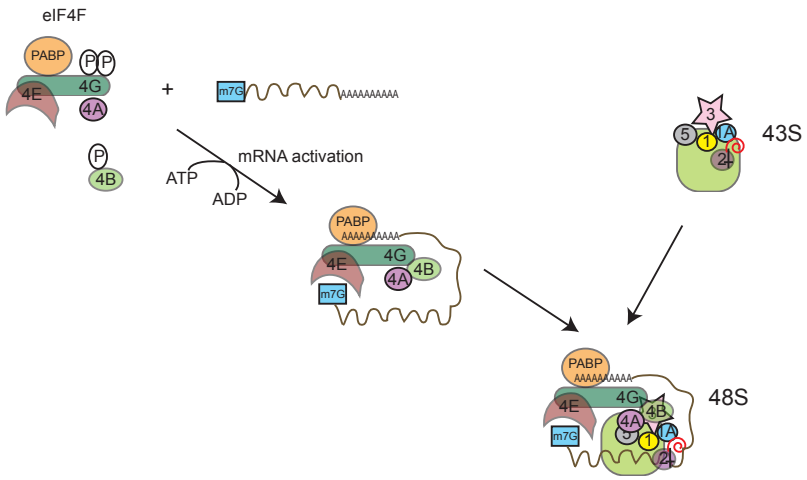

Figure 6. mRNA activation and formation of the 48S compex.
mRNA is activated through the binding of eIF4F. Once bound, secondary structure of the 5' end is removed by the RNA helicase activity of eIF4A. The bound mRNA is now ready to join with the 43S, via interactions between eIF4G and eIF3, to form the 48S compex. Adapted from (63).

Once the 48S complex is formed, the ribosome begins scanning along the mRNA in the 3'direction until an initiation AUG codon is reached (Figure 7). Once the initiating AUG codon is recognized, eIF2-GDP is released along with eIF1. The association of the 60S subunit follows leading to formation of the 80S ribosome. Translational elongation can now proceed, followed eventually by translational termination when a stop codon is encountered (63,64).

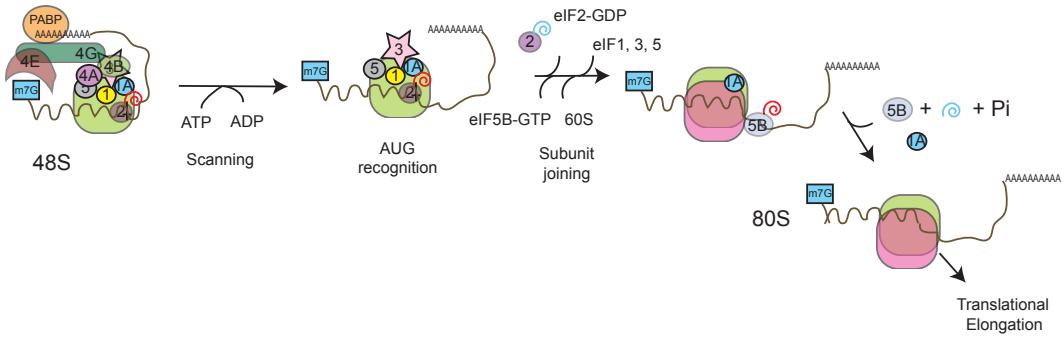

Figure 7. Formation of the 80S ribosome.
The 48S ribosome scans the mRNA in a 3' direction. Once an initiating AUG codon is recognized, eIF2-GDP is released while eIF5B-GTP is bound. Binding of the 60S subunit occurs followed by eventual formation of the 80S ribosome. The 80S ribosome is now ready to perform translational elongation.

1.7.2 Regulators of translational initiation

Regulation of protein synthesis is an essential component of the overall regulation of gene expression. Translational control is largely exerted at the level of translation initiation. Translation initiation is strongly influenced by the features of the mRNA to be translated. These include the 5' methyl^7 G-cap, 5' leader secondary structure and length, splicing, polyadenylation, eIF state, the Kozak consensus sequence of the initiating codon, and upstream 5'UTR open reading frames (uORFs).

The presence and accessibility of the 5'methyl^{7}G-cap is not an absolute requirement but has been shown to stimulate translation through recruitment of eIF4E and eIF4G (65). Binding of eIF4G allows for the associated helicase-activity of eIF4A, which reduces secondary structure of the 5' leader.

The length and overall secondary structure of the 5' leader has been shown to directly influence translational initiation. The closer the secondary structure is found to the 5'end of the mRNA, the more chance it has to inhibit ribosome entry (66,67). Additionally, increased secondary structure can impede a scanning ribosome, due to limits of the ribosome's ability to disrupt base pairing structures (68). One method for physically reducing the amount of secondary structure is by removal of RNA sequence through splicing.

Splicing out of introns in pre-mRNA removes length and secondary structure, in essence, truncating the 5' leader, which leads to enhanced translational initation (69,70). In addition, splicing allows for export of mRNAs out of the nucleus and into the cytoplasm, where the translational machinery is found (71). Once in the cytoplasm, spliced mRNAs display an enhanced ability to interact with ribosomes, which may be attributed to the presence of exon junction complexes (EJCs) on the mRNA (71,72). Exon junction complex proteins are placed upstream of exon-exon junctions following splicing, and play a role in detection of defective mRNAs in nonsense mediated decay (73). Another checkpoint for mRNA quality involves polyadenylation of the mRNA.

Improper polyadenylation of pre-mRNA can result in immediate targeting for degradation within the nucleus (71,74). In addition, polyadenylation has been suggested to influence translational initiation in the cytoplasm through the interaction

of the polyadenylated mRNA tail with polyA binding protein (PABP). Binding of PABP to the polyA tail allows for circularization of the RNA via eventual binding of PABP to eIF4G (75,76), leading to enhanced cooperation amongst the translational machinery components.

Altering the phosphorylation states of eIFs can have negative consequences for translational initiation. Phosphorylation of eIFs can prevent these factors from engaging with the translational machinery (64). Phosphorylation of eIF2α induces binding to eIF2B, which prevents eIF2α recycling, while phosphorylation of eIF4E induces binding to eIF4E binding proteins, preventing association of eIF4E with eIF4F (64). While the role of eIFs in translational initiation is important, the context of the initiating methionine codon within the mRNA is crucial to translational initiation.

An initiation codon can be part of an upstream open reading frame (uORF), which is composed of an AUG, coding sequence, and a termination codon (77). uORFs are found in the 5'UTR and have been shown to modulate initiation at downstream initiation codon. This modulation primarily occurs through two mechanisms (77). The first is by allowing ribosomes to scan through the uORF, a process termed leaky scanning. The second is by initiating translation at the uORF, and then reloading the post-termination ribosome to allow initiation at a downstream initiation site, a process termed reinitiation. Both mechanisms can be regulated to increase or decrease protein production. The leaky scanning mechanism is dependent on the Kozak sequence context of an initiating methionine codon.

The Kozak sequence encompasses nucleotides found immediately upstream (nts -6 to -1) and downstream (nt + 4) of the initiating methionine codon (A of AUG is + 1) (Table 1). Mutations within the Kozak sequence have been shown to modulate translation, with the identity of nucleotides at positions -3, -1, and + 4 drastically affecting translational initiation (78,79). The presence of a purine nucleotide (A preferred over G) at position -3 results in a 10-fold increase in translational efficiency compared to a sequence containing a pyrimidine nucleotide (78). Indeed, a poor Kozak context can result in leaky scanning whereby the ribosome scans past the first initiating methionine to a downstream initiating methionine with a stronger Kozak

sequence (80). This can result in truncated forms of a protein that differ from the full-length protein at the N-terminus. However, leaky scanning is not the only initiation method that results in truncated protein isoforms.

Table 1. The Kozak sequence.
Sequences obtained from 699 vertebrate mRNAs were compiled and then analyzed for nucleotide composition immediately upstream and downstream of the initiating AUG. Frequency of each nucleotide at each position is represented in percentage. Arrows indicate those nucleotides that have the largest influence on translational initiation within the Kozak sequence. Nucelotides at position -3 and +4 have the strongest effects. Adapted from (81).

	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4
	N	N	N	N	N	N	A	U	G	N
position:	-6	-5	-4	-3	-2	-1				+4
percent A:	17	18	25	**61**	27	15				23
percent C:	19	39	53	2	49	**55**				16
percent G:	44	23	15	36	13	21				**46**
percent U:	20	20	7	1	11	9				15

1.7.3 Types of translational initiation

Two main types of translation initiation have been identified in eukaryotes. Recognition of the 5'methyl^{7}G-cap is necessary for one type of initiation but not the other. Cap-dependent initiation is the predominant mode of translational initiation in the eukaryotic cell. The 43S ribosome recognizes the bound 5'methyl^{7}G-cap via interactions between eIF4G/eIF4B and eIF3/eIF5. Binding to the 5' end of the mRNA occurs, allowing the 48S to scan through the 5'UTR in a 3' direction. Upon recognition of the initiation AUG codon, the 48S ribosome can then proceed to form the 80S complex and begin actively translating the mRNA. In contrast, internal ribosome entry site (IRES) translational initiation does not require the presence of the 5'-methyl^{7}G-cap, instead the IRES directly recruits the translational machinery via secondary structures formed by the mRNA.

IRES translational initiation was first identified in the polio virus (82). The 5'UTR of the polio viral RNA is long, highly structured, and lacks a canonical 5' cap (83,84). Based on these characteristics, cap-dependent translation of polio RNA would be predicted to be poor. In addition, polio protease activity shuts down cap-

dependent host cell translation (85). Therefore, polio was hypothesized to initiate translation through a cap-independent mechanism. Studies determined that a sequence contained within the polio 5'UTR was sufficient for internal translational initiation. This sequence, when fused to other mRNAs, could confer IRES activity to those normally cap-dependent mRNAs (82). Once IRES activity was identified in polio, identification of IRES activity in other viruses soon followed (86). Among the identified viral IRESes, it was shown that the number of initiation factors needed to recruit the 40S ribosome varies (Figure 8) (87). Examples range from the highly structured Cricket paralysis virus (requiring only the 40S ribosome) (88) to the less structured foot and mouth disease virus (FMDV), which requires all of the initiation factors with the exception of eIF4E (87,89).

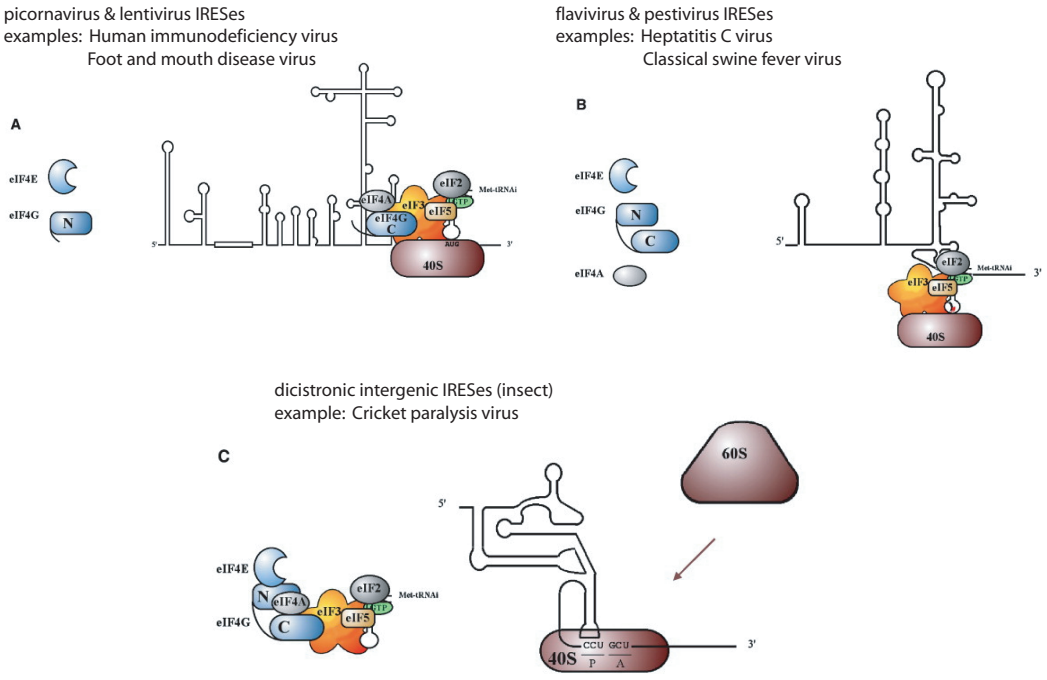

Figure 8. Types of IRES.
Three main groups of IRESes are presented above. The groups are defined by the initiation factors that are not needed for translational initiation. Viruses belonging to each group are listed, as are examples of each type of virus. Those initiation factors not needed are to the left of each IRES structure. A) IRESes that require all initiation factors except eIF4E and eIF4G. B) IRESes that require all initiation factors except eIF4E, eIF4G, and eIF4A. C) IRESes that require no initiation factors. Adapted from (86).

Two groups have independently proposed that HIV-1 contains an IRES. The Siliciano laboratory proposed that HIV-1 contains an IRES within the Gag coding

region (Figure 9) (90). Translation initiation using this IRES can produce either a full-length Gag protein (with backscanning to the initiating AUG) or a truncated Gag protein. The Sonenberg laboratory proposed that HIV-1 contains an IRES within the 5'UTR (91). Translation initiation using this IRES could produce the full-length Gag protein. However, a consensus has not been reached with regard to the presence of an IRES in HIV-1 (92).

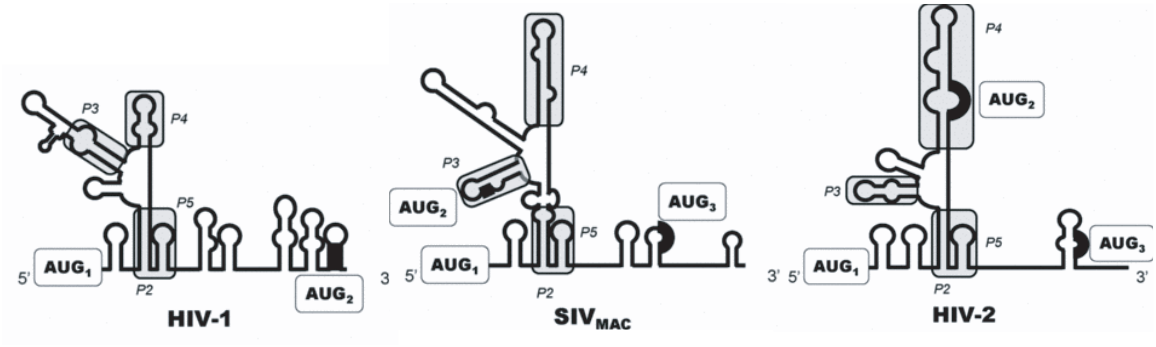

Figure 9. Retrovirus IRESes located in the Gag coding region.
A schematic of IRES structures for HIV-1, SIV_{mac}, and HIV-2, respectively. The IRESes represented above are encompassed entirely within the Gag coding region for each virus. Shared structures are shown in light grey boxes. Adapted from (93).

Although an IRES has been proposed to occur in both the 5'UTR and the Gag coding region (Figure 9) of the closely related SIV_{mac239} (94,95), a 5'UTR IRES has not been proposed to occur in HIV-2. HIV-2 has been suggested to contain an IRES located downstream of the *gag* initiation codon (96), formed entirely from nucleotides within the *gag* open reading frame (Figure 9). According to this model, three isoforms of Gag are attributed to HIV-2 IRES activity, with translation initiation able to occur at three distinct methionine codons (97). The HIV-2 IRES is proposed to comprise three distinct and independent segments that direct translational initiation at each of the respective methionine start codons (98). In addition, Weill et al. have suggested that the HIV-2 IRES is capable of recruiting three initiation complexes onto the same RNA, one positioned at each of the respective initiating methionine (99). This ability would put the HIV-2 IRES in a class of its own.

1.7.4 HIV translation

Numerous mechanisms for modulating translation of HIV mRNAs have been proposed. Some have been mentioned above, including secondary structure at the 5' end (48,100), splicing (52), and IRES (43,91,96). Others are more specific to HIV replication, including the actions of TAR, Rev, and Protease.

Because of HIV-2 TAR's significant secondary structure and location, a role in modulation of translation was obvious. However, TAR also modulates translation by another means, through activation of PKR (101). PKR, a double stranded RNA dependent kinase, when activated will phosphorylate eIF2α, preventing recycling of this elongation factor. Interestingly, HIV circumvents this self-induced inhibition through competitive binding of PKR by the HIV Tat protein, thus inhibiting phosphorylation of eIF2α (102). This is not the only HIV protein that modulates translation

The Rev protein also plays a critical role in translation, primarily through its interaction with unspliced and incompletely spliced HIV mRNAs. In eukaryotic cells, unspliced and incompletely spliced mRNAs are not efficiently exported out of the nucleus. HIV evades this quality control step through the actions of Rev. Rev proteins enter the nucleus and bind to and coat mRNAs containing the Rev responsive element (RRE). This coating allows for the export of HIV unspliced and incompletely spliced mRNAs out of nucleus and into the cytoplasm. Moreover, association of the Rev protein with the HIV mRNAs has been shown to increase the association of the mRNAs with polyribosomes, in turn modulating translation (103).

HIV protease is another viral protein that can regulate HIV translation by cleaving eIF4G (104) and PABP (105) . Removal of functional eIF4G from the initiation factor pool prevents recognition of the 5'methyl^{7}G-cap and therefore activation of mRNA. Cleavage of PABP prevents its association with the polyA tail of mRNAs as well as its interaction with initiation factors. Consequently, HIV protease activity leads to down-regulation of cap-dependent translation initiation and up-regulation of cap-independent IRES initiation.

1.8 Hypotheses

TAR (transcription and translation), pal (packaging), the C-box (dimerization) and the G-box (dimerization) all play roles in regulation of HIV-2 replication. As illustrated above, a regulator may exert influence during more than one step in the replication cycle. We propose that this is also the case for the CGI. We have previously shown that the CGI regulates *in vitro* dimerization. Based on the composition of the G-box (Kozak sequence and initiation codon of *gag*), the formation or disruption of the CGI may also influence Gag synthesis. In conjunction, the position of the C-box within the 5'UTR may play a role in Gag regulation. The C-box is contained within the 5'UTR intron. If the 5'UTR intron were spliced out of a *gag* RNA, the C-box, the polyA signal domain, and half of TAR would be removed. Indeed, secondary structure and length would be reduced in tandem with the disruption of the CGI. Therefore, a *gag* mRNA undergoing 5'UTR splicing should have an improved 5'UTR with regard to ribosomal scanning. To that end, this study focuses on the roles of the 5'UTR intron and the proposed long distance base pairing interaction (CGI) between the C-box and the G-box.

The overall hypotheses for this research are: 1) A 5'UTR-spliced *gag* mRNA isoform exists, specialized for producing Gag protein and 2) Maintaining the ability to form a proposed CGI between the C-box and G-box is critical for one or more steps of HIV-2 replication. This study employs biochemical, molecular, and cell culture techniques to test these hypotheses. I have identified functional roles and requirements of the 5'UTR intron and the proposed CGI, providing further insight into how to target various steps of the HIV-replication cycle. Previous research on HIV-2 5'UTR splicing neglected to address its effect on *gag* mRNA. Our studies have remedied this omission. In addition, our studies on the proposed CGI have finally put to rest whether or not this proposed interaction occurs in HIV-2 RNA.

1.8.1 Specific Aim 1. Characterize the role of 5'UTR splicing in Gag production.

HIV RNAs are grouped into three classes based on alternative splicing (40,106,107). The first class consists of multiply spliced HIV RNAs including *tat*, *rev*, and *nef*. The

second class consists of partially spliced HIV RNAs including *env*, *vif*, *vpu*, and *vpr*. The third class consists of unspliced HIV RNAs including *gag* and *gag-pol*. Alternative splicing in HIV-1 occurs in a sequential manner, where the most 5' intron is removed first (108). The splice donor site of the most 5' intron in HIV-1 is the major splice donor site (SD). The major SD is located at the end of the 5'UTR. Like HIV-1, HIV-2's major splice donor site is also located at the end of the 5'UTR. However, unlike HIV-1, HIV-2 contains an intron upstream of the major SD. The 5'UTR intron is contained entirely within the 5'UTR. The location of the 5'UTR intron suggest that is possible to have a splicing event and still retain the entire *gag* and *gag-pol* reading frame. In essence, having 5'UTR spliced and 5'UTR unspliced *gag* and *gag-pol* RNA isoforms is possible in HIV-2. 5'UTR splicing of *gag* and *gag-pol* does occur in SIV_{mac239}, a closely related retrovirus (109). The spliced leader of SIV_{mac239} *gag*/*gag-pol* mRNAs display enhanced translational efficiency compared to the unspliced leader (109). Whether this is applicable in HIV-2 is not known.

I hypothesized that 5'UTR splicing of *gag/gag-pol* mRNA occurs during HIV-2 infection. Removal of the 5'UTR should increase translational efficiency, producing more Gag from the 5'UTR spliced isoform than its unspliced counterpart. This hypothesis was tested by comparing translation of 5'UTR unspliced and 5'UTR intronless constructs in cell culture and translation experiments.

1.8.2 Specific Aim 2. Characterize the sequence and/or base pairing requirements of the C-box and G-box.

The requirement of the presence of both the C-box and G-box RNA motifs to modulate RNA dimerization *in vitro*, in HIV-2, suggested that a long distance base pairing interaction (CGI) occurs between the two motifs during viral replication. I hypothesized that the sequence requirements of both motifs are dependent on their ability to form a long distance base pairing interaction (CGI) with their respective partner. I tested this hypothesis by randomizing the C-box and G-box motifs using a modified SELEX (Systematic Evolution of Ligands through Exponential Enrichment) technique to produce randomized libraries of proviral DNAs. These libraries were

used in cell culture experiments to determine sequence and/or base pairing requirements.

CHAPTER TWO: A 5'UTR-SPLICED MRNA ISOFORM IS SPECIALIZED FOR ENHANCED HIV-2 *GAG* TRANSLATION

Christy L. Strong, Jean-Marc Lanchy, Abdoulaye Dieng-Sarr, Phyllis Kanki, and J. Stephen Lodmell.

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

Corresponding author: J. Stephen Lodmell

This chapter is a modified version of the manuscript published in the Journal of Molecular Biology in August 2009

Strong CL, Lanchy JM, Dieng-Sarr A, Kanki PJ, and Lodmell JS. 2009. A 5'UTR-Spliced mRNA Isoform is Specialized for Enhanced HIV-2 *gag* Translation. J. Mol. Biol. **39**: 426-437.

This manuscript represents the experiments done by the author of this dissertation with the exception of cell culture studies of HIV-2, which were performed by Dr. Jean-Marc Lanchy (Figures 11, 12, 14C, and 17.)

2.1 Abstract

Full-length unspliced genomic RNA plays critical roles in HIV replication, serving both as mRNA for synthesis of the key viral polyproteins Gag and Gag-Pol and as genomic RNA for encapsidation into assembling viral particles. We show that a second *gag* mRNA species that differs from the genomic RNA molecule by the absence of an intron in the 5' untranslated region (5'UTR) is produced during HIV-2 replication in cell culture and in infected patients. We developed a cotransfection system in which epitopically tagged Gag proteins can be traced back to their mRNA origins in the translation pool. We show that a disproportionate amount of Gag is translated from 5'UTR intron-spliced mRNAs, demonstrating a role for the 5'UTR intron in the regulation of *gag* translation. To further characterize the effects of the HIV-2 5'UTR on translation, we fused wild type, spliced, or mutant leader RNA constructs to a luciferase reporter gene and assayed their translation in reticulocyte lysates. These assays confirmed that leaders lacking the 5'UTR intron increased translational efficiency compared to that of the unspliced leader. In addition, we found that removal or mutagenesis of the C-box, a pyrimidine-rich sequence located in the 5'UTR intron and previously shown to affect RNA dimerization, also strongly influenced translational efficiency. These results suggest that the splicing of both the 5'UTR intron and the C-box element have key roles in regulation of HIV-2 *gag* translation *in vitro* and *in vivo*.

2.2 Introduction

Unspliced genomic-length RNA plays at least two central roles in HIV replication. It serves both as messenger RNA for the synthesis of the key viral proteins Gag and Gag-Pol and as genomic RNA for encapsidation into virus particles. Regulation of the use of this full-length viral RNA as either mRNA or genomic RNA (or both) is likely to be critical for efficient viral replication. In fact, regulation of many viral RNA functions, including translation and encapsidation, occurs through the 5' untranslated region (UTR), a highly conserved region in human immunodeficiency virus (HIV) (46).

Previous research has shown that the 5'UTR influences translation of HIV-1 mRNAs. The overall length of the HIV-1 5'UTR was shown to contribute to a decrease in initiation efficiency by increasing the requirement for ribosomal scanning in advance of the AUG start codon (110). In addition, the position and stability of the highly structured transactivation responsive element (TAR) at the very 5' end of the UTR plays a role in impeding 5' cap accessibility and ribosomal scanning (48).

HIV-2 and HIV-1 share a common overall organization of their 5'UTRs; thus, it is likely that HIV-2 mRNA translation is influenced by the 5'UTR. However, differences between the HIV-1 and HIV-2 UTRs may necessitate alternative modes of translational regulation. First, the HIV-1 (LAI isolate) leader region is significantly shorter at 335 nucleotides (nt) long compared to the HIV-2 (ROD isolate) leader region, which is 545 nt long. Second, HIV-1 contains a 59-nt TAR structure consisting of one major stem-loop, while HIV-2 contains a larger 123-nt TAR structure consisting of three stem-loops (49-51). Third, unlike HIV-1, the HIV-2 5'UTR contains a predicted intron (52). Splicing of this intron results in the removal of 140 nucleotides of 5' leader sequence including half of the large TAR structure sequence, the entire poly (A) signal domain, and the C-box region (Figure 10; compare A to B).

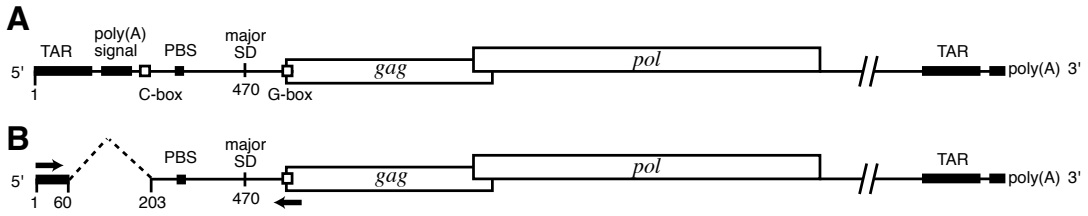

Figure 10. Schematic of the two forms of HIV-2 gag (and gag-pol) mRNAs.
The two forms are 5'UTR-unspliced (A) and 5'UTR-spliced (B), respectively. TAR, poly(A) signal, C-box, PBS, major SD, G-box, gag, and pol represent the transactivation region, the poly(A) signal domain, the core of the C-box, the tRNA primer binding site, the major SD site, the core of the G-box, the Gag protein coding region, and the Pol protein coding region, respectively.

Evidence of 5'TR intron splicing was first revealed in a simian immunodeficiency virus (SIV_{mac239}), a phylogenetic relative of HIV-2 (111). Most viral mRNA species found in SIV_{mac239} infected cells were present in both the 5'UTR intron spliced and the unspliced forms (109,111). In HIV-2, evidence of 5'UTR splicing was revealed in multiply spliced mRNA species, but its occurrence in *gag/gag-pol* coding mRNA species was not investigated (52).

The effects of 5'UTR intron presence or absence on translation were first studied using partial leader SIV_{mac239} constructs truncated at the major splice donor (SD) site, thus mimicking the 5'UTR of multiply spliced mRNA species (109). Translation of a downstream reporter gene was found to be more efficient if the 5'UTR intron was not present (109). However, a direct examination of the role of the HIV-2 5'UTR and its splicing in *gag* (and *gag-pol*) translational regulation has not been conducted and is of unique interest for several reasons. First, the 5'UTR spliced *gag* is singly rather than multiply spliced at a site that does not employ the otherwise ubiquitous major SD site. Second, it is plausible the 5'UTR splicing could regulate translation by shortening and removing secondary structure in the 5'UTR. Finally, this splicing event could regulate translation by removal of RNA signals that have importance in other regulatory events, including the long-range interaction between the 5'UTR element known as the C-box and the G-box that overlaps the *gag* translation initiation codon (47,58).

In this study, we demonstrate the presence of both 5'UTR spliced and unspliced *gag* mRNA species in transfected and infected cells and in peripheral blood

mononuclear cells (PBMCs) isolated from HIV-2 infected patients. We show in transfected cells that a 5'UTR unspliced construct yielded less Gag compared to its 5'UTR spliced counterpart. To further characterize the effect of the HIV-2 5'UTR on translation, we tested the translation of a luciferase reporter gene fused to various 5'UTR leader constructs in reticulocyte lysates. We find that leaders lacking the 5'UTR intron increased translational efficiency compared to constructs harboring the unspliced leader. Furthermore, our *in vitro* and cell culture studies implicate the C-box, which is part of the intronic sequence removed by 5'UTR splicing, as a contributor to translational regulation. Taken together, our results underscore the importance of the 5'UTR in the coordinated regulation of several essential viral replicative functions *in vitro* and *in vivo*.

2.3 MATERIALS AND METHODS

2.3.1 Plasmid construction

To engineer mutations in the 5' untranslated region or to insert a tag in the C-terminal domain of the Gag protein, we used plasmids derived from the modified pROD10 containing the full-length HIV-2 ROD (112-114). The 12,828-bp-long modified plamsid pROD10 was provided by the EU Programme EVA/MRC Centralised Facility for AIDS Reagents, NIBSC, UK (Grant Number QLK2-CT-1999-00609 and GP828102). The AatII-XhoI fragment of modified pROD10 digestion was subcloned in pGEM-7Zf+ (Promega). This plasmid, pGRAXS, contains the upstream long terminal repeat (LTR) and most of the Gag coding region, up to the XhoI site (115). The numbering is based on the sequence of the HIV-2 ROD isolate (GenBank no. M15390), which is contained with modified pROD10. Mutations were introduced in pGRAXS using either a ligation-PCR protocol (to delete the 5'UTR intron, nucleotides 61-202) or the QuikChange II XL site-directed mutagenesis kit (Stratagene) and the appropriate primers (to mutate the 5'UTR SD site in the TAR structure or the C-box). The 5'UTR SD(-) mutations removes the SD site while keeping the second TAR stem structure intact (nucleotides 61-62 GG changed to CC and nucleotides 78-79 CC changed to GG). The AatII-XhoI fragment of plasmids bearing the mutations was then reinserted in the modified pROD10 backbone.

To introduce the same mutation [5'UTR SD(-)] in both LTRs, the mutation was separately introduced in the unique upstream LTR of pGRAXS as described above and in the unique downstream LTR of the pAVR plamsid using the QuikChange II XL kit. The pAVR plasmid corresponds to modified pROD10 missing an internal AvrII-AvrII fragment encompassing the upstream LTR and *gag* region and thus contains only the downstream LTR. After mutagenesis, the AvrII-AvrII fragment containing the mutated downstream LTR in a pGRAXS backbone was reinserted in the AvrII site of pAVR mutated in the upstream LTR. To insert the FLAG-tag coding sequence (5'-gat tac aag gac gac gac gac aag-3', N-terminus-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-C-terminus) near the end and in frame with the *gag* gene, we used a single PCR step followed by replacement of the FLAG-tag-containing XhoI-Bsu361 fragment in wild type and mutated HIV-2 plasmids. All constructs were checked by DNA sequencing.

2.3.2 Cell culture and transfection

COS-7 cells ere maintained in Dulbecco's modified Eagle's (DMEM) medium supplemented with 10% fetal calf serum, penicillin, and streptomycin (Invitrogen). Transient transfection of COS-7 cells was performed using Trans-IT-COS transfection kit (Mirus). Cells and media were harvested two days posttransfection. HIV-2 capsid protein levels in the media were quantified by a p27 ELISA (SIV p27 ELISA; Zeptometrix).

2.3.3 Cell culture and infection

C8166 cells (116) were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% fetal calf serum, penicillin, streptomycin, and glutamine (Invitrogen). Media from transfected COS-7 cells containing 10 ng of HIV-2 p27 capsid protein, as determined by ELISA, was used to infect 2 x 10^4 C8166 cells (NIH AIDS reagent, no. 4040). After 12 hours, cells were washed twice and resuspended in RPMI/fetal bovine serum. A media aliquot was then taken and served as a reference (ELISA p27). Aliquots of media were taken daily, spun to remove cells, and the supernatant frozen at -80˚C. The viral replication was followed for one or two weeks by quantifying the level of Cap27 protein in the supernatant. The

Cap27 protein concentration was determined by ELISA (Retro-Tek SIV p27 Antigen ELISA kit from Zeptometrix).

2.3.4 Protein isolation and analysis

Intracellular COS-7 RNA and proteins were harvested 48 hours after transfection by washing and scraping cells in phosphate-buffered saline. To prevent Gag cleavages, an HIV protease inhibitor, saquinavir, was added to the media of the transfected cells 24 hours before harvesting (10 μM final concentration). One half of the cells was harvested in RNA lysis buffer for genomic RNA quantitation (see below), while the other half was lysed in radioimmunoprecipitation assay buffer (Santa Cruz Biotechnology Inc.; Tris-HCl, 50 mM, pH 7.4, 150 mM NaCl, 0.1% SDS, 1% Nonidet P-40, 0.5% sodium deoxycholate with protease inhibitors). After lysis on ice, cells were passaged six times through a 27-gauge needle, centrifuged for 10 minutes at 21,000 g and 4˚C, and supernatants were transferred to new tubes. Samples were fractionated on SDS-8% polyacrylamide gels. After electrotransfer onto polyvinylidene difluoride (PVDF) membranes, Gag proteins were visualized by Western blotting with one of the following primary antibodies: biotin-labeled anti-capsid antibody (p27 ELISA kit from Zeptometrix) or goat anti-FLAG tag antibody (sc-807G from Santa Cruz Biotechnology). Bound primary antibodies were detected using either streptavidin-horseradish peroxidase (p27 ELISA kit from Zeptometrix) or chicken anti-goat IgG-horseradish peroxidase (sc2953 from Santa Cruz Biotechnology) by electrochemiluminescence (ECL Plus Western Blotting Detection Reagents from GE Healthcare) on a Fujifilm LAS-3000.

For virion fractions, medium from transfected or infected cells was carefully removed without disturbing the cell layer, filtered through 0.45 μm filters to remove cells and debris, and centrifuged at 21,000 g for 2 hours at 4˚C to pellet virus. The pellet was gently washed and resuspended in radioimmunoprecipitation buffer. Aliquots were then treated as above for electrophoresis, blotting, and visualization.

2.3.5 PBMC sample acquisition

The PBMCs of HIV-2-infected individuals were obtained from a cohort of commercial sex workers in Senegal that has been followed since 1985. Detailed

clinical and epidemiological data concerning this cohort have been published elsewhere (32,117). The patients were all antiretroviral therapy-naïve, were asymptomatic, and had $CD4^+$ T-cell counts above 200 per microliter at the time of sample acquisition. All subjects signed informed consent documents and participated in protocols approved by the Conseil National de Lutte Contre le SIDA Comite Ethique et Juridique (Senegal) and the Harvard School of Public Health Human Subjects Committee.

2.3.6 RNA isolation and analysis

The total cellular RNA from transfected COS-7 or infected C8166 cells and the extracellular viral RNA fractions were purified using Stratagene's Absolutely RNA mini-prep and micro-prep kits, respectively. To pellet the viral particles, a fraction of the media was centrifuged for 2 hours at 4˚C at 21,000 g. Purified RNAs from cells and viruses were used for RT-PCR (OneStep RT-PCR kit, Qiagen), or Rnase protection assay (RPA III assay, Ambion), as described by the manufacturers. Briefly, the Rnase protection assay was done using an antisense RNA probe complementary to the 1988-2137 region of HIV-2 ROD isolate containing the FLAG tag insertion at the end of the *gag* gene. The antisense region was cloned into the pGEM7Zf(+) vector (Novagen) so that the T7 transcript has 41 nt of vector origin at its 5' end. This non-HIV-2 tail is used as a marker of the RNase's digestion efficiency during the RPA experiment. Up to three protected bands are expected: one 174-nt-long band corresponding to the FLAG-tagged *gag* gene and two protected bands corresponding to the untagged *gag* RNA, 49 and 101 nt long.

For RNA derived from human PMBCs, RNA was extracted as described in MacNeil *et al* (117). $PolyA^+$ RNA was reverse transcribed using random hexamer primers. cDNAs were amplified first using primers P1 and P2. The P2 primer selectively targets and amplifies HIV-2 *gag* species, since its binding site is located downstream of the major SD site. The P1/P2 PCR products encompassed nucleotides 2 to 555. A second, nested PCR was performed on an aliquot of the first P1/P2 PCR reaction using the P3/P4 primer pair. The P3/P4 PCR products encompassed nucleotides 40-400. The final products are predicted to be either 361-bp-long DNA,

for the unspliced *gag* species, or 219-bp-long DNA if the 142-nt-long 5'UTR intron was spliced out. These PCR products were cloned into pGEM7-Zf(+) and sequenced to verify their identities.

2.3.7 Luciferase template construction for in vitro transcription

The full-length and truncated HIV-2 inserts (5'UTR plus the first eight codons of *gag*) were generated using PCR. A sense primer (sNHErod) containing a NheI site and an antisense primer (asBsoBIrod) containing a BsoBI site were used to amplify the first 1-569 nucleotides of the HIV-2 genomic RNA, ROD isolate (GenBank M15390; genomic RNA sequence starts at 1) and the first 1-569 nucleotides minus the 5'UTR intron (61-202) of HIV-2 genomic RNA, respectively (55). Sense primers (sNHE186, sNHE198, sNHE486) containing a NheI site in combination with an antisense primer (asBsoBIrod) containing a BsoBI site were used to amplify 186-569, 198-569, and 486-569 of HIV-2 genomic RNA, ROD isolate, respectively. The PCR products were digested and then inserted via NheI/BsoBI sites into the pHRL-CMV vector (just upstream of the second codon of the hRLuc reporter gene) (Promega). The C-box mutants were generated using the sNHErod primer and asBsoBIrod primer to amplify the first 569 nucleotides of mutant A and the mutant G constructs. All constructs were checked by DNA sequencing.

2.3.8 In vitro transcription

The plasmids were linearized via BamHI digestion. The linearized DNA templates were then used in mScript (Epicentre) transcription reactions to synthesize RNA. Following incubation, the transcription reactions were treated with Rnase-free Dnase followed by ammonium acetate precipitation. The RNAs underwent 5' capping and polyadenylation (mScript mRNA Production System, Epicentre) followed by ammonium acetate precipitation.

2.3.9 In vitro translation

Equimolar amounts (46 nM) of the various mRNAs were used in comparison translation reaction sets. A 25.2-µL volume of mRNA in water was denatured at 65˚C for 3 minutes followed by snap cooling on ice for 3 minutes. The translation

mix (70 µL of rabbit reticulocyte lysate, 1 µL of amino acid mixture minus leucine, 1 µL of amino acid mixture minus methionine, and 2.8 µL of KCl 2.5M; Flexi Rabbit Reticulocyte Lysate System, Promega) was added to the mRNA sample to initiate the reaction. The translation reactions were incubated at 30˚C for 1 hour. A 22.0-µL aliquot was taken from each sample at various time points and immediately placed on ice. The aliquot tubes stayed on ice until the luciferase assays were performed.

2.3.10 Luciferase assays

Twenty microliters of each translation reaction aliquot was plated (in sets) on a 96-well plate. A Tropix TR717 microplate luminometer was used to autoinject each sample well with 100 µL of *Renilla* Luciferase Assay Reagent (Promega) and quantify/record the resulting chemiluminescence product.

2.4 RESULTS

2.4.1 5'UTR splicing of gag mRNA in cell culture

In HIV-1, there is only one type of *gag* and *gag-pol* mRNA species, the unspliced genome-length RNA species. Because HIV-2 RNA contains a predicted intron in the 5' UTR (52), we sought to detect the presence of a second 5'UTR-spliced *gag* mRNA species in transfected and infected cells. RNAs derived from PBMCs from HIV-2-seropositive but asymptomatic patients in Senegal (117) were also analyzed using nested PCR (RT-PCR) for the presence of 5'UTR spliced RNA.

COS-7 (monkey kidney fibroblast) cells were transfected with full-length HIV-2 proviral DNA to produce viral particles. We examined the 5'UTR of *gag* mRNA species in both the intracellular RNA fraction and the extracellular viral particles of the transfected cells by RT-PCR. The intracellular RNA yielded two distinct products (Figure 11). The primer asECO561 allowed us to amplify only the leader region of *gag* mRNA molecules that were not spliced at the major SD site.

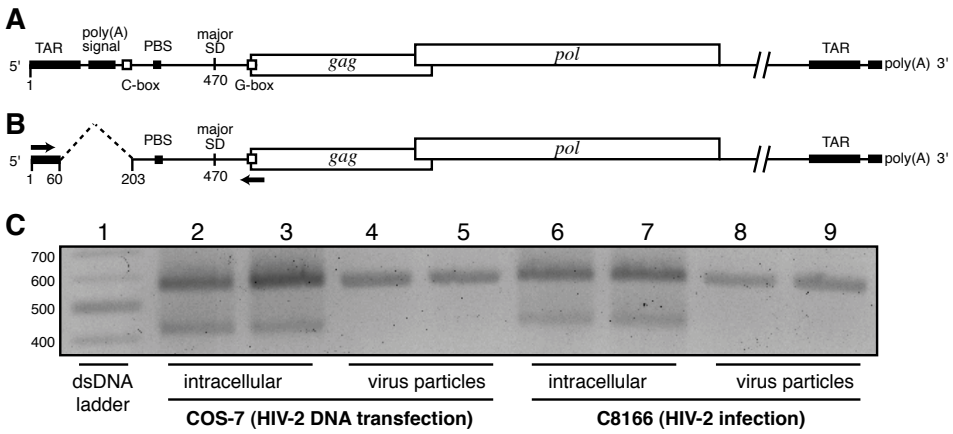

Figure 11. Analysis of HIV-2 gag mRNA species in cell culture.
Schematic of the two forms of HIV-2 *gag* (and *gag-pol*) mRNAs, 5'UTR-unspliced (A) and 5'UTR spliced (B). (C) Gel electrophoresis analysis of RT-PCR products from COS-7 cells transfected with wild type HIV-2 plasmid DNA (lanes 2-5) or from C8166 cells infected with wild type HIV-2 (lanes 6-9). The origin of the RNA used in the RT-PCRs, intracellular or extracellular (i.e., viral particles), is indicated below the gel. The primers used for the RT-PCRs are represented by arrows in (B).

Sequencing of the gel-purified bands revealed that the larger band corresponded to the leader region of the unspliced genomic RNA species, while the smaller band corresponded to a leader region with nucleotides 61-202 missing. The smaller band was clearly the 5'UTR spliced *gag* mRNA, since the missing sequence corresponds exactly to the predicted 5'UTR intron and mutation of the splice donor or splice acceptor site resulted in disappearance of the smaller band (data not shown). Furthermore, RT-PCR analysis of the viral particles produced from COS-7 cells revealed that the unspliced genomic RNA was encapsidated, while the 5'UTR spliced variant was excluded (Figure 11C).

Viral particles produced from the transfected COS-7 cells were used to infect permissive C8166 (human T lymphocyte) cells. As with the COS-7 cells, both spliced and unspliced *gag* mRNA species were found inside the infected C8166 cells, while only the unspliced genomic RNA was seen in C8166-produced viral particles (Figure 11C), suggesting that the 5'UTR spliced *gag* mRNA species lacks a functional packaging signal.

2.4.2 5'UTR splicing of gag mRNA in PBMCs from HIV-2-infected persons

Lysates from PBMCs from five HIV-2 asymptomatic seropositive patients were subjected to nested RT-PCR wherein the first PCR primer set was designed to

amplify *gag* mRNAs, and the second primer set to further amplify a fragment within the 5'UTR containing a predicted intron (Figure 12).

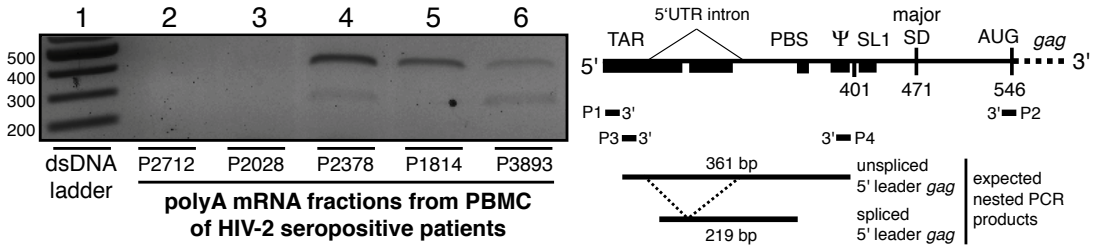

Figure 12. Gel electrophoresis analysis of RT-PCR products from HIV-2-seropositive patients. After reverse transcription, a first PCR was performed with primers P1 and P2. An aliquot of the P1/P2 PCR product was then used as template for a second nested PCR with primers P3 and P4 (lanes 2-6). Sample identification numbers from the original reference are indicated (118). Primers and expected products are described in the right panel.

Three of the five samples produced bands corresponding to HIV-2 leader RNA. HIV-2 RNA was not seen in all samples, perhaps due either to sequence variation in these isolates that preventer primer hybridization, or the level of the viral mRNA was below our detection limit. Of the three samples that yielded PCR product, two samples produced obvious bands, indicating 5' UTR-spliced RNA was present; a third sample had lower or no obvious 5'UTR spliced RNA (Figure 12). The PCR products were sequenced and it was determined that the *gag* mRNAs were derived from sources other than the stock HIV-2 ROD isolates used in our laboratory (data not shown). The presence of the 5'UTR spliced *gag* mRNA in samples derived from human patients confirms that this splicing event is not limited to experimental infections or transfections.

2.4.3 Effect of the 5'UTR intron on Gag production in cell culture

Since two *gag* mRNA species are found in the cytoplasm of both transfected and infected cells, we wanted to assess their relative contributions to the overall Gag protein production. To this end, we cotransfected COS-7 cells with two HIV-2 DNA plasmids. One plasmid provided only the unspliced *gag* mRNA species because of a mutation in the 5'UTR SD site that prevents splicing [called SD(-)] (Figure 13A). The other plasmid provided the *gag* mRNA species lacking the 5'UTR intron through

deletion of nucleotides 61-202 in the proviral DNA (called Δ intron) (Figure 13B). To differentiate which Gag proteins came from which *gag* mRNA species, we inserted a FLAG tag at the end of the *gag* gene of one or the other of the cotransfected plasmids (Figure 13).

RNA and proteins were harvested from the cotransfected cells, with half of the intracellular fraction processed for protein analysis and the other half for RNA analysis. Because of the eight amino acid addition in the FLAG-tagged proteins, an electrophoretic migration difference could be discerned between FLAG-tagged and untagged Gag proteins (Figure 13C, top panel). Those with a FLAG tag migrated slightly slower, and all proteins were recognized by an antibody directed against the capsid domain (anti-capsid antibody). As a control, we probed the western blot with an anti-FLAG antibody to directly verify the presence and position of the FLAG-tagged proteins (Figure 13C, bottom panel). The 5'UTR Δ-intron mRNA produced more Gag protein than did the 5'UTR SD (-) RNA (Figure 13C, lanes 4 and 5 in top panel). We also confirmed that Gag expression was not significantly affected by the presence or absence of the FLAG tag (Figure 13C, lanes 4 and 5 in top panel).

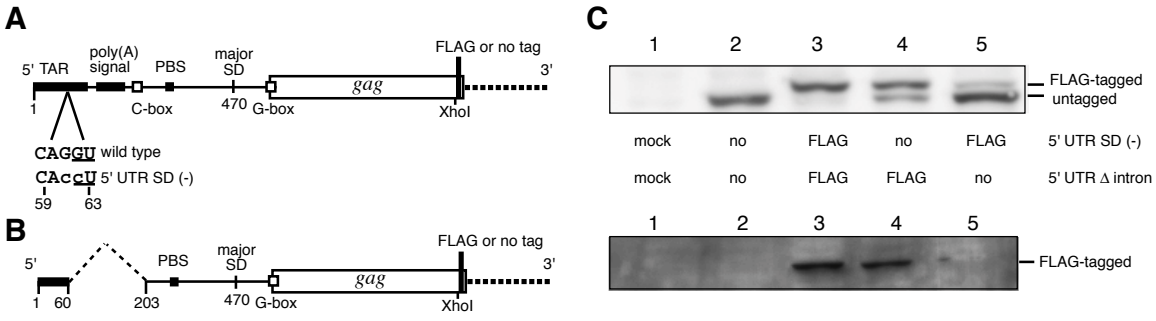

Figure 13. Effects of the 5'UTR intron on HIV-2 Gag protein species.
(A) The unique *gag* mRNA species produced upon transfection of HIV-2 5'UTR SD(-) plasmid DNA in COS-7 cells. The HIV-2 5'UTR SD(-) DNA contains mutations designed to abrogate the underlined 5'UTR SD site. The substitutions are in lower case letters. (B) The unique *gag* mRNA species produced upon transfection of HIV-2 5'UTR Δ-intron plasmid DNA in COS-7 cells. The HIV-2 5'UTR Δ-intron DNA contains a deletion of nucleotides 61-202, which corresponds to the 5'UTR intron. Two versions of 5'UTR SD(-) and Δ-intron plasmid DNA were used, one without a tag and one with a FLAG tag inserted downstream of the XhoI site in the *gag* gene. (C) Western blot analysis of intracellular proteins from COS-7 cells cotransfected with equimolar amounts of 5'UTR SD(-) FLAG-tagged or untagged DNA and 5'UTR Δ-intron FLAG-tagged or untagged DNA. Blots were probed with either anti-capsid primary antibody (top) or and anti-FLAG primary antibody (bottom). Lane 1 contains the mock-transfected sample. DNA templates used for each cotransfection are designated below each respective lane (lanes 2-5).

Moreover, RNase protection assay (RPA) revealed that Gag expression levels varied according to whether or not the 5'UTR was spliced, and not simply with mRNA concentrations (Figure 14A). The protein yields (corrected for mRNA concentrations) were quantified and are reported in Figure 14B. To determine whether the Gag (FLAG-tagged or untagged) translated from mRNAs containing or missing the 5'UTR intron was incorporated into virions, aliquots of the cell culture media were centrifuged (21,000 g for 2 hours) and the viral pellets were subjected to protein analysis as with the intracellular fractions. Gag derived from all mRNA sources was found in the virion fraction, and, as above, the mRNA species lacking the 5'UTR intron was consistently the predominant source of the protein (Figure 14C). These results strongly support our hypothesis that the 5'UTRs of the two *gag* mRNA species lead to different translational efficiencies of the downstream open reading frame.

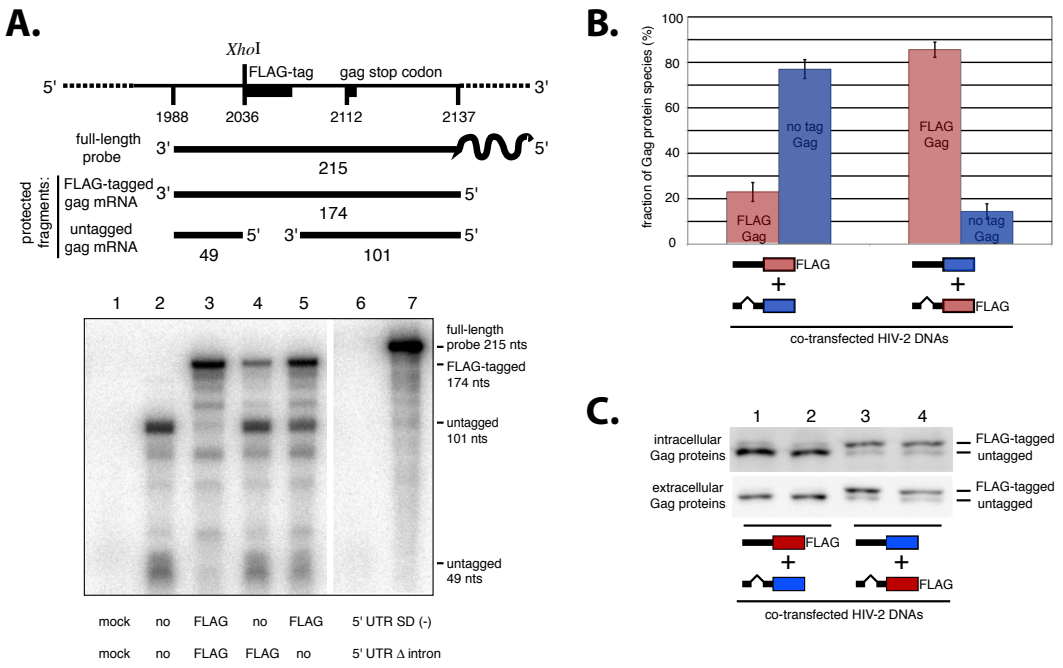

Figure 14. Effects of the 5'UTR intron on HIV-2 *gag* mRNA and protein species in cell culture. **(A) (Top) Schematic of the radioactive probe and predicted protected fragments based on the presence or absence of the FLAG tag. (Bottom) Rnase protection assay of intracellular RNAs. Lane 1 contains the mock-transfected sample. DNA templates used for each cotransfection are designated below each respective lane (lanes 2-5). Lanes 6 and 7 represent yeast RNA used as a nonspecific target RNA control, incubated with and without Rnases, respectively. (B) Quantitation of Gag protein species produced in the cytoplasm of cells cotransfected with 5'UTR SD(-) and Δ-intron HIV-2 plasmid DNA, with the FLAG-tag inserted in one or the other plasmid. The levels of the Gag protein species produced, either FLAG-tagged or untagged, were quantified from a Western blot assay (Figure 13C, lanes 4 and 5) and then corrected for the level of their corresponding mRNA (A, lanes 4 and 5). Three experiments were used to determine the fraction of Gag protein species in each cotransfection. A schematic of the cotransfected plasmid DNAs used is shown below the graph. (C) Western blots of intracellular and extracellular (virion fraction centrifuged from cell culture media) Gag protein. A schematic of the cotransfected plasmid DNAs used is shown below the blots. The blots show both the intracellular and extracellular fractions contain a majority of Gag derived from the Δ-intron construct, and that both the FLAG-tagged and untagged Gag are present without discrimination in the virus fraction.**

2.4.4 In vitro translation of luciferase mRNA harboring 5'UTR spliced or unspliced HIV-2 leaders

We tested the translation of HIV-2 leaders fused to a *Renilla* luciferase reporter gene in reticulocyte lysates. This enabled us to specifically examine the effect of the 5'UTR on translation independently of other viral functions. Briefly, two luciferase constructs were built containing either the first 569 nucleotides of the HIV-2 genomic RNA sequence (unspliced) or the first 569 nucleotides minus 61-202 nt (spliced) fused to the second codon of the *Renilla* luciferase reporter gene (Figure 15A). Following transcription, 5' capping, and polyadenylation, the luciferase RNA constructs were introduced in equimolar amounts into individual rabbit reticulocyte

lysate translation reactions. Aliquots were taken at set time points and luciferase activity at each time point was quantified using a microplate luminometer. The spliced HIV-2 5'UTR luciferase construct translated approximately five times more efficiently than the unspliced HIV-2 5'UTR luciferase construct, illustrating that the removal of the 5'UTR intron significantly increased the efficiency of HIV-2 translation (Figure 15B). All mRNAs demonstrated similar stability in the translation reactions as assayed by RPA (Figure 15C).

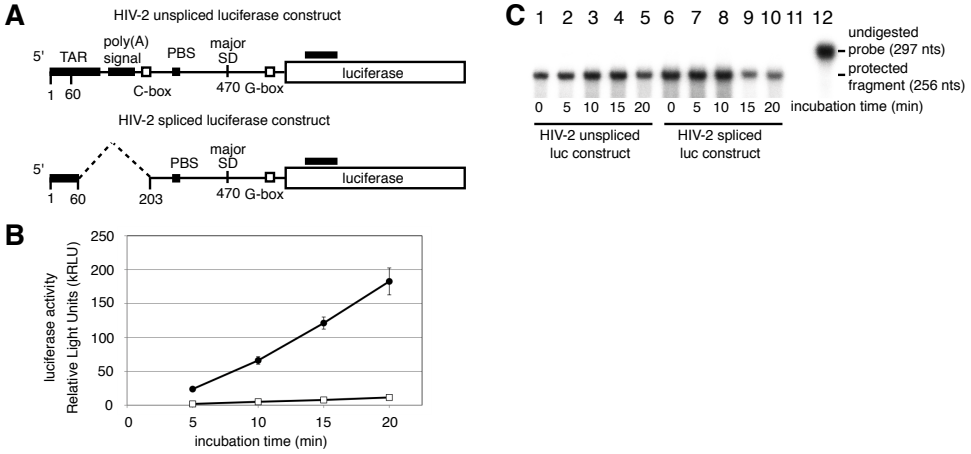

Figure 15. Effects of the 5'UTR intron on translation of HIV-2 luciferase chimeras *in vitro*.
(A) Wild type unspliced (top) or 5'UTR Δ-intron (bottom) HIV-2 sequences were fused to a *Renilla* luciferase open reading frame. The absent 5'UTR intron is represented by dotted lines. The black bar above the luciferase open reading frame represents the binding site of the antisense RPA probe used in (C). (B). Luciferase assays of translation reactions using the wild type unspliced luciferase construct (open squares) or the 5'UTR Δ-intron luciferase construct (closed circles). The y-axis represents the *Renilla* luciferase activity resulting from translation of the full-length or 5'UTR spliced constructs. The x-axis represents sample incubation times of the complete translation reactions. Aliquots of the same reactions used to quantify luciferase activity in (B) were extracted to isolate RNA, which was then subjected to RPA. Decay profiles unspliced (lanes 1-5) or spliced (lanes 6-10) were similar for these RNAs. Lanes 11 and 12 represent yeast RNA used as a nonspecific target RNA control, incubated with and without Rnases, respectively.

2.4.5 In vitro translation of luciferase mRNAs harboring serially truncated HIV-2 leaders

Several structural elements are affected by the 5'UTR intron splicing including the TAR and poly(A) signal domains and the C-box (Figure 10; compare A to B). To probe the influence of these structural elements on translation, we constructed chimeric luciferase reporter genes fused with the HIV-2 5'UTR containing serial truncations of the 5' HIV-2 leader. The constructs produced mRNAs by T7 RNA

polymerase transcription beginning at nucleotide 1, 186, 198, or 486 (HIV-2 ROD numbering). The 186 truncation lacked the entire TAR sequence as well as the poly(A) signal domain. The 198 truncation was similar to 186 truncation with just 12 additional nucleotides corresponding to the C-box deleted (55). The 486 truncation lacked TAR, the poly(A) signal domain, the C-box, and the PBS (primer binding site) domain. Translation efficiencies of these constructs were analyzed as described above.

The truncation constructs exhibited progressively higher translational efficiencies than the full-length 5' leader construct, but notably the 198 construct, which lacked the C-box, showed a disproportionately higher translational efficiency than the incrementally longer 186 construct, suggesting that the C-box sequence exerts a specific effect on translation (data not shown). To test this, we made mutations in the C-box, described below.

2.4.6 In vitro translation of luciferase mRNAs harboring a mutated HIV 5'UTR leader

We next sought to determine whether the observed increased translational efficiency caused by removal of the C-box could be recapitulated by substitution of nucleotides within this element. Two *Renilla* luciferase constructs were built containing substitutions that would test the effect of mutating the C-box sequence in the context of a full-length leader background (Figure 16A). The cytidine nucleotides in the 189 to 196 sequence of mutants A and G were substituted with adenine and guanine, respectively. Both C-box mutants exhibited higher translational efficiencies than the wild type 5' HIV-2 leader driven luciferase construct, suggesting the C-box sequence itself exerts translational regulation (Figure 16B).

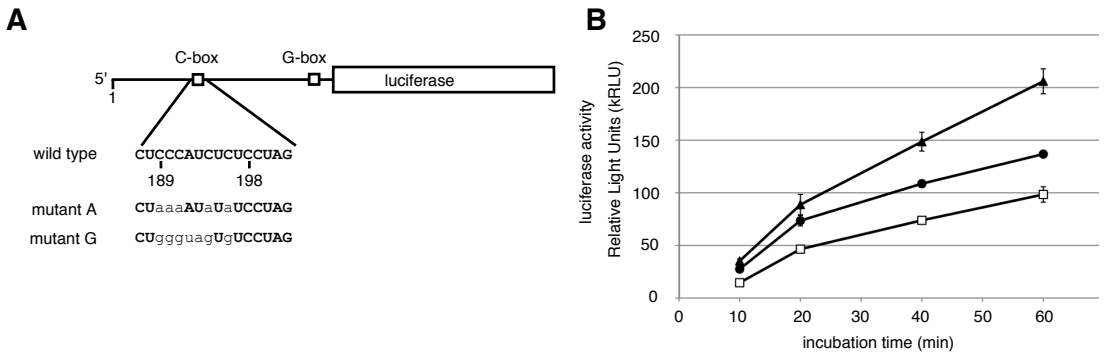

Figure 16. Effects of C-box mutations on *in vitro* HIV-2 luciferse translation.
(A) RNA constructs containing C-box mutations used in the *in vitro* translation assays. The substitutions are in lowercase letters. (B) Luciferase activity of *in vitro* translation reactions using the wild type luciferase construct (open squares) or C-box mutated constructs including mutant G (filled circles) or mutant A (filled triangles).

2.4.7 HIV-2 C-box mutants in cell culture

After identifying a role of the C-box region in translational regulation *in vitro*, we investigated whether mutations in the C-box could affect HIV-2 replication in cell culture. To obtain viral particles, COS-7 cells were transfected with full-length HIV-2 plasmid DNA containing the A and G substitutions in the C-box (Figure 17A).

We first examined 5'UTR splicing of *gag* mRNA species in both the intracellular fraction and the extracellular viral particles of transfected COS-7 cells. RT-PCR of the total intracellular RNA yielded only one distinct band for each mutant, instead of the two bands observed with wild type HIV-2 transfections (Figure 17B). Sequencing of the gel-purified bands confirmed the lack of 5'UTR splicing for each mutant (data not shown). In addition, RT-PCR analysis of the extracellular RNA confirmed the presence of the unspliced genomic RNA in viral particles (Figure 17C). These results demonstrated that there is only one species of intracellular *gag* mRNA for both C-box mutants.

Viral particles produced from the transfected COS-7 cells were used to infect permissive C8166 cells. Viral replication of both C-box mutated viruses was attenuated compared to wild type (Figure 17D), demonstrating that in addition to affecting translation, the C-box also plays an important role in viral replication.

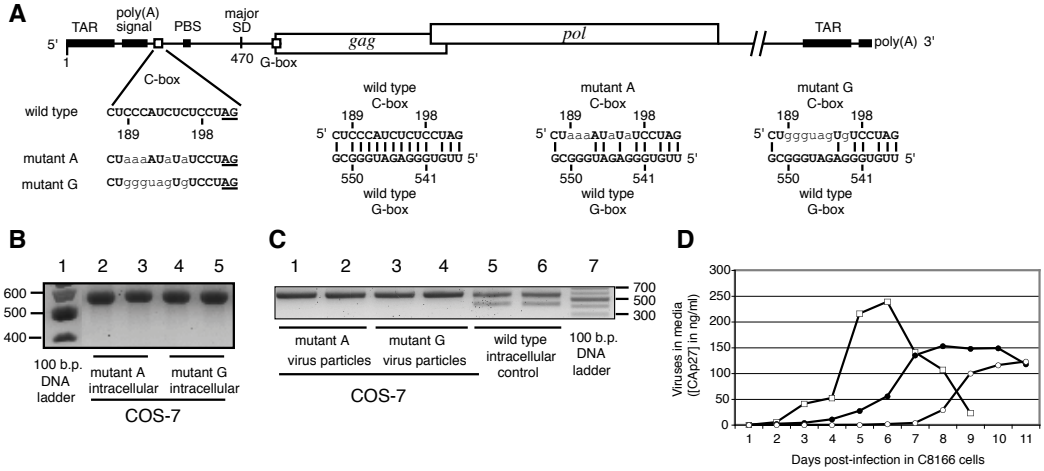

Figure 17. Effects of C-box mutations on HIV-2 replication.
(A) Gag mRNA produced following COS-7 transfection of HIV-2 DNA mutant A or G. The substitutions are in lowercase letters with the splice acceptor site underlined. Predicted C-G box wild type base pairing is shown, with mismatches introduced in the A and G mutants shown for comparison. (B) Gel electrophoresis analysis of intracellular RT-PCR products from COS-7 cells transfected with either mutant A (lanes 2-3) or mutant G (lanes 4-5) DNA plasmids. Lane 1 contains a 100-bp DNA ladder. (C) Gel electrophoresis analysis of viral particles, RT-PCR products from COS-7 cells transfected with either mutant A (lanes 1-2), mutant G (lanes 3-4), or wild type (lanes 5-6) DNA plasmids. Lane 7 contains a 100-bp DNA ladder. (D) Replication of either wild type (open squares), mutant A (filled circles), or mutant G (open circles) viruses in permissive C8166 cells. Viral production was measured by ELISA quantitation of viral p27 capsid protein.

2.5 Discussion

The impetus for studying the use of different *gag* mRNA species and their roles in HIV-2 replication came from several observations. First, the long and highly structured leader RNA in HIV-2 would not be expected to be efficiently translated by a cap-dependent scanning mechanism. Second, Lever and coworkers have proposed that the specificity of HIV-2 genomic RNA packaging depends on regulated *gag* translation in which viral RNA selection is mediated by a cotranslational interaction of the genomic RNA with its nascent Gag protein product (119). Third, the absence of a 5'UTR intron was shown to favor translation of model SIV RNAs, although the 5'UTR constructs used in this prior study were truncated ~150 nt upstream of the *gag* AUG start codon (109). Here we show that two *gag* mRNA species exist in HIV-2 in culture and in patients, differing only by an intron in the 5'UTR. The species missing this intron is a better template for *gag* translation.

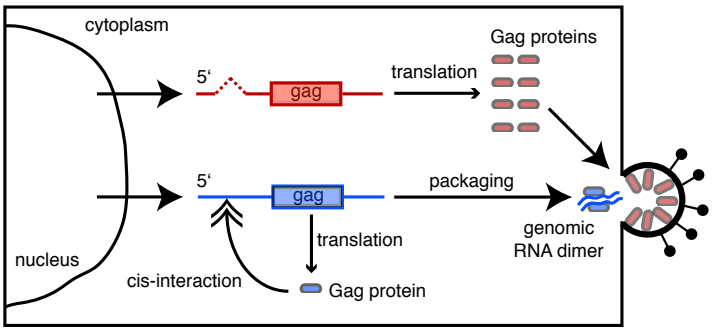

Figure 18. Model of the roles of the two forms of *gag* (or *gag-pol*) RNA species during HIV-2 replication. Both 5'UTR-spliced and –unspliced *gag* HIV-2 RNA molecules are exported from the nucleus into the cytoplasm where they serve as mRNA for Gag and Gag-Pol protein expression (small arrowheads). One functional difference between the two RNA species is that the 5'UTR spliced molecule is a better substrate for *gag* translation (this work). However, only the 5'UTR unspliced molecule is a substrate for packaging into assembling viral particles (this work). Furthermore, prior translation of the unspliced *gag* HIV-2 RNA appears to be necessary for its selection for packaging (in dimeric form) (119). A cotranslational binding event (double arrowhead) between the unspliced *gag* mRNA and the nascent Gag protein helps to select this RNA for packaging (119).

Introns entirely contained within the 5'UTR occur in about 25% of metazoan mRNAs and average between 100 and 200 nt in length (120,121). For example, removal of a 5'UTR intron has been shown to have a significant translational stimulatory effect for various cellular and viral mRNAs including platelet-derived growth factor B chain/c-*sis* and β1,4 galactosyltransferase (69,122,123). 5'UTR splicing can affect translation in several ways. Removal of secondary structures within the 5'UTR may increase translation efficiency at several stages of the translation initiation process, including scanning by the initiating ribosome and AUG recognition (65). In HIV-2, splicing of the 5'UTR intron removes half of the TAR sequence, the entire poly(A) signal, and the C-box region, thus decreasing the size of the 5'UTR spliced leader RNA to a size similar to that of the HIV-1 leader (from 545 to 405 nt, as compared to 335 nt for HIV-1) (Figure 10B). No internal 5'UTR splicing has been shown to exist in HIV-1, where the most proximal splicing event is initiated using the major SD site (107). We observed in the course of this study that translational efficiency of a reporter construct fused to an HIV-1 full-length leader was comparable to the 5'UTR Δ-intron HIV-2 leader construct (data not shown); thus, the translationally inefficient HIV-2 leader can be converted to one with higher efficiency (i.e., comparable to HIV-1) through 5'UTR splicing. Splicing may also contribute to translational efficiency through increased accessibility of the 5' methyl-

G cap in HIV-2 mRNA, since the 3' half of the TAR structure (nucleotides 61-123), which normally base pairs to the 5' end of the mRNA, is intronic. A more accessible 5' cap structure would increase translation initiation efficiency (124), and *in vitro* studies using HIV-1 5'UTR have shown that TAR has a significant inhibitory effect on translation (48,100,125,126).

It is notable that 5'UTR splicing removes a pyrimidine-rich sequence known as the C-box. The C-box was previously shown to influence HIV-2 RNA dimerization through base pairing with the G-box, an area encompassing the *gag* start codon (54,60). A similar base pairing interaction has been demonstrated in HIV-1, where it affects the equilibrium between two proposed 5'UTR conformations. It was demonstrated to influence genomic RNA dimerization both *in vitro* (61) and in cell culture (56) although evidence for 5'UTR conformational control of HIV-1 translation was not observed using a transfected reporter construct (53). In the present study, we showed that viral replication of C-box substitution mutants is impaired. Our attempts to make compensatory mutations in the G-box were confounded by the overlap of the G-box with a consensus Kozak sequence surrounding the *gag* AUG start codon as well as the obligatory myristoylation signal at the subsequent codons of *gag* (data not shown). Interestingly, the same C-box mutations increased translational efficiency of 5'UTR unspliced RNA construct *in vitro*, suggesting a role for the C-box in translational regulation of the downstream gene. In addition to their effects on translation, the C-box substitution mutants do not undergo 5'UTR splicing in transfected cells. Examination of the C-box region and surrounding sequence revealed that it encompasses the conserved polypyrimidine tract of the splice acceptor site whose disruption likely explains the absence of 5'UTR splicing in these mutants (Figure 17A)(127-129).

Surprisingly, despite the conservation of the 5'UTR splice sites among HIV-1 and SIV strains, abrogation of 5'UTR splicing by mutagenesis of the UTR splice acceptor or donor sites did not strongly affect replication of HIV-2 in immortal C8166 cells (data not shown). This suggests either that translational efficiency is not the replication rate-limiting factor in viruses grown in this cell line (which is used because it is one of the few cell lines that support robust HIV-2 replication), or that

5'UTR splicing becomes important only during a *bona fide* infection in a host. This prompted us to examine whether 5'UTR splicing was evident in clinical isolates. The observation of 5'UTR splicing in the PBMCs of HIV-2-infected individuals (Figure 12) attests to its physiological relevance, and the role(s) and regulation of 5'UTR splicing will continue to be explored. It is also possible that several mechanisms of translational activation are available to HIV-2/SIV, including the use of internal ribosome entry sites as recently reported (96,98). In the case where UTR splicing was inhibited by splice-site mutations, it is possible that increased internal ribosome entry site activity could compensate. The interplay between these two proposed mechanisms may well represent another level of regulation of translation and replication.

Unspliced genome-length HIV RNA can serve either as mRNA or a packageable genomic RNA, or both (Figure 18)(130). The present study suggests a specialized translation-only role for HIV-2 *gag* RNA following 5'UTR splicing since it augments translation, is not encapsidated to detectable levels (Figure 11), and coexpression of homologous 5'UTR spliced and unspliced *gag* mRNAs showed a strong translational bias toward the spliced species (Figure 13C). Previous research has shown that recognition of the packaging signal (Ψ) by Gag is necessary for initiation of HIV genomic RNA encapsidation and suggests that the transition between translation and packaging of the HIV genomic RNA relies on a switch-like mechanism that affects packaging signal presentation (112). Removal of secondary structure in the 5'UTR by splicing could result in an improperly presented packaging signal. In fact, the C-box, through its interaction with the G-box, has been shown to influence the presentation of the encapsidation and dimerization signals *in vitro* (58,60). Taken together, our results suggest a role for the 5'UTR splicing and the C-box in the concomitant regulation of *gag* translation and packaging during the HIV-2 replication cycle.

2.6 Acknowledgements

We thank Tayyba Baig, Leila Sears, and Mary Ellenbecker for critical reading of the manuscript. This work was funded by the National Institute of Allergy and Infectious

Diseases grant AI45388 to J.S.L., AI6274-04 to P.J.K, and a Grant-In-Aid of Research from Sigma Xi, the Scientific Research Society, to C.L.S. The plasmid pROD10 was provided by the EU Programme EVA/MRC Centralised Facility for AIDS Reagents, NIBSC, UK (Grants QLK2-CT-1999-00609 and GP828102). The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: C8166-145 (Cat# 404) from Dr. Robert Gallo and saquinavir (Cat# 4658).

CHAPTER THREE: VIRAL SELEX REVEALS INDIVIDUAL AND COOPERATIVE ROLES OF THE C- AND G-BOXES IN HIV-2 REPLICATION

Christy L. Strong, Jean-Marc Lanchy and J. Stephen Lodmell

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

Corresponding author: J. Stephen Lodmell

This chapter is a modified version of the manuscript submitted to the RNA Journal.

This manuscript represents the experiments done by the author of this dissertation with the exception of the forced evolution, 2 nt and 5 nt mutation cell culture studies of HIV-2, which were performed by Dr. Jean-Marc Lanchy (Figures 20 and 21, Table 6).

3.1 Abstract

The 5'UTR of HIV-2 genomic RNA contains signaling motifs that regulate specific steps of the replication cycle. Two motifs of interest are the C- and G-boxes. The C-box is found in the 5' untranslated region upstream of the primer binding site, while the G-box is found downstream of the major splice donor site, encompassing the *gag* start codon and flanking nucleotides. Together the C- and G-boxes form a long-range base pairing interaction called the CGI. We and others have previously shown that formation of the CGI affects RNA dimerization *in vitro* and the positions of the C- and G-boxes are suggestive of potential roles of the CGI in other steps of HIV-2 replication. Therefore, we attempted to elucidate the role of the CGI using a viral SELEX approach. We constructed proviral DNA libraries containing randomized regions of the C-box or G-box paired with wild type or mutant base pairing partners. These proviral DNA libraries were transfected into COS-7 cells to produce viral libraries that were then used to infect permissive C8166 cells. The "winner" viruses were sequenced and further characterized. Our results demonstrate that there is strong selective pressure favoring viruses that can form a branched CGI. In addition, we show that the mutation of the C-box alone can enhance RNA encapsidation and mutation of the G-box can alter the levels of Gag protein isoforms. These results suggest coordinated regulation of RNA translation, dimerization, and encapsidation during HIV-2 replication.

3.2 Introduction

HIV replication is highly orchestrated, requiring tight regulation of each step of the viral cycle to ensure successful virus synthesis and propagation. The 5' untranslated region (5'UTR) of the HIV RNA genome contains several regulatory motifs including TAR (transactivation responsive element), the polyadenylation signal, the C-box, the primer binding site (PBS), stem loop 1 (SL1), and the G-box (Figure 19) (46,47). The C-box, a pyrimidine-rich region, has been implicated in the regulation of genomic RNA dimerization and translation through a base pairing interaction with the G-box, a purine-rich region encompassing the *gag* start codon and flanking nucleotides (Figure 19) (47,131).

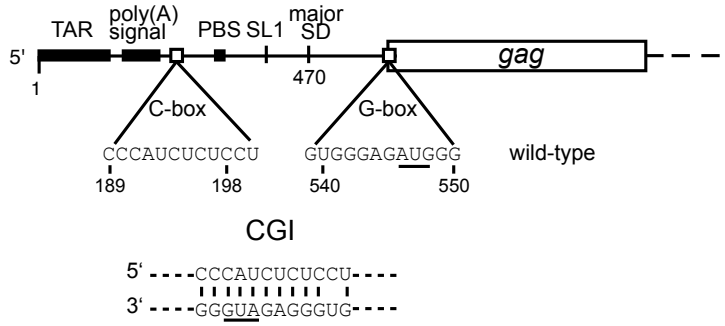

Figure 19. Schematic of HIV-2 genomic RNA.
TAR, poly(A) signal, C-box, PBS, SL1, major SD, G-box, and *gag* represent the trans-activation region, the poly(A) signal domain, the core of the C-box, the tRNA primer binding site, the stem loop one dimerization signal, the major splice donor site, the core of the G-box, and the Gag protein coding region, respectively. A proposed C-box-G-box interaction (CGI) is illustrated below. The AUG initiation codon of Gag, found in the G-box, is underlined.

Several previous studies suggested that both the presence of the C- and G-boxes and their ability to base pair affected dimerization in HIV-2 (47,55,58,60,132) and in HIV-1 (53,56,61,62). It was postulated that the C-box-G-box interaction termed CGI in HIV-2 (54) or U5-AUG duplex in HIV-1 (61) could modulate translation through occlusion of the *gag* start codon encompassed by the G-box (47,60-62). Mutational studies designed to test two proposed structural conformations of the HIV-1 leader, namely the branched multiple-hairpin (BMH) and

the long distance interaction (LDI) conformations, did not appear to support a role for the U5-AUG duplex, which would be disrupted in a large-scale LDI-BMH conversion, in translational control (53). However, the main focus of that study was the BMH-LDI interconversion and did not directly probe the U5-AUG duplex. In HIV-2, directed mutagenesis of the leader suggested that the C-box does play a role in translational regulation, as deletion or mutation of the C-box to disrupt the CGI led to enhanced translation of HIV-2 chimeric reporter constructs (131).

HIV-2 and SIV_{mac} are unique among the primate lentiviruses in that they have a conserved intron contained entirely within the 5' untranslated leader region (52,109,111,131). The C-box is located at the 3'end of the intron and is completely removed upon 5'UTR splicing. Previously, we have shown that removal of this intron results in increased translation of *gag* mRNA, apparently through the combined effects of overall shortening of the leader and removal of secondary structure that could impede ribosomal scanning (131). Detailed *in vitro* translation studies demonstrated that mutations or deletions of the C-box that were designed to disrupt the proposed CGI resulted in increased translational efficiency. However, attempts to directly test the CGI by making compensatory mutations in the G-box were confounded by the exquisite sensitivity of the G-box to changes, apparently because the G-box encompasses the Kozak context for the *gag* start codon (131). Thus the importance and the role(s) of the CGI have been difficult to address experimentally.

Here we used a viral SELEX technique (132,133) to generate viable virus variants at the C-box or the G-box. When either of these sequences was randomized alone, wild type-like sequences rapidly emerged as the dominant species. This led us to question whether this was the result of local sequence constraints or the requirement to maintain base pairing compatibility with another viral sequence. To address this, we built two additional proviral DNA libraries, one that was simultaneously randomized at both the C- and the G-box, and one that contained a lethal mutated G-box sequence combined with a randomized C-box. Each of these libraries was designed to test, with the use of non-native sequence in the C- and/or G-box, whether base pairing between these sites was important for virus viability. These co-randomization and forced evolution experiments reproducibly demonstrated

that the requirement for base pairing in the CGI is finely tuned, with lack of base pairing being as deleterious as overly stable extended base pairing arrangements between these elements. Sequence analysis of winning viruses suggested that the CGI is directly involved in a branched three-way RNA junction, consistent with a previously proposed model (47).

3.3 Material and Methods

3.3.1 Construction of plasmids for generation of randomized C-box, G-box, co-randomized, and forced evolution libraries

To prevent contamination by wild type proviral DNA plasmid in the C-box, G-box, co-randomized, and forced evolution libraries, we used parent plasmids called pSCR2, pAUG1, pCboxg, and pGboxc, respectively. The pSCR2, pAUG1, pCboxg, and pGboxc plasmids harbor deleterious mutations in the encapsidation signal (132), the *gag* initiation codon, the C-box region (194-GGUGG-199), and the G-box region (540-CCACC-545), respectively.

To create a vector for generating the C-box, co-randomized, and forced evolution libraries, a derivative of pSCR2 plasmid called pSCR2AfeIΔ(173-2030) was constructed, in which a fragment encompassing nt 173 to 2030 of the viral genome, including part of the noncoding region, and most of the Gag coding region up to the *Xho*I site was deleted and substituted with eight nucleotides to introduce a unique *Afe*I restriction site and a spacer region. All nucleotide numbering in the present study is referenced to the RNA sequence of HIV-2 (ROD isolate, GenBank no. M15390). The cloning region in the pSCR2AfeIΔ(173-2030) vector is as follows: 5'-$_{160}$GCCAGTTAGAAGCgctaagtc$_{180}$/ /$_{2031}$CTCGAG-3', where AfeI and XhoI sites are underlined and the lowercase letters indicate the changed nucleotides. pSCR2AfeIΔ(173-2030) was constructed as follows. A fragment containing the long terminal repeat (LTR) sequence up to nt 172, followed by a 14-nt sequence containing the remaining half of the *Afe*I site, the spacer region, and the XhoI site, was amplified using as sense primer (M13 forward (-41) binding upstream of a unique AatII site) and an antisense primer (asXhoCbox, containing the XhoI site, spacer region, and AfeI site). The amplified product was digested with AatII and

XhoI, and ligated into pSCR2 plasmid vector missing the original AatII-XhoI fragment. The ligated product was transformed in *Escherichia coli* DH5α cells, and ampicillan-resistant colonies were selected, followed by plasmid purification and sequencing.

To create a vector generating the G-box randomized library, a derivative of pAUG1 plasmid called pAUG1AfeΔ was constructed, in which a fragment encompassing nt 498 to 2030 was deleted and substituted with eight nucleotides to introduce a unique AfeI restriction site and a spacer region. The cloning region of the pAUG1AfeΔ(498-2030) vector is as follows: 5'-$_{481}$ACACCAAAAACTGTAGCgctaagtc$_{505}$//$_{2031}$CTCGAG-3', where *Afe*I and *Xho*I sites are underlined and the lowercase letters indicate the changed nucleotides. pAUG1AfeΔ(498-2030) was constructed as follows. A fragment containing the LTR sequence up to nt 498, followed by a 14-nt sequence containing the remaining half of the *Afe*I site, the spacer region, and the *Xho*I site, was amplified using a sense primer (M13 forward (-41) binding upstream of a unique *Aat*II site) and an antisense primer (asXhoGbox, containing the XhoI site, spacer region, and AfeI site). The amplified product was digested with *Aat*II and *Xho*I, and ligated into the pAUG1 plasmid vector missing the original *Aat*II-*Xho*I fragment. The ligated product was transformed in *Escherichia coli* DH5α cells, and ampicillin-resistant colonies were selected, followed by plasmid purification and sequencing.

3.3.2 Generation of the C-box, G-box, co-randomized, and forced evolution proviral DNA libraries

To generate the proviral DNA libraries, we used a single-step ligation dependent DNA cloning protocol (In-fusion; Clontech) as described previously (132). Both the pSCR2AfeΔ and pAUG1AfeΔ vectors were prepared by an *Afe*I and *Xho*I digestion, followed by gel electrophoresis and gel extraction (5PRIME).

To construct the C-box randomized insert, a PCR product (nt 157 to 2050) was generated using a mutagenic primer (sCboxrandom) with degenerate nucleotides at positions 194 to 199, an antisense primer (asROD2050), and the pCboxg plasmid as a template. To construct the G-box randomized insert, a PCR product (nt 483 to

2050) was generated using a mutagenic primer (sGboxrandom) with degenerate nucleotides at positions 540 to 545, an antisense primer (asROD2050), and the pSCR2 plasmid as a template. To contruct the co-randomized insert, overlap extension two PCR products was performed, followed by a PCR using a sense primer (sROD157) and an antisense primer (asROD2050). The first PCR product (157-571) was generated using a mutagenic sense primer (sCboxrandom) with degenerate nucleotides at positions 194 to 199, a mutagenic antisense primer (asGboxrandom) with degenerate nucleotides at positions 540 to 545, and the pCboxg plasmid as a template. The second PCR product (551-2050) was generated using a sense primer (sROD551), an antisense primer (asROD2050), and the pCboxg plasmid as a template. To construct the forced evolution randomized insert, a PCR product (nt 157-2050) was generated using a mutagenic sense primer (sCboxrandom) with degenerate nucleotides at positions 194 to 199, an antisense primer (asROD2050), and the pGboxc plasmid as a template. All PCR products (randomized inserts) were purified by gel electrophoresis.

The C-box randomized, co-randomized, or forced evolution randomized insert was combined with the digested pSCR2AfeIΔ(173-2030) vector at a 2:1 ratio in an infusion reaction. The G-box randomized insert was cloned into the digested pAUG1AfeIΔ(498-2030) vector at a 2:1 ratio using the Infusion Clontech kit. Each of the resulting recombined products were transformed into *E. coli* DH5α cells and DNA was extracted from bulk transformation reactions to obtain each of the randomized proviral DNA libraries. Degeneracy of the proviral DNA libraries was checked by sequencing.

3.3.3 Construction of individual plasmids with C-box and G-box mutations

To construct full-length plasmids with mutated and non-mutated C-box and G-box regions, we used the In-Fusion strategy described previously with the pSCR2AfeIΔ(173-2030) vector but with inserts that contained the desired C-box and G-box sequences. Each insert was produced using overlap extension with two PCR products as previously described. All overlap extensions contained a PCR product (551-2050) generated using a sense primer (sROD551), an antisense primer

(asROD2050), and the pCboxg plasmid as a template. The Cbox119 pcr product was generated with a sense primer (sCboxCLS119), an antisense primer (asGboxwt), and the pCboxg plasmid as a template. The Gbox119 PCR product was generated with a sense primer (sCboxwt), an antisense primer (asGboxCLS119), and the pGboxc plasmid as a template. The CLS-119 PCR product was generated with a sense primer (sCboxCLS119), an antisense primer (asGboxCLS119), and the pGboxc plasmid as a template. The CLS-120, CLS121, and CLS-005 proviral DNAs were constructed as described for CLS-119, using their designated primers.

3.3.4 Cell culture and transfection

COS-7 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, glutamine, penicillin, and streptomycin (Invitrogen). COS-7 cells were transiently transfected using the Trans-IT-COS transfection kit (Mirus). Cells and media were harvested 24-48 hours post-transfection. The amount of HIV-2 particles in the media was determined by measurement of the viral capsid (Cap27) protein levels (Retro-Tek SIV p27 Antigen ELISA kit; Zeptometrix).

3.3.5 Cell culture and infection

C8166 cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% fetal calf serum, glutamine, penicillin, and streptomycin (Invitrogen). C8166 cells (2×10^4 cells) were infected with media from COS-7 transfected cells (10 ng of HIV-2 capsid as determined by ELISA). The C8166 cells were washed twice 12 hours post-infection, followed by resuspension in RPMI/fetal bovine serum. A media aliquot (day 0) was taken to serve as a reference. A daily aliquot of media was harvested, spun to remove cells, and then the supernatant frozen at -80˚C. Viral replication was followed by quantification of the Cap27 protein (Retro-Tek SIV p27 Antigen ELISA kit; Zeptometrix) or reverse transcriptase (Amersham Biosciences Quan-T-RT assay system) levels in the supernatant sample.

3.3.6 Protein isolation and analysis

Intracellular COS-7 RNA and proteins were harvested 48 hours after transfection by washing and scraping cells in phosphate-buffered saline. To prevent Gag cleavages, 10 μM (final concentration) of saquinavir (an HIV protease inhibitor) was added to the media of the transfected COS-7 cells 24 hours before harvesting. One half of the cells were harvested in RNA lysis buffer for genomic RNA quantitation (see below). The remaining half of the cells was lysed in radioimmunoprecipitation assay buffer (Santa Cruz Biotechnology Inc.; Tris-HCl, 50 mM, pH 7.4, NaCl 150 mM, SDS 0.1%, Nonidet P-40 0.1%, 0.5% sodium deoxycholate with protease inhibitors). Following lysis on ice, samples were fractionated on SDS-8% polyacrylamide gels. Electrotransfer onto polyvinylidene difluoride (PVDF) membranes was then performed. To visualize Gag proteins, the western blot was probed using the primary antibody biotin-labeled anti-capsid antibody (p27 ELISA kit from Zeptometrix) followed by incubation with streptavidin-horseradish peroxidase (p27 ELISA kit from Zeptometrix). Detection was performed by electrochemiluminescence (ECL Plus Western Blotting Detection Reagents from GE Healthcare) using a Fujifilm LAS-3000.

Supernatants from transfected cells were filtrated through 0.45 μm filters and virions were collected by a 1-hour centrifugation at 21,000 g/4˚C. Pelleted virions were resuspended in radioimmunoprecipitation assay buffer and Gag proteins were visualized as described above.

3.3.7 RNA isolation and analysis

Total cellular RNA and extracellular viral RNA fractions from transfected COS-7 cultures were purified with the Absolutely RNA mini-prep kit (Stratagene). Virions were pelleted by centrifuging an aliquot of media for 1-hour at 21,000 g/4˚C. Purified RNAs from the cells and viruses were used in Rnase protection assays (RPA III kit, Ambion). The Rnase protection assay was performed using an antisense RNA probe complementary to the 401-562 region of HIV-2 ROD isolate. The antisense region was cloned into the pGEM7Zf(+) vector (Novagen) allowing for 41 nts of vector origin at the 5' end of the T7 transcript. This non-HIV-2 tail was used to

confirm Rnase's digestion efficiency during the RPA experiment. Three protected bands are expected: one 162 nt band corresponding to the 401-562 region (wild type G-box), one 139 nt band corresponding to the 401-539 region (mutant G-box), and one 17 nt band corresponding to the 546-562 region (mutant G-box).

3.3.8 Luciferase template construction for in vitro transcription

Luciferase chimeric constructs were built containing the first 920 nucleotides of HIV-2 fused to a *Renilla* luciferase reporter gene. The HIV-2 inserts (1-920 nt) were generated using PCR. A sense primer (sNHErod) containing a *Nhe*I site and an antisense primer (asBsoBI920) containing a *Bso*BI site were used to amplify the first 1-920 nucleotides of the C-box119, G-box119, and CLS-119 constructs as well as wild type HIV-2 genomic RNA, ROD isolate (GenBank no. M15390; genomic sequence start at 1). All constructs were checked by DNA sequencing.

3.3.9 In vitro transcription

The HIV-2-luciferase plasmids were linearized via *Cla*I digestion. The linearized DNA templates were then used in transcription reactions to synthesize RNA (mScript mRNA Production System, Epicentre). Following completion of the incubation, the reactions were treated with RNase-free DNase, followed by ammonium acetate precipitation. The RNAs were then 5' capped and polyadenylated (mScript mRNA Production System, Epicentre), followed by ammonium acetate precipitation.

3.3.10 In vitro translation and protein analysis

Comparison translation reactions were performed using equimolar amounts (46 nM) of the tested RNAs. The RNAs were initially denatured in water (final volume of 25.2 µL) at 65˚C, then snap-cooled on ice for 3 minutes. The translation mix (70 µL rabbit reticulocyte lysate, 1 µL amino acid mixture minus methionine, 1 µL [^{35}S] methionine (Perkin Elmer), and 2.8 µL KCl 2.5M; Flexi Rabbit Reticulocyte Lysate System, Promega) was added to each mRNA sample to initiate the reaction. The translation reactions were then incubated at 30˚C. Five minutes into the 45-minute incubation, each reaction was spiked with 1 mM non-radioactive methionine. The samples were again placed at 30˚C. Upon completion of the incubation, reactions

were placed on ice. A 10 µL aliquot of each reaction was combined with 20 µL of 4M urea protein loading dye. The samples were denatured at 90˚C for 2 minutes and run on a SDS-polyacrylamide gel. The radioactive protein isoforms were visualized using a Fuji FLA-3000.

3.3.11 Prediction of RNA secondary structures

Mfold version 3 (134) was used to predict the secondary structures for the wild type HIV-2 ROD and the 34 individual clonal RNA sequences isolated from the co-randomized viral library at day 40 and 56. The analyzed RNA fragments represent the first 580 nucleotides of the genomic RNA sequences. The software used is found on the mfold server (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) (134). The ten most stable secondary structures for each RNA were visually analyzed and the base pairing partners of the C-box and G-box, presence or absence of the CGI, and the type of CGI were recorded.

3.4 Results

3.4.1 Mutations in the C-box and G-box are deleterious to HIV-2 in cell culture

We first addressed the role of the CGI in HIV-2 replication by investigating the effects of mutations and presumed compensatory mutations of the C and G-boxes. Mutant viruses containing five altered nucleotides in either the C-box (positions 194-198) or the G-box (positions 541-545) were replication-defective in cell culture (Figure 20). The compensatory mutant, containing simultaneously mutated C- and G-boxes, failed to rescue replication (Figure 20).

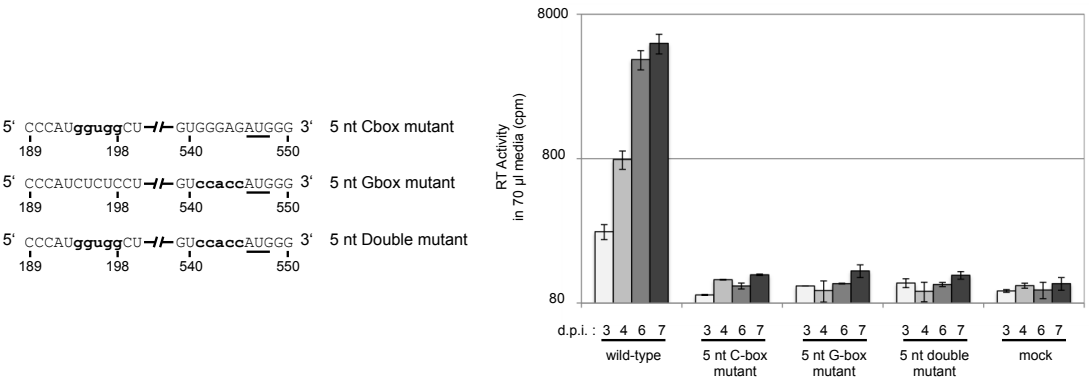

Figure 20. Five nucleotide compensatory mutation analysis of the CGI.
(Left panel) Schematic of the 5 nt mutation series. Only the C-box and G-box regions of the unspliced genomic RNAs are represented. Nucleotide positions that are mutated are in bold lowercase letters. (Right panel) Replication kinetics of the wild type, the 5 nt C-box mutant, the 5 nt G-box mutant, and the 5 nt double mutant viruses in C8166 cells. Equal amounts of viral particles (10 ng of p27 capsid, as determined by ELISA), produced during COS-7 transfection were used to infect C8155 cells. Viral replication was measured using reverse transcriptase quantification. Numbers below the x-axis represent days post-infection. The y-axis error bars represent the standard error of duplicate experiments.

Because five-nucleotides of the C- or G-box were lethal, we attempted more subtle two-nucleotide mutations in the C-box (positions 197-198), G-box (positions 541-542), and compensatory mutations in both (Figure 21). The 2 nt C-box mutant reached close to wild type replication levels (Figure 21). The 2 nt G-box mutant showed reduced levels of replication. Interestingly, combining the C-box and G-box mutations resulted in viruses that were replication defective (Figure 21).

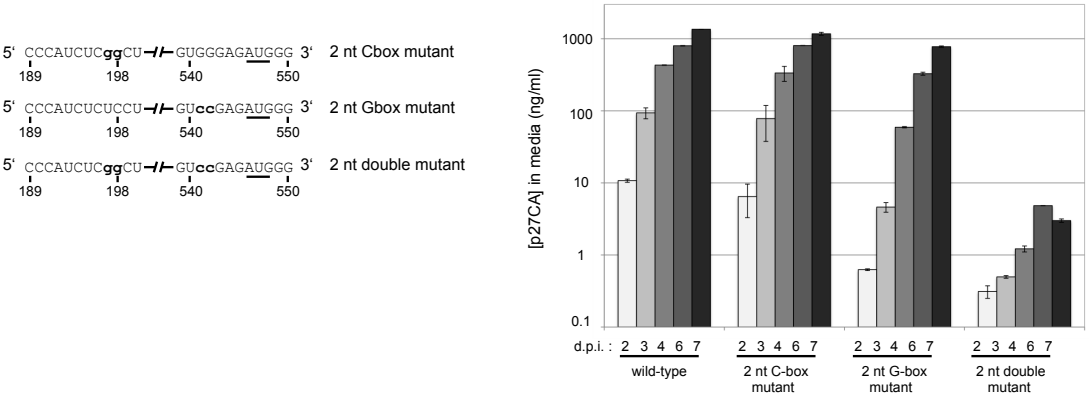

Figure 21. Two nucleotide compensatory mutation analysis of the CGI.
(Left panel) Schematic of the 2 nt mutation series. Only the C-box and G-box regions of the unspliced genomic RNAs are represented. Nucleotide positions that are mutated are in bold lowercase letters. (Right panel) Replication kinetics of wild type, the 2 nt C-box mutant, the 2 nt G-box mutant, and the 2 nt double mutant viruses in C8166 cells. Equal amounts of viral particles (10 ng of p27 capsid, as determined by ELISA), produced during COS-7 transfection were used to infect C8166 cells. Viral replication was measured using p27 quantification. Numbers below the x-axis represent days post-infection. The y-axis bars represent the standard error of duplicate experiments.

Thus our first attempts to create a non-wild type CGI using compensatory mutations resulted in replication-defective viruses, hampering analysis of sequence requirements using cell culture studies. To circumvent the limitations of building and testing substitution mutants one at a time, we adopted a viral SELEX approach.

3.4.2 Viral SELEX of the C-box and G-box randomized libraries

Six nucleotides of the C-box (nucleotides 194-199) or six nucleotides of the G-box (nucleotides 540-545) were randomized using degenerate primers to make PCR libraries containing up to 4096 possible sequence combinations of either the C-box or the G-box region of the genome. Each PCR library was used to construct a library of full-length infectious HIV-2 proviral DNA clones using a ligation-independent protocol (see Material and Methods). The degeneracy of each randomized proviral DNA library was verified through pool sequencing. The C-box and G-box proviral DNA libraries were independently transfected into COS-7 cells to generate virus. COS-7 cytoplasmic and media fractions for both libraries were harvested. Viral RNAs purified from the media fraction were examined for degeneracy of the randomized region (Figure 22).

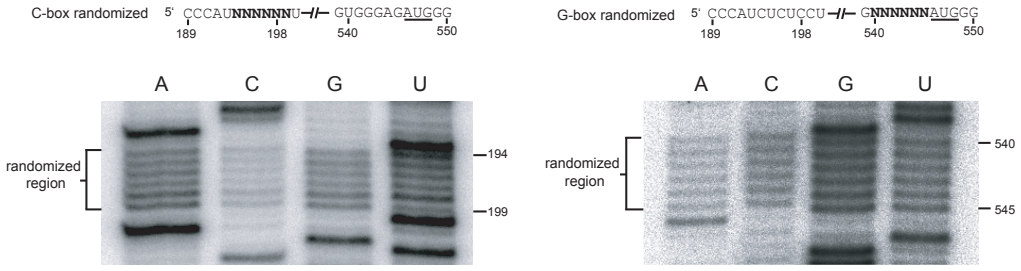

Figure 22. C-box and G-box randomized libraries.
(Top panel) Schematic of the C-box and G-box elements in the two randomized libraries. Only the C-box and G-box regions of the unspliced genomic RNAs are represented. Nucleotide positions containing an "N" have an equal probability of being A, C, G, or U. (Bottom panel) Viral RNAs purified from the media fraction of transfected COS-7 cells were reverse transcribed into cDNA. Primer extension of the different cDNA showed degeneracy of the C-box randomized region in the C-box randomized library (Bottom; left) and the G-box randomized region in the G-box randomized viral library (Bottom; right).

Virus harvested from the media fraction was used to infect permissive C8166 cells. Emergence of dominant virus sequences was quite rapid, with the two winning sequences for the G-box library (5' 540-UGGGRG, where R is G or wt A)

dominating by day 13 (compare Table 2 to Table 3) and the wild type C-box sequence (5'194-CUCUCC) dominating the C-box library by day 20 (Table 4).

Table 2. Sequences obtained from the G-box COS-7 viral library.
Individual clone sequencing of the six-nt randomized G-box virus library contained within the media fraction of the transfected COS-7 culture. This virus library was used to infect permissive C8166 cells.

Category and clone number	Nucleotide at position: 540	541	542	543	544	545
Extracellular vRNA (day 0)						
wild type	U	G	G	G	A	G
CLS-1	U	A	G	A	G	U
CLS-2	G	A	C	A	C	G
CLS-3	G	C	A	G	A	A
CLS-4	C	G	G	U	G	U
CLS-5	A	U	C	A	U	G
CLS-6	U	A	C	A	A	C
CLS-7	A	A	A	A	G	C
CLS-8	C	G	G	A	A	G
CLS-9	A	A	G	A	A	G
CLS-10	U	G	G	A	G	G
CLS-11	G	G	U	A	G	A
CLS-12	C	A	G	A	G	A
CLS-13	C	U	G	C	A	A
CLS-14	G	G	G	C	G	U
CLS-15	G	A	C	U	G	U

Table 3. Pool sequencing of the G-box region of viruses from the six-nt randomized G-box library that were selected for after several rounds of infection in C8166 cells.

Time post infection	Nucleotide at position: 540	541	542	543	544	545
Day 5	G/U	G	G	G	A/G	G
Day 9	G/U	G	G	G	A/G	G
Day 13	U	G	G	G	A/G	G
Day 19	U	G	G	G	A/G	G
Day 27	U	G	G	G	A/G	G

Table 4. Pool sequencing of the C-box region of viruses from the six-nt randomized C-box virus library that were selected for after several rounds of infection in C8166 cells.

Time post infection	Nucleotide at position: 194	195	196	197	198	199
Day 6	C/U	C/U	N	N	N	N
Day 12	C	U	C	U	N	N
Day 14	C	U	C	U	C	N
Day 20	C	U	C	U	C	C
Day 26	C	U	C	U	C	C

The more rapid emergence of dominant sequences suggested there was stronger selective pressure on the G-box than the C-box. By day 5, four out of the six positions in the randomized G-box were wild type, with the two remaining positions 540 and 544 being U or G and A or G, respectively (Table 3). The C-box, on the other hand, contained pyrimidines at positions 194 and 195 by day 6, with the remaining four positions (196-199) still relatively diverse (Table 4). However, only the wild type C-box sequence was found by day 20.

The rapid emergence of wild type sequence seen in both the C-box and G-box randomized libraries made us ask whether this was because of local sequence requirements, mutual base pairing requirements (i.e. CGI), or both. To address this question we built a new library (Figure 23) with both the C-box (positions 194-199) and G-box (positions 540-545) randomized on the same proviral DNA.

Figure 23. Co-randomized viral library.
Only the C-box and G-box regions of the unspliced genomic RNAs are represented. Nucleotide positions containing an "N" have an equal probability of being A, C, G, or U.

3.4.3 Viral SELEX of the co-randomized library

The co-randomized viral library (Figure 23) was constructed and tested using the same techniques as described for the C-box and G-box libraries. Mfold (134) prediction of secondary structure of all winner viral sequences revealed several key characteristics. First, non-wild type viruses represented a large proportion of selected viruses on day 40 (12 out of 17) and day 56 (9 out of 17) (Table 5).

Table 5. C-box and G-box sequences of viruses that were selected for in the co-randomized viral library following several rounds of infection in C8166 cells. The CGI found in the lowest energy structure is either (a) the branched CGI or (b) the extended CGI.

Time post infection and clone number	Nucleotide at position:												CGI type	
	194	195	196	197	198	199	**540**	**541**	**542**	**543**	**544**	**545**	a	b
40 dpi														
wild type	C	U	C	U	C	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-88	U	G	A	G	A	A	**C**	**U**	**G**	**G**	**G**	**G**	X	
CLS-89:93	C	U	C	U	C	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-94	C	C	A	A	C	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-95	G	G	A	U	U	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-96	C	C	A	A	C	C	**U**	**G**	**G**	**U**	**A**	**A**	X	
CLS-97	U	G	C	C	G	G	**G**	**U**	**U**	**G**	**C**	**A**	X	
CLS-98	U	G	C	C	G	G	**U**	**G**	**G**	**G**	**A**	**G**		X
CLS-99	C	C	A	A	C	C	**U**	**G**	**G**	**U**	**U**	**G**	X	
CLS-100:104	G	U	C	A	G	C	**U**	**G**	**G**	**G**	**A**	**U**	X	
56 dpi														
Wild type	C	U	C	U	C	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-105/106	G	U	C	A	G	C	**U**	**G**	**G**	**G**	**A**	**U**	X	
CLS-107	G	U	C	A	G	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-108/109	C	U	C	U	C	C	**U**	**G**	**G**	**G**	**A**	**U**	X	
CLS-110:117	C	U	C	U	C	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-118	C	U	C	U	C	C	**C**	**U**	**G**	**G**	**G**	**G**	X	
CLS-119	G	G	A	U	C	C	**U**	**G**	**G**	**U**	**U**	**C**	X	
CLS-120	C	C	A	A	C	C	**U**	**G**	**G**	**G**	**A**	**G**	X	
CLS-121	C	G	G	C	A	C	**U**	**G**	**G**	**G**	**A**	**G**	X	

Second, all of the selected viruses were predicted by Mfold to form two nearly isoenergetic conformations of CGI, an extended form and a branched form (Figure 24). The extended CGI consisted of a continuous base pairing through the entirety of the C-box and G-boxes (Figure 24C). The branched CGI consisted of a three way RNA junction with base pairing of the 5' proximal C-box with the 3' proximal G-box, while the 5' proximal G-box was base paired with the nucleotide 385 region, similar to a structure previously proposed (47) (Figure 24A & B). The branched CGI was found in lowest energy structures for all but one of the viral sequences (Table 5).

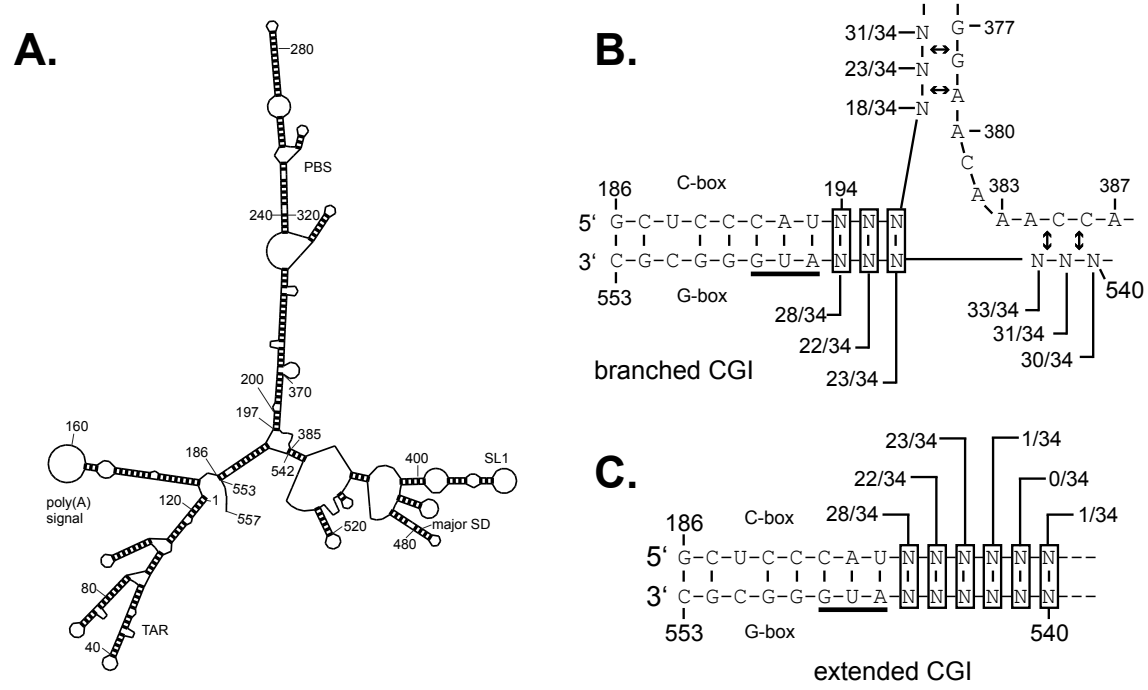

Figure 24. Summary of the base pairing interactions involving the C-box and G-box found in the predicted optimal RNA structures from the co-randomized viral selection.
(A) The most stable secondary structure predicted by Mfold for nucleotides 1-557 of the wild type HIV-2 genomic RNA. (B) and (C) When analyzed similarly, the 34 viral sequences from the evolved co-randomized library (Table 5) grouped into two CGI conformations, the branched CGI (B) and the extended CGI (C). Each potential base pairing interaction of the randomized nucleotides is shown. Each ratio indicates the number of times this base pairing is found in the most stable CGI containing structure for each of the 34 viral sequences. The double arrows indicate a level of flexibility with regard to the base pairing partner for the randomized nucleotide. Depending on the identity of the randomized nucleotide, the base pairing partner in the designated region may shift by one or two positions.

Selection for wild type sequence at the 5' side of the proximal G-box (positions 540-542) was readily apparent, represented by 15 out of 17 viruses on day 40 and 16 out of 17 viruses by day 56 (Table 5).

Although a significant percentage of viruses with wild type sequence emerged from our co-randomized library, it also produced CGI-forming viruses with non-wild type sequences (12 out of 17 sequences by day 40). To determine whether these non-wild type viruses were viable, three late round viruses (CLS-119, CLS-120, and CLS-121) and one early round virus (CLS-005) were selected for individual testing to compare their viral replication to that of wild type (Figure 25). To show that the C-box and G-box mutations are solely responsible for the observed viral phenotypes, we cloned these mutations into the original wild type HIV-2 backbone. All of the late round viruses were predicted to form the branched CGI while the CLS-005 was not. As shown in Figure 25B, the early round CLS-005 displayed defective viral replication while late round CLS-119, CLS-120, and CLS-121 were replication competent when compared to wild type. These results showed first that the CGI is important for viral replication and second that it can be studied outside of a wild type context.

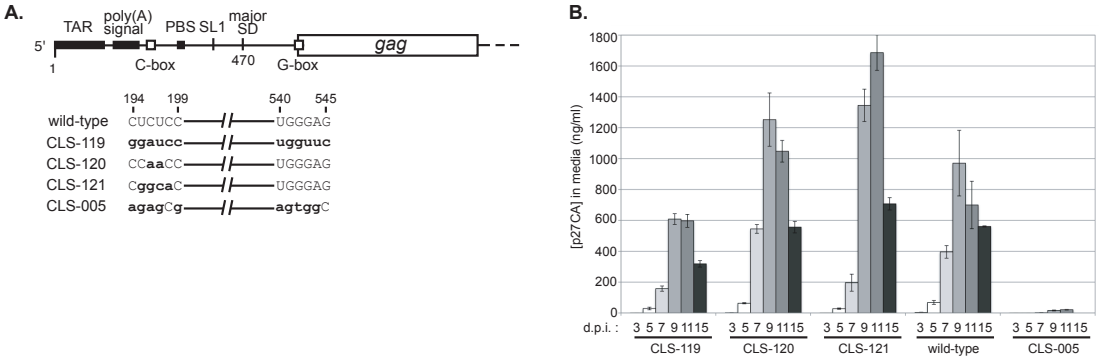

Figure 25. Replication kinetics of mutant HIV-2 viruses.
(A) Schematic of the five HIV-2 genomic RNAs produced in this study. The CLS-119 RNA contains a mutated C-box and mutated G-box. The CLS-120 RNA contains a mutated C-box and a wild type G-box. The CLS-121 RNA contains a mutated C-box and a wild type G-box. The CLS-005 RNA contains a mutated C-box and a mutated G-box. The mutations are in bold lowercase letters. (B) Replication kinetics of wild type, CLS-119, CLS-120, CLS-121, and CLS-005 viruses in C8166 cells. Equal amounts of viral particles (10 ng of p27 capsid, as determined by ELISA), produced during COS-7 transfection were used to infect C8166 cells. Viral replication was measured using p27 capsid quantification. The y-axis bars indicate standard error for two experiments.

3.4.4 Forced evolution of C-box randomized viruses with a defective G-box

To further evaluate the relationship of C-box and G-box sequences, we built a new randomized viral library that forced the viruses to compensate for a defective mutant G-box sequence with novel sequence combinations at the C-box. Six nucleotides of

the C-box (nucleotides 194-199) were randomized using degenerate primers in an HIV-2 background containing a mutated G-box that is lethal by itself (Figure 26 and Figure 20).

Figure 26. Forced evolution library.
Schematic of the C-box and G-box element in the forced evolution library. Only the C-box and G-box regions of the unspliced genomic RNAs are represented. Nucleotide positions containing an "N" have an equal probability of being A, C, G, or U. The G-box mutation is in lowercase letters. The Gag start codon, found in the G-box, is underlined.

The proviral DNA library was transfected into COS-7 cells and the virions produced were used to infect C8166 cells. Emergence of viable virus occurred after three weeks. The mutated G-box sequence (5'541-CCACC) was conserved in all sequenced clones (data not shown). All selected viruses contained guanosines at positions 194 and 195 of the C-box (Table 6). There was more variability seen in position 196 although the wild type cytosine was completely absent in all selected viruses. Wild type sequence (UCC) was found at positions 197-199 for most of the selected viruses, which would not be expected to base-pair with the mutated G-box nucleotides 540-542, located upstream of the *gag* AUG start codon.

Table 6. C-box sequences of viruses that were selected for in the forced evolution virus library following several rounds of infection in C8166 cells.

Time post infection and clone number	Nucleotide at position:					
	194	195	196	197	198	199
24 d.p.i						
Wild type	C	U	C	U	C	C
CLS-143	G	G	A	C	G	A
CLS-132	G	G	A	U	G	A
CLS-130	G	G	G	U	C	C
CLS-131	G	G	G	U	C	C
CLS-140	G	G	G	U	C	C
CLS-142	G	G	U	A	A	A
CLS-133	G	G	U	A	C	C
CLS-151	G	G	U	U	C	C
CSL-153	G	G	U	U	C	U
CLS-161	G	G	U	U	C	C
CLS-162	G	G	U	U	C	C
CLS-163	G	G	U	U	C	C

This experiment showed that mutations in the C-box rescued a lethal mutant G-box mutation. Moreover, these results corroborated results from the co-randomized library, demonstrating selective pressure for base pairing between the 5' proximal C-box and 3' proximal G-box.

3.4.5 Effects of artificially evolved C-box and G-box sequences on viral replication

Although the co-randomized and forced evolution library results supported a role in viral replication for the C-box and G-box through their base pairing interaction, our five and two nt mutational studies suggested that the sequence of each region may also have a CGI-independent role in viral replication. To probe the individual roles of the C-box and G-box, a viral sequence (CLS-119) from the co-randomized library was selected for testing. We chose CLS-119 for three reasons. First, CLS-119 was a selected (viable) virus, predicted to form the branched CGI (Figure 24A & B). Second, the CLS-119 C- and G-boxes were non-native, composed equally of both wild type and non-wild type nucleotides (Table 5). The CLS-119 C-box contained both purines (positions 194-196, GGA) and pyrimidines (positions 197-199, UCC) as did the G-box with mostly purines (positions 540-542, UGG) and pyrimidines (positions 543-545, UCC). Third, although represented in the winning pool, CLS-119 displayed delayed viral replication compared to the other selected viruses as well as wild type (Figure 25). We therefore hypothesized that this delay may be a result of the individual contributions of the C-box and the G-box.

To determine the individual effects of the non-native CLS-119 C-box and G-box sequences, we built two additional clones and compared viral replication to that of wild type (Figure 27).

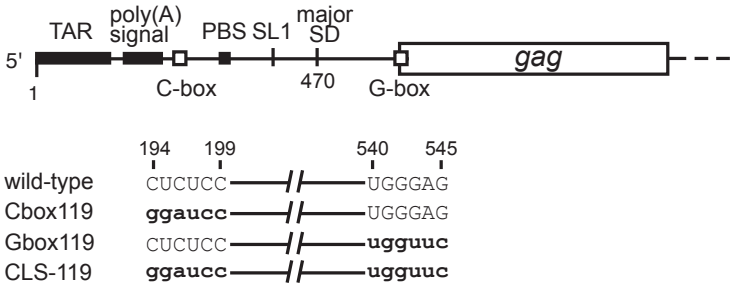

Figure 27. Mutant HIV-2 viruses.
Schematic of the four HIV-2 genomic RNAs produced in this study. The C-box119 RNA contains a mutated C-box and a wild type G-box. The G-box119 RNA contains a wild type C-box and a mutated G-box. The CLS-119 RNA contains a mutated C-box and a mutated G-box. The mutations are in bold lowercase letters.

All of the clones were predicted to form the branched CGI, which allowed us to concentrate on individual effects of the C-box and G-box regions instead of effects of the CGI. The CLS-119 clone contained both the mutated C- and G-boxes. The C-box119 and G-box119 clones contained either the CLS-119 C-box or the CLS-119 G-box, respectively. COS-7 cells were transfected with wild type, Cbox119, G-box119, and CLS-119 full-length HIV-2 plasmids. Virions from the COS-7 supernatant were harvested and used to infect permissive C8166 cells. Viral replication of the C-box119 mutant, as determined by ELISA, reached wild type like levels, whereas the G-box119 and CLS-119 mutants exhibited attenuated replication (Figure 28).

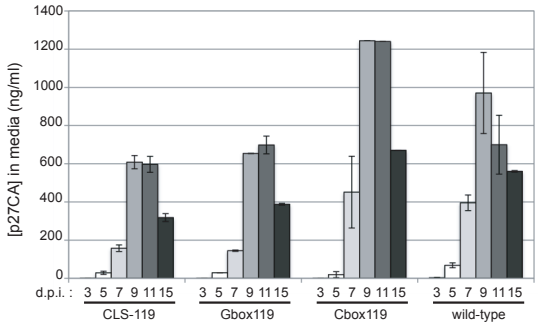

Figure 28. Mutant HIV-2 viruses harboring G-box mutations display attenuated replication.
Replication kinetics of wild type, C-box119, G-box119, and CLS-119 viruses in C8166 cells. Equal amounts of viral particles (10 ng of p27 capsid, as determined by ELISA), produced during COS-7 transfection, were used to infect C8166 cells. Viral replication was measured using p27 quantification. The y-axis error bars indicate standard error for two experiments.

The shared feature of the G-box119 and CLS-119 mutants is the presence of the mutated G-box. To determine whether the decrease in replication was a result of a packaging deficiency linked to the presence of the G-box, intracellular and virion RNA levels were examined.

We measured intracellular viral RNA levels and virion RNA content using an RNase protection assay (RPA) (Figure 29). Intracellular *gag* RNA levels were very similar, while RNA levels found in the virion fractions differed (Figure 29). Surprisingly, the C-box119 virus had higher levels of genomic RNA encapsidated (normalized to extracellular capsid protein p27) than the other viruses. This suggested that in addition to affecting dimerization, translation, and 5'UTR splicing, the C-box also plays a role in encapsidation.

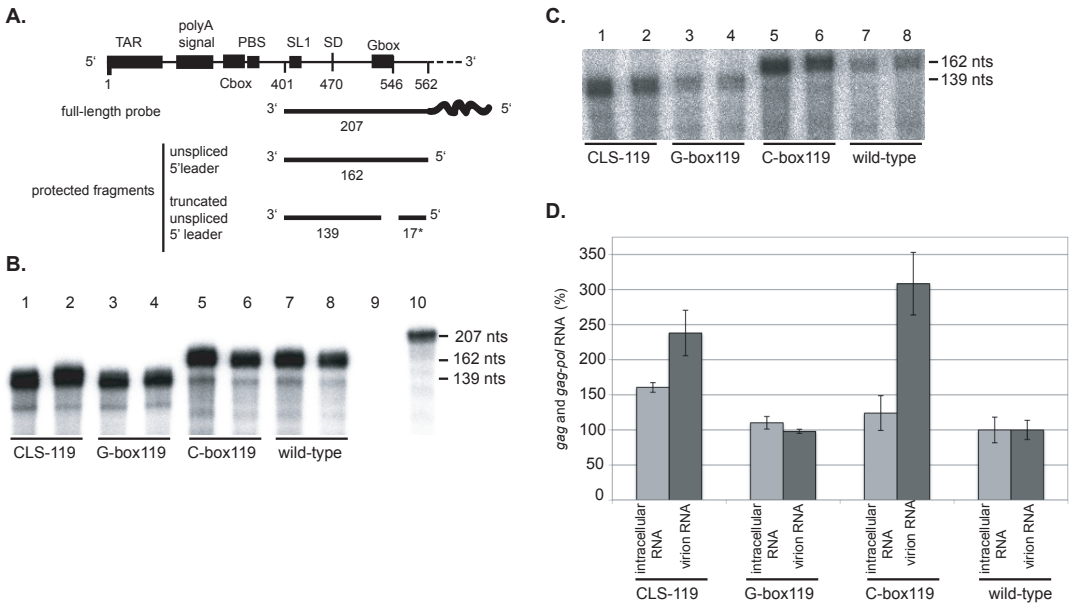

Figure 29. C-box mutation results in increased RNA encapsidation.
(A) A schematic of the radioactive probe and predicted protected fragments based on the presence of a wild type or mutant G-box. (B) RNase protection assay of intracellular RNAs obtained from duplicate experiments. DNA templates used for each transfection are listed below each lane (lanes 1-8). Lanes 9 and 10 contain yeast RNA used as a non-specific target control, incubated with or without RNases, respectively. (C) RNase protection assay of virion RNAs produced from duplicate HIV-2 DNA transfected COS-7 cell experiments. DNA templates used for each transfection are listed below each land (lanes 1-8). (D) Quantification of (B) and (C). RNA levels were normalized to wild type levels, set at 100% for both intracellular and virion samples. Identities of RNA source, intracellular or virion, and DNA templates used for each transfection are listed on the x-axis. The y-axis represents *gag* and *gag-pol* RNA levels in percent. The y-axis error bars indicate standard error for the two experiments.

However, the RPA results could not explain why the presence of the G-box resulted in decreased replication. Because the mutated G-box encompasses the Kozak sequence of the *gag* initiation site, we hypothesized that perhaps this mutation altered the translation of *gag* mRNA. Therefore, we then sought to determine if the attenuated viral replication seen with the G-box119 and CLS-119 viruses was the result of decreased translational efficiency. To that end, we examined both intracellular and virion fractions of transfected COS-7 cells for total Gag levels (Figure 30). The intracellular and virion protein fractions were examined via western blotting, using a primary antibody against p27 capsid. Both the CLS-119 and G-box119 viruses showed a marked decrease in total Gag levels both inside and outside the cell compared to C-box119 and wild type viruses (Figure 30C and D). In addition to differences in total Gag levels, there were also differences in the relative expression of three isoforms of Gag between viruses with a mutant G-box and those with a wild type G-box. Ohlmann and coworkers have shown that these three isoforms of Gag are expressed during HIV-2 infection, which they attributed to alternative initiation events (96,99). The first isoform, p58, initiates synthesis form a start codon located at nucleotides 546-548 (AUG1). The second isoform, p51, initiates from a start codon located at positions 744-746 (AUG2). The third isoform, p45, initiates synthesis from a start codon located at positions 897-899 (AUG3). Interestingly, clones containing the mutant G-box, i.e., G-box119 and CLS-119, exhibited an increased level of the p45 Gag isoform initiated from AUG3 compared to wild type and to the C-box119 mutant (Figure 30C; compare lanes 2-5 with lanes 6-9).

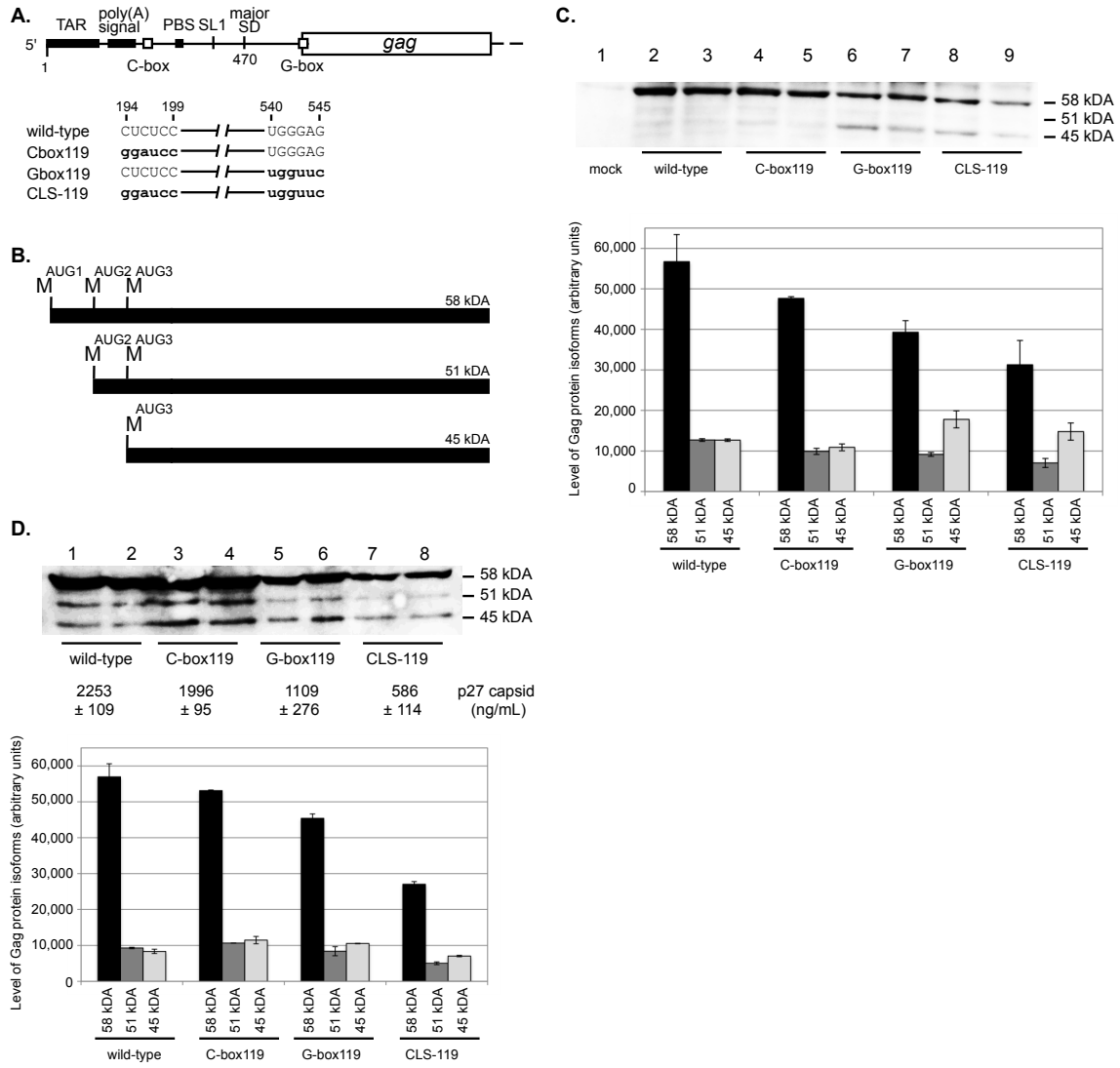

Figure 30. Effects of a mutated G-box on Gag production in cell culture.
(A) A schematic of the four HIV-2 genomic RNAs produced in this study. The mutations are in bold lowercase letters. (B) A schematic of the three HIV-2 Gag protein isoforms that are synthesized in cell culture. The 58, 51, and 45 kDa isoforms are synthesized when translation initiates from methionine codons at positions 546, 744, and 897, respectively. (C) Western blot analysis of intracellular proteins from duplicate experiments where COS-7 cells were transfected with equimolar amounts of the wild type, C-box119, G-box119, or CLS-119 proviral DNAs. Lane 1 contains the mock-transfected sample. DNA templates used in each transfection are designated below each lane. The relative proportion of each isoform band was quantified and is graphically represented. The y-axis error bars represent the standard error of duplicate experiments. (D) Western blot analysis of virion proteins produced from duplicate experiments where COS-7 cells were transfected with equimolar amounts of the wild type, C-box119, G-box119, or CLS-119 proviral DNAs. DNA templates used in each transfection are designated below each lane. The averaged level of capsid p27 protein in the COS-7 media was quantified (ELISA) and is indicated below the western blot. The relative proportion of each isoform band was quantified and is graphically represented. The y-axis error bars represent the standard error of the duplicate experiments.

3.4.6 In vitro translation of luciferase RNAs containing mutated C-box, G-box, or both

To test whether the different levels of Gag isoforms seen in clones containing the mutant G-box were caused by altered start codon usage, we assayed the translation of the HIV-2 C-box and G-box mutant constructs in an *in vitro* rabbit reticulocyte translation assay. As shown in Figure 31, constructs containing a mutated G-box demonstrated differences in *gag* start codon usage compared to constructs with a wild type G-box. Both CLS-119 and G-box119 *gag*-luciferase constructs demonstrated a slight increase in usage of the AUG3 start codon, similar to what was seen in cell culture. These data suggest that the different levels of the isoforms are caused by altered start codon usage, and that this difference is directly influenced by the mutation in the G-box.

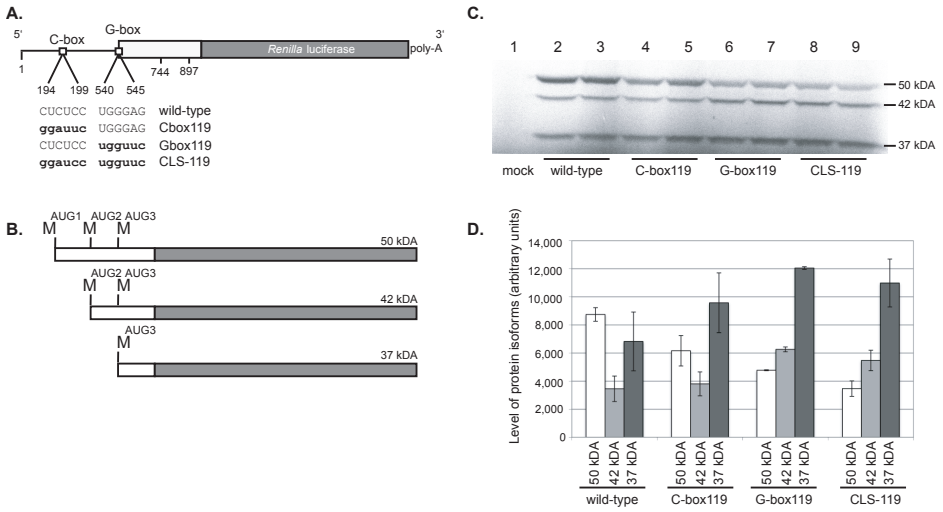

Figure 31. *In vitro* translation of luciferase mRNAs harboring mutated HIV-2 5' leaders. (A) A schematic of the four mRNA constructs used in S^{35} *in vitro* translation assays. The mutated nucleotides are in bold lowercase letters. (B) A schematic of the three HIV-2 Gag-luciferase fusion protein isoforms that are synthesized *in vitro*. The 50, 42, and 37 kDa isoforms are synthesized when translation initiates from methionine codons located at positions 546, 744, and 897, respectively. (C) Duplicate S^{35} *in vitro* translation assays of the wild type, C-box119, G-box119, and CLS-119 mRNAs. Lane 1 contains a mock where no mRNA was added to the translation reaction. RNAs used in each translation reaction are indicated below each lane. (D) Quantification of protein isoform expression levels. The RNAs used are indicated on the x-axis, along with the protein isoforms produced and quantified. The level of protein isoforms in arbitrary units is indicated on the y-axis. The y-axis error bars represent the standard error of the duplicate experiments.

3.5 Discussion

In this study, we employed viral SELEX to determine the role of the CGI in viral replication. Generating a collection of unique viruses that differed only in their C-box and G-box regions allowed us to circumvent the experimental limitations of building and testing one unique mutant at a time. Testing these viruses in a competitive cell culture setting allowed us to examine in detail viable viruses with CGI sequence combinations not found in nature. Our results using this viral evolution technique emphasize the importance of the CGI for viral replication and fine-tune our understanding of base pairing arrangements in this region. Furthermore, we show that in addition to a CGI-dependent role, the C-box and G-box have CGI-independent roles in viral replication.

We demonstrate here that mutation of the C-box affects HIV-2 encapsidation levels (Figure 29), suggesting a role for the C-box in the packaging process. Mutation of the C-box may contribute to encapsidation through structural rearrangements of the 5'UTR, exposing previously hidden motifs while trapping others (135). Indeed, in HIV-2, destabilization of the CGI was shown to free RNA elements essential for dimerization (SL1) and encapsidation (Ψ) from structural entrapment (54). Another possibility is that the enhanced encapsidation phenotype may be due to enhanced binding of nucleocapsid protein (NC) to the mutated C-box. It is known that NC mediates genomic RNA encapsidation by binding RNA elements found in the 5' leader region (136,137). Studies of HIV-1 leader RNA have suggested a role for the C-box as a binding partner of NC, wherein the binding of NC disrupts the CGI (138,139). Additionally, *in vitro* HIV-1 mutant RNAs harboring guanosines in the C-box displayed enhanced NC binding affinity compared to wild type (138). Here, the mutated CLS-119 C-box differs from the wild type C-box at the 5' proximal positions 194-196, with the sequence GGA instead of CUC. It is possible that the presence of two non-wild type guanosines in the mutant C-box may influence NC binding and encapsidation in HIV-2. It is conceivable that wild type C-box functions as a good but not ideal encapsidation enhancer, similar to what we have shown for the G-box and its modulation of translation. However, we cannot rule out

an indirect effect of the CLS-119 C-box mutation on the presentation of RNA signals such as encapsidation elements.

Our studies on the G-box demonstrated the extreme sensitivity of this region to mutation. Besides its CGI-dependent context, the G-box sequence is also constrained by the basic requirement for a start codon and an adjacent glycine codon to serve as the myristoylation signal (140). In addition, G-box nucleotides are involved in translation initiation as part of the Kozak sequence. It is notable that according to the Kozak consensus rules, the context of the wild type start codon AUG1 is good but not ideal (78). Interestingly, when we mutated the G-box sequence closer to an ideal consensus sequence, resulting mutant viruses were not viable. This suggested that enhancement of AUG1 translational efficiency is not necessarily beneficial in the overall scheme of HIV-2 viral replication. Conversely, we show that viruses displaying decreased AUG1 translational efficiency compared to wild type, are viable (Figure 28) but display altered ratios of the three isoforms of Gag. The decrease in p58 Gag production is expected due to the mutation of the AUG1 Kozak consensus sequence, however the slight increase in p45 Gag production (AUG3 initiation) is notable. The p45 Gag isoform was previously shown to be necessary for efficient viral replication, but its role remains obscure (96).

Previous studies on HIV-2 translation have suggested that the genomic RNA contains an IRES located downstream of the *gag* AUG1 initiation codon (96,99). The HIV-2 IRES recruits ribosomes for translational initiation at AUG1 and downstream counterparts AUG2 and AUG3, generating the three Gag isoforms. IRES initiation at each of the three AUGs occurs independently of one another (98). Interestingly, we observed that mutation of nucleotides immediately upstream of AUG1 resulted in altered Gag isoform ratios. Although it is possible that the three nucleotide G-box mutation affected the IRES-dependent *gag* initiations, a more likely scenario is leaky scanning at the AUG1. It has been recently shown that initiation at the HIV-1 AUG1 *gag* initiation site primarily occurs through ribosomal scanning (92). Since our HIV-2 G-box mutation results in a weaker Kozak context for AUG1, leaky scanning at this initiation site enhances the use of a least one of the downstream start codons. Indeed, leaky scanning is a common mechanism used in HIV-1 translation. Several

laboratories have shown that the strength of upstream AUGs can influence initiation at downstream AUGs (141-143). Mutation of a weak upstream HIV-1 *rev* Kozak sequence to a strong initiating sequence resulted in decreased leaky scanning and poor initiation at the downstream *vpu* AUG (142). Alternately, here the CLS-119 AUG1 mutation is a relatively poor Kozak context compared to the wild type AUG1, resulting in increased initiation at the downstream AUG3. Of the three potential *gag* initiation codons in CLS-119, AUG3 is predicted to have the best Kozak context, thus a scanning ribosome could scan past AUG1 and 2 and initiate translation at AUG3 more often than when the Kozak context of AUG1 is wild type (78,81). Taken together, we propose that the strength of the Kozak context sequence for the HIV-2 *gag* AUG1 start codon is evolutionarily fine-tuned to allow balanced synthesis of Gag isoforms and that, in combination with the C-box, there could be a higher order structure component to the use of the AUG1 start codon as well.

The main focus for this study was to determine the necessity of the CGI in HIV-2 replication. Our viral SELEX work supports a role for the association of the C-box with the G-box with the caveat that both local and long distance constraints govern the CGI. Local sequence constraints limit nucleotide identity in both the C-box and G-box. Sequence identities in these regions appear to affect 5'UTR splicing, Gag or NC binding, and the Kozak consensus sequence. However, our co-randomization and forced evolution experiments demonstrated reproducibly that partial base pairing across this region is necessary, as seen in the branched CGI conformation (Figure 24A & B). Indeed, the forced evolution experiment clearly illustrates that a relationship between the C-box and G-box exists in HIV-2 replication. A lethal viral replication phenotype associated with the presence of the mutant G-box (CCACC) was rescued by the presence of the selected C-boxes. The selected C-boxes in this library all contain guanosines at positions 194 and 195, which are compatible for base pairing with the mutant G-box cytosines at positions 544 and 545. Interestingly, extension of the CGI through additional base pairing did not emerge significantly among non-wild type surviving viruses from both the forced evolution and co-randomized libraries. Instead, it seems that selective pressure may be exerted at these remaining positions (197-199) to keep the CGI from becoming too

stable and to allow the 5' proximal G-box to form a long distance base pairing interaction with the nucleotide 385 region. In all sequences forming the branched CGI, nucleotides 200-204 (just downstream of the randomized C-box) were base-paired with nucleotides 379-376, irrespective of the identity of randomized nucleotides 197-199. It is notable that the base pairing interaction of the 5' proximal G-box with the nucleotide 385 region could vary through shifting of one or two nucleotides depending on the identity of the randomized G-box nucleotides.

Our two-and five-nucleotide mutations experiments support the pressure for CGI flexibility, as we saw that high stability base pairing across the entire mutated CGI results in a lethal phenotype. The replication-defective two-nucleotide double mutant is predicted to be unable to form two of the three branches of the branched CGI. Similarly, the five-nucleotide double mutant is also predicted to be unable to form two of the three branches of the branched CGI. Only when the five-nucleotide G-box was paired with a randomized C-box that did not extend base pairing through the CGI, were viable viruses obtained. Furthermore, analysis of published HIV-2/SIV sequences (144) shows several examples of nucleotide variations that maintain base pairing that is consistent with our "artificial phylogeny" and supports a branched three-way RNA junction (Figure 32).

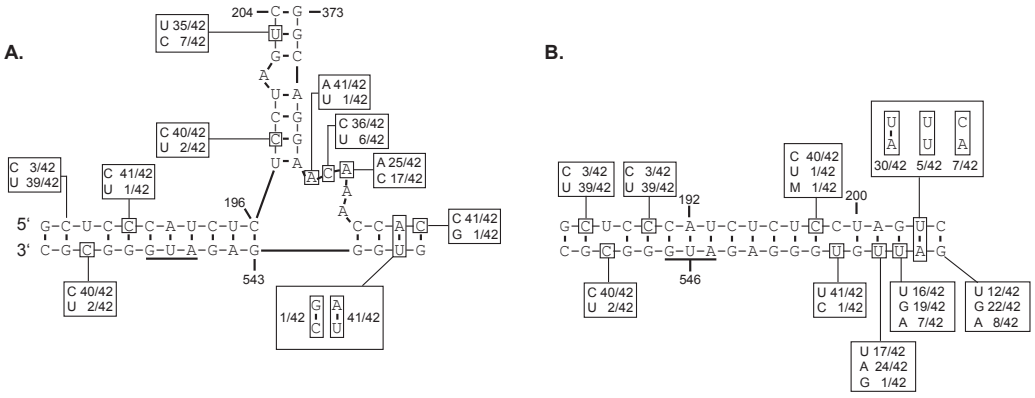

Figure 32. Nucleotide identities of non-conserved position among natural isolates of HIV-2 and SIV superimposed on the two potential CGI conformations.
The consensus branched (A) and extended (B) CGI structures from the analysis of 42 HIV-2 and SIV sequences from the HIV Sequence Compendium 2010 is indicated, with the proportions of alternative base pairs shown (144). Those nucleotides not absolutely conserved are boxed.

In particular, a key element in the branching CGI structure is the base pairing of some of the G-box nucleotides with the 385 region. These four base pairs are highly conserved, which is further illustrated by a conservative compensatory mutation in the isolate $SIV_{smm}SL92$.

Our data emphasize the importance of the CGI *in vivo* and suggest a model in which some degree of CGI association is necessary, as seen in the branched model (Figure 24A & B) (47), but because of roles in multiple steps of the HIV-2 replication cycle, the CGI needs to be dynamic. The dynamic nature of base pairing in the 5' leader has previously been supported in the context of the LDI/BMH riboswitch affecting the BMH and LDI conformational equilibrium in HIV-1 RNA (53). Although it has been shown that the base pairing between the C-box and G-box in mature HIV-1 particles is maintained in the open conformation (139), its behavior in HIV-2 still needs to be investigated. Mapping the dynamic range of the CGI during select replication steps would provide further insight into the CGI's mode of action.

3.6 Acknowledgements

We thank Michalee Moen for critical reading of the manuscript. This work was funded by the National Institute of Health grant AI45388 to J.S.L. and a Grant-In-Aid of Research from Sigma Xi, the Scientific Research Society to C.L.S. The plasmid pROD10 was provided by the EU Programme EVA/MRC Centralised Facility for AIDS Reagents, NIBSC, UK (Grant no. QLK2-CT-1999-00609 and GP828102). The following reagent was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: C8166-145 (Cat# 404) from Dr. Robert Gallo.

CHAPTER FOUR: DISCUSSION

4.1 Summary of work

The 5'UTR intron and the CGI are two elements within the 5'UTR that regulate multiple steps in HIV-2 replication. When the 5'UTR intron is removed or the base pairing partners in the CGI are mutated, HIV-2 *gag/gag-pol* translation, encapsidation, and Gag isoform ratios are altered.

We have shown that 5'UTR splicing of *gag* and *gag-pol* RNA species occurs *in vitro* (*i.e.*, in cell culture) and *in vivo* (from samples obtained from asymptomatic HIV-2 infected patients) (131). The removal of secondary structure by 5'UTR splicing results in specialized isoforms of *gag*/*gag-pol* RNAs that display increased translational efficiency compared to their unspliced counterparts (131). In addition to removing local secondary structures such as the poly (A) signal domain, 5'UTR splicing also disrupts a long-range base-pairing interaction, the CGI.

The C-box is involved in a long-range base-pairing interaction with the G-box, termed the CGI. The CGI is disrupted by 5'UTR splicing because the C-box is positioned within the 5'UTR intron and encompasses the poly-pyrimidine tract of the 5'UTR splice acceptor site. Disruption of the CGI was predicted to enhance translational efficiency by eliminating potentially inhibitory base pairing of the *gag* start codon. *In vitro* translation studies demonstrated that disruption of the CGI by deletion or mutation of the C-box did result in increased translational efficiency (131). Additional studies to determine the base pairing and nucleotide identity requirements of the CGI were performed using viral SELEX.

A branched CGI was selected for in the HIV-2 randomized viral SELEX experiments (Figure 33). In addition to identifying this novel secondary structure in HIV-2, additional roles for the C-box and G-box were uncovered. Our CLS-119 studies illustrated that mutating the C-box can result in enhanced genomic RNA encapsidation while mutating the G-box can lead to decreased viral replication and altered Gag isoform ratios. In essence, we have demonstrated a role for the CGI in encapsidation and in translation.

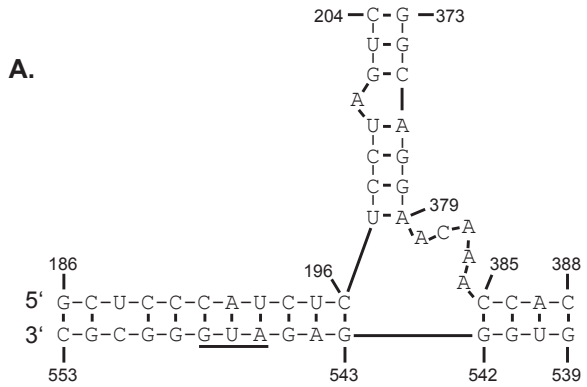

Figure 33. The branched HIV-2 CGI.
The structure of the HIV-2 CGI is supported by observed patterns of base pairing arrangements that dominated after randomization of this region and selection for viable viruses. The *gag* initiation codon is underlined.

4.2 Retroviral RNA translation and encapsidation

Multiple RNA species are generated during the HIV replication cycle. In HIV-1, the 47 unique RNA species produced are assigned to three classes based on size (107). The RNA classes include 1.8 kilobases, 4.0 kilobases, and 9.0 kilobases (Figure 34). All HIV-1 RNA classes except for one contain multiple members. The 9.0 kb class contains only one RNA species, the unspliced *gag-pol* RNA. The unspliced *gag-pol* RNA (i.e. the genomic RNA) is unique among the HIV-1 RNA species because it contains the complete encapsidation signal. All other HIV-1 RNAs do not contain the encapsidation signal because this element is located downstream of the major splice donor site and is thus removed during splicing of the 1.8 and 4.0 kb classes. Therefore, only the full-length genomic RNA is encapsidated based on the presence of the primary encapsidation signal. This is not the case in HIV-2.

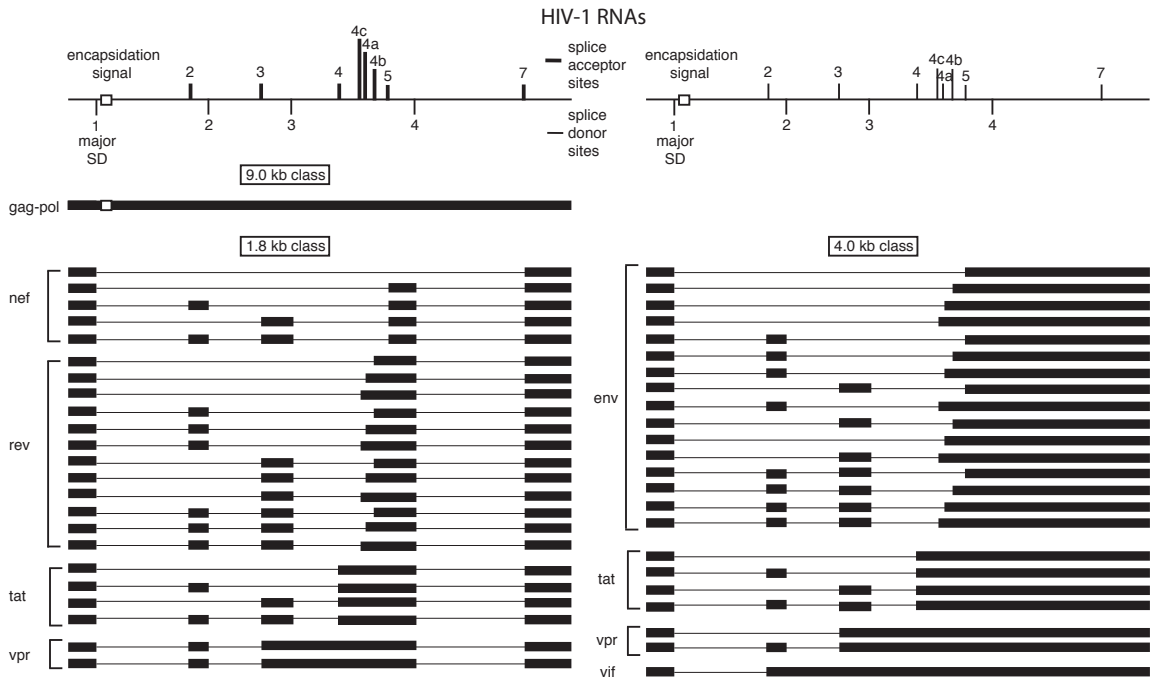

Figure 34. HIV-1 RNA species.
A representation of the viral RNA species generated during HIV-1 replication. The viral RNAs are grouped into size classes including 9.0 kb, 1.8 kb, and 4.0 kb. The splice acceptor sites used are indicated above the sample viral RNA while the splice donor sites are found below. A white box represents the encapsidation signal. Adapted from (107).

In HIV-2, the primary encapsidation signal, ψ, is located upstream of the major splice donor site. Thus, all HIV-2 RNA species contain this part of the apparently multipartite encapsidation signal. To complicate matters, there are twice as many RNA species in HIV-2 than in HIV-1. The doubling of RNA species results from 5'UTR splicing so instead of approximately 47 unique RNAs, there are now approximately 94 unique RNAs. Although all HIV-2 RNAs contain ψ, only the genomic RNA is encapsidated (115). The selectivity of the encapsidation is based on a cotranslational encapsidation mechanism. The genomic HIV-2 RNA is first translated, producing Gag/Gag-Pol proteins. Once synthesized, the Gag/Gag-Pol proteins bind preferentially, through their nucleocapsid domains, to the encapsidation signal of the translated genomic RNA (115). Hence, multiply spliced HIV-2 RNAs are not encapsidated because they do not produce Gag. However, the cotranslational mechanism does not explain why 5'UTR spliced *gag-pol* RNAs are not encapsidated.

Both 5'UTR unspliced and spliced *gag-pol* RNAs produce Gag and Gag-Pol proteins (131). Both *gag-pol* RNA species contain the primary encapsidation signal. Yet, 5'UTR spliced *gag-pol* RNAs are not encapsidated to detectable levels (131). So, how does HIV-2 preferentially encapsidate unspliced *gag-pol* RNAs over 5'UTR spliced *gag-pol* RNAs? A closer examination of both RNA species sequences reveals a key difference between the two that may influence encapsidation. The 5'UTR spliced *gag-pol* RNA lacks the C-box. As was shown in Chapter 3, the sequence identity of the C-box plays a role in encapsidation. Indeed, the C-box (by way of the CGI) forms a novel secondary structure that more than likely also affects the presentation of the primary encapsidation signal. Based on the results in Chapters 2 and 3, one could hypothesize that 5'UTR spliced *gag-pol* RNAs are not encapsidated because of the absence of the C-box and therefore the CGI. Instead these specialized RNAs are used to produce large quantities of Gag and Gag-Pol proteins. These proteins are then used in virion construction and encapsidation of unspliced *gag-pol* RNAs. This seems likely based on previous encapsidation studies by the Lever laboratory. Their studies demonstrated that Gag proteins produced by encapsidation-defective *gag-pol* HIV-2 RNAs were used in *trans* to encapsidate wild-type *gag-pol* RNAs (119). This seems the mechanism likely to be employed by HIV-2 when 5'UTR splicing is activated. To fully appreciate these dynamics during HIV-2 replication, a closer examination of 5'UTR splicing and the CGI is warranted.

4.2.1 The role of 5'UTR splicing in gag/gag-pol RNA translation

In retroviruses, Gag and Gag-Pol were previously thought to only be synthesized from the unspliced genomic RNA (130). Although this criterion was applied to all retroviruses, genomic RNA encapsidation differences were identified and resulted in the proposal of more than one model of retroviral encapsidation (130). The main models at the time differed as to whether or not two pools of unspliced *gag/gag-pol* RNAs existed in the cytoplasm (Figure 35). In model 1, two pools of unspliced *gag/gag-pol* RNAs exist in the cytoplasm, one designated for translation and the other for encapsidation. This model is supported by research on murine leukemia virus (MLV) encapsidation (145,146). In model 2, there is only one pool of unspliced

gag/gag-pol RNAs, which is translated before encapsidation. Model 2 is supported by research on HIV-1 and SIV_{cpz} encapsidation (145,146). Unlike HIV-1 and SIV_{cpz}, HIV-2 and SIV_{mac239} encapsidation do not exactly fit either model. Instead, the identification of 5'UTR splicing in both HIV-2 (131) and SIV_{mac239} (109) suggested an alternate mechanism of regulating translation and encapsidation. The alternate mechanism called for a third model, which was proposed in Chapter 2. In model 3, there are two pools of *gag/gag-pol* RNAs, similar to what is seen in model 1 (Figure 35). However, unlike model 1, the two RNA pools differ based on the presence of the 5'UTR intron. In model 3, there is a RNA pool specialized for translation in which the 5'UTR intron has been removed. The other RNA pool contains unspliced *gag/gag-pol* RNAs that undergo translation and encapsidation, similar to model 2 (131). As discussed above, 5'UTR splicing plays an indirect role in HIV-2 encapsidation. However, how 5'UTR splicing correlates to the overall cycle of HIV-2 infection in patients requires further analysis.

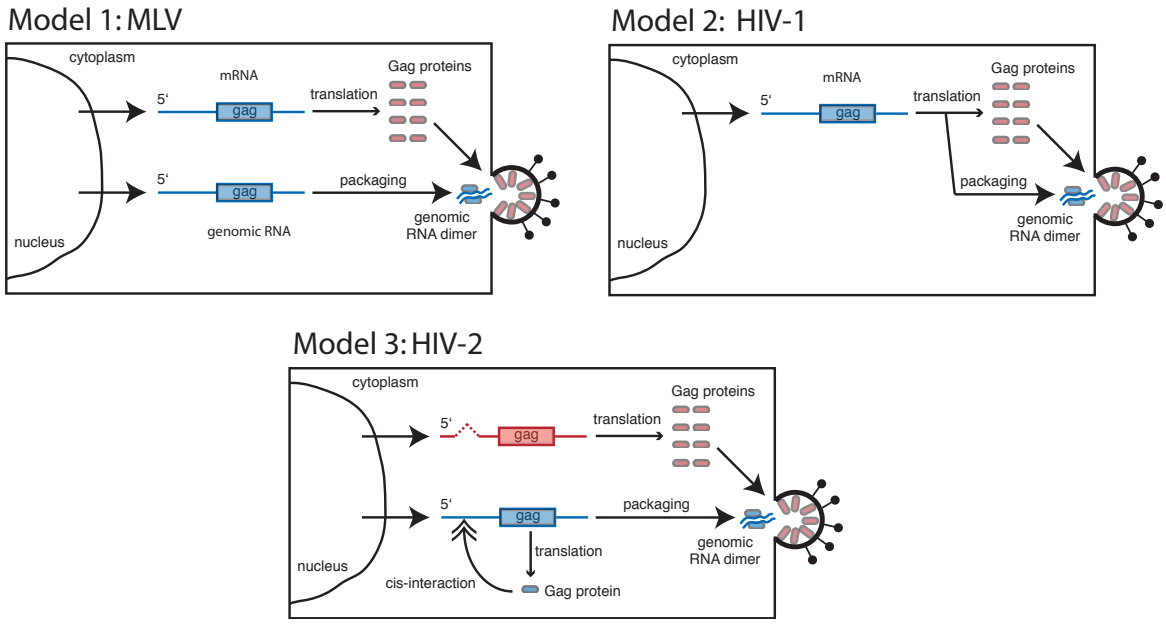

Figure 35. Models of translation and encapsidation of retrovirus *gag/gag-pol* RNAs.
In model 1, two pools of unspliced *gag/gag-pol* RNAs exist. The first pool is designated for translation-only. The second pool is designated for packaging-only. In model 2, one pool of unspliced *gag/gag-pol* RNAs exists. The unspliced *gag/gag-pol* RNAs are translated prior to packaging. In model 3, two pools of *gag/gag-pol* RNAs exist. The first pool is composed of *gag/gag-pol* RNAs that have undergone 5'UTR splicing. The 5'UTR spliced *gag/gag-pol* RNAs are specialized for translation. The second pool is composed of unspliced *gag/gag-pol* RNAs that are translated prior to encapsidation. Adapted from (130) and (131).

Results presented in Chapter 2 show that the 5'UTR spliced *gag/gag-pol* RNAs exist in asymptomatic HIV-2 infected patient samples. In cell culture and *in vitro*, the 5'UTR-spliced *gag/gag-pol* RNAs are specialized for Gag/Gag-Pol synthesis (131). Based on these 5'UTR splicing results, one might hypothesize that a balanced increase in translation of HIV-2 proteins through 5'UTR splicing may lead to an increase in virion production. Initially, increasing virion production may help to establish a persistent HIV-2 infection *in vivo*. In the long term, this increased virion production would lead to progression of HIV-2 infection to AIDS. To test this hypothesis, cell culture and animal model studies could be employed. Experiments designed to characterize 5'UTR splicing using PBMCs and a mouse model system are described below.

Experiment 1: Characterize 5'UTR splicing in PBMCs.

In Chapter 2, PBMCs of asymptomatic HIV-2 infected individuals were analyzed for the presence of 5'UTR spliced *gag/gag-pol* RNAs. Two of the five asymptomatic patient samples contained 5'UTR-spliced *gag/gag-pol* RNAs. At this time, analysis of 5'UTR splicing in symptomatic HIV-2 infected patient samples has not been performed. Because 5'UTR splicing may contribute to a surge in overall HIV replication, one might expect to see 5'UTR splicing occurring in the majority of PBMC samples of symptomatic HIV-2-infected individuals. Therefore, using RT-PCR to determine the presence/absence of 5'UTR splicing within the HIV-2 viral *gag/gag-pol* RNA populations of PBMC samples of symptomatic HIV-2 infected individuals would establish (or not) the correlation of prevalence of 5'UTR splicing and the transition to a symptomatic stage. In addition, two primary HIV-2 replication studies in PBMCs could provide further insight into the role of 5'UTR splicing in HIV-2 replication. The first would compare the viral replication kinetics of wild type and SD (-) in PBMCs. The second would determine viral fitness through competition experiments. Various ratios of wild type and SD(-) would be used to jointly infect PBMCs and each competition would be followed to determine the dominant virus. Viral replication kinetics of the viruses would be compared using ELISA, reverse transcriptase assays, RT-PCR, RPA and Western blot.

Experiment 2: Determining the role of HIV-2 5'UTR splicing using a mouse model system.

To elucidate how 5'UTR splicing functions during a bona fide HIV infection, an animal model could be used. Previous research has shown that the genetically modified $Rag2^{-/-}\gamma c^{-/-}$(Rag-hu) mice are capable of maintaining multi-lineage human hematopoiesis. The human T-cells produced by the Rag-hu mice are targets and therefore susceptible to HIV infection (147). Moreover, the mice can mount an immune response and maintain chronic HIV infection with continued CD4 T-cell depletion (148). Using this animal model would allow one to examine the role of 5'UTR splicing in early, mid-, and late HIV-2 infection. To test the role of 5'UTR splicing during a bona fide HIV-2 infection, mice will be infected with either HIV-2 wild-type or 5'UTR SD(-) viruses. HIV-2 infection will be followed over time and compared for both mice series using RNA *in situ* hybridization, CD4 T-cell depletion assays, PCR, and RT-PCR.

4.2.2 The role of the CGI in genomic RNA encapsidation

The CGI in HIV-1 is found in a structure that is capable of dimerization (53), suggesting a positive role in dimer formation. Although associated with dimer formation, the HIV-1 CGI is not found in mature virions. Instead, in the mature virion, the HIV-1 C-box interacts with nucleocapsid protein (139). In contrast to HIV-1, the CGI in HIV-2 is associated with inhibition of dimerization (47). Formation of the HIV-2 CGI is thought to result in structural entrapment of dimerization (SL1) and encapsidation (Ψ) signals (54). Interestingly, addition of nucleocapsid protein can relieve this inhibition in HIV-2 (47). Unfortunately, unlike HIV-1, there are no studies probing the CGI in HIV-2 mature virions. In spite of this, studies suggest a role for the C-box in HIV-2 encapsidation.

Studies described in Chapter 3 demonstrated that some mutations within the C-box result in a gain of function, i.e. enhanced genomic encapsidation. Why the C-box mutations resulted in enhanced encapsidation has not been definitively determined. Perhaps enhanced binding of nucleocapsid to the mutated HIV-2 C-box results in the enhanced encapsidation phenotype. The presence of guanosines within

the C-box of HIV-1 is associated with enhanced NC binding (138), which suggests that the guanosine mutations in the C-box119 virus may also enhance NC binding. A caveat to this is that whether this gain of function confers increased viral fitness during viral competition has not been determined. Further studies are needed to elucidate the fitness of the enhanced encapsidation phenotype and the mechanism behind the gain of function. A set of potential experiments to address this is described below.

Experiment 1: Determine the role of the C-box in NC binding and fitness. Mutation of the three nucleotides at positions 194-196 (CUC to GGA) in the C-box resulted in increased genomic HIV-2 encapsidation. The presence of guanosines in the HIV-1 C-box has been shown to be associated with increased NC binding (138). Therefore a comparison of the Gag/NC binding affinities of C-box119 to HIV-2 wild type could be performed. This would determine if the presence of guanosines in the C-box is associated with increased Gag/NC binding. In addition, an examination of whether the C-box119 gain of function contributes to overall fitness of the mutant virus during replication is worthwhile. This would entail performing competition cell culture studies using various ratios of C-box119 and HIV-2 wild type viruses. C8166 cells would be jointly infected with varying ratios of the two viruses and each competition would be followed until dominance of one virus occurred.

4.2.3 A HIV-2 model linking translation, dimerization, and encapsidation

Based on the results described in Chapters 2 and 3, a model that links translation, dimerization, and encapsidation can be proposed for HIV-2. Key elements of this model include 5'UTR splicing, nucleocapsid binding, and formation and disruption of the CGI. The first step in the model involves the role of 5'UTR splicing in translation.

5'UTR spliced *gag* and *gag-pol* RNAs display high levels of translational efficiency resulting in faster accumulation of Gag and Gag-Pol (Figure 36). Translation that produces more Gag and Gag-Pol equates to more nucleocapsid because it is a component of these poly-proteins. Because the 5' UTR spliced *gag-pol* RNAs lack the CGI, the genomic RNA pool uses the proteins synthesized. A

larger population of Gag and Gag-Pol would increase the probability of the binding of nucleocapsid domain to the C-box of the genomic RNAs during translation. The binding event would prevent the C-box from base pairing with the G-box. If the CGI does not form in HIV-2, SL1 and ψ will no longer be masked through a forced base pairing interaction. SL1 would be free to initiate dimerization and ψ would be available to initiate encapsidation. Thus, the increased translational efficiency from 5'UTR splicing would negatively affect the formation of the CGI, resulting in dimerization and encapsidation of the genomic RNA (Figure 36).

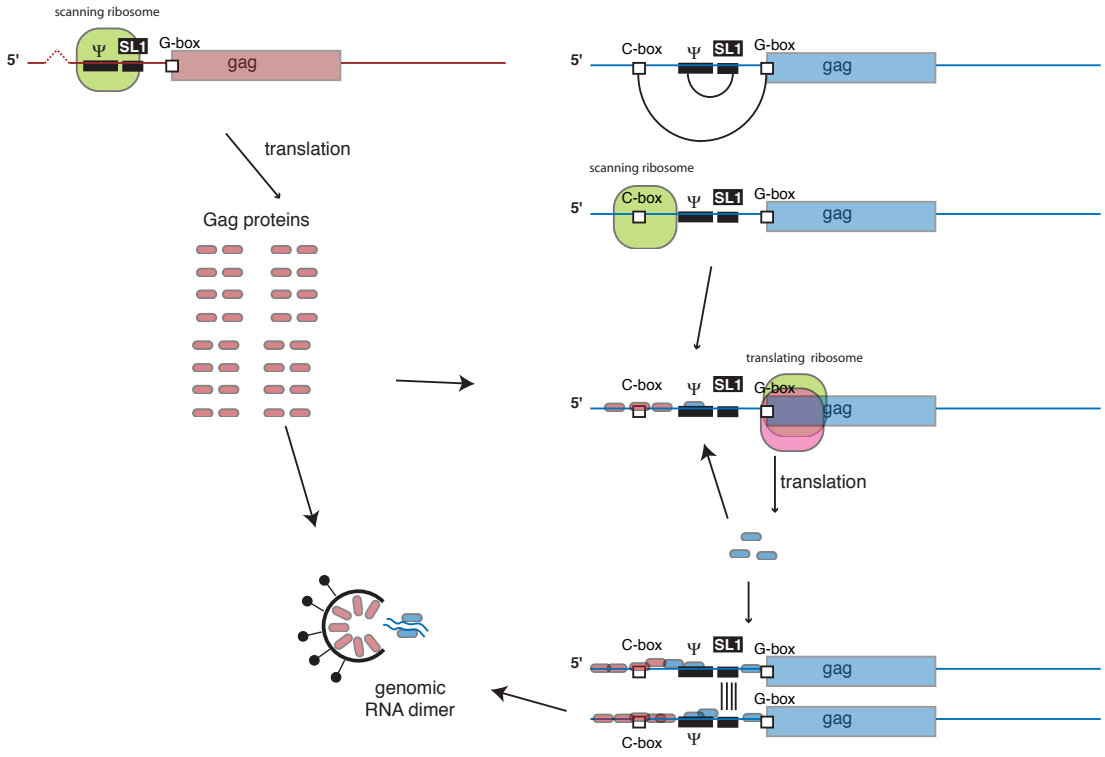

Figure 36. Coupling translation, dimerization and encapsidation in HIV-2.
The 5'UTR spliced *gag* RNA is shown in red. The 5'UTR unspliced genomic RNA is shown in blue. Gag proteins translated from the 5'UTR spliced RNA or unspliced genomic RNA are represented by red ovals or blue ovals, respectively. Base pairing interactions between elements are shown in black. Scanning and translating ribosomes disrupt the CGI. Accumulation of Gag proteins synthesized from 5'UTR spliced RNAs leads to binding of Gag to the C-box of the genomic RNA, preventing reformation of the CGI. Gag proteins synthesized from the unspliced genomic RNA can now bind to the displayed packaging signal. The SL1 element of the genomic RNA cannot interact with the SL1 element in another genomic RNA. Dimerization and packaging follow.

4.3 Implications of current work

The research described in Chapters 2 and 3 provides insight into mechanisms regulating several steps of the HIV replication cycle. These mechanisms include 5'UTR splicing and CGI formation/disruption. We have shown that 5'UTR splicing regulates translational efficiency of HIV-2 *gag/gag-pol* RNAs. We have also shown that the C-box and G-box have individual and cooperative roles in the viral cycle. Based on this research, the C-box and G-box are strong candidates for anti-retroviral targeting design.

Both the C-box and G-box are highly conserved RNA motifs in HIV-1 and HIV-2. In addition, these RNA motifs have been shown to regulate multiple steps of the HIV replication cycle. The C-box was shown to regulate encapsidation, 5'UTR

splicing, CGI formation, and translation in HIV-2. The G-box was shown to regulate HIV-2 Gag translation, Gag isoform ratios, and CGI formation. Because of their high level of sequence conservation and multiple roles in HIV replication, the C-box and G-box would be worthwhile targets for anti-retroviral drug design. By targeting just one motif, an effective drug could potentially affect the individual and collective role(s) of the motif and therefore may affect multiple steps of the viral replication cycle simultaneously.

References

1. CDC. (1981) Pneumocystis Pnuemonia-Los Angeles. *MMWR*, **30**, 250-252.
2. Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A. and 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**, 1425-1431.
3. Quagliarello, V. (1982) The Acquired Immunodeficiency Syndrome: current status. *Yale J Biol Med*, **55**, 443-452.
4. Hanrahan, J.P., Wormser, G.P., Maguire, G.P., Delorenzo, L.J. and Gavis, G. (1982) Opportunistic infections in prisoners. *N Engl J Med*, **307**, 498.
5. Masur, H., Michelis, M.A., Wormser, G.P., Lewin, S., Gold, J., Tapper, M.L., Giron, J., Lerner, C.W., Armstrong, D., Setia, U. *et al.* (1982) Opportunistic infection in previously healthy women. Initial manifestations of a community-acquired cellular immunodeficiency. *Ann Intern Med*, **97**, 533-539.
6. CDC. (1982) Opportunistic infections and Karposi's sarcoma among Haitians in the United States. *MMWR*, **31**, 353-361.
7. CDC. (1982) Pneumocystis carninii pneumonia among persons with hemophilia A. *MMWR*, **31**, 365-367.
8. Jaffe, H.W., Bregman, D.J. and Selik, R.M. (1983) Acquired immune deficiency syndrome in the United States: the first 1,000 cases. *J Infect Dis*, **148**, 339-345.
9. Dozier, N., Ballentine, R., Adams, S.C. and Okafor, K.C. (1983) Acquired immune deficiency syndrome and the management of associated opportunistic infections. *Drug Intell Clin Pharm*, **17**, 798-807.
10. Kornfeld, H., Vande Stouwe, R.A., Lange, M., Reddy, M.M. and Grieco, M.H. (1982) T-lymphocyte subpopulations in homosexual men. *N Engl J Med*, **307**, 729-731.
11. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B. *et al.* (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, **224**, 500-503.
12. Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C. *et al.* (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
13. Levy, J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A., Shimabukuro, J.M. and Oshiro, L.S. (1984) Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science*, **225**, 840-842.
14. Montagnier, L., Chermann, J.C., Barre-Sinoussi, F., Klatzmann, D., Wain-Hobson, S., Alizon, M., Clavel, F., Brun-Vezinet, F., Vilmer, E., Rouzioux, C. *et al.* (1984) Lymphadenopathy associated virus and its etiological role in AIDS. *Princess Takamatsu Symp*, **15**, 319-331.

15. Klatzmann, D., Barre-Sinoussi, F., Nugeyre, M.T., Danquet, C., Vilmer, E., Griscelli, C., Brun-Veziret, F., Rouzioux, C., Gluckman, J.C., Chermann, J.C. *et al.* (1984) Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science*, **225**, 59-63.
16. Gallo, R.C., Sarngadharan, M.G., Popovic, M., Shaw, G.M., Hahn, B., Wong-Staal, F., Robert-Guroff, M., Zaki Salahuddin, S. and Markham, P.D. (1986) HTLV-III and the etiology of AIDS. *Prog Allergy*, **37**, 1-45.
17. Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S. *et al.* (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, **220**, 865-867.
18. Gelmann, E.P., Popovic, M., Blayney, D., Masur, H., Sidhu, G., Stahl, R.E. and Gallo, R.C. (1983) Proviral DNA of a retrovirus, human T-cell leukemia virus, in two patients with AIDS. *Science*, **220**, 862-865.
19. Rabson, A.B. and Martin, M.A. (1985) Molecular organization of the AIDS retrovirus. *Cell*, **40**, 477-480.
20. Karpas, A. (2004) Human retroviruses in leukaemia and AIDS: reflections on their discovery, biology and epidemiology. *Biol Rev Camb Philos Soc*, **79**, 911-933.
21. Cheingsong-Popov, R., Weiss, R.A., Dalgleish, A., Tedder, R.S., Shanson, D.C., Jeffries, D.J., Ferns, R.B., Briggs, E.M., Weller, I.V., Mitton, S. *et al.* (1984) Prevalence of antibody to human T-lymphotropic virus type III in AIDS and AIDS-risk patients in Britain. *Lancet*, **2**, 477-480.
22. Curran, J.W. (1985) The epidemiology and prevention of the acquired immunodeficiency syndrome. *Ann Intern Med*, **103**, 657-662.
23. (1985) Screening for AIDS. *Med Lett Drugs Ther*, **27**, 29-30.
24. Serwadda, D., Mugerwa, R.D., Sewankambo, N.K., Lwegaba, A., Carswell, J.W., Kirya, G.B., Bayley, A.C., Downing, R.G., Tedder, R.S., Clayden, S.A. *et al.* (1985) Slim disease: a new disease in Uganda and its association with HTLV-III infection. *Lancet*, **2**, 849-852.
25. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C. *et al.* (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science*, **233**, 343-346.
26. Clavel, F., Brun-Vezinet, F., Guetard, D., Chamaret, S., Laurent, A., Rouzioux, C., Rey, M., Katlama, C., Rey, F., Champelinaud, J.L. *et al.* (1986) [LAV type II: a second retrovirus associated with AIDS in West Africa]. *C R Acad Sci III*, **302**, 485-488.
27. Reeves, J.D. and Doms, R.W. (2002) Human immunodeficiency virus type 2. *J Gen Virol*, **83**, 1253-1265.
28. Weiss, R.A. (2000) Getting to know HIV. *Trop Med Int Health*, **5**, A10-15.
29. Marlink, R., Kanki, P., Thior, I., Travers, K., Eisen, G., Siby, T., Traore, I., Hsieh, C.C., Dia, M.C., Gueye, E.H. *et al.* (1994) Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science*, **265**, 1587-1590.

30. Marlink, R.G., Ricard, D., M'Boup, S., Kanki, P.J., Romet-Lemonne, J.L., N'Doye, I., Diop, K., Simpson, M.A., Greco, F., Chou, M.J. *et al.* (1988) Clinical, hematologic, and immunologic cross-sectional evaluation of individuals exposed to human immunodeficiency virus type-2 (HIV-2). *AIDS Res Hum Retroviruses*, **4**, 137-148.
31. Marlink, R. (1996) Lessons from the second AIDS virus, HIV-2. *Aids*, **10**, 689-699.
32. Kanki, P., M'Boup, S., Marlink, R., Travers, K., Hsieh, C.C., Gueye, A., Boye, C., Sankale, J.L., Donnelly, C., Leisenring, W. *et al.* (1992) Prevalence and risk determinants of human immunodeficiency virus type 2 (HIV-2) and human immunodeficiency virus type 1 (HIV-1) in west African female prostitutes. *Am J Epidemiol*, **136**, 895-907.
33. Kanki, P.J., Travers, K.U., S, M.B., Hsieh, C.C., Marlink, R.G., Gueye, N.A., Siby, T., Thior, I., Hernandez-Avila, M., Sankale, J.L. *et al.* (1994) Slower heterosexual spread of HIV-2 than HIV-1. *Lancet*, **343**, 943-946.
34. Gottlieb, G.S., Hawes, S.E., Kiviat, N.B. and Sow, P.S. (2008) Differences in proviral DNA load between HIV-1-infected and HIV-2-infected patients. *Aids*, **22**, 1379-1380.
35. Gueudin, M., Benard, A., Chene, G., Matheron, S. and Simon, F. (2008) Significant differences in DNA viral load between HIV-1 and HIV-2 infected patients. *Aids*, **22**, 1519-1520.
36. Gueudin, M., Damond, F., Braun, J., Taieb, A., Lemee, V., Plantier, J.C., Chene, G., Matheron, S., Brun-Vezinet, F. and Simon, F. (2008) Differences in proviral DNA load between HIV-1- and HIV-2-infected patients. *Aids*, **22**, 211-215.
37. Pepin, J., Morgan, G., Dunn, D., Gevao, S., Mendy, M., Gaye, I., Scollen, N., Tedder, R. and Whittle, H. (1991) HIV-2-induced immunosuppression among asymptomatic West African prostitutes: evidence that HIV-2 is pathogenic, but less so than HIV-1. *Aids*, **5**, 1165-1172.
38. Andersson, S., Norrgren, H., da Silva, Z., Biague, A., Bamba, S., Kwok, S., Christopherson, C., Biberfeld, G. and Albert, J. (2000) Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Arch Intern Med*, **160**, 3286-3293.
39. de Silva, T.I., Cotten, M. and Rowland-Jones, S.L. (2008) HIV-2: the forgotten AIDS virus. *Trends Microbiol*, **16**, 588-595.
40. Freed, E.O. (2001) HIV-1 replication. *Somat Cell Mol Genet*, **26**, 13-33.
41. Kiss-Laszlo, Z. and Hohn, T. (1996) Pararetro- and retrovirus RNA: splicing and the control of nuclear export. *Trends Microbiol*, **4**, 480-485.
42. Tang, H., Kuhen, K.L. and Wong-Staal, F. (1999) Lentivirus replication and regulation. *Annu Rev Genet*, **33**, 133-170.
43. Balvay, L., Lopez Lastra, M., Sargueil, B., Darlix, J.L. and Ohlmann, T. (2007) Translational control of retroviruses. *Nat Rev Microbiol*, **5**, 128-140.
44. Paillart, J.C., Shehu-Xhilaga, M., Marquet, R. and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol*, **2**, 461-472.

45. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B. and 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**, 4945-4949.
46. Berkhout, B. (1996) Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol*, **54**, 1-34.
47. Lanchy, J.M., Rentz, C.A., Ivanovitch, J.D. and Lodmell, J.S. (2003) Elements located upstream and downstream of the major splice donor site influence the ability of HIV-2 leader RNA to dimerize in vitro. *Biochemistry*, **42**, 2634-2642.
48. Parkin, N.T., Cohen, E.A., Darveau, A., Rosen, C., Haseltine, W. and Sonenberg, N. (1988) Mutational analysis of the 5' non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *EMBO J*, **7**, 2831-2837.
49. Arya, S.K. and Gallo, R.C. (1988) Human immunodeficiency virus type 2 long terminal repeat: analysis of regulatory elements. *Proc Natl Acad Sci U S A*, **85**, 9753-9757.
50. Emerman, M., Guyader, M., Montagnier, L., Baltimore, D. and Muesing, M.A. (1987) The specificity of the human immunodeficiency virus type 2 transactivator is different from that of human immunodeficiency virus type 1. *EMBO J*, **6**, 3755-3760.
51. Pachulska-Wieczorek, K., Purzycka, K.J. and Adamiak, R.W. (2006) New, extended hairpin form of the TAR-2 RNA domain points to the structural polymorphism at the 5' end of the HIV-2 leader RNA. *Nucleic Acids Res*, **34**, 2984-2997.
52. Chatterjee, P., Garzino-Demo, A., Swinney, P. and Arya, S.K. (1993) Human immunodeficiency virus type 2 multiply spliced transcripts. *AIDS Res Hum Retroviruses*, **9**, 331-335.
53. Abbink, T.E., Ooms, M., Haasnoot, P.C. and Berkhout, B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*, **44**, 9058-9066.
54. Baig, T.T., Strong, C.L., Lodmell, J.S. and Lanchy, J.M. (2008) Regulation of primate lentiviral RNA dimerization by structural entrapment. *Retrovirology*, **5**, 65.
55. Lanchy, J.M., Szafran, Q.N. and Lodmell, J.S. (2004) Splicing affects presentation of RNA dimerization signals in HIV-2 in vitro. *Nucleic Acids Res*, **32**, 4585-4595.
56. Song, R., Kafaie, J. and Laughrea, M. (2008) Role of the 5' TAR stem--loop and the U5-AUG duplex in dimerization of HIV-1 genomic RNA. *Biochemistry*, **47**, 3283-3293.
57. Huthoff, H. and Berkhout, B. (2001) Two alternating structures of the HIV-1 leader RNA. *RNA*, **7**, 143-157.
58. Dirac, A.M., Huthoff, H., Kjems, J. and Berkhout, B. (2002) Regulated HIV-2 RNA dimerization by means of alternative RNA conformations. *Nucleic Acids Res*, **30**, 2647-2655.

59. Ooms, M., Huthoff, H., Russell, R., Liang, C. and Berkhout, B. (2004) A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol*, **78**, 10814-10819.
60. Lanchy, J.M., Ivanovitch, J.D. and Lodmell, J.S. (2003) A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA*, **9**, 1007-1018.
61. Abbink, T.E. and 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**, 11601-11611.
62. Damgaard, C.K., Andersen, E.S., Knudsen, B., Gorodkin, J. and Kjems, J. (2004) RNA interactions in the 5' region of the HIV-1 genome. *J Mol Biol*, **336**, 369-379.
63. Sonenberg, N. and Hinnebusch, A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, **136**, 731-745.
64. Merrick, W.C. (2003) Mini-Series: Modern Metabolic Concepts Initiation of Protein Biosynthesis in Eukaryotes. *Biochemistry and Molecular Biology Education*, **31**, 378-385.
65. Merrick, W.C. (2004) Cap-dependent and cap-independent translation in eukaryotic systems. *Gene*, **332**, 1-11.
66. Goossen, B. and Hentze, M.W. (1992) Position is the critical determinant for function of iron-responsive elements as translational regulators. *Mol Cell Biol*, **12**, 1959-1966.
67. Kozak, M. (1989) The scanning model for translation: an update. *J Cell Biol*, **108**, 229-241.
68. Kozak, M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*, **83**, 2850-2854.
69. Charron, M., Shaper, J.H. and Shaper, N.L. (1998) The increased level of beta1,4-galactosyltransferase required for lactose biosynthesis is achieved in part by translational control. *Proc Natl Acad Sci U S A*, **95**, 14805-14810.
70. Han, B., Dong, Z. and Zhang, J.T. (2003) Tight control of platelet-derived growth factor B/c-sis expression by interplay between the 5'-untranslated region sequence and the major upstream promoter. *J Biol Chem*, **278**, 46983-46993.
71. Moore, M.J. and Proudfoot, N.J. (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, **136**, 688-700.
72. Diem, M.D., Chan, C.C., Younis, I. and Dreyfuss, G. (2007) PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol*, **14**, 1173-1179.
73. Le Hir, H., Nott, A. and Moore, M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci*, **28**, 215-220.
74. Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell*, **21**, 239-248.

75. Tarun, S.Z., Jr., Wells, S.E., Deardorff, J.A. and Sachs, A.B. (1997) Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. *Proc Natl Acad Sci U S A*, **94**, 9046-9051.
76. Mangus, D.A., Evans, M.C. and Jacobson, A. (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol*, **4**, 223.
77. Wethmar, K., Smink, J.J. and Leutz, A. (2010) Upstream open reading frames: molecular switches in (patho)physiology. *Bioessays*, **32**, 885-893.
78. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283-292.
79. Kozak, M. (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *Embo J*, **16**, 2482-2492.
80. Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13-37.
81. Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*, **15**, 8125-8148.
82. Pelletier, J., Kaplan, G., Racaniello, V.R. and Sonenberg, N. (1988) Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Mol Cell Biol*, **8**, 1103-1112.
83. Pilipenko, E.V., Blinov, V.M., Chernov, B.K., Dmitrieva, T.M. and Agol, V.I. (1989) Conservation of the secondary structure elements of the 5'-untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Res*, **17**, 5701-5711.
84. Nomoto, A., Lee, Y.F. and Wimmer, E. (1976) The 5' end of poliovirus mRNA is not capped with m7G(5')ppp(5')Np. *Proc Natl Acad Sci U S A*, **73**, 375-380.
85. Etchison, D., Milburn, S.C., Edery, I., Sonenberg, N. and Hershey, J.W. (1982) Inhibition of HeLa cell protein synthesis following poliovirus infection correlates with the proteolysis of a 220,000-dalton polypeptide associated with eucaryotic initiation factor 3 and a cap binding protein complex. *J Biol Chem*, **257**, 14806-14810.
86. Balvay, L., Soto Rifo, R., Ricci, E.P., Decimo, D. and Ohlmann, T. (2009) Structural and functional diversity of viral IRESes. *Biochim Biophys Acta*, **1789**, 542-557.
87. Filbin, M.E. and Kieft, J.S. (2009) Toward a structural understanding of IRES RNA function. *Curr Opin Struct Biol*, **19**, 267-276.
88. Jan, E. and Sarnow, P. (2002) Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J Mol Biol*, **324**, 889-902.
89. Stassinopoulos, I.A. and Belsham, G.J. (2001) A novel protein-RNA binding assay: functional interactions of the foot-and-mouth disease virus internal ribosome entry site with cellular proteins. *Rna*, **7**, 114-122.
90. Buck, C.B., Shen, X., Egan, M.A., Pierson, T.C., Walker, C.M. and Siliciano, R.F. (2001) The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J Virol*, **75**, 181-191.

91. Brasey, A., Lopez-Lastra, M., Ohlmann, T., Beerens, N., Berkhout, B., Darlix, J.L. and Sonenberg, N. (2003) The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J Virol*, **77**, 3939-3949.
92. Berkhout, B., Arts, K. and Abbink, T.E. (2011) Ribosomal scanning on the 5'-untranslated region of the human immunodeficiency virus RNA genome. *Nucleic Acids Res*.
93. Chamond, N., Locker, N. and Sargueil, B. (2010) The different pathways of HIV genomic RNA translation. *Biochem Soc Trans*, **38**, 1548-1552.
94. Ohlmann, T., Lopez-Lastra, M. and Darlix, J.L. (2000) An internal ribosome entry segment promotes translation of the simian immunodeficiency virus genomic RNA. *J Biol Chem*, **275**, 11899-11906.
95. Nicholson, M.G., Rue, S.M., Clements, J.E. and Barber, S.A. (2006) An internal ribosome entry site promotes translation of a novel SIV Pr55(Gag) isoform. *Virology*, **349**, 325-334.
96. Herbreteau, C.H., Weill, L., Decimo, D., Prevot, D., Darlix, J.L., Sargueil, B. and Ohlmann, T. (2005) HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat Struct Mol Biol*, **12**, 1001-1007.
97. Ricci, E.P., Soto Rifo, R., Herbreteau, C.H., Decimo, D. and Ohlmann, T. (2008) Lentiviral RNAs can use different mechanisms for translation initiation. *Biochem Soc Trans*, **36**, 690-693.
98. Ricci, E.P., Herbreteau, C.H., Decimo, D., Schaupp, A., Datta, S.A., Rein, A., Darlix, J.L. and Ohlmann, T. (2008) In vitro expression of the HIV-2 genomic RNA is controlled by three distinct internal ribosome entry segments that are regulated by the HIV protease and the Gag polyprotein. *RNA*, **14**, 1443-1455.
99. Weill, L., James, L., Ulryck, N., Chamond, N., Herbreteau, C.H., Ohlmann, T. and Sargueil, B. (2010) A new type of IRES within gag coding region recruits three initiation complexes on HIV-2 genomic RNA. *Nucleic Acids Res*, **38**, 1367-1381.
100. Geballe, A.P. and Gray, M.K. (1992) Variable inhibition of cell-free translation by HIV-1 transcript leader sequences. *Nucleic Acids Res*, **20**, 4291-4297.
101. Edery, I., Petryshyn, R. and Sonenberg, N. (1989) Activation of double-stranded RNA-dependent kinase (dsl) by the TAR region of HIV-1 mRNA: a novel translational control mechanism. *Cell*, **56**, 303-312.
102. Silverman, R.H. and Sengupta, D.N. (1990) Translational regulation by HIV leader RNA, TAT, and interferon-inducible enzymes. *J Exp Pathol*, **5**, 69-77.
103. D'Agostino, D.M., Felber, B.K., Harrison, J.E. and Pavlakis, G.N. (1992) The Rev protein of human immunodeficiency virus type 1 promotes polysomal association and translation of gag/pol and vpu/env mRNAs. *Mol Cell Biol*, **12**, 1375-1386.
104. Alvarez, E., Menendez-Arias, L. and Carrasco, L. (2003) The eukaryotic translation initiation factor 4GI is cleaved by different retroviral proteases. *J Virol*, **77**, 12392-12400.

105. Alvarez, E., Castello, A., Menendez-Arias, L. and Carrasco, L. (2006) HIV protease cleaves poly(A)-binding protein. *Biochem J*, **396**, 219-226.
106. Vaishnav, Y.N. and Wong-Staal, F. (1991) The biochemistry of AIDS. *Annu Rev Biochem*, **60**, 577-630.
107. Purcell, D.F. and Martin, M.A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol*, **67**, 6365-6378.
108. Bohne, J., Wodrich, H. and Krausslich, H.G. (2005) Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. *Nucleic Acids Res*, **33**, 825-837.
109. Viglianti, G.A., Rubinstein, E.P. and Graves, K.L. (1992) Role of TAR RNA splicing in translational regulation of simian immunodeficiency virus from rhesus macaques. *J Virol*, **66**, 4824-4833.
110. SenGupta, D.N., Berkhout, B., Gatignol, A., Zhou, A.M. and Silverman, R.H. (1990) Direct evidence for translational regulation by leader RNA and Tat protein of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*, **87**, 7492-7496.
111. Viglianti, G.A., Sharma, P.L. and Mullins, J.I. (1990) Simian immunodeficiency virus displays complex patterns of RNA splicing. *J Virol*, **64**, 4207-4216.
112. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. and Alizon, M. (1987) Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature*, **326**, 662-669.
113. Ryan-Graham, M.A. and Peden, K.W. (1995) Both virus and host components are important for the manifestation of a Nef- phenotype in HIV-1 and HIV-2. *Virology*, **213**, 158-168.
114. McCann, E.M. and Lever, A.M. (1997) Location of cis-acting signals important for RNA encapsidation in the leader sequence of human immunodeficiency virus type 2. *J Virol*, **71**, 4133-4137.
115. Kaye, J.F. and Lever, A.M. (1998) Nonreciprocal packaging of human immunodeficiency virus type 1 and type 2 RNA: a possible role for the p2 domain of Gag in RNA encapsidation. *J Virol*, **72**, 5877-5885.
116. Salahuddin, S.Z., Markham, P.D., Wong-Staal, F., Franchini, G., Kalyanaraman, V.S. and Gallo, R.C. (1983) Restricted expression of human T-cell leukemia--lymphoma virus (HTLV) in transformed human umbilical cord blood lymphocytes. *Virology*, **129**, 51-64.
117. MacNeil, A., Sarr, A.D., Sankale, J.L., Meloni, S.T., Mboup, S. and Kanki, P. (2007) Direct evidence of lower viral replication rates in vivo in human immunodeficiency virus type 2 (HIV-2) infection than in HIV-1 infection. *J Virol*, **81**, 5325-5330.
118. MacNeil, A., Sankale, J.L., Meloni, S.T., Sarr, A.D., Mboup, S. and Kanki, P. (2007) Long-term intrapatient viral evolution during HIV-2 infection. *J Infect Dis*, **195**, 726-733.
119. Griffin, S.D., Allen, J.F. and Lever, A.M. (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**, 12058-12069.
120. Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F. and Liuni, S. (2001) Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, **276**, 73-81.
121. Hong, X., Scofield, D.G. and Lynch, M. (2006) Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol*, **23**, 2392-2404.
122. Pickering, B.M. and Willis, A.E. (2005) The implications of structured 5' untranslated regions on translation and disease. *Semin Cell Dev Biol*, **16**, 39-47.
123. Hughes, T.A. (2006) Regulation of gene expression by alternative untranslated regions. *Trends Genet*, **22**, 119-122.
124. Babendure, J.R., Babendure, J.L., Ding, J.H. and Tsien, R.Y. (2006) Control of mammalian translation by mRNA structure near caps. *RNA*, **12**, 851-861.
125. Bannwarth, S. and Gatignol, A. (2005) HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. *Curr HIV Res*, **3**, 61-71.
126. Yilmaz, A., Bolinger, C. and Boris-Lawrie, K. (2006) Retrovirus translation initiation: Issues and hypotheses derived from study of HIV-1. *Curr HIV Res*, **4**, 131-139.
127. Frendewey, D. and Keller, W. (1985) Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell*, **42**, 355-367.
128. Ruskin, B. and Green, M.R. (1985) Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. *Nature*, **317**, 732-734.
129. Green, M.R. (1986) Pre-mRNA splicing. *Annu Rev Genet*, **20**, 671-708.
130. Butsch, M. and Boris-Lawrie, K. (2002) Destiny of unspliced retroviral RNA: ribosome and/or virion? *J Virol*, **76**, 3089-3094.
131. Strong, C.L., Lanchy, J.M., Dieng-Sarr, A., Kanki, P.J. and Lodmell, J.S. (2009) A 5'UTR-spliced mRNA isoform is specialized for enhanced HIV-2 gag translation. *J Mol Biol*, **391**, 426-437.
132. Baig, T.T., Lanchy, J.M. and Lodmell, J.S. (2009) Randomization and in vivo selection reveal a GGRG motif essential for packaging human immunodeficiency virus type 2 RNA. *J Virol*, **83**, 802-810.
133. Berkhout, B. and Klaver, B. (1993) In vivo selection of randomly mutated retroviral genomes. *Nucleic Acids Res*, **21**, 5020-5024.
134. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.
135. D'Souza, V. and Summers, M.F. (2004) Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature*, **431**, 586-590.
136. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci*, **23**, 297-301.
137. D'Souza, V. and Summers, M.F. (2005) How retroviruses select their genomes. *Nat Rev Microbiol*, **3**, 643-655.

138. Spriggs, S., Garyu, L., Connor, R. and Summers, M.F. (2008) Potential intra- and intermolecular interactions involving the unique-5' region of the HIV-1 5'-UTR. *Biochemistry*, **47**, 13064-13073.
139. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C. and Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol*, **6**, e96.
140. Gottlinger, H.G., Sodroski, J.G. and Haseltine, W.A. (1989) Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*, **86**, 5781-5785.
141. Schwartz, S., Felber, B.K. and Pavlakis, G.N. (1992) Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol Cell Biol*, **12**, 207-219.
142. Anderson, J.L., Johnson, A.T., Howard, J.L. and Purcell, D.F. (2007) Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs. *J Virol*, **81**, 4664-4676.
143. Krummheuer, J., Johnson, A.T., Hauber, I., Kammler, S., Anderson, J.L., Hauber, J., Purcell, D.F. and Schaal, H. (2007) A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG. *Virology*, **363**, 261-271.
144. Kuiken, C., Foley, B., Leitner, T., Apetrei, C., Hahn, B., Mizrachi, I., Mullins, J., Rambaut, A., Wolinsky, S. and Korber, B. (2010) *HIV Sequence Compendium 2010*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.
145. Messer, L.I., Levin, J.G. and Chattopadhyay, S.K. (1981) Metabolism of viral RNA in murine leukemia virus-infected cells; evidence for differential stability of viral message and virion precursor RNA. *J Virol*, **40**, 683-690.
146. Dorman, N. and Lever, A. (2000) Comparison of viral genomic RNA sorting mechanisms in human immunodeficiency virus type 1 (HIV-1), HIV-2, and Moloney murine leukemia virus. *J Virol*, **74**, 11413-11417.
147. Berges, B.K., Wheat, W.H., Palmer, B.E., Connick, E. and Akkina, R. (2006) HIV-1 infection and CD4 T cell depletion in the humanized Rag2-/-gamma c-/- (RAG-hu) mouse model. *Retrovirology*, **3**, 76.
148. Berges, B.K., Akkina, S.R., Remling, L. and Akkina, R. (2010) Humanized Rag2(-/-)gammac(-/-) (RAG-hu) mice can sustain long-term chronic HIV-1 infection lasting more than a year. *Virology*, **397**, 100-103.