Analysis of the relationship between the structure and function of the HIV-1 Rev Response Element (RRE)

Chringma Sherpa
Germantown, Maryland

Master of Science in Medical Microbiology, Tribhuvan University, Nepal, 2008

A Dissertation presented to the Graduate Faculty
Of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Microbiology, Immunology, and Cancer Biology

University of Virginia
November, 2014

ABSTRACT

HIV-1 is a retrovirus of global health interest as the causative agent of Acquired Immunodeficiency Syndrome (AIDS). It is also used as a molecular tool to study various eukaryotic cellular processes. For instance, major discoveries on the Crm1 cellular nuclear export were made using the HIV Rev protein and its RNA binding partner, the Rev Response Element (RRE). The RRE is a *cis*-acting RNA element with multiple stem-loops present in all intron-retaining HIV mRNAs. Binding of the Rev protein to the primary binding site, and Rev multimerization on other regions of the RRE, is required for the nucleo-cytoplasmic export of these mRNAs. This is an essential step in HIV replication. However, the precise secondary structure of the HIV-1 RRE remains controversial. Studies have reported that the RRE has either 4 or 5 stem-loops, which differ only in the rearrangement of regions that lie outside of the primary Rev binding site. To understand the role played by these regions, we have examined the relationship between these two structures and Rev-RRE activity.

In vitro transcribed NL4-3 RRE was found to migrate as a “doublet” band on a native polyacrylamide gel. Using in-gel *Selective 2’ Hydroxyl Acylation analyzed by Primer Extension* (SHAPE), we found that one of these bands consisted of an RRE with a 5 stem-loop structure, whereas the other was a 4 stem-loop structure. Thus, our data demonstrate, for the first time, that the NL4-3 RRE exists in two alternative structures.

To study the significance of these alternative structures, we made RRE mutants predicted to allow only one or the other of the structures to form. The predictions were confirmed using SHAPE. We then compared the activity of the two forms to each other and to the wt RRE. Analysis of the complexes that each RRE formed with purified Rev protein *in vitro* showed no significant difference in Rev binding affinity between the two structures. However, we observed differences in the migration rates of these complexes on native gels, suggesting structural differences. The RREs were also tested for their abilities to promote viral replication, by inserting each RRE into the Nef region of an RRE-defective provirus which contained RRE mutations that do not change the Env protein. Growth kinetics and competition assays showed that the virus with the 5 stem-loop RRE had a higher replicative fitness than the virus with either the wt or the 4 stem-loop RRE. Between the wt and the 4 stem-loop RRE-containing viruses, the virus with the wt RRE appeared more fit. These results suggest that HIV may use two alternative RRE secondary structures to modulate replication, potentially allowing adaptation to environmental demands in time and/or space, analogous to the use of riboswitches in bacteria.

Dedicated to my loving parents,
Nimi Sherpa and Pasang Norbu Sherpa

ACKNOWLEDGEMENT

At the very outset, I express my deep gratitude to Dr. David Rekosh and Dr. Marie-Louise Hammarskjöld for allowing me to do science under their mentorship. Their enthusiasm for their research has taught me how exciting the scientific process can be, and I hope this excitement lives-on in my future scientific endeavors. I am also very grateful to Dr. Ann Beyer, Dr. Barbie Ganser-Pornillos, Dr. Dean H. Kedes, Dr. Ian Glomski, and Dr. Owen Pornillos for giving me their valuable time as the faculty of my PhD dissertation committee. Their advice and expertise were extremely helpful especially in streamlining my research.

My thanks also goes to Dr. Stuart LeGrice, our collaborater at NCI, NIH for providing me the opportunity to learn SHAPE technology in his lab under the guidance of Dr. Jason Raush. I would also like to take this opportunity to thank all the past and present members of HamRek lab during my time as a graduate student. I have, directly or indirectly, learnt a lot from each of them. In particular, I would like to thank Dr. Yeou-cherng Bor, who was always willing to advise me not just on my experiments but also on life matters outside the lab. I also thank my fellow PhD graduate student, Siripong Tongjai, for being a great friend. I would also like to thank Benson Iweriebor, Laurie Gray, Eileen Trainum, Susan Prasad, and Dr. Denis Tebit, Dr. Emily Sloan for helping me with their expertise.

I consider myself very lucky to have had many great mentors in my educational journey leading to the completion of my PhD. I am deeply indebted to one such mentor, Dr. Ronald Bauerle. I am grateful to him for all his help during my cultural assimilation process during the first few years in the US, and for never losing interest in both my personal and my professional development. I would also like especially thank Judy Bauerle and Anusa Thapa for their friendship and support.

Finally, without a strong family support, my PhD journey would not have been possible. I thank my husband Phudorji Sherpa, especially, for happily accompanying me to the lab during many of my wee lab-hours, listening patiently to my disappointments at failed experiments, and filling me with renewed optimism to try again. My thanks also go to my sisters - Dr. Dakki Sherpa, Mingma Sherpa, and Lakpa Sherpa for all their encouragement and support. Last but not the least, I am beyond words in thanking my parents, Pasang Norbu Sherpa and Nimi Sherpa; after all, I am who I am today because of them.

TABLE OF CONTENTS

CHAPTER II – MATERIALS AND METHODS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1- INTRODUCTION

In the first half of this chapter, a general overview of HIV and its replication cycle is presented. This part of the chapter introduces the Rev-RRE pathway as a step in the HIV life cycle. Using the viral replication cycle as the anchor, this chapter also describes the functions of different HIV macromolecules so that the readers can analyze the Rev-RRE data presented in this dissertation more holistically. The second half of this chapter focuses on the Rev-RRE system and presents the rationale for the study.

HIV-1 classification

Human Immunodeficiency Virus (HIV) is placed in the family retroviridae, subfamily orthoretrovirinae, and genus lentivirus by the *International Committee on Taxonomy of Viruses* (http://www.ictvonline.org). It was placed under the family retroviridae because of its reverse transcriptase activity, which was observed for the first time at the Pasteur Institute (193). The genus lentivirus (*lente*-Latin for "*slow*") contains viruses with characteristic long clinical incubation periods. Besides HIV, the genus includes equine infectious anemia virus (EIAV) and simian immunodeficiency virus (SIV). The five other genera in the subfamily orthoretrovirinae are: alpharetrovirus (avian sarcoma and leucosis virus), betaretrovirus [mouse mammary tumor virus (MMTV)], gammaretrovirus [murine leukemia virus (MLV)], deltaretrovirus [human T-lymphtropic virus (HTLV)], and epsilonretrovirus (walleye dermal sarcoma virus).

Phylogenetic analysis of HIV isolates based on sequence homology has

identified two types of HIV, namely HIV-1 and HIV-2. It is believed that both originated in West-Central Africa in non-human primates before the zoonotic transmission to humans (122, 175). HIV-1 was first isolated in 1983 (11) and proven to be the etiological agent of Acquired Immunodeficiency Syndrome (AIDS) (89, 187, 207). Subsequently, HIV-2 was discovered in 1986 in patients in West Africa (34). HIV-2 infected individuals can also develop AIDS but usually with an even longer incubation period, and HIV-2 has a lower morbidity compared to HIV-1 (175). HIV-1 isolates are further grouped into three distinct genetic subtypes: M (main), O (outlier), and N (non-M, non-O) (197, 226). According to the World Health Organization (WHO) estimate, between 33.1 and 37.2 million people were living with HIV in 2013. More than 95% of viral isolates from these infected people fall into the group M subtype. Group M consists of at least 9 different subtypes or clades (A, B, C, D, F, G, H, J, and K) and more than 60 circulating recombinant forms have also been described (175, 226). The high genetic variabilty of HIV-1 manifest in its several groups, subtypes and an increasing number of circulating forms is a unique characteristic of HIV. This results mainly from the error-prone reverse transcription by the viral reverse transcriptase (3.4×10^{-5} mutations per base pair per replication cycle) coupled with the high replication rate of the virus (baseline rate of viral production is approximately 10^{10} virions per day in an actively infected person) (177). Thus, in a single person, millions of viral variants are produced in just a few days. Recombination between HIV genomic RNA strands from different viral strains in a co-infected cell is another factor contributing to the HIV diversity (18).

The genetic diversity of the HIV isolates necessitated the use of some reference HIV strains to facilitate comparisons among different studies on HIV. The most widely used reference strains are HXB2 and NL4-3, which are laboratory-adapted strains derived from HIV-1 isolates belonging to the subgroup B of the group M. HXB2 was the first reported HIV-1 molecular clone (213). It was derived from HIV-1 isolate LAI/IIB (formerly called HTLV-IIIB) from a patient's blood sample that have been expanded and maintained in immortalized human T cells. The HXB2 clone is defective in HIV accessory proteins Vpr, Vpu, and Nef. One the other hand, the HIV sequence of NL4-3 is a chimera of 5.8 kb long 5' region of NY5 HIV-1 isolate and the 3.8 kb long 3' region of the LAV HIV-1 isolate joint at the EcoRI site unique to both the isolates (2). Unlike HXB2, this molecular clone is not defective in any of the known HIV proteins. Throughout this study, NL4-3 is used as reference or the WT virus.

HIV-1 structure and genomic organization

The HIV-1 consists of an RNA genome and fifteen proteins. Many of the HIV-1 proteins are cleaved products of the viral polyproteins. Table 1 shows the full names and abbreviations of the HIV-1 proteins and the polyproteins.

Table 1: List of HIV-1 proteins

	Types of HIV-1 proteins (abbreviation)	HIV-1 proteins (abbreviation)
1.	Viral core structural proteins – cleavage products of Gag-specific antigen (Gag or Pr55) polyprotein	Matrix protein (MA or p17)
		Capsid protein (CA or p24)
		Spacer peptide 2 (SP2)
		Nucleocapsid protein (NC or p7)
		Spacer peptide 1 (SP1)
		P6 protein
2.	Viral envelop structural proteins – cleavage products of Env (envelope or gp160) polyprotein	Surface glycoprotein (SU or gp120)
		Transmembrane glycoprotein (TM or gp41)
3.	Essential viral enzymes – cleavage products of GagPol polyprotein	Reverse transcriptase (RT). It has two functional domains – the polymerase domain (p51) and RNase H domain (p15)
		Integrase (IN or p31)
		Protease (Pr or p10)
4.	Essential regulatory proteins	Regulator of viral transcription (Rev)
		Trans-activator of transcription (Tat)
5.	Accessory proteins	Negative regulatory factor (Nef)
		Virion infectivity factor (Vif)
		Viral protein R (Vpr)
		Viral protein unique (Vpu)

The HIV virion is nearly spherical in shape and is ~120 nm in diameter (84, 93, 187). As shown in Figure 1, the core of the virion contains the viral genome, consisting of two copies of non-covalently linked single stranded RNA. Each genomic RNA is ~9.7 kb long (191, 237) and is bound tightly to nucleo-capsid protein (NC or p7) and the protein p6 within a conical capsid (CA or p24) (93). The capsid is, in turn, enclosed within an inner layer of matrix protein (MA or p17) and an outer envelope layer. The envelope layer is derived mostly from the host plasma membrane and contains viral envelope proteins [the surface glycoprotein (SU or gp120) and the trans-membrane protein (TM or gp41)] (84). The viral core also contains three viral enzymes: reverse transcriptase (RT), integrase (IN), and protease (PR). The virion also contains small amounts of some of the accessory HIV proteins: Vif, Vpr, and Nef (26, 229).

The genomic RNA is produced from the un-spliced mRNA transcribed from the integrated form of the virus (the provirus) (160). The provirus has a repeated sequence called the long terminal repeats (LTRs) at both its ends. The HIV ORFs are located in between the two LTRs (Figure 1) (38, 236). The 5' LTR contains the HIV promoter and the 3' LTR contains the poly-adenylation signal (84). Early nucleotide sequence analysis of HIV-1 isolates revealed an important feature of HIV-1 that was different from the prototypic retroviruses (e.g., ASLV, MLV) (5, 6, 191). Unlike the genome of the prototypes which contain only the *gag*, *pol*, and *env* ORFS, the HIV-1 genome included several additional and overlapping open reading frames (ORFs) located between the 3' end of the *pol* ORF and the 3' LTR.

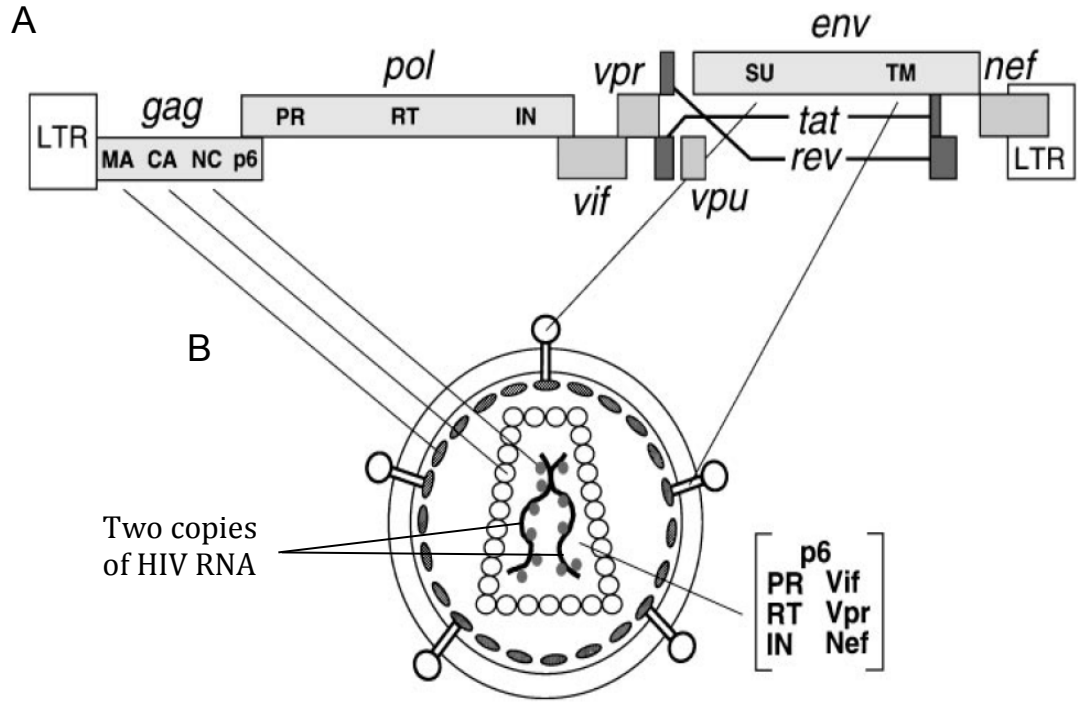


Figure 1: Genomic organization of HIV-1 Adapted from (84).

.

(A) The proviral DNA showing the HIV-1 open reading frames represented as boxes. The HIV-1 has many different overlapping open reading frames that encode 15 different viral proteins.

(B) The structure of a mature HIV virion. The viral genome along with some viral and cellular components is enclosed within the conical viral core. The core is enclosed within the matrix protein layer, which in turn, is covered by the envelope layer containing spikes of viral glycoproteins.

Later, it was shown that other retroviruses like HTLV-1 and HIV-2 also shared a genomic complexity similar to HIV-1. This led to the classification of retroviruses into simple and complex retroviruses based on their genomic organization (38, 236). The simple retroviruses (e.g. MLV) contain *gag*, *pol*, and *env* ORFs that are present in all retroviruses. They direct the synthesis of Gag, GagPol, and Env poly-proteins from the unspliced and singly spliced mRNAs. These poly-proteins are cleaved into elementary structural proteins and enzymes by viral or cellular proteases (38, 236). In HIV, both the Gag and GagPol poly-proteins are translated from the un-spliced mRNA. The Gag poly-protein is cleaved into four Gag proteins (MA, CA, NC, and p6). The GagPol poly-protein is cleaved into three essential HIV enzymes (RT, IN, and PR) in addition to the proteins in the Gag domain. The Env poly-protein is translated from a singly spliced *env* mRNA. It is cleaved into gp120 (SU) and gp41 (TM) (84). The complex retroviruses (e.g. HIV), on the other hand, contain many additional ORFs (38, 236). In HIV, these encode for 4 accessory proteins and 2 regulatory proteins from singly spliced and multiply spliced mRNAs. Accessory proteins are not necessary for viral growth in cell culture, while the regulatory proteins are essential for viral growth. The accessory proteins are Vpr, Vpu, Vif and Nef. The regulatory proteins are Rev and Tat (84). In this study, we have compared the functional activity of the different structural variants of the binding partner of Rev called the Rev Response Element (RRE).

Each LTR is 640-bp long and is composed of three sub-regions namely the unique 3’ (U3) region, the repeat (R) region, and the unique 5’ (U5) region

(219). The U3 region, which occupies the 5' end of each LTR is 453-bp long. The central region of each LTR contains the 98-bp R region, which is followed by the 83-bp long U5 region. Transcription initiates at the first base of the R region in the 5' LTR and poly-adenylation occurs immediately after the last base of R in the 3' LTR. Therefore, all the viral mRNAs including the genomic RNA have "R" followed by "U5" at its 5' end and "U3" followed by "R" at its 3' end. The 5' LTR contains many *cis*-acting elements upstream and downstream of the transcription initiation site at the 5' end of R that provide DNA binding sites for various transcription factors. The upstream elements contain the core promoter region, the enhancer region, and the modulatory region (176, 199, 245). The core promoter region contains the TATA-box, and three Sp1 sites. The enhancer region includes two nuclear factor-κB sites. The modulatory region contains three CCAAT/enhancer binding protein (C/EBP) sites, the activating transcription factor/cyclic AMP response element binding (ATF/ CREB) region, and an upstream stimulatory factor binding site. The downstream elements include AP-1 motifs, an AP-3-like (AP-3L) motif, a C/EBP/ NFAT (nuclear factor for activated T cells) downstream binding site (DS3), two downstream sequence element (DSE) sites and one downstream binding factor (DBF-1). There are also two Sp1 sites in the U5 and *gag* leader sequence regions (44, 65, 66, 176, 189, 198, 200).

The 5' untranslated region (5' UTR) region in HIV genomic RNA is highly structured and includes the R and U5 from the 5' LTR region and regions immediately downstream of the 5' LTR (Figure 2).

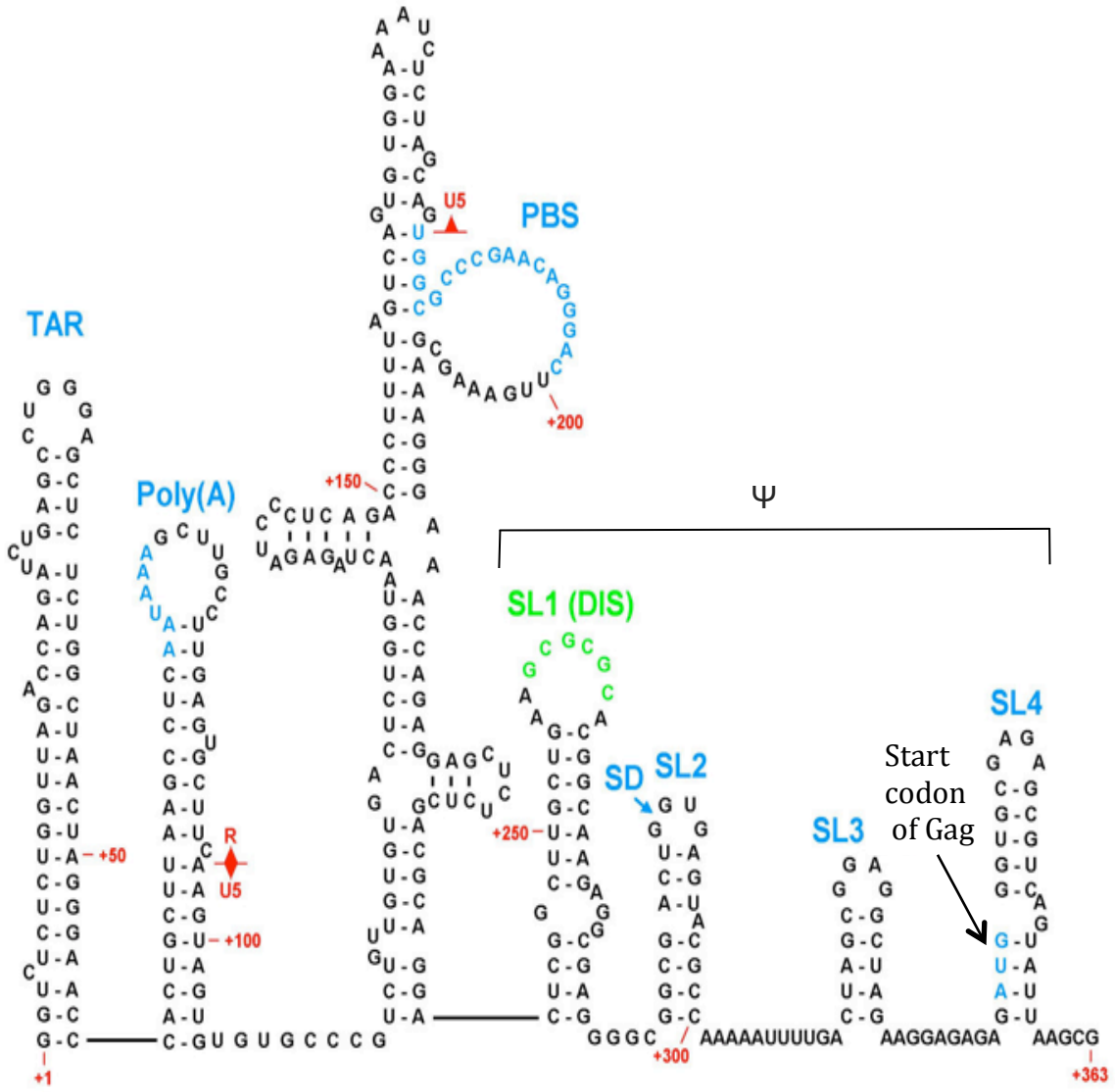


Figure 2: Secondary structure model of the HIV-1 5'UTR. Nucleotides and numbering correspond to the HIV-1 HXB2 sequence. See text for discussion of 5' UTR regions. The AUG at the start of Gag is shown in blue. Adapted from (35) and (16).

The regions downstream of the 5' LTR include the primer binding site (PBS), the packaging signal (Ψ), the dimerization site (DIS), and the major 5' splice donar (SD). The PBS is positioned just after the last nucleotide of U5 (Figure 2). It binds $tRNA^{Lys,3}$, which acts as the primer for reverse transcription. The packaging signal consists of four different stem-loops (SL1, SL2, SL3, and SL4). This region binds NC to facilitate the packaging of viral genome into the virion (36). The DIS is located within the SL1 of Ψ. The DIS sequence of the two genomic RNAs form a kissing loop, which facilitates RNA incorporation into the virion (29, 36). The major 5' splice site (SD), located within SL2, acts as the first or only 5' splice site in all of the sub-genomic HIV mRNAs. The SL4 region contains the Gag start codon (Figure 2).

The 5' end of HIV RNA transcribed from R contains the TAR (trans-activation region) hairpin (Figure 2), which provides a binding site for Tat. R also contains the AAUAAA poly-adenylation signal for binding the cleavage and poly-adenylation factor (CPSF) and the GU/U rich binding site for the 3' terminal cleavage stimulation factor (CstF). Only the 3' LTR poly-adenylation signal is active. It is speculated that the suppression of the 5' LTR polyadenylation signal may be due to the factors binding to the neighboring regions of 5'UTR. Ashe *et al.* have shown that the binding of U1 snRNP to the major 5' splice site (SD) leads to the inactivation of the 5' LTR poly A site (8). They hypothesize two possible mechanisms for this U1 snRNP mediated inhibition of the use of this poly-A site. U1 snRNP might either occupy the poly A site making it inaccessible to the poly A factors or interact with the poly A factors making them less available

to bind poly A site.

Replication cycle

Viral entry

The natural host cells of HIV-1 are human T-cells and macrophages expressing CD4 (the receptor) and CCR5/CXCR4 (the co-receptors) on their surfaces (23). HIV also infects dendritic cells (DC), which then present the viral particles to other CD4+ cells. DCs facilitate transmission of the virus especially from a mucosal site of infection (60, 130).

The virus enters into the host cell by binding to the receptors and co-receptors on the cell surface through its envelope spikes (Figure 3). Each viral envelope spike consists of a trimer of heterodimers of gp120 and gp41 (141, 255). The gp120, first, makes specific contact with the CD4 receptor. This induces a conformational change in gp120 that exposes the binding site for the co-receptors on gp120 (134, 196). Co-receptor engagement leads to the unfolding of gp41, revealing the two heptad repeats in gp41 - heptad repeat 1 (HR1) and heptad repeat 2 (HR2).

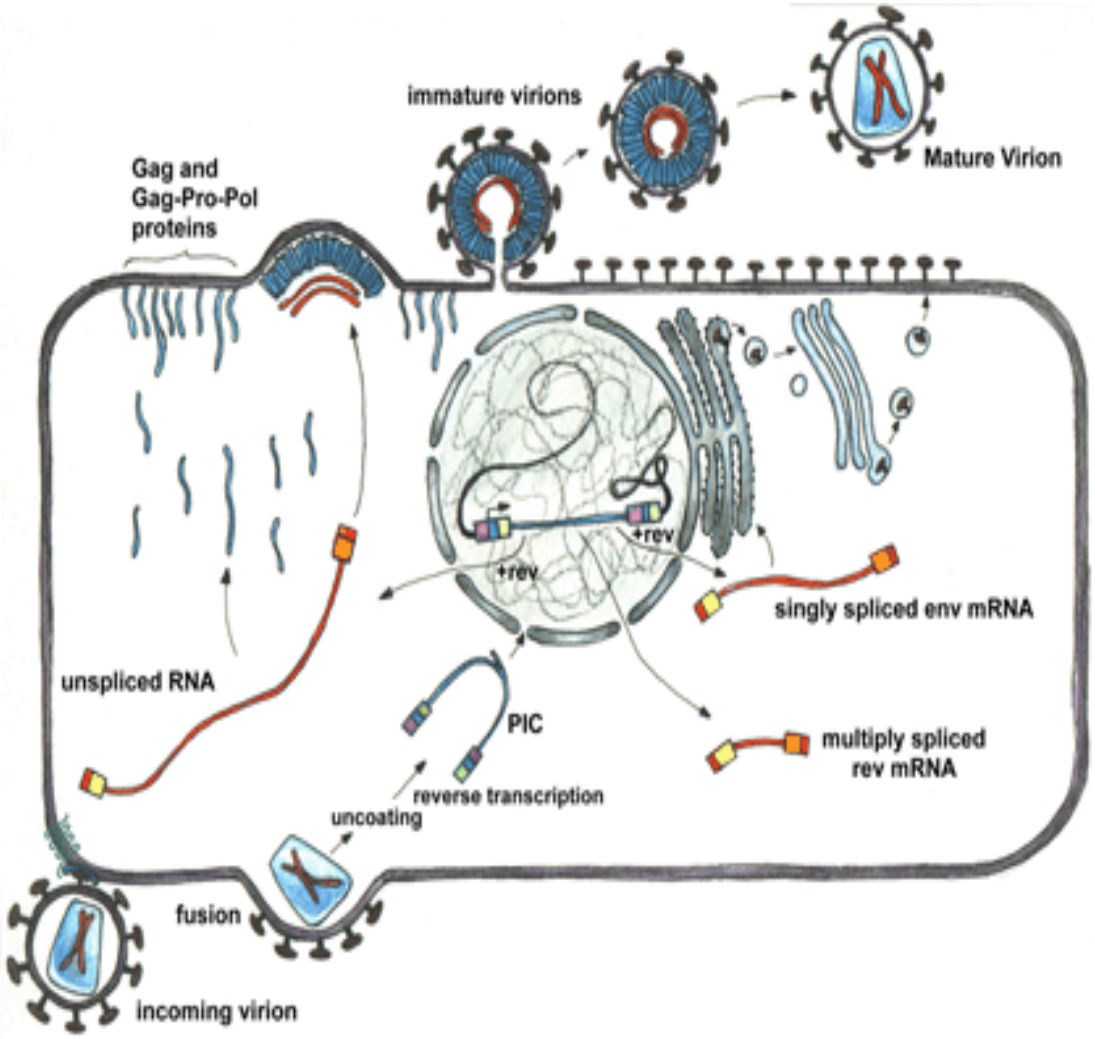


Figure 3: Overview of the HIV-1 life cycle: Following the binding of the virus to CD4 and CCR5 or CXCR4 on the surface of the host cell, the virus envelope fuses to the cell membrane releasing the viral core into the host cell cytoplasm. The viral genome is then reverse transcribed to form double stranded DNA, which enters into the nucleus and gets integrated into the host cell DNA. The proviral DNA is transcribed into the viral mRNAs, which are then exported into the cytoplasm, where they are expressed into the viral proteins and form the viral genome. These viral components assemble at the cell-membrane to form the immature viral particle, which buds out of the cell and undergoes cleavage by the viral protease to form the mature virus. Adapted from Eric M. Poeschla lab website, Mayo Clinic.

Upon unfolding, HR1 inserts into the cell membrane of the host cell. HR2 then folds onto itself which leads to the fusion of the viral membrane with the cell membrane (28, 30). Upon fusion, the viral core containing the viral genome and viral proteins (RT, IN and others) is released into the cell cytoplasm.

Some studies have shown that prior to binding specifically to CD4 molecules on the host cell during entry, the viral envelope makes nonspecific interactions with the cell-surface glycans or adhesion factors (40). It is believed that such non-specific interaction increases the infectivity by pre-concentrating the virion particles at the cell surface. Other modes of viral entry into the host cells have also been documented. These include entry through virological synapses and through membrane bridges (222).

Reverse transcription and integration

Following entry, partial dissolution of the capsid shell takes place to facilitate reverse transcription. Premature dissolution of CA, however, at this early stage of replication impairs viral replication (81, 83). Reverse transcription of the genomic RNA is initiated using $tRNA^{Lys,3}$ as the primer. The RNA-dependent and DNA-dependent DNA polymerization activity of the RT reverse transcribes the viral RNA to double stranded viral DNA. This process is very complex and involves two DNA strand transfer reactions. The genomic RNA is degraded by the RNase H activity of the RT in this process (84). The HIV nucleoprotein complex, containing mostly the linear double stranded viral DNA and several viral proteins including Vpr, RT, IN, and MA, is called the pre-

integration complex (PIC) (21).

This complex is then transported to the nucleus, where the viral DNA gets integrated into the host chromosome. To date, no viral factor or cellular factor has been shown to be indispensible for the nuclear import of the complex. However, various studies have suggested that the karyophilic elements in the PIC proteins like Vpr, IN, and MA and a three-stranded DNA flap, a structure present in neo-synthesized viral DNA, specified by the central polypurine tract-central termination sequence (cPPT-CTS) mediate the nuclear localization of this complex (59, 99, 148, 194). Recently, Riviere *et al*. (194) studied the aforementioned determinants implicated in nuclear import. They found that none of these elements, individually or in combination, were absolutely required for nuclear import either in primary cells (both dividing and non-dividing) or in Hela cells. Besides being implicated in nuclear import, Vpr also blocks cell division by arresting the cells in G2 phase of the cell cycle (120). However, the role of this activity of Vpr in viral life cycle is still not clear.

The enzyme IN catalyzes two well-characterized catalytic steps, referred to as 3'-processing and strand transfer, resulting in the integration of the viral DNA into the cellular chromatin (27). The 3' processing occurs in the cytoplasm. Here, the enzyme uses its 3' processing activity to trim the two 3' nucleotides from each end of the linear viral DNA duplex, leaving 5' CA overhangs at each end. The strand transfer step takes place in the nucleus. In this step, the enzyme catalyzes the 3' hydroxyl group of the overhangs to create a double stranded break at the integration sites, joining the 3' end of the overhangs to the 5'

phosphate group of the target DNA (70, 73). Finally, the un-joined 5' end of the viral DNA is ligated to the target DNA by host enzymes (248).

IN can catalyze both of the catalytic steps in vitro in the absence of any other viral or cellular co-factors. Nevertheless, various potential cellular co-factors of integrase have been reported using *in-vitro* reconstitution of enzymatic activity of salt-stripped PICs (72, 136), yeast two-hybrid assays (68, 121), and co-immunoprecipitation assays (33, 140). These cellular factors include LEDGF/p75, integrase interactor 1(INI-1), high mobility group chromosomal protein A1(HMGA1), and barrier-to-autointegration factor (BAF). Of all these cellular factors, the function of LEDGF/p75 in retroviral integration is the most well-defined. It has been shown that siRNA knockdown of LEDGF/p75 dramatically decreases the nuclear localization of HIV-1 DNA (147). In direct nuclear-import assays, it was later shown that the recombinant HIV-1 integrase is still actively imported in the nucleus in the absence of LEDGF/p75 (68). Therefore, the most accepted role of LEDGF/p75 is the tethering of viral DNA to the host chromosome for integration. It is believed that the integration takes place at sites where the chromosomal DNA in the chromatin is more accessible (161). Not surprisingly, HIV integration is mostly reported in and around active genes (208).

Synthesis of viral mRNAs

Following integration, in most cells, the integrated proviral DNA gets replicated and transcribed along with the cellular DNA. However, in latently

infected cells, the provirus can remain dormant within the cellular chromosome for many months or years (203). Upon activation, transcription initiates from the promoter at the 5' LTR of the proviral DNA using the host RNA polymerase II. Early viral transcripts are mostly abortive transcription products. A few polyadenylated and multiply spliced mRNAs, however, are also produced (as described below). They get transported to the nucleus and express Rev, Tat, and Nef. The Tat protein shuttles to the nucleus to promote efficient transcription elongation from the 5'LTR. However, unlike cellular transcription factors that bind DNA, Tat is actually an RNA binding protein (76, 202). It binds to the TAR region of the nascent viral mRNA and recruits the host cell transcription elongation factor b (P-TEFb) to the TAR region (17, 87). P-TEFb is made of cyclin-dependent kinase 9 (CDK9) and cyclin T1. The CDK9 component of P-TEFb phosphorylates the CTD tail of RNA polymerase II promoting transcription elongation. It has been reported that the binding of Tat to TAR activates transcription from the HIV LTR by at least 1000-fold (109). With the help of Tat, the virus makes a full length ~9 kb long pre-mRNA using the promoter at the 5'LTR and the poly-adenylation signal at the 3' LTR. This pre-mRNA undergoes alternative splicing to produce three classes of mRNAs (Figure 4). They are as follows:

1. Unspliced mRNA: This is a 9 kb long mRNA that encodes the Gag and Gag-Pol polyproteins. It also serves as genomic RNA and gets packaged into the virion.

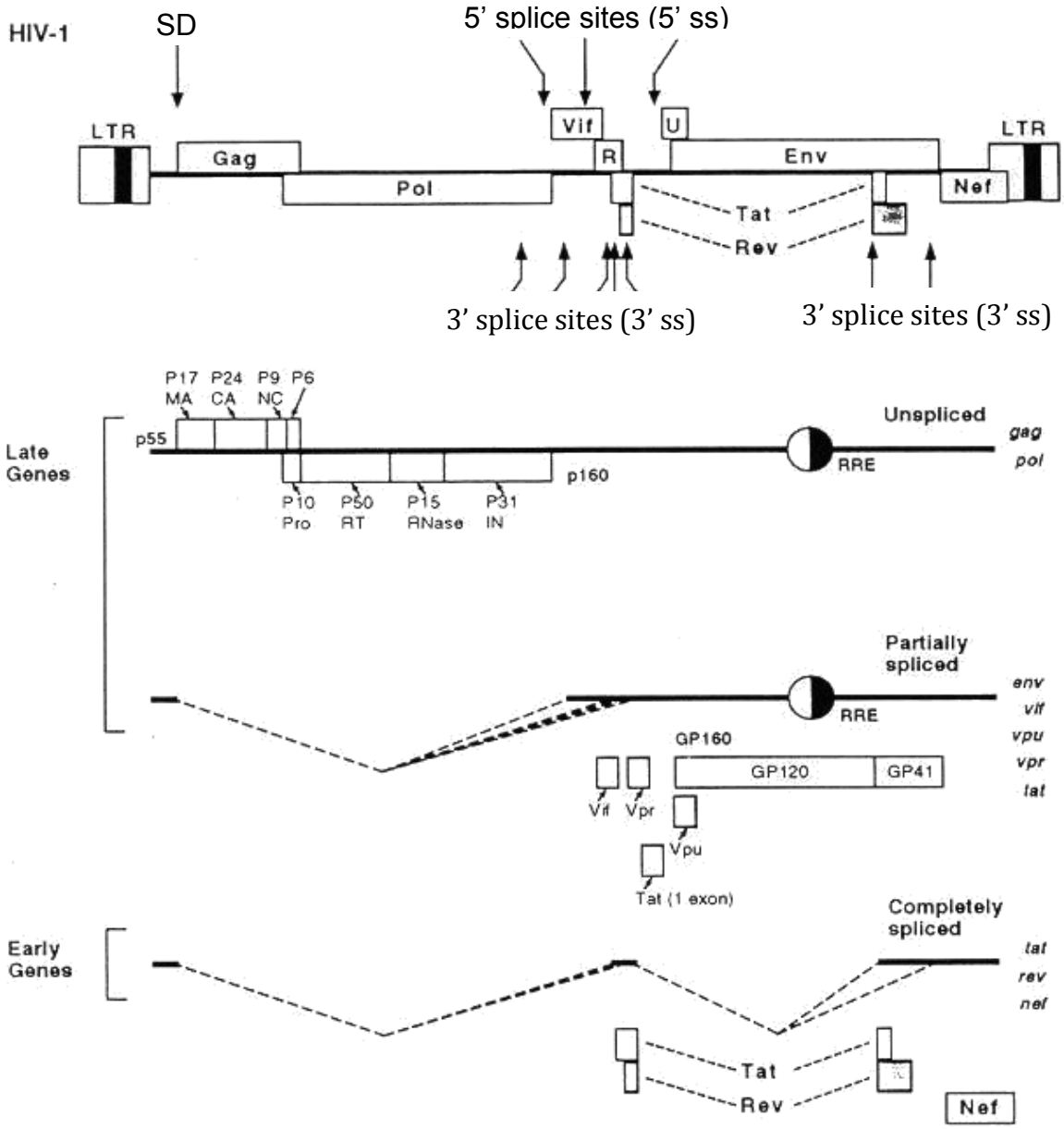


Figure 4: Alternative Splicing of HIV-1 RNA. The predominant 5' splice sites and 3' splice sites are shown on the proviral DNA. Using the promoter in the 5' LTR and the polyA site on the 3' LTR, viral pre-mRNA is transcribed. The pre-mRNA undergoes alternative splicing to form three classes of viral mRNAs, using different combination of 5' and 3' splice sites. Only the predominant species of splice variants and their products are shown (http://hivinsite.ucsf.edu). Refer to Table 1 for the full names of most of the abbreviations in the figure)

2. Singly spliced mRNA (also called partially spliced mRNA). These mRNAs are about 4-5 kb long and encode Env, Vif, Vpu, and Vpr. They are generated using the major 5' splice site (SD) in the 5' UTR in the genome in combination with different 3' splice sites.
3. Multiply (or completely) spliced RNA. These mRNAs are about 2kb long and do not contain any introns. They encode Rev, Tat, and Nef.

Thus, the virus uses intron retention and other forms of alternative splicing to increase the coding capacity of a small 9 kb viral genome. Although 15 different viral proteins are produced in the virus infected cells, RT-PCR analysis of viral mRNAs from these cells has revealed the existence of at least 30 different mRNAs (188, 209). This mRNA diversity arises from the variable usage of five 5'splice sites and more than ten 3' splice sites present in the viral genome; some of these RNAs encode the same protein. The functional importance of this phenomenon is still not clear.

Another trick used by the virus probably to increase its genomic capacity takes place at the level of translation. This process is called ribosomal frame-shifting. It allows both Gag and Gag-Pol poly-proteins to be made from the same genomic RNA. Frame-shifting results in the production of the Gag-Pol poly-protein and is mediated by a heptameric UUUUUUA slippery sequence within the *gag* gene and a downstream short stem-loop. The ribosome undergoes -1 translational frame-shifting when it encounters this motif, shifting from the *gag* to *pol* reading frame without interrupting translation (173). Such frame-shifting happens 5% of the time and explains why the ratio of Gag to Gag-Pol poly-

protein is 20:1.

Nuclear export of viral mRNAs and the Rev-RRE pathway

Following transcription and splicing, the fully spliced mRNAs are exported into the cytoplasm where they are translated into Tat, Nef, and Rev. The Tat protein shuttles back into the nucleus, where it increases transcription efficiency, as described below. Unlike the fully spliced mRNAs, the un-spliced and singly spliced mRNAs are exported and expressed along with their introns. Intron retention is generally not allowed for cellular mRNAs but is essential for the viral replication, as these mRNAs make up the progeny viral genome and encode essential viral proteins. It is now widely believed that the cell often recognizes incompletely spliced mRNAs with retained introns as faulty and degrades them (25). All retroviruses, including HIV, must overcome the host cellular restriction on the nucleo-cytoplasmic export of their intron-retaining mRNAs.

One of the mechanisms used by cellular mRNAs to overcome this restriction, is through the use of a *cis*-acting RNA element called the constitutive transport element (CTE), that directs these incompletely spliced mRNAs to the Nxf1 mRNA export pathway (139).

Many simple retroviruses such as MPMV (22), MLV (183, 206), ALV (246) contain a *cis*-acting CTE or a CTE-like element in their RNAs to facilitate the export of the RNAs via the Nxf1 mRNA export pathway. Most of the complex retroviruses like HTLV (105, 211), MMTV (158), EIAV(86, 156), and Jaagsiekte Sheep virus (107, 167) encode a specific regulatory protein that interacts with a

cis element in the viral transcripts to support nuclear export of the intron-retaining viral transcripts through the CRM1 pathway. However, there is little sequence similarity among the regulatory proteins and among the RNA elements across these viruses. Interestingly, foamy viruses are unique in that they use a cellular RNA-binding protein, HuR, instead of a viral regulatory protein to facilitate nuclear export of the unspliced viral transcripts through the CRM1 pathway (13).

HIV uses the Rev protein and its RNA binding partner called the RRE (Rev response element) to dodge the cellular RNA surveillance mechanisms (98, 103, 151). The RRE is a *cis*-acting RNA element present in all the intron-retaining viral mRNAs. The current model of the Rev-RRE pathway is as follows: Rev, produced from a fully spliced mRNA, shuttles back to the nucleus and binds specifically as a monomer to the primary Rev binding site on the RRE (12, 104, 115, 127, 227). Next, more Rev molecules (about 6-12) bind co-operatively along the other regions of the RRE in a process that involves protein-protein as well as protein-RNA interactions (41, 46, 51, 52, 62, 104, 149, 155, 249, 250). Rev binding to the primary Rev binding site and Rev oligomerization on the RRE are both essential for Rev-RRE function (47, 149, 168, 249). While the former allows the Rev to make specific contacts with the RRE (13), the latter increases the affinity of Rev-RRE complex by ~500 fold (51). Therefore, it appears that the tight cooperative nature of Rev-RRE binding explains how Rev manages to specifically and preferentially bind to the RRE RNA in the presence of many other RNAs within the cell.

Previously, there were reports that suggested that Rev molecules might

also bind to the RRE as a preformed oligomer. These findings were mostly based on the fact that Rev has the tendency to self-associate (168, 249). However, many Rev-RRE assembly studies, performed under ensemble-average conditions, employing assays like filter binding, gel shift, chemical foot-printing, mass spectrometry (137), fluorescence resonance energy transfer (FRET) (235), and surface plasma resonance (234), have indicated that Rev incrementally binds to the RRE. A recent single molecule Rev-RRE binding study done by using total internal reflection fluorescence (TIRF) microscopy has also shown that Rev binds to the RRE in monomeric increments (186).

The Rev-RRE complex is recognized by CRM1 and RAN-GTP forming an export competent ribonucleoprotein (RNP) complex (82, 170). The complex is then targeted to the nuclear pore where it interacts with the nucleoporins, resulting in its translocation to the cytoplasmic side. Once in the cytoplasm, the RRE containing mRNAs are translated into Gag, Gag-Pol, Env polyproteins, and accessory viral proteins.

Another mechanism employed by retroviruses to avoid the removal of these introns is the incorporation of sub-optimal splice sites in the viral genome (31). Thus, the introns are inefficiently removed by the splicing machinery, which in turn, results in the accumulation of mRNAs with retained introns. Various other positive and negative *cis*-acting elements within the viral genome can also regulate the levels of intron-retaining viral mRNAs in the nucleus by regulating splicing (146).

Viral assembly

The Gag polyprotein ($Pr55^{Gag}$) and the Gag-Pol polyprotein ($Pr160^{Gag-Pol}$) are synthesized on free ribosomes, from where they are targeted to the plasma membrane (Figure 3). The trafficking of the $Pr55^{Gag}$ to the cell membrane is both sufficient and necessary for the assembly and release of the viral particles (94). From the N-terminus, the Gag poly-protein ($Pr55^{Gag}$) contains domains for MA, CA, spacer peptide SP1, NC, spacer peptide SP2, and p6 (refer to table 1 for the full names of the viral components). Membrane targeting of the $Pr55^{Gag}$ is mediated by the MA domain through its myristic acid at the N-terminal Gly residue (24, 96) and a patch of conserved basic residues (106, 190, 253). The MA domain initially binds to phosphatidylinositol-4,5-bisphosphate [PI(4,5)P2] in the cytoplasmic aspect of the plasma membrane. This binding exposes the amino-terminal myristoyl group of MA, allowing $Pr55^{Gag}$ to stably anchor in the plasma membrane (204, 205, 254). PI(4,5)P2 is concentrated only in the inner leaflet of plasma membrane. This differential distribution of PI(4,5)P2 has been suggested as the reason behind the preferentially trafficking of $Pr55^{Gag}$ to plasma membrane rather than to intracellular membranes (69).

The critical determinant for $Pr160^{Gag-Pol}$ incorporation into particles has been controversial. Some groups (111, 218) have reported that the CA domain of $Pr160^{Gag-Pol}$ is essential for the incorporation of the $Pr160^{Gag-Pol}$ into viral particles. Another group has shown that CA is not required for $Pr160^{Gag-Pol}$ incorporation into the virion (152).

The Env poly-protein (gp160) is translated on ribosomes associated with the endoplasmic reticulum (ER). It is glycosylated co-translationally. On the ER, it also oligomerizes predominantly into trimers, which interact with newly synthesized CD4 (HIV receptors on host cells). Vpu degrades the CD4 to release the trimers (20). Another role of Vpu is the down-regulation of the cell surface expression of MHC class I proteins, thereby preventing the recognition of the infected cells by CTLs (124). The gp160 trimer is then transported to the golgi complex, where it is cleaved by a cellular protease to form mature gp120 and gp41 (100). In most cell lines and primary cells like PMBC and MDM, the Env glycoproteins are targeted to the lipid rafts of the plasma membrane through the gp41 cytoplasmic tail. On the plasma membrane, the gp120-gp41 complex interacts directly with the MA domain, resulting into its incorporation into the virion (63, 85).

Incorporation of Vpr takes place through specific interaction with the p6 domain (131, 144). The protein p6 also facilitates particle release (112). Various cellular factors like ICAM-1, HLA-II, actin, actin-binding proteins, cyclophilin A, ubiquitin, lysyl-tRNA-synthetase, t-RNA and many RNA-binding proteins are also found incorporated in the virion although the precise role of many of them in the viral life cycle is not clear (171).

Two copies of un-spliced RNA non-covalently held at the DIS loop are incorporated into the virion as genomic RNA (119). This dimerization of the genomic RNA is necessary for packaging and viral infectivity (159). It requires specific binding of the RNA with the NC domain. RNA packaging requires 5'UTR

sequences, particularly the Ψ signal, which spans from the major 5' splice site to the Gag start codon (42) (Figure 2). The requirement of the sequence downstream of the 5' splice site for packaging may explain why only the un-spliced mRNAs out of the three classes of viral mRNAs get packaged into the virion.

Budding and maturation

Viral budding from the cell membranes requires interaction between the p6 domain of Gag and the cellular E vacuolar protein sorting (VPS) proteins. VPS proteins normally mediate multi-vescicular body formation in the cells. The p6 domain contains two L (late) domains called P(T/S)AP and LYPX1-3L(95, 220). The former binds to ESCRT-I while the latter binds to ALIX (92, 220). Engagement of these two VPS proteins by p6 initiates a series of events that results in viral budding.

The final release of the viral particles from the plasma membrane is restricted by a host cell trans-membrane protein, tetherin. HIV-1 Vpu binds to tetherin which facilitates viral particle release (165, 233). Interaction between the CD4 molecules and the viral Env on the cell surface can also prevent particle release. The viral Nef protein inhibits this interaction by re-routing CD4 molecules from the cell surface to lysosomes (153). Like Vpu, Nef also facilitates the degradation of CD4 attached to the viral Env on the golgi apparatus and down-modulates the expression of MHC class I molecules on the cell surface (135,

153). Hence, major functions of Nef are to promote particle assembly and release and to prevent CTL killing of the infected cells.

The final step in the HIV life cycle is the proteolysis of Gag and Gag-Pol poly-proteins by PR into their individual protein components. The cleavage between MA and CA disassembles the immature lattice and releases CA-SP1. Cleavage of CA from SP1 triggers the CA to assemble into a geometric structure called “the fullerene cone” (58, 90). The mature cell-fee virus can then infect new host cells.

Rev and the RRE

The discovery of Rev and the RRE

The *rev* gene was discovered serendipitously during the genetic delineation studies of *tat*. It dates back to 1986 when two back-to-back papers by Feinberg *et al.* (74) and Sodroski *et al*. (216) reported that the disruption of the neighboring sequences of *tat* in infectious proviral clones resulted in replication defective viruses with severely reduced ability to produce Gag and Env. These sequences were shown to contain a new gene (now known as *rev*) that overlaps *tat* and encodes a 116 aa protein in many B HIV subtypes. To study the underlying mechanism behind the functional effect, both groups performed northern blot analyses of total viral mRNA from transfected cells. The former group reported that in the absence of Rev, the levels of intron containing 9kb and 4kb mRNA species were greatly reduced, whereas the levels of fully spliced 2kb mRNAs were increased. They hypothesized that the protein acted by regulating

splicing and hence named the gene trs (transregulator of splicing). In contrast, the latter group found that the absence of “Rev” did not change the levels of the different mRNA species and postulated that the protein acted post-transcriptionally by supressing *cis*-acting negative regulatory sequences present in the Gag and Env mRNAs. Thus, they named the gene art (antirepression transactivator). To prevent the confusion arising due to different names of the same gene, three years later, the gene was renamed *rev* ((regulator of expression of virion proteins) (88).

Subsequent studies (67, 75, 91, 103, 210) on Rev function looked at the distribution of different HIV mRNA species in subcellular compartments of cells transfected with Rev-responsive subgenomic or proviral reporter plasmids and Rev cDNA expression vectors. These studies used more refined Northern blotting methods or in situ hybridization techniques. Collectively, these studies demonstrated that Rev increases the steady state levels of intron-retaining HIV mRNAs (the 4kb and the 9kb RNA species) in the cytoplasm, by mediating their nucleocytoplasmic export, thus providing a mechanistic explanation for the Rev phenotype observed in the previously mentioned two studies.

The discovery of Rev immediately led to the search for its functional target. Extensive mutational analyses of the *gag*, *pol*, and *env* regions that were absent in the 2 kb mRNAs but were present in the intron-retaining mRNAs resulted in the discovery of the Rev responsive element, called the RRE (Rev response element) in the *env* region (98, 103, 151, 201). Hammarskjold *et al.* (103), using a construct that contained the entire *tat*, *rev*, *vpu*, and *env* genes

between the SV40 late promoter and an intron from the rabbit β globin gene and a polyadenylation signal, showed that the RRE region, indeed, increased the steady state cytoplasmic levels of *env* mRNA. Their northern blot analysis of the *env* mRNAs also showed that the Rev-responsiveness of this element was independent of the position of the RRE but required the RRE to be in the correct orientation (sense-strand orientation).

The mechanisms underlying the role of the Rev-RRE system in increasing the levels of intron-retaining mRNAs were initially debated. Various models were proposed. According to one model, Rev would function by inhibiting splicing or spliceosome assembly (31, 128, 129). Another model proposed that Rev would bind directly to these RNAs and facilitate export independent of splicing (75, 80, 150).

To address these and other issues, Rev function in RNA export was studied in *Xenopus laevis* oocytes. This system was particularly useful as it is very simple to separate the nuclear and cytoplasmic fractions of the oocytes. The experiments were done by co-injecting the oocytes with the purified recombinant Rev protein and the RRE containing pre-mRNA molecules. Consistent with the results from the transfection studies, these studies demonstrated that the intron-retaining RRE mRNAs were seen in the cytoplasmic fraction only in the presence of Rev (80).

In this system, it was also shown by different groups that the Rev-RRE system functioned independent of splicing. Hamm *et al.* (101) demonstrated this using a synthetic RNA capped at its 5’ terminus with the unusual ApppG instead

of the m7GpppN cap, making it a poor substrate for export. This RNA also did not contain any recognizable introns or splice sites and hence would not engage the splicing machinery. They found that the RNA was efficiently exported into the cytoplasm only in the presence of Rev. Fischer *et al.* (80) employed splice-able intron containing pre-mRNA fused to RRE RNA. They found that even in the presence of Rev, spliced mRNA with the fused RRE derived from the pre-mRNA were present in the cytoplasm, suggesting that Rev did not inhibit splicing. While providing some insights into Rev-RRE export, it should be recognized that studies performed in the Xenopus oocyte system may not completely reflect the situation present in mammalian cells.

Indeed studies by Chang and Sharp (31) and Lu *et al*.(143) provided evidence that Rev does not function independently of the splicing machinery in mammalian cells. They showed that interaction between some components of the splicing machinery and the intron containing mRNA is essential for Rev-responsiveness of the intron-retaining mRNA.

Chang and Sharp studied the effect of the intactness of the 5' and the 3' splice sites on Rev-responsiveness of rabbit β-globin construct containing an RRE within an intron and driven by an SV-40 early promoter. They transiently transfected COS cells with the β-globin construct along with a Rev expression plasmid and looked at the nuclear and cytoplasmic distribution of un-spliced β-globin pre-mRNA by nuclease S1 protection and northern blotting. They found that when either the 5' splice site or the 3' splice site were attenuated, the β-globin pre-mRNA accumulated in the nucleus in the absence of Rev. The

nuclear export of the mRNA was restored in the presence of Rev. But when both the 5' splice site and the 3' splice site were attenuated, the β-globin pre-mRNA accumulated in the cytoplasm both in the absence and in the presence of Rev. This study demonstrated that the intact 5' splice site and/or the intact 3' splice site acted as *cis* repressor sequence(s) (CRS) to retain the intron retaining mRNA inside the nucleus. The study also suggested a link between Rev function and the splicing machinery.

Lu *et al.* studied the role of splice sites in the context of an HIV mRNA, i.e. *env* mRNA. Specifically, they used a construct that contained the HIV *env* gene with or without some upstream sequences including the 1st exon of *rev* and *tat*. The *env* gene was placed under the transcriptional control of the SV-40 late promoter. The 3' end of the construct contained an intron and a poly A site from the rabbit β-globin gene. The construct was transiently co-transfected with a Rev expression plasmid into CMT3 cells and the Env protein levels in the cell lysates were determined. It was found that an intact tat/rev 5' splice site was required for Rev-dependent expression of Env protein. Mutation or deletion of the 5' splice site led to a significant reduction or abolishment of Env expression. More importantly, they demonstrated that when the cells were co-transfected with a U1 small nuclear RNA (snRNA) mutant that restored base-pairing at the 5' splice site, the rev-responsiveness of the construct with the mutated 5' splice site was restored. The study showed that the binding of U1 snRNA to the 5' splice site is essential for Rev-dependent expression of Env. Thus, this study provided stronger evidence that interaction or cross-talk between the Rev-RRE

system and the splicing machinery is required for Rev-RRE function.

The structure of the HIV-1 RRE

The RRE is a highly structured *cis*-acting RNA element with multiple stem-loops and bulges. It is located in a highly conserved and functionally important region of *env* at the N-terminus of gp41 (151). At this position, it is present only in the unspliced or the singly spliced RNAs. The minimal size of a functional RRE, also called the short RRE, is a 234-nt (151) region. However, Mann *et al.* (155) have shown that a fully functional RRE, also referred to as the long RRE, is 351-nt long (Figure 5).

The precise secondary structure of the RRE has remained controversial. The early studies on the secondary structure of RRE were done on the subtype B HXB2 RRE. Dayton *et al.* proposed the first secondary structure of an *in-vitro* transcribed and folded RRE using a combination of extensive mutagenesis and computational predictions methods (57). The RRE was shown to form a 5 stem-loop structure in this study. This structure was largely confirmed by Kjems *et al.* by nuclease cleavage and chemical protection assays. In their RRE structure, the 5’ end and the 3’ end of RRE base-pair to form the central base of the RRE - the stem-loop I/I’. The stem-loop I/I’ opens into central loop of the RRE, which gives rise to four other stem-loops, namely the multi-branched stem-loop II, and the stem-loops III, IV, and V. The multi-branched stem-loop II consists of the stem region (IIA) branching out of the central loop and opening into a three-way junction. The junction opens into two different stem-loops namely IIB and IIC.

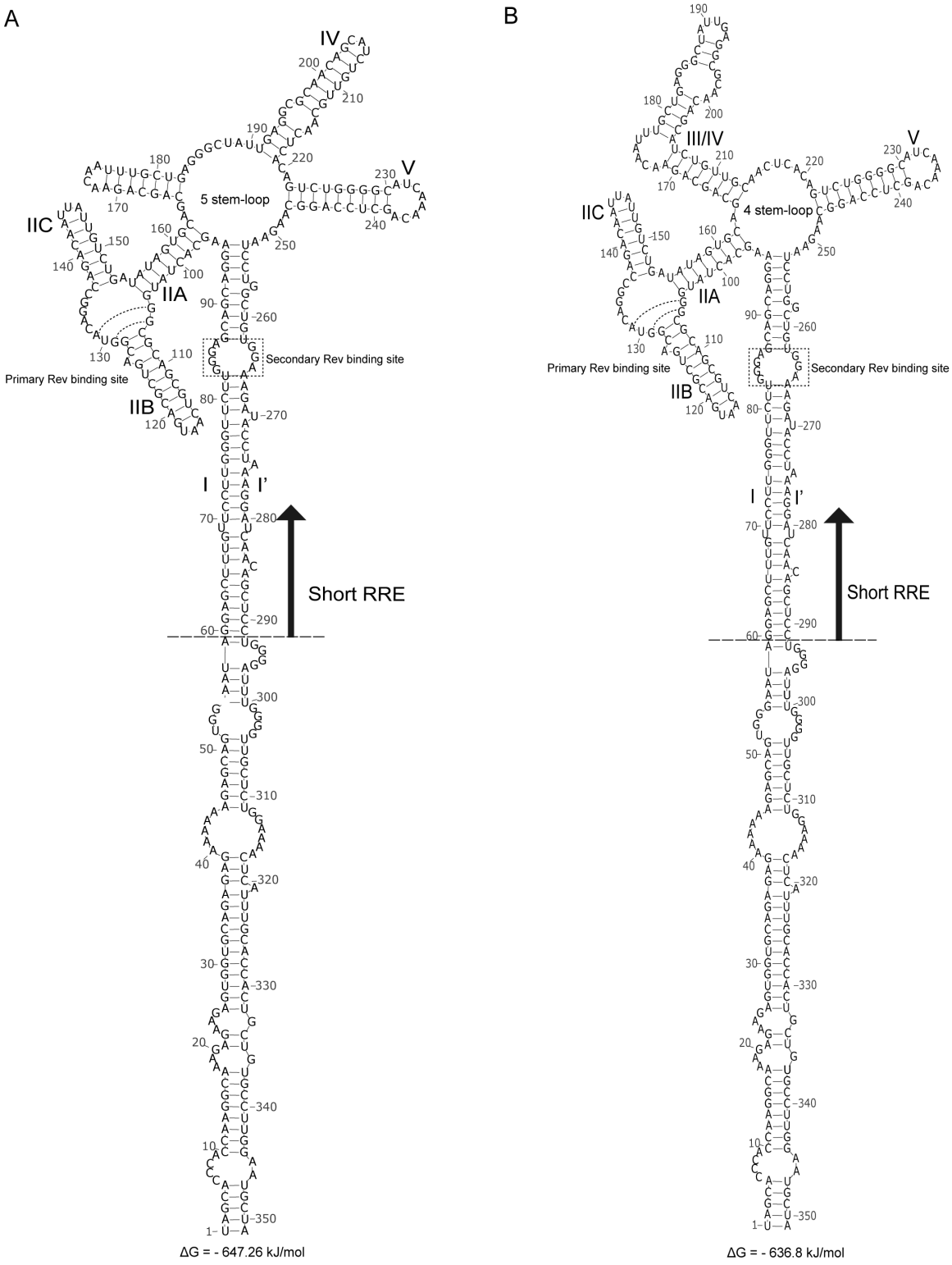

A
IV
V
5 stem-loop
IIC
IIA
IIB
Primary Rev binding site
Secondary Rev binding site
I
I'
Short RRE
ΔG = - 647.26 kJ/mol
B
III/IV
V
4 stem-loop
IIC
IIA
IIB
Primary Rev binding site
Secondary Rev binding site
I
I'
Short RRE
ΔG = - 636.8 kJ/mol

Figure 5

Figure 5: The two secondary structures of the 351-nt NL4-3 RRE.

(A) The 5 stem-loop structure of the long RRE.

(B) The 4 stem-loop structure of the long RRE.

The 351-nt long RRE structure, established by Mann *et al.* (155), is defined by nucleotides 7701-8050 of pNL4-3 (Gen-Bank accession number AF324493). The dashed arcs represent the two non-Watson-Crick base-pairings that allow specific binding of the Rev ARM domain to the primary Rev binding site on the RRE. Secondary structures were generated by RNAstructure 5.5 software.

As NL4-3 increasing become the standard HIV molecular clone studied in labs, subsequent structural studies of the RRE were performed on the NL4-3 RRE. The NL4-3 RRE sequence used in these studies was different from the HXB2 RRE sequence used in the previous two studies only at positions 154 (A➔G), and 236 (A➔G). The secondary structure of the RRE in genomic RNA from purified NL4-3 virions was determined by Watt *et al.* using a more advanced chemical probing technique called SHAPE (Selective 2' hydroxyl acylation analyzed by primer extension) (238). To obtain the purified virions, SupT1 cells were first infected with virus inoculum generated by transfection of 293 T cells with a pNL4-3 proviral plasmid. The virus from the infected cells was then purified by ultracentrifugation of the SupT1 infection supernatant onto a sucrose cushion. Unlike the previous studies, this study probed the RRE structure in the context of full genomic RNA. This is important as other sequences of HIV genome might influence the RRE structure by short-range or long-range interactions. The study reported that the full structure of the RRE was largely that of the long RRE and that the RRE itself exists in the 5 stem-loop structure (Figure 5A).

Interestingly, the secondary structure determination of an *in-vitro* transcribed and folded 232-nt short NL4-3 RRE by the same SHAPE technology gave a different result (137). This study reported that the RRE forms an alternative 4 stem-loop structure, which structurally differs from the 5 stem-loop structure only in the rearrangement of stem-loops III and IV (Figure 5B). In this alternative structure, stem-loops III and IV of the 5 stem-loop RRE combine to form a single stem-loop III/IV that opens into the central loop. Two other studies

supported the 4 stem-loop structure of the NL4-3 RRE (32, 155). In both these studies, the secondary structure of the 351-nt long NL4-3 RRE was investigated using a combination of computational structure prediction, and chemical, and or enzymatic structure probing methods. In one of the two studies, the structure of *in-vitro* transcribed RRE was investigated (155). In the second study, the structure of the RRE was investigated *in-vivo* in yeast cells transformed with the Rev-responsive reporter plasmid containing the RRE and a Rev expression plasmid (32).

Attempts have also been made to look at the tertiary structure of the full-length pNL4-3 RRE. The first attempt was made by Pallesen *et al*., who studied the structure of an in-*vitro* transcribed 370-nt long RRE by high-resolution atomic force microscopy (172). They reported that the tertiary structure of the RRE appeared very similar to its secondary structure in dimensions and shape. It included a head like structure, which they presumed consisted of stem-loops II-V, sitting on top of a stalk like structure presumably consisting of stem-loop I. The length of the whole structure was about 43.4 nm. The stalk was about 22.2 nm long, while the head structure was 21.4 nm long and 17 nm wide. They also showed that their renatured RRE samples migrated as two discrete bands on a native PAGE gel. They speculated that these two bands might represent the 5 stem-loop and the 4 stem-loop structures. However, they reported that even though the RRE samples for the AFM were renatured under the gel-shift assay conditions, the structures of the RRE in the AFM images appeared very

homogenous. They reconciled this discrepancy by concluding that the differences in the two structures were too minute to be resolved by this technique.

A more detailed study on the 3 dimensional structure of the RRE has been recently published (71). In this study, the three dimensional structure of the 233-nt NL4-3 RRE was determined by small angle X-ray scattering (SAXS). According to the study, the RRE forms an “A”-like structure (Figure 6). One of the legs of the “A” is comprised of stem loops II, III, and IV while the other is made of stem loops I and V. By modeling the 4 stem-loop secondary structure of the RRE on the SAXS envelope, the authors deduced a SAXS structure model where the distance between the Rev binding sites on the stem loop IIB (the primary Rev binding site) and on the stem loop I (the secondary Rev binding site) is about ~55 Å. This distance matches the previously shown distance between the two RRE binding motifs in a Rev dimer (52, 62). To further support their result, they showed that increasing the distance between the legs of the “A” by introducing RNA helical turns between them diminished the Rev-RRE activity *ex-vivo* and the Rev-RRE binding affinity *in-vitro*. The study, however, did not try to fit the 5 stem-loop structure of the RRE into their RRE SAXS envelope. Since the past studies support both the 5 stem-loop and the 4 stem-loop RRE structures, a complete and accurate SAXS analysis of the RRE will require the modeling of both the RRE secondary structures on the RRE SAXS envelope.

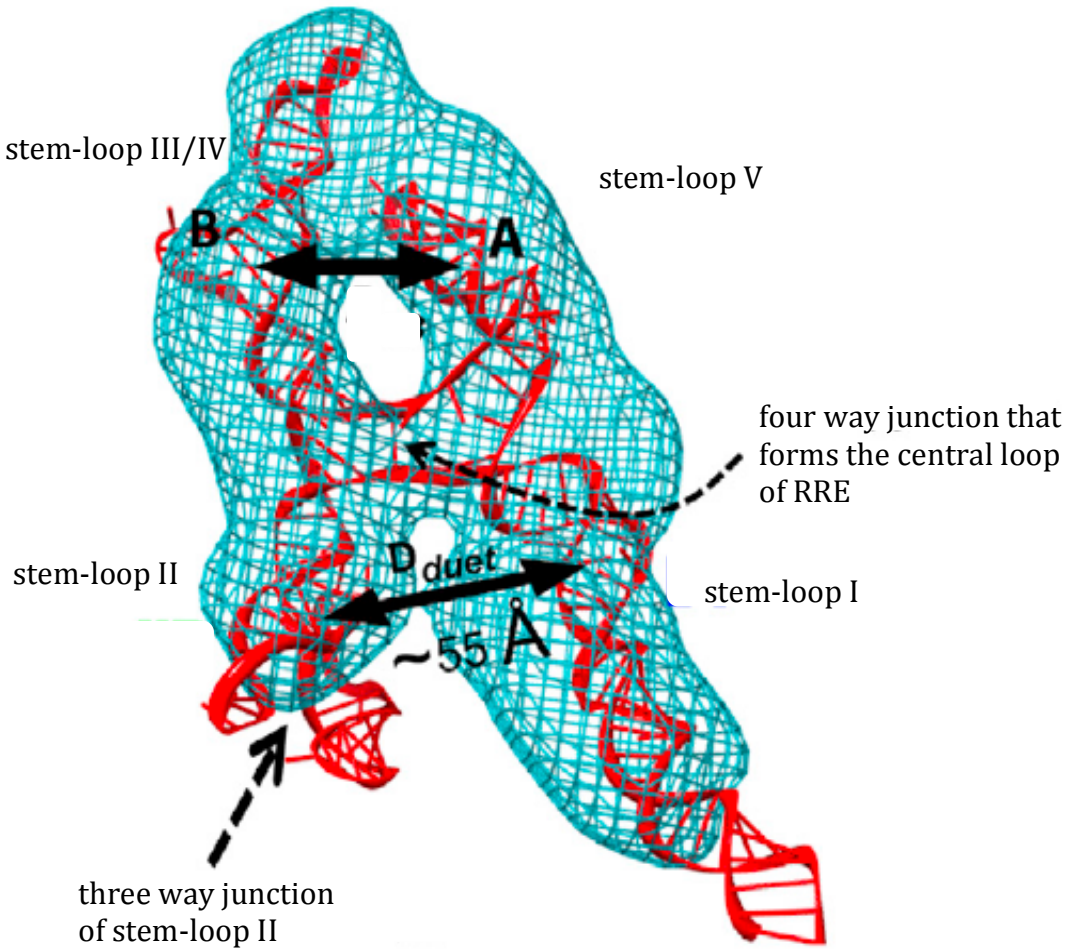


Figure 6: The 4 stem-loop secondary structure of the RRE modeled into the SAXS envelope of 233-nt NL4-3. The tertiary structure of the RRE adopts an "A" shape. One leg of the "A" is contributed by the stem-loops II, III/IV, while the other leg is made of stem-loops V and I. Adapted from (71)

The functions of the RRE

The most widely accepted function of the RRE is to act as a scaffold for Rev binding, facilitating nuclear export of the RRE containing RNA. Within the RRE itself, there are only two functionally defined regions – one within stem-loop IIB and the other within stem-loop I. Stem IIB has been mapped as the primary binding site for Rev binding by several mutagenesis studies (55), in vitro binding and chemical modification interference assays (126). Although there seems to be only one high affinity primary Rev binding site on the RRE, nuclease protection (154), nuclease protection gel retardation (169), and nitrocellulose filter binding (108) experiments have demonstrated that Rev can also bind to other regions of the RRE. Mann *et al.* (154), showed that the progressive truncations of stem loop I of the 351-nt RRE leads to a progressive reduction in the total number of Rev molecules that can bind co-operatively to the RRE and also in the levels of *gag* and *env* mRNA in the cytoplasm. A recent study identified a secondary Rev binding site in stem loop I of the RRE by site directed mutagenesis (50) (Figure 5).

The importance of stem loops III, IV, and V in Rev-RRE function remains controversial. Olsen et al. (169) showed that the deletion of stem loop III significantly reduced the Rev binding ability of the RRE as measured by a nuclease protection gel retardation assay. A Rev responsive CAT expression assay of this mutant showed that the deletion led to 5 times lower Rev-RRE function than with the wild type RRE. By employing an *in-vitro* gel retardation assay and an *ex-vivo* Rev-responsive Gag reporter assay in COS cells, Dayton

et al. (57) showed that the clean individual deletion of each of these stem loops minimally affected the Rev-RRE function in a reporter assay. In this assay, the Rev was supplied in supersaturated levels from a plasmid containing SV40 origin. But in HIV infection, the Rev expression is very low. Therefore, the high levels of Rev used in their assay could have masked the potential importance of these stem-loops. In other studies, mutations predicted to disrupt stem loop III, IV, and V have been found to significantly lower Rev-RRE activity (57, 77). Such disruption mutations might alter the secondary structure of the other regions of the RRE by releasing sequences of these regions that might interact with the neighboring stem-loop regions.

Other mechanisms for the reduced Rev-RRE function of these stem-loop disruption mutations might also exist. Oligonucleotides complementary to the top-most distal loop of stem loop V blocked the Rev-RRE function equally as well as those directed to stem loop II (77). Interestingly, these oligonucleotides did not block the binding of the Rev to the RRE but were capable of completely disrupting preformed Rev-RRE complexes. One possibility is that stem loop V may act as a binding site for a cellular co-factor(s) that may serve to enhance Rev mediated RNA transport or other aspects of Rev function. This site may become available only after binding of Rev to the RRE. In support of this notion, secondary structural analysis of the RRE indicated that stem-loop V became more sensitive to modification with DMS and kethoxal when Rev was bound to the RRE (125).

In addition, previous work from our own group has shown that single

nucleotide changes in the stem loop III-V regions of the RRE can overcome resistance to a trans-dominant negative *rev* allele, Rev M10 (102). Although the mechanism of resistance mediated by single nucleotide changes remains somewhat mysterious, our recent studies have unexpectedly shown that the changes cause a rearrangement of stem loops III, IV and V of the RRE into an alternative stable secondary structure that is more similar to the 5 stem-loop structure than the 4 stem-loop structure (137). Together these studies again highlight the point that regions in the RRE outside of the primary Rev binding site are important for Rev function.

Rev structure and functional domains

The prototype Rev (NL4-3) from HIV clade B is a ~ 18 kd phospho-protein composed of 116 amino acids and is encoded by two exons as 26 and 90 amino acids respectively (14). The first coding exon of Rev overlaps the first coding exon of Tat in a different reading frame. The second exon is encoded in an alternative open reading frame that overlaps the gp41 coding sequence of the *env* gene and the second coding exon of Tat. The Rev protein consists of several well-characterized functional domains as delineated by extensive mutagenesis (Figure 7A). One of these domains (aa 34-50) is a basic, arginine rich motif (ARM) that functions as a nuclear/nucleolar (NLS/NOS) localization signal (14, 37, 132). This domain also serves as an RNA-binding domain (RBD) that binds specifically to the RRE. It is flanked on both sides by the oligomerization domains that are required for Rev multimerization.

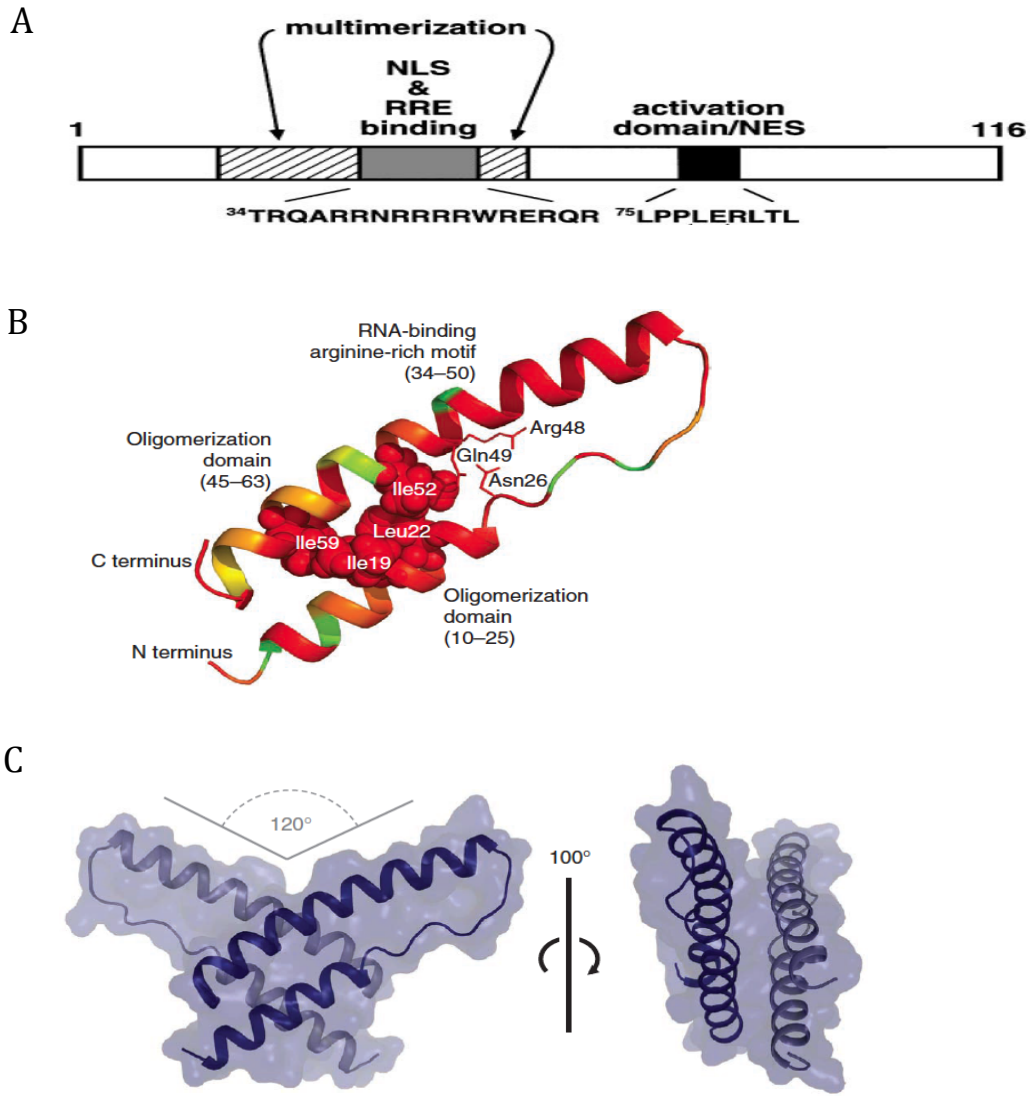


Figure 7: Structure of NL4-3 Rev.

(A) Linear representation of the full length NL4-3 Rev showing the functional domains. Adapted from (84)

(B) Crystal structure of NL4-3 Rev monomer at 2.5 Å resolution. The protein shown is truncated at aa-71 and lacks the last 46 aa including the NES domain. The folded core of a Rev monomer with its functional domains highlighted. The amino acids that mediate the stability of the monomer core structure are shown. The different colors of the ribbon indicate amino acid conservation among 1201 HIV-1 isolates in the Los Alamos Sequence Database, with green representing the least conservation (26%) and red the highest (100%). Adapted from (52)

(C) Two views of a surface representation of the Rev dimer. The two 71-aa Rev monomers are shown as ribbons in light and dark blue. Adapted from (52)

Oligomerization is mediated mainly by Leu18, Leu55, Val16, Leu12, and Leu60 (117). Towards the carboxy terminal is a leucine rich "activation" domain. This domain functions as a specific nuclear export signal (NES) (79, 241). The NLS and NES together enable Rev to actively shuttle between the nucleus/nucleolus and the cytoplasm by accessing cellular pathways for nuclear import and export.

A series of genetic, biochemical, and biophysical studies have contributed to our understanding of Rev structure and some molecular details of the assembly of NL4-3 Rev and the RRE (12, 13, 49, 50, 53, 116, 118, 154, 185, 228). For a long time, the only structural information of Rev came from an NMR study (13) and circular dichroism experiments (225) using a 23 amino acid synthetic polypeptide corresponding to the ARM domain of Rev (residue 34 to 50) bound to the stem IIB of RRE containing the high affinity Rev binding site (13). Both of the studies showed that the ARM sequence of Rev forms an α-helix. The NMR study also reported that the binding of ARM peptide to the stem IIB is accompanied by the formation of G106:G129 and G105:A131 base-pairs (Figure 5) in the internal loop. These two non-Watson-Crick base-pairs widen the major groove in that region of the RRE by 5 Å, allowing the helical ARM to make specific contact with the RRE (12, 13). Unlike in DNA, the major groove of the normal A-form RNA has a deep and narrow major groove and widening of the major groove might be important for specific interaction between the RNA and the protein (239, 240). Specifically, the carboxamide moiety of Asn40 makes hydrogen bond with two purines of the G105:A131 base pair. Mutation of Asn40

leads to a ~40 fold reduction in the binding affinity between ARM and stem IIB (225).

NMR analysis of the contact surface between the ARM peptide and the secondary Rev binding site on stem-loop I of the RRE (51) was also performed. The authors (51) found that, to make contact with this RRE site, Rev uses an ARM interface (Arg38, Arg41, and Arg46) that is different from that used to make contact with the high affinity stem IIB region. When each of these residues was mutated, the interaction between the secondary Rev binding site and ARM peptide was disrupted but the same mutations did not affect stem IIB binding. Conversely, mutation of Asn40 did not affect stem-loop I binding (51).

A crystal structure of Rev was lacking for a long time due to the propensity of purified Rev to oligomerize into filaments or even form insoluble aggregates, which made it difficult to crystallize (39, 104, 244). This self association property of Rev has been attributed to hydrophobic interactions (117) and to the basic nature of the protein. To obtain Rev crystals, Daugherty *et al.* (52) partly disrupted the hydrophobic interactions that mediate higher order Rev oligomerization by mutating Rev residues Leu12 to Ser and Leu 60 to Lys. In addition, they also removed the last 46 residues, including the Leu-rich activation domain from the C' terminus of Rev. To neutralize the basic charge of the protein, they purified Rev as a fusion protein with the negatively charged B1 domain of streptococcal protein G (GB1). At the final purification step, Rev protein was cleaved off the GB1 and released into solutions of 100mM potassium phosphate or sodium sulfate. These solutions further provided oxyanions to

maintain Rev solubility after GB1 cleavage.

The structure of the crystal so obtained was resolved at 2.5 Å resolution (Figure 7B and 7C). The disruption of the higher order oligomerization interface was not very successful as the unit crystal cell contained two Rev dimers. The crystal structure revealed that the N-terminal 1-70 amino acid residues of the Rev monomer adopts a planar antiparallel helix–loop-helix structure (Figure 7B). The N-terminal arm of this structure consists of a shorter helix that includes one of the two oligomerization domains followed by a loop region. The C- terminal arm consists of a longer helix made up of the ARM at its N-terminal half and the second oligomerization domain at its C-terminal half. The two oligomerization domains of a monomer face each other. The apposed surfaces of oligomerization domains are rich in hydrophobic residues (Leu22, Ile52, and Ile59) which along with some polar residues (Asn26, Arg48, and Gln49) provide stability to the monomer structure. Some of these residues also participate in dimer formation with other Rev monomers. The Rev dimer is held together mainly by hydrophobic interactions mediated by Leu18, Phe21, Ile55, but as a consequence of this, dimerization residues Leu22, and Ile59, which are exposed in the monomer get buried. The burial of these residues has been reported to stabilize the dimeric form by >22 $kcalmol^{-1}$ (230). The authors use this observation to explain cooperative binding and propose that the building block of the Rev-RRE complex is a Rev dimer. In a Rev dimer, the two Rev monomers are positioned face to face by their dimerization interface at an angle of ~ 120^{o} to form a V shaped Rev dimer (Figure 7C). In this V shaped dimer, the ARM domain of each monomer is

free to grasp the RNA. Each Rev monomer of the dimer interacts with Rev monomer of the adjacent dimer by its higher order oligomerization interface. This interface is stabilized by hydrophobic interactions mediated by Leu12, Val16, and Leu60.

The group also attempted to generate a model of the RRE bound to the Rev dimer (Figure 8). To achieve this, they placed the IIB RNA hairpin from the NMR structure of the ARM peptide–IIB34 RRE (13) into the Rev dimer crystal structure. The RNA was extended by 4-nt to accommodate the second Rev monomer of the dimer. Consistent with a previous study from the same group (51), in the model, the two ARM domains of the dimer bind the RNA using two different interfaces, namely the inner face and the outer face. The ARM of one of the monomers contacts IIB using Asn40 on the “inner face” of the helix. The ARM of the second monomer binds to the other end of the extended IIB by its the 'outer face'. The distance between the two ARM domains in the dimer was found to be 55 Å. As previously mentioned, in the SAXS structure of the RRE too (Figure 6), the distance between the primary Rev binding region on stem-loop II and the secondary Rev binding region on stem-loop I was found to be 55 Å.

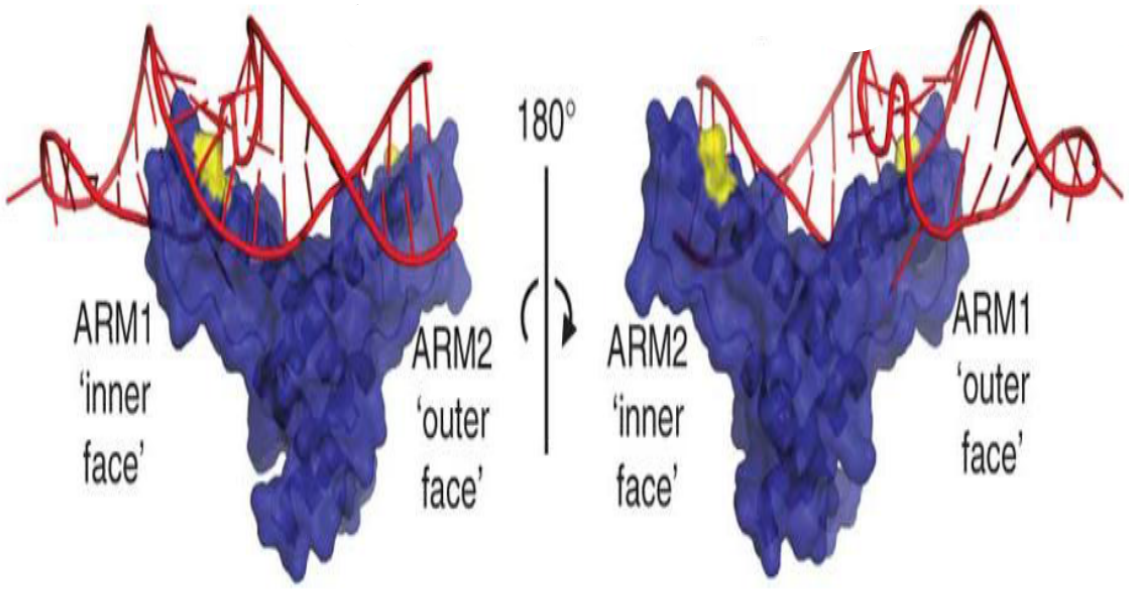


Figure 8: Stem loop IIB (extended by 4-nt) modeled on the Rev dimer. In this model, the first Rev monomer binds to the high affinity Rev binding site through Asn40 (yellow residue) located in the inner surface of the ARM domain (ARM1). The second monomer binds the RNA using a different surface (outer face) of the ARM domain (ARM2). Adapted from (52)

The crystal structure of Rev was also solved by an independent group (62), using Rev linked to a Fab fragment at 3.2 Å resolution. The Fab fragment was specially engineered to prevent Rev aggregation. Each crystal unit cell contained six Rev monomers and six Fab fragments. The structure and the orientation of the Rev monomer in each Rev dimer were similar to those reported by Daugherty *et al.* (52). Interestingly, in this study, the residues that were reported to mediate higher order oligomerization by Daugherty *et al.* (52) mediated Rev dimerization. Conversely, the residues that mediated Rev dimerization in Daugherty *et al.* (52) crystal structure were actually occupied by the Fab fragment.

According to the Rev-RRE assembly model proposed by Daugherty *et al*., the first event in the Rev-RRE assembly is the specific binding of the N40 amino acid of the first Rev monomer to the primary high affinity Rev binding site of stem IIB of RRE. This exposes the dimerization interface of that Rev, which allows the binding by a second Rev monomer on a secondary Rev binding site on RRE. The angular orientation of the Rev monomer in a dimer necessitates that the second Rev binding site on the RRE be correctly oriented and spaced relative to the stem IIB site. As more Rev monomers are cooperatively added to the RRE the affinity of Rev-RRE complex increases. Up to 8 Rev monomers can bind the ~350-nt RRE. The hexameric Rev homooligomer-RRE RNP has 500 times higher affinity than monomer Rev- RRE complex (50) and can bind two Crm1 molecules through its NES sequences. Therefore, oligomerization of Rev may

serve at least two purposes: increasing specificity for RRE binding and facilitating Crm1 binding.

Cellular factors interacting with the Rev and the RRE

Crm1 or Exportin-1 is a member of the β-importin family of transport receptors and has been demonstrated to specifically interact with many proteins containing NES domains (157). Normally employed for the nucleo-cytoplasmic export of proteins, snRNAs, and rRNAs, the Crm1 pathway is hijacked by HIV to facilitate export of intron-retaining HIV RNAs. Crm1 bridges Rev and Ran-GTP interaction to form a functional export complex (9, 43, 133). It also bridges the interaction between Rev and nucleoporins like Rip/Rab (166) allowing the Rev-RRE ribonucleoprotein (RNP) to ferry across the nucleopore complex into the cytoplasm. Here, the RNP disassembles with the hydrolysis of GTP bound to Ran, releasing the RRE containing RNAs into the cytoplasm. The Rev-RRE-Crm1 pathway reportedly also requires the binding of a helicase DDX3 (247) to Crm1, although the exact function of DDX3 in the pathway has not been elucidated. Various other cellular factors like eIF-5A, SF2/ASF, B23, p32, Sam68, RHA, and DDX1 have also been implicated to modulate Rev-RRE function, though their roles remain either unclear or controversial (184, 221).

Rev and RRE variation in infected patients

In an infected individual, CTL (cytotoxic T lymphocyte) pressure is one of the major factors driving variation of the virus. A Rev specific CTL response has

been observed in patients' PBMCs (peripheral blood mononuclear cells) tested with overlapping synthetic Rev peptides' epitopes (3, 4). Baalen *et al.* (231) demonstrated an inverse correlation between Rev-specific CTL response and rapid progression to AIDS. They also found that in patients with identical HLA alleles, but with significantly different Rev specific CTL frequency, there was sequence variation in the Rev epitopes, suggesting the ability of different Rev epitopes to elicit different degrees of CTL recognition. The same group also reported the evolution of Rev in cell culture due to CTL pressure (232).

Another study has shown the ability of different Rev alleles to indirectly modulate immune evasion. Bobbitt *et al.* (19) found that in infected patients with well functioning immune systems, infected cells were more resistant to anti-Gag and anti-Env CTL killings compared to infected cells from late stage AIDS patients. The decrease in susceptibility to CTL killing was attributed to lower expression levels of Gag and Env resulting from attenuated Rev alleles suggesting that loss of CTL function as the infection progresses might select for more active Rev alleles. Unlike mechanisms that affect overall transcription and translation, down-modulation of gene expression by less active Rev alleles primarily affects viral structural protein expression and is not expected to affect HIV Nef expression. Thus this mechanism would allow the immune evasion activities of Nef to still persist, while the lower structural protein levels would make the infected cells less susceptible to CTL killing, Other studies have also reported the contribution of less-active *rev* alleles to long-term survival of HIV-1 infection in some patients (110, 113).

Another important factor driving evolution of HIV is anti-HIV drug pressure. Inhibitors of Rev and the RRE have not been approved for clinical use to date. However, because of the overlap of the RRE and the 2nd exon of Rev with gp41 of *env* gene, it is conceivable that HIV drugs targeting gp41 of *env* will also select for mutations in the Rev and the RRE sequence. In fact, ENF (enfutivirtide or T20), the first fusion inhibitor to be used for HIV treatment, acts by binding to a region on gp41 and has been reported to select for secondary mutations in Rev and the RRE. The primary mutations reported to be associated with ENF resistance are located within the ENF target region and map to 36-45 amino acids of gp41 (61), which lies within the RRE. Secondary mutations in the RRE were found to restore the RRE structure predicted to be disrupted by the primary mutations. Such “structure conservation mutations” were observed in stem loop IIC (163) and III (223), which underscores the importance of structural integrity of regions of the RRE outside the primary Rev binding site. Secondary mutations in Rev were also found in both ARM and oligomerization domains (223). Those in the ARM domain accompanied mutation in the primary Rev binding site of RRE, suggesting that these Rev mutations were selected to correct impaired Rev-RRE binding. Furthermore, mutations observed in the oligomerization domain of Rev were predicted to regain the Rev-RRE activity level altered by RRE mutation. The authors, however, presented no experimental data to test the predicted functional significance of the secondary mutations. Nevertheless, the studies point to the importance of taking into account Rev-RRE interaction while

designing such drugs and more importantly the need of evolutionary conservation of Rev-RRE function by the virus.

Rev and RRE variations are also seen among HIV subtypes. Consensus Rev sequences across different clades are considerably different especially at regions outside the NLS/ARM and NES domains of Rev. Most clade C isolates lack 15 amino acids at the C-terminus of Rev relative to clade B and there exists variable deletions in all of the clades. Comparison of 894 HIV-1 RRE sequences of different clades available from Los Alamas HIV database with HXB2 RRE revealed sequence conservation of almost all bases in stem IIA, the high affinity Rev binding site, and most of the stem region of stem-loop loop III. The most conserved region of the RRE within and across HIV clades appears to be the high affinity binding site of RRE (179, 212). The high sequence specificity of this region has also been demonstrated by site directed mutagenesis (108). Secondary structure conservation was observed in stem IIA, stem IIB, stem III, and distal parts of stem IV and V (179).

There are just a few studies on the Rev and the RRE sequence variation in a single patient and even fewer studies on Rev and RRE functional variation. One of these studies was done by Phuphuakrat *et al.* (178), who studied 10 RRE variants from Thailand and observed differential Rev-RRE activity when patients' RREs were paired with NL4-3 Rev. Further functional studies of these Rev-RRE pairs showed that the differential Rev-RRE activity resulted from both differential RNA export and translation enhancement efficiency of these pairs. RRE functional variation has also been studied in a longitudinal cohort (182). The

study presented a positive correlation between rates of CD4+ decline and RRE activity at the late time points while such correlation was absent at early time points. These data suggest that RRE heterogeneity may be important in viral pathogenesis and disease progression. A major caveat in these functional studies was that Rev and the RRE were paired with heterologous control Rev or control RRE rather than their cognate partners.

A recent study from our lab has shown a marked functional difference between autologous and heterologous Rev-RRE pairs (214). In this study, multiple autologous/cognate Rev and RRE sequences were obtained by single genome sequencing of blood plasma samples. These plasma samples were collected within 6 months after seroconversion and at a later time. The Rev-RRE activity of the Rev-RRE pairs was tested in a transient transfection assay using a CMV driven GagPol reporter construct. Each patient Rev-NL4-3 RRE pair and NL4-3 Rev-patient RRE pair were also tested for their Rev-RRE activity using the same system. In general, it was observed that different levels of Rev-RRE activity were obtained for each of these cognate or non-cognate pairs of Rev-RRE, suggesting that each pair served as a single functional unit.

Furthermore, it was also observed that activity of the cognate pairs from the same time point in a patient clustered around a single activity level, which was termed the set point. The set point changed significantly in 3/5 patients over the time period studied. These changes in set point were mainly driven by small changes in the RRE sequence. For instance, patient SC3, who showed a progressive decline in CD4 counts and a rapid increase in blood viral load, also

exhibited a significant increase in the Rev-RRE activity at a later time-point. This increase in activity was entirely driven by the RRE as the activity of Rev in this patient did not change over the course of study. In fact, in many virions isolated at the late time-point, the Rev sequence remained unchanged. Interestingly, there were only four nucleotide changes in the late time-point RRE compared to the early time-point RRE and all of these changes were in regions outside the primary Rev binding site. It was also shown using a gel shift assay that the late time-point RRE formed higher order complexes with Rev at lower Rev concentration compared to the early time-point RRE. Therefore, the study demonstrated that the Rev and the RRE appeared to co-evolve to attain a range of activity. Such variation in activity might regulate viral replication to allow the virus to adapt to different selection pressures within the host.

Rationale

HIV transcribes various intron-retaining mRNAs that make up the progeny viral genome and encode most of the viral proteins. The nucleo-cytoplasmic export of these mRNAs is essential for viral replication. This export is initiated by the binding of HIV Rev protein to the RRE, a *cis*-acting RNA element found in a highly conserved region of *env* in these mRNAs (67, 75, 151, 251). Functionally, the RRE has been modeled as a scaffold for binding Rev molecules (48). The primary Rev binding site on the RRE has been mapped to stem IIB (41, 56, 125), and stem I/I' has been shown to provide secondary Rev binding sites (51, 155). Upto 8 Rev monomers bind co-operatively to the whole RRE (45, 46, 155)

forming the Rev-RRE oligomeric complex. The complex recruits the Crm1/Ran-GTP (7, 10) nuclear export machinery resulting in the RNA export.

Despite our knowledge of Rev-RRE mediated RNA export, the precise secondary structure of the RRE has remained controversial. Various computational modeling, chemical and enzymatic probing methods, alone or in combination, have been used to investigate the secondary structure of the RRE. Some of these studies (57, 125, 238) have shown that the RRE forms a 5 stem-loop structure (Figure 5A). Other studies (32, 137, 155) have reported that the RRE forms an alternative 4 stem-loop structure (Figure 5B), which structurally differs from the 5 stem-loop structure only in the rearrangement of stem-loops III and IV. In the 5 stem-loop structure, the stem-loop III and IV form separate stem-loop structure whereas in the alternative structure, stem-loops III and IV combine to form a single stem-loop III/IV.

Although there is evidence (77, 137, 224) that stem loops III, IV, V of the RRE contribute to Rev-RRE function, the precise functional role of these regions is not well understood. In this thesis, we have re-examined the secondary structure of the RRE and compared the functional importance of the 5 stem-loop and the 4 stem-loop RRE structures.

CHAPTER II - MATERIALS AND METHODS

Numbering of the RRE

Both the short and the long versions of the RRE were tested in the study. In the Rev-RRE electrophoretic mobility shift assays, the activities of a short version (234-nt) of the RREs were compared. The coordinates of these short RREs map to the nucleotides 59-292 on the long RRE (Figure 5). The 234-nt WT RRE sequence was originally defined by Malim *et al.* (151) as the minimal length functional RRE. In the SHAPE and the RRE gel migration experiments, the 232-nt RREs, which map to nucleotides 60-291 on the long RRE, were tested. In all the cell-culture assays, the long RRE was studied. The WT long RRE is the 351-nt long RRE established by Mann *et al.* (155) and is defined by nucleotides 7701-8050 of pNL4-3 (Gen-Bank accession no. AF324493).

Plasmid constructs reference number

All the plasmid constructs used in the study are referenced by their Hammarskjöld-Rekosh laboratory archive number with a prefix pHR (plasmid Hammarskjöld-Rekosh).

Description of the plasmid constructs

pNL4-3 (pHR 1145): It is an infectious proviral clone of HIV-1 with pUC18 backbone. As mentioned in the introduction, its HIV sequence is a chimera of 5.8Kb long 5' region of NY5 HIV-1 isolate and the 3.8 Kb long 3' region of the LAV HIV-1 isolate joined at the EcoRI site unique to both the isolates (2).

Constructs used in the hygromycin resistance assay

1. pNL4-3 Nef- (pHR 1272): It is a Nef defective form of pNL4-3. It has pNL4-3 sequence except for the 97 nucleotides (nt 8789-8886) deletion in the *nef* gene that places the unique XhoI site of pNL4-3 immediately downstream of the start codon of Nef.

2. pTR 167 (pHR 1266): This construct has been previously described (195). It was created by deleting the entire region (5491-bp) between the two NsiI sites located between nucleotides 1251 to 6742 of pNL4-3 (HIV NL4-3[GenBank accession number M19921]). The deleted region includes part of *gag*, entire *pol*, and part of *env* genes. A cassette containing the hygromycin resistant gene (Hyg^r, Hygromycin B phosphotransferase) driven by SV40 early promoter – enhancer was inserted into the NheI site in the remaining *env* sequence at nucleotide 7520. It makes three species of mRNAs - 1. the Rev-RRE dependent genomic RNA which contains the SV40- Hyg^r gene cassette, 2. the Rev-RRE independent SV40 driven subgenomic Hyg^r mRNA and 3. the Rev-RRE independent spliced mRNA that makes Nef.

3. pTR 167 Nef- w/ WT RRE in its native position (pHR 4494): It is a derivative of pHR 1266 with the 97 nucleotide deletion in the *nef* gene. It was created by replacing the MfeI to XhoI fragment of pHR 4910 (see below) with that of pHR 1272.

All the pTR 167 Nef- w/ RRE constructs (described in number 3 and number 4) make only three species of mRNAs - 1. the Rev-RRE dependent genomic RNA which contains the SV40- Hyg^r gene cassette, 2. the Rev-RRE independent SV40 driven subgenomic Hyg^r mRNA, and 3. the Rev-RRE independent mRNA that encodes a non-functional N-terminally deleted Nef protein.

4. pTR 167 Nef- w/ mutant RRE: This series of plasmids are derivatives of pTR 167. Each plasmid is Nef negative and contains a different mutant RRE in the native RRE location. The constructs with the mutant A RRE, mutant B RRE, mutant C RRE, mutant D RRE, and mutant E RRE were referenced as pHR 4910 (pTR 167 Nef- w/ mutant A RRE), pHR 5054 (pTR 167 Nef- w/ mutant B RRE), pHR 4406 (pTR 167 Nef- w/ mutant C RRE), pHR 4438 (pTR 167 Nef- w/ mutant D RRE), pHR 4408 (pTR 167 Nef- w/ mutant E RRE) respectively. To create these constructs, different mutant RREs were first created as a portion of MfeI and XhoI fragment of pHR 4494 by SOE (Splicing by overlap extension)-PCR. Briefly, three different ds-DNA fragments (the 5' fragment with a MfeI site at its 5' end, the middle 2^{nd} fragment containing the desired RRE mutations, and the 3' fragment with a XhoI site at its 3' end) were first PCR generated. All three fragments were then joined at the overlapping regions to create a 1351-bps long fragment with a MfeI site at its 5' end and a XhoI site at its 3' end. This was done in two steps. First, the 5' fragment and the middle fragment were joined to create

a 455-bps fragment. Second, the 455-bps first SOE product was joined to the 3' fragment to create the final 1351-bps final SOE product.

To generate the MfeI-XhoI fragment containing the mutant B RRE, the 5' fragment was generated using pHR 1272 as the template and oligo-pair 2087 and 2567 as the primers. The middle 109-bp fragment was created using oligo 2567 as the template and oligo-pair 2568 and 2091 as the primers. For the MfeI-XhoI fragment containing the other mutant RREs, template pHR 1272 and primer-pair 2087 and 2089 were used to generate the 5' fragments. The 71-bps middle fragments were generated using primer-pair 2088 and 2091. The template for the middle fragment was oligo 2437 for mutant A, oligo 2093 for mutant C, oligo 2095 for mutant D, and oligo 2094 for mutant E. The 3' fragment of all the MfeI-XhoI fragments was created using pHR 1272 as the template and oligo-pair 2090 and 2092 as the primers.

Sequence information of oligos used to create the SOE products. Mutated nuclotides are shown in bold and larger font.

Oligo# 2437: Oligo that introduces mutations on the both sides of top stem of stem-loop IV of the NL4-3 RRE to create mutant A RRE (71-nt)

5' gaacaatttgctgagggctattgaggc**agcc**cagcatctg**ggct**aactcacagtctggggcatcaaaca
gc

Oligo# 2567: Oligo that introduces mutations in the base stem of combined stemloop III/IV to create mutant B (109-nt long)

5’ caggccagacaattattgtctgatatagtgca**cgttgtc**aacaatttgctgagggctattgaggcgcaacag cat**gacgacg**aactcacagtctggggcatcaaacagc

Oligo# 2093: Oligo that introduces mutations on theleft side of the top stem of stem loop IV of the NL4-3 RRE to create mutant C RRE (71-nt)

5’ gaacaatttgctgagggctattgaggc**agcc**cagcatctgttgcaactcacagtctggggcatcaaaca gc

Oligo# 2095: Oligo that introduces mutations on both sides of the top stem of stem loop IV of the NL4-3 RRE to create mutant D RRE (71-nt)

5’ gaacaatttgctgagggctattgaggc**agcc**cagcatctg**aacg**aactcacagtctggggcatcaaaca gc

Oligo# 2094: Oligo that introduces mutations on the right side of the top stem of stem loop IV of the NL4-3 RRE to create mutant E RRE (71-nt)

5’ gaacaatttgctgagggctattgaggcgcaacagcatctg**aacg**aactcacagtctggggcatcaaaca gc

Oligo# 2088: Oligo with the 5’ end sequence of 71-nt mutant NL4-3 RRE oligo

5’ gaacaatttgctgagggctattg

Oligo# 2089: Oligo with sequence reverse complimentary to the 5' end sequence of 71-nt NL4-3 RRE mutant oligo

5' caatagccctcagcaaattgttc

Oligo# 2090: Oligo with the 3' end sequence of 71-nt mutant NL4-3 RRE oligo

5' cagtctggggcatcaaacagc

Oligo# 2091: Oligo with sequence reverse complimentary to the 3' end sequence of 71-nt NL4-3 RRE mutant oligo

5' gctgtttgatgccccagactg

Oligo# 2568: Oligo with the 5' end sequence of the 109-nt mutant NL4-3 RRE oligo (oligo# 2567)

5' caggccagacaattattgtctg

Oligo# 2569: Oligo with sequence reverse complimentary to the 5' end sequence of the109-nt mutant NL4-3 RRE oligo (oligo# 2567)

5' cagacaataattgtctggcctg

Oligo# 2087: Forward oligo that anneals upstream of MunI (MfeI) and BsaBI sites in NL4-3 *env*

5' aaacatgtggcaggaagtagg

Oligo# 2092: Reverse oligo that anneals downstream of XhoI site in NL4-3 *nef*

5’ cacaagtagcaatacagcagc

5. pCMVΔR9 (pHR 2003): The plasmid, pCMVΔR9 (162) contains the human cytomegalovirus (hCMV) immediate early promoter, which drives the expression of all viral proteins except the viral envelope. Some of the *cis*-acting sequences have been removed from the plasmid, which renders the transcripts derived from the plasmid defective in packaging, reverse transcription, and integration.

6. pCMV_VSV-G (pHR 2004): The pCMV_VSV-G (162) plasmid encodes the G glycoprotein of vesicular stomatitis virus (VSV G) under the control of hCMV immediate early promoter. The plasmid supplies the heterologous envelope protein for pseudo-typing the particles produced from pCMVΔR9. The VSV-G pseudo-typing enables the viral particles to exhibit broad cell tropism. VSV-G has robust and pantropic infectivity because the cell receptors for VSV-G are highly ubiquitous low-density lipoprotein (LDL) receptors (78).

Constructs used in replication kinetic assays and competition assays

1. pNL4-3 RRE- (pHR 1284) is a version of pNL4-3 with a non-functional RRE generated by multiple third base mutations with the RRE that do not change the coding sequence of the overlapping *env* (164).

2. pNL4-3 RRE- Nef- w/polylinker (pHR 2772) is a derivative of pHR 1284 with the 97 nt between the start codon of Nef and the unique XhoI site replaced with a polylinker. The polylinker includes MluI, SmaI, and XhoI restriction sites (5'➔3').

3. pNL4-3 Nef- RRE- /long RRE: These plasmids were constructed using the larger MluI and XhoI double digest fragment of pHR 2772 as the vector backbone. The long 351-nt WT/mutant NL4-3 RRE along with 24-nt of adjacent viral sequence upstream of the RRE and 17-nt adjacent viral sequence downstream of the RRE was PCR amplified. This 393-nt RRE region corresponds to nt 7676 to 8069 in pHR 1145 and pHR 1272. The respective pTR 167 Nef- w/ WT or mutant RRE plasmid served as the PCR template and oligo-pair 2249 and 2250 served as the primers. The oligos were designed to introduce a MluI site and XhoI site at the 5' end and the 3' end of the PCR product respectively. The PCR products were double digested with MluI and XhoI and ligated into the MluI and XhoI ends of pHR2772 to obtain the final plasmid construct. This version of the plasmids with WT RRE, mutant A, mutant B, mutant C, mutant D, and mutant E were archived as pHR 4874, pHR 4884, pHR 5121, pHR 4876, pHR 4880, and pHR 4878 respectively.

Sequence information of oligos used to create pNL4-3 Nef- RRE- /long RRE constructs:

Oligo# 2249: Forward oligo that anneals to the NL4-3 sequence upstream of the 351-nt RRE. It has a MluI site at it 5' end.

5’ gacacgcgtgtagtaaaaattg

Oligo# 2250: Reverse oligo that anneals to the NL4-3 sequence downstream of the 351-nt RRE. It has a XhoI site at it 5’ end.

5’ gacctcgagagatttattac

Constructs used in the Rev-dose response assay using the Gag-Pol reporter construct

1. pCMV (pHR 16): This plasmid contains the promoter-enhancer from the simian cytomegalovirus (sCMV) IE94 gene (nt -650 to +30, relative to the transcription start site). The region is immediately followed by a spliceable intron and a polyadenylation signal derived from the rabbit β-globin gene (138, 217). Because of the lack of any ORF between the promoter-enhancer and beta-globin gene region, this plasmid is also called empty pCMV plasmid.

All the other constructs used in this assay are derivatives of pHR 16 and contain extra sequence in between the sCMV promoter-enhancer and the rabbit beta-globin gene region.

2. pCMV_SEAP (pHR 1831): This plasmid contains the gene that encodes a secreted form of human placental alkaline phosphatase called SEAP in between the sCMV promoter and β-globin gene. SEAP is not produced from eukaryotic cells under normal conditions. Transient transfection of an SV-40 version of this plasmid into COS cells yields SEAP expression that is qualitatively and quantitatively similar both at the mRNA and at the protein levels (15).

3. pCMV_Rev (pHR 3821): This construct contains a codon-optimized NL4-3 Rev cDNA in between the sCMV promoter and the β-globin gene. This Rev cDNA was synthesized by Genescript company.

4. pCMV_GagPol_U2 (pHR 2739): Between the sCMV promoter-enhancer and rabbit beta-globin gene region, it contains *gag, pol, and vif* ORF from HIV-1 BH10 clone followed by a XhoI-BamHI fragment containing human U2 snRNA specific A' protein gene. The XhoI-BamHI (5'➔3') fragment is also called the stuffer fragment as this fragment is usually replaced by a specific sequence, whose effect on Gag and Pol production needs to be determined.

5. pCMV_GagPol_long RRE: This series of constructs were used as the reporter constructs in the assay. They were directly derived from pHR 2739 by swapping the stuffer fragment with PCR generated and Xho-BamHI double digested long RRE region. The long WT/mutant RRE region tested in this assay were the same 393-nt long WT/mutant RRE regions tested in the spreading infections (viral growth assays). The RRE regions was amplified using the respective pTR 167 Nef- w/ WTor mutant RRE plasmid as template and the oligo-pair 2246 and 2247 as PCR primers. The oligos were designed to introduce an XhoI site and BamHI site at the 5' end and the 3' end of the 393-nt RRE respectively. The pCMV_GagPol_long RRE constructs containing the WT RRE, mutant A, mutant B, mutant C, mutant D, and mutant E were archived as pHR 4824, pHR 4908, pHR 5119, pHR 4808, pHR 4812, and pHR 4810 respectively.

Sequence information of oligos used to create pCMV_GagPol_long RRE constructs:

Oligo# 2246: Forward oligo that anneals to the NL4-3 sequence upstream of the 351-nt RRE. It has a XhoI site at it 5' end.

5' ggcctcgaggtagtaaaaattg

Oligo# 2247: Reverse oligo that anneals to the NL4-3 sequence downstream of the 351-nt RRE. It has a BamHI site at it 5' end.

5' ggcggatccagatttattac

Constructs used in the Rev-RRE gel-shift assay

1. pCR 2.1 TOPO TA _ short RRE: These constructs were created by inserting the 234-nt short RRE containing the T7 promoter sequence at it 5'end into pCR 2.1 TOPO TA vectors (Invitrogen) by TOPO TA cloning. The short RRE sequence spans from nt-7759 to nt-7992 in pHR 1145. The RRE insert was PCR amplified using a primer set (oligo 2363 and oligo 2367) designed to introduce T7 promoter sequence at the 5' end of the RREs. The PCR products were cloned into the EcoRI site of pCR 2.1 TOPO TA vector. The pCR 2.1 TOPO TA _ short RRE construct containing the WT RRE, mutant A, mutant B, mutant C, mutant D, and mutant E were archived as pHR 4854, pHR 5083, pHR 5091, pHR 5081, pHR 5089, and pHR 5085 respectively.

Sequence information of oligos used to create pCR 2.1 TOPO TA _ short RRE constructs:

Oligo# 2363: Forward oligo that anneals to the 5' end of the 234-nt NL4-3 RRE. It has a T7 promoter sequence (shown in bold) at it 5' end.

5' agcgtac**ttaatacgactcactataggg**aggagctttgttccttg

Oligo# 2367: Reverse oligo that anneals to the 3' end of the 234-nt NL4-3 RRE.

5' aggagctgttgatcctttagg

TOPO-TA cloning

TOPO-TA cloning was done by TOPO-TA cloning kit containing pCR^{TM}2.1-TOPO vector (Invitrogen) to create the RRE constructs used in the Rev-RRE gel shift assay. The vector provided in the kit has Topoisomerase I from Vaccinia covalently bound to it. The Topoisomerase binds to duplex DNA of the vector at specific sites and cleaves the phosphodiester backbone after 5'CCCTT in one strand. This sequence is present at the region in the vector where the PCR product will be inserted. The energy from the broken phosphodiester backbone is conserved by the formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue of the topoisomerase I. Therefore, the vector provided in the kit is a linearized vector with single overhanging 3' deoxythymidine (T) residues. When PCR product with an overhanging A is added to the vector, the PCR product ligates efficiently with the vector via TA base-pairing.

The first step in the cloning was to generate a Taq polymerase amplified-PCR product. The PCR product was purified using Qiagen PCR purification kit. The PCR product was used in the TOPO-TA reaction the same day to avoid falling off of the TA overhangs introduced by Taq. Each TOPO-TA reaction contained 4 µl of the fresh PCR product, I µl of the salt solution (provided in the kit), 1 µl of pCR^{TM}2.1-TOPO vector in a PCR tube. The reaction was incubated for 25 minutes at room temperature.

The TOPO-TA reaction mixture was transformed into electro-competent TOP10 cells (Invitrogen). These cells can be used for blue/white screening without adding IPTG to the culture plate. To carry out the transformation, 50 µl of the competent cells in a 2 ml microfuge tube was thawed on ice for 2 mins. 2 µl of TOPO cloning reaction was added to the cells and mixed with the cells by gently flicking the base of the tube. The tube was then incubated on ice for 30 minutes. The cells were then heat shocked by placing the tube at 42°C water bath for 30 seconds without shaking. The tube was immediately placed on ice and 250 µl of room temperature SOC medium was added to the cells. The tube was capped well and the tube was shaken horizontally at 200 rpm at 42°C for 1 hour. Meanwhile, 40 µl of 40 mg/ml of X-gal was added to 110 µl of SOC medium. The X-gal mixture was then spread evenly on the LB-agar plate containing 50 µg/ml kanamycin. The LB-plate was then pre-warmed at 37°C until use. 50 µl of each transformation mixture was added to 100 µl of SOC medium. This mixture was then spread on the pre-warmed LB-plate. Upon drying, the plate was incubated overnight at 37°C. White or very light blue colonies were

picked and the plasmid DNA transformed into the cells was checked for correct sequence and orientation by sequencing the DNA using M13 forward oligo. All the constructs used in the study had the PCR product in the same orientation.

Directional Cloning

DNA digestion

In most of the cloning, both the DNA backbone and the insert were created by digesting the appropriate plasmids with restriction endonucleases. The backbone and the insert were gel purified. In some cases, the insert was generated by PCR amplification using specific primers that were designed to introduce restriction sites at the ends of the PCR product. The PCR product was digested and purified by Qiagen gel or PCR purification kit. Both the backbone and the insert were quantitated by running them on a gel along with certain concentration of DNA ladder containing fragments of known concentration [typically hyperladder I (Fermentas) or λ-HindIII (Invitrogen)].

DNA ligation

Each ligation reaction was performed in 20 µl volume using 200ng of vector backbone and 3 times more insert molecules (i.e. backbone : insert in 1:3 mole ratio). The ligation reaction contained the vector backbone DNA, insert DNA, 2 µl of 10X T4 DNA ligase buffer (NEB), 1 µl of 20mM ATP and ddH2O to bring the volume to 20 µl. The self-ligation control reaction contained equal

amount of ddH2O in the place of insert. Ligation reactions were incubated overnight at 14°C in a thermocycler and then transformed into bacteria.

Bacterial cell transformation

DH5α cells: This strain of *E.coli* contains some mutations that make it suitable for cloning. For instance, it contains a *recA1* mutation that increases the insert stability. It also contains a mutation in *endA1* gene, which encodes an inactive endonuclease that otherwise degrades plasmid DNA. For transformation, 100 µl of chemically competent house-made DH5α cells was transferred to pre-chilled 13 ml falcon tube on ice. 13 µl of ligation mixture was added to the cells and mixed by tapping the base of the tube. The tube was left on ice for 30 minutes after which the competent cells were heat-shocked in a 42°C water bath for 90 seconds. The cells were immediately transferred to ice for 2 minutes. 900 µl of pre-warmed (at 37°C) bacterial SOC medium or LB broth was added to each tube. The tubes were shaken at 37°C at 250 rpm for 45 minutes. 250 µl of each transformation mixture was plated evenly on an individual LB-agar plate containing 50 µg/ml ampicillin. Upon drying, the plate was incubated at 37°C overnight. To the remaining 750 µl mixture, 250 µl of 60% sterile glycerol was added and the mixture was stored at - 80°C. After incubation, discrete colonies were picked from the agar-plate and the desired clone was identified using restriction digestion and or sequencing of the plasmid DNA obtained from the clone. DH5α cells take overnight to enter the logarithmic phase of their growth cycle.

Stbl2 cells: Besides mutations in *recA1* and *endA1* genes as in DH5α cells, the genotype of these cells also contain changes that stabilize direct repeat and retroviral sequences. Therefore, all the full-length, LTR containing proviral constructs were transformed into these cells. Transformation of the Stbl2 cells was done as described above for DH5α cells with the following changes: First, the cells were heat-shocked for 25 seconds instead of 90 seconds. Second, the LB broth or SOC medium added to the cells after the heat-shock step was brought to room temperature and not pre-warmed to 37°C. Third, all incubations were done at 30°C instead of 37°C. The lower incubation temperature decreases the growth rate of the bacteria. Therefore colonies were picked after 24 hours. The reason behind growing the bacteria at lower temperature is to further reduce the possibility of recombination in the plasmid; the rate of recombination between the LTRs (or direct repeats) in the plasmid is slower in slowly replicating bacteria.

Cell lines and their maintenance

293T cells

This human cell line was originially generated by Dubridge *et al.* (64) by stably expressing simian virus 40 large T antigen in the standard HEK293 cell line. The presence of T antigen allows for episomal replication of transfected plasmids containing the SV40 origin of replication. This, in turn, leads to the amplification of transfected plasmids. HEK 293 cell line is supposedly derived from human embryonic kidney cells, although it might be of neuronal origin because of the presence of mRNA and gene products typically found in neurons.

These cells are transformed with sheared adenovirus 5 DNA resulting in the incorporation of approximately 4.5 kb of viral DNA into human chromosome 19 (97, 142). In the cell line name, “293” stands for the experiment number and “T” stands for T antigen. Our lab uses the clone 17 of 293T cells. This clone was selected for its high tranfectibility and its capability of producing high titers of infectious retrovirus (174). 293T cells were maintained in IMDM (Iscove’s modified Dulbecco’s medium) supplemented with 10% bovine calf serum (BCS), and 10μg/ml of gentamycin.

Hela cells

It is an immortal human cell line derived from cervical cancer cells from Henrietta Lacks. Cells were maintained in IMDM supplemented with 10% BCS, and 10μg/ml of gentamycin.

SupT1 cells

This cell line was originally developed by Smith *et al.* (215). It was derived from malignant cells collected from the pleural effusion of an 8-year old child with T-Cell acute lymphoblastic lymphoma. The term SUP stands for Stanford University Pediatric Department, where this cell line was generated. SupT1 cells were maintained in RPMI (Roswell Park Memorial Institute) medium supplemented with10% fetal bovine serum (FBS), and 10 μg/ml of gentamycin.

p24 Assay

This assay was done to determine the levels of HIV replication by determining the amount of the viral capsid (p24) protein in cell free culture medium. p24 assays were performed using an in-house p24 sandwich ELISA system. The primary antibody used in the assay was the monoclonal antibody against p24 generated by hybridoma technology [hybridoma 183-H12-5C (NIH cat# 1513)]. Pooled anti-HIV-1 human immunoglobulin G (NIH AIDS Research and Reference Reagents, cat# 3957) was used as the secondary antibody. The secondary antibody was biotinylated to allow binding by streptavidin linked to peroxidase (ICN Biomedicals #191394). The antibody-antigen-antibody sandwich was detected by adding the substrate [*o*-phenylenediamine dihydrochloride (OPD)(Sigma #P-3804)] for peroxidase. The assay was performed as follows: 5.35 µl of 9.34 mg/ml of p24 primary monoclonal antibody was mixed with 10 ml of 1XPBS [Life technologies, Gibco Dulbecco's phosphate buffered saline (DPBS) $CaCl_2$(-), $MgCl_2$(-) Ref#14190-144]. 100 µl of the diluted primary antibody was added to each well of a 96 well round-bottom Immunolon II plate (VWR#62402-954) and sealed with a film sealant. After overnight incubation at 37°C, the unbound primary antibody was washed off with 1X PBS in a plate washer. All the washing steps in this assay were done using the plate washer. The plate washer was programmed to wash each well with the buffer four times at each step. The plate was then blocked by incubating it at 37°C for 1 hour with 250 µl of blocking buffer (1X PBS with 5% BSA)/well. During this blocking step, p24 standards were prepared by making two fold serial dilutions of p24

(1600pg/ml➔50ng/ml) solution using the tissue culture medium. The plate was washed with the wash buffer (1X PBS with 0.5% tween-20). 10 µl of lysis buffer (1X PBS with 10% trition X-100 and 0.05% trypan blue) was added to each well except the blanks. 100 µl of p24 standards and diluted samples (made in tissue culture medium used to maintain the cells) were added to the wells and the plate was further incubated at 37°C for a minimum of 2 hours or up to overnight. The plate was then washed with the wash buffer after which it was incubated at 37°C for 1 hour with 100 µl of biotinylated secondary antibody/well. The secondary antibody was diluted at 1:800 in assay buffer (1X PBS with 10% BCS and 0.5%% trition X-100) before use. Again, the secondary antibody was not added to the blanks. The plate was subsequently washed with the wash buffer to remove any unbound secondary antibody. Then 100 µl of peroxidase-streptavidin solution (diluted 1:4000 in assay buffer) was added to each well except the blanks and the plate was incubated at 37°C for 30 minutes. The plate was washed with wash buffer and 100 µl of peroxidase substrate solution (OPD) was added to all the wells including the blanks. The plate was incubated in dark at room temperature for 30 minutes. 50 µl of 2M H_2SO_4 stop solution was added to each well and the optical density (OD) of the solution in each well was measured at 492nm using Biotek's Synergy 2 plate-reader. The concentration of p24 in the sample was extrapolated from its OD value using the p24 standard linear curve equation, factoring in the dilution of the samples. The standard linear curve equation for the p24 standards was generated by performing linear regression analysis of the

standard curve created by plotting the OD values of the standards against standards' concentrations.

SEAP assay

This assay was used to determine the amount of secreted form of human placental alkaline phosphatase (SEAP) in cell-free transfection supernatant. SEAP production was used as a control for transfection efficiency. pCMV_SEAP (pHR 1831) plasmid was co-transfected with other plasmids in the Rev-dose reponse assay as a control for transfection efficiency. SEAP produced from this plasmid was quantified by a chemiluminescent SEAP kit (Tropix cat# T1015). First, 10 fold dilutions of the samples were made in the IMDM tissue culture medium supplemented with 10% BCS, and 10 µg/ml of gentamycin. 10 µl of the diluted sample was mixed with 30 µl of 1XSEAP buffer (diluted in sterile distilled water) in a 96 well plate. Since the sample diluent (culture medium) contained serum, the diluted samples with the SEAP buffer were incubated at 65°C for 30 minutes to heat inactivate serum endogenous phosphatase in a thermocycler.

Meanwhile, 10 µl of the assay buffer was added to each desired well of the 96-well white polystyrene plate (COSTAR). The plate with the samples was briefly spun and 10 µl of each sample was added in duplicate to the wells containing the assay buffer. The samples were mixed and incubated at room temperature for 30 minutes. 10 µl of substrate buffer (CSPD:dilution buffer,1:19) was added to each well and mixed together. The samples were then incubated in dark at room temperature for 20 minutes. The chemi-luminescence reading from

each well was determined using the BrightGlo protocol (integration time = 5 seconds/sample) of the Glomax-96 well luminometer.

Calcium phosphate transfection

A day prior to transfection, the cells were plated such that they were logarithmically growing on the day of transfection (i.e. 60-70% confluent at the time of transfection). Specifically, $3.5X10^6$ cells were plated on a 10 cm diameter tissue culture plate or 75 cm^2 tissue culture flask. The appropriate dilution of DNA to be transfected was made in Tris (pH 7.2) and the total DNA volume was brought to 450 µl with Tris (pH 7.2) in a sterile microfuge tube. Then, 50 µl of CaCl2 solution (2.5M $CaCl_2$ in 10mM HEPES, pH 7.2) was added to the DNA solution. The DNA and $CaCl_2$ were mixed by pipetting up and down with a 1 ml pipet. This mixture was added drop-wise to 2X HEBS in a 15 ml tube after which it was incubated at room temperature for 30 minutes. The DNA-$CaPO_4$ co-precipitate was added drop-wise to the surface of the medium containing the cells. The plate was gently swirled to allow uniform distribution of the DNA-CaPO4 co-precipitate. The cells were then incubated at 37°C in a CO_2 incubator. After 5-6 hours, the medium was replaced with fresh culture medium and the incubation was resumed.

$TCID_{50}$ (50% tissue culture infective dose) determination and MOI calculation

10^4 SupT1 cells in 150 µl RPMI full medium were seeded on each well of

a 96 well plate with clear flat bottom. The wells at the border of each plate were not used for the experiment as the rate of evaporation is higher from these wells. These well were instead filled with 150 µl of 1X PBS. The plate was placed in the incubator until infection. The viral stocks were thawed and 10 fold serial dilutions of the viral stocks were made in RPMI full medium. 100 µl of each serial dilutions was added in triplicate to the appropriate wells. Some of the wells were left uninfected as negative control for viral stock. The plate was placed back in the incubator. After 3 days, 150 µl of the cell-free culture was removed carefully without disturbing the cells. 150 µl of fresh culture medium was added back to each well and the plate was placed back in the incubator. On days 6, 8, 10, and 12 post-infection, 10 µl of cell-free supernatants were removed and the p24 levels in the supernatants were determined. The $TCID_{50}$ values in infectious units per ml (IU/ml) were calculated for each virus using the Reed-Muench technique. The $TCID_{50}$ values were used to determine the multiplicity of infection (MOI). MOI is the ratio between the number of infectious units (IU) to the number of cells.

RNA preparation for SHAPE experiments and RRE gel migration experiments

The RNAs were prepared by in vitro transcription using the MegaShortScript kit (Ambion/Life Technologies). The template for the transcription reaction consisted of the double stranded DNA corresponding to the 232-nt wt/mutant RRE (short RRE) sequence with T7 promoter at the 5’ end and a structure cassette at the 3’ end. The template was PCR amplified using oligo 5’

T7RRE and oligo 3' SCRRE. The RNAs were PAGE purified (5% polyacrylamide (19:1), 1x TBE, 7 M urea), quantified (Abs at 260 nm) and stored at -20°C in TE light 7.6 buffer (10 mM Tris, pH 7.6; 0.1 mM EDTA) prior to use.

Sequence information of oligos used:

Oligo 5' T7RRE designed to introduce T7 promoter sequence at the 5' end of the 232-nt RRE. T7 promoter sequence is shown in bold and larger fonts.

5' **gctaatacgactcactatagg**ggagctttgttccttgggttc

Oligo 3' SCRRE designed to introduce the structure cassette sequence at the 3' end of 232-nt RRE. Structure cassette sequence is shown in bold and larger fonts

5' **gaaccggaccgaagcccgatttggatccggcgaaccggatcga**aggagctgttgatcct ttaggtatc

RRE gel migration assay

This assay was used for two purposes. First, it allowed us to determine the structural homogeneity of the folded RNA. Second, it provided a preliminary estimate of the RNA structure by comparison of the migration rates of different RNAs. At first, an 8% native polyacrylamide gel (29:1) was prepared by mixing 6 ml of 40% Acrylamide (29:1), 3 ml of 10X TBE, 150 µl of 1 M $MgCl_2$ (final concentration = 5 mM), and 20.85 ml of cold Millipore (deionized) water to get a final volume of 30 ml. 300 µl of 10% APS, 30 µl of TEMED were added to this mixture. Immediately, the mixture was poured on the gel mold with 0.8 mm thick

spacers and then 0.8 mm thick gel comb with 4.5 mm wide wells was place on the gel. The distance between the bottom of the well to the bottom of the gel was 15.8 cm. The gel was left to polymerize. The gel was pre-run at constant 200 V at 4°C for at least 45 minutes. The running electrophoresis buffer (1XTBE) also contained 5 mM $MgCl_2$.

While the gel was pre-running, the RNA was folded. For this, approximately 20 pico-moles of RNA was mixed with re-naturation buffer (10 mM Tris - pH 8.0, 100 mM KCl, 0.1 mM EDTA) in a volume of 5 µl. The RNA solution was then refolded by heating it to 85°C for 2 minutes, followed by slow cooling to 25°C for 15 minutes (ramp rate 0.1°C/sec). The refolded RNA was incubated with 5 µl of 2X RNA folding buffer containing 70 mM Tris pH 8.0, 180 mM KCl, 0.3 mM EDTA, 8 mM $MgCl_2$, 5% of glycerol at 37°C for 30 min and then placed on ice. The folded RRE was loaded on the wells of the gel. For tracking the progress of the run, xylene-cyanol alone was loaded on a separate distant well. The gel was then run at constant 200V for 22 hours at 4°C. The running buffer was changed at least once (preferably twice) during the run. After 22 hours, the RNA had travelled ~ 13 cm from the bottom of the well and xylene cyanol had run off the gel. The RRE bands were then visualized by UV shadowing.

Conventional SHAPE

It was used to determine the secondary structure of structurally homogenous RNA in solution**.**

RNA folding and NMIA modification

Approximately 20 picomoles of RNA was mixed with 2 µl of 10X re-naturation buffer (100 mM Tris - pH 8.0, 1 M KCl, 1 mM EDTA) in a volume of 20 µl adjusted with nuclease free water (Ambion). The RNA solution was then refolded by heating it to 85°C for 2 minutes, followed by slow cooling to 25°C for 15 minutes (ramp rate 0.1°C/sec). 100 µl of nuclease free water (Ambion) was added to the RNA mixture. The mixture was then incubated with 30 µl of 5X RNA folding buffer (200 mM Tris - pH 8.0, 650 mM KCl, 2.5 mM EDTA, 25 mM $MgCl_2$) at 37°C for 20 minutes. The resulting 150 µl of folded RNA was divided into two 72 µl tubes. One of the tubes was labeled “NMIA+ or +” and the other labeled “NMIA- or -“. Eight µl of 30 mM anhydrous NMIA (*N*-methylisatoic anhydride) in DMSO was added to the “+” tube and 8 µl of DMSO was added to “-“ tube. Both the tubes were then incubated at 37°C for 50 minutes.

The chemically modified RNA from the “+” tube and the unmodified RNA from the “-“ tube were precipitated by treating them with 8 µl of 3M NaOAc-pH 5.2, 0.5 µl of glycerol and 240 µl of pre-chilled 95% ethanol for at least 1 hour at -20°C. The RNA precipitate was collected by centrifuging the tubes at 4°C at 13,000 g for 30 minutes. The supernatant was removed and the pellet was air dried for ~ 10 minutes. Ten µl of nuclease free water (Ambion) was added to the pellet and the RNA suspension was stored at -20°C.

Reverse transcription and cDNA fractionation

The RNA was reverse-transcribed using Superscript III kit (Invitrogen). To achieve this, 5 µl of the RNA (1-3 picomoles) was mixed with 1 µl of Cy5-labeled oligo SC (for "+" RNA) or Cy5.5-labeled oligo SC (for "-" RNA), 1 µl of 2 mM EDTA, 5 µl nuclease free water in a 0.5 ml PCR tube (Sarstedt # 72735002). The mixture was incubated in a thermo-cycler with the heating program: 85°C for 1 minute, 65°C for 5 minutes, 4°C for 10 minutes, 50°C for 50 minutes. During the 4°C step, 8 µl of RT mix was added to each tube. The RT mix contained 4 µl of 5X RT buffer, 1 µl of 100 mM DTT, 1.5 µl of water, 1 µl 10 mM dNTPs, and 0.5 µl of Superscript III per reaction. The remaining RNA in the cDNA preparation was hydrolyzed by adding 1 µl of 2 M NaOH and then incubating the tube at 95°C for 5 minutes in heating block. The cDNA tube was then placed on ice immediately and the solution neutralized by adding 1 µl of 2 N HCl.

The cDNA preparations from the "+" tube and the "-" tube were mixed in a single tube and the cDNA mixture was precipitated by adding 56 µl of water, 10 µl of 3 M NaOAc, 0.5 µl glycogen, and 300 µl of pre-chilled 95% ethanol. The resulting mixture was mixed well and incubated at -20°C for at least 1 hour. The pellet was centrifuged at 13,000g for 30 minutes. The pellet was washed twice with pre-chilled 70% ethanol. The supernatant was discarded and the pellet air-dried on Speed-Vac for 5 min. The pellet was vortexed in 40 µl of de-ionized formamide for 20 minute after which it was incubated at 65°C for 5 minutes and then vortexed again for 10 minutes. At this point, the cDNA mixture was either stored at -20°C or fractionated immediately. Fractionation was done using

capillary electrophoresis (CE). Ten µl of ddA and ddC sequencing ladders were added to the cDNA mixture just before fractionation.

Generation of sequencing ladder

The sequencing ladder was prepared by cycle-sequencing using USB cycle sequencing kit (# 78500). Each reaction contained 2.1 µl of water, 4 µl of ddA or ddC, 0.5 µl of fluorescent labeled SC primer (10 picomoles/µl) (D2-labeled for ddA reaction and Licor labeled for ddC reaction), 0.45 µl of purified short wt RRE DNA template (same as that used as template in *in-vitro* transcription of RRE RNA for SHAPE experiment), 0.45 µl of 10X Taq sequencing buffer, 0.5 µl of Taq DNA polymerase (1.25 units). The reaction was incubated in a thermo-cycler using the following cycling conditions:

96°C for 2 minutes

25 cycles of 96°C for 20 seconds, 55°C for 20 seconds, 72°C for 1 minute

72°C for 1 minute

4°C forever

After the completion of cycle sequencing, 1.6 µl of stop mix was added and vortexed after which 20 µl of pre-chilled 95% ethanol was added and then vortexed again. The stop mix was prepared during the cycle sequencing step. The stop mix contains 100 µl of 3 M NaOAc (pH 5.2), 80 µl of water, 20 µl of 500 mM EDTA, and 2 µl of glycogen. The cycle sequencing mixture was centrifuged immediately at 14,000 g for 30 minutes. The pellet was washed twice with 400 µl of pre-chilled 70% ethanol and then dried in Speed-Vac for 5 minutes. The pellet

was re-suspended in 100 µl of deionized formamide, then vortexed for 30 minutes, heated for 10 minutes at 65°C and then vortexed again for 10 minutes. The sequencing ladder so prepared was then stored at -20°C until use.

Oligo 3' SC: Reverse oligo that anneals to the 3' end of the structure cassette of the RRE cDNA or ds-DNA with the 3' structure cassette tag.

5' **gaaccggaccgaagcccgattt**

SHAPE data analysis and presentation

SHAPE data analysis was done as previously described (145). Briefly, the CE electropherograms from the SHAPE experiment were imported into Shape-Finder software (version 1). Using this software, the CE traces were adjusted to correct for i) fluorescent background, ii) spectral overlap between fluorescent channels, iii) mobility shifts of each tagged primer, iv) signal decay from premature termination of reverse transcription. After these adjustments, the nucleotide identity of each peak of the electropherogram was determined. NMIA reactivity value of each nucleotide was generated by the software by substracting the peak area of the "NMIA-" reaction from the peak area of the "NMIA+" reaction. The reactivity values from the software were imported into Excel spreadsheet where these values were normalized. The normalization step includes i) excluding outliers, ii) determining the "effective maximum" reactivity (i.e., the average of the highest 8% of reactivity value, and iii) normalization by dividing all reactivity values by the "effective maximum". The normalized reactivity values were imported into the RNAstructure 5.5 software. This software

was used to generate a graphical representation of the energetically most favorable secondary structure of the RRE under the SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. The dot-bracket format of the so-generated secondary structure was imported into an RNA structure viewer software called VARNA (version 1). The VARNA generated structure was further imported into Adobe Illustrator (CS6) to create the final RRE secondary structure figures showing color-coded reactivity values of each nucleotide. Reactivity values <0.3 were assigned as “no reactivity”, those between 0.3 to 0.5 were grouped under “low reactivity”, those between 0.5 to 0.7 were assigned “moderate reactivity” and those > 0.7 were grouped under “high reactivity”.

In-gel SHAPE

It was used to determine the secondary structure of structurally heterogenous WT RRE. The WT RRE RNA was folded, and run and visualized on a native gel exactly as done in the RRE migration assay except for the following changes: i) The RRE folding reaction was scaled up 4 times, ii) 20 µl of the re-folded WT RNA was loaded on two alternate wells. Each of the two wt RRE bands were excised and placed into separate microfuge tube. The RRE bands from one well were NMIA modified while those from the other well were left unmodified (used as NMIA control). The “NMIA +” RRE bands were modified by soaking them in 1XTBE containing 10% of 100mM NMIA. The “NMIA-” RRE bands were soaked in 1XTBE containing 10% of DMSO. The bands were soaked

in their respective solutions at 37°C for 45 minutes. The gel slices were then washed three times in 1X TAE after which they were crushed into small pieces using sterile RNase-free pipette tips. The crushed gel pieces were placed into an individual Elutrap (Whatman) channel containing 1× TAE. The RNA in the gel pieces was electroeluted overnight at 100 V into a chamber bordered by Whatman BT-1 and BT-2 membranes. The electro-eluted RNA was precipitated exactly as the precipitation of the RNA immediately after NMIA treatment in the conventional SHAPE. The RNA was re-suspended in 13 µl of water. 2 µl of the RNA was removed to determine the concentration of RNA by nano-drop. The downstream procedure of reverse transcription, cDNA fractionation, SHAPE data analysis and structure presentation were done exactly as in conventional SHAPE.

Rev-RRE gel shift

This assay allowed us to study the Rev binding and Rev mutlimerization pattern of different RREs. First, a 4% polyacrylamide gel was prepared. To prepare the gel, in a 50ml falcon tube, 3.323 ml of 30% acrylamide/bis solution (19:1; BIORAD #161-0154), 5 ml of 5XTBE, and 7.25 g of urea were combined and the volume was brought to 25 ml. The mixture was mixed well and filtered through 0.45µM filter. Just before making the gel sandwich, 250 µl of freshly prepared 10% APS and 20 µl of TEMED were added to the mixture and the mixture was poured into the gel mold and allowed to

polymerize. The gel was pre-run at constant 20mA current (4 Watts power) for 30 minutes at room temperature.

While the gel was polymerizing, internally ^{32}P labeled 234-nt long RRE RNAs were prepared by in-vitro transcription using T7 polymerase (NEB). pCR 2.1 TOPO TA _ short RRE constructs were used to generate the templates for the transcription. These constructs contain the 234-nt RRE between two EcoRI sites. Each construct was digested with EcoRI to release the RRE ds DNA fragment from the construct. The RRE DNA fragments were then gel purified and used as template for the transcription reaction. Each transcription reaction contained 1X T7 RNA Pol buffer (NEB), 7.5 mM DTT, 20 Units of RNasin Plus Rnase Inhibitor (Promega) 0.5mM rNTP's (-UTP), 5 mM UTP, 50μCi (0.84 mM) of $\alpha^{32}P$ UTP (Perkin Elmer), 40 ng template DNA and 50 U T7 RNA Pol in a volume of 20 μl. The transcription reaction mixes were incubated at 37°C for 1 hour, treated with DNaseI (Promega), further incubated for another 1 hour. 20 μl of RNA loading dye (80% formamide, 20% TE, 0.025% bromophenol blue, 0.025% xylene cyanol) was added to each reaction. One μl of this mix was drawn out to record its cpm using the gieger counter for later determination of the concentration of the final RRE probe. The rest of the mix was heated at 80°C for 3 minutes, gently spun and loaded on a 4% denaturing acylamide gel. The gel was run at constant 20mA current (4 Watts power) for 1 hour at room temperature until the first line of dye (bromophenol blue) was near the bottom of the gel.

The 234-nt long RRE band was located by autoradiography. The RRE

RNA band was excised and the RNA was passively eluted from the band by gently shaking the gel slice in 300 µl TE buffer overnight at 4°C. The TE containing the RNA was transferred to another clean microcentrifuge tube. The eluted RNA was purified by phenol/chloroform extraction and concentrated by ethanol precipitation. This was done by adding 300 µl of chloroform followed by 300 µl of phenol pH6.8. Extraction was done by adding the reagent and inverting the tube 6 times and then centrifuging the tube at 13000 rpm for 1 min at RT. The aqueus phase was collected in a new tube. Then 30 µl of Na0Ac(pH5.5) and 600 µl of 100% ethanol were added. The mixture was mixed by inverting 6 times and then placed at -80°C for 1 hour. The tube was centrifuged at 4°C for 30 minutes at 13000 rpm. A small dot of pellet was seen. The supernatant was removed and the pellet washed with 80% ethanol. The supernatant was discarded and the pellet was air dried on ice for 30 minutes. The pellet was then suspended in 20 µl of RNase free water. The RNA suspension was passed through the DEPC-Chromaspin 30 column (B.D Biochemicals) to get rid of the unincorporated nucleotides and stored in small aliquots at -80°C.

The Rev-RRE gel shift was carried out in a 4% non-denaturing acrylamide gel using PROTEAN II xi cell vertical electrophoresis unit (BIORAD). To make the gel, in a 250 ml clean sterile glass bottle, 13.3 ml of 30% Acrylamide/Bis solution (19:1; BIORAD #161-0154), 20 ml of 5XTBE, and 66.7 ml of RNase free water were combined. The mixture was filtered through a 0.45 µM filter and immediately before pouring the gel, 1ml of freshly prepared 10% APS and 80 µl

of TEMED were added to the gel. The gel mold was made using spacers of 1.5 mm thickness and a 15 well-comb of 1.5mm thickness and the gel was allowed to fully polymerize. The gel cassette and 1X TBE running buffer were added to the unit and the unit was left at 4°C overnight. Next day, the gel was pre-run at constant 20 mA current (4 Watts power) for at least 30 minutes at 4°C.

Meanwhile, the Rev-RRE complexes were prepared. To prepare the Rev-RRE complexes, serial dilutions of Rev protein (10 ng/µl, 25 ng/µl , 50 ng/µl , 100 ng/µl , 200 ng/µl) were made in Rev storage buffer (50 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA) immediately before use. 1 µl of each Rev dilution or 1 µl of Rev storage buffer alone was incubated with Rev binding buffer [10 mM Hepes/KOH, pH7.8, 20 mM KCL, 2 mM MgCl2, 0.5 mM EDTA, 1 mM DTT, 10% glycerol, 5 µg/ml yeast tRNA (Invitrogen) and 20 U RNase inhibitor (Promega)], in a volume of 10.6 µl, on ice for at least 10 mins. Meantime, internally labeled RNA (at a final concentration of 1 nM/µl) was folded by incubating it in re-naturation buffer (50 mM NaCl, 10 mM Hepes/KOH, pH 7.6, 2 mM $MgCl_2$) at 85°C for 3 minutes and then at room temperature for 15 mins. One µl of the renatured RNA (~1 nM/µl) mix was added to the pre-incubated Rev binding buffer mix with or without Rev. The resulting mixture was incubated on ice for 10 minutes before loading it onto a 4% native polyacrylamide gel. To indirectly ensure that the bands were migrating on the gel, 10 µl of the RNA loading DNA was also to one of the wells of the gel and the migration of the dye tracked. The gel was run at constant 20 mA current (4 Watts power) at 4°C until the xylene cyanol dye migrated to the end of the gel. The gel was then vacuum dried at

80°C for 90 minutes. The gel was exposed to a phosphorimaging screen for >16 hours and visualized with a phosphorimager (Molecular Dynamics) and ImageQuant software.

Rev dose response assay

This assay was used to study the efficiency of different RREs to mediate Gag expression at different concentration of Rev expression plasmid. One day before the transfection, 22 mm diameter wells of 12 well tissue culture plates were seeded with $2x10^5$ 293T cells in 2 ml of culture medium (IMDM supplemented 10% BCS, and 10ug/ml of Gentamycin). Just before the transfection, transfection DNA mixtures were prepared. Each transfection mixture contained 2 µg of pCMV_GagPol-RRE reporter, 100 ng of secreted placental alkaline phosphatase (SEAP), 2-fold increasing concentrations (0-32 ng) of pCMV_Rev plasmid, and empty pCMV plasmid to equalize the total amount of DNA in all the mixtures. The final volume of each transfection DNA mixture was brought to 200 µl using un-supplemented IMDM medium. 3 µl of turbofect reagent (Thermo Scientific) was added to each mixture. The mixtures were mixed well and then left at room temperature for 20 minutes after which each mixture was added drop-wise to the cells on the respective well. 48 hours later, transfection supernatants were collected. The supernatants were spun at 3000rpm for 3 minutes at 4°C. The levels of p24 and SEAP in the supernatants were determined as described above. The Rev dose response curves for Rev-

RRE activity were generated by plotting the SEAP normalized p24 values against Rev plasmid concentration for each pCMV_GagPolRRE.

Hygromycin Resistance Assay

This assay was used to study the Rev-RRE activity of different RREs in their native position. It measures the packaging efficiency of the genomic mRNA as a function of Rev-RRE activity. Transducing viral stocks were produced in 293T cells. For this, $3.5x10^6$ 293T cells in 10 ml of culture medium (IMDM supplemented 10% Bovine Calf Serum, and 10 µg/ml of Gentamycin) were seeded on 100 mm diameter tissue culture plates. After about 24 hours, cells were co-transfected with 20 µg of pTR 167 Nef- w/ RRE, 15 µg of pCMVΔR9, and 5 µg of pCMV_VSV-G by calcium phosphate method. 48 hours post transfection, culture medium was harvested and spun at 3000 rpm for 3 minutes at 4°C. Cell-free transfection supernatant (the viral stock) containing the transducing viruses were stored at -80°C for future use.

The viral stock was used to infect the target Hela cells. Twenty-four hours prior to infection, 60 mm tissue culture plates were seeded with $5 x10^5$ Hela cells in 4 ml of culture medium. Just before infection, the medium was sucked out of the plates and the cells were infected with 1 ml of 10-fold serial dilutions of the viral stocks (prepared in the IMDM full medium). DEAE dextran was added to each plate at a concentration of 8 µg/ml to facilitate viral adsorption. After 6 hours at 37°C, 3 ml of fresh medium was added to each plate and the incubation was resumed. After 2 days of infection, the medium was replaced with fresh medium

containing 200 µg/ml of Hygromycin B. The hygromycin medium was changed after every 3rd day. After 14 days under hygromcyin selection, the cells were washed twice with PBS and then fixed and stained using 0.5% crystal violet in 50% methanol. The titer of hygromycin resistant colonies was determined. The differences in the titer were analyzed for statistical significance at 99% confidence level by one way unrepeated ANOVA assuming homoscedasticity. The multiple comparison error was corrected by two-sided Dunnett post-hoc test. The titer values were confirmed to follow normal distribution by residual plots and by D'Agostino & Pearson omnibus normality test. The values were tested for homoscedasticity by Levene statistics. Both normality and homoscedasticity tests were done at 95% confidence level prior to ANOVA analysis. These statistical analysis were performed using IBM SPSS Statistics 21and PRISM Graphpad 5.

Replication kinetic assay

The Rev-RRE activities of different RREs were also studied using this assay (also called spreading infection) that more closely mimics the natural HIV infection. Here, viral stocks made from proviral constructs were tested in SupT1 cells. The proviral constructs used were the pNL4-3 Nef- RRE- /long RRE constructs. Transfection viral stocks were prepared by seeding 3 X 10^6 293T cells in a 75 cm^2 flask a day prior to transfection. These cells were transfected with 5 µg of the proviral DNA by calcium phosphate method. Transfection supernatant was collected after 48 hrs and was spun briefly at 2500Xg for 5 minutes at 4°C to get rid of cell debris. SupT1 cells were infected with the

transfection viral stocks by DEAE dextran method and by unaided infection. DEAE dextran mediated infections were done by adding 100 ng p24 equivalent of viral stocks to $6X10^6$ SupT1 cells in 1 ml of culture medium (RPMI supplemented 10% fetal bovine serum, and 10 ug/ml of gentamycin) containing 8 µg/ml of DEAE. Infections were carried out in a loosely capped 50 ml falcon tube at 37°C. Six hours after infection, the infected cultures were centrifuged at 2500Xg for 5 minutes at 4°C and the supernatants were discarded. The infected cells were washed with 1X PBS twice and then gently suspended in 5 ml of medium. The culture was then transferred to 25 cm2 flask for further incubation. The unaided infections were carried out by infecting $6X10^6$ SupT1 cells in 10 ml of culture medium with 300 ng p24 equivalent of viral stocks in 25 cm^2 flasks and then incubating the cultures at 37°C.

In both the DEAE mediated and unaided infections, after every 3-4^{th} day, 2/3^{rd} of the culture was replaced with fresh medium needed to maintain the culture volume to 5 ml or 10 ml respectively. The replaced culture was spun at 2500 g for 5 minutes at 4°C to remove the cell debris. The secreted p24 in the cell-free replaced culture was determined by ELISA to generate the growth curve.

The transfer replication kinetic assay: The infection culture from the peak p24 day of the replication kinetic assay was spun at 2500 g for 5 minutes at 4°C. The TCID50 value of the cell-free infection supernatant was determined. The MOIs of the SupT1 passaged viral stocks were determined as described (1) with p24 levels used as the measure of viral replication instead of RT activity

(Table 2 and Table 3). $6X10^6$ SupT1 cells were infected with $5X10^{-6}$ MOI (100ng p24 equivalent) of SupT1 passaged viral stocks by DEAE dextran method.

Table 2: SupT1 infection to determine TCID50 for set A viruses. p24 levels in the supernatant of triplicate SupT1 cultures 12 days after infection with p24 normalized SupT1-passaged virus are shown. 10-fold serial dilutions of the A set (set A) of WT, mutant A, and mutant B viruses were used as input virus such that all the dilution 0 cultures were infected with 100 µl of 100 ng p24/100 µl of viral stock. These p24 values (in picograms/ml) were used to determine the TCID50 values of these set A viruses.

	WT			
Dilutions	Infection-1	Infection-2	Infection-3	Percent infected
0	Overflow*	overflow*	overflow*	100
10^1	8842.2	8967.2	12573.9	100
10^2	<0.0	<0.0	<0.0	0
10^3	<0.0	<0.0	<0.0	0
10^4	<0.0	<0.0	<0.0	0
10^5	<0.0	<0.0	<0.0	0
10^6	<0.0	<0.0	<0.0	0
10^7	<0.0	<0.0	<0.0	0
Dilutions	Mutant A			Percent infected
0	overflow*	overflow*	overflow*	100
10^1	11169.7	13679.9	11823.7	100
10^2	<0.0	<0.0	<0.0	0
10^3	<0.0	<0.0	<0.0	0
10^4	<0.0	<0.0	<0.0	0
10^5	<0.0	<0.0	<0.0	0
10^6	<0.0	<0.0	<0.0	0
10^7	<0.0	<0.0	<0.	0
	Mutant B			
Dilutions	1	2	3	Percent infected
0	overflow*	overflow*	overflow*	100
10^1	16420.9	16517.1	overflow*	100
10^2	<0.0	<0.0	<0.0	0
10^3	<0.0	<0.0	<0.0	0
10^4	<0.0	<0.0	<0.0	0
10^5	<0.0	<0.0	<0.0	0
10^6	<0.0	<0.0	<0.0	0
10^7	<0.0	<0.0	<0.0	0

overflow* = p24 value higher than the linear limit of detection

Table 3: SupT1 infection to determine TCID50 for set B viruses. p24 levels in the supernatant of triplicate SupT1 cultures 12 days after infection with p24 normalized SupT1-passaged virus are shown. 10-fold serial dilutions of the B set (set B) of WT, mutant A, and mutant B viruses were used as input virus such that all the dilution 0 cultures were infected with 100 µl of 100 ng p24/100 µl of viral stock. These p24 values were used to determine the TCID50 values of these set B viruses.

Dilutions	WT			Percent infected
	1	2	3	
0	Overflow*	Overflow*	Overflow*	100
10^1	18151.8	20397.5	20867.8	100
10^2	<0.0	<0.0	<0.0	0
10^3	<0.0	<0.0	<0.0	0
10^4	<0.0	<0.0	<0.0	0
10^5	<0.0	<0.0	<0.0	0
10^6	<0.0	<0.0	<0.0	0
10^7	<0.0	<0.0	<0.0	0
Dilutions	Mutant A			Percent infected
	1	2	3	
0	Overflow*	Overflow*	Overflow*	100
10^1	17744.5	22309.0	21212.3	100
10^2	<0.0	<0.0	<0.0	0
10^3	<0.0	<0.0	<0.0	0
10^4	<0.0	<0.0	<0.0	0
10^5	<0.0	<0.0	<0.0	0
10^6	<0.0	<0.0	<0.0	0
10^7	<0.0	<0.0	<0.0	0
Dilutions	Mutant B			Percent infected
	1	2	3	
0	Overflow*	Overflow*	Overflow*	100
10^1	21786.8	23520.7	22089.7	100
10^2	<0.0	<0.0	<0.0	0
10^3	<0.0	<0.0	<0.0	0
10^4	<0.0	<0.0	<0.0	0
10^5	<0.0	<0.0	<0.0	0
10^6	<0.0	<0.0	<0.0	0
10^7	<0.0	<0.0	<0.0	0

overflow* = p24 value higher than the linear limit of detection

MOI calculation for the viruses:

Since the number of infected cultures at a certain dilution was identical for all the viruses, the TCID50 and the MOI values were also identical for all the viruses. These values were calculated as follows for each virus using Reed Muench method(192).

$$\text{Index(h)} = \frac{\text{(\% infected at dilution immediately above 50\% -50\%)}}{\text{(\% infected at dilution immediately above 50\%) – (\% dilution immediately below 50\%)}} \times 100$$

=(100-50)/(100-0)

=50/100

=0.5

Apply the index calculated using this formula to the dilution that produced the infection rate immediately above 50 percent = $10^{-1.5}$

This dilution of the virus suspension contained one $TCID_{50}$ unit of virus in 0.1 ml

Therefore, infectivity titre of virus suspension in $TCID_{50}$/mL = $10 \times 10^{1.5} = 10^{2.5}$ = 316.227766 infectious dose.
Or, 100 ng of the viral stock contained 31.6 infectious units.

1 ng of the viral stock contained 0.316 infectious units

For the transfer replication assay, 6,000,000 SupT1 cells were infected with 100 ng p24. Therefore, the MOI used in this experiment was $5X10^{-6}$.

For the competition assay, 60,000 SupT1 cells were infected with 1 ng p24 equivalent virus.

Or, 60,000 cells were infected with 0.316 IU. Therefore, the multiplicity of infection was $5X10^{-6}$.

Growth Competition Assay:

Here, the growth kinetic of two different viruses containing different RREs were measured in a dual infection assay that provided identical culture condition to both the viruses. The viral stocks used in this assay were the cell-free infection cultures from the peak p24 day from the replication kinetic assay. Thirty-six well plates were seeded with 6 $X10^4$ SupT1 cells/well in a volume of 1ml RPMI full medium. $5X10^{-6}$ MOI (p24 equivalent) of each of the two different SupT1 passaged viral stocks were added to the appropriate well. The same MOI of each of the viral stocks was also added separately to appropriate wells for use as mono-infection controls. After 3 days, 500 µl of the cell-free culture was carefully removed and replaced with 500 µl of fresh culture medium in the place of RT activity. Ten µl of cell-free infection culture was taken out on days 6, 8, and 10 (optional) post-infection and the p24 levels in the culture were determined. On day 10, the cells were harvested and the proviral DNA from the cells was extracted using DNeasy blood and tissue DNA kit (Qiagen) according to manufacturer's instruction. The amount of proviral DNA coming from each of the two viruses in the dual infections was determined by hetero-duplex tracking assay. For this, the proviral DNA was amplified by nested PCR using outer PCR oligo pair (forward oligo 2760 and reverse oligo 2761) and inner PCR oligo pair (forward oligo 2767 and reverse oligo 2768). The final nested PCR product was 216-nt long starting at nt-116 and ending at nt-331of the 351-nt long RRE.

To prepare the HTA probe amplicons, the forward oligo 2767 was first end-radiolabeled. The end-labelling reaction contained 20 µl of oligo 2767, 10 µl

of 5X forward reaction buffer, 3 µl T4 PNK (NEB), 3 µl of γ^{32}P dATP (3000 Ci/mmol, 10 mCi/mL), and 14 µl of water. The mixture was incubated at 37°C for 60 minutes after which it was passed through G-25 column resin (Roche) according to the manufacturer's manual. Next, the labeled oligo was used to PCR amplify a part of the RRE region. The PCR reaction contained 2.5 µl of 10ng/µl of nested outer PCR product of pHR 4874, 10 µl of 10X PCR buffer (with 1.5 mM final MgCl2), 1.5 µl of 10 mM dNTPs, 7 µl of γ^{32}P labeled forward oligo 2767 (4 pmol/µl), 3 µl of reverse oligo 2768 (10 pmol/ µl), 0.5 µl Taq DNA polymerase (5 U/µl), and water in 100 µl volume.

The proviral DNA amplicons were annealed with the probe. This was accomplished by setting up the HTA reaction. Each reaction contained 10 µl of nested PCR product, 1 µl of 10X annealing buffer (100mM Tris-HCl pH 7.8, 10mM NaCl, 2mM EDTA), 1 µl of probe reaction (~300–400 CPM final or 0.1 pmol per reaction). The mixture was heated at 95°C for 3 minutes, 37°C for 5 min, and then held at 4°C in a thermocycler. The reaction mixtures were placed on ice and 3 µl of HTA loading dye (5ml of glycerol, 0.03g of Tris-HCl, 5 µl of 1M DDT, 0.25% of bromophenol blue, 0.25% of xylene cyanol in a final volume of 10 ml) was added to each tube. The tubes were briefly spun and the tube contents were loaded on separate wells of the 8% non-denaturing polyacrylamide gel. The gel was run at constant 200V at room temperature for 3 hours till the xylene cyanol dye had migrated to ¾ down the gel. The gel was then vacuum dried at 80°C for 60 minutes and exposed to a phosphorimaging screen for >16 hours. The bands on the gel were visualized with a phosphorimager (Molecular

Dynamics) and ImageQuant software and the intensity of each band was quantitated using ImageQuant software. The relative fitness of each virus (say A and B) in a dual infection was determined as follows:

Let the intensity of the virus A signature band in the mono-infection be M_A

Let the intensity of the virus A signature band in the dual-infection be D_A

Similarly,

Let the intensity of the virus B signature band in the mono-infection be M_B

Let the intensity of the virus B signature band in the dual-infection be D_B

$$\text{Then, relative fitness of virus A (in \%)} = \frac{(D_A/M_A)}{(M_A+M_B)} \text{ X } 100$$

$$\text{Relative fitness of virus B (in \%)} = \frac{(D_B/M_B)}{(M_A+M_B)} \text{ X } 100$$

The 0.75 mm thick HTA gel was prepared before starting the HTA reaction. To prepare the gel, 15 ml of 40% acrylamide/bis solution (19:1), 10 ml of 10X non-denaturation gel shift buffer (0.25 M Tris-HCl, 1.9 M electrophoresis-grade glyxine, 10 mM EDTA), and 75 ml of water were combined. Two ml of 10% APS, 120 µl of TEMED were added to the acrylamide solution just before pouring the mixture into the gel mold. The gel comb with 15 teeth and 0.75 mm thickness was used to create the wells. The polymerized gel cassette was placed into the protean II xi (Bio-Rad) vertical gel electrophoresis unit and filled with appropriate

volume of running buffer. 1X non-denaturation gel shift buffer was used as the running buffer for the PAGE.

Sequence information of oligos used in HTA

Oligo # 2760: Forward outer primer for HTA that anneals to position nt-8560 to nt-8581 of pHR 4874

5’ cgaggattgtggaacttctggg

Oligo # 2761: Reverse outer primer for HTA that anneals to position nt-9288 to nt-9308 of pHR 4874

5’ gtgactggaaaacccacctc

Oligo # 2767: Forward inner primer for HTA that anneals to position nt-8943 to nt-8962 of pHR 4874

5’ aatgacgctgacggtacaggccagac

Oligo# 2768: Reverse inner primer for HTA that anneals to position nt-9155 to nt-9180 of pHR 4874

5’ caactagcattccaaggcacagcagtgg

CHAPTER III - RESULTS

Evidence for two secondary structures of the WT NL4-3 RRE

In a previous joint study from our lab and Dr. Stuart Le-Grice's lab (NIH, Frederick), (137), using conventional *selective 2' hydroxyl acylation analyzed by primer extension* (SHAPE), we have shown that the WT NL4-3 RRE could adopt a homogeneous 4 stem-loop structure which combined stem-loops III and IV, and that single base changes could shift the structure into a 5 stem-loop form with separate stem-loops III and IV. Conventional SHAPE is an RNA secondary structure chemical probing technique that interrogates the local flexibility of ribo-nucleotides of RNA in an aqueous solution on the basis of their reactivity to a hydroxyl-selective electrophile. The 2'-hydroxyl group of unconstrained or non base-paired nucleotides (bulge loop, junction, hairpin loop, single stranded region) have enhanced nucleophilicity and are therefore more reactive to the chemical than the base-paired or constrained nucleotides (242). The reaction forms a 2' bulky adduct and hence the chemically modified nucleotide can be identified as stops to primer extension of the RNA by reverse transcriptase. Since the reactivity is not significantly affected by the identity of the nucleo-base (243), this technique provides structural information of RNA at single nucleotide resolution.

Shortly after our report, a second conventional SHAPE study (238), performed in another laboratory, examined the structure of the entire HIV genome prepared from virions, and concluded that within this context the WT NL4-3 RRE folded as a 5 stem-loop structure. All other studies on the structure

of the HIV-1 RRE also support either the 4 stem-loop structure or the 5 stem-loop structure (32, 57, 125, 155). Both the 4 and 5 stem-loop RREs retain the same secondary structure in the regions other than stem-loops III and IV of the RRE including the primary Rev binding site. Both the structures fold with nearly equal predicted free energy. The Gibbs free energy change values of the 5 stem-loop structure RRE and the 4 stem-loop structure RRE were found to be -647.26 kJ/mol and -636.8kJ/mol respectively by RNAStucture v-5.5 (Figure 5). These findings led us to hypothesize that the RRE is a metastable structure that can readily switch between these two forms. To test this hypothesis, we revisited the secondary structure of the RRE.

The secondary structure of the 232-nt RRE sequence ("short RRE" Figure 5) was determined using the SHAPE technique. This RRE sequence consisted of pNL4–3 nucleotides 7760–7992 (GenBank accession number AF324493). It is numbered as nucleotides 60–292 based on the RRE sequence of Charpentier *et al*. (32). To determine the secondary structure, the RRE, adjacent to a short primer binding cassette, was synthesized *in vitro* with T7 polymerase, heated to 85°C, and cooled slowly to 25°C to allow refolding. After refolding, it was analyzed on a native 8% polyacrylamide gel at 4°C run slowly at 200V constant voltage for 22 hrs. Figure 9C shows that the RRE migrated as two bands. Similar results have been observed previously by others (172). Because conventional SHAPE cannot distinguish the reactivity signals from different structural forms of an RNA population, the RRE structure was determined by in-gel SHAPE technique.

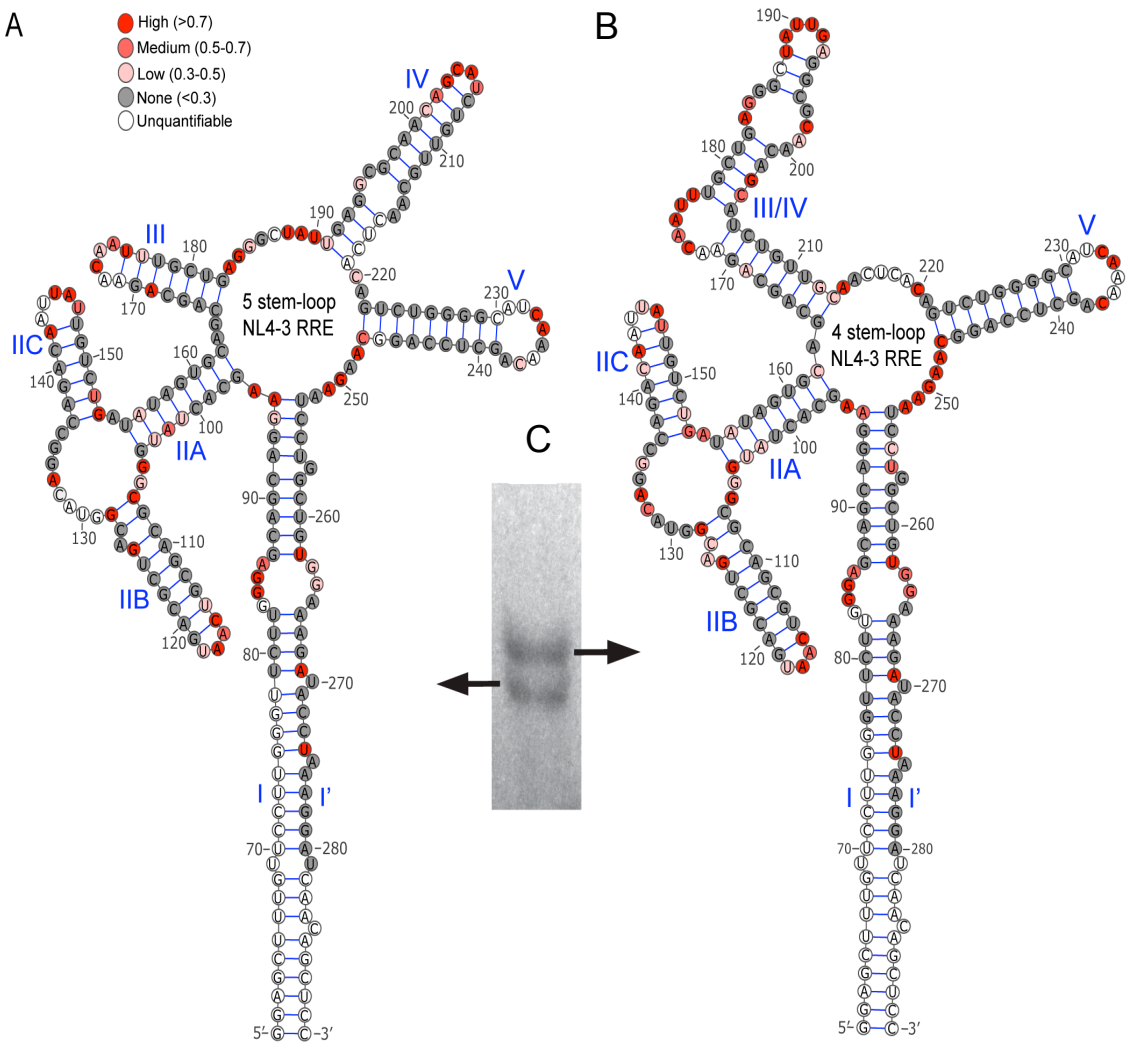


Figure 9: Secondary structure of the 232-nt short NL4-3 RRE determined by in gel-SHAPE.

(A). The energetically most favorable SHAPE structure of the WT RRE (the 5 stem-loop structure) in the faster migrating band shown in part C.

(B). The energetically most favorable SHAPE structure of the WT RRE (the 4 stem-loop structure) in the slower migrating band shown in part C.

(C). The two bands of the RRE on a native polyacrylamide gel visualized by UV shadowing.

Approximately 75 picomoles of *in-vitro* transcribed and PAGE purified RRE RNAs were heated to 85°C in renaturation buffer (10 mM Tris pH 8.0, 100 mM KCl, 0.1 mM EDTA) and slow cooled to 25°C. The folded RNAs were run on a 8% native polyacrylamide gel at 200V at 4°C for about 22 hours in the presence of 5 mM $MgCl_2$ both in the gel and in the 1XTBE running buffer. The two RRE bands were excised and the RREs in the gel matrix chemically modified by 100 mM NMIA. Secondary structure was generated by RNAstructure (v5.5) software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. Representative data from two replicates. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

In-gel SHAPE is a variation of SHAPE, where different structural conformers of the RNA are resolved into different bands on a native gel so that the structure of each conformer trapped within the gel matrix can be probed separately (123). To examine the structure of the RNA in each RRE band, they were both excised separately and subjected to in-gel SHAPE analysis using *N*-methylisatoic anhydride (NMIA) as the SHAPE reagent. The reactivity values of the first 17 nucleotides at the 5' end and last 11 nucleotides at the 3' end of the RRE were unquantifiable due to the loss of signal resolution of the longer cDNA products and abortive transcription initiation by reverse transcriptase. Out of the remaining RRE nucleotides, reactivity values were obtained for > 92 % (Figure 9A and 9B). Almost all of the unscored nucleotides (nt 216 to 219, nt 171-172, nt 129-131, nt 143-144, 231-232) were found consistently unquantifiable, which might reflect some inherent property of the enzyme to pause at these bases, possibly due to higher order RRE structure at these positions. The results indicate that the faster moving band folded as a 5 stem-loop structure, while the slowly moving band folded as the 4 stem-loop structure. Figure 9A and 9B shows the SHAPE reactivity for each base placed on the most energetically favorable RRE structure predicted by RNAStructure 5.5, using the SHAPE reactivity constraints, for each gel band. Thus, we conclude that the NL4-3 RRE can exist as an approximately equal mixture of the two structures.

Creation of RRE mutants with altered secondary structures.

We next attempted to create RRE mutations that favored one or the other of the secondary structures. Mutant A (Figure 10) was expected to disrupt the base pairing at the base of the combined stem-loop III/IV structure in the 4 stem-loop structure, but maintain base pairing in stem-loop IV in the 5 stem-loop structure, and thus likely adopt only the 5 stem-loop structure. Mutant B (Figure 11) was expected to disrupt the base pairing in both stem-loop III and stem-loop IV of the 5 stem-loop structure, but keep the combined stem-loop III/IV intact, and thus likely adopt only a 4 stem-loop structure. Mutant C (Figure 12) created aberrant base pairing within the combined stem-loop III/IV structure and thus likely would form a severely altered stem-loop III/IV but maintain a 4 stem-loop structure due to disruption of base pairing in stem-loop IV. Mutants D (Figure 13) and E (Figure 14) were expected to disrupt both the combined stem-loop III/IV in the 4 stem-loop structure and stem-loop IV in the 5 stem-loop structure and thus likely would not form either the 4 stem-loop or 5 stem-loop structure.

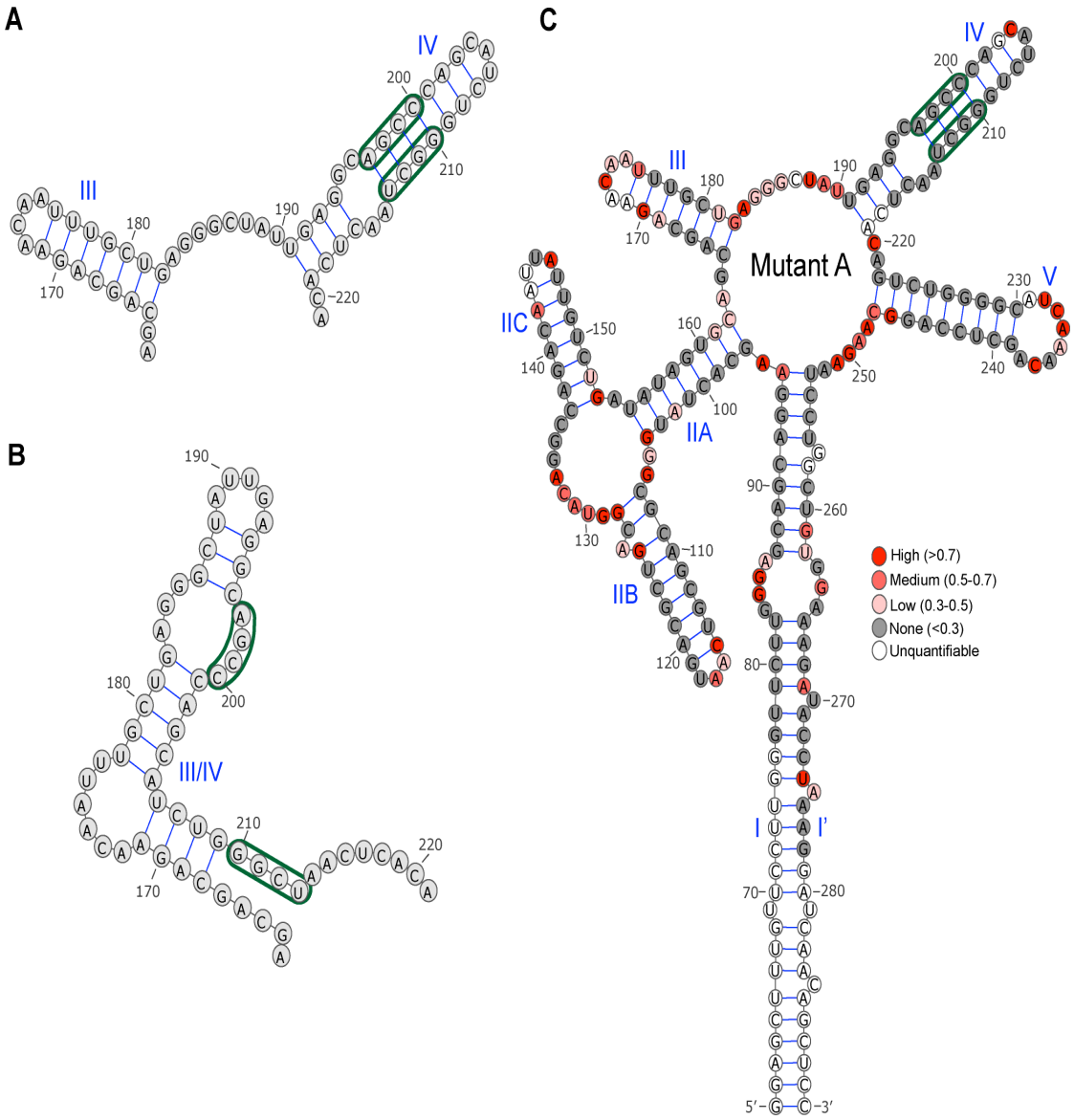


Figure 10: SHAPE derived secondary structure of RRE mutant A.
(A). Mutations of mutant A mapped on the 5 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(B). Mutations of mutant A mapped on the 4 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(C). SHAPE derived secondary structures of RRE mutant A.

In-vitro transcribed mutant A RRE RNA, folded by slow cooling at 25°C, was probed in solution by NMIA. Secondary structure was generated by RNAstructure 5.5 software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. The mutated nucleotides are enclosed by green border. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

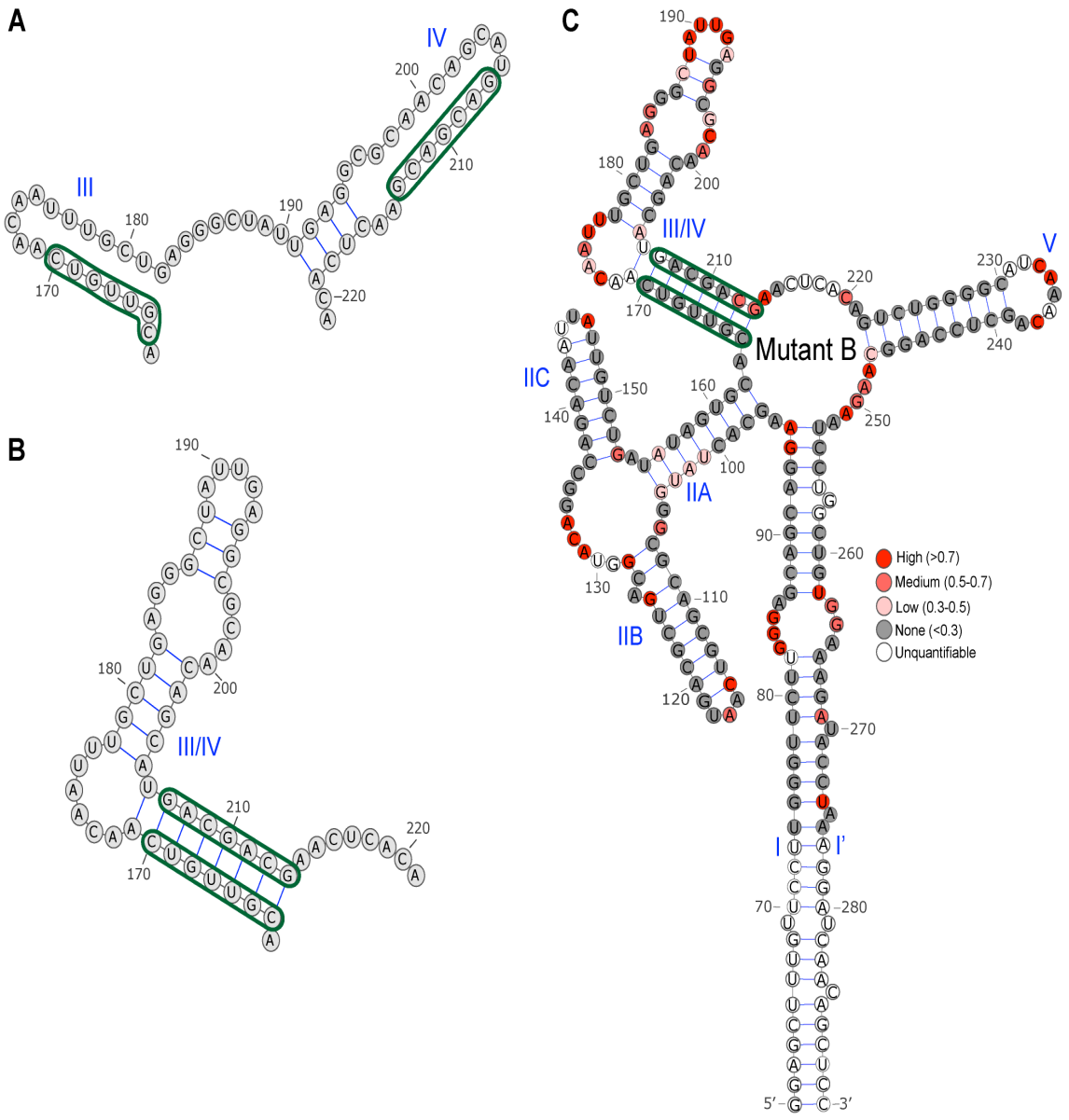


Figure 11: SHAPE derived secondary structure of RRE mutant B.
(A). Mutations of mutant B mapped on the 5 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(B). Mutations of mutant B mapped on the 4 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(C). SHAPE derived secondary structures of RRE mutant B.

In-vitro transcribed mutant B RRE RNA, folded by slow cooling at 25°C, was probed in solution by NMIA. Secondary structure was generated by RNAstructure 5.5 software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. The mutated nucleotides are enclosed by green border. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

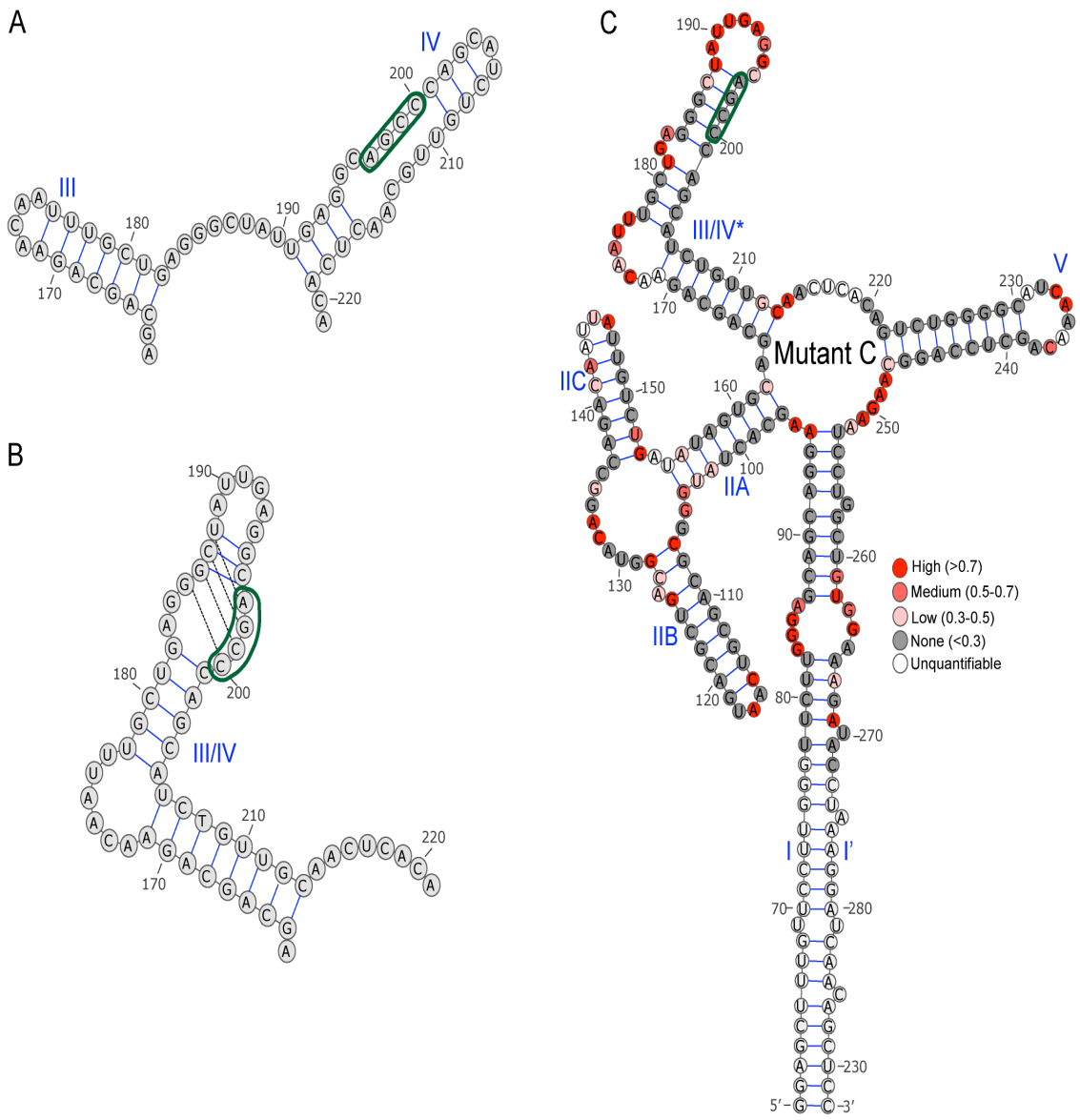


Figure 12: SHAPE derived secondary structure of RRE mutant C.
(A). Mutations of mutant C mapped on the 5 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(B). Mutations of mutant C mapped on the 4 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(C). SHAPE derived secondary structures of RRE mutant C.

In-vitro transcribed mutant C RRE RNA, folded by slow cooling at 25°C, was probed in solution by NMIA. Secondary structure was generated by RNAstructure 5.5 software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. The mutated nucleotides are enclosed by green border. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

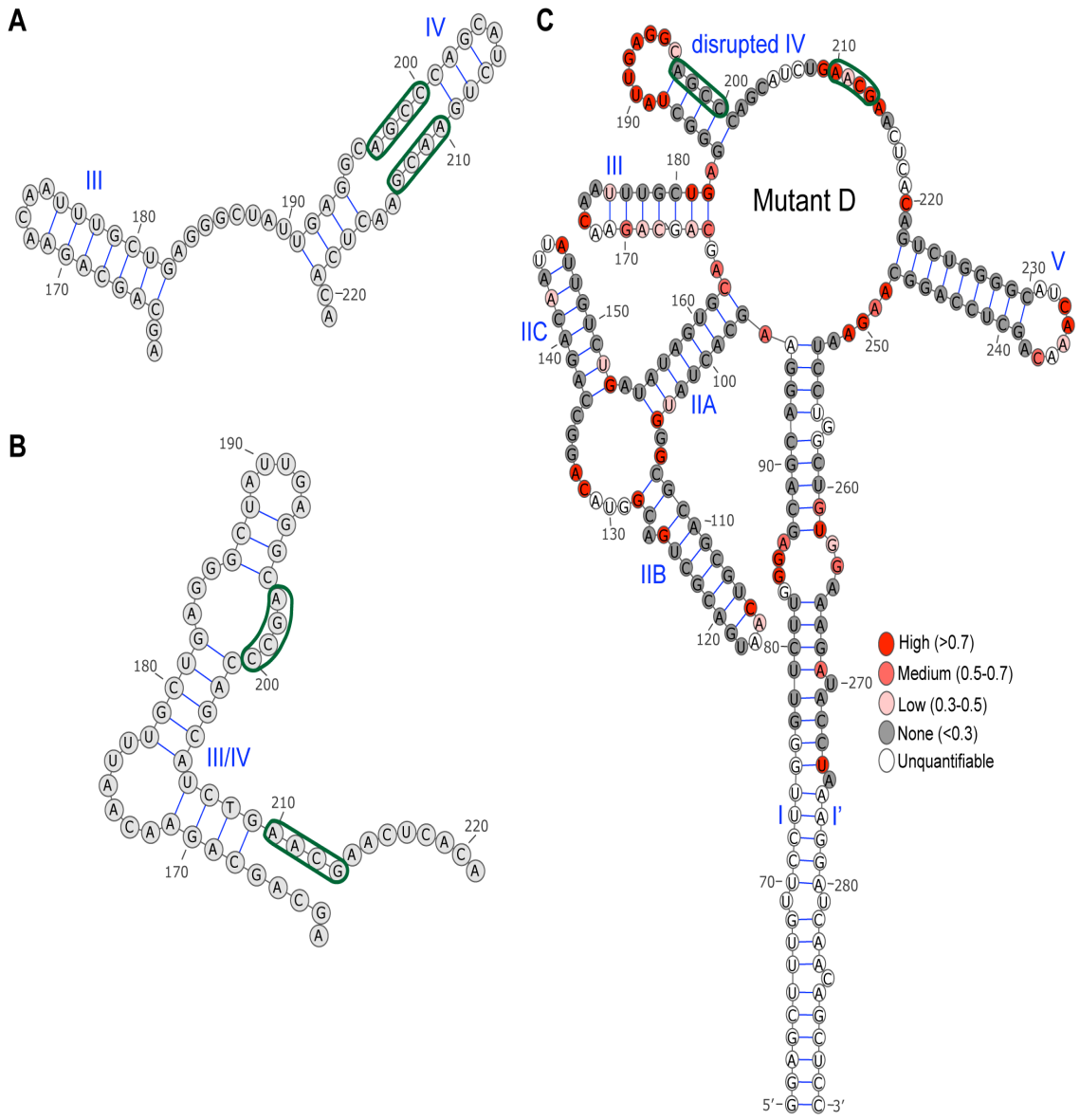


Figure 13: SHAPE derived secondary structure of RRE mutant D.
(A). Mutations of mutant D mapped on the 5 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(B). Mutations of mutant D mapped on the 4 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(C). SHAPE derived secondary structures of RRE mutant D.

In-vitro transcribed mutant D RRE RNA, folded by slow cooling at 25°C, was probed in solution by NMIA. Secondary structure was generated by RNAstructure 5.5 software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. The mutated nucleotides are enclosed by green border. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

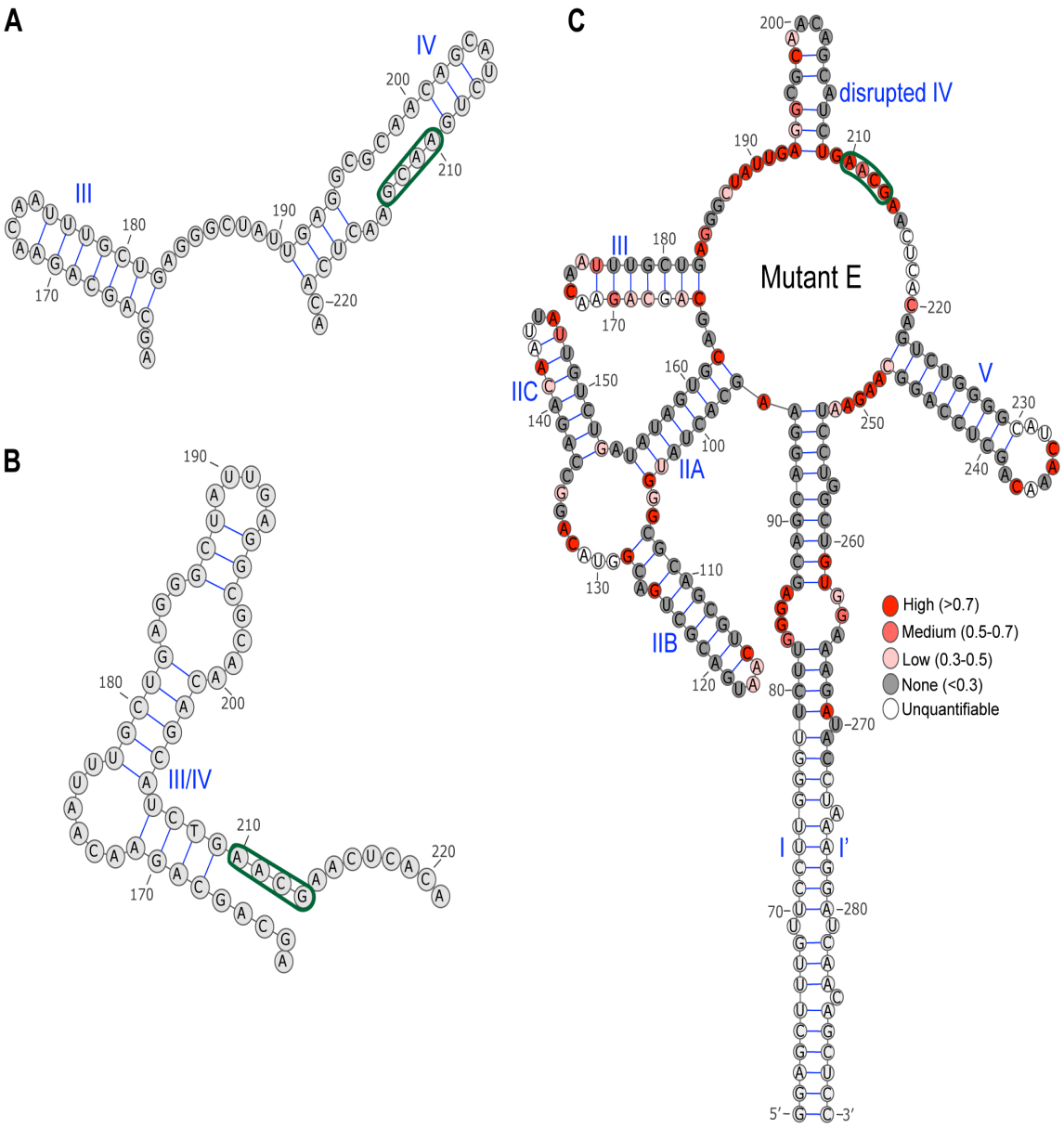


Figure 14: SHAPE derived secondary structure of RRE mutant E.
(A). Mutations of mutant E mapped on the 5 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(B). Mutations of mutant E mapped on the 4 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(C). SHAPE derived secondary structures of RRE mutant E.

In-vitro transcribed mutant E RRE RNA, folded by slow cooling at 25°C, was probed in solution by NMIA. Secondary structure was generated by RNAstructure 5.5 software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. The mutated nucleotides are enclosed by green border. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

Since these mutants were used to study the functional importance of defined RRE structures, it was imperative that all the mutant RRE RNAs included in the study form homogenous structures. To test the structural homogeneity of these mutants, the mutant RRE RNAs, with same 3' tag used with the WT RRE, were prepared by *in vitro* transcription, re-folded and run on a native 8% polyacrylamide gel, together with the WT RRE, using similar conditions utilized for the in-gel SHAPE experiment described above. Figure 15F shows that each of the five mutant RREs formed only a single homogeneous band. Importantly, Mutant A RRE migrated to the position of the 5 stem-loop WT RRE and the mutant B RRE migrated to the same position as the 4 stem-loop WT RRE as predicted. Mutant C migrated close to the mutant B band and mutants D and E migrated to higher positions on the gel. On one hand, this experiment indicated that each of the mutant RREs formed only one structure. On the other hand, it also provided preliminary evidence that these mutants indeed form their respective predicted structures.

Since all the mutant RRE RNAs appeared structurally homogenous on the native gel, we used conventional SHAPE to determine the secondary structure of each of these mutants (Figures 10-14; Figure 15A-E).

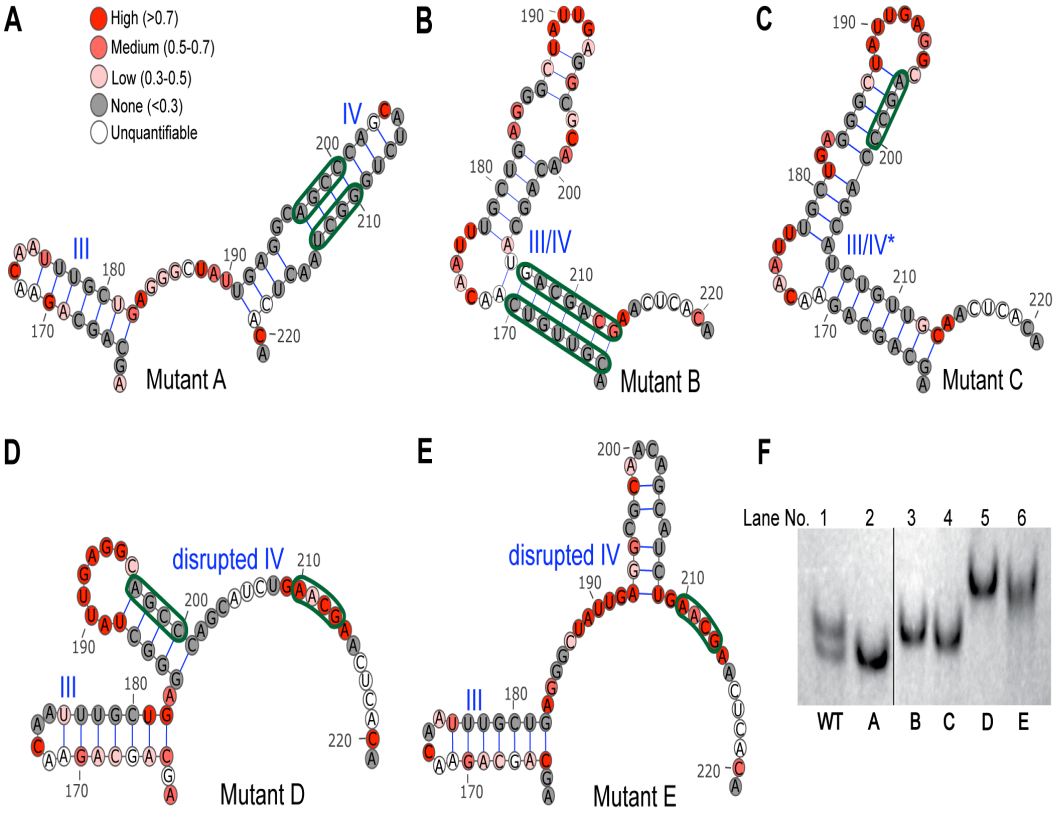


Figure 15: SHAPE Structure and native gel migration pattern of all the RRE mutants.

(A). SHAPE derived secondary structures of mutant A.
(B). SHAPE derived secondary structures of mutant B.
(C). SHAPE derived secondary structures of mutant C.
(D). SHAPE derived secondary structures of mutant D.
(E). SHAPE derived secondary structures of mutant E.
In Figure 15 A-E, only the stem-loop III and stem-loop IV regions are shown; the SHAPE structure of the other regions of the RRE were identical to that of WT RRE.
(F). Migration pattern of the RRE mutants along with WT RRE bands on a native polyacrylamide gel visualized by UV shadowing.
Approximately 20 picomoles of in-vitro transcribed and PAGE purified mutant RRE RNAs were heated to 85°C in re-naturation buffer (10 mM Tris pH 8.0, 100 mM KCl, 0.1 mM EDTA) and slow cooled to 25°C. The folded RNAs were ran on a 8% non-denaturing polyacrylamide gel for at 200V at 4°C in the presence of 5mM $MgCl_2$ both in the gel and in the 1XTBE running buffer. The folded RNAs were probed in solution by 30 nM NMIA. Secondary structure was generated by RNAstructure 5.5 software under SHAPE reactivity constraint using a slope of 1.8 kcal/mol and an intercept of -0.6 kcal/mol. Consistent data (not shown) were obtained under two other conditions 1. RNAs were folded by snap cooling at 4°C and NMIA was used as the electrophile and 2. RNAs were folded by snap cooling at 4°C and 1M7 was used as the electrophile. [SHAPE experiments and data analysis by Chringma Sherpa and Jason Rausch (Le-Grice's Lab, NIH)]

Graphs showing the NMIA reactivity plots for all of the RREs are shown in Figure 16. Essentially, the same results were obtained when these RNAs were folded by snap cooling and probed with either NMIA or 1M7 (data not shown). Hence, both the non-denaturing gel-shift data and the SHAPE analysis data supported the predicted secondary structure of the RRE mutants.

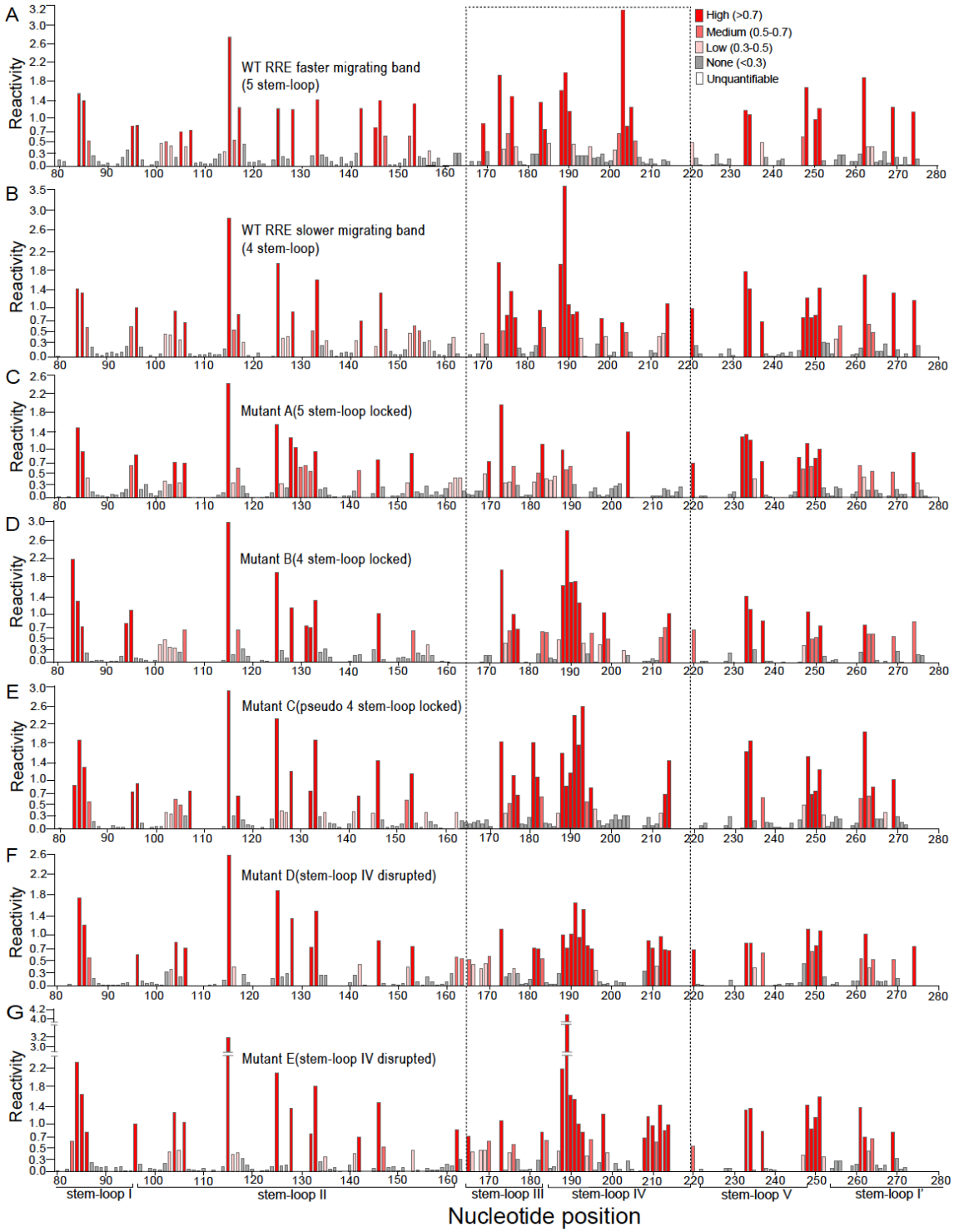


Figure 16: Comparison of the SHAPE reactivity plots of the two structural forms of the WT RRE derived by in-gel SHAPE (A, B) and of the mutant RREs derived by conventional SHAPE (C-G).

Primary binding of Rev to the RRE and multimerization are largely not affected by the alterations in stem-loop III/IV secondary structure.

We next tested the ability of each RRE mutant to bind Rev *in vitro* and act as a scaffold for Rev multimerization. To do this, increasing amounts of NL4-3 Rev protein, purified from *E.coli*, were incubated in a binding reaction at 4°C with a constant amount of the different radiolabeled RREs. The resulting complexes were separated on a polyacrylamide gel and detected using a phosphorimager (Figure 17).

In order to establish the specificity of our binding assay, Rev protein was incubated with the WT NL4-3 RRE or with an RRE containing a mutation (A131G) in stem loop IIB, the primary Rev binding site. The nt-131 of the RRE is one of the four nucleotides involved in forming the two non-Watson Crick base pairs that mediate primary Rev binding. This A131G RRE mutation has been shown to reduce the Rev affinity by 40-100 fold (51, 117). It is also reported to impair Gag expression from pCMV-GagPol construct by ~3.3 fold (51). In our study, this mutant RRE was used as a negative control for primary Rev binding. Primary binding of Rev to the WT RRE, as measured by the formation of slower migrating complex was observed with 25 ng of Rev in the reaction mix. Slower migrating bands started to appear at 50 ng. In contrast, for the mutated RRE, weak Rev binding was only detected at the highest Rev concentrations (Figure 17).

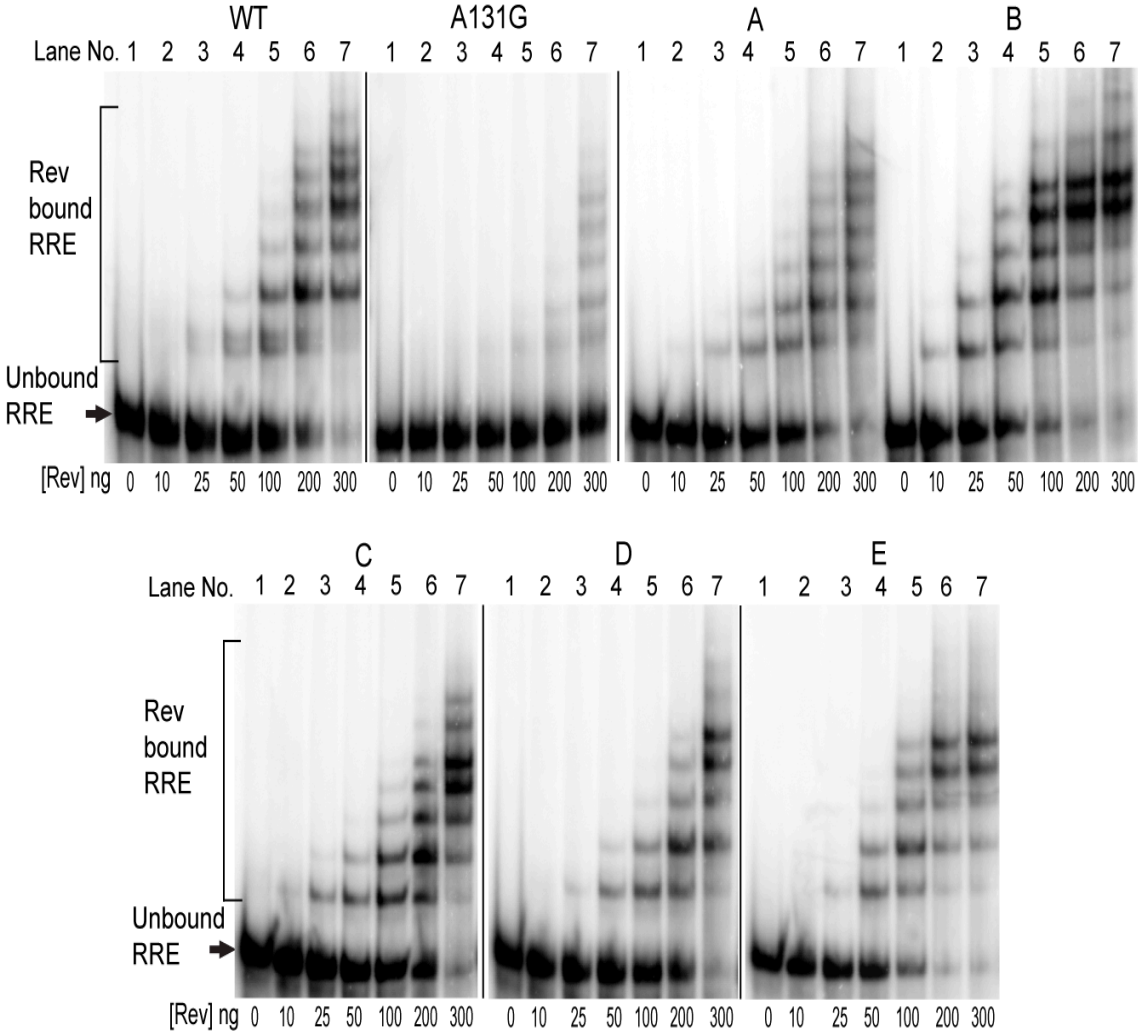


Figure 17: Rev primary binding and multimerization on the 234-nt RRE RNAs determined by electrophoretic mobility shift assay (EMSA). In-vitro transcribed and PAGE purified 1 nM α^{32}P-internally labeled RRE RNA was heated at 85°C for 3 min in renaturation buffer (50 mM NaCl, 10 mM Hepes/KOH, pH 7.6, 2 mM MgCl2) followed by cooling at room temperature for 15 min. The folded RNA was then either incubated without Rev or with increasing concentration of bacterially derived purified, codon optimized GB1-full-length NL4-3 Rev protein in Rev binding buffer (10 mM Hepes/KOH, pH7.8, 20 mM KCL, 2 mM MgCl2, 0.5 mM EDTA, 1 mMDTT, 10% glycerol, 5 µg/ml yeast tRNA and 20 U RNase inhibitor) on ice for at least 10 minutes. The Rev-RRE complexes were resolved on a 4% native polyacrylamide gel run at 20 mA constant current for 4 hours at 4°C. Representative EMSA gel-image shown. N (independent experiments) = 5 for WT, 1 for A131G, 4 for A, 4 for B, 3 for C, 4 for D, 5 for E.

Having validated the binding assay, we next examined the binding of NL4-3 Rev to each of the RRE mutants (A-E). These assays were repeated multiple times with different preparations of Rev protein and RRE and each time essentially the same results were obtained. The images shown in Figure 17 are representative of the data (see figure legend). Examination of the different gels reveals that within the limitations of this assay, very little difference in either primary binding or multimer formation was observed. However the pattern observed with the WT RRE clearly shows the presence of noticeable doublet bands, especially in the faster migrating Rev-RRE complexes. These doublet bands might have resulted from the two structural conformers of the WT RRE. In the slower migrating complexes, the higher molecular mass of the complexes might have prevented the resolution of the complexes differing only in the RRE structure. The doublet bands, however, were not present in the mutants. In addition, a comparison of the mutant A and mutant B gel shifts shows reproducible spacing differences that are especially obvious between the lower bands. Thus these results are consistent with our finding that the WT RRE is composed of a mixture of 4 stem-loop and 5 stem-loop structures, while each of the mutant RREs has a single structure. The results also strongly suggest that there is little difference in the ability of each RRE structure to bind Rev and act as a platform for Rev multimerization, although small differences in the rates of migration of the comparable complexes formed by each RRE can be observed.

Alterations in SLIII/IV secondary structure affect the ability of the RRE to mediate expression of HIV GagPol from a reporter construct.

Each RRE was then tested for its ability to provide functional activity by inserting it into a GagPol reporter construct that has been used previously to quantitate Rev/RRE function (214). In this assay, each HIV-1 GagPol reporter plasmid containing the different mutant RREs was transfected into 293T cells, together with differing amounts of a plasmid that expressed the NL4-3 Rev protein. A third plasmid that expressed secreted alkaline phosphatase (SEAP) was also transfected as an internal control (see Materials and Methods). Under these conditions, Rev/RRE activity could be quantified by measuring the amount of p24 capsid protein secreted into the media. The results of this experiment (Figure 18) show that each RRE promoted a different level of Gag and GagPol expression in response to each concentration of Rev plasmid used. Thus, we conclude that although the results of the *in-vitro* Rev binding assay revealed no reproducible differences in Rev binding to the various RRE mutants, the alterations made to RRE secondary structure by the mutations affected their activity in cells. In this assay the RRE with 5 stem-loop structure (mutant A) appeared to be more active than WT RRE and the RRE with the 4 stem-loop structure (mutant B) appeared to be less active. The mutants that severely altered the stem-loop III/IV structure (mutant C) or that abolished stem-loop IV (mutant D and mutant E) were even less active.

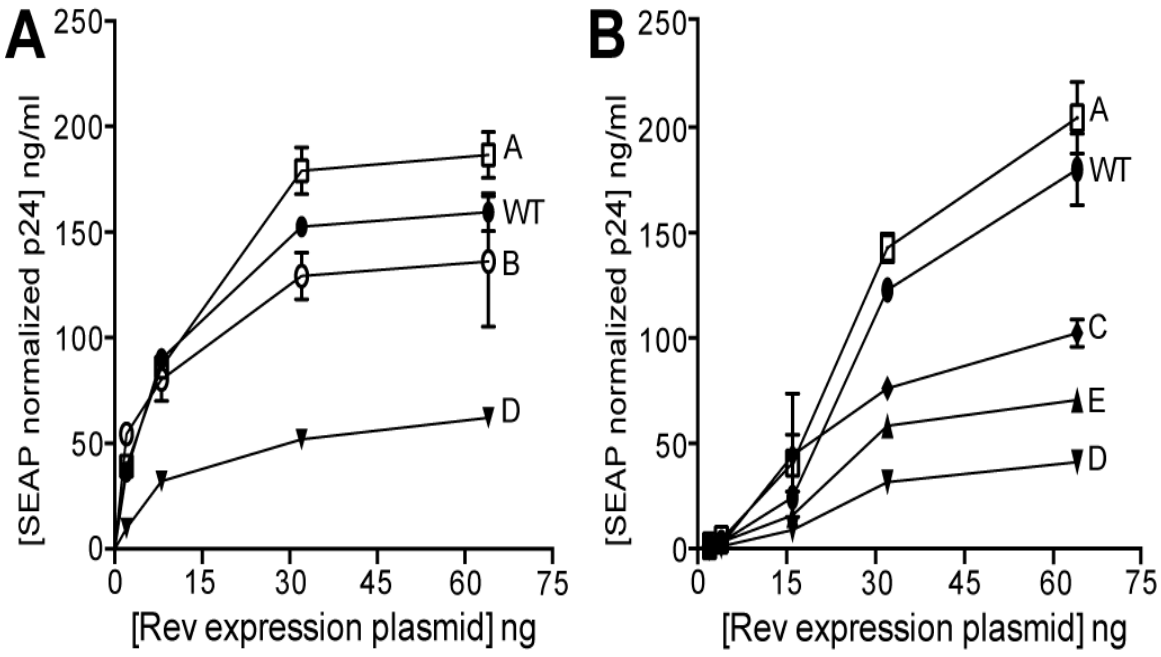


Figure 18: Rev dose response curves for WT and mutant RREs. 2 X 10^5 293T cells were co-transfected with 2 µg of pCMVGag-Pol reporter plasmid containing the WT or mutant RRE, increasing concentration (0-32ng) of codon optimized NL4-3 Rev expression plasmid, 100 ng of SEAP expression plasmid, and compensatory amount of empty plasmid. The p24 and SEAP levels in the culture medium were measured 48 hours post transfection. Error bars represent standard error of mean from two replicates. **A and B panels represent two independent experiments.**

Alterations in RRE stem-loop III/IV secondary structure affect HIV replication rates in a spreading infection.

We next tested the ability of each mutant RRE to promote HIV replication in the context of a spreading infection in SupT1 cells. Normally mutations in the RRE would be expected to alter or disrupt the envelope protein coding region, making a comparison of the replication rates of viruses with different RRE mutations problematic. To overcome this problem, we inserted each RRE into the Nef region of an NL4-3 proviral clone that had its normal RRE inactivated by mutations that did not change the envelope protein sequence. Since Nef is not required for HIV replication in SupT1 cells, this allowed us to create a series of replication competent proviral clones that were isogenic except for the mutations contained within the RRE that was being tested. These proviral clones were transfected into 293T cells and all produced enough p24 to allow us to prepare viral stocks for infection of SupT1 cells. SupT1 cells were then infected with each transfection supernatant containing equal amounts of p24. The cells were maintained in long-term culture by replacing three-fifths of the culture every 3rd or 4th day with the fresh medium and p24 released into the supernatant was monitored. The results of this experiment show that each mutant virus replicated at a different rate, with the order of efficiency being mutant A > WT >mutant B >mutantC. Additionally, no detectable replication was observed for either mutant D or mutant E. These results were the same whether 300 ng (Figure 19A) or 100 ng (Figure 19C) of input p24 were used.

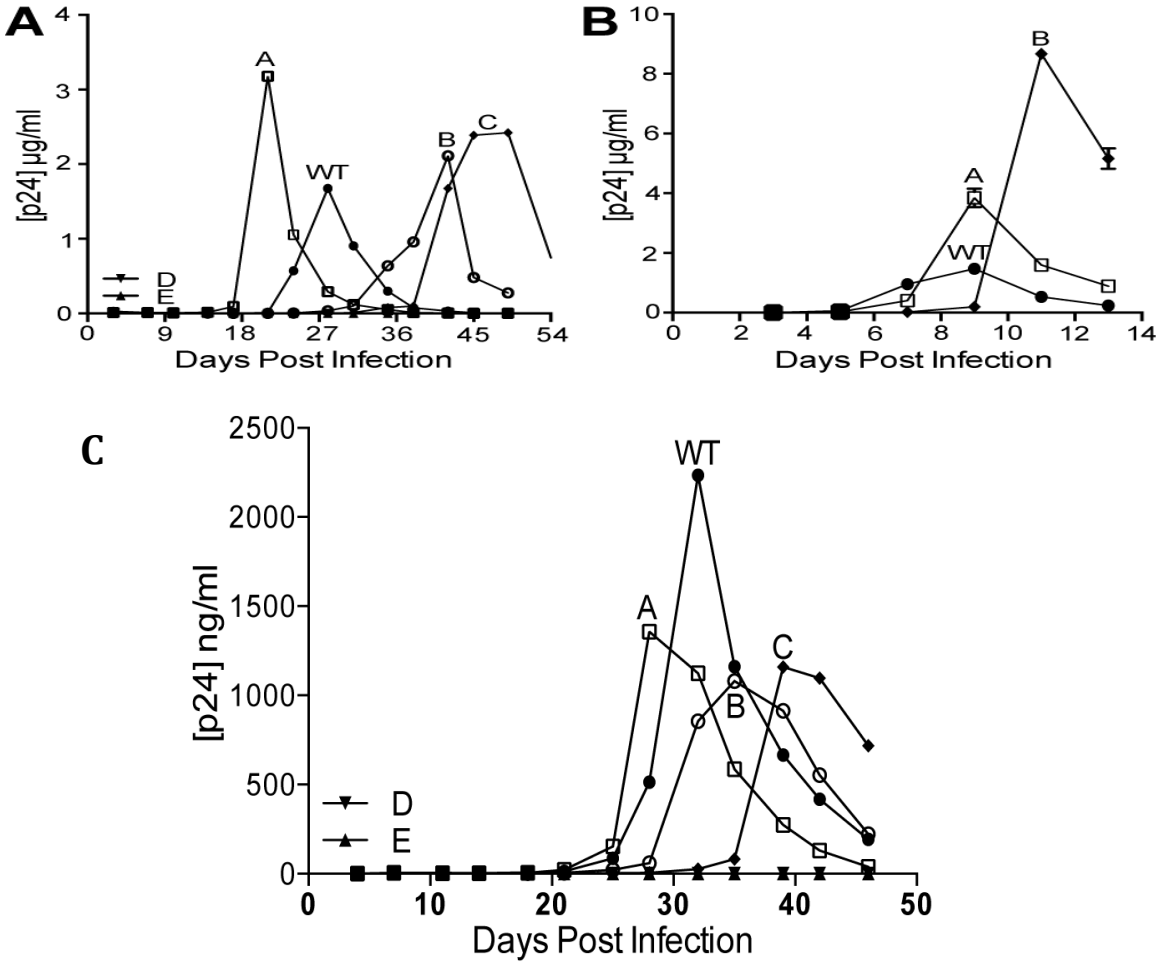


Figure 19: Replicative fitness of HIV with WT or mutant RREs in spreading infection assay.

(A). Growth curves using 300 ng p24 equivalent of transfection viral stocks. 300 ng p24 equivalent of viral stocks, prepared by transfection of 293T cells with 5 µg NL4-3 Nef- RRE-/long RRE plasmids, were used to infect 6 X 10^6 SupT1 cells. The peaks of viral replication were monitered by determining the p24 levels in the infection supernatant after every 3-4 days

(B). Growth curves using 5 X 10^{-6} MOI of SupT1 passaged viral stocks from the peak days of Figure 19A. Fresh 6 X 10^6 SupT1 cells were infected with 5 X 10^{-6} MOI (100 ng) of SupT1 passaged viral stocks using DEAE-dextran. The peaks of viral replication were monitered by determining the p24 levels in the infection supernatant after every 2-3 days.

(C). Growth curves using 100 ng p24 of transfection viral stocks. Viral stocks were prepared by transfection of 293T cells with NL4-3 Nef- RRE-/long RRE plasmids. 100 ng p24 equivalent of viral stocks were used to infect 6 X 10^6 SupT1 cells by DEAE-dextran method. The peaks of viral replication were monitered by determining the p24 levels in the infection supernatant after every 3-4 days.

To further compare the replication efficiency of some of these viruses, the peak fractions of the SupT1 supernatants from the WT, mutant A and mutant B infections shown in Figure 19A were harvested and used to infect fresh SupT1 cells. Infections were done with $5X10^{-6}$ MOI of input virions (p24 equivalent). Essentially, the same relative replication kinetics were observed with these viral stocks (Figure 19B). Thus, in accordance with the data obtained from the p24 reporter assay (Figure 18), this experiment shows that the secondary structure of stem loops III and IV strongly influence the efficiency of viral replication. The data also show that the 5 stem-loop RRE promotes HIV replication in SupT1 cells more efficiently than either the WT or the 4 stem-loop RRE. A comparison of the growth curves generated using 100 ng p24 equivalent transfection viral stocks (Figure 19C) and SupT1 passaged viral stocks (Figure 19B), showed that the peak replication day was reached much later in the former. Such difference in the replication peak days might indicate the possibility of a higher p24/infectivity ratio in the transfection viral stocks.

Abrogation of a correct SLIII/IV secondary structure in the RRE diminishes the titer of a Rev-dependent HIV vector.

We next tested each of the mutated RREs in the context of a modified proviral vector (pTR167nef$^-$). This vector (Figure 20A) was created by deletion of 5,491 nucleotides from the central region of pNL4-3, followed by the insertion of an SV40 early promoter-driven hygromycin resistance gene (Hyg^r) in the middle of the *env* region. It also has a small deletion that inactivates the Nef gene and

the vector transcribes a modified HIV genomic RNA whose trafficking is dependent on Rev-RRE function. When co-transfected with an HIV packaging system, the genomic RNA gets packaged into virus-like particles that are capable of transducing Hyg^r to target cells. Thus the magnitude of the virus titer derived from the packaging cell is directly reflective of Rev-RRE functional activity.

Each mutant RRE was exchanged with the WT NL4-3 RRE to create a series of isogenic vectors, which differed only in the RRE mutations they contained. In these vectors, each RRE resided in its normal position within the *env* intron. The vectors were transfected into 293T cells together with plasmids that supplied Gag/GagPol, Tat, Rev and VSV-G. Supernatants were collected after 48 hrs, assayed by p24 ELISA, and titered for hygromycin resistant colony formation on HeLa cells. Figure 20B shows the results of this experiment. It is clear that disruption of the normal stem-loop IV structure (mutants C and E) greatly decreased the vector titer. However, titer differences between the WT and the A and B mutants did not reach statistical significance, although the data trended towards Mutant A having a slightly higher titer than Mutant B. Thus, this experiment clearly shows that the stem-loop III/IV region, a region lying outside of the primary Rev binding site, must be folded into either the 4 stem-loop or 5 stem-loop structure to allow efficient RNA trafficking and packaging into virus particles, but it did not show a clear difference in activity between the WT and the mutants with the 4 stem-loop and 5 stem-loop structures.

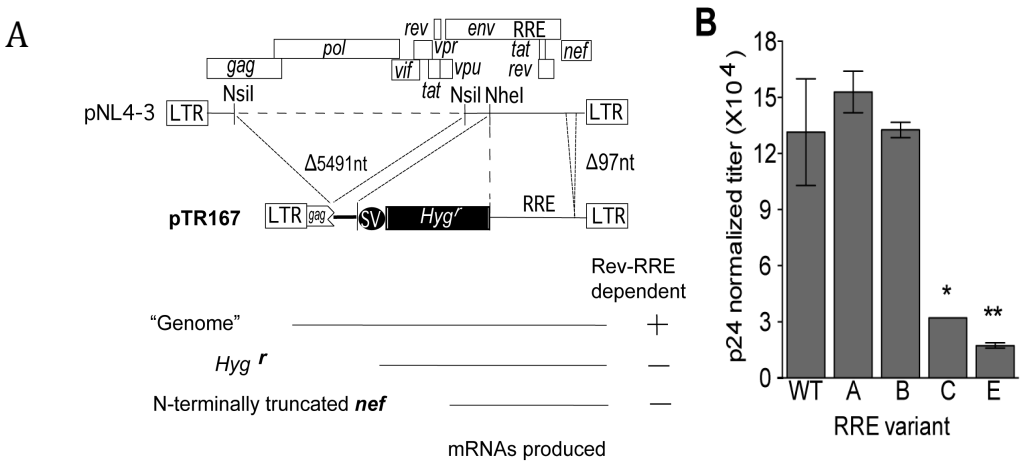


Figure 20: Packaging efficiency of virus with WT or mutant RREs.
(A). Schemetic representation of pTR nef- construct. The construct was derived from pNL4-3 by deleting a part of *gag, env* (excluding the RRE region) and *nef*, and the entire *pol*. A cassette containing the hygromycin resistant gene (hyg^r) driven by SV40 early promoter-enhancer was inserted in the remaining *env* sequence. Each test RRE was cloned into the native RRE position.
(B). p24 normalized - hygromycin resistant colony titer for each RRE. The viral stocks were prepared by co-transfecting the producer 293T cells with pTR nef- transducing plasmid and two helper plasmids (pCMVΔR9 and pCMV_VSV-G). 1 ml of serially diluted transfection supernatant was used to infect 5 X 10^5 Hela cells and the cells were hygromycin selected from day 3 to day 15 post-infection and the hygromycin resistant colonies were counted The statistical significance of the differences in the relative activities of the RRE mutants with respect to wt RRE was determined at 99% confidence level by one way unrepeated ANOVA assuming homoscedasticity and corrected for multiple comparison error by two-sided Dunnett post-hoc test. Here * denotes p value < 0.05, ** denotes p value < 0.01.

Competition growth experiments demonstrate that a virus containing a 5 stem-loop RRE is more fit than a virus contains a 4SL RRE.

The data presented thus far suggest that the 5 stem-loop RRE is a more active form of RRE compared to the 4 stem-loop and WT structures. However, the differences observed are slight, so it was of interest to directly compare the viruses capable of forming only the 4 stem-loop or 5 stem-loop RRE to each other and to WT in competition growth assays. To perform these assays, two SupT1-passaged viruses, each with a different RRE inserted into the Nef region, were added to the same culture of SupT1 cells at an equal and low multiplicity of infection. Input virions for the infection were normalized for MOI. The viruses were then allowed to replicate and spread throughout the cultures until all cells became infected. At that time, the cultures were harvested and the amount of proviral DNA that was produced by each virus was measured using a PCR-based heteroduplex tracking assay (see materials and methods). The assay exploits the fact that each of the mutants differ significantly in sequence from each other and wild type in the stem-loop III/IV region. As a result of these sequence differences, heteroduplexes, formed by a PCR product amplified from the stem-loop III/IV region in each proviral DNA and a ^{32}P-labeled WT probe, migrate significantly differently on a polyacylamide gel.

To test the resolution of this assay, we performed a control experiment (Figure 21) where different ratios of proviral constructs (pNL4-3 Nef-/long RRE plasmids) containing the test RRE (WT or mutant A or mutant B) were mixed.

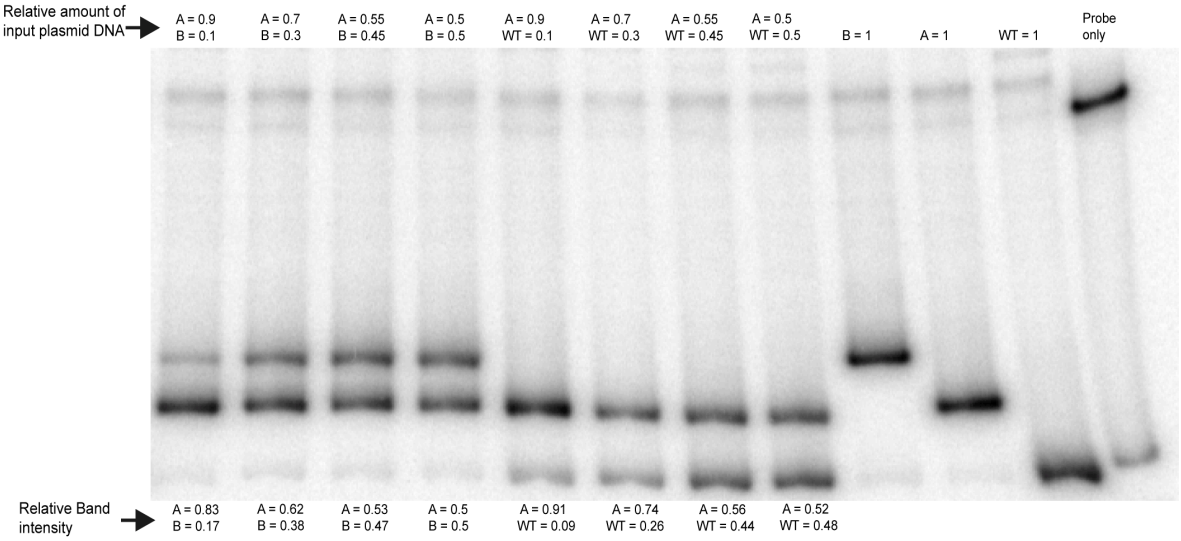


Figure 21: Proof of principle HTA experiment. Different ratios and combinations of WT, mutant A, and mutant B RRE containing proviral plasmids (pNL43 nef- RRE-/long RRE) were annealed to γ^{32}P-end labeled WT RRE probe and the mixtures were run on a 8% native polyacrylamide gel run at room temperature for 3 hours.

The oligos used to amplify the RRE region from the proviral DNA in the competition assay were then used to amplify the same RRE region from each of the plasmid mixtures. The resulting PCR amplicons were annealed to the WT probe, and run on a native gel. As shown in Figure 21, the relative intensities of the heteroduplexes formed by the PCR product of a plasmid mixture were equal or almost equal to the relative ratio of the corresponding plasmids in that mixture. Thus, our competition assay protocol could be reliably used to determine the amount of proviral DNAs coming from different viruses in the dual infections. Figure 22A and 22B show the results of several representative experiments. The left most lanes in each figure show the gel bands generated from mono infections of the WT, mutant A and mutant B viruses and establish the migration position of each specific heteroduplex. The reminder of each figure shows the various dual infection competitions with infection by MOI equivalent of two viruses. Each of the specific bands produced by each virus in the dual infections was then quantified and used to calculate a relative replicative “fitness” in a manner similar to assays that have been previously used to determine the replicative “fitness” of viruses between different HIV clades. From these results we conclude that the 5 stem-loop RRE conveyed a selective growth advantage over both the WT and 4 stem-loop structures and that the WT RRE conveyed a selective growth advantage over the 4 stem-loop structure.

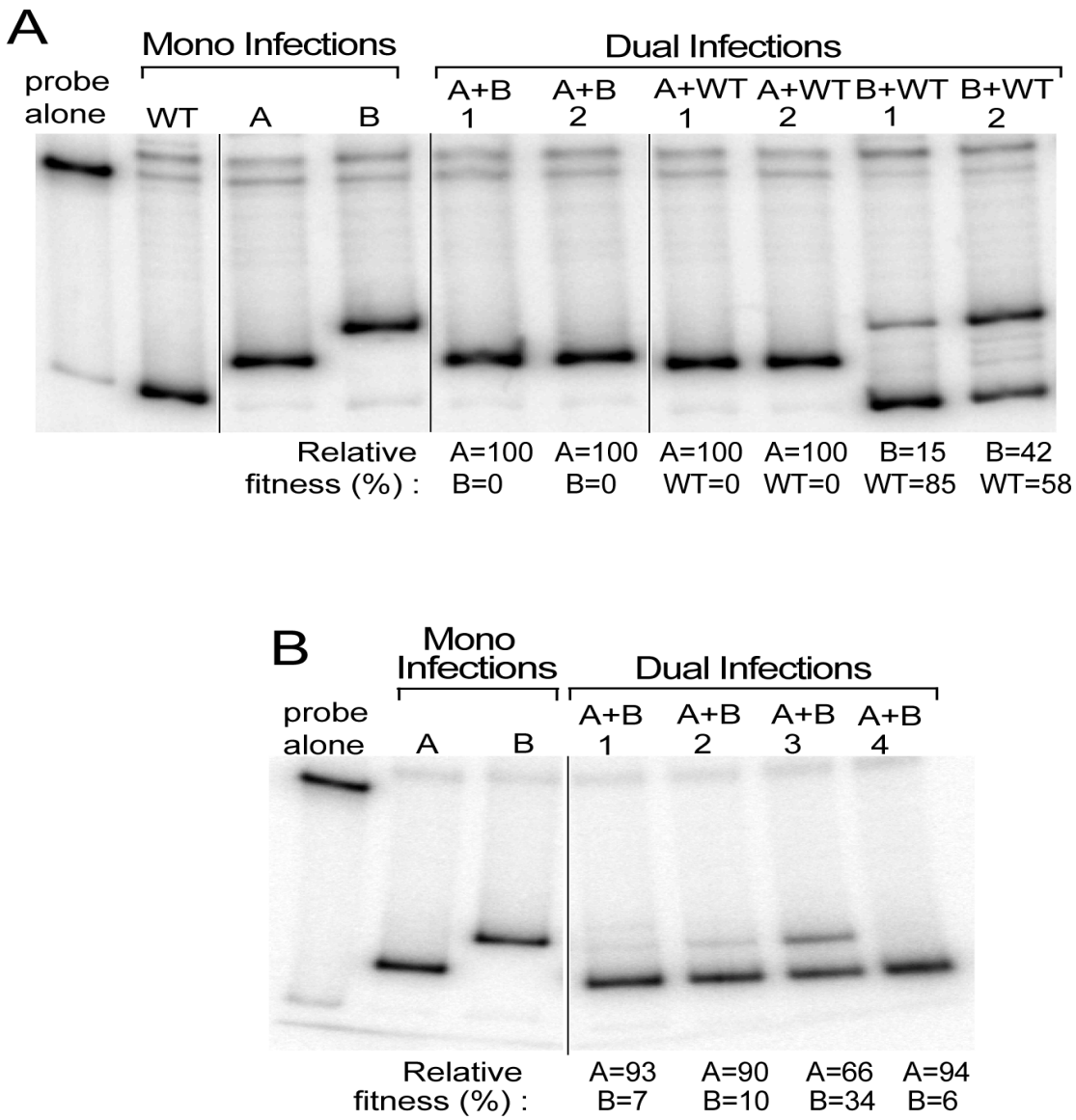


Figure 22: Replicative fitness of HIV with WT or mutant RREs in growth competition /HTA assay. Proviral DNA bands of WT, mutant A, and mutant B RREs from the competition assay are shown. SupT1 cells were dually infected with 0.000005 MOI of each SupT1 passaged viruses in three combinations namely 1. virus with the WT RRE and virus with the mutant A RRE, 2. virus with the WT RRE and virus with the mutant B RRE, and 3. virus with the mutant A RRE and virus with the mutant B RRE. The proviral DNA was extracted 10 days post infection and quantitated by HTA using γ^{32}P-end labeled WT RRE probe on a 8% native polyacrylamide gel run at room temperature for 3 hours. (**A) and (B) panels represent two independent experiments carried out using two independently produced set A and set B viruses. (See materials and methods)**

Table 4: Summary of results

	Predicted structure	RRE native gel migration	SHAPE structure	Rev-RRE gel shift	GagPol expression	Spreading infection	Packaging efficiency	Growth competition
WT	both 5 and 4	doublet (both 5 and 4)	both 5 and 4	same Rev affinity /doublet seen				
A	5	5	5	same Rev affinity/	slightly more active than WT	replication peaks in the order A>WT>B>C	almost same as WT	A>WT>B
B	4	4	4	same Rev affinity	slightly less active than WT	replication peaks in the order A>WT>B>C	almost same as WT	A>WT>B
C	4	close to 4	pseudo 4	same Rev affinity	much less than WT	replication peaks in the order A>WT>B>C	much less than WT	A>WT>B
D	neither 5 nor 4	much slower than either 5 or 4	neither 5 nor 4	same Rev affinity	much less than WT	does not peak	much less than WT	
E	neither 5 nor 4	much slower than either 5 or 4	neither 5 nor 4	same Rev affinity	much less than WT	does not peak	much less than WT	

Structural and functional analysis of mutant F

Mutant F was one of the RRE mutants that was designed to form only the 5 stem-loop structure. This mutant was created by introducing mutations at the top stem of stem-loop IV of the 5 stem-loop structure such that base-pairing in this region remained intact. These mutations, however, were expected to disrupt the base-pairing at the bottom stem of the combined stem-loop III/IV of the 4 stem-loop structure - thus favoring the formation of the 5 stem-loop structure alone (Figure 23). However, when analyzed, this mutant migrated as two distinct bands on a native polyacrylamide gel (Figure 24), suggesting that the RRE folded into two distinctly different conformers. This mutant is a good example of

why RNA secondary structure prediction solely on the basis of base-pairing can fail at times. Therefore, RNA structure prediction should be confirmed using biochemical and biophysical methods. However, the faster migrating band of this mutant RRE did have a migration rate equal to the 5 stem-loop band of the wt-RRE. Because of the structural heterogeneity of this mutant, the experimental data obtained with it is difficult to interpret. However, for the record, it is presented separately in this section.

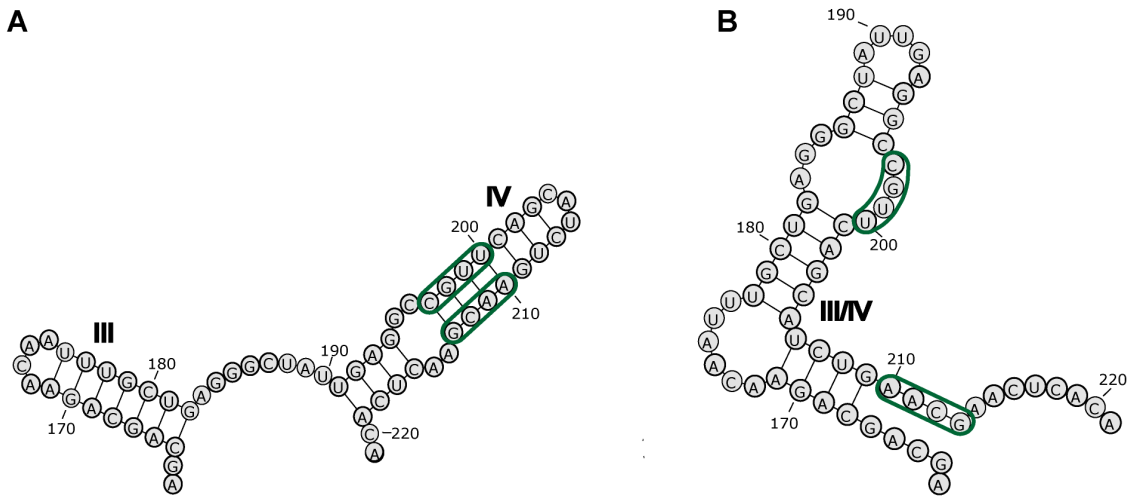


Figure 23: Predicted secondary structure of RRE mutant F.
(A). Mutations of RRE mutant F mapped on the 5 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).
(B). Mutations of RRE mutant F mapped on the 4 stem-loop WT RRE structure (only the stem-loop III and stem-loop IV regions are shown).

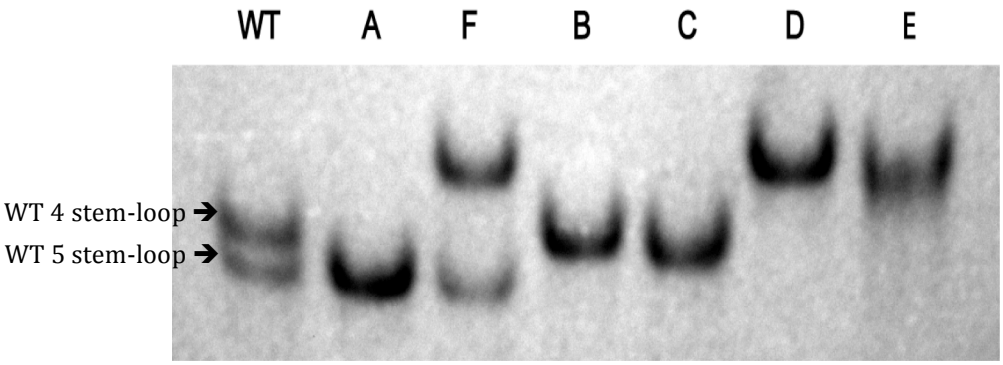


Figure 24: Comparison of the migration pattern of mutant F RRE RNA with other RRE RNAs on a native 8% polyacrylamide gel. (Data provided by Jason Raush, (LeGrice lab, NIH)

Mutant F was tested in the transient transfection assay using GagPol constructs and we observed that mutant F was significantly more active than the RRE mutants with disrupted stem-loop III and IV (mutant D and mutant E) at 32 ng and 64 ng of Rev expression plasmid (Figure 25A). Furthermore, the activity level of this RRE was very comparable or even higher than mutant A. This indicated that even though this RRE did not exclusively form the 5 stem-loop structure, its activity level was similar to the 5 stem-loop locked RRE.

Although we were able to observe a clear difference in the Rev-RRE activity among the RRE mutants (including mutant F) in the Rev dose response GagPol expression assay and in our viral replication studies, it was puzzling that we failed to see any difference in particle production when 293T cells were transfected with the proviral constructs containing the mutant RREs (Figure 25B). We hypothesized that, in the latter assay, there was high overexpression of Rev since very high numbers of genomes get expressed during transient expression using the proviral clones. As the proviral constructs contain an intact Rev ORF, the level of Rev protein cannot be titrated independently of the proviral plasmid. In contrast, in the GagPol reporter assay, Rev can be expressed at much lower levels since it is made from a separate plasmid whose concentration can be greatly reduced (ng versus μg).

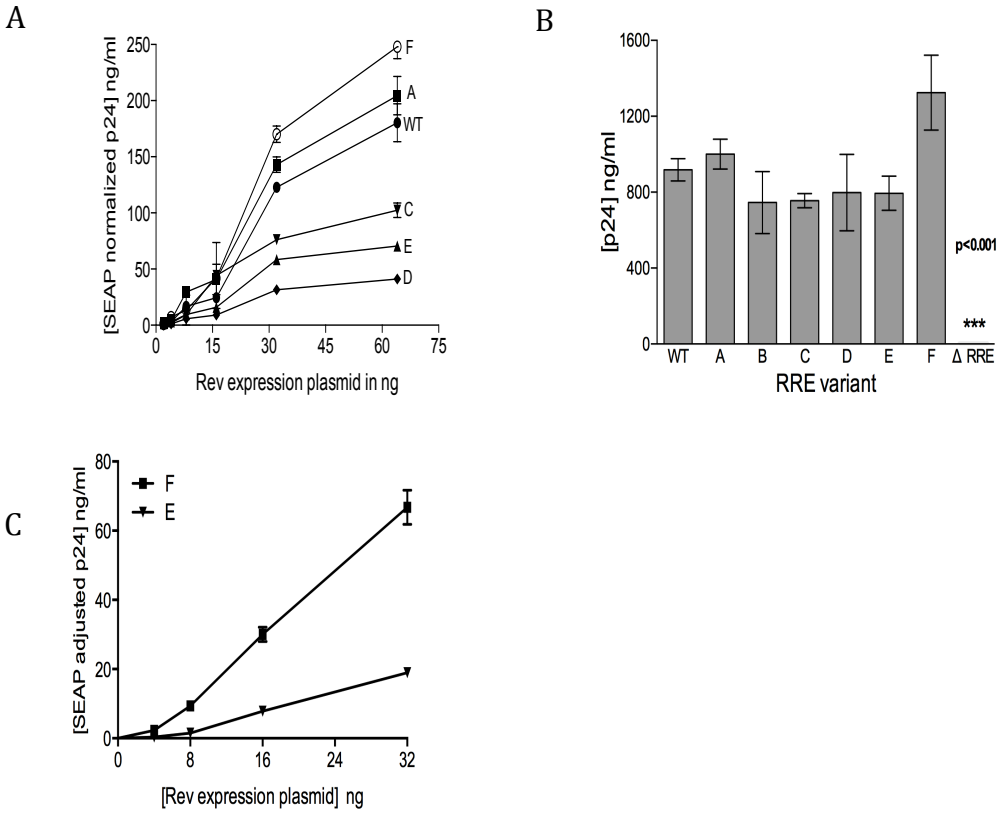


Figure 25: Functional assays of mutant F.

(A). Rev dose response curves for mutant F along with other RREs generated from GagPol reporter assay. 2 X 10^5 293T cells were co-transfected with 2 µg of pCMVGag-Pol reporter plasmid containing the WT or mutant RRE, increasing concentration (0-32ng) of codon optimized NL4-3 Rev expression plasmid, 100 ng of SEAP expression plasmid, and compensatory amount of empty plasmid (pCMV). The p24 and SEAP levels in the culture medium were measured 48 hours post transfection. Error bars represent standard error of mean from two replicates.

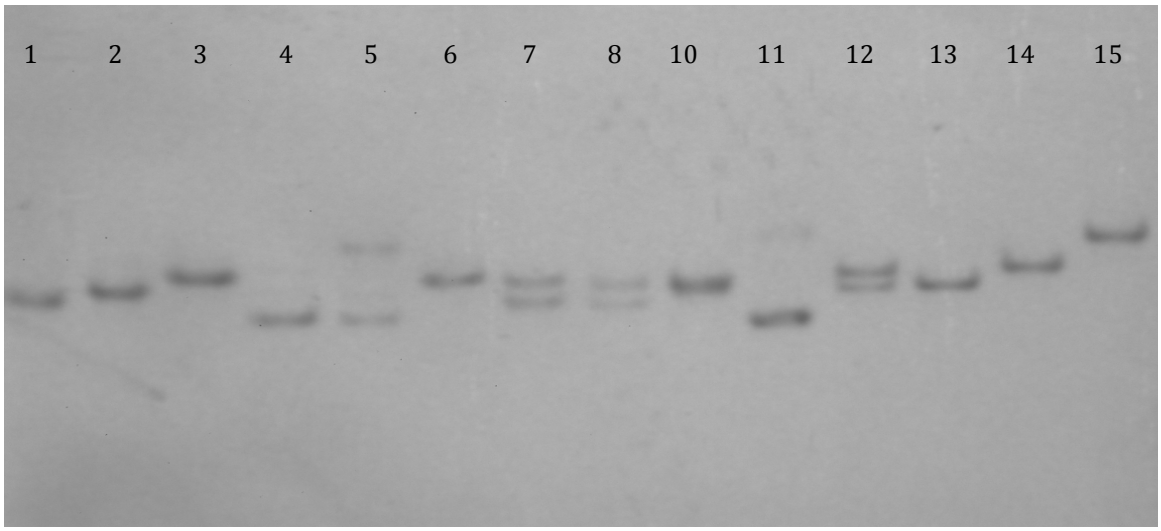
(B). Transient transfection of 293T cells with proviral DNA. 3 X 10^6 293T cells were transfected with 5 µg of proviral plasmid and the p24 in the culture medium was measured 48 hours post transfection.

(C). Rev dose response curves for mutant F and mutant E RRE using proviral plasmids. 2 X 10^5 293T cells were co-transfected with 2 µg of pCMVGag-Pol reporter plasmid containing the WT or mutant RRE, increasing concentration (0-32 ng) of codon optimized NL4-3 Rev expression plasmid, 100 ng of SEAP expression plasmid, and compensatory amount of empty plasmid (pUC18). The p24 and SEAP levels in the culture medium were measured 48 hrs post transfection. Error bars represent standard error of mean from two replicates.

To test this idea, we created Rev- versions of the proviral constructs and performed the proviral construct transfection assay almost exactly like the GagPol reporter assay, supplying Rev at very low levels from a separate plasmid. In this assay, we observed that the mutant F was now significantly more active than mutant E (Figure 25C). Thus, we conclude that, in a transient transfection, the differential Rev-RRE activity among the RRE mutants can only be observed at low Rev concentrations. This would be expected to more closely mimic what goes on in an HIV infected cell where only a few viral genomes are present to produce Rev.

Structural heterogeneity of patients RRE

We also tested if the RREs from clinical HIV isolates have a propensity to display structural heterogeneity like NL4-3 RRE on a native polyacrylamide gel (Figure 26). The RREs tested were from a previously published longitudinal study of sequence and functional variation of patients' Revs and RREs in plasma samples (214). These RREs were *in-vitro* transcribed without the 3' structure cassette tag, heated to 85°C and then slow cooled to 25°C. As controls, we ran the WT RRE, the 5 stem-loop locked RRE (mutant A), 4 stem-loop locked RRE (mutant B) and RRE61 RRE (137) (lane 12 to lane 15). All the control RREs' did not contain the structure cassette. We found that the controls migrated as expected – the WT RRE formed two bands and among the latter three RREs, mutant A migrated the fastest followed by the mutant B and then the RRE61 RRE (137).



Lane #	RRE name
1.	1005-M0
2.	1005-M41-B
3.	SC3-MO-A
4.	SC3-M57-A
5.	SC3-M57-B
6.	SC3-M0-B
7.	3024-M0
8.	3024-M43-A
10.	SC1-M0
11.	SC1-M121-A
12.	WT-RRE
13.	5 SL RRE
14.	4 SL RRE
15.	RRE61

Figure 26: Migration pattern of patients' RREs on a native polyacrylamide gel. Approximately 20 picomoles of in-vitro transcribed and PAGE purified mutant RRE RNAs were heated to 85°C in re-naturation buffer (10 mM Tris pH 8.0, 100 mM KCl, 0.1 mM EDTA) and slow cooled to 25°C. The folded RNAs were run on a 8% non-denaturing polyacrylamide gel at 200V at 4°C in the presence of 5mM $MgCl_2$ both in the gel and in the 1XTBE running buffer. In the RRE names, the numbers 1005, SC3, 3024, and SC1 are patients' code number. M"0" indicates the RRE observed at the earliest time point. M"X" indicates the late RRE, where X refers to the months after M0. The letters A and B refers to different RRE sequence from the same patient and from the same time-point.

Since RRE migration pattern for the WT RRE, mutant A, and mutant B was identical to what we previously observed with the structure cassette tagged RREs, we conclude that, indeed, the structure cassette does not change the migration pattern of the RREs

In one (SC3) out of the four patients, we observed that one of the late time-point RRE (lane 5; SC3-M57B) appeared to have two distinctly different structural forms, one of which co-migrated with an another late time-point RRE band from the same patient [SC3-M57A (lane 4)]. Additionally, both the early and late RREs from patient 3024 appeared to migrate as doublets (lanes 7 and 8), with migration rates very similar to the WT RRE bands. Patients' 1005 and SC3 RREs appeared to form predominantly a single band, hence a single structure. It will be interesting to determine if the patients' RREs are some variations of the 4 and the 5 stem-loop structures or if they form some unique RRE structures.

CHAPTER IV - DISCUSSION

Past studies (32, 57, 128, 137, 155, 238) on the secondary structure of the subtype B RRE have either supported a 5 stem-loop structure or a 4 stem-loop structure. Early studies on the secondary structure of the RRE were done on the HXB2 short RRE, which supported a 5 stem-loop structure. Mann *et al.* (155) showed for the first time that the 351-nt long NL4-3 RRE adopts an alternative 4 stem-loop structure. However, since the NL4-3 and the HXB2 RRE sequences differ only by two nucleotides, the authors suggested that the longer stem-loop I region of the long RRE might have allowed the RRE to fold into an alternative structure. In contrast to their suggestion, more recent studies using SHAPE technology have shown that the short RRE forms the alternative 4 stem-loop structure (137) and the long RRE forms the 5 stem-loop structure (238). Therefore, a consensus on a definitive secondary structure of the RRE has been still lacking.

In this study, we have shown that the short NL4-3 RRE (both 232-nt RRE with the SHAPE structure cassette and the 234-nt RRE without the SHAPE structure cassette) migrate as two bands on a native polyacrylamide gel, indicating that the RRE might exist in two structural forms. This data is consistent with a recent report by Pallesen *et al.* (172) who have shown that the long NL4-3 RRE (370-nt) forms a doublet band on a native gel. The authors speculated that the two bands of the RRE represent the 5 and the 4 stem-loop structures. They further studied the 3D structure of the RRE using atomic force microscopy but the resolution of this technique was not high enough to separate the two

structural forms of the RRE. In our study, we have successfully applied the in-gel SHAPE technique to show that the RREs in the faster migrating band form the 5 stem-loop structure, while the slower migrating RRE band consists of an RRE with the 4 stem-loop structure. Thus, we have shown, for the first time, that an *in-vitro* transcribed NL4-3 RRE can exist in both the 5 stem-loop and the 4 stem-loop structure.

Since both bands of the short NL4-3 RRE in our study and those of the long NL4-3 RRE [from Pallesen *et al.* (172) study] were of equal intensity, we also conclude that the RRE distributes equally into the 5 stem-loop and the 4 stem-loop structures. This observation is further supported by the fact that the Gibbs free energy of formation predicted by the RNAStructure program is very similar for both structures. As such, it is conceivable that both the structures might exist in dynamic equilibrium at least under the conditions used in these studies. Furthermore, the data does not support the Mann *et al.* (155*)* prediction that the length of stem-loop I might play a role in determining which of the two structures predominates.

By creating mutants capable of forming only one or the other of these structures, we also demonstrated functional differences between the two structures. We have shown that the 5 stem-loop structure was more active in promoting viral replication than its 4 stem-loop counterpart. The replication results cannot be easily explained by either differences in RRE-Rev binding affinities or in Rev binding co-operativity between either the wild type RRE or any of the RRE mutants, since our electrophoretic mobility gel shift experiments failed

to indicate consistent quantitative differences. Closer inspection of the gels reveal slight differences in the relative migration rates of some complexes, suggesting that differences in conformation might account for the effects on replication by potentially altering the ability to recruit additional proteins needed for Rev function *in vivo*. A detailed understanding of why these RREs behave differently in promoting viral replication will require structural studies beyond those reported here, including both *in vivo* and *in vitro* structural determinations.

A detailed 3 dimensional structure of the RRE, as determined by small angle X-ray scattering (SAXS), was recently reported (71). By modeling the 4 SL secondary structure of the RRE on the SAXS envelope, the authors deduced a SAXS structural model where the distance between the Rev binding sites on stem loop IIB (the primary Rev binding site) and stem loop I (the secondary Rev binding site) is ~55 Å. This distance matches the previously shown distance between the two RRE binding motifs in a Rev dimer (52, 62). However, the study did not attempt to fit the 5 SL RRE structure into the SAXS envelope; thus, it sheds no light on the basis for the replication differences observed here.

Cell-type specific effects of RRE mutations on viral replication kinetics have been previously documented by Dayton *et al.* (54). They studied the replication kinetics of RRE mutants in different cell types representative of monocytic and lymphocytic lineages, namely A3.01 (transformed T cells derived from CEM cells), the lymphoid lines Jurkat, MT4, and SupTl and the monocytoid lines THP1 and U937. They found that many of their RRE mutants showed severely impaired viral replication in some cell type while producing only a

minimal effect on viral replication in other cell-types. Interesting, like both the 5 and 4 stem-loop locked RRE mutants used in our study, the RRE mutants used in this study, had mutations in regions outside the primary Rev binding region of the RRE. Therefore, such cell type specific effects might be attributed to cell-type specific cellular factors, which might be interacting directly with the RREs outside the primary Rev binding regions. It is also possible that such factors might promote secondary Rev binding events or other Rev-RRE dependent steps in the viral life cycle. It would be interesting to determine if our RRE mutants behave differently in different cell types.

Contrary to our findings, a previously published short study (252) has shown that both structural forms of the RRE were functionally equivalent. In the study, different mutant RREs designed to distort or maintain the 5 stem-loop and the 4 stem-loop structures were created and the Rev-RRE function of these mutants were determined in a transient transfection assay. The study, however, has many shortcomings. First, the structures of the RRE mutants used in the study were not confirmed experimentally. Second, the functional data of the study came only from one assay. Moreover, the assay consisted of determining Rev-RRE activity from transient transfection of COS-7 cells with GagPol reporter plasmid containing the test RRE and the Rev expression plasmid containing the SV40 promoter. Since the COS-7 cells expressed T-antigen, in this assay, the Rev-RRE activity was measured at very high levels using a replicating Rev expression plasmid. We have shown in this study that the Rev-RRE activity difference among the RRE mutants in a transient transfection was observed only

at very low level of Rev plasmid, as occurs during HIV infection. Therefore, the use of higher levels of the Rev expression plasmid in their assay could explain why they were unable to notice any functional difference between the two structural forms.

As described in the Introduction section, several studies have indicated potential links between HIV Rev/RRE function and pathogenesis. For example, Bobbitt *et al.* (19) found that in asymptomatic infected patients with an active immune response, infected cells were more resistant to anti-Gag and anti-Env CTL killing compared to infected cells from advanced AIDS patients. This was attributed to attenuated *rev* alleles in the viral DNA of the healthy patients lowering cellular expression levels of Gag and Env. Other studies have also reported the contribution of less-active *rev* alleles to long-term survival of HIV-1 infection in some patients (110, 114) and there have also been a few studies on the evolution of Rev and RRE function in individual patients. These studies have concluded that RREs of varying activity can evolve over time (178, 180). RRE functional variation has also been studied in a longitudinal cohort (178, 181, 188), where a correlation was observed between the rates of CD4+ decline and RRE activity at the late time points. Thus, RRE evolution may be an important regulator of HIV pathogenesis and disease progression.

In the present study, we have shown that HIV can achieve differential replication rates simply by alternating RRE structure, which presumably might occur in the absence of virus evolution. The ability of the RRE to adopt different structural conformations that promote different replication activities may allow

HIV to modulate its rate of replication under distinct conditions, or in distinct compartments, to better survive in the environment of the host. It is striking that HIV has evolved an RNA export element that is larger and more complex than the CTEs found in the simpler retroviruses and it has remained a puzzle as to why such a large RNA element has been conserved. It may be that the complex structure of the RRE is necessary to confer the ability to rearrange and promote differential Rev activity. This might provide a novel heretofore unstudied mechanism to regulate virus growth in response to various cellular cues. While factors promoting formation of different RRE conformers remain to be elucidated, it is striking that the free energies of each are extremely similar, making it plausible that minor differences in the concentration of yet unidentified cellular factors could dynamically influence RRE folding in vivo.

In this study, we also tested if the RREs from clinical HIV isolates have a propensity to display structural heterogeneity like NL4-3 RRE on a native polyacrylamide gel. While most of the RREs migrated as single bands under the assay conditions, we did observe that in one patient, one of the late-time point RRE appeared to be a mixture of the corresponding early-time point RRE structure and a structurally distinct late-time point RRE. What was even more interesting was that some of the patients' RREs from early and late time-points produced bands with strikingly different migration rates that were also distinct from the two NL4-3 RRE bands. Evidently, these RREs form structures distinctly different from both the NL4-3 RRE and from each other. It will be interesting to see what regions of these RREs vary structurally. Such studies will provide us

with insights into the structural plasticity of functional RREs. Since our current understanding of RRE structure comes mainly from studies of the HIV molecular clones NL4-3 and HXB2 RREs, there is a clear need in the field to expand the RRE structural studies to patients' RREs.

REFERENCES

1. **Abraha, A., R. M. Troyer, M. E. Quinones-Mateu, and E. J. Arts.** 2005. Methods to determine HIV-1 ex vivo fitness. Methods Mol. Biol. **304:**355-368.

2. **Adachi, A., H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin.** 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. **59:**284-291.

3. **Addo, M. M., M. Altfeld, E. S. Rosenberg, R. L. Eldridge, M. N. Philips, K. Habeeb, A. Khatri, C. Brander, G. K. Robbins, G. P. Mazzara, P. J. Goulder, and B. D. Walker.** 2001. The HIV-1 regulatory proteins Tat and Rev are frequently targeted by cytotoxic T lymphocytes derived from HIV-1-infected individuals. Proc Natl Acad Sci U S A **98:**1781-1786.

4. **Addo, M. M., X. G. Yu, E. S. Rosenberg, B. D. Walker, and M. Altfeld.** 2002. Cytotoxic T-lymphocyte (CTL) responses directed against regulatory and accessory proteins in HIV-1 infection. DNA Cell Biol **21:**671-678.

5. **Alizon, M., and L. Montagnier.** 1986. Lymphadenopathy/AIDS virus: genetic organization and relationship to animal lentiviruses. Anticancer Res. **6:**403-411.

6. **Alizon, M., P. Sonigo, F. Barre-Sinoussi, J. C. Chermann, P. Tiollais, L. Montagnier, and S. Wain-Hobson.** 1984. Molecular cloning of lymphadenopathy-associated virus. Nature **312:**757-760.

7. **Arnold, M., A. Nath, J. Hauber, and R. H. Kehlenbach.** 2006. Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. J. Biol. Chem. **281:**20883-20890.

8. **Ashe, M. P., L. H. Pearson, and N. J. Proudfoot.** 1997. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. EMBO J. **16:**5752-5763.

9. **Askjaer, P., T. H. Jensen, J. Nilsson, L. Englmeier, and J. Kjems.** 1998. The specificity of the CRM1-Rev nuclear export signal interaction is mediated by RanGTP. J Biol Chem **273:**33414-33422.

10. **Askjaer, P., T. H. Jensen, J. Nilsson, L. Englmeier, and J. Kjems.** 1998. The specificity of the CRM1-Rev nuclear export signal interaction is mediated by RanGTP. J. Biol. Chem. **273:**33414-33422.

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

12. **Bartel, D. P., M. L. Zapp, M. R. Green, and J. W. Szostak.** 1991. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell **67:**529-536.

13. **Battiste, J. L., H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, A. D. Frankel, and J. R. Williamson.** 1996. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. Science **273:**1547-1551.

14. **Berger, J., C. Aepinus, M. Dobrovnik, B. Fleckenstein, J. Hauber, and E. Bohnlein.** 1991. Mutational analysis of functional domains in the HIV-1 Rev trans-regulatory protein. Virology **183:**630-635.

15. **Berger, J., J. Hauber, R. Hauber, R. Geiger, and B. R. Cullen.** 1988. Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. Gene **66:**1-10.

16. **Berkhout, B., and J. L. van Wamel.** 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. RNA **6:**282-295.

17. **Bieniasz, P. D., T. A. Grdina, H. P. Bogerd, and B. R. Cullen.** 1998. Recruitment of a protein complex containing Tat and cyclin T1 to TAR governs the species specificity of HIV-1 Tat. EMBO J. **17:**7056-7065.

18. **Blackard, J. T., D. E. Cohen, and K. H. Mayer.** 2002. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. Clin. Infect. Dis. **34:**1108-1114.

19. **Bobbitt, K. R., M. M. Addo, M. Altfeld, T. Filzen, A. A. Onafuwa, B. D. Walker, and K. L. Collins.** 2003. Rev activity determines sensitivity of HIV-1-infected primary T cells to CTL killing. Immunity **18:**289-299.

20. **Bour, S., U. Schubert, and K. Strebel.** 1995. The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. J. Virol. **69:**1510-1520.

21. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3:**469-478.

22. **Bray, M., S. Prasad, J. W. Dubay, E. Hunter, K. T. Jeang, D. Rekosh, and M. L. Hammarskjold.** 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. Proc. Natl. Acad. Sci. U. S. A. **91:**1256-1260.

23. **Broder, C. C., and R. G. Collman.** 1997. Chemokine receptors and HIV. J. Leukoc. Biol. **62:**20-29.

24. **Bryant, M., and L. Ratner.** 1990. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc. Natl. Acad. Sci. U. S. A. **87:**523-527.

25. **Buckley, P. T., M. Khaladkar, J. Kim, and J. Eberwine.** 2014. Cytoplasmic intron retention, function, splicing, and the sentinel RNA hypothesis. Wiley interdisciplinary reviews. RNA **5:**223-230.

26. **Bukrinskaya, A. G.** 2004. HIV-1 assembly and maturation. Arch. Virol. **149:**1067-1082.

27. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV integration protein in vitro. Science **249:**1555-1558.

28. **Buzon, V., G. Natrajan, D. Schibli, F. Campelo, M. M. Kozlov, and W. Weissenhorn.** 2010. Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. PLoS Pathog. **6:**e1000880.

29. **Cassan, M., N. Delaunay, C. Vaquero, and J. P. Rousset.** 1994. Translational frameshifting at the gag-pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. J. Virol. **68:**1501-1508.

30. **Chan, D. C., D. Fass, J. M. Berger, and P. S. Kim.** 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell **89:**263-273.

31. **Chang, D. D., and P. A. Sharp.** 1989. Regulation by HIV Rev depends upon recognition of splice sites. Cell **59:**789-795.

32. **Charpentier, B., F. Stutz, and M. Rosbash.** 1997. A dynamic in vivo view of the HIV-I Rev-RRE interaction. J. Mol. Biol. **266:**950-962.

33. **Cherepanov, P., G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq, and Z. Debyser.** 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. **278:**372-381.

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

35. **Clever, J., C. Sassetti, and T. G. Parslow.** 1995. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. J. Virol. **69:**2101-2109.

36. **Clever, J. L., and T. G. Parslow.** 1997. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. J. Virol. **71:**3407-3414.

37. **Cochrane, A. W., A. Perkins, and C. A. Rosen.** 1990. Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function. Journal of virology **64:**881-885.

38. **Coffin, j. m.** 1996. The Viruses and Their Replication. Lippincott-Raven, philadelphia.

39. **Cole, J. L., J. D. Gehman, J. A. Shafer, and L. C. Kuo.** 1993. Solution oligomerization of the rev protein of HIV-1: implications for function. Biochemistry **32:**11769-11775.

40. **Connell, B. J., and H. Lortat-Jacob.** 2013. Human immunodeficiency virus and heparan sulfate: from attachment to entry inhibition. Front. Immunol. **4:**385.

41. **Cook, K. S., G. J. Fisk, J. Hauber, N. Usman, T. J. Daly, and J. R. Rusche.** 1991. Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. Nucleic Acids Res. **19:**1577-1583.

42. **D'Souza, V., and M. F. Summers.** 2005. How retroviruses select their genomes. Nature reviews. Microbiology **3:**643-655.

43. **Daelemans, D., S. V. Costes, S. Lockett, and G. N. Pavlakis.** 2005. Kinetic and molecular analysis of nuclear export factor CRM1 association with its cargo in vivo. Mol Cell Biol **25:**728-739.

44. **Dahiya, S., Y. Liu, M. R. Nonnemacher, W. Dampier, and B. Wigdahl.** 2014. CCAAT enhancer binding protein and nuclear factor of activated T cells regulate HIV-1 LTR via a novel conserved downstream site in cells of the monocyte-macrophage lineage. PLoS One **9:**e88116.

45. **Daly, T. J., K. S. Cook, G. S. Gray, T. E. Maione, and J. R. Rusche.** 1989. Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature **342:**816-819.

46. **Daly, T. J., R. C. Doten, P. Rennert, M. Auer, H. Jaksche, A. Donner, G. Fisk, and J. R. Rusche.** 1993. Biochemical characterization of binding of multiple HIV-1 Rev monomeric proteins to the Rev responsive element. Biochemistry **32:**10497-10505.

47. **Daly, T. J., J. R. Rusche, T. E. Maione, and A. D. Frankel.** 1990. Circular dichroism studies of the HIV-1 Rev protein and its specific RNA binding site. Biochemistry **29:**9791-9795.

48. **Daugherty, M. D., D. S. Booth, B. Jayaraman, Y. Cheng, and A. D. Frankel.** 2010. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. Proc. Natl. Acad. Sci. U. S. A. **107:**12481-12486.

49. **Daugherty, M. D., D. S. Booth, B. Jayaraman, Y. Cheng, and A. D. Frankel.** 2010. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. Proc Natl Acad Sci U S A **107:**12481-12486.

50. **Daugherty, M. D., I. D'Orso, and A. D. Frankel.** 2008. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. Mol Cell **31:**824-834.

51. **Daugherty, M. D., I. D'Orso, and A. D. Frankel.** 2008. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. Mol. Cell **31:**824-834.

52. **Daugherty, M. D., B. Liu, and A. D. Frankel.** 2010. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. Nat. Struct. Mol. Biol. **17:**1337-1342.

53. **Daugherty, M. D., B. Liu, and A. D. Frankel.** 2010. Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. Nat Struct Mol Biol **17:**1337-1342.

54. **Dayton, E. T., D. A. Konings, S. Y. Lim, R. K. Hsu, L. Butini, G. Pantaleo, and A. I. Dayton.** 1993. The RRE of human immunodeficiency virus type 1 contributes to cell-type-specific viral tropism. J. Virol. **67:**2871-2878.

55. **Dayton, E. T., D. A. Konings, D. M. Powell, B. A. Shapiro, L. Butini, J. V. Maizel, and A. I. Dayton.** 1992. Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein. J Virol **66:**1139-1151.

56. **Dayton, E. T., D. A. Konings, D. M. Powell, B. A. Shapiro, L. Butini, J. V. Maizel, and A. I. Dayton.** 1992. Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein. J. Virol. **66:**1139-1151.

57. **Dayton, E. T., D. M. Powell, and A. I. Dayton.** 1989. Functional analysis of CAR, the target sequence for the Rev protein of HIV-1. Science **246:**1625-1629.

58. **de Marco, A., B. Muller, B. Glass, J. D. Riches, H. G. Krausslich, and J. A. Briggs.** 2010. Structural analysis of HIV-1 maturation using cryo-electron tomography. PLoS Pathog. **6:**e1001215.

59. **De Rijck, J., and Z. Debyser.** 2006. The central DNA flap of the human immunodeficiency virus type 1 is important for viral replication. Biochem. Biophys. Res. Commun. **349:**1100-1110.

60. **de Witte, L., A. Nabatov, and T. B. Geijtenbeek.** 2008. Distinct roles for DC-SIGN+-dendritic cells and Langerhans cells in HIV-1 transmission. Trends Mol. Med. **14:**12-19.

61. **Desmezieres, E., N. Gupta, R. Vassell, Y. He, K. Peden, L. Sirota, Z. Yang, P. Wingfield, and C. D. Weiss.** 2005. Human immunodeficiency virus (HIV) gp41 escape mutants: cross-resistance to peptide inhibitors of HIV fusion and altered receptor activation of gp120. J Virol **79:**4774-4781.

62. **DiMattia, M. A., N. R. Watts, S. J. Stahl, C. Rader, P. T. Wingfield, D. I. Stuart, A. C. Steven, and J. M. Grimes.** 2010. Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element. Proc. Natl. Acad. Sci. U. S. A. **107:**5810-5814.

63. **Dorfman, T., F. Mammano, W. A. Haseltine, and H. G. Gottlinger.** 1994. Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. J. Virol. **68:**1689-1696.

64. **DuBridge, R. B., P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller, and M. P. Calos.** 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol. Cell. Biol. **7:**379-387.

65. **el Kharroubi, A., and M. A. Martin.** 1996. cis-acting sequences located downstream of the human immunodeficiency virus type 1 promoter affect its chromatin structure and transcriptional activity. Mol. Cell. Biol. **16:**2958-2966.

66. **el Kharroubi, A., and E. Verdin.** 1994. Protein-DNA interactions within DNase I-hypersensitive sites located downstream of the HIV-1 promoter. J. Biol. Chem. **269:**19916-19924.

67. **Emerman, M., R. Vazeux, and K. Peden.** 1989. The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization. Cell **57:**1155-1165.

68. **Emiliani, S., A. Mousnier, K. Busschots, M. Maroun, B. Van Maele, D. Tempe, L. Vandekerckhove, F. Moisant, L. Ben-Slama, M. Witvrouw, F. Christ, J. C. Rain, C. Dargemont, Z. Debyser, and R. Benarous.** 2005. Integrase mutants defective for interaction with LEDGF/p75 are

impaired in chromosome tethering and HIV-1 replication. J. Biol. Chem. **280:**25517-25523.

69. **Engelman, A., and P. Cherepanov.** 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. Nature reviews. Microbiology **10:**279-290.

70. **Engelman, A., K. Mizuuchi, and R. Craigie.** 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell **67:**1211-1221.

71. **Fang, X., J. Wang, I. P. O'Carroll, M. Mitchell, X. Zuo, Y. Wang, P. Yu, Y. Liu, J. W. Rausch, M. A. Dyba, J. Kjems, C. D. Schwieters, S. Seifert, R. E. Winans, N. R. Watts, S. J. Stahl, P. T. Wingfield, R. A. Byrd, S. F. Le Grice, A. Rein, and Y. X. Wang.** 2013. An unusual topological structure of the HIV-1 Rev response element. Cell **155:**594-605.

72. **Farnet, C. M., and F. D. Bushman.** 1997. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell **88:**483-492.

73. **Farnet, C. M., and W. A. Haseltine.** 1990. Integration of human immunodeficiency virus type 1 DNA in vitro. Proc. Natl. Acad. Sci. U. S. A. **87:**4164-4168.

74. **Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong-Staal.** 1986. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. Cell **46:**807-817.

75. **Felber, B. K., M. Hadzopoulou-Cladaras, C. Cladaras, T. Copeland, and G. N. Pavlakis.** 1989. rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. U. S. A. **86:**1495-1499.

76. **Feng, S., and E. C. Holland.** 1988. HIV-1 tat trans-activation requires the loop sequence within tar. Nature **334:**165-167.

77. **Fenster, S. D., R. W. Wagner, B. C. Froehler, and D. J. Chin.** 1994. Inhibition of human immunodeficiency virus type-1 env expression by C-5 propyne oligonucleotides specific for Rev-response element stem-loop V. Biochemistry **33:**8391-8398.

78. **Finkelshtein, D., A. Werman, D. Novick, S. Barak, and M. Rubinstein.** 2013. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. Proc. Natl. Acad. Sci. U. S. A. **110:**7306-7311.

79. **Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Luhrmann.** 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell **82:**475-483.

80. **Fischer, U., S. Meyer, M. Teufel, C. Heckel, R. Luhrmann, and G. Rautmann.** 1994. Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. EMBO J. **13:**4105-4112.

81. **Fitzon, T., B. Leschonsky, K. Bieler, C. Paulus, J. Schroder, H. Wolf, and R. Wagner.** 2000. Proline residues in the HIV-1 NH2-terminal capsid domain: structure determinants for proper core assembly and subsequent steps of early replication. Virology **268:**294-307.

82. **Fornerod, M., M. Ohno, M. Yoshida, and I. W. Mattaj.** 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. Cell **90:**1051-1060.

83. **Forshey, B. M., U. von Schwedler, W. I. Sundquist, and C. Aiken.** 2002. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. J. Virol. **76:**5667-5677.

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

85. **Freed, E. O., and M. A. Martin.** 1995. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. J. Virol. **69:**1984-1989.

86. **Fridell, R. A., K. M. Partin, S. Carpenter, and B. R. Cullen.** 1993. Identification of the activation domain of equine infectious anemia virus rev. J. Virol. **67:**7317-7323.

87. **Fujinaga, K., T. P. Cujec, J. Peng, J. Garriga, D. H. Price, X. Grana, and B. M. Peterlin.** 1998. The ability of positive transcription elongation factor B to transactivate human immunodeficiency virus transcription depends on a functional kinase domain, cyclin T1, and Tat. J. Virol. **72:**7154-7159.

88. **Gallo, R., F. Wong-Staal, L. Montagnier, W. A. Haseltine, and M. Yoshida.** 1988. HIV/HTLV gene nomenclature. Nature **333:**504.

89. **Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, and 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.

90. **Ganser, B. K., S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist.** 1999. Assembly and analysis of conical models for the HIV-1 core. Science **283:**80-83.

91. **Garrett, E. D., L. S. Tiley, and B. R. Cullen.** 1991. Rev activates expression of the human immunodeficiency virus type 1 vif and vpr gene products. J. Virol. **65:**1653-1657.

92. **Garrus, J. E., U. K. von Schwedler, O. W. Pornillos, S. G. Morham, K. H. Zavitz, H. E. Wang, D. A. Wettstein, K. M. Stray, M. Cote, R. L. Rich, D. G. Myszka, and W. I. Sundquist.** 2001. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell **107:**55-65.

93. **Gelderblom, H. R., E. H. Hausmann, M. Ozel, G. Pauli, and M. A. Koch.** 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. Virology **156:**171-176.

94. **Gheysen, D., E. Jacobs, F. de Foresta, C. Thiriart, M. Francotte, D. Thines, and M. De Wilde.** 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell **59:**103-112.

95. **Gottlinger, H. G., T. Dorfman, J. G. Sodroski, and W. A. Haseltine.** 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. Proc. Natl. Acad. Sci. U. S. A. **88:**3195-3199.

96. **Gottlinger, H. G., J. G. Sodroski, and W. A. Haseltine.** 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. U. S. A. **86:**5781-5785.

97. **Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. **36:**59-74.

98. **Hadzopoulou-Cladaras, M., B. K. Felber, C. Cladaras, A. Athanassopoulos, A. Tse, and G. N. Pavlakis.** 1989. The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. J. Virol. **63:**1265-1274.

99. **Haffar, O. K., S. Popov, L. Dubrovsky, I. Agostini, H. Tang, T. Pushkarsky, S. G. Nadler, and M. Bukrinsky.** 2000. Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. J. Mol. Biol. **299:**359-368.

100. **Hallenberger, S., V. Bosch, H. Angliker, E. Shaw, H. D. Klenk, and W. Garten.** 1992. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature **360:**358-361.

101. **Hamm, J., and I. W. Mattaj.** 1990. Monomethylated cap structures facilitate RNA export from the nucleus. Cell **63:**109-118.

102. **Hamm, T. E., D. Rekosh, and M. L. Hammarskjöld.** 1999. Selection and characterization of human immunodeficiency virus type 1 mutants that are resistant to inhibition by the transdominant negative RevM10 protein. Journal of virology **73:**5741-5747.

103. **Hammarskjold, M. L., J. Heimer, B. Hammarskjold, I. Sangwan, L. Albert, and D. Rekosh.** 1989. Regulation of human immunodeficiency virus env expression by the rev gene product. J. Virol. **63:**1959-1966.

104. **Heaphy, S., J. T. Finch, M. J. Gait, J. Karn, and M. Singh.** 1991. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. Proc. Natl. Acad. Sci. U. S. A. **88:**7366-7370.

105. **Hidaka, M., J. Inoue, M. Yoshida, and M. Seiki.** 1988. Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. EMBO J. **7:**519-523.

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

107. **Hofacre, A., T. Nitta, and H. Fan.** 2009. Jaagsiekte sheep retrovirus encodes a regulatory factor, Rej, required for synthesis of Gag protein. J. Virol. **83:**12483-12498.

108. **Holland, S. M., N. Ahmad, R. K. Maitra, P. Wingfield, and S. Venkatesan.** 1990. Human immunodeficiency virus rev protein recognizes a target sequence in rev-responsive element RNA within the context of RNA secondary structure. J Virol **64:**5966-5975.

109. **Hope Thomas J., T. D.** 2000, posting date. Structure, Expresson, and Regulation of the HIV genome. HIV InSite Knowledge Base Chapter, University of California, San Francisco. [Online.]

110. **Hua, J., J. J. Caffrey, and B. R. Cullen.** 1996. Functional consequences of natural sequence variation in the activation domain of HIV-1 Rev. Virology **222:**423-429.

111. **Huang, M., and M. A. Martin.** 1997. Incorporation of Pr160(gag-pol) into virus particles requires the presence of both the major homology region and adjacent C-terminal capsid sequences within the Gag-Pol polyprotein. J. Virol. **71:**4472-4478.

112. **Huang, M., J. M. Orenstein, M. A. Martin, and E. O. Freed.** 1995. p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. J. Virol. **69:**6810-6818.

113. **Iversen, A. K., E. G. Shpaer, A. G. Rodrigo, M. S. Hirsch, B. D. Walker, H. W. Sheppard, T. C. Merigan, and J. I. Mullins.** 1995. Persistence of attenuated rev genes in a human immunodeficiency virus type 1-infected asymptomatic individual. J Virol **69:**5743-5753.

114. **Iversen, A. K., E. G. Shpaer, A. G. Rodrigo, M. S. Hirsch, B. D. Walker, H. W. Sheppard, T. C. Merigan, and J. I. Mullins.** 1995. Persistence of attenuated rev genes in a human immunodeficiency virus type 1-infected asymptomatic individual. J. Virol. **69:**5743-5753.

115. **Iwai, S., C. Pritchard, D. A. Mann, J. Karn, and M. J. Gait.** 1992. Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. Nucleic Acids Res. **20:**6465-6472.

116. **Jain, C., and J. G. Belasco.** 2001. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. Mol Cell **7:**603-614.

117. **Jain, C., and J. G. Belasco.** 2001. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. Mol. Cell **7:**603-614.

118. **Jain, C., and J. G. Belasco.** 1996. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. Cell **87:**115-125.

119. **Johnson, S. F., and A. Telesnitsky.** 2010. Retroviral RNA dimerization and packaging: the what, how, when, where, and why. PLoS Pathog. **6:**e1001007.

120. **Jowett, J. B., V. Planelles, B. Poon, N. P. Shah, M. L. Chen, and I. S. Chen.** 1995. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. J. Virol. **69:**6304-6313.

121. **Kalpana, G. V., S. Marmon, W. Wang, G. R. Crabtree, and S. P. Goff.** 1994. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science **266:**2002-2006.

122. **Keele, B. F., F. Van Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, and B. H. Hahn.** 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science **313:**523-526.

123. **Kenyon, J. C., L. J. Prestwood, S. F. Le Grice, and A. M. Lever.** 2013. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. Nucleic Acids Res. **41:**e174.

124. **Kerkau, T., I. Bacik, J. R. Bennink, J. W. Yewdell, T. Hunig, A. Schimpl, and U. Schubert.** 1997. The human immunodeficiency virus type 1 (HIV-1) Vpu protein interferes with an early step in the biosynthesis of major histocompatibility complex (MHC) class I molecules. J. Exp. Med. **185:**1295-1305.

125. **Kjems, J., M. Brown, D. D. Chang, and P. A. Sharp.** 1991. Structural analysis of the interaction between the human immunodeficiency virus

Rev protein and the Rev response element. Proc. Natl. Acad. Sci. U. S. A. **88:**683-687.

126. **Kjems, J., M. Brown, D. D. Chang, and P. A. Sharp.** 1991. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. Proc Natl Acad Sci U S A **88:**683-687.

127. **Kjems, J., B. J. Calnan, A. D. Frankel, and P. A. Sharp.** 1992. Specific binding of a basic peptide from HIV-1 Rev. EMBO J. **11:**1119-1129.

128. **Kjems, J., A. D. Frankel, and P. A. Sharp.** 1991. Specific regulation of mRNA splicing in vitro by a peptide from HIV-1 Rev. Cell **67:**169-178.

129. **Kjems, J., and P. A. Sharp.** 1993. The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6.U5 small nuclear ribonucleoprotein in spliceosome assembly. J. Virol. **67:**4769-4776.

130. **Knight, S. C., S. E. Macatonia, and S. Patterson.** 1990. HIV I infection of dendritic cells. Int. Rev. Immunol. **6:**163-175.

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

132. **Kubota, S., T. Nosaka, B. R. Cullen, M. Maki, and M. Hatanaka.** 1991. Effects of chimeric mutants of human immunodeficiency virus type 1 Rev and human T-cell leukemia virus type I Rex on nucleolar targeting signals. Journal of virology **65:**2452-2456.

133. **Kuersten, S., M. Ohno, and I. W. Mattaj.** 2001. Nucleocytoplasmic transport: Ran, beta and beyond. Trends Cell Biol **11:**497-503.

134. **Kwong, P. D., R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature **393:**648-659.

135. **Le Gall, S., J. M. Heard, and O. Schwartz.** 1997. Analysis of Nef-induced MHC-I endocytosis. Res. Virol. **148:**43-47.

136. **Lee, M. S., and R. Craigie.** 1994. Protection of retroviral DNA from autointegration: involvement of a cellular factor. Proc. Natl. Acad. Sci. U. S. A. **91:**9823-9827.

137. **Legiewicz, M., C. S. Badorrek, K. B. Turner, D. Fabris, T. E. Hamm, D. Rekosh, M. L. Hammarskjold, and S. F. Le Grice.** 2008. Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. Proc. Natl. Acad. Sci. U. S. A. **105:**14365-14370.

138. **Lewis, N., J. Williams, D. Rekosh, and M. L. Hammarskjold.** 1990. Identification of a Cis-Acting Element in Human-Immunodeficiency-Virus Type-2 (Hiv-2) That Is Responsive to the Hiv-1 Rev and Human T-Cell Leukemia-Virus Type-I and Type-Ii Rex Proteins. J. Virol. **64:**1690-1697.

139. **Li, Y., Y. C. Bor, Y. Misawa, Y. Xue, D. Rekosh, and M. L. Hammarskjold.** 2006. An intron with a constitutive transport element is retained in a Tap messenger RNA. Nature **443:**234-237.

140. **Lin, C. W., and A. Engelman.** 2003. The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. J. Virol. **77:**5030-5036.

141. **Liu, J., A. Bartesaghi, M. J. Borgnia, G. Sapiro, and S. Subramaniam.** 2008. Molecular architecture of native HIV-1 gp120 trimers. Nature **455:**109-113.

142. **Louis, N., C. Evelegh, and F. L. Graham.** 1997. Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. Virology **233:**423-429.

143. **Lu, X. B., J. Heimer, D. Rekosh, and M. L. Hammarskjold.** 1990. U1 small nuclear RNA plays a direct role in the formation of a rev-regulated human immunodeficiency virus env mRNA that remains unspliced. Proc. Natl. Acad. Sci. U. S. A. **87:**7598-7602.

144. **Lu, Y. L., R. P. Bennett, J. W. Wills, R. Gorelick, and L. Ratner.** 1995. A leucine triplet repeat sequence (LXX)4 in p6gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles. J. Virol. **69:**6873-6879.

145. **Lusvarghi, S., J. Sztuba-Solinska, K. J. Purzycka, J. W. Rausch, and S. F. Le Grice.** 2013. RNA secondary structure prediction using high-throughput SHAPE. Journal of visualized experiments : JoVE**:**e50243.

146. **Madsen, J. M., and C. M. Stoltzfus.** 2005. An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. J. Virol. **79:**10478-10486.

147. **Maertens, G., P. Cherepanov, W. Pluymers, K. Busschots, E. De Clercq, Z. Debyser, and Y. Engelborghs.** 2003. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J. Biol. Chem. **278:**33528-33539.

148. **Mahalingam, S., V. Ayyavoo, M. Patel, T. Kieber-Emmons, and D. B. Weiner.** 1997. Nuclear import, virion incorporation, and cell cycle arrest/differentiation are mediated by distinct functional domains of human immunodeficiency virus type 1 Vpr. J. Virol. **71:**6339-6347.

149. **Malim, M. H., and B. R. Cullen.** 1991. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. Cell **65:**241-248.

150. **Malim, M. H., and B. R. Cullen.** 1993. Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. Mol. Cell. Biol. **13:**6180-6189.

151. **Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen.** 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature **338:**254-257.

152. **Mammano, F., A. Ohagen, S. Hoglund, and H. G. Gottlinger.** 1994. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. J. Virol. **68:**4927-4936.

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

154. **Mann, D. A., I. Mikaelian, R. W. Zemmel, S. M. Green, A. D. Lowe, T. Kimura, M. Singh, P. J. Butler, M. J. Gait, and J. Karn.** 1994. A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. J Mol Biol **241:**193-207.

155. **Mann, D. A., I. Mikaelian, R. W. Zemmel, S. M. Green, A. D. Lowe, T. Kimura, M. Singh, P. J. Butler, M. J. Gait, and J. Karn.** 1994. A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. J. Mol. Biol. **241:**193-207.

156. **Martarano, L., R. Stephens, N. Rice, and D. Derse.** 1994. Equine infectious anemia virus trans-regulatory protein Rev controls viral mRNA stability, accumulation, and alternative splicing. J. Virol. **68:**3102-3111.

157. **Mattaj, I. W., and L. Englmeier.** 1998. Nucleocytoplasmic transport: the soluble phase. Annu Rev Biochem **67:**265-306.

158. **Mertz, J. A., M. S. Simper, M. M. Lozano, S. M. Payne, and J. P. Dudley.** 2005. Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. J. Virol. **79:**14737-14747.

159. **Moore, M. D., and W. S. Hu.** 2009. HIV-1 RNA dimerization: It takes two to tango. AIDS reviews **11:**91-102.

160. **Muesing, M. A., D. H. Smith, C. D. Cabradilla, C. V. Benton, L. A. Lasky, and D. J. Capon.** 1985. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature **313:**450-458.

161. **Muller, H. P., and H. E. Varmus.** 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:**4704-4714.

162. **Naldini, L., U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono.** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science **272:**263-267.

163. **Nameki, D., E. Kodama, M. Ikeuchi, N. Mabuchi, A. Otaka, H. Tamamura, M. Ohno, N. Fujii, and M. Matsuoka.** 2005. Mutations conferring resistance to human immunodeficiency virus type 1 fusion inhibitors are restricted by gp41 and Rev-responsive element functions. J Virol **79:**764-770.

164. **Nasioulas, G., A. S. Zolotukhin, C. Tabernero, L. Solomin, C. P. Cunningham, G. N. Pavlakis, and B. K. Felber.** 1994. Elements distinct from human immunodeficiency virus type 1 splice sites are responsible for the Rev dependence of env mRNA. J. Virol. **68:**2986-2993.

165. **Neil, S. J., T. Zang, and P. D. Bieniasz.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451:**425-430.

166. **Neville, M., F. Stutz, L. Lee, L. I. Davis, and M. Rosbash.** 1997. The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. Curr Biol **7:**767-775.

167. **Nitta, T., A. Hofacre, S. Hull, and H. Fan.** 2009. Identification and mutational analysis of a Rej response element in Jaagsiekte sheep retrovirus RNA. J. Virol. **83:**12499-12511.

168. **Olsen, H. S., A. W. Cochrane, P. J. Dillon, C. M. Nalin, and C. A. Rosen.** 1990. Interaction of the human immunodeficiency virus type 1 Rev protein with a structured region in env mRNA is dependent on multimer formation mediated through a basic stretch of amino acids. Genes Dev. **4:**1357-1364.

169. **Olsen, H. S., P. Nelbock, A. W. Cochrane, and C. A. Rosen.** 1990. Secondary structure is the major determinant for interaction of HIV rev protein with RNA. Science **247:**845-848.

170. **Otero, G. C., M. E. Harris, J. E. Donello, and T. J. Hope.** 1998. Leptomycin B inhibits equine infectious anemia virus Rev and feline immunodeficiency virus rev function but not the function of the hepatitis B virus posttranscriptional regulatory element. J. Virol. **72:**7593-7597.

171. **Ott, D. E.** 2008. Cellular proteins detected in HIV-1. Rev. Med. Virol. **18:**159-175.

172. **Pallesen, J., M. Dong, F. Besenbacher, and J. Kjems.** 2009. Structure of the HIV-1 Rev response element alone and in complex with regulator of virion (Rev) studied by atomic force microscopy. The FEBS journal **276:**4223-4232.

173. **Parkin, N. T., M. Chamorro, and H. E. Varmus.** 1992. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. J. Virol. **66:**5147-5151.

174. **Pear, W. S., G. P. Nolan, M. L. Scott, and D. Baltimore.** 1993. Production of high-titer helper-free retroviruses by transient transfection. Proc. Natl. Acad. Sci. U. S. A. **90:**8392-8396.

175. **Peeters, M., M. D'Arc, and E. Delaporte.** 2014. Origin and diversity of human retroviruses. AIDS reviews **16:**23-34.

176. **Pereira, L. A., K. Bentley, A. Peeters, M. J. Churchill, and N. J. Deacon.** 2000. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Nucleic Acids Res. **28:**663-668.

177. **Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science **271:**1582-1586.

178. **Phuphuakrat, A., and P. Auewarakul.** 2005. Functional variability of Rev response element in HIV-1 primary isolates. Virus Genes **30:**23-29.

179. **Phuphuakrat, A., and P. Auewarakul.** 2003. Heterogeneity of HIV-1 Rev response element. AIDS Res Hum Retroviruses **19:**569-574.

180. **Phuphuakrat, A., and P. Auewarakul.** 2003. Heterogeneity of HIV-1 Rev response element. AIDS Res. Hum. Retroviruses **19:**569-574.

181. **Phuphuakrat, A., R. M. Paris, S. Nittayaphan, S. Louisirirotchanakul, and P. Auewarakul.** 2005. Functional variation of HIV-1 Rev Response Element in a longitudinally studied cohort. J. Med. Virol. **75:**367-373.

182. **Phuphuakrat, A., R. M. Paris, S. Nittayaphan, S. Louisirirotchanakul, and P. Auewarakul.** 2005. Functional variation of HIV-1 Rev Response Element in a longitudinally studied cohort. J Med Virol **75:**367-373.

183. **Pilkington, G. R., K. J. Purzycka, J. Bear, S. F. Le Grice, and B. K. Felber.** 2015. Gammaretrovirus mRNA expression is mediated by a novel, bipartite post-transcriptional regulatory element. Nucleic Acids Res. **42:**11092-11106.

184. **Pollard, V. W., and M. H. Malim.** 1998. The HIV-1 Rev protein. Annu Rev Microbiol **52:**491-532.

185. **Pond, S. J., W. K. Ridgeway, R. Robertson, J. Wang, and D. P. Millar.** 2009. HIV-1 Rev protein assembles on viral RNA one molecule at a time. Proc Natl Acad Sci U S A **106:**1404-1408.

186. **Pond, S. J., W. K. Ridgeway, R. Robertson, J. Wang, and D. P. Millar.** 2009. HIV-1 Rev protein assembles on viral RNA one molecule at a time. Proc. Natl. Acad. Sci. U. S. A. **106:**1404-1408.

187. **Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo.** 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science **224:**497-500.

188. **Purcell, D. F., and M. A. Martin.** 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J. Virol. **67:**6365-6378.

189. **Rabbi, M. F., M. Saifuddin, D. S. Gu, M. F. Kagnoff, and K. A. Roebuck.** 1997. U5 region of the human immunodeficiency virus type 1 long terminal repeat contains TRE-like cAMP-responsive elements that bind both AP-1 and CREB/ATF proteins. Virology **233:**235-245.

190. **Rao, Z., A. S. Belyaev, E. Fry, P. Roy, I. M. Jones, and D. I. Stuart.** 1995. Crystal structure of SIV matrix antigen and implications for virus assembly. Nature **378:**743-747.

191. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, and et al.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature **313:**277-284.

192. **Reed, L. J., Muench,H.** 1983. A simple method of estimating fifty percent endpoints. American Journal of Hygiene **27:**493-497.

193. **Rey, M. A., B. Spire, D. Dormont, F. Barre-Sinoussi, L. Montagnier, and J. C. Chermann.** 1984. Characterization of the RNA dependent DNA polymerase of a new human T-lymphotropic retrovirus (lymphadenopathy associated virus). Biochem. Biophys. Res. Commun. **121:**126-133.

194. **Riviere, L., J. L. Darlix, and A. Cimarelli.** 2010. Analysis of the viral elements required in the nuclear import of HIV-1 DNA. J. Virol. **84:**729-739.

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

196. **Rizzuto, C. D., R. Wyatt, N. Hernandez-Ramos, Y. Sun, P. D. Kwong, W. A. Hendrickson, and J. Sodroski.** 1998. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. Science **280:**1949-1953.

197. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber.** 2000. HIV-1 nomenclature proposal. Science **288:**55-56.

198. **Roebuck, K. A., D. A. Brenner, and M. F. Kagnoff.** 1993. Identification of c-fos-responsive elements downstream of TAR in the long terminal repeat of human immunodeficiency virus type-1. J. Clin. Invest. **92:**1336-1348.

199. **Rohr, O., C. Marban, D. Aunis, and E. Schaeffer.** 2003. Regulation of HIV-1 gene transcription: from lymphocytes to microglial cells. J. Leukoc. Biol. **74:**736-749.

200. **Romanchikova, N., V. Ivanova, C. Scheller, E. Jankevics, C. Jassoy, and E. Serfling.** 2003. NFAT transcription factors control HIV-1 expression through a binding site downstream of TAR region. Immunobiology **208:**361-365.

201. **Rosen, C. A., E. Terwilliger, A. Dayton, J. G. Sodroski, and W. A. Haseltine.** 1988. Intragenic cis-acting art gene-responsive sequences of the human immunodeficiency virus. Proc. Natl. Acad. Sci. U. S. A. **85:**2071-2075.

202. **Roy, S., U. Delling, C. H. Chen, C. A. Rosen, and N. Sonenberg.** 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes Dev. **4:**1365-1373.

203. **Ruelas, D. S., and W. C. Greene.** 2013. An integrated overview of HIV-1 latency. Cell **155:**519-529.

204. **Saad, J. S., E. Loeliger, P. Luncsford, M. Liriano, J. Tai, A. Kim, J. Miller, A. Joshi, E. O. Freed, and M. F. Summers.** 2007. Point mutations in the HIV-1 matrix protein turn off the myristyl switch. J. Mol. Biol. **366:**574-585.

205. **Saad, J. S., J. Miller, J. Tai, A. Kim, R. H. Ghanam, and M. F. Summers.** 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc. Natl. Acad. Sci. U. S. A. **103:**11364-11369.

206. **Sakuma, T., J. I. Davila, J. A. Malcolm, J. P. Kocher, J. M. Tonne, and Y. Ikeda.** 2014. Murine leukemia virus uses NXF1 for nuclear export of spliced and unspliced viral transcripts. J. Virol. **88:**4069-4082.

207. **Sarngadharan, M. G., A. L. DeVico, L. Bruch, J. Schupbach, and R. C. Gallo.** 1984. HTLV-III: the etiologic agent of AIDS. Princess Takamatsu Symp. **15:**301-308.

208. **Schroder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:**521-529.

209. **Schwartz, S., B. K. Felber, D. M. Benko, E. M. Fenyo, and G. N. Pavlakis.** 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J. Virol. **64:**2519-2529.

210. **Schwartz, S., B. K. Felber, and G. N. Pavlakis.** 1991. Expression of human immunodeficiency virus type 1 vif and vpr mRNAs is Rev-dependent and regulated by splicing. Virology **183:**677-686.

211. **Seiki, M., J. Inoue, M. Hidaka, and M. Yoshida.** 1988. Two cis-acting elements responsible for posttranscriptional trans-regulation of gene expression of human T-cell leukemia virus type I. Proc. Natl. Acad. Sci. U. S. A. **85:**7124-7128.

212. **Sharma, Y., U. Neogi, V. Sood, S. Banerjee, S. Samrat, A. Wanchu, S. Singh, and A. C. Banerjea.** 2010. Genetic and functional analysis of HIV-1 Rev Responsive Element (RRE) sequences from North-India. AIDS Res Ther **7:**28.

213. **Shaw, G. M., B. H. Hahn, S. K. Arya, J. E. Groopman, R. C. Gallo, and F. Wong-Staal.** 1984. Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science **226:**1165-1171.

214. **Sloan, E. A., M. F. Kearney, L. R. Gray, K. Anastos, E. S. Daar, J. Margolick, F. Maldarelli, M. L. Hammarskjold, and D. Rekosh.** 2013. Limited nucleotide changes in the Rev response element (RRE) during HIV-1 infection alter overall Rev-RRE activity and Rev multimerization. J. Virol. **87:**11173-11186.

215. **Smith, S. D., M. Shatsky, P. S. Cohen, R. Warnke, M. P. Link, and B. E. Glader.** 1984. Monoclonal antibody and enzymatic profiles of human malignant T-lymphoid cells and derived cell lines. Cancer Res. **44:**5657-5660.

216. **Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine.** 1986. A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature **321:**412-417.

217. **Srinivasakumar, N., N. Chazal, C. Helga-Maria, S. Prasad, M. L. Hammarskjold, and D. Rekosh.** 1997. The effect of viral regulatory

protein expression on gene delivery by human immunodeficiency virus type 1 vectors produced in stable packaging cell lines. J. Virol. **71:**5841-5848.

218. **Srinivasakumar, N., M. L. Hammarskjold, and D. Rekosh.** 1995. Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that affect particle formation and Gag-Pol precursor incorporation. J. Virol. **69:**6106-6114.

219. **Starcich, B., L. Ratner, S. F. Josephs, T. Okamoto, R. C. Gallo, and F. Wong-Staal.** 1985. Characterization of long terminal repeat sequences of HTLV-III. Science **227:**538-540.

220. **Strack, B., A. Calistri, S. Craig, E. Popova, and H. G. Gottlinger.** 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell **114:**689-699.

221. **Suhasini, M., and T. R. Reddy.** 2009. Cellular proteins and HIV-1 Rev function. Curr HIV Res **7:**91-100.

222. **Sundquist, W. I., and H. G. Krausslich.** 2012. HIV-1 assembly, budding, and maturation. Cold Spring Harb. Perspect. Med. **2:**a006924.

223. **Svicher, V., C. Alteri, R. D'Arrigo, A. Lagana, M. Trignetti, S. Lo Caputo, A. P. Callegaro, F. Maggiolo, F. Mazzotta, A. Ferro, S. Dimonte, S. Aquaro, G. di Perri, S. Bonora, C. Tommasi, M. P. Trotta, P. Narciso, A. Antinori, C. F. Perno, and F. Ceccherini-Silberstein.** 2009. Treatment with the fusion inhibitor enfuvirtide influences the appearance of mutations in the human immunodeficiency virus type 1 regulatory protein rev. Antimicrob Agents Chemother **53:**2816-2823.

224. **Svicher, V., C. Alteri, R. D'Arrigo, A. Lagana, M. Trignetti, S. Lo Caputo, A. P. Callegaro, F. Maggiolo, F. Mazzotta, A. Ferro, S. Dimonte, S. Aquaro, G. di Perri, S. Bonora, C. Tommasi, M. P. Trotta, P. Narciso, A. Antinori, C. F. Perno, and F. Ceccherini-Silberstein.** 2009. Treatment with the fusion inhibitor enfuvirtide influences the appearance of mutations in the human immunodeficiency virus type 1 regulatory protein rev. Antimicrob. Agents Chemother. **53:**2816-2823.

225. **Tan, R., L. Chen, J. A. Buettner, D. Hudson, and A. D. Frankel.** 1993. RNA recognition by an isolated alpha helix. Cell **73:**1031-1040.

226. **Tebit, D. M., and E. J. Arts.** 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect. Dis. **11:**45-56.

227. **Tiley, L. S., M. H. Malim, H. K. Tewary, P. G. Stockley, and B. R. Cullen.** 1992. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. Proc. Natl. Acad. Sci. U. S. A. **89:**758-762.

228. **Tovanabutra, S., M. de Souza, N. Sittisombut, S. Sriplienchan, V. Ketsararat, D. L. Birx, C. Khamboonrueng, K. E. Nelson, F. E. McCutchan, and M. L. Robb.** 2007. HIV-1 genetic diversity and compartmentalization in mother/infant pairs infected with CRF01_AE. Aids **21:**1050-1053.

229. **Turner, B. G., and M. F. Summers.** 1999. Structural biology of HIV. J. Mol. Biol. **285:**1-32.

230. **Vallone, B., A. E. Miele, P. Vecchini, E. Chiancone, and M. Brunori.** 1998. Free energy of burying hydrophobic residues in the interface between protein subunits. Proc. Natl. Acad. Sci. U. S. A. **95:**6103-6107.

231. **van Baalen, C. A., O. Pontesilli, R. C. Huisman, A. M. Geretti, M. R. Klein, F. de Wolf, F. Miedema, R. A. Gruters, and A. D. Osterhaus.** 1997. Human immunodeficiency virus type 1 Rev- and Tat-specific cytotoxic T lymphocyte frequencies inversely correlate with rapid progression to AIDS. J Gen Virol **78 (Pt 8):**1913-1918.

232. **Van Baalen, C. A., M. Schutten, R. C. Huisman, P. H. Boers, R. A. Gruters, and A. D. Osterhaus.** 1998. Kinetics of antiviral activity by human immunodeficiency virus type 1-specific cytotoxic T lymphocytes (CTL) and rapid selection of CTL escape virus in vitro. J Virol **72:**6851-6857.

233. **Van Damme, N., D. Goff, C. Katsura, R. L. Jorgenson, R. Mitchell, M. C. Johnson, E. B. Stephens, and J. Guatelli.** 2008. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell host & microbe **3:**245-252.

234. **Van Ryk, D. I., and S. Venkatesan.** 1999. Real-time kinetics of HIV-1 Rev-Rev response element interactions. Definition of minimal binding sites on RNA and protein and stoichiometric analysis. J. Biol. Chem. **274:**17452-17463.

235. **Vercruysse, T., S. Pawar, W. De Borggraeve, E. Pardon, G. N. Pavlakis, C. Pannecouque, J. Steyaert, J. Balzarini, and D. Daelemans.** 2011. Measuring cooperative Rev protein-protein interactions on Rev responsive RNA by fluorescence resonance energy transfer. RNA Biol. **8:**316-324.

236. **Vogt, V. M.** 1997. retroviral virions and genomes. cold spring harbor press.

237. **Wain-Hobson, S.** 1989. HIV genome variability in vivo. AIDS **3 Suppl 1:**S13-18.

238. **Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess, Jr., R. Swanstrom, C. L. Burch, and K. M. Weeks.** 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:**711-716.

239. **Weeks, K. M., and D. M. Crothers.** 1993. Major groove accessibility of RNA. Science **261:**1574-1577.

240. **Weeks, K. M., and D. M. Crothers.** 1991. RNA recognition by Tat-derived peptides: interaction in the major groove? Cell **66:**577-588.

241. **Wen, W., J. L. Meinkoth, R. Y. Tsien, and S. S. Taylor.** 1995. Identification of a signal for rapid export of proteins from the nucleus. Cell **82:**463-473.

242. **Wilkinson, K. A., E. J. Merino, and K. M. Weeks.** 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. **1:**1610-1616.

243. **Wilkinson, K. A., S. M. Vasa, K. E. Deigan, S. A. Mortimer, M. C. Giddings, and K. M. Weeks.** 2009. Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. RNA **15:**1314-1321.

244. **Wingfield, P. T., S. J. Stahl, M. A. Payton, S. Venkatesan, M. Misra, and A. C. Steven.** 1991. HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. Biochemistry **30:**7527-7534.

245. **Wu, Y., and J. W. Marsh.** 2003. Gene transcription in HIV infection. Microbes and infection / Institut Pasteur **5:**1023-1027.

246. **Yang, J., and B. R. Cullen.** 1999. Structural and functional analysis of the avian leukemia virus constitutive transport element. RNA **5:**1645-1655.

247. **Yedavalli, V. S., C. Neuveut, Y. H. Chi, L. Kleiman, and K. T. Jeang.** 2004. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. Cell **119:**381-392.

248. **Yoder, K. E., and F. D. Bushman.** 2000. Repair of gaps in retroviral DNA integration intermediates. J. Virol. **74:**11191-11200.

249. **Zapp, M. L., T. J. Hope, T. G. Parslow, and M. R. Green.** 1991. Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. Proc. Natl. Acad. Sci. U. S. A. **88:**7734-7738.

250. **Zemmel, R. W., A. C. Kelley, J. Karn, and P. J. Butler.** 1996. Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. J. Mol. Biol. **258:**763-777.

251. **Zhang, G., M. L. Zapp, G. Yan, and M. R. Green.** 1996. Localization of HIV-1 RNA in mammalian nuclei. J. Cell Biol. **135:**9-18.

252. **Zhang, M. J., and A. I. Dayton.** 1996. Two secondary structures for the RRE of HIV-1? J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. **13:**403-407.

253. **Zhou, W., L. J. Parent, J. W. Wills, and M. D. Resh.** 1994. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. J. Virol. **68:**2556-2569.

254. **Zhou, W., and M. D. Resh.** 1996. Differential membrane binding of the human immunodeficiency virus type 1 matrix protein. J. Virol. **70:**8540-8548.

255. **Zhu, P., J. Liu, J. Bess, Jr., E. Chertova, J. D. Lifson, H. Grise, G. A. Ofek, K. A. Taylor, and K. H. Roux.** 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature **441:**847-852.